



UNIVERSIDADE DE LISBOA
FACULDADE DE MOTRICIDADE HUMANA



**A MEDIDA E AS ESCALAS DE AVALIAÇÃO DA SAÚDE
DAS POPULAÇÕES NEONATAIS E PEDIÁTRICAS
Estudo de Simulação de Monte Carlo com Variáveis Ordinais**

Dissertação Elaborada com vista à obtenção do Grau de Doutor no Ramo de Motricidade Humana na Especialidade de Comportamento Motor

Orientador: Professor Doutor João Manuel Pardal Barreiros

Coorientadores: Professora Doutora Júlia Maria Vitorino Teles

Professor Doutor João Paulo Marôco Domingos

Júri:

Presidente: Reitor da Universidade de Lisboa

Vogais: Doutor Carlos Alberto Ferreira Neto
Professor Catedrático
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor João Manuel Pardal Barreiros
Professor Catedrático
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Pedro Augusto de Melo Lopes Ferreira
Professor Associado com Agregação
Faculdade de Economia da Universidade de Coimbra

Doutora Maria Adriana Pereira Henriques
Professora Coordenadora
Escola Superior de Enfermagem de Lisboa

Doutora Ana Isabel Andrade Diniz Carita
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutora Júlia Maria Vitorino Teles
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Maria Alice dos Santos Curado

2016



UNIVERSIDADE DE LISBOA
FACULDADE DE MOTRICIDADE HUMANA



**A MEDIDA E AS ESCALAS DE AVALIAÇÃO DA SAÚDE
DAS POPULAÇÕES NEONATAIS E PEDIÁTRICAS
Estudo de Simulação de Monte Carlo com Variáveis Ordinais**

Dissertação Elaborada com vista à obtenção do Grau de Doutor no Ramo de Motricidade Humana na Especialidade de Comportamento Motor

Orientador: Professor Doutor João Manuel Pardal Barreiros

Coorientadores: Professora Doutora Júlia Maria Vitorino Teles

Professor Doutor João Paulo Marôco Domingos

Júri:

Presidente: Reitor da Universidade de Lisboa

Vogais: Doutor Carlos Alberto Ferreira Neto
Professor Catedrático
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor João Manuel Pardal Barreiros
Professor Catedrático
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Pedro Augusto de Melo Lopes Ferreira
Professor Associado com Agregação
Faculdade de Economia da Universidade de Coimbra

Doutora Maria Adriana Pereira Henriques
Professora Coordenadora
Escola Superior de Enfermagem de Lisboa

Doutora Ana Isabel Andrade Diniz Carita
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutora Júlia Maria Vitorino Teles
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Maria Alice dos Santos Curado

2016

**À Minha Família e aos Amigos
Pelo Apoio em Todos os Momentos**

AGRADECIMENTOS

O percurso de elaboração de dissertação de doutoramento é um processo muito solitário, dizem os investigadores mais experientes. Apesar de saber desta máxima, nunca pensei passar por tamanha solidão! Quero por isso deixar os meus sinceros agradecimentos a todos que de alguma forma me acompanharam e contribuíram com a sua competência científica para que eu conseguisse ultrapassar as minhas dificuldades e conseguisse chegar ao fim desta primeira fase do trabalho.

Aos meus orientadores, Professores Doutores João Barreiros, Júlia Teles e João Marôco, agradeço pelo acompanhamento a apoio no desenvolvimento deste trabalho. Um especial agradecimento à Professora Júlia, pela disponibilidade e pelo rigor com que acompanhou o meu trabalho e ao Professor João Marôco, pois sem o seu apoio teria sido muito difícil chegar ao fim. Ao João Barreiros por estar ao meu lado nos momentos mais difíceis.

À Direção da Escola Superior de Enfermagem de Lisboa (ESEL) agradeço o apoio que me deu, sobretudo na disponibilização da equipa de informática e da sala de computadores para eu trabalhar durante as simulações, sem esta ajuda teria sido impossível terminar o estudo de simulação.

Agradeço à equipa de informática da ESEL, Diogo, Mário, António, Armando e André. Um agradecimento especial ao Rui Palmeirinha da Reitoria da Universidade de Lisboa pela ideia fabulosa dos computadores virtuais, que resolveram muitos dos meus problemas durante as longas horas de simulação.

Aos funcionários da ESEL que dão apoio à receção da escola e da residência e me apoiaram durante dezassete meses, quando ia verificar as simulações, fosse aos fins de semanas ou durante o período de férias, bem hajam pela disponibilidade e carinho.

Às Professoras Isabel Fraga Alves, Antónia Turkman e Marília Antunes por terem permitido que assistisse às aulas das disciplinas que coordenam na Faculdade de Ciências da UL.

Às colegas que me acompanharam a jornada dos dezassete meses de simulação e não sabendo o que fazer para me ajudar, me ouviam e alimentavam fartamente com fruta da quinta, bolachinhas muito boas e chocolates.

Agradeço às equipas de pesquisa que amavelmente me disponibilizaram os dados das suas investigações para eu poder fazer as aplicações práticas do meu trabalho.

Quero deixar um agradecimento muito especial à minha FAMÍLIA. Aos meus filhos João e Joel e à família que eles construíram e acrescentaram à minha, famílias com AMOR. Que mulher feliz que eu sou por vos ter, pelos homens que são, pelo carinho com que me tratam. A alegria de viver, que transmitem, faz-me sentir que a vida tem valido a pena, apesar das pedras no caminho. À Patrícia e à Sofia agradeço por em conjunto convosco me darem bens tão preciosos, como os mais queridos netinhos do mundo, Gui, Magui, Kiko e Alex e a outros bebés que possam chegar. São eles

que me fazem viver momentos maravilhosos seja quando estamos juntinhos ou através daquela solução maravilhosa, que inventaram para minimizar as distâncias, chamada *skype*.

Agradeço aos meus Pais pela educação que me deram, pelas orientações para a vida e por me fazerem sempre acreditar que sou capaz...Que chego lá! Apesar de não estarem...Eu sinto que estão muito perto! Às minhas manas, Rosa e Lourdes por me tratarem sempre como a menina, mesmo na bela idade em que me encontro, e se preocuparem comigo como com os seus próprios filhos! Aos meus sobrinhos por me ouvirem sempre que preciso e desculparem as minhas ausências.

Peço perdão pela falta de tempo para estar mais presente, por ter estado centrada tempo demais neste projeto e me alhear completamente do mundo. Mas eu sei que já me perdoaram, pois, família é isso é perdoar tudo, é amar incondicionalmente, é estar lá quando é preciso. Família, nós não escolhemos, mas se eu escolhesse não escolheria melhor, era esta que eu queria!

Quero agradecer aos meus amigos, Nair, Carlos, Ana, Soraia, Ana Cristina, Júlio, Belita, Lurdes, Maria Isabel, Thereza, Cláudia, Dulce, Jorge, Xanoco, Emília, John, Helena, Isabel, Cristina, Graça, Marília, pelo apoio ao longo deste percurso tortuoso. Agradeço as discussões, as reflexões, as construções e desconstruções, as correções, o saber ouvir, o secar as lágrimas. Como sou feliz por vos ter como AMIGOS.

Como não sou muito criativa, vou usar as palavras de Oscar Wilde que diz muitas coisas com as quais me identifico, e vou misturar as palavras dele com as minhas para vos dizer o quanto prezo a vossa amizade. A escolha dos amigos dizia ele, deve ser feita pela pupila e, esta tem que ter uma tonalidade perturbadora e um brilho especial. A escolha dos amigos deve ser feita pela alma lavada e exposta, não devemos querer deles só o ombro, devemos querer também o riso, a alegria, a felicidade de rir em conjunto; pois só os amigos que conseguem rir juntos conseguem sofrer da mesma forma e, apoiar-se nos bons e nos maus momentos. Eu concordo com ele quando diz que os amigos têm que ser tontos e sérios, realistas, mas, sem deixarem de lutar pelos seus sonhos e fantasias. É assim que vos vejo e é assim que eu gosto de todos vós.

Os amigos têm que ser loucos, com as suas dúvidas e angústias e tolerantes com o pior do seu amigo e também têm que ser santos para não duvidarem das diferenças e pedirem perdão pelas injustiças. Os amigos têm que ser crianças para não esquecerem a importância do vento a tocar o rosto e velhos para não terem pressa. Por tudo isto, ter AMIGOS é um privilégio, pois através deles conhecemo-nos melhor e identificamos quem somos. Isto é o que acontece comigo! Tal como com Oscar Wilde, quando vos observo e, vejo uns mais loucos, outros mais santos, outros mais sérios, outros mais tontos, outros mais crianças e outros mais velhos, sinto e digo, são os meus amigos! Tudo isto me faz muito feliz e, por isso, me revejo na frase dele, quando diz que “a normalidade é uma ilusão néscia e estéril”.

Comecei por dizer que o processo de desenvolvimento deste trabalho tinha sido muito solitário. Talvez tenha exagerado... Não foi tanto assim! O vosso apoio e carinho compensaram a minha solidão!

Índice

AGRADECIMENTOS.....	vi
Lista de Figuras	xi
Lista de Tabelas.....	xvii
Lista de Quadros.....	xx
RESUMOS.....	1
Resumo.....	2
Abstract.....	5
INTRODUÇÃO GERAL.....	10
Capítulo I. Apresentação Geral do Trabalho.....	10
1.1 A Escolha da Área Temática da Investigação.....	10
1.2 Contextualização do Estudo	13
1.3 Justificação do Problema	19
1.4 Questões e Objetivos da Investigação.....	21
1.5 Organização do Trabalho.....	26
1.6 Referências Bibliográficas.....	27
PARTE I. Contextualização Teórica.....	33
Introdução.....	34
Capítulo II. Investigação e Avaliação da Saúde da População Neonatal e Pediátrica.....	35
2.1 Escalas de Avaliação do Desenvolvimento e da Saúde da Criança e Jovem	39
2.2 Da Investigação aos Cuidados de Saúde Baseados na Evidência	54
2.3 Referências Bibliográficas.....	60
Capítulo III. A Medida e as Técnicas Estatísticas.....	68
3.1 O Conceito de Medida e as Escalas de Medida	68
3.2 A Análise da Variância Multivariada	74
3.2.1 MANOVA Paramétrica a Um Fator.....	78
3.2.2 MANOVA não Paramétrica a um Fator.....	82
3.3 Referências Bibliográficas.....	86
Capítulo IV. A Simulação e a Geração Aleatória de dados.....	90
4.1 O Método de Monte Carlo	90
4.2 A Geração Aleatória de Variáveis Ordinais.....	92
4.3 Referências Bibliográficas.....	94
PARTE II. Aplicações da Teoria.....	95
Introdução.....	96
Capítulo V. Análise Estatística de Escalas Ordinais: Aplicações na área da Saúde Infantil e Pediatria.....	97
Resumo	97
5.1 Introdução	98
5.2 Metodologia	102
5.2.1 Estratégias de Pesquisa	103
5.3 Resultados	103
5.4 Conclusões	108
5.5 Referências Bibliográficas.....	109
Capítulo VI. Análise de Variáveis Não Diretamente Observáveis: Influência na Tomada de Decisão Durante o Processo de Investigação.....	112
Resumo	112
6.1 Introdução	113

6.2 As Variáveis e a Medida.....	115
6.3 Dimensão da Amostra	121
6.4 Considerações Finais	123
6.6 Referências Bibliográficas.....	125
PARTE III. Estudo de Simulação de Monte Carlo.....	127
Introdução.....	128
Capítulo VII. MANOVA Paramétrica e não Paramétrica com Variáveis Ordinais: A Probabilidade de Erro de Tipo I e a Potência de Teste.....	128
Resumo	128
7.1 Introdução	131
7.2 Metodologia	136
7.3 Apresentação dos Resultados	141
7.3.1 Probabilidade de Erro de Tipo I.....	141
7.3.2 Potência dos Testes	157
7.4 Discussão dos Resultados	190
7.5 Conclusões	196
7.6 Referências Bibliográficas.....	199
PARTE IV. Aplicações Práticas.....	203
Introdução	204
Capítulo VIII. Escalas de Avaliação da criança e jovem.....	206
8.1 Metodologia	208
8.2 Avaliação das Competências na Alimentação Oral do Neonato	210
8.2.1 Escala de Observação das Competências Precoces na Alimentação Oral ..	215
8.2.2 Procedimentos.....	218
8.2.3 Apresentação dos Resultados	218
8.2.4 Discussão dos Resultados	223
8.2.5 Considerações Finais	224
8.3 Avaliação do Risco de Lesão da Pele em Neonatos.....	225
8.3.1 Escala de Observação do Risco de Lesão da Pele.....	228
8.3.2 Procedimentos.....	230
8.3.3 Apresentação dos Resultados	231
8.3.4 Discussão dos Resultados	235
8.3.5 Considerações Finais	236
8.4 Avaliação da Independência Funcional da Criança/Jovem com Espinha Bífida .	236
8.4.1 Medida de Independência Funcional	241
8.4.2 Procedimentos.....	244
8.4.3 Apresentação dos Resultados	244
8.4.4 Discussão dos Resultados	252
8.4.5 Considerações Finais	253
8.5 Conclusões	254
8.6 Referências Bibliográficas.....	255
PARTE V: Discussão Geral.....	263
Introdução.....	264
Capítulo IX. Discussão Geral dos Resultados.....	265
9.1 Implicações Teóricas.....	265
9.2 Implicações Práticas.....	268
9.3 Referências Bibliográficas.....	270

Apêndice I: Programas em R.....	275
Apêndice II: Tabelas com resultados das médias da Assimetria e Curtose.....	287
Apêndice III: Tabelas com resultados das médias das estatísticas de Kruskal-Wallis...	325
Anexo I: Pedidos de Autorização.....	331

Lista de Figuras

Figura 2. 1 - Modelo Conceptual de Cuidados de Saúde Baseados na Evidência.	55
Figura 2. 2 - Produção de Evidência e Cuidados de Saúde.	56
Figura 5. 1 - Número (Nº) de Artigos por escala de avaliação, nos 127 artigos, que cumpriam todos os critérios de inclusão.....	104
Figura 7. 1 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada no ponto central da escala (distribuições iguais (p_1,p_1,p_1) , ou seja H_0 verdadeira).....	143
Figura 7.2 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada no extremo inferior da escala (com distribuições iguais (p_2,p_2,p_2) , ou seja H_0 verdadeira).	144
Figura 7.3 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada no	

extremo superior da escala (com distribuições iguais (p_3, p_3, p_3) , ou seja H_0 verdadeira).

..... 145

Figura 7.4 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens igualmente distribuída na escala (com distribuições iguais (p_4, p_4, p_4) , ou seja H_0 verdadeira). 146

Figura 7.5 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada uniformemente e simétrica na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_4, p_4, p_1) ou seja H_0 falsa). 158

Figura 7.6 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100, 30, 60, 90, 120, 240, e 300$ com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa magnitude ($r=0.10$), média ($r=0.40$) e elevada ($r=0.70$) e com a maior probabilidade de resposta aos itens com localização simétrica e uniforme na escala (distribuições iguais em dois grupos e diferentes do terceiro grupo (p_1, p_1, p_4) , ou seja H_0 falsa). 159

Figura 7.7 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com um formato de medida com 5 pontos (Likert 5

pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens com localização simétrica e uniforme na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_1, p_1, p_4), ou seja H_0 falsa). 159

Figura 7.8 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada uniformemente e simétrica na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_4, p_4, p_1) ou seja H_0 falsa). 159

Figura 7.9 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 e 4 pontos (Likert 3 e tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada no ponto central e nos extremos da escala de medida (com distribuições diferentes em cada um dos três grupos (p_1, p_2, p_3), ou seja H_0 falsa). 162

Figura 7.10 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 e 7 pontos (Likert 5 e 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada no ponto central e nos extremos da escala de medida (com distribuições diferentes em cada um dos três grupos (p_1, p_2, p_3), ou seja H_0 falsa). 163

Figura 7.11 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e

nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 e 5 pontos (Likert 3 e 5 pontos), com correlações de baixa ($r=0.10$ média), ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada nos extremos da escala de medida (com distribuições iguais para dois grupos e diferentes para o terceiro grupo (p_2, p_2, p_3), ou seja H_0 falsa)..... 165

Figura 7.12 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos da escala de medida (com distribuições iguais para dois grupos e diferentes para o terceiro grupo (p_2, p_2, p_3), ou seja H_0 falsa)..... 166

Figura 7.13 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa)..... 167

Figura 7.14 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 pontos (Likert 5 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e

uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa). 167

Figura 7.15 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa). 169

Figura 7.16 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens-Pillai NP, lambda de Wilks aplicado às ordens-Wilks NP e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no ponto central e no extremo inferior na escala de medida (com distribuições iguais em dois dos grupos e diferentes no terceiro grupo, (p_1, p_1, p_2), ou seja H_0 falsa). 171

Figura 7.17 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$), e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no ponto central e uniformemente distribuída na escala de medida (com distribuições de frequência diferentes em cada um dos três grupos, (p_1, p_2, p_4), ou seja H_0 falsa). 176

Figura 7.18 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks-W, traço de Hotelling-H, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão

$n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 pontos (Likert 5 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$), e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no extremo inferior e no ponto central da escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_2, p_2, p_1) , ou seja H_0 falsa). 181

Figura 7.19 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada uniformemente e no extremo inferior da escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_4, p_4, p_2) , ou seja H_0 falsa). 185

Figura 7.20 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), com maior probabilidade de resposta aos itens localizada no extremo inferior e uniformemente distribuída na escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_2, p_2, p_4) , ou seja H_0 falsa). 189

Lista de Tabelas

Tabela 3. 1 - MANOVA para comparação de vetores médios (Notação matricial).....	81
Tabela 5. 1 - Artigos científicos analisados por tipo de escala de avaliação/psicométrica	105
Tabela 7. 1 - Probabilidades marginais (<i>p.m.</i>) por formato de medida da escala ordinal usadas no estudo de simulação.	139
Tabela 7. 2 - Diferença máxima entre as correlações de <i>input</i> e as correlações amostrais.	140
Tabela 7.3 - Condições testadas nas diferentes simulações (432 cenários) com quatro tipos de probabilidades marginais (<i>p</i>), seis dimensões amostrais (<i>n</i>), correlações com três tipos de magnitude (<i>r</i>), e escalas de avaliação com quatro formatos de medida (3, 4, 5, e 7 pontos).	141
Tabela 7.4 - Condições testadas nas diferentes simulações (216 cenários) com dois tipos de probabilidades marginais (<i>p</i>), seis dimensões amostrais (<i>n</i>), correlações com três tipos de magnitude (<i>r</i>), e escalas de avaliação com quatro formatos de medida (3, 4, 5, e 7 pontos).	158
Tabela 7. 5 – Condições testadas nas diferentes simulações (324 cenários) com três tipos de probabilidades marginais (<i>p</i>), seis dimensões amostrais (<i>n_i</i>), correlações com três tipos de magnitude (<i>r</i>), e quatro escalas de avaliação com um formato de medida com 3, 4, 5, e 7 pontos.....	160
Tabela 7. 6 - Correlações entre o valor observado da estatística de teste de Kruskal- Wallis média, das três variáveis, nas mil réplicas, para <i>n</i> =30, 60, 120, 300, nas escalas com 3 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica e respectivo <i>p</i> -value.	194
Tabela 7. 7- Correlações entre o valor observado da estatística de teste de Kruskal- Wallis média, das três variáveis, nas mil réplicas, para <i>n</i> =30, 60, 120, 300, nas escalas com 4 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica e respectivo <i>p</i> -value.	195

Tabela 7. 8 - Correlações entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para $n=30, 60, 120, 300$, nas escalas com 5 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica e respetivo p-value.	195
Tabela 7. 9 - Correlações entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para $n=30, 60, 120, 300$, nas escalas com 7 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica e respetivo p-value.	196
Tabela 8.2. 1 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2), nos itens 4, 5, e 11 da EFS para $n_i=10, 20$ e 30	220
Tabela 8.2. 2 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) com um erro padrão de assimetria de 0.241, nos itens 4, 5, e 11 da EFS para $n_i=40, 80$ e 100	220
Tabela 8.2.3 - Traço de Pillai nas diferentes dimensões amostrais na escala EFS	221
Tabela 8.2.4 - Lambda de Wilks nas diferentes dimensões amostrais na escala EFS..	221
Tabela 8.2.5 - Traço de Hotelling nas diferentes dimensões amostrais na escala EFS	221
Tabela 8.2.6 - Maior raiz de Roy nas diferentes dimensões amostrais na escala EFS .	222
Tabela 8.2.7 - Traço de Pillai e Lambda de Wilks aplicado às ordens na escala EFS .	222
Tabela 8.2.8 - p -value das estatísticas da MANOVA paramétrica e não paramétrica .	222
Tabela 8.3. 1 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 1, 3, e 6 da NSRAS para $n_i=10, 20, 30$ e 40	232
Tabela 8.3.2 - Traço de Pillai nas diferentes dimensões amostrais na escala NSRAS.	233
Tabela 8.3.3 - Lambda de Wilks nas diferentes dimensões amostrais na escala NSRAS	233
Tabela 8.3.4 - Traço de Hotelling nas diferentes dimensões amostrais na escala NSRAS	233
Tabela 8.3.5 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala NSRAS	233
Tabela 8.3.6 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens com a NSRAS	234

Tabela 8.3.7 - Valores do <i>p-value</i> das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais na escala NSRAS	234
Tabela 8.4. 1 - Mediana (<i>Me</i>), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 6, 10, e 13 da MIF motora modificada para $n_i=10$ e 20.	246
Tabela 8.4.2 - Traço de Pillai nas diferentes dimensões amostrais na escala.....	246
Tabela 8.4.3 - Lambda de Wilks nas diferentes dimensões amostrais na escala.....	247
Tabela 8.4.4 - Traço de Hotelling nas diferentes dimensões amostrais na escala.....	247
Tabela 8.4.5 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala.....	247
Tabela 8.4.6 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens na escala FIM dimensão motora modificada (itens 6, 10, 13)	247
Tabela 8.4.7 - Valores do <i>p-value</i> das estatísticas da MANOVA paramétrica e não paramétrica nas duas dimensões amostrais na MIF-Motora modificada (itens 6, 10, 13)	248
Tabela 8.4.8 - Mediana (<i>Me</i>), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 1, 2 e 4 da MIF dimensão motora para $n_i=10, 20$ e 30.	249
Tabela 8.4.9 - Traço de Pillai nas diferentes dimensões amostrais na escala MIF Motora (itens 1, 2, 4).....	250
Tabela 8.4.10 - Lambda de Wilks nas diferentes dimensões amostrais na escala MIF Motora (itens 1, 2, 4)	250
Tabela 8.4.11 - Traço de Hotelling nas diferentes dimensões amostrais na escala MIF Motora itens (itens 1, 2, 4).....	250
Tabela 8.4.12 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala MIF Motora itens (itens 1, 2, 4).....	251
Tabela 8.4.13 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens na escala MIF dimensão motora itens (itens 1, 2, 4).....	251
Tabela 8.4.14 - Valores do <i>p-value</i> das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais na escala MIF motora (itens 1, 2, 4)	251

Lista de Quadros

Quadro 2. 1 - Características principais de algumas escalas de avaliação do desenvolvimento da Criança e Jovem.	42
Quadro 2. 2 - Características principais de algumas escalas de avaliação da saúde da Criança e Jovem.....	51
Quadro 8.2. 1 - Descritores da dimensão Capacidade em organizar o funcionamento Oro-motor	216
Quadro 8.2. 2 - Descritores da dimensão Capacidade em Coordenar a Deglutição.....	217
Quadro 8.2. 3 - Descritores da dimensão Capacidade em Manter a Estabilidade Fisiológica	217
Quadro 8.3. 1 - Descritores da Escala de Observação do Risco de Lesão da Pele de Neonatos	229
Quadro 8.4. 1- Formato de medida dos itens da escala MIF	242
Quadro 8.4. 2 - Medida de Independência Funcional (MIF).....	243

RESUMOS

Resumo

A investigação na área da saúde e a utilização dos seus resultados tem funcionado como base para a melhoria da qualidade de cuidados, exigindo dos profissionais de saúde conhecimentos na área específica onde desempenham funções, conhecimentos em metodologia de investigação que incluam as técnicas de observação, técnicas de recolha e análise de dados, para mais facilmente serem leitores capacitados dos resultados da investigação. Os profissionais de saúde são observadores privilegiados das respostas humanas à saúde e à doença, podendo contribuir para o desenvolvimento e bem-estar dos indivíduos muitas vezes em situações de grande vulnerabilidade. Em saúde infantil e pediatria o enfoque está nos cuidados centrados na família privilegiando-se o desenvolvimento harmonioso da criança e jovem, valorizando os resultados mensuráveis em saúde que permitam determinar a eficácia das intervenções e a qualidade de saúde e de vida.

No contexto pediátrico realçamos as práticas baseadas na evidência, a importância atribuída à pesquisa e à aplicação dos resultados da investigação nas práticas clínicas, assim como o desenvolvimento de instrumentos de mensuração padronizados, nomeadamente as escalas de avaliação, de ampla utilização clínica, que facilitam a apreciação e avaliação do desenvolvimento e da saúde das crianças e jovens e resultem em ganhos em saúde. A observação de forma sistematizada das populações neonatais e pediátricas com escalas de avaliação tem vindo a aumentar, o que tem permitido um maior equilíbrio na avaliação das crianças e também uma observação baseada na teoria e nos resultados da investigação. Alguns destes aspetos serviram de base ao desenvolvimento deste trabalho que pretende dar resposta a 3 objetivos fundamentais.

Para dar resposta ao primeiro objetivo, “Identificar na literatura científica, os testes estatísticos mais frequentemente utilizados pelos investigadores da área da saúde infantil e pediatria quando usam escalas de avaliação” foi feita uma revisão sistemática da literatura, que tinha como objetivo analisar artigos científicos cujos instrumentos de recolha de dados fossem escalas de avaliação, na área da saúde da criança e jovem, desenvolvidas com variáveis ordinais, e identificar os testes estatísticos aplicados com estas variáveis. A análise exploratória dos artigos permitiu-nos verificar que os

investigadores utilizam diferentes instrumentos com diferentes formatos de medida ordinal (com 3, 4, 5, 7, 10 pontos) e tanto aplicam testes paramétricos como não paramétricos, ou os dois em simultâneo, com este tipo de variáveis, seja qual for a dimensão da amostra. A descrição da metodologia nem sempre explicita se são cumpridas as suposições dos testes. Os artigos consultados nem sempre fazem referência à distribuição de frequência das variáveis (simetria/assimetria) nem à magnitude das correlações entre os itens. A leitura desta bibliografia serviu de suporte à elaboração de dois artigos, um de revisão sistemática da literatura e outro de reflexão teórica. Apesar de terem sido encontradas algumas respostas às dúvidas com que os investigadores e os profissionais, que trabalham com estes instrumentos, se deparam, verifica-se a necessidade de desenvolver estudos de simulação que confirmem algumas situações reais e alguma teoria já existente, e trabalhem outros aspetos nos quais se possam enquadrar os cenários reais de forma a facilitar a tomada de decisão dos investigadores e clínicos que utilizam escalas de avaliação.

Para dar resposta ao segundo objetivo “Comparar a performance, em termos de potência e probabilidade de erro de tipo I, das 4 estatísticas da MANOVA paramétrica com 2 estatísticas da MANOVA não paramétrica quando se utilizam variáveis ordinais correlacionadas, geradas aleatoriamente”, desenvolvemos um estudo de simulação, através do Método de Monte Carlo, efetuado no Software R. O delineamento do estudo de simulação incluiu um vetor com 3 variáveis dependentes, uma variável independente (fator com três grupos), escalas de avaliação com um formato de medida com 3, 4, 5, e 7 pontos, diferentes probabilidades marginais (p_1 para distribuição simétrica, p_2 para distribuição assimétrica positiva, p_3 para distribuição assimétrica negativa e p_4 para distribuição uniforme) em cada um dos três grupos, correlações de baixa, média e elevada magnitude ($r=0.10$, $r=0.40$, $r=0.70$, respetivamente), e seis dimensões de amostras ($n=30, 60, 90, 120, 240, 300$).

A análise dos resultados permitiu dizer que a maior raiz de Roy foi a estatística que apresentou estimativas de probabilidade de erro de tipo I e de potência de teste mais elevadas. A potência dos testes apresenta comportamentos diferentes, dependendo da distribuição de frequência da resposta aos itens, da magnitude das correlações entre itens, da dimensão da amostra e do formato de medida da escala. Tendo por base a

distribuição de frequência, considerámos três situações distintas: a primeira (com probabilidades marginais p_1, p_1, p_4 e p_4, p_4, p_1) em que as estimativas da potência eram muito baixas, nos diferentes cenários; a segunda situação (com probabilidades marginais p_2, p_3, p_4 ; p_1, p_2, p_3 e p_2, p_2, p_3) em que a magnitude das potências é elevada, nas amostras com dimensão superior ou igual a 60 observações e nas escalas com 3, 4, 5 pontos e potências de magnitude menos elevada nas escalas com 7 pontos, mas com a mesma magnitude nas amostras com dimensão igual a 120 observações, seja qual for o cenário; a terceira situação (com probabilidades marginais p_1, p_1, p_2 ; p_1, p_2, p_4 ; p_2, p_2, p_1 ; p_4, p_4, p_2 e p_2, p_2, p_4) em que quanto maiores, a intensidade das correlações entre itens e o número de pontos da escala, e menor a dimensão das amostras, menor a potência dos testes, sendo o lambda de Wilks aplicado às ordens mais potente do que todas as outras estatísticas da MANOVA, com valores imediatamente a seguir à maior raiz de Roy. No entanto, a magnitude das potências dos testes paramétricos e não paramétricos assemelha-se nas amostras com dimensão superior a 90 observações (com correlações de baixa e média magnitude), entre as variáveis dependentes nas escalas com 3, 4 e 5 pontos; e superiores a 240 observações, para correlações de baixa intensidade, nas escalas com 7 pontos.

No estudo de simulação e tendo por base a distribuição de frequência, concluímos que na primeira situação de simulação e para os diferentes cenários, as potências são de baixa magnitude devido ao facto de a MANOVA não detetar diferenças entre grupos pela sua similaridade. Na segunda situação de simulação e para os diferentes cenários, a magnitude das potências é elevada em todos os cenários cuja dimensão da amostra seja superior a 60 observações, pelo que é possível aplicar testes paramétricos.

Na terceira situação de simulação, e para os diferentes cenários quanto menor a dimensão da amostra e mais elevada a intensidade das correlações e o número de pontos da escala, menor a potência dos testes, sendo a magnitude das potências mais elevadas no teste de Wilks aplicado às ordens, seguido do traço de Pillai aplicado às ordens. No entanto, a magnitude das potências dos testes paramétricos e não paramétricos assemelha-se nas amostras com maior dimensão e correlações de baixa e média magnitude.

Para dar resposta ao terceiro objetivo “Enquadrar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados reais provenientes de escalas de avaliação com um formato de medida com 3, 4, 5 e 7 pontos, nos resultados do estudo de simulação estatística” utilizaram-se dados reais que emergiram da observação de recém-nascidos com a escala de avaliação das competências para a alimentação oral, Early Feeding Skills (EFS), o risco de lesões da pele, com a Neonatal Skin Risk Assessment Scale (NSRAS), e a avaliação da independência funcional em crianças e jovens com espinha bífida, com a *Functional Independence Measure* (FIM). Para fazer a análise destas escalas foram realizadas 4 aplicações práticas que se enquadrassem nos cenários do estudo de simulação. A idade, o peso, e o nível de lesão medular foram as variáveis independentes escolhidas para selecionar os grupos, sendo os recém-nascidos agrupados por “classes de idade gestacional” e por “classes de peso” as crianças e jovens com espinha bífida por “classes etárias” e “níveis de lesão medular”. Verificou-se um bom enquadramento dos resultados com dados reais no estudo de simulação.

Descritores: variáveis ordinais, simulação de Monte Carlo, probabilidade de erro de tipo I, potência, MANOVA paramétrica e não paramétrica, criança, jovem, escalas de avaliação.

Abstract

Research in the healthcare area has worked as a basis for the improvement of caregiving quality, demanding from healthcare professionals more specific knowledge in the area in which they perform their functions, including knowledge in research methodology which involves observation, data collection and analysis, so that they can become competent readers of research outcomes. Healthcare professionals are privileged observers of human responses to health and sickness, and can therefore contribute for the development and well-being of individuals often in situations of great vulnerability. In infant and paediatric health, the focus is on family care, highlighting the harmonious development of the infant and young child, valuing the measurable results in health which allow for the determination of intervention efficiency and quality of health and life. In the paediatric context we enhance the evidence based practices, the

importance of research and the application of results in clinical practice research, as well as the development of standard measurement instruments, including assessment scales, wide clinical use, which can facilitate the observation and evaluation of the development and health of infants and young children and result in health gains. The systematic observation of neonatal and paediatric populations with assessment scales has been rising, which has allowed for a higher balance in children assessment and also for observation based on theory and on research results. Some of these aspects were the basis for the development of this work that aims to respond to three fundamental objectives.

In order to provide an answer to the first objective of this research, “Identifying in scientific literature the most frequent statistical tests used by researchers in the area of paediatrics and infant health whenever evaluation scales are used”, and a systematic review of the literature was first made. It aimed at analysing scientific articles whose data gathering instruments were assessment scales, in the area of paediatric and infant health, developed with ordinal variables, as well as identifying the statistical tests applied to those variables. The exploratory analysis of the articles allowed us to verify that researchers use different instruments with different formats of ordinal measure (3, 4, 5, 6, 7, 9, 10 points) and that they either use parametric or non-parametric tests or even both simultaneously with the ordinal data, regardless of the sample dimension.

The description of the methodology is not always explicit as to whether the test assumptions are met or not. In many reviewed articles the points of scales, the skew, and the magnitude of correlations between items were not mentioned. The reading of this bibliography served as support to the elaboration of two articles, i) a systematic review of literature and ii) a reflection on theoretical concepts. Although some answers to questions that researchers and professionals face were found, there still is the need to develop more simulation studies in order to confirm some real situations and some of the already existing theory as well as to work on other aspects in which we can frame real scenarios so that the decision making process of researchers and practitioners who use these instruments can be facilitated.

In order to provide an answer to the second objective “Comparing the performance, in terms of power and type I error, of the 4 parametric MANOVA tests

(Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root) and 2 non-parametric tests (Pillai's Trace and Wilks' Lambda applied to the rank transformed data), using correlated ordinal variables randomly generated". A Monte Carlo resampling simulation (1000 replications) was carried out to estimate the power and type I error rate from the proportion of rejected null hypothesis at a 5% significance level of parametric and non-parametric one-way MANOVA, with ordinal data. In the data generation process, scales with three, four, five and seven points were considered and a vector with three dependent variables, one independent variable with three levels (groups), different marginal distributions in groups (symmetric, p_1 ; positive skew, p_2 ; negative skew, p_3 ; and uniform, p_4); different correlation structures (low, $r=0.10$; average, $r=0.40$ and high, $r=0.70$), and different sample sizes were accounted for ($n=30, 60, 90, 120, 240, 300$).

The analysis of the simulation results allows us to conclude that Roy's largest root was the statistic that presented the highest values of the probability of type I error, and was also one of the statistics with the biggest statistical power rate. The power presents different behaviours, depending on the distribution, the magnitude of correlation between independent variables (items), the sample size, and the points of the scale. Based on the frequency distribution, the data analysis of the power results allows us to identify three distinct situations. In the first, for the different scenarios (with marginal probabilities p_1, p_1, p_4 e p_4, p_4, p_1) the power is of low magnitude due to the fact that MANOVA does not detect differences between groups due to their similarity. In the second situation (with marginal probabilities p_2, p_3, p_4 ; p_1, p_2, p_3 e p_2, p_2, p_3), the magnitude of the power rate of MANOVA depends on a combination of the correlation between dependent variables the sample size, and the number of scale points, and in different scenarios, the power has a higher magnitude if the sample size, is larger (dimension greater or equal to sixty observation, and scales with 3, 4, 5 points and with 120 observations for scales with 7 points). In the third situation (with marginal probabilities p_1, p_1, p_2 ; p_1, p_2, p_4 ; p_2, p_2, p_1 ; p_4, p_4, p_2 e p_2, p_2, p_4), the magnitude of the power rate of MANOVA depends on a combination of the correlation between dependent variables, the sample size, and the number of scale points. As the correlations and number of scale points increased and the sample size decreased, the power rate of

MANOVA decreased, and Wilks' lambda applied to the rank transformed data had a higher power rate than the other statistics. If the samples size increased (with 90, 120 and 240 observations), and the magnitude decreased (with low, average and high magnitude correlations) between the dependent variables for scales with 3, 4, and 5 points, and for scales with 7 points (samples with 240 observations and average correlations), the magnitude of the power rate of MANOVA is similar in all statistics, parametric and non-parametric.

The data analysis of the power results allows us to conclude that in the first situation, the power is of low magnitude due to the fact that MANOVA does not detect differences between groups. In the second situation, the power has higher magnitude if the sample size is larger; the magnitude of the power rate of MANOVA is similar in all statistics, parametric and non-parametric. In the third situation, as the magnitude of correlations and number of scale points increased and the sample size decrease, the power rate of MANOVA decreased, and the Wilks's Lambda applied to the ranks' transformed data have higher power rate than the other statistics. But the power magnitude is similar in all statistics, parametric and non-parametric if the sample size is larger and the magnitude correlation decreased to low and average magnitude).

To address the third objective "Framing the results of the application of the parametric MANOVA and the non-parametric MANOVA to real data, from assessment scales with three, four, five, and seven points, in the results of the statistical simulation study", real data was used which emerged from the observation of new-borns with the "Early Feeding Skills Assessment Scale", the "Neonatal Skin Risk Assessment Scale" and the "Functional Independence Measure" in children and youth with spine bifida.

In order to carry out the analysis of these scales, three practical applications that fitted the study simulation scenarios are presented. Weight and level of spinal cord injuries were the independent variables chosen for the selection of groups, whereas new-borns were grouped by "weight classes" and children and young people with spina bifida were grouped by "level of spinal cord injuries" and age group. There was a good framework of the results of practical applications in the simulation study.

Descriptors: ordinal data, Monte Carlo simulation, power, type I error probability, parametric and non-parametric MANOVA, child and youth, assessment scales

INTRODUÇÃO GERAL

INTRODUÇÃO GERAL

Esta introdução geral tem como finalidade expor as motivações para a realização deste trabalho de investigação sobre a medida e as escalas de avaliação da saúde das populações neonatais e pediátricas, bem como discutir a sua pertinência em Ciências Sociais e em Ciências da Saúde e, especificamente, em investigação nessas áreas. Procuramos, ainda, apresentar a lógica de todo o processo de investigação desenvolvido, da revisão teórica e do plano metodológico, justificando as diferentes decisões no domínio da investigação de forma a tornar compreensível os princípios e procedimentos que nortearam a pesquisa que passamos a expor.

Capítulo I. Apresentação Geral do Trabalho

No primeiro capítulo apresentamos as razões da escolha da área temática da investigação sobre a medida e as escalas de avaliação, das populações neonatais e pediátricas, centrando o problema na área da avaliação da saúde da criança e do jovem. Faremos, também, a contextualização teórica e a justificação do problema, assim como apresentamos as questões e os objetivos da pesquisa e a organização geral deste trabalho.

1.1A Escolha da Área Temática da Investigação

Neste trabalho procura-se relatar, de forma sistematizada, todo o percurso de investigação e as questões inerentes a um tema que se apresenta como problemático há longos anos e que tem gerado alguma discussão. Usar instrumentos de medida com as características das escalas de avaliação, compostas por itens com formato de medida ordinal não é uma tarefa fácil, devido ao carácter qualitativo e subjetivo dos itens que as compõem e, também, devido à forma como os dados que emergem destes instrumentos são analisados estatisticamente.

Nas áreas de Ciências Sociais e Ciências da Saúde confrontamo-nos, demasiadas vezes, com problemas desta natureza, não só ao medir, mas também ao analisar os

dados que emergem das variáveis ordinais. O tipo de medida destas variáveis pode perpetrar alguns erros estatísticos pela dificuldade na validação dos pressupostos metodológicos e, pelo facto de apenas os seus efeitos ou manifestações poderem ser observáveis.

A motivação pessoal para desenvolver este trabalho está relacionada com os contextos em que tenho trabalhado, na formação de enfermeiros, como professora da disciplina de investigação e na intervenção como profissional de enfermagem e, como investigadora inserida em alguns grupos de pesquisa na área da saúde.

A importância atribuída à área da criança e do jovem teve na sua base, por um lado, o facto de trabalhar com crianças e por outro lado, ser uma área onde se aplicam muitas das escalas desenvolvidas com variáveis com formato de medida ordinal. A utilização destes instrumentos é feita, muitas vezes, sem recorrer à evidência científica e sem tomar as decisões baseadas nessa mesma evidência e nos estudos de simulação já desenvolvidos. É também problemática a escolha de soluções estatísticas para o tratamento de dados, sobretudo quando se consideram as variações possíveis associadas à dimensão da amostra, ao tipo de escala e outros fatores. Assim, surgiu o interesse por estudar as variáveis ordinais e os instrumentos desenvolvidos com estas variáveis para dar algum contributo na tomada de decisão aquando do desenvolvimento e da tradução e adaptação de escalas para uma determinada situação e para um cliente específico.

Na literatura científica está disponibilizado um grande número de instrumentos que apresentam muitas vezes os mesmos objetivos de avaliação, pelo que, a seleção dos instrumentos que melhor servem as decisões a serem tomadas, quer na avaliação específica do desenvolvimento, quer da saúde individual, nem sempre são fáceis por parte de quem os procura, quer seja por serem de diferentes tipologias, quer por falta de uma fundamentação adequada.

Enquanto docente e investigadora fui questionada sobre a problemática da medida, técnicas estatísticas, e validade e fiabilidade dos instrumentos utilizados na avaliação e observação da criança e jovem e nem sempre as justificações teóricas ou metodológicas permitiram uma melhor adequação da decisão.

O recurso à estatística como ferramenta facilitadora da tomada de decisão faz com que o investigador se depare com a problemática do erro mesmo antes de iniciar a

realização dos testes estatísticos (Loureiro & Gameiro, 2011). O estudo da potência dos testes estatísticos está associado a dois erros que podem ser cometidos quando se tomam decisões incorretas. Se o investigador rejeita a hipótese nula sendo ela verdadeira, estamos perante um erro de tipo I (α) ou se não rejeita a hipótese nula sendo ela falsa, estamos perante o erro de tipo II (β), sendo a potência de um teste o seu complemento ($1-\beta$). A investigação de Pearson (1930, 1931) sobre a potência dos testes paramétricos perante a violação dos seus pressupostos é um exemplo da importância atribuída à potência estatística desde os anos trinta do século passado. Estes estudos começaram por analisar a probabilidade de erro de tipo I e só depois a probabilidade do erro de tipo II e a potência de teste.

A simulação estatística tem sido uma mais-valia na comparação da performance dos testes paramétricos e não paramétricos (Blair & Higgins, 1985; Lehman, 1975), mas os estudos com variáveis ordinais são em menor quantidade (Nath & Pavur, 1985; Katz & MacSweeney, 1980). Os estudos de simulação, sob o ponto de vista científico e metodológico, apresentam especificidades muito próprias e usam-se pela impossibilidade de se obterem analiticamente resultados. Assim, podemos dizer que uma das âncoras deste trabalho é o estudo de simulação pelo método de Monte Carlo, no qual foram utilizadas duas metodologias de geração aleatória de dados ordinais. No delineamento do estudo de simulação, houve a necessidade de criar vários cenários nos quais pudessem ser enquadradas diferentes situações reais. Assim, foram utilizadas várias dimensões amostrais, 3 tipos de intensidade de correlação entre os itens, escalas de medida com números de pontos diferentes, e também distintas distribuições de frequência.

Em teoria, com variáveis com formato de medida ordinal não deviam ser utilizadas estatísticas paramétricas e, a serem usadas deviam ter pelo menos 5 pontos (Lozano, García-Cueto, & Muñiz, 2008, Temme, 2006). Em situações específicas em que as escalas têm menos de três pontos, nem as estatísticas não paramétricas deviam ser aplicadas (Katz & MacSweeney, 1980). Na prática, tanto se usam testes paramétricos como não paramétricos com este tipo de variáveis, havendo situações em que a justificação está enquadrada em alguns estudos de simulação e outras em que tal

não acontece, não sendo, em muitos casos, valorizada a dimensão da amostra, a correlação entre os itens, nem a distribuição de frequência.

No estudo de simulação desenvolvido no âmbito deste trabalho, a seleção dos 4 formatos de medida com 3, 4, 5, 7 pontos, justifica-se por estes serem o número de pontos mais utilizados nas escalas de avaliação em saúde infantil e pediatria.

A exclusão de escalas com um formato de medida com 2 pontos prende-se com o facto de estarmos a usar variáveis ordinais. Esta decisão exclui globalmente as escalas de avaliação de desenvolvimento da criança cuja finalidade é verificar a ausência ou presença de um atributo na criança, e como tal apresentam-se de forma dicotómica. A opção por escalas com um formato de medida igual ou superior a 3 pontos fez com que o foco fosse em instrumentos que avaliam a saúde da criança (avaliação de desenvolvimento oro-motor, de risco de lesão da pele, da independência nas atividades de vida diária).

O percurso de trabalho para a investigação que agora se apresenta foi bastante sinuoso, não só pelas mudanças e obstáculos que tiveram que ser ultrapassados, mas também pelo próprio estudo de simulação realizado. Numa primeira fase deparámo-nos com as normais dificuldades da elaboração do programa de simulação que apresentava especificidades muito próprias por necessitar de um algoritmo para gerar dados ordinais; numa segunda fase, surge o problema relacionado com o tempo de execução das simulações, pelo facto de um dos dois métodos de geração de dados utilizado ser muito lento, o que exigiu um longo período de simulação (17 meses) com recurso contínuo a 7 computadores virtuais e periodicamente (9 meses) a mais 23 computadores fixos de forma a simular todos os cenários previstos.

No ponto seguinte faremos a contextualização teórica da temática sob estudo.

1.2 Contextualização do Estudo

Durante alguns anos, a tomada de decisão relacionada com a prática de cuidados foi sustentada pelas crenças e valores e pelas competências clínicas baseadas na experiência individual dos profissionais de saúde. Este processo, apesar da sua importância, tem vindo a demonstrar ser insuficiente para responder aos problemas de saúde de forma sustentada e fundamentada. O facto de as decisões não serem suportadas

pelo conhecimento científico, e serem baseadas na tradição e, no senso comum, reduziu a credibilidade do processo de decisão. Esta descredibilização, associada às exigências da sociedade, para uma resposta eficaz aos problemas de saúde, reforça a necessidade de a tomada de decisão ser mais orientada por princípios científicos, ou seja, por princípios suportados pela investigação e pelos seus resultados. O conceito de práticas baseadas na evidência dá mais relevância à investigação, sem, no entanto, menosprezar outros fatores fundamentais relacionados com os cuidados ao cliente. Assim, as práticas baseadas na evidência determinam que a decisão clínica deve ser fundamentada na evidência científica disponível, no reconhecimento da preferência do cliente, no contexto dos cuidados prestados e na perícia clínica dos profissionais. Este tipo de práticas facilita o cuidado individualizado e contribui para a melhoria na qualidade de cuidados o que, conseqüentemente se traduz, em ganhos em saúde e bem-estar para o cliente e em satisfação para o profissional.

As práticas de cuidados exigem dos profissionais uma constante observação e/ou avaliação do cliente, em que obrigatoriamente se recorre a técnicas de mensuração. Medir, em saúde, pode incluir duas facetas, uma mais ligada à medida global da população e outra mais ligada à medida individual (seja ela objetiva ou subjetiva). Este tipo de medida objetiva ou subjetiva tem a ver com o tipo de variáveis a medir: as diretamente observáveis, relacionadas com características físicas e biológicas (e.g., peso, altura, hemoglobina), e como tal mais fáceis de medir; e as não diretamente observáveis (e.g., desenvolvimento, desenvolvimento motor oral, dor, satisfação, bem-estar, depressão, distúrbios da linguagem, risco de úlceras de pressão), cuja medição é feita através das suas manifestações e, por isso, colocando outros problemas ao processo de medida.

Nas áreas das ciências sociais e ciências da saúde, os investigadores e os profissionais confrontam-se, demasiadas vezes, com problemas relacionados com a mensuração e com escalas de avaliação, fazendo a sua discussão segundo dois ângulos. O primeiro, relativo à investigação empírica, em que se discutem metodologias de análise de dados ordinais e em que se avalia a adequação das técnicas (univariadas vs. multivariadas e paramétricas vs. não paramétricas) aos problemas a investigar e, às variáveis e à dimensão das amostras. O segundo, relativo à intervenção clínica, em que

os profissionais de saúde têm dificuldade na escolha de instrumentos de observação e de avaliação para usar nas suas práticas, o que se traduz numa escassez de instrumentos analisados e validados para as crianças e jovens e por vezes uma avaliação menos rigorosa dos clientes e uma disparidade nas intervenções dentro das equipas de saúde. O enfoque dos cuidados deve assentar na observação sistematizada da criança e nos resultados mensuráveis da sua evolução ao longo do desenvolvimento, e tendo como objetivo a sua saúde física e psicossocial, pois só através da medida, seja ela objetiva ou subjetiva, e da análise dos resultados que dela emergem se pode avaliar com rigor o desenvolvimento, a saúde e o comportamento destes clientes tão vulneráveis e as suas necessidades em saúde.

No caso dos recém-nascidos e crianças mais pequenas esta avaliação é feita, normalmente, pelos pais ou “pessoa significativa”, por educadores e por profissionais de saúde; nas crianças mais velhas e nos jovens a avaliação muitas vezes é feita tendo por base a autoavaliação. A avaliação subjetiva, com recurso a escalas, tem-se revelado uma mais-valia na avaliação da saúde individual (Theme Filha, Szwarcwald & Souza Junior, 2008), sobretudo em áreas como a neonatologia e pediatria. A medição do estado de saúde da criança permite-nos comparar, em termos de condição de saúde, grupos definidos à custa de variáveis como o género, a idade, com as condições de saúde, as condições sociais, psicológicas e comportamentais (Hicks et al., 2001).

As escalas de avaliação (Hill & Hill, 2009) são instrumentos que permitem avaliar de forma sistematizada constructos de quantificação mais difícil e têm na sua origem muito do trabalho desenvolvido até 1930 na área da psicologia e da psicometria por psicólogos como J. M. Cattell, Alfred Binet e Theodore Simon e estatísticos como Sir Francis Galton, Karl Pearson e Charles Spearman, entre outros, pela sua vertente associada à psicometria (Pasquali, 2010). A avaliação em psicologia seguiu orientações muito próprias e serviu de recurso na construção de escalas a outras áreas de conhecimento, o que deu origem à diversificação das escalas de avaliação mediante o contexto em que eram desenvolvidos e aplicados (sociologia, educação, saúde, etc.) (Ribeiro, 2007).

Diferentes profissionais de distintas áreas do conhecimento têm trabalhado no desenvolvimento e adaptação de escalas, inventários e *checklists* de itens (Anastasi,

1990; Agresti, 2002). Os problemas da adaptação e validação destes instrumentos são um tópico recorrente em todas as áreas do conhecimento que as utilizam, e colocam problemas complexos aos profissionais que as aplicam. A aproximação dos grupos profissionais de distintas áreas fez com que a investigação deixasse de se centrar apenas nos testes psicológicos para se estudarem outros constructos de fronteira interdisciplinar que pudessem cobrir um leque mais abrangente de fenómenos a observar em contexto real (Leal & Marôco, 2010). Assim, começam-se a avaliar não somente as características psicológicas e disfuncionais, mas também de desenvolvimento, como as alterações específicas do desenvolvimento, do comportamento, da linguagem e constructos específicos da área da saúde, os quais exigiam diferentes formas de avaliação e, como tal, de medida.

As escalas para avaliar desenvolvimento podem ter como finalidade avaliar o desenvolvimento geral (Bayley, 1933, 2006; Provost, 2004; Bellman, Lingam & Aukett, 2003; Huntley, 1996; Frankenburg et al., 1992; Frankenburg & Dodds 1967; Griffiths, 1954; Provost, Crowe, & McClain, 2000); o desenvolvimento motor (Russell et al., 2002; Folio, & Fewell, 2000; Piper & Darrah, 1994); o desenvolvimento da linguagem (Zimmerman & Castilleja, 2005; Zimmerman, Steiner & Pond, 2002; Coplan & Gleason, 1993; Coplan, 1985; Rescorla, 1989); o desenvolvimento cognitivo (Swanson et al., 2009); o comportamento (Brazelton, 1973; Korner et al., 1987, 2000; Dubowitz & Dubowitz, 1981) e alterações específicas do desenvolvimento (Robins, 2008, Robinn et al., 2001; Eaves, Wingert & Ho, 2006). Na sua maioria estes instrumentos avaliam a presença ou a ausência do atributo (escalas dicotómicas) e permitem fazer uma apreciação por *scores* ou perfis de desenvolvimento.

Têm sido também desenvolvidas escalas que avaliam constructos específicos da saúde, como escalas que avaliam a dor (Grégoire & Finley, 2008), como por exemplo a *Premature Infant Pain Profile* (Stevens et al., 1996), a *Neonatal Pain, Agitation and Sedation Scale* (Hummel, 2009), ou a *Faces Pain Scale-Revised* (Hicks et al., 2001); escalas que avaliam o risco de lesões da pele no recém-nascido e crianças, como por exemplo a *Neonatal Skin Risk Assessment Scale* (Huffines & Lodgson, 1997); escalas que avaliam as competências para alimentação oral, por exemplo a *Early Feeding Skills* (Thoyre, Shaker & Pridham, 2007, 2005); escalas que avaliam a independência

funcional, por exemplo a *Functional Independence Measure*, (Keith, et al., 1984, 1987; Granger et al., 1986); escalas que avaliam a depressão, por exemplo o *Beck Depression Inventory* (Beck, Steer & Garbin, 1988; Beck et al., 1961), a *Depression Self-Rating Scale* (Birlleson et al., 1987, Birlleson, 1981); escalas que avaliam a ansiedade, por exemplo o *State-Trait Anxiety Inventory for Children*, (Spielberger, 1973, Spielberger et al., 1973); escalas que avaliam stress, por exemplo a *Perceived Stress Scale* (Cohen, Kamarck & Mermelstein, 1983); as escalas que avaliam autoperceção, por exemplo a *Self-Perception Profile for Adolescents*, a *Self-Perception Profile for Children* (Harter, 1988, 1985); e as escalas que avaliam autoestima e autoconceito, por exemplo a *Rosenberg Self-Esteem Scale* (Rosenberg, 1965, 1962). Estas escalas, contrariamente às escalas desenvolvidas para avaliar desenvolvimento, apresentam formatos de medida com mais de três pontos. Este foi o facto que levou a que algumas delas fossem usadas neste estudo, pois o seu formato de medida ajusta-se ao estudo de simulação desenvolvido neste trabalho.

A medição dos constructos anteriormente enumerados passa pelo facto de ser possível observar comportamentos num processo de heteroavaliação (como é o caso da observação de populações neonatais), ou autoavaliação (como é o caso das populações pediátricas). É neste contexto que emergem as escalas de avaliação compostas por variáveis qualitativas com um formato de medida ordinal. As variáveis que constituem estes instrumentos são muitas vezes analisadas como se fossem de tipo intervalar, o que tem levantado alguma controvérsia (Urdan, 2005; Jamieson, 2004), ainda que a sua utilização seja justificada, empiricamente, por estudos de simulação (Carifo & Perla, 2008, 2007; Finch, 2005; Glass, Peckham & Sanders, 1972).

Na prática, verifica-se que muitos investigadores usam estas variáveis como se fossem variáveis contínuas (Curado, Teles & Marôco, 2013), justificando esta decisão com estudos de simulação em que os autores consideram que os resultados obtidos são de confiança se o formato de medida for superior a 5 pontos e a distribuição de frequências for próxima da distribuição normal (Marôco, 2010b; Temme, 2006).

O formato da medida e respetivo número de pontos dos itens ordinais surge na literatura com uma grande diversidade (2 a 11 pontos) (Curado, Teles & Marôco, 2013; Preston & Colman, 2000; Alwin, 1997; Jacoby & Matell, 1971) dependendo do objetivo

da investigação, da população sob estudo e do tipo de análise de dados a aplicar. Se o investigador vai usar a soma algébrica de itens (*summated scales*), apesar de não haver definições concretas são usados formatos de medida desde os dois pontos, se a análise é feita item a item, as linhas orientadoras são no sentido de um número de pontos mais elevado.

Muitos estudos têm tido como finalidade relacionar o número de pontos de resposta ao item com a consistência interna dos itens, havendo diferentes interpretações desta questão. Há autores que não valorizam esta relação (Green & Rao 1970, Jacoby & Matell, 1971; Aiken 1983) e outros que não só a valorizam (Krosnick and Fabrigar, 1997) como consideram que entre 5 e 7 pontos está o número ótimo para avaliar validade e fiabilidade dos itens. Outros dizem ainda que, a fiabilidade pode ir aumentando nas escalas e nos 7 pontos estabilizar, pelo que os itens com formato de medida superior a 7 pontos podem ser menos fiáveis (Alwin & Krosnick, 1991; Dawes, 2008).

Lozano, García-Cueto e Muñiz (2008) desenvolveram um estudo de simulação, através do Método de Monte Carlo com o objetivo de determinar o número de alternativas de resposta aos itens que maximiza as propriedades psicométricas fundamentais de uma escala: fiabilidade e validade.

The results show that as the number of response alternatives increases, both reliability and validity improve. The optimum number of alternatives is between four and seven. (Lozano, García-Cueto & Muñiz, 2008, p. 73).

Os estudos de Moors (2007) acerca da utilização de um formato com um número de pontos par (seis pontos) ou um número de pontos ímpar (cinco) mostraram que não havia grandes diferenças na utilização dos dois formatos:

In both samples a latent-class confirmatory factor analysis revealed an extreme response style factor. The 6-response categories version, however, revealed the more consistent set of effects. As far as the content latent-class factors, i.e. familistic values and ethnocentrism, are concerned, results were fairly similar. However, a somewhat deviant pattern regarding the familistic values items in the 6-response categories version suggested that this set of items is less homogeneous than the set of ethnocentric items. (Moors, 2007, p.779).

No que diz respeito à aplicação de testes paramétricos com variáveis ordinais, embora teoricamente isso seja pouco aceitável, na prática verifica-se que é muito comum entre os investigadores das ciências sociais e da saúde. Algumas das justificações, que se encontram na literatura, para esta prática são suportadas pelo formato de medida, havendo autores que consideram que os resultados são de confiança desde que o formato de medida tenha 5 pontos ou mais e os itens apresentem distribuição de frequência normal:

Simulation studies have indeed shown that single-group confirmatory analysis is relatively robust against violating the assumptions of continuous variables if categorization is based on at least five answer categories and the data does not show excessive skewness and/or Kurtosis (Temme, 2006, p. 457).

No entanto relativamente aos testes não paramétricos, Katz e McSweeney (1980) consideram que para aplicar a MANOVA não paramétrica a variáveis ordinais, seria importante que as escalas utilizadas tivessem um formato de medida superior a três pontos

The nonparametric analogue to the one-way MANOVA requires only that the distributions be continuous distributions. The use of the technique would be questionable for data on a two or three-point scale since this is a clear violation of the assumption of continuity; however, a few ties in the data can be tolerated. (Katz & McSweeney, 1980, p. 283).

1.3 Justificação do Problema

Ao longo do processo de investigação, o investigador desenvolve instrumentos de recolha de dados, deparando-se com o conceito de variável e com as condições da sua medição. A importância atribuída ao tipo de variáveis e aos instrumentos a elas subjacentes tem incentivado a investigação sobre as escalas de avaliação, no sentido de diminuir a subjetividade da observação.

Os investigadores e os clínicos passaram a privilegiar instrumentos cuja construção ou adaptação cultural e validação estatística tenha sido desenvolvida de forma criteriosa (consistência interna, fiabilidade e validade dos itens) (Marôco, 2010b;

Severo et al., 2006). Associado a estes aspetos, tem-se enfatizado o uso de dois observadores na observação ou avaliação com estes instrumentos (com validação inter-observadores) e também feito a comparação entre instrumentos. Jacobson, Groot e Samson (1994) consideram que a utilização simultânea de um instrumento por dois observadores e a utilização de mais do que um instrumento de recolha de informação podem constituir-se como uma abordagem suplementar e mais abrangente.

Mas, se por um lado, a forma estruturada e sistematizada das escalas de avaliação tem facilitado a aplicação prática pelos profissionais ou pelos clientes, por outro lado, o seu formato de medida (ordinal) e o número de pontos da escala têm levantado alguns problemas na fase de análise de dados. Esta problemática está especificamente relacionada com o tipo de testes estatísticos que devem ser utilizados, isto é, se a opção deve ser por testes paramétricos ou não paramétricos.

A reflexão sobre as práticas tem focado dois aspetos fundamentais. Um, ao nível da investigação empírica, em que os investigadores discutem as metodologias a utilizar na análise de dados com variáveis ordinais e se usam testes paramétricos ou não paramétricos. Outro, ao nível da prática, em que os profissionais de saúde têm dificuldade na seleção de instrumentos a usar com os seus clientes, pois, ao pesquisarem a melhor evidência científica, no sentido de replicarem estudos ou de traduzirem e validarem estatisticamente os instrumentos, deparam-se com uma grande diversidade de “escalas de avaliação”, com diferentes formatos de medida (3, 4, 5, 6, 7, 9 e 10 pontos).

Relativamente à análise estatística, verifica-se também na literatura que tem sido tradição, quer nas ciências sociais, quer nas ciências da saúde, os investigadores usarem a estatística paramétrica na análise de variáveis ordinais, em vez da estatística não paramétrica. Contudo, estudos de simulação (Zimmerman, 2000) têm sugerido que alguns testes não paramétricos (e.g. Mann-Whitney e Kruskal-Wallis) são tão sensíveis à violação de algumas condições (por exemplo, a homocedasticidade), quanto os paramétricos, na inflação das probabilidades dos erros de tipo I e de tipo II. Assim sendo, a utilização da estatística não paramétrica, em alternativa à paramétrica, pode não ser a melhor estratégia de análise dos dados. Por isso, muitas vezes, na literatura científica, deparamo-nos com instrumentos constituídos por variáveis ordinais, as quais são analisadas com recurso à estatística paramétrica e, outras vezes, verificamos que,

com os mesmos instrumentos, se aplicam testes não paramétricos (Curado, Teles & Marôco, 2013).

Os problemas referidos anteriormente, associados a uma prática de utilização das escalas de avaliação, incentivaram a elaboração de um estudo de simulação, através do método de Monte Carlo, no sentido de analisar a potência dos testes e a probabilidade de erro de tipo I, e determinar quais os testes a aplicar com variáveis ordinais em alguns dos cenários mais comuns com que se deparam os investigadores e profissionais de saúde. Em seguida foram analisados dados reais, provenientes de escalas de avaliação, e feito o seu enquadramento no estudo de simulação.

Se, por um lado, é importante que a tomada de decisão do investigador na escolha do instrumento seja a correta e a análise estatística seja a adequada de modo a obter resultados fiáveis, por outro lado, é fundamental que a tomada de decisão dos profissionais aquando da escolha dos instrumentos a usar na avaliação dos clientes (crianças ou adultos) seja a mais credível.

1.4 Questões e Objetivos da Investigação

Este trabalho organizou-se em torno das diferentes fases do processo de investigação. Na fase inicial de conceção construiu-se o projeto de investigação, identificou-se o problema e fez-se a revisão teórica que permitiu situar o domínio da investigação. Assim, no caso da investigação que agora se apresenta, o confronto com questões que surgem na prática de cuidados e da investigação nesta área, com diferentes perspetivas teóricas e com o conhecimento e observação da realidade social conduziram-nos à elaboração das questões de investigação que servem de suporte à orientação da pesquisa:

1) Quais os testes estatísticos mais frequentemente utilizados pelos investigadores da área da saúde infantil e pediatria, quando usam escalas de avaliação compostas por variáveis (itens) com um formato de medida ordinal?

2) Qual a performance em termos de potência de teste e probabilidade de erro de tipo I das quatro estatísticas da MANOVA paramétrica comparadas com duas da MANOVA não paramétrica quando se analisam variáveis (geradas aleatoriamente) com formato de medida ordinal?

3) Quais os testes estatísticos mais adequados para analisar os dados ordinais oriundos da observação de populações neonatais e pediátricas, os paramétricos ou os não-paramétricos?

As três questões de investigação formuladas são indissociáveis dos três objetivos da investigação, que se descrevem e fundamentam em seguida.

Para fundamentar o primeiro objetivo, “Identificar na literatura científica os testes estatísticos mais frequentemente utilizados pelos investigadores da área da saúde infantil e pediatria quando usam escalas de avaliação”, foi feita uma revisão sistemática da literatura, onde foram incluídos os artigos científicos que continham os seguintes aspetos: instrumentos de recolha de dados (escalas de avaliação) usados na área de saúde infantil e pediatria (desde o recém-nascido ao jovem adulto, com e sem deficiência motora), compostos por variáveis ordinais com diferentes formatos de medida, tendo sido usados na análise de dados testes paramétricos e/ou não paramétricos. Esta revisão deu origem a um artigo de revisão sistemática da literatura (Curado, Teles & Marôco, 2013). Os contributos teóricos associados a uma reflexão crítica sobre esta temática deram origem a um segundo artigo sobre a problemática do uso de variáveis não diretamente observáveis (Curado, Teles & Marôco, 2014).

Para fundamentar o segundo objetivo, “Comparar a performance, em termos de potência e probabilidade de erro de tipo I, das 4 estatísticas da MANOVA paramétrica com 2 da MANOVA não paramétrica quando se utilizam variáveis ordinais correlacionadas, geradas aleatoriamente” foi feito um estudo de simulação através do método de Monte Carlo, efetuado no *Software R (R Development Core Team, 2012)*. Foram gerados dados de natureza ordinal correlacionados através de dois métodos: a conversão binária e o mapeamento de médias. (Kaiser, Träger & Leisch, 2011; Ramsey & Ramsey, 2012; Ramsey et al., 2011; Demirtas, 2006; Bilodeau & Micheaux, 2005; Zwick, 1985; Nath & Pavur, 1985).

No delineamento do estudo de simulação foi considerado o caso balanceado com dimensão global da amostra igual a n (com $n=30, 60, 90, 120, 240, 300$); 3 grupos ($g=3$), sendo a dimensão da amostra em cada grupo dada por $n_i=n/3$; 4 tipos de vetores de probabilidades marginais, fazendo variar a localização da maior probabilidade de resposta, no ponto central (distribuição simétrica), nos extremos inferior e superior da

escala (distribuição assimétrica positiva e negativa) ou uniforme na escala de medida (distribuição uniforme); 3 valores de correlação para a matriz de simetria composta ($r=0.10$, $r=0.40$, $r=0.70$); e 4 formatos de medida com 3, 4, 5 e 7 pontos.

No caso da MANOVA paramétrica avalia-se a hipótese nula com utilização das estatísticas Lambda de Wilks, traço de Pillai, traço de Hotelling, maior raiz de Roy (Todorov & Filzmoser, 2010; Marôco, 2010a; Sheskin, 2007; Finch, 2005; Zwick, 1985; Katz & McSweeney, 1980; Puri & Sen, 1969, 1974) e para a MANOVA não paramétrica avaliam-se as estatísticas Lambda de Wilks e traço de Pillai aplicadas às ordens (Van Aelst & Willems, 2011; Todorov & Filzmoser, 2010; Nath & Pavur, 1985; Katz & McSweeney, 1980; Puri & Sen, 1974).

Para fundamentar o terceiro objetivo, “Enquadrar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados reais provenientes de escalas de avaliação com quatro formatos de medida (3, 4, 5 e 7 pontos), nos resultados do estudo de simulação estatística” foram realizadas 4 aplicações práticas. As estatísticas utilizadas nas simulações relativas às diferentes propostas de MANOVA paramétrica e não paramétrica foram aplicadas a variáveis ordinais com um formato de medida com 3, 4, 5 e 7 pontos. Esta escolha fundamenta-se no facto de serem os formatos mais utilizados na construção de escalas de observação especificamente na área da saúde da criança (Curado, Teles & Marôco, 2013; Pereira da Silva & Justo da Silva, 2010).

Os instrumentos selecionados foram aplicados a crianças e jovens (no período neonatal e adolescência/jovens adultos) e para construir os grupos considerou-se a variável idade, a variável peso e a variável nível de lesão medular. Para os recém-nascidos considerou-se a idade gestacional e o peso, por serem duas das variáveis que mais interferem no bem-estar do recém-nascido, quer sob o ponto de vista fisiológico, quer sob o ponto de vista comportamental. Para as crianças e jovens consideraram-se os escalões etários e o nível de lesão medular (sagrada, lombar e dorsal).

As escalas selecionadas para fazer as aplicações práticas foram: a Early Feeding Skills (EFS); a Neonatal Skin Risk Assessment Scale (NSRAS); a Functional Independence Measure (FIM). A primeira tem um formato de medida com 3 pontos, a segunda com 4 pontos e a terceira com 7 e a MIF motora modificada com 5 pontos.

A Early Feeding Skills (EFS) foi traduzida e adaptada para português como Escala de Observação das Competências Precoces na Alimentação Oral - Versão Modificada (Curado et al., 2013) e avalia desenvolvimento motor oral do recém-nascido (Thoyre, Shaker & Pridham, 2007, 2005) até às 52 semanas de vida, sendo composta por vinte e dois itens ordinais, com um formato de medida com 3 pontos. Esta escala avalia os comportamentos de regulação da criança em 3 domínios fundamentais: capacidade para organizar o funcionamento motor oral, capacidade para coordenar a deglutição e a capacidade para manter a estabilidade fisiológica (compostos por 7, 5 e 10 itens, respetivamente). A escolha desta escala está implicitamente ligada ao facto de a alimentação ser um dos problemas mais comuns na população neonatal e especificamente nos recém-nascidos prematuros.

De um modo geral, depois do quadro respiratório estabilizado, a prioridade para os recém-nascidos pré-termo passa a ser o aspeto nutricional. (...) que favoreça o seu crescimento e desenvolvimento, assim como a redução dos déficits motores orais (...) (Figueiredo et al., 2007, p. 3).

Na aplicação prática foi utilizado o mesmo número de observações utilizadas no estudo de simulação. A seleção aleatória das amostras foi feita com recurso à variável independente idade gestacional (em classes) ($n_i=100, 80, 40, 30, 20$ e 10).

A Neonatal Skin Risk Assessment Scale (NSRAS) (Huffines & Lodgson, 1997) avalia o risco de úlceras de pressão no recém-nascido, nos primeiros vinte e oito dias de vida. A escala encontra-se em fase de validação estatística, com recém-nascidos portugueses, como Escala de Observação do Risco de Lesão da Pele de Neonatos (Martins, Curado & Marçal, 2013). Esta escala tem um formato de medida ordinal com 4 pontos e é composta por 6 itens: condição física geral; estado mental; mobilidade; atividade; nutrição e humidade. A escolha desta escala deve-se ao facto de as úlceras de pressão serem consideradas pela Organização Mundial de Saúde como um grave problema de saúde pública e ao mesmo tempo serem um dos melhores indicadores de avaliação da qualidade de cuidados em saúde (OMS, 2007). Em Portugal, as úlceras são um problema recorrente o qual é considerado como um fator causal de dor e sofrimento (DGS, 2011) e cuja prevenção pode evitar a morte em recém-nascidos de risco (EPUAP-NPUAP, 2009). Na aplicação prática com esta escala foram utilizadas

observações de cento e vinte recém-nascidos, sendo a seleção aleatória das amostras feita com recurso à variável independente peso (em classes) com as mesmas dimensões das utilizadas no estudo de simulação ($n_i=40, 30, 20$ e 10).

The Functional Independence Measure (FIM) (Guide for the Uniform Data System for Medical Rehabilitation (Adult FIM), 1993; Guide for the Functional Independence Measure for Children (WeeFIM), 1993; Keith et al., 1984, 1987; Granger et al., 1986), traduzida e adaptada para a língua portuguesa como Medida de Independência Funcional (MIF) (Laíns, 1991) é um instrumento que permite fazer o acompanhamento de pessoas em processo de reabilitação. Este instrumento não se focaliza na capacidade de realizar tarefas, mas sim na realização efetiva e de forma independente das atividades de vida diárias. A MIF permite avaliar o desempenho da criança/jovem na realização de um conjunto de 18 tarefas incluídas em 2 domínios, o motor (4 subescalas com 13 itens) e o cognitivo (2 subescalas com 5 itens). O domínio motor é composto pelas subescalas: autocuidado (6 itens), controlo esfíncteriano (2 itens), mobilidade e transferências (3 itens), locomoção (2 itens). O domínio cognitivo é composto pelas subescalas: comunicação (2 itens) e cognição social (3 itens). Cada item pode ser classificado numa escala de graus de dependência com formato de medida de 7 pontos (de dependência total a independência total). Malheiro, Gaspar e Barros (2013) usaram o domínio motor da MIF para comparar com outra escala, transformando o seu formato de medida de 7 para 5 pontos. A modificação da escala de medida teve em conta as necessidades de ajuda das crianças com espinha bífida, agregando os 3 pontos mais elevados da escala num ponto único, o ponto 5-Independentes e mantendo os outros quatro pontos da escala de medida, 4-Ajuda Mínima, 3-Ajuda Moderada, 2-Ajuda Máxima, 1-Ajuda Total.

Nas aplicações práticas com a MIF modificada (escala com 5 pontos) foram selecionadas aleatoriamente 40 observações com recurso à variável independente nível de lesão medular ($n_i=20$ e 10) e na MIF com formato de medida com 7 pontos foram selecionadas 90 observações ($n_i=30, 20$ e 10) de crianças/jovens com *spina bífida* com recurso à variável grupos etários.

1.5 Organização do Trabalho

O trabalho está estruturado em cinco partes principais, as quais são precedidas pela Introdução Geral.

As cinco partes que integram o trabalho são constituídas por oito capítulos, cada uma delas precedida de uma nota introdutória. A parte I reporta à contextualização teórica, a parte II apresenta dois artigos, já publicados, baseados na aplicação da teoria e investigação divulgada, a parte III apresenta o estudo de simulação, a parte IV apresenta as aplicações práticas e seu enquadramento no estudo de simulação e a parte V a discussão geral dos resultados e respetivas implicações teóricas e práticas.

No âmbito da Introdução Geral, explicitam-se as razões da escolha da área temática, faz-se a contextualização do estudo, justifica-se o problema, apresentam-se as questões de investigação e os objetivos e a organização do trabalho.

Na Contextualização Teórica, o Capítulo II e Capítulo III integram a revisão da teoria que fundamenta todo o processo de investigação. O Capítulo II inclui a avaliação da saúde da população neonatal e pediátrica no que concerne às escalas de avaliação de saúde e desenvolvimento e destaca a importância dos cuidados baseados na evidência.

O terceiro capítulo coloca em relevo o formato de medida (escalas ordinais) e as técnicas estatísticas e o quarto capítulo destaca a simulação estatística pelo método de Monte Carlo, na qual é utilizada a geração aleatória de dados ordinais. O quinto e sexto capítulo incluem dois artigos, uma revisão sistemática da literatura intitulada “Análise Estatística de Escalas Ordinais: Aplicações na área da Saúde Infantil e Pediatria”¹ e um texto de reflexão teórica baseado nos resultados da investigação, “Análise de Variáveis Não Diretamente Observáveis: Influência na Tomada de Decisão Durante o Processo de Investigação”². Os resultados do estudo de simulação pelo método de Monte Carlo, “MANOVA paramétrica e não paramétrica com variáveis ordinais: A probabilidade de Erro de Tipo I e a potência de Teste” são apresentados no capítulo VII.

¹ Curado, M.A.S., Teles, J., & Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, Abril. <http://www.um.es/eglobal>.

² Curado, M.A.S., Teles, J., & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process. *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48(1), 149-56.

No capítulo VIII faz-se a apresentação dos resultados das aplicações práticas com dados das escalas Early Feeding Skills, Neonatal Skin Risk Assessment Scale e The Functional Independence Measure.

Na Parte V apresenta-se a discussão geral que enfatiza as implicações teóricas e práticas do processo geral de investigação e salienta os resultados do estudo de simulação estatística e das aplicações práticas com dados reais. A finalizar apresentamos os anexos.

1.6 Referências Bibliográficas

- Agresti A. (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons, Inc.
- Alwin, D.F. & Krosnick, J.A (1991). The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods and Research*, 20(1), 139-181.
- Anastasi A. (1990). *Psychological Testing*. New York: Macmillan Publishing Company.
- Bayley, N. (1933). *The California first-year mental Scale*. Berkeley: University of California Press.
- Bayley, N. (2005). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: Psychological Corporation.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: Psychological Corporation.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Beck, A.T., Steer, R.A., & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77-100.
- Bellman, M., Lingam, S., & Aukett, A. (2003). SGS II. Escala de Avaliação das Competências no Desenvolvimento Infantil – dos 0 aos 5 anos. Lisboa: CEGOC – TEA.
- Bilodeau, M., & Micheaux, P.L. (2005). A multivariate empirical characteristic function test of independence with normal marginal. *Journal of Multivariate Analysis*, 95, 345-369.
- Birleson, P. (1981). The Validity of Depressive Disorder in Childhood and the Development of a Self-Rating Scale: A Research Report. *Journal of Child Psychology and Psychiatry*, 22, 73-88.
- Birleson, P., Hudson, I., Grey-Buchanan, D., & Wolff, S. (1987). Clinical Evaluation of a Self-Rating Scale for Depressive Disorder in Childhood (Depression Self-Rating Scale). *Journal of Child Psychology and Psychiatry*, 28, 43-60.
- Blair, R.C., & Higgins, J.J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon's signed-rank test under various population shapes. *Psychological Bulletin*, 97, 119-128.
- Brazelton, T.B. (1973). Neonatal Behavioral Assessment Scale. *Clinics in Developmental Medicine*, 50. London: Willian Heinemann Medical Books.

- Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396.
- Coplan, J., & Gleason, J.R. (1993). Test-retest and interobserver reliability of the Early Language Milestone Scale, 2nd Ed. *Journal of Pediatric Health Care*, 7(5), 212-219.
- Coplan J. (1985). Evaluation of the child with delayed speech or language. *Pediatric Annals*, 14, 203-208.
- Curado, M.A.S., Teles, J., Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud Infantil y Pediatría. *Enfermería Global*, 30, 446-457. <http://www.um.es/eglobal>.
- Curado, M.A.S., Teles, J. & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process, *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48(1), 149-156.
- Curado, M.A.S., Mâroco, J., Vasconcellos, T., Marques, L.C., Oliveira, S.R., & Nicolau, C.S. (2013). Versão Reduzida da Escala de Observação de Competências Precoces na Alimentação Oral. Um estudo de validação estatística com crianças prematuras Portuguesas, *Actas do 1º Encontro de Enfermeiros de Neonatologia da Área de Lisboa: "Cuidar para o Desenvolvimento"*, 15-16 novembro 2013 no Auditório do Hospital S. Francisco Xavier, Lisboa, resumo 4.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales, *International Journal of Market Research*, 50(1), 61-77.
- Demirtas, H. (2006). A method for multivariate ordinal data generation given marginal distributions and correlations. *Journal of Statistical Computation and Simulation*, 76(11), 1017-1025.
- Direção Geral de Saúde (DGS) (2011). *Escala de Braden: Versão Adulto e Pediátrica* (Braden Q). Retrieved October 22, 2012, from: <http://www.dgs.pt/upload/membro.id/ficheiros/i015800.pdf>.
- Dubowitz L.M.S., & Dubowitz V. (1981). The neurological assessment of the preterm and full term newborn infant. *Clinics in Development Medicine*, 79.
- Eaves, L.C., Wingert, H., & Ho, H.H. (2006). Screening for autism. *Autism*, 10(3), 229-242.
- Eccles, M., Grimshaw, J., Walker, A., Johnston, M., Pitts, N. (2005). Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology*, 58, 107-112.
- European Pressure Ulcer Advisory Panel and National Pressure Ulcer Advisory Panel (EPUAP-NPUAP) (2009). *Prevention and Treatment of Pressure Ulcers*. Washington DC: National Pressure Ulcer Advisory Panel.
- Figueiredo, H., Barreiros, J., Gonçalves, I., & Cunha, M. (2007). Avaliação de um programa de intervenção oral em recém-nascidos pré-termo de muito baixo peso. *ESSFISIONLINE*, 3(2), 3-12.
- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions Are Violated. *Methodology*, 1(1), 27-38.
- Folio, M.R., & Fewell, R.R. (2000). *Peabody Developmental Motor Scales, (PDMS-2)* 2nd Ed. Austin, Tex: PRO-ED, Inc.

- Frankenburg, W.K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics*, 89, 91-97.
- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42, (3), 237-288.
- Granger, C.V., Hamilton, B.B., Keith, R.A., Zielezny, M., & Sherwin, F.S. (1986). Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation*, 1, 59-74.
- Green, P.E. & Rao, V.R. (1970). Rating Scales and Information Recovery: How Many Scales and Response Categories to Use? *Journal of Marketing*, 34, 33-39.
- Grégoire, M.C., & Finley, A.G. (2008). Doctor, I Think my Baby is in Pain: The assessment of infants' pain by health professionals. *Journal of Pediatric*, 84, 6-8.
- Griffiths, R. (1954). *The Abilities of Babies*. London: University of London Press.
- Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM) (1993). Uniform Data System for Medical Rehabilitation: Procedures for scoring the functional independence measure (FIM), version 4.0, Buffalo (NY), State University of New York at Buffalo.
- Guide for the Functional Independence Measure for Children (WeeFIM) (1993). Uniform Data System for Medical Rehabilitation, version 4.0, *Community/Outpatient*. Buffalo (NY), State University of New York at Buffalo.
- Guyatt G.H., Oxman A.D., Vist G., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J., for the GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336, 924-926.
- Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver: University of Denver Press.
- Harter, S. (1988). *Manual for the Adolescent Self-Perception Profile*. Denver: University of Denver Press.
- Hicks, C.L., Von Baeyer, C.L., Spafford, P.A., Van Korlaar, I, Goodenough, B. (2001). The Faces Pain Scale-Revised: Toward a common metric in pediatric pain measurement. *Pain*, 93(173), 83-86.
- Hill, M.M. & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Huffines, B., & Lodgson, M.C. (1997). The Neonatal Skin Risk Assessment Scale for predicting skin breakdown in neonates. *Issues Comprehensive Pediatrics Nursing*, 20, 103-114.
- Hummel, P. (2009). *N-PASS Neonatal pain, agitation and sedation scale*. [16 de Setembro de 2012]; Disponível em: www.n-pass.com/assessment_table.html.
- Hunt, M. (1987). The process of translating research findings into nursing practice. *Journal of Advanced Nursing*, 12, 101-110.
- Huntley, M. (1996). *The Griffiths Mental Development Scales. From Birth to 2 Years. Manual. (Revision)*. London, UK: Association for Research in Infant and Child Development.
- JACOBSON, A.M., GROOT, M., SAMSON, J.A. (1994). THE EVALUATION OF TWO MEASURES OF QUALITY OF LIFE IN PATIENTS WITH TYPE I AND TYPE II DIABETES. *DIABETES CARE*, 17(4), 267-274.
- Jacoby, J. & Matell, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, VIII, 495-500.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Kaiser, S., Träger, D., & Leisch, F. (2011). *Generating Correlated Ordinal Random Values. Technical Report Number 94*. Munich: Department of Statistics, University of Munich.

- Katz, B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Keith, R., Granger, C., Hamilton, B., & Sherwin, F. (1984). *Functional Independence Measure, FIM. Uniform Data System for Medical Rehabilitation*. State University of New York: Buffalo.
- Keith, R.A., Granger, C.V., Hamilton, B.B., & Sherwin, F.S. (1987). The Functional Independence Measure: A new tool for rehabilitation. *Advanced Clinical Rehabilitation*, 1, 6-18.
- Krosnick, J.A. & Fabrigar, L.R. (1997). Designing rating scales for effective measurement in surveys. In L.E. Lyberg, P. Biemer, M. Collins and E.D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin. *Survey measurement and process quality*, Hoboken, NJ. Wiley, 14–164.
- Krosnick, J.A. & Presser, S. (2010). Question and Questionnaire Design. In P.V. Marsden & J.D. Wright, *Handbook of Survey Research*, Vol. 2. Bingley, U.K. Emerald Group Publishing, Ltd., 263-314.
- Laíns, J. (1991). Guia para o uso do Banco Uniformizado de Dados para Reabilitação Médica - MIF. Versão Portuguesa de Guide for the Uniform Data System for Medical Rehabilitation, FIM (1984). Produção Sandoz.
- Leal, I. & Marôco, J. (2010). *Avaliação em Sexualidade e Parentalidade*. Porto: Legis, Ed.
- Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
- Loureiro, L.M.J., & Gameiro, M.G.H. (2011). Interpretação crítica dos resultados estatísticos: para lá da significância estatística. *Revista de Enfermagem Referência*, 3, 151-162.
- Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73-79.
- Malheiro, M.I.C., Gaspar, M.F.M., & Barros, L. (2013). The effects of a therapeutic camp: program for promotion of self-management of disability on functionality and self-concept of children with spina bífida. In *Abstracts Book of 57th Annual Meeting of Society for Research into Hydrocephalus and Spina Bífida*, 26-29 June, Cologne, paper 11.
- Marôco, J. (2010a). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Marôco, J. (2010b). *Análise de Equações Estruturais: Fundamentos teóricos, Software & Aplicações*. Pêro Pinheiro: ReportNumber, Lda.
- Martins, C.O.A., Curado, M.A.S., & Marçal, T. O. (2014). *Adaptation and Validation of the Observation Scale of Skin Injury Risk in New-borns: A Statistical Study of Pré-validation with Portuguese New-Borns*. Abstracts of 13th International Conference of Nursing Research: Person Centered Culture. Fundação Calouste Gulbenkian, Lisboa: APE, 25.
- Moors, G. (2007). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42(6), 779-794.
- Nath, R., & Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics & Data Analysis*, 2, 297-315.
- ORGANIZAÇÃO MUNDIAL DE SAÚDE (OMS) (2007). *INTERNATIONAL STATISTICAL CLASSIFICATION OF DISEASES AND RELATED HEALTH PROBLEMS*, 10TH REVISION. Retrieved October 22, 2013, from: www.who.int/classifications/apps/icd/icd10online/
- Pasquali, L. (2010). *História dos instrumentos psicológicos*. In Luiz Pasquali e Colaboradores, *Instrumentação Psicológica. Fundamentos e Prática*. Porto Alegre: Artmed.
- Pearson, E.S. (1930). A further development of tests of normality. *Biometrika*, 22, 239-249.
- Pearson, E.S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- Pearson, A., Wiechula, R., Court, A., Lockwood, C. (2005). The JBI model of evidence-based healthcare. *International Journal of Evidence-Based Healthcare*, 3(8), 207-215.

- Pereira da Silva, T., & Justo da Silva, L. (2010). Escalas de Avaliação da dor utilizadas no recém-nascido. Revisão Sistemática. *Acta Médica Portuguesa*, 23, 437-454.
- Piper M.C., Darrah J. (1994). *Motor assessment of the developing infant*. EUA, WB: Saunders Company.
- Presser, S. & Schuman H. (1980). *The Measurement of a Middle Position in Attitude Surveys*. *Public Opinion Quarterly*, 44(1), 70-85.
- Provost, B., Crowe, T., & McClain, C. (2000). Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales in two-year old children. *Physical Occupational Therapy in Pediatrics*, 20, 5-18.
- Provost, B., Heimerl, S., McClain, C., Kim, N., Lopez, B., & Kodituwakku, P. (2004). Concurrent validity of the Bayley Scales of Infant Development II, Motor Scale and the Peabody Developmental Motor Scales-2 in children with developmental delays. *Pediatric Physical Therapy*, 16(3), 149-156.
- Puri, M.L., & Sen, P.K. (1974). *Nonparametric methods in multivariate analysis*. New York: John Wiley & Sons.
- Puri, M.L., & Sen, P.K. (1969). Class of Rank Order Tests for a General Linear Hypothesis. *Annals of Mathematical Statistics*, 40(4), 1325-1343.
- Ramsey, P.H., & Ramsey, P.P. (2013). Pairwise testing of group mean vectors in MANOVA with small samples. *Journal of Statistical Computation and Simulation*, 83(6), 1037-1057.
- Ramsey, P.H., Ramsey, P.P., Hachimine, P., Andiloro, N. (2011). Robustness, Power and Interpretability of Pairwise Tests of Discriminant Functions in MANOVA. *Journal of Modern and Applied Statistical Methods*, 10(2), 403-417.
- R Development Core Team (2012). *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (<http://www.R-project.org>).
- Rescorla, L. (1989). The Language Development Survey: a screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54(4), 587-599.
- Ribeiro, J.P. (2007). *Avaliação em Psicologia da Saúde*. Lisboa: Climepsi.
- Robins, D.L. (2008). Screening for Autism Spectrum in primary care settings. *Autism*, 12(5), 537-556.
- Robins, D.L., Fein, D., Barton, M.L., & Green, J.A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive development disorders. *Journal of Autism and Development Disorders*, 31, 2, 131-144.
- Rosenberg, M. (1962). The association between self-esteem and anxiety. *Journal of Psychiatric Research*, 1, 135-152.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Russell, D.J., Rosenbaum, P.L., Avery, L.M., & Lane, M. (2002). *Gross Motor Function Measure (GMFM-66 & GMFM-88) user's manual*. Cambridge (UK): Cambridge Univ. Press.
- Sackett, D.L., Rosenberg, W., Grover, F., Haynes, R., Richards, W. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71-72.
- Severo, M., Santos, A. C., Lopes, C., Barros, H. (2006). Fiabilidade e Validade dos Conceitos Teóricos das Dimensões de Saúde Física e Mental da Versão Portuguesa do MOS sf-36*. *Acta Médica Portuguesa*, 19, 281-288.
- Sheridan, M.D. (1984). *From birth to five years: children development progress*. Hong Kong: The Nfer-Nelson Publishing Company, 1-74.
- Sheskin, D.J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures* (4th ed.). Boca Raton, Florida: Chapman & Hall/CRC.

- Spielberger, C.D. (1973). *State-Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C.D., Edwards, C.D., Lushene, R.E., Montuori, J., & Platzek, D. (1973). *Preliminary manual for the State-Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists Press.
- Stevens, B., Johnston, C., Petryshen, P., Taddio A. (1996). Premature infant pain profile: Development and initial validation. *Clinical Journal of Pain*, 12(1), 13-22.
- Swanson, J.R., Bradley-Johnson, S., Johnson, C.M., & O'Dell, A.R. (2009). The Cognitive Abilities Scale-Second Edition Preschool Form. Studies of Concurrent Criterion-Related, Construct, and Predictive Criterion-Related Validity. *Journal of Psychoeducational Assessment*, 27(1), 46-56.
- Temme, D. (2006). Assessing measurement invariance of ordinal indicators in cross-national research. In S. Diehl & R. Terlutter (Eds.), *International advertising and communication*. Wiesbaden, Germany: DUV, 452-472.
- Theme Filha, M.M., Szwarcwald, C.L., Souza J., P.R.B. (2008). Medidas de morbidade referida e inter-relações com dimensões de saúde. *Revista de Saúde Pública*, [online], 42(1), 73-81.
- Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistics and Data Analysis*, 54, 37-48.
- Thoyre, S.M.; Shaker C.S. & Pridham, K.F. (2007). *The Manual of Early Feeding Skills - EFS*, Chapel Hill, The University of North Carolina.
- Thoyre, S.M., Shaker C.S., & Pridham, K.F. (2005). The early feeding skills assessment for preterm infants. *Neonatal Network*, 24(3), 7-16.
- Tucker, J., & McGuire, W. (2004). ABC of preterm birth: Epidemiology of preterm birth. *British Medical Journal*, 329, 675-677.
- Urdan, T.C. (2005). *Statistics in Plain English*. London: Laurence Erlbaum Associates.
- Van Aelst, S., & Willems, G. (2011). Robust and Efficient One-Way MANOVA Tests. *Journal of the American Statistical Association*, 106(494), 706-718.
- Zimmerman, D.W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127, 354-364.
- Zimmerman, I.L., & Castilleja, N.F. (2005). The role of language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities*, 11, 238-246.
- Zimmerman, I.L., Steiner, V. G., & Pond R. E. (2002). *Preschool Language Scale*, 4th Ed. San Antonio, TX: The Psychological Corporation.
- Zwick, R. (1985). Nonparametric One-Way Multivariate Analysis of Variance: A Computational Approach Based on the Pillai Bartlett Trace. *Psychological Bulletin*, 97(1), 148-152.

PARTE I. Contextualização Teórica

Introdução

Neste capítulo damos relevância à investigação em saúde e aos seus resultados e também à necessidade de os profissionais de saúde recorrerem a práticas de cuidados baseados na evidência. Somente com a investigação como suporte se podem desenvolver cuidados de “saúde avançados” (*advanced care*) à pessoa em “processo de saúde e doença”. Neste contexto consideramos a avaliação em saúde e valorizamos a observação subjetiva da saúde e a sua importância para a saúde individual dos clientes. Dá-se particular destaque à observação sistematizada de populações neonatais e pediátricas e, à avaliação do seu desenvolvimento e de alguns aspetos da sua saúde com escalas de avaliação, dando importância à sua aplicação, à sua estrutura e à escala de medida (ordinal) subjacente aos itens que as constituem.

Num processo de investigação e de análise dos resultados da investigação o recurso à estatística é fundamental. A utilização dos métodos e técnicas estatísticas começa com o planeamento do projeto e atravessa todo o planeamento da investigação, na decisão da dimensão da amostra, na decisão do instrumento de recolha de dados, na organização dos dados e na análise e interpretação dos resultados. No que diz respeito à análise estatística dos dados, tendo em conta que interessa analisar simultaneamente diversas variáveis que estão relacionadas, é usual o recurso a técnicas multivariadas: análise fatorial, análise de *clusters*, análise de regressão linear múltipla, análise de regressão logística múltipla, análise discriminante e análise de variância (Hair et al., 2010; Hill & Hill, 2009).

A análise de observações multivariadas numa determinada amostra pode ser feita com recurso a diferentes metodologias, porém neste trabalho, damos relevo à MANOVA paramétrica e à MANOVA não paramétrica a um fator. A opção na contextualização teórica deve-se ao facto de estes testes serem muito utilizadas em análise de dados em ciências sociais e da saúde (Curado, Teles, & Marôco, 2013).

Capítulo II. Investigação e Avaliação da Saúde da População Neonatal e Pediátrica

A inovação e o avanço tecnológico na área da saúde, associado à investigação e às exigências da sociedade para uma resposta eficaz aos problemas de saúde dos indivíduos e das populações, têm mostrado que o processo de tomada de decisão deve ser conduzido por princípios científicos. Só o conhecimento científico e a aplicação dos resultados da investigação permitem a adequação das intervenções em situações específicas, obtendo-se assim melhores resultados em saúde.

O conhecimento científico é organizado, metódico, sistemático, analítico, cumulativo, comunicável e preciso na sua fundamentação, possuindo características de racionalidade, de objetividade, de ser empírico e replicável.

A produção de conhecimento científico assenta em dois elementos informativos básicos: os dados ou observações e as teorias ou postulados teóricos. (...) O primeiro mais heurístico [procura de soluções] e exploratório, o segundo volta-se declaradamente para a réplica, testagem, explicação e controlo de fenómenos. (Almeida & Freire, 2003, p. 32).

Perante cuidados de saúde cada vez mais diferenciados e intervenções complexas é exigido dos profissionais de saúde que as suas práticas se baseiem em competências fundamentais:

a) Expert guidance and coaching of patients, families, and other care providers; b) Consultation; c) Research skills, including use, evaluation, and conduct; d) Clinical and professional leadership, which includes competence as a change agent; e) Collaboration; f) Ethical decision-making skills. (Smith, Vezina & Samost, 2009, p. 212).

Estas competências profissionais incluem a orientação especializada e o treino de clientes, famílias e outros prestadores de cuidados, consultadoria, competências de investigação (transferência dos resultados para as práticas), liderança profissional e clínica (o profissional como agente de mudança), colaboração e ética na tomada de decisão. A aplicação dos resultados da investigação na prática de cuidados suporta o desempenho profissional, através de tomadas de decisão fundamentadas cientificamente e de práticas baseadas na evidência.

For the Advanced Practice Nursing to guide and shape practice, research must be integrated. Whether we serve as consumers of research by reading and applying results of scientific reports or actually conduct studies to determine the answer to a clinical question, research must be evident as an important element of everyday practice. (Cipriano & Burns, 2009, p. 306)

A investigação torna-se, assim, uma peça fundamental dos cuidados em duas dimensões que se complementam: a do investigador e a do cuidador que participa em equipas de investigação e precisa de saber interpretar os resultados da investigação para assim fazer a sua utilização nas práticas. O desenvolvimento das práticas baseadas na evidência começou por estar relacionado com a ciência médica e só depois passou para outras áreas da saúde. Foi com Archie Cochrane e com a publicação em 1972 de *Effectiveness and Efficiency: Random Reflections on Health Services* que se admitiu o desajuste entre o conhecimento científico e as práticas de cuidados. Assim, começaram a desenvolver-se práticas baseadas na evidência, multiplicando-se as ações formativas e criando-se linhas orientadoras, no sentido de reduzir a variabilidade nas práticas médicas. Nos anos noventa do século passado, o Professor David Sackett da Universidade de Oxford fomentou o seu desenvolvimento e o termo “baseado na evidência” extravasou do campo da medicina e passou a ser aplicado na enfermagem pelos decisores de políticas de saúde e outros profissionais de saúde. Em saúde, o conceito de Prática Baseada na Evidência (PBE), como processo dinâmico que é, combina as quatro componentes da decisão clínica: *A Evidência Científica disponível; A Perícia Clínica; O Reconhecimento da Preferência do Cliente* e o *Contexto de Cuidados*. As práticas baseadas na evidência são a base da qualidade de cuidados e consequentemente traduzem-se em ganhos em saúde para o cliente.

A Evidência Científica disponível reporta à investigação relevante na área clínica, cujo enfoque está no cliente e nas competências que o profissional tem para analisar os aspetos teóricos, metodológicos, de análise e interpretação dos relatórios de investigação e da informação que daí pode ser retirada. Os resultados que advêm da investigação permitem condutas e tomadas de decisão mais seguras e eficazes, daí a importância da capacitação dos profissionais e dos investigadores para a sua

interpretação e implementação. Esta implementação deve ser monitorizada e avaliada de forma a melhorar o desempenho profissional.

A *Perícia Clínica* incorpora as capacidades individuais de utilização do conhecimento e da experiência clínica, e a sua aplicação prática, na identificação do estado de saúde do cliente (observação, diagnóstico, riscos e benefícios das intervenções e avaliação). *O Reconhecimento da Preferência do Cliente* implica que o profissional de saúde respeite os seus valores, crenças, expectativas e preocupações, incluindo estes aspetos no plano de cuidados e integrando o cliente nas decisões clínicas. *O Contexto de Cuidados* está em constante mudança, pelo que o profissional de saúde deve estar preparado para oferecer espaços de intimidade com todas as condições de respeito e igualdade que os clientes necessitam, atendendo às suas vulnerabilidades (Guyatt et al., 2008; Simões, 2005; Pearson & Craig, 2002; Sackett et al., 2000).

A resolução de problemas complexos em saúde tem subjacente a prática baseada na evidência e resoluções que podem incluir contribuições de diferentes áreas como a psicologia e a sociologia. Podemos dizer que os problemas com que nos deparamos na atualidade são cada vez mais complexos pelo que a sua abordagem deve ser multidisciplinar, para que se possa chegar a uma melhor solução. Esta ideia é reforçada pela *European Union Research Advisory Board* (ERAB, 2004) que afirma que a maioria dos problemas complexos ambientais, de segurança e de saúde devem ser abordados de forma multidisciplinar. Deste modo, importa definir o que realmente se entende por intervenções complexas, em saúde. Estas intervenções são assim denominadas por englobarem múltiplos componentes, os quais podem atuar independentemente ou de forma interdependente, que, quando aplicados a uma determinada população produzem um conjunto muito variável de possíveis resultados (Campbell et al., 2007; Hallberg, 2009, 2006).

Os cuidados à criança, jovem e família são cuidados que na sua grande maioria são classificados como cuidados complexos, daí a recomendação para que os profissionais de saúde da área da criança e jovem fundamentem as suas práticas no conhecimento procedente da investigação. Também o *International Council of Nursing* (ICN, 1999) tem acompanhado estes processos e tecido considerações fundamentais de suporte à PBE. O ICN considera a investigação em enfermagem como um instrumento

de suporte à atuação do enfermeiro, sem o qual as decisões não serão fundamentadas. A investigação e o processo de implementação dos seus resultados têm envolvido a produção de conhecimento, a sua disseminação e a sua utilização, promovendo mudanças nas práticas.

A valorização da pesquisa e da aplicação nas práticas dos resultados da investigação permitiu uma melhoria na qualidade de cuidados e uma maior valorização das profissões da saúde, considerando-se que na área da saúde a natureza da evidência é influenciada pela:

(...) experience of clinicians in everyday practice who, in using the evidence, assert that there are diverse sources of research-based and non-research-based evidence and that the process of evidence-based practice should be placed within a broader context that is grounded in practice; recognizes different evidentiary bases; and is directed towards improving global health across vastly different practice contexts. (Pearson et al., 2005, p. 207).

Estes aspetos fizeram com que as escolas de enfermagem e os profissionais de saúde considerassem a prestação de cuidados baseada em evidência, uma das suas habilidades fundamentais dos enfermeiros e de todos os profissionais da área da saúde (Cullen & Adams, 2010; Sackett et al., 2000; Culham, 1998; Richards, Lawrence, Sackett, 1997; Sackett et al, 1996; Dawes, 1996; Hunt, 1987), adequando a formação e as intervenções em saúde à situação do utente, fosse ele idoso, adulto, jovem ou criança, orientando a sua decisão clínica pelos valores do cliente e envolvendo-o o mais possível na tomada de decisão sobre a sua saúde (Larrabee, 2009; Pearson et al., 2005).

No caso específico da intervenção com crianças, os pais ou pessoa significativa devem ter um papel preponderante, como responsáveis pela sua saúde, mas este papel não substitui o dos profissionais, que devem procurar a melhor evidência (Craig, 2002). A aplicação dos resultados da investigação nas práticas de cuidados e a observação/avaliação do cliente pediátrico suportada por instrumentos de medida adaptados e validados a esta população facilitam a satisfação das necessidades da criança e famílias (Flemming, 2007). É com base no conhecimento, que emerge da investigação e pela sua devida apropriação pelos profissionais que se avalia com rigor, se faz a otimização dos resultados em saúde e se favorecem as “boas práticas” em saúde

(Cochrane Collaboration, 2011; OE, 2006; Fineout-Overholt & Melnyk, 2005; JBI, 2001).

No capítulo seguinte daremos relevância às questões relacionadas com o formato de medida das variáveis (itens) presentes em inúmeros instrumentos como escalas de avaliação de desenvolvimento e escalas de avaliação de constructos da área da saúde que permitem a observação/avaliação do comportamento, da saúde física e psicossocial de populações neonatais e pediátricas. Iremos também refletir sobre os testes estatísticos utilizados na análise dos dados que emergem da aplicação destas escalas nos contextos de saúde, os quais são também aplicados em outras áreas do conhecimento.

2.1 Escalas de Avaliação do Desenvolvimento e da Saúde da Criança e Jovem

A investigação na área da criança, associada às técnicas de observação, teve o seu início no século XIX, com as pesquisas de Darwin sobre o estudo de desenvolvimento ontogenético e, a análise do comportamento individual (Barbosa, 2004). Um outro contributo fundamental para a investigação e para a medição, na área da criança foi o trabalho desenvolvido, entre 1822 e 1930, por psicólogos como J. M. Cattell, Alfred Binet e Theodore Simon e estatísticos como Sir Francis Galton, Karl Pearson e Charles Spearman, pela sua vertente associada à psicometria (Pasquali, 2010). O interesse de Sir Francis Galton pela área psicopedagógica e clínica fez com que a psicologia se interessasse pela aplicação de testes de aptidão académica e pela avaliação do “atraso mental” em crianças e adultos. Galton acreditava que as operações intelectuais podiam ser medidas através da avaliação sensorial e como tal desenvolveu a teoria de discriminação sensorial, que utilizava escalas com formato de medida qualitativa (com pontos), e simplificou os métodos estatísticos, os quais foram desenvolvidos pelo seu discípulo Karl Pearson e por Charles Spearman.

Os estudos de Alfred Binet na área da psicopedagogia e da saúde foram mais abrangentes, pois o investigador considerava, contrariamente a Sir Francis Galton, que as medidas sensoriais eram importantes, mas não tinham relação com a função intelectual e também eram medidas de avaliação muito específicas. Assim, as suas pesquisas eram mais globais e abrangiam funções como a memória, a atenção, a

compreensão. Em associação com Théodore Simon, desenvolve uma escala de avaliação da inteligência, a escala de Binet-Simon, que permitia avaliar o perfil cognitivo de crianças que necessitavam de apoio pedagógico. Esta escala, denominada *Binet-Simon Scale*, era inicialmente constituída por trinta itens que envolviam três grandes domínios, a compreensão, o julgamento e o raciocínio. Estas foram as grandes bases dos muitos testes e escalas desenvolvidos para avaliar inteligência, das quais salientamos a *Stanford-Binet Intelligence Scale* desenvolvida por psicólogos da Universidade de Stanford e as escalas *Wechsler Intelligence Scale for Children (WISC)* e *Wechsler Preschool and Primary Scale of Intelligence (WPPSI)* desenvolvidas pelo psicólogo David Wechsler (Kaufman, 1994, 1990). A avaliação em psicologia seguiu orientações muito próprias e serviu de recurso nesta especificidade instrumental a outras áreas de conhecimento, o que deu origem à diversificação de instrumentos de avaliação e/ou de observação mediante o contexto (em que esses mesmos instrumentos são desenvolvidos e aplicados, sociologia, educação, saúde, etc.) (Ribeiro, 2007).

Diferentes profissionais, de diferentes áreas do conhecimento, têm trabalhado no desenvolvimento e, adaptação de escalas, inventários e *checklist* de itens (Anastasi, 1990; Agresti, 2002), o que tem levado a uma aproximação dessas áreas num trabalho conjunto para analisar, adaptar e validar estes instrumentos. A aproximação de grupos profissionais, de diferentes áreas, fez com que a investigação deixasse de se centrar apenas nos testes psicológicos para se estudarem outros constructos de fronteira interdisciplinar que pudessem cobrir um leque mais abrangente de fenómenos a observar em contexto real (Leal & Marôco, 2010). Assim, começa-se a avaliar não somente as características psicológicas e disfuncionais, mas também desenvolvimento, alterações específicas do desenvolvimento, comportamento, linguagem e constructos mais específicos da área da saúde, os quais exigiam diferentes formas de avaliação e como tal com diferentes formatos de medida. Apesar da problemática associada ao formato de medida das variáveis qualitativas, estes instrumentos tiveram uma grande aceitação e têm sido amplamente utilizados (Thurstone, 1928; Likert, 1932; Guttman 1950).

Relativamente às muitas escalas para avaliar desenvolvimento, estas podem ter como finalidade avaliar o desenvolvimento geral, desenvolvimento motor,

desenvolvimento da linguagem, desenvolvimento cognitivo, comportamento e alterações específicas do desenvolvimento de entre muitas outras áreas. Estes instrumentos de observação do desenvolvimento da criança, na sua maioria, avaliam a presença ou a ausência de um atributo (escalas dicotómicas) e permitem fazer uma apreciação global do desenvolvimento por scores ou perfis de desenvolvimento.

No Quadro 2. 1, apresentam-se alguns dos instrumentos mais usados para avaliar desenvolvimento e, destacamos as escalas recomendadas, pela Direção Geral de Saúde (DGS), pela Secção de Neonatologia e Sociedade Portuguesa do Neurodesenvolvimento da Sociedade Portuguesa de Pediatria (SNSPN-SPP) e pela Mesa do Colégio da Especialidade de Enfermagem de Saúde Infantil e Pediátrica da Ordem dos Enfermeiros (MCEESIP-OE), para observar as populações pediátricas e neonatais portuguesas.

A DGS (2013), a SNSPN-SPP (2012) e a MCEESIP-OE (2011) preconizam a utilização das escalas de *Mary Sheridan, Schedule of Growing Skills II* e a *Modified Checklist for Autism in Toddlers*, e a SNSPN-SPP recomenda também o uso das escalas de Griffiths, Denver, Bayley e a Gross Motor Function Classification System.

Quadro 2. 1 - Características principais de algumas escalas de avaliação do desenvolvimento da Criança e Jovem.

Instrumento Sigla ou Acrónimo	Idade da População Alvo	Domínio	Nº de Itens	Nº de Pontos	Autores
Gesell Test GT	De 1 aos 36 meses	Desenvolvimento Geral – motricidade fina e grosseira, comportamento adaptativo, linguagem e adaptação social	27	2	Gesell (1940); Gesell & Amatruda (1947)
Bayley Scales Infant and Toddler Development Bayley II	De 1 aos 42 meses	Desenvolvimento Geral – cognitivo (sensoriomotor, exploração, manipulação, memória), linguagem, motor (fino e grosseiro) e socio emocional, comportamento adaptativo	289 e 163	5	Bayley (1933, 2005, 2006); Provost et al. (2004); Barbosa (2004); Provost, Crowe, & McClain (2000)
Mary Sheridan Scale MSS	Dos 0 aos 5 anos	Desenvolvimento geral – postura e motricidade global, visão e motricidade fina, audição e linguagem, comportamento e adaptação social	Varia entre 7 e 14 (depende da idade)	2	Sheridan, Sharma & Cockerill (2007); Sharma & Cockerill (2014)
Schedule of Growing Skills II SGS II	Dos 0 aos 5 anos	Desenvolvimento Geral – função motora grosseira e postura (controlo postural passivo e ativo, locomoção), função motora fina (manipulação), visão, audição, fala e linguagem, social (interação), autonomia pessoal	179	2	Bellman, Lingam & Aukett, (2003)
The Griffiths Mental Development Scale Griffiths	Dos 2 aos 8 anos	Desenvolvimento Geral	120	2	Amod, Cockcroft & Soellaart (2007); Stewart (1997); Huntley (1996); Griffiths (1954, 1967, 1970)
Denver Developmental Screening Test Denver II	Dos 0 aos 6 anos	Desenvolvimento Geral – motricidade grosseira, fina, comportamento pessoal e social, linguagem	125	2	Frankenburg et al. (1992); Frankenburg & Dodds (1967)
Alberta Infant Motor Scale AIMS	Dos 0 aos 18 meses	Desenvolvimento Motor – alinhamento postural, padrões motores	58	2	Piper & Darrah (1994); Kolobe & Bulanda (2006)
Test of Infant Motor Performance TIMP	Dos 0 aos 4 meses (idade gestacional superior a 32 semanas)	Desenvolvimento Motor – controlo motor e alinhamento postural, equilíbrio e coordenação	27 25	2 5/6	Campbell et al. (1995)

Quadro 2.1 - Características principais de algumas escalas de avaliação do desenvolvimento da Criança e Jovem (Continuação).

Peabody Developmental Motor Scale-2 PDMS-2	Dos 0 aos 5 anos (atraso de desenvolvimento)	Desenvolvimento Motor – motor fino (pressão da mão, destreza manual, coordenação mão-olho), motor grosseiro (reflexos, equilíbrio, locomoção)	282	2	Folio & Fewell (2000)
Gross Motor Function Measure GMFM-66 GMFM-88	Dos 5 meses aos 16 anos (paralisia cerebral)	Desenvolvimento Motor grosseiro – deitada a rolar, sentada, gatinhar, joelhos, de pé e a andar, correr e saltar	88	4	Piper & Darrah (1994)
			66	4	Russell et al. (2002)
Preschool Language Scale PLS-5	Dos 0 aos 6 anos e 11 meses	Desenvolvimento da Linguagem (comunicação social, semântica, sintaxe, articulação)	68	2	Zimmerman & Castilleja (2005); Zimmerman, Steiner & Pond (2002)
Early Language Milestone Scale ELMS-2	Dos 0 aos 36 meses	Desenvolvimento da Linguagem – comunicação e linguagem (receptiva e expressiva)	43	2	Coplan & Gleason (1993); Coplan (1985)
The Language Development Survey LDS	Dos 18 aos 35 meses	Desenvolvimento da Linguagem – categorias semânticas	310	2	Rescorla (1989)
The Cognitive Abilities Scale CAS-2	Dos 3 meses aos 3 anos	Desenvolvimento Cognitivo – linguagem oral, escrita, leitura, matemática	88	2	Swanson et al. (2009)
The Graham Test GT	De 1 ao 14 dias	Neuro-motor do recém-nascido – comportamento motor e força, resposta auditiva, visual, tátil e a estímulos dolorosos, irritabilidade	55	2	Graham (1956); Graham, Matarazzo & Caldwell (1956)
Neonatal Behavioral Assessment Scale NBAS	Dos 3 dias ao mês	Neuro-motor do recém-nascido – comportamento (estabilidade autonômica, organização motora e dos estados, atenção e interação), reflexos	28	9	Brazelton (1973)
			18	4	
The Neurological Assessment of Preterm and Full-term Newborn Infant NAPI e NAPIFTI	Dos 0 aos 12 meses (idade gestacional superior a 32 semanas)	Neuro-comportamental – Três domínios (tônus muscular, neuro-comportamento, reflexos)	71	3/4/5	Korner et al. (1987, 2000)
			34	3/4/5	Dubowitz & Dubowitz (1981); Dubowitz, Ricci,Mercuri (2005)
The Modified Checklist for Autism in Toddlers MCHAT	Dos 16 aos 30 meses	Desenvolvimento da criança com Autismo – interação social, linguagem, comunicação, comportamentos repetitivos	23	2	Robins (2008); Robins et al. (2001); Eaves, Wingert & Ho (2006)
Escala de Avaliação Desenvolvimento e Funcionalidade da Criança com Paralisia Cerebral EADFCPC	Dos 12 aos 18 anos	Desenvolvimento e funcionalidade da criança com Paralisia Cerebral. Função motora global e fina, habilidade manual, o desempenho na alimentação, comunicação e inclusão escolar	30	2	Andrada et al. (2009)

As mudanças sociais, os avanços tecnológicos e do conhecimento, problemas económicos e sociais têm vindo a influenciar a saúde das famílias, sendo a prematuridade uma das principais causas de risco nas populações neonatais com consequências ao longo do desenvolvimento da criança. O período neonatal (primeiros 28 dias de vida) torna-se, assim, um dos períodos fundamentais para a saúde da criança, pelo que os profissionais de saúde têm investido, no sentido de salvaguardar a vida e a saúde do recém-nascido de risco, não só pelas repercussões imediatas, mas também pelas repercussões futuras na saúde da criança.

Em Portugal, apesar de termos uma baixa taxa de mortalidade neonatal (0.21%) e infantil (0.33%) (Machado, Alves & Couceiro, 2011), sabemos que a prematuridade é responsável por cerca de 73% da mortalidade neonatal. Os recém-nascidos pré-termo e de baixo-peso, especificamente com uma idade gestacional inferior a 32 semanas e/ou peso inferior a 1500 g, são dos mais vulneráveis. Associados a estes problemas surgem, muitas vezes, novas morbilidades, decorrentes da prematuridade e da doença social (Machado, Alves & Couceiro, 2011). Assim, a promoção da saúde do recém-nascido e da criança é uma prioridade para os profissionais de saúde já que:

(...) fornece a oportunidade de reduzir as diferenças no status de saúde entre os membros de diferentes grupos e assegura a igualdade de oportunidades e recursos para capacitar todas as crianças a alcançar o seu potencial máximo de saúde. (Hockenberry & Barrera, 2014, p. 2).

O desenvolvimento dos cuidados intensivos neonatais associado ao rápido aperfeiçoamento da tecnologia de apoio ao recém-nascido criticamente doente e/ou nascido prematuro têm favorecido a sua sobrevivência e originando consequentemente um número acrescido de crianças com problemas comportamentais e de desenvolvimento. Relativamente ao peso, são considerados recém-nascidos de baixo peso ao nascer (BPN), crianças com peso entre 1501g e 2500g, de muito baixo peso ao nascer (MBPN), crianças com peso entre 1001 e 1500g e extremo baixo peso ao nascer (EBPN), crianças com peso entre 751g e 1000g, independentemente da sua idade gestacional. Relativamente à idade gestacional, são considerados recém-nascido pré-termo os que apresentam idade gestacional inferior a 37 semanas de gestação e, em particular os pré-termo tardio, que apresentam uma idade gestacional entre as 34 e as 36

semanas de gestação, independentemente do peso ao nascer (Engle, 2006; Engle, et al., 2007).

As taxas de sobrevivência aumentaram para 93% nos recém-nascidos com muito baixo peso e para 85% nos recém-nascidos de extremo baixo peso. Os prematuros tardios representam 70% do total da população pré-termo e apresentam uma taxa de mortalidade significativamente mais elevada do que as crianças nascidas a termo (7.9 para 2.4 por 1000 nados-vivos, respetivamente) (Askin & Wilson, 2013; Tomashek, et al., 2007). Dentro das populações neonatais, salienta-se o pré-termo tardio que aparentemente apresenta características de um recém-nascido a termo. As suas características físicas e a transição eficaz da vida intrauterina para a extrauterina, sem que se verifique qualquer problema respiratório, fazem com que os profissionais de saúde o tratem como uma criança de termo o que pode acrescer o risco para a sua saúde. A avaliação da idade gestacional é fundamental para que estas crianças possam ser sinalizadas e também porque a idade gestacional está descrita como fator de risco para problemas relacionados com a alimentação, neuro-desenvolvimento, termorregulação, hipoglicemia, hiperbilirrubinémia, infeções e alterações da função respiratória (Bakewell-Sachs, 2007; Darcy, 2009). Muito tem sido feito ao longo das últimas décadas para melhorar as taxas de mortalidade neonatal e infantil e também para melhorar o conhecimento e atendimento da criança, mas ainda há muito trabalho a desenvolver na área da neonatologia para que sejam prestados cuidados baseados na evidência e suportados por teorias do desenvolvimento.

A evolução da saúde infantil, no país, tem mostrado muito bons resultados (WHO, 2008). A taxa de mortalidade infantil diminuiu de 77.5‰ em 1960 para 3.6‰ em 2009, resultado da melhoria das condições socioeconómicas dos portugueses e das reformas globais na saúde com intervenção específica nas populações pediátricas (Machado, 2008). A taxa de mortalidade abaixo dos 5 anos apresentou entre 2001 e 2009 uma evolução semelhante à taxa de mortalidade infantil, com um decréscimo de 6.2 para 4.5 óbitos por 1000 nados-vivos. O grupo etário dos 5 aos 9 anos apresentou a taxa de mortalidade ainda mais baixa com um decréscimo de 45.1% relativamente ao verificado em 2001 (20.4 por 100 000). Na faixa etária de 1 a 19 anos a mortalidade por causas externas apresenta peso relativo crescente, à medida que aumenta a idade: de

28% dos óbitos para 55% dos óbitos entre 15 e 19 anos (Machado, Alves & Couceiro, 2011).

Os profissionais de saúde deparam-se diariamente com problemáticas de saúde nos diferentes grupos etários, daí a importância de usar instrumentos, baseados nos resultados da investigação, que facilitem a observação e a avaliação sistematizada da criança e jovem privilegiando os cuidados centrados na família.

As escalas de avaliação têm gerado alguma discussão. Para os investigadores a sua utilização tem sido considerada essencial para o aprofundamento do conhecimento e para a validação das teorias que lhe estão subjacentes. Nos contextos de prática clínica, as escalas de avaliação têm permitido observações mais sistematizadas e menos subjetivas, embora muitas das escalas usadas nas práticas de cuidados não estejam validadas para as populações específicas.

Cuidar de recém-nascidos e crianças exige observação física, observação de comportamentos, questionamento, reflexão e investigação. Mas também exige que se faça uma integração dos resultados da observação em teorias consistentes ou, mesmo, obrigue o profissional a procurar novas teorias de apoio à avaliação do comportamento de populações com tantas especificidades como as populações neonatais (Gomes-Pedro, 1986) e pediátricas. Nos anos trinta do século passado, os profissionais de saúde começaram a interessar-se pela avaliação neurológica comportamental dos recém-nascidos, apesar da avaliação sistematizada do comportamento só surgir nos anos cinquenta com o Graham Test (Graham, 1956; Graham, Matarazzo & Caldwell, 1956). Este teste foi considerado como um dos primeiros instrumentos para avaliar o desenvolvimento neuro-comportamental e contemplava, na observação da criança, o comportamento motor, a força, a irritabilidade, a resposta auditiva, a resposta visual, a resposta tátil e a resposta a estímulos dolorosos, o que veio incentivar o interesse pelo desenvolvimento deste e de novos instrumentos. Nesta conjuntura, começam a desenvolver-se várias escalas de observação, quer para a investigação, quer para a utilização nos contextos de prática clínica, muitas das quais para avaliar dor neonatal e pediátrica. Rosenblith e Lipsitt (1959) desenvolveram um estudo de validação dos resultados do teste de *Graham* com o objetivo de verificar a aplicabilidade do teste na população neonatal em diferentes contextos. A investigação baseada nos estudos de

Graham originou uma nova escala de observação, *The Modified Graham Test for Neonates* (Rosenblith, 1961) que incluía uma graduação das respostas do neonato aos estímulos.

As escalas de avaliação desenvolvidas no início dos anos sessenta permitiam observar, sobretudo, a modelação do comportamento da criança pelo ambiente, ou seja, a influência do ambiente no comportamento da criança. As escalas subsequentes passaram a avaliar em que medida o recém-nascido influenciava o ambiente e os cuidadores, para além de ser também influenciado por eles, pois muito do ambiente foi mudado em prol do bem-estar da criança (ruído, luminosidade, oxigenoterapia).

Neste contexto, Barry Brazelton e colaboradores começaram a desenvolver uma “escala de avaliação” com características diferentes das anteriores que, para além da avaliação neurológica, avaliava também o comportamento da criança nas seguintes dimensões: estabilidade autonómica, organização motora, organização dos estados, e qualidade das capacidades relativas à atenção/interação, descrevendo como a criança se integra na interação com o ambiente (Brazelton, 1973). Enquanto as escalas anteriores se centravam na patologia e nas respostas reflexas a estímulos negativos, a *Neonatal Behavioral Assessment Scale* (NBAS) tinha como objetivo avaliar a capacidade de organização do recém-nascido perante estímulos sensoriais positivos, partindo do princípio que a criança modela mais o ambiente do que é modelada por ele, ideia que veio a revolucionar os cuidados neonatais (Gomes-Pedro, 1986) e pediátricos.

A NBAS foi também usada por Als et al., (1979) na operacionalização do Modelo da Teoria Sinactiva do Desenvolvimento. Este modelo teórico identifica as respostas comportamentais da criança ao ambiente, integrando o sistema autónomo ou fisiológico (padrão respiratório, coloração da pele), de autorregulação (comportamento na manutenção dos outros sistemas), motor (postura, padrões de movimento, atividade), estado ou organizacional (sono, alerta, choro), e atenção ou interação (reação a estímulos externos) (Als, 1982, 2004). Nesta sequência, surge o programa de cuidados individualizados para o desenvolvimento. Este modelo avalia o desenvolvimento neuro-comportamental do recém-nascido prematuro no ambiente extrauterino, através de observação sistematizada. O observador humano é o instrumento mais sensível para

identificar pistas através do ambiente e da prestação de cuidados. *The Newborn Individualized Developmental Care and Assessment Program* (NIDCAP)

(...) has lead to a greater emphasis on development care for high-risk infants and their families to enhance neuro-development outcomes. Developmental care is an approach that uses a range of medical and nursing interventions that aim to decrease the stress of pré-term infants in NICU. [Neonatal Intensive Care Unit] (Sehgal & Stack, 2006, p. 1007).

A especificidade deste modelo e, a necessidade de pessoal especializado para o colocar em prática impossibilitou a sua utilização por parte de algumas unidades de neonatologia. Neste contexto, começa a ser desenvolvido o Modelo do Universo dos Cuidados para o Desenvolvimento (*Universe of Developmental Care* (UDC)) (Gibbins et al., 2008) que não exige formação específica com credenciais de aplicação do modelo, mas sim formação teórica em desenvolvimento da criança e prática clínica em neonatologia e pediatria, o que permitiu uma revisão destas teorias:

“(...) the first major reformulations of neonatal developmental care theory since Als' Synactive theory. On both theoretical and empirical grounds, the concept of a shared surface is central to this model.” (Gibbins et al., 2008, p. 141).

No UDC, os cuidados estão estruturados para dar resposta às necessidades de cada criança e sua família e focalizam-se na proteção desta díade relativamente às suas vulnerabilidades físicas, psicossociais e emocionais, minimizando os efeitos do ambiente das unidades de cuidados intensivos neonatais e pediátricos. Neste modelo considera-se que o sistema nervoso se organiza como um *feedback* em *looping*, em que a pele é a interface entre as terminações nervosas e o ambiente:

Developmental care has its roots in developmental neurobiology, with Als sinactive theory and in the discipline of professional nursing, with Florence Nightingale's theories on the importance of the nurturing environment for patient care and restoration of health. (Gibbins et al., 2008, p. 142).

O modelo reconhece a pele como fronteira e superfície partilhada entre o prestador de cuidados e a pessoa e os componentes do modelo suportam os cuidados centrados na família: *The model emphasizes the shared surface of the caregiver and infant as the main neurodevelopment interface* (Gibbins et al., 2008, p. 143), dando

importância ao ambiente físico (microambiente), ao ambiente da unidade (luz, ruído), à interação família e criança através do “toque positivo”. O núcleo central do modelo (*central core*) coloca a criança como elemento central dos cuidados, legitimando a díade criança-família, o ambiente, os prestadores de cuidados e as práticas de cuidados (*care practices*). Os sistemas centrais do modelo são: o nervoso, o músculo-esquelético, o gastrointestinal, o tegumentar, o imunológico, o metabólico, o hematológico, o cardíaco e o pulmonar, os quais são influenciados pela idade gestacional e pela situação de saúde e doença da criança. O modelo enfatiza a importância da monitorização e da observação da criança, salientando como ações fundamentais dos profissionais: a observação da alimentação (avaliação das competências e estimulação oral), a promoção de um posicionamento correto (promover o alinhamento à linha média, flexão e contenção da criança), o controlo e prevenção da infeção, a segurança da criança (medidas de segurança apropriadas), o conforto (avaliação e tratamento da dor), a termorregulação (manutenção de ambiente térmico neutro), os cuidados com a pele (avaliação da integridade cutânea e prevenção de lesões da pele), e a respiração (monitorização e ventilação não-invasiva).

Os profissionais de saúde, como observadores privilegiados das respostas humanas à saúde e doença, podem dar contributos fundamentais para a investigação e utilizar os resultados desta nas práticas de cuidados. A ênfase está na avaliação da criança e nos resultados mensuráveis pelo facto de, através da medida, serem determinados os resultados positivos para o cliente e para o prestador de cuidados. A tendência para a utilização dos resultados da investigação e a transferência desses resultados para a prática é uma realidade, especialmente em áreas altamente especializadas como a neonatologia e a pediatria. Dar consistência teórica às práticas pela via da investigação é um dos objetivos dos profissionais de saúde, considerando que ao fundamentarem as práticas clínicas na ciência e na investigação, e ao (...) *documentarem os seus resultados clínicos, eles serão capazes de validar as suas contribuições para a saúde, bem-estar e cura (...) dos clientes (...)* (Hockenberry & Barrera, 2014, p. 15).

As escalas de avaliação (Hill & Hill, 2009) com um formato de medida ordinal, são um exemplo muito comum dos instrumentos de observação utilizados em pediatria.

Estes instrumentos apresentam-se com diferentes objetivos de avaliação que estão relacionados com a população alvo (e.g. neonatos, crianças e jovens), tipo de escalas (unidimensionais ou multidimensionais com várias dimensões ou subescalas), número de itens (indicadores), formato de medida (2, 3, 4, 5, 6, 7, ..., pontos). Os diferentes objetivos da medição fazem com que, na prática pediátrica e neonatal, sejam muitos os instrumentos utilizados para avaliar o desenvolvimento, comportamento e saúde individual, como foi exemplificado anteriormente. Por um lado, os instrumentos de avaliação têm finalidade própria, de acordo com os objetivos da observação e da colheita de dados, por outro lado, sendo instrumentos para usar na clínica, devem obedecer a normas e critérios de referência, explícitos e baseados na teoria e na investigação, permitindo uma medida fiável e eficaz (Rocha, Dornelas & Magalhães, 2013). Assim, devem ser adaptados às populações específicas e testados estatisticamente para que possam medir o que realmente se pretende medir. As práticas de cuidados avançados em neonatologia e pediatria dependem (...) *on “acquiring”, analyzing, synthesizing and applying evidence to inform the practice process* (Dickenson-Hazard, 2002, p. 1). Na impossibilidade de destacar todos os instrumentos desenvolvidos na área da saúde da criança, vamos dar relevância a alguns dos mais utilizados (Quadro 2. 2).

Quadro 2. 2 - Características principais de algumas escalas de avaliação da saúde da Criança e Jovem

Instrumento Sigla ou Acrônimo	Idade da População Alvo	Domínio	Nº Itens	Nº de Pontos	Autores
Echelle Douleur et d'Inconfort du Nouveau-Né EDIN	Dos 0 aos 12 meses (idade gestacional superior a 25 semanas)	Dor aguda prolongada (pós-operatório)	5	4	Debillon et al. (2001)
Premature Infant Pain Profile PIPP	Recém-nascidos pré-termo e a termo	Dor aguda (pós-operatório)	4	4	Stevens et al. (1996); Grégoire & Finley (2008)
Neonatal Infant Pain Scale NIPS	Recém-nascidos pré-termo e a termo	Dor aguda no prematuro (pós-operatório)	6	3	Lawrence et al. (1993)
Neonatal Pain, Agitation and Sedation Scale N-PASS	Recém-nascidos pré-termo e a termo	Dor e sedação	5	3	Hummel (2009)
The Faces Pain Scale-Revised FPSR	Crianças com idade superior a 4 anos	Todos os tipos de dor	1	10	Hicks et al. (2001)
Scale of Pain Face, Legs, Activity, Cry, Consolability FLACC	Crianças e jovens, com idade inferior a 19 anos	Dor aguda (pós-operatório)	5	3	Merker et al. (1997); Nilsson, Finnström, & Kokinsky (2008); Batalha et al. (2009)
Comfort Scale CS	Recém-nascidos em cuidados intensivos	Distresse	8	5	Ambuel et al. (1992); Bear & Ward-Smith (2006).
Douleur Enfant San Salvador DESS	Criança com deficiência motora e paralisia cerebral	Todos os tipos de dor	10	3	Collignon (1995, 1997); Garcia & Fernandes (2007)
Neonatal Skin Risk Assessment Scale NSRAS	Recém-nascidos até aos 28 dias	Risco de lesão da pele no recém-nascido	6	4	Huffines & Lodgson (1997)
Braden Scale Braden Q	Crianças com idade superior a 28 dias	Risco de úlceras de pressão na criança	7	4	Braden & Bergstrom (1987); Ferreira et al. (2007)
Early Feeding Skills EFS	Das 0 às 52 semanas	Competências para a alimentação oral	22	3	Thoyre, Shaker & Pridham (2007, 2005)

Quadro 2.2 - Características principais de algumas escalas de avaliação da saúde da Criança e Jovem (Continuação).

Functional Independence Measure FIM	Crianças e jovens	Independência funcional e atividades de vida diária	18	7	Keith, et al. (1984, 1987); Granger et al. (1986)
Beck Depression Inventory BDI	Crianças e jovens	Depressão	21	4	Beck, Steer & Garbin (1988); Beck et al. (1961)
Child Depression Inventory CDI	Dos 7 aos 17 anos	Depressão	27	3	Kovacs (1992)
Depression Self-Rating Scale DSRS	Crianças e jovens	Depressão	18	3	Birleson et al. (1987); Birleson (1981)
State-Trait Anxiety Inventory for Children STAIC	Crianças e jovens	Ansiedade	40	3	Spielberger (1966, 1983); Spielberger, Gorsuch & Lushere (1973)
Self-Perception Profile for Adolescents e Self-Perception Profile for Children SPPA e SPPC	Jovens	Auto-percepção	51	4	Harter (1988, 1985, 1982); Peixoto & Mata (1983)
	Crianças		36	4	
Adolescent Social Self Efficacy Scale S-EFF	Jovens	Percepção da eficácia em situações sociais	25	7	Connolly (1989)
Rosenberg Self-Esteem Scale RSES	Crianças dos 7 aos 12 anos	Autoestima global	10	4	Rosenberg (1965, 1962); Byrne (1996)
Multidimensional Scale of Perceived Social Support MSPSS	Jovens	Suporte social	12	7	Zimet et al. (1988, 1990); Ekbäck et al. (2013)
			24	5	
Perceived Stress Scale PSS	Crianças e jovens	Stress	10	5	Cohen, Kamarck & Mermelstein (1983)
Posttraumatic Stress Disorder scale PTSD	Crianças e jovens	Stress pós-traumático	20	5	Blake et al. (1995)

No modelo de *Joanna Briggs Institute* (JBI) os cuidados de saúde fazem parte de um processo cíclico que começa com questões, dúvidas, preocupações, indecisões sobre uma determinada intervenção clínica ou problema de saúde. No caso de instrumentos de observação, como as “escalas de avaliação”, é necessário ter em consideração o modo como foram construídas ou adaptadas e validadas estatisticamente para as populações específicas. Se nos focamos na intervenção clínica, por exemplo na alimentação da criança, e se precisarmos de um instrumento que nos permita avaliar as competências da criança para iniciar a alimentação, é necessário, primeiro, identificar os diferentes instrumentos disponíveis. Posteriormente deve-se prosseguir com uma reflexão sobre as práticas no sentido de tomar a melhor decisão sobre a escala mais apropriada à situação. Ou seja, decidir, de entre as muitas escalas encontradas, qual a melhor para cumprir o objetivo inicial, fundamentando essa decisão na experiência, na teoria e nos resultados da investigação. Por isso, importa verificar se a intervenção ou atividade a desenvolver é prática e praticável, ou seja, exequível do ponto de vista prático, se financeiramente é possível no contexto de cuidados e adequada à experiência de alimentação da criança e se é eficaz, ou seja, se nos permite atingir o objetivo para o qual foi planeada. Quando a evidência sugere o uso de uma intervenção específica tem que haver também planeamento e tomada de decisão ao nível organizacional, pois estas questões exigem, para além da informação, a medição e o apoio organizacional, dado que se não houver apoio institucional, dificilmente se podem implementar medidas de melhoria de cuidados de saúde. A utilização destas escalas veio facilitar a observação e avaliação da criança de forma sistematizada, apesar de acarretar também algumas dificuldades. Usar instrumentos suportados pela investigação e pela prática baseada na evidência exige, dos profissionais, trabalho acrescido. Não podem intervir junto das crianças com suposições baseadas na tradição e, como tal, a formação passa a ser uma constante no seu dia-a-dia. É importante que os profissionais valorizem a formação contínua, quer através da permanente atualização e estudo da investigação disponível, quer pela formação especializada. No subcapítulo seguinte o enfoque está nas práticas de cuidados baseados na evidência e no modelo utilizado pelo Instituto Joanna Briggs, destacando-se os resultados da investigação e a sua transferência para as práticas de cuidados.

2.2 Da Investigação aos Cuidados de Saúde Baseados na Evidência

O modelo de cuidados de saúde baseado na evidência preconizado pelo JBI criado em 1996, na Universidade de Adelaide – Austrália, utiliza o conceito de PBE conceptualizando as práticas de cuidados em decisões baseadas na investigação:

The JBI model of evidence-based healthcare conceptualizes evidence-based practice as clinical decision-making that considers the best available evidence; the context in which the care is delivered; client preference; and the professional judgment of the health professional. (Pearson, et al., 2005, p. 209)

A evidência pode ser definida como a investigação clínica centrada no cliente, neste caso, centrada em populações neonatais e pediátricas. Mas o desenvolvimento de uma investigação na área da criança não implica que se tenha que transferir os resultados, de imediato, para o contexto de cuidados. Sobretudo em contextos tão especiais como as unidades de cuidados intensivos neonatais, ou unidades de internamento pediátrico, ou na comunidade, os estudos individuais devem ser analisados e confrontados com outros estudos antes de iniciar mudanças nas práticas (Pearson & Craig, 2002). Nas unidades pediátricas, em que a criança está dependente da família e dos profissionais, não nos deparamos com a preferência do cliente recém-nascido, criança, jovem, mas sim com a preferência da criança/família como cliente de cuidados (Flemming, 2007). A filosofia dos cuidados centrados na família reconhece a família como constante na vida da criança, por isso os profissionais de saúde devem apoiar, respeitar e potenciar as competências da família através do desenvolvimento de parcerias com os pais (Hockenberry & Barrera, 2014) durante o internamento e preparação para a alta.

O modelo JBI tem como finalidade desenvolver programas de investigação consistentes que permitam a articulação entre a investigação em enfermagem e a comunidade internacional de investigadores e enfermeiros investigadores, com o objetivo de melhorar a utilização dos resultados da investigação e consequentemente a saúde individual e coletiva. O modelo tem subjacente o conceito de tomada de decisão informada, pelo que os documentos de boas práticas do JBI têm um carácter informativo e não prescritivo e focam-se na saúde global. Os cuidados globais à criança e família fazem parte de um processo cíclico que começa com questões, dúvidas, preocupações,

indecisões sobre uma intervenção clínica ou sobre um problema e prossegue com a reflexão sobre as práticas no sentido da melhor decisão clínica e resolução de problemas.

A Saúde global é a força mobilizadora dos cuidados de saúde baseados na evidência científica (cf. Figura 2. 1) e parte da suposição de que as decisões apoiadas na evidência exigem avaliação, síntese e transferência de conhecimento para as práticas, o que pressupõe preparação em áreas específicas. A leitura dos resultados da investigação exige da parte dos profissionais conhecimento teórico, metodológico, estatístico e em análise de dados, para que possam interpretar corretamente os resultados da investigação e fazer a sua aplicação nas práticas de cuidados.

The term 'evidence' is used in the model to mean the basis of belief; the substantiation or confirmation that is needed in order to believe that something is true. Health professionals seek evidence to substantiate the worth of a very wide range of activities and interventions and thus the type of evidence needed depends on the nature of the activity and its purpose. (Pearson, et al., 2005, p. 210).



Figura 2. 1 - Modelo Conceptual de Cuidados de Saúde Baseados na Evidência*.

Adaptado de Pearson et al. (2005), The Model of evidence-based healthcare, pp. 207-215.

*Traduzido e adaptado por: Góis Paixão, M.J. (2013). Aulas de Investigação do curso de Licenciatura em Enfermagem, ESEL.

No modelo são abordados quatro tipos de evidência, a da Viabilidade, a da Adequação, a da Significância e a da Eficácia (VASE) (*Feasibility, Appropriateness, Meaningfulness, Effectiveness* (FAME)). São identificadas como fontes de evidência o discurso, a experiência e a investigação (cf. Figura 2.2), em que esta engloba estudos qualitativos e quantitativos (e.g., indo do estudo de caso ao ensaio clínico aleatorizado).

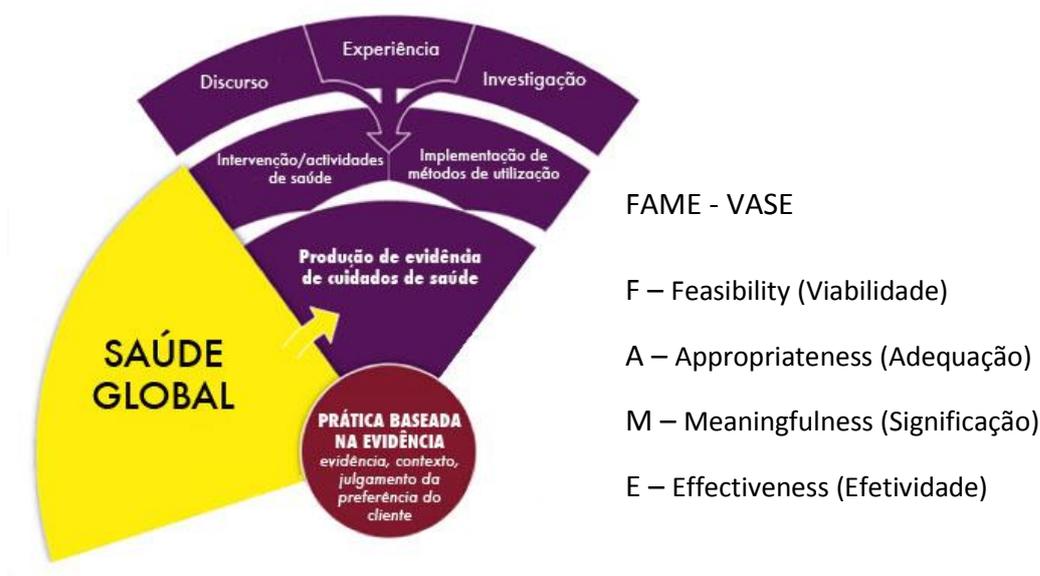


Figura 2.2 - Produção de Evidência e Cuidados de Saúde*.

Adaptado de Pearson et al. (2005), *The Model of evidence-based healthcare*, pp. 207-215.

A *Evidência de Viabilidade* permite verificar se uma intervenção ou atividade é prática e praticável, ou seja, exequível do ponto de vista prático. A viabilidade clínica alude a uma intervenção ou atividade física ou não, cultural ou financeiramente exequível ou possível no contexto específico de cuidados.

A *Evidência de Adequação* permite verificar se uma atividade ou intervenção se adequa a uma situação específica. A adequação clínica espelha a relação entre uma atividade ou intervenção e o seu contexto de cuidados.

A *Evidência de Significação* permite verificar se uma intervenção ou atividade é positivamente experienciada pelo cliente. A significação considera opiniões, valores, pensamentos, crenças e interpretações dos clientes associada à sua experiência pessoal.

*Traduzido e adaptado por: Góis Paixão, M.J. (2013). Aulas de Investigação do curso de Licenciatura em Enfermagem, ESEL.

A *Evidência de Efetividade* permite verificar se uma intervenção, quando utilizada de forma adequada, atinge o efeito para a qual foi planeada. A relação entre a intervenção e os resultados clínicos diz respeito à eficácia clínica (Pearson et al., 2005).

A produção de evidência científica em cuidados de saúde está intimamente ligada aos cuidados ao cliente, à investigação científica na área e à sua divulgação. Os profissionais de saúde devem ser os grandes produtores de conhecimento nesta área, pelo que a sua formação em métodos e técnicas de pesquisa são fundamentais para que, através da investigação, colaborem na melhoria de cuidados ao cliente.

Ao longo do processo de investigação o investigador torna-se perito em dois tipos de teorias. A teoria de apoio ao objeto de estudo, ou teoria principal ou de referência, permite-nos fazer antecipações dos resultados e colocar questões e hipóteses. A teoria de apoio aos métodos e técnicas de investigação, ou teoria auxiliar de pesquisa, permite ao investigador tornar-se perito nos métodos e técnicas de investigação. Ligados aos métodos estão dois procedimentos fundamentais, a construção ou adaptação do instrumento de recolha de dados e as variáveis que o compõem e o tipo de análise mais adequada (qualitativa ou quantitativa) com suporte ou não da estatística e *softwares*. Caso o investigador tenha optado por uma abordagem quantitativa, vai necessitar de usar como ferramenta a Estatística. Este pressuposto pode trazer alguns problemas associados. Altamiro da Costa Pereira diz que, apesar do progresso na área da saúde, existe alguma dificuldade na utilização de uma ferramenta fundamental como a Estatística, o que tem sido comprovado na prática:

(...) ao longo de mais de um século de descobertas e aperfeiçoamentos metodológicos, muitos dos que mais dela precisam continuam a ter dificuldades tanto em a compreender como, sobretudo, em a utilizar (...) (Costa Pereira in Prefácios; Oliveira, 2009).

Acrescenta ainda que, muito foi feito desde Florence Nightingale (enfermeira, investigadora e analista de dados quantitativos, século XIX) a Sir Austin Bradford Hill (epidemiologista e estaticista, século XX). Muitos investigadores deram contributos para facilitar a aplicação e interpretação da metodologia científica e da estatística. Importa, por isso, que os profissionais de saúde se capacitem para trabalhar em equipas de investigação e que, para tal, adquiram as competências necessárias à utilização da

metodologia científica, à interpretação dos resultados da investigação desenvolvida e façam a sua melhor síntese e a sua aplicação na prática de cuidados.

No modelo JBI, a síntese da evidência apresenta-se como uma metodologia de pesquisa que permite analisar a investigação desenvolvida sobre um determinado problema ou área temática salientando teorias, metodologias, instrumentos, resultados e conclusões (revisão sistemática da literatura e meta-análise). A revisão sistemática da literatura identifica todos os estudos que respondem aos critérios de elegibilidade, permite a apreciação da validade dos resultados e faz a apresentação sistemática e síntese das características e dos resultados dos estudos incluídos. Mesmo quando a existência de evidência proveniente da investigação é limitada ou inexistente, as revisões sistemáticas sumarizam a evidência possível sobre um tópico específico, fornecendo ferramentas para a tomada de decisão clínica, assim como identificando as necessidades de novas investigações (JBI, 2001). A meta-análise utiliza vários métodos estatísticos para combinar os resultados de múltiplos estudos de forma a produzir conclusões mais fiáveis. Faz uma análise estatística dos resultados provenientes de diferentes estudos, examinando as fontes das diferenças nos resultados entre estudos, e terminando num sumário quantitativo dos resultados (Porta, 2008).

Systematic reviews and meta-analyses are used to synthesize the available evidence for a given question to inform policy, as in (...) medicine, social science, business, ecology, and other fields. (...) Systematic reviews and meta-analyses can play a role in designing new research. (Borenstein et al., 2009, p. xxv).

A transferência da evidência diz respeito ao ato de transferência do conhecimento, de forma global, para os profissionais de saúde, instituições e sistemas de saúde:

Evidence transfer is seen to involve more than disseminating or distributing information and to include careful development of strategies that identify target audiences – such as clinicians, managers, policy-makers and consumers – and designing methods to package and transfer information that is understood and used in decision-making. (Pearson, et al., 2005, p. 213).

O modelo JBI apresenta três elementos fundamentais de transferência da evidência, através de sistemas organizacionais e de equipa da saúde, nomeadamente: a educação e

formação, a divulgação da informação e a transferência do conhecimento. Para que o conhecimento seja transferido para os profissionais é fundamental que eles tenham acesso à informação, façam formação contínua e tenham suporte nos contextos de cuidados e nas organizações, pois qualquer mudança nas práticas requer envolvimento dos profissionais e dos gestores. São vários os estudos que descrevem as barreiras às práticas baseadas na evidência, salientando que estas são de ordem individual, organizacional e dos contextos das práticas. As questões de ordem individual salientam a falta de tempo para fazer investigação ou ler os resultados da investigação, não dominar as metodologias de investigação e, como tal, serem incapazes de avaliar, por exemplo, a qualidade dos instrumentos (escalas), das técnicas de análise, incapacidade para avaliar a qualidade dos estudos, incapacidade para elaborar projetos de investigação e fazer uma análise estatística. Relativamente às questões organizacionais, os profissionais consideram que a maioria das organizações não tem uma cultura facilitadora destes processos. Nos serviços, sentem falta de cooperação dentro da equipa de saúde, dificuldade na implementação de resultados da investigação e por vezes, devido à falta de autoridade, não conseguem implementar a mudança (Kajermo et al., 2010; Cumming et al., 2010; MacGuire, 2006). A utilização da evidência tem sido fortemente influenciada por fatores, como recursos disponíveis, perícia e formação do prestador de cuidados, preferências do cliente e investigação disponível. Quando a evidência sugere o uso de uma intervenção específica e os profissionais de saúde desejam implementar tal intervenção, isso requer processos de planeamento e de tomada de decisão organizacionais que implicam a medição e a informação. Estes dois aspetos são os alicerces da criação de indicadores de saúde que permitem comparar resultados e assim melhorar os padrões de qualidade dos cuidados e qualidade dos serviços de saúde. Pearson, Jordan & Munn (2012), reforçam estes aspetos através do conceito de “translação do conhecimento”, para os autores, a translação do conhecimento projeta garantir que em todos os níveis de decisão em saúde, os consumidores, os clientes, os profissionais, os gestores e os decisores políticos estão conscientes do acesso e uso da evidência científica que suporte a tomada de decisão em saúde.

No capítulo seguinte o enfoque está no conceito de medida e especificamente no formato de medida ordinal e nas técnicas de análise de dados com este tipo de variáveis.

2.3 Referências Bibliográficas

- Agresti, A. (2002). *Categorical Analysis* (2nd ed.). New Jersey: John Wiley & Sons, Inc.
- Almeida, L.S. & Freire, T. (2003). *Metodologia da Investigação em Psicologia e Educação*. Braga: Psiquilíbrios.
- Als, H. (2004). Early Experience Alters Brain Function and Structure. *Pediatrics*, 113(4), 845-857. <http://www.pediatrics.org/cgi/content/full/113/4/846>
- Als, H. (1982). Toward a synactive theory of development: promise for the assessment and support of infant individuality. *Infant Mental Health Journal*, 3(4), 229-243.
- Als, H., Tronick, E.C., Lester, B.M., & Brazelton, T.B. (1979). The Brazelton Neonatal Behavioral Assessment Scale, In J. Osofsky ed. *Handbook of Infant Development*. New York: Wiley, 185-216.
- Ambuel, B., Hamlet, K.W., Marx, C.M., & Blumer, J.L. (1992). Assessing distress in pediatric intensive-care environments, Comfort Scale. *Journal of Pediatric Psychology*, 17, 95-109.
- Anastasi, A. (1990). *Psychological Testing* (6th ed.). New York: Macmillan Publishing Company.
- Andrada, G., Folha, T., Calado, E., Gouveia, R., & Virella, D. (2009). *Paralisia Cerebral aos 5 anos de idade em Portugal. Crianças com Paralisia Cerebral nascidas em 2001*. Coimbra: Federação das Associações Portuguesas de Paralisia Cerebral.
- Askin, D.F., & Wilson, D. (2014). *Recém-nascido de Alto Risco e Família*, In Marilyn J. Hockenberry & M. Wilson. D. Wong: *Enfermagem da Criança e do Adolescente*. Loures: Lusociência.
- Bakewell-Sachs, S. (2007). Near-term/late preterm infants. *Newborn and Infant Nursing Reviews*, 7(2), 67-71.
- Batalha, L.M.C., dos Reis, G.M.R., Costa, L.P.S., Carvalho, M.D.R., Miguens, A.P.M. (2009). Adaptação cultural e validação da reprodutibilidade da versão Portuguesa da escala de dor Face, Legs, Activity, Cry, Consolability (FLACC) em crianças. *Revista Referência*, II Série, 10, 7-14.
- Bayley, N. (1933). *The California first-year mental Scale*. Berkeley: University of California Press.
- Bayley, N. (2005). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: Psychological Corporation.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development, Administration Manual* (3rd ed.). San Antonio: Psychological Corporation.
- Bear, L.A., & Ward-Smith, P. (2006). Interrater Reliability of the COMFORT Scale. *Pediatric Nursing*, 32(5), 427-434.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Beck, A.T., Steer, R.A., & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77-100.
- Bellman, M., Lingam, S., & Aukett, A. (2003). *SGS II. Escala de Avaliação das Competências no Desenvolvimento Infantil – dos 0 aos 5 anos*. Lisboa: CEGOC – TEA.
- Birleson, P. (1981). The Validity of Depressive Disorder in Childhood and the Development of a Self-Rating Scale: A Research Report. *Journal of Child Psychology and Psychiatry*, 22, 73-88.
- Birleson, P., Hudson, I., Grey-Buchanan, D., & Wolff, S. (1987). Clinical Evaluation of a Self-Rating Scale for Depressive Disorder in Childhood (Depression Self-Rating Scale). *Journal of Child Psychology and Psychiatry*, 28, 43-60.

- Byrne, B.M. (1996). *Measuring self-concept across the life-span*. Washington, DC: American Counseling Assoc.
- Blake, D.D., Weathers, F.W., Nagy, L.M., & Kaloupek, D.G. (1995). The development of a Clinician-Administered PTSD scale. *Journal of Traumatic Stress*, 8, 75-90.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley and Sons, Ltd.
- Brazelton, T.B. (1973). Neonatal Behavioral Assessment Scale. *Clinics in Developmental Medicine*, 50. London: Willian Heinemann Medical Books.
- Campbell, S.K., Kolobe, T.H.A., Osten, E.T., Lenke, M., & Girolami, G.L. (1995). Performance Construct Validity of the Test of Infant Motor. *Physical Therapy*, 75, 585-596.
- Campbell, N.C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., Guthrie, B., Lester, H., Wilson, P., & Kinmonth, A.L. (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal*, 3, 334(7591), 455-459.
- Cipriano, P.F., & Burns, S.M. (2009). The Advanced Practice Nurse and Research, In Lucille A. Joel, *Advance Practice Nursing: Essential for Role Development*. Philadelphia: F.A. Davis Company, 295-308.
- Cochrane Collaboration (2011). Retrieved July 05, 2013, from:
<http://www.cochrane.org/resources/general.htm>.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396.
- Collignon, P., Giusiano, B., Boutin, A.M., & Combe, J.C. (1997). Utilisation d'une échelle d'hétéro-évaluation de la douleur chez le sujet sévèrement polyhandicapé. *Douleur et Analgésie*, 1, 27-32.
- Collignon, P., Giusiano, B., Porsmoguer, E., Jimeno, M.T., & Combe, J.C. (1995). Difficultés du diagnostic de la douleur chez l'enfant polyhandicapé. *Annales de Pédiatrie*, 42(2), 123-126.
- Connolly, J. (1989). Social self-efficacy in adolescence: Relations with selfconcept, social adjustment, and mental health. *Canadian Journal of Behavioral Science*, 21, 258-269.
- Coplan, J., & Gleason, J.R. (1993). Test-retest and inter-observer reliability of the Early Language Milestone Scale, 2nd Ed. *Journal of Pediatric Health Care*, 7(5), 212-219.
- Coplan J. (1985). Evaluation of the child with delayed speech or language. *Pediatric Annals*, 14, 203-8.
- Craig, J.V. (2002). Skills for evidence-based practice, In Jean V. Craig, Rosalind L. Smyth (ed.). *The Evidence-Based Practice: Manual for Nurses*. London: Churchill Livingstone.
- Culham, E. (1998). Evidence based practice and professional credibility (Editorial), *Physiotherapy Theory and Practice*, 14, pp. 65-67.
- Cullen, L., & Adams, S. (2010). An Evidence-Based Practice Model, *Journal of PeriAnesthesia Nursing*, 25(5), 307-310.
- Cummings, G.G., Hutchinson, A.M., Scott, S.D., Norton, P.G., & Estabrooks, C.A. (2010). The relationship between characteristics of context and research utilization in a pediatric setting. *BMC Health Services Research*, 10, 168-178.
- Curado, M.A.S., Teles, J., Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, 446-457.
<http://www.um.es/eglobal>.
- Darcy, A.E. (2009). Complications of the late preterm infant. *Journal of Perinatal & Neonatal Nursing*, 23(1), 78-86.
- Dawes, M.G. (1996). On the need for evidence-based general and family practice. *Evidence-Based Medicine*, 1, 68-69.

- Debillon, T., Zupan, V., Ravault, N., Magny, J-F., & Dehan, M. (2001). Development and initial validation of the EDIN scale, a new tool for assessing prolonged pain in preterm infants. *Archives of Disease in Childhood. Fetal and Neonatal*, 85, 36-41.
- DiCenso, A., Guyatt, G., & Ciliska, D. (2005). *Evidence-based nursing: A Guide to Clinical Practice*. St. Louis: Elsevier Mosby.
- Dickenson-Hazard, N. (2002). Evidence-based practice: the right approach. *Leadership*, 28(2), 1-6.
- Direção Geral de Saúde (DGS), Ministério da Saúde (2013). *Programa Nacional de Saúde Infantil e Juvenil*. Lisboa: DGS.
- Direção Geral de Saúde (DGS), Ministério da Saúde (2011). *Escala de Braden: Versão adulto e Pediátrica (Braden Q)*. Retrieved August 09, 2013, from: <http://www.dgs.pt/upload/membro.id/ficheiros/i015800.pdf>.
- Dubowitz L.M.S., & Dubowitz V. (1981). *The neurological assessment of the preterm and full term newborn infant*. Clinics in Development Medicine, vol. 79, London, UK: Heinemann.
- Dubowitz, L., Ricci, D., & Mercuri, E. (2005). The Dubowitz Neurological Examination of the Full-Term Newborn. *Mental Retardation And Developmental Disabilities Research Reviews*, 11, 52-60.
- Eaves, L.C., Wingert, H., & Ho, H.H. (2006). Screening for autism. *Autism*, 10(3), 229-242.
- Ekbäck, M., Benzein, E., Lindberg, M., & Årestedt, K. (2013). The Swedish version of the multidimensional scale of perceived social support (MSPSS) a psychometric evaluation study in women with hirsutism and nursing students. *Health and Quality of Life Outcomes*, 11(1), 168.
- Engle, W.A. (2006). A recommendation for the definition of “late preterm” (near-term) and the birth weight-gestational age classification system. *Seminars in Perinatology*, 30(1), 2-7.
- Engle, W.A., Tomashek, K.M., Wallman, C., Committee on Fetus and Newborn, American Academy of Pediatrics (2007). Late preterm infants: a population at risk. *Pediatrics*, 120(6), 1390-1401.
- European Union Research Advisory Board (EURAB) (2004). *Interdisciplinarity in Research*. Stanford: Stanford University.
- Ferreira, P., Miguéns, C., Gouveia, J., & Furtado, K. (2007). *Risco de Desenvolvimento de Úlceras de Pressão: Implementação Nacional da Escala de Braden*. Coimbra: Lusociência.
- Fineout-Overholt, E., & Melnyk, B. (2005). Building a Culture of Best Practice. *Nurse Leader*, 3(6), 26-30.
- Flemming, K. (2007). The knowledge base for evidence-based nursing: a role for mixed methods research? *Advances in Nursing Science*, 30(1), 41-51.
- Folio, M.R., & Fewell, R.R. (2000). *Peabody Developmental Motor Scales (PDMS-2)* (2nd ed). Austin, Tex: PRO-ED, Inc.
- Frankenburg, W.K., & Dodds, J.B. (1967). The Denver Developmental Screening Test. *Journal of Pediatrics*, 71, 181-91.
- Frankenburg, W.K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics*, 89, 91-97.
- Gesell, A. (1940). *The First Five Years of Life*. London: Methuen.
- Gesell, A., & Amatruda, C.S (1947). *Developmental Diagnosis* (2nd ed.). New York: Hoeber.
- Gibbins, S., Hoath, S., Coughlin, M., Gibbins, A., & Franck, L. (2008). The universe of developmental care: a new conceptual model for application in the neonatal intensive care unit. *Advances in Neonatal Care*, 8(3), 141-147.

- Gomes-Pedro, J. (1986). Competências e comportamentos do recém-nascido: modo e significado da sua avaliação. *Acta Médica Portuguesa*, 7, 207-214.
- Graham, F.K. (1956). Behavioral differences between normal and traumatized newborns. The test procedures. *Psychological Monographs: General and Applied*, 70(20), 1-16.
- Graham, F.K., Matarazzo, R.G., & Caldwell, B.M. (1956). Behavioral differences between normal and traumatized newborns. Standardization, Reliability, and Validity. *Psychological Monographs: General and Applied*, 70(21), 17-33.
- Granger, C.V., Hamilton, B.B., Keith, R.A., Zielesny, M., & Sherwin, F.S. (1986) Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation*, 1, 59-74.
- Grégoire, M.C., & Finley, A.G. (2008). Doctor, I Think my Baby is in Pain: The assessment of infants' pain by health professionals. *Journal of Pediatric*, 84, 6-8.
- Griffiths, R. (1954). *The Abilities of Babies: A Study in a Mental Measurement*. London: University of London Press.
- Griffiths, R. (1967). *The abilities of babies: A Study in a Mental Measurement* (4th ed.). London: University of London Press.
- Griffiths, R. (1970). *The abilities of young children. A comprehensive System of Mental Measurement for the First Eight Years of Life*. London: Child Development Research Center.
- Guttman, L. (1950). The basis for Scalogram analysis. In Stouffer et al. *Measurement & Prediction, The American Soldier* (Vol. 4). New York: Wiley.
- Guyatt G.H., Oxman A.D., Vist G., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J., for the GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336, 924-926.
- Hair Jr., J.F., Black, W.C., Babin, B.J., Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hallberg, I.R. (2009). Moving nursing research forward towards a stronger impact on health care practice? *International Journal of Nursing Studies*, 46(4), 407-412.
- Hallberg, I.R. (2006). Challenges for future nursing research: Providing evidence for health-care practice. *International Journal of Nursing Studies*, 43, 923-927.
- Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, 53, 87-97.
- Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver: University of Denver Press.
- Harter, S. (1988). *Manual for the Adolescent Self-Perception Profile*. Denver: University of Denver Press.
- Hicks, C.L., Von Baeyer, C.L., Spafford, P.A., Van Korlaar, I, Goodenough, B. (2001). The Faces Pain Scale-Revised: Toward a common metric in pediatric pain measurement. *Pain*, 93(173), 83-86.
- Hill, M.M., & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Hockenberry, M.J., & Barrera, P. (2014). *Perspetivas de Enfermagem Pediátrica*, In Marilyn J. Hockenberry & M. Wilson. D. Wong: *Enfermagem da Criança e do Adolescente*. Loures: Lusodidata.
- Huffines, B., & Lodgson, M.C. (1997). The Neonatal Skin Risk Assessment Scale for predicting skin breakdown in neonates. *Issues Comprehensive Pediatrics Nursing*, 20, 103-114.
- Hummel, P. (2009). *N-PASS Neonatal pain, agitation and sedation scale*. [16 de Setembro de 2012]; Disponível em: www.n-pass.com/assessment_table.html.
- Hunt, M. (1987). The process of translating research findings into nursing practice, *Journal of Advanced Nursing*, 12, 101-110.

- International Council of Nurses (ICN) (1999). *ICN Position Statement on Nursing Research*.
- Joanna Briggs Institute (JBI) (2001). *An Introduction to Systematic Review. Changing Practice: evidence-based practice information sheets for health professionals*. Supp. 1, 1-6.
- Kajermo, K.N., Boström, A-M., Thompson, D.S., Hutchinson, A.M., Estabrooks, C.A., & Wallin, L. (2010). The Barriers scale. The barriers to research utilization scale: A systematic review. *Implementation Science*, 5(1), 32-54.
- Kaufman, A.S. (1990). *Assessing Adolescent and Adult Intelligence*. Boston: Allyn & Bacon. (Revised 2002 and edited in 2006 with Elizabeth Lichtenberger, published by Wiley & Sons).
- Kaufman, A.S. (1994). *Intelligent Testing with the WISC-III*. New York: John Wiley & Sons.
- Keith, R., Granger, C., Hamilton, B., & Sherwin, F. (1984). *Functional Independence Measure - FIM. Uniform Data System for Medical Rehabilitation*. State University of New York: Buffalo.
- Keith, R.A., Granger, C.V., Hamilton, B.B., & Sherwin, F.S. (1987). The Functional Independence Measure: A new tool for rehabilitation. *Advanced Clinical Rehabilitation*, 1, 6-18.
- Kolobe, T.A., & Bulanda, M. (2006). Diagnostic accuracy and consistency of the Alberta Infant motor Scale in a longitudinal sample. *Pediatric Physical Therapy*, 18, 76-77.
- Korner, A.F., Kraemer, H.C., Read, E.P., Forrest, T., Dimiceli, S., & Thom, V.A. (1987). A methodological approach to developing an assessment procedure for testing the neurobehavioral maturity of preterm infants. *Child Development*, 58(6), 1478-1487.
- Korner, A.F., Brown, J.V., Thom, V.A., & Constantinou, J.C. (2000). *The neurobehavioral assessment of the preterm infant. Manual revised* (2nd ed.). Van Nuys, CA: Child Development Media.
- Kovacs, M. (1992). *Children's Depression Inventory (CDI)*. New York: Multi-health Systems, Inc.
- Larrabee, J.H. (2009). *Nurse to nurse: Evidence-based practice*. New York: McGraw-Hill.
- Lawrence, J., Alcock, D., McGrath, P., MacMurray, S.B., & Dulberg, C. (1993). The development of a tool to assess neonatal pain. *Neonatal Network*, 12(6), 59-66.
- Leal, I. & Marôco, J. (2010). *Avaliação em Sexualidade e Parentalidade*. Porto: Legis, Ed.
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1-50.
- MacGuire, J.M. (2006). Putting nursing research findings into practice: Research utilization as an aspect of the management of change. *Journal of Advanced Nursing*, 53(1), 65-74.
- Machado, M.C. (2008). *Saúde Infantil: uma história de sucesso*. In Sakelarides C., Alves M.V., Ed. Saúde e Inovação. Do renascimento aos dias de hoje. Lisboa, Edição Gradiva, 197-203.
- Machado, M.C., Alves, M.I., Couceiro, M.L. (2011). Saúde Infantil e Juvenil em Portugal: Indicadores do Plano Nacional de Saúde. *Acta Pediátrica Portuguesa*, 42(5), 195-204.
- Merker, S.I., Voepel-Lewis, T., Shayevitz, J.R., & Malviya, S. (1997). The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing*, 23(3), 293-297.
- Mesa do Colégio da Especialidade de Enfermagem da Saúde Infantil e Pediátrica – Ordem dos Enfermeiros (MCEESIP-OE) (2010). Parecer n°04/2010: *Escala de Avaliação das Competências de Desenvolvimento Infantil dos 0-5 anos*, Retrieved July 05, 2014, from: http://www.ordemenfermeiros.pt/documentos/Documents/Parecer%2004_MCEESIP_18.11.2010.pdf.

- Nilsson, S., Finnström, B., & Kokinsky, E. (2008). The FLACC behavioral scale for procedural pain assessment in children aged 5-16 years. *Paediatric Anaesthesia*, 18(8), 767-774.
- Oliveira, A.G. (2009). *Bioestatística, Epidemiologia e Investigação. Teoria e Aplicações*. Lisboa: Lidel, Ed. Técnicas.
- Ordem dos Enfermeiros (OE) (2006). *Investigação em Enfermagem: Tomada de posição*. Retrieved July 05, 2013, from:
http://www.ordemenfermeiros.pt/tomadasposicao/Documents/TomadaPosicao_26Abr2006.pdf.
- Ordem dos Enfermeiros (OE) (2010). *Guias orientadores de boa prática em enfermagem de saúde infantil e pediátrica*. Lisboa, Retrieved July 05, 2014, from:
http://www.ordemenfermeiros.pt/publicacoes/Documents/GuiasOrientadores_BoaPratica_Saude Infantil_Pediatria_volume1.pdf.
- Pasquali, L. (2010). *História dos instrumentos psicológicos*. In Luiz Pasquali e Colaboradores, Instrumentação Psicológica. Fundamentos e Prática. Porto Alegre: Artmed.
- Pearson, A., Jordan, Z., & Munn, Z. (2012). Translational Science and Evidence – Based Healthcare: A clarification and reconceptualization of how knowledge is generated and used in healthcare. *Nursing Research and Practice*, 1, 1-6.
- Pearson, A., Wiechula, R., Court, A., & Lockwood, C. (2005). The Model of evidence-based healthcare. *International Journal of Evidence Based Health*, 3, 207-215.
- Pearson, M., & Craig, J.V. (2002). Evidence-based practice in nursing. In J.V. Craig & R.L. Smith. *The evidence-based practice manual for nurses*. New York: Churchill Livingstone, 3-20.
- Peixoto, F., & Mata, L. (1983). Efeitos da Idade, Sexo e Nível Sócio-Cultural no Auto-Conceito. *Análise Psicológica*, 3(11), 401-413.
- Piper, M.C., Darrah J. (1994). *Motor assessment of the developing infant*. EUA, WB: Saunders Company.
- Porta, M. (2008). *A Dictionary of Epidemiology* (5th ed.). Oxford: Oxford University Press.
- Provost, B., Crowe, T., & McClain, C. (2000). Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales in two-year old children. *Physical Occupational Therapy in Pediatrics*, 20, 5-18.
- Provost, B., Heimerl, S., McClain, C., Kim, N., Lopez, B., & Kodituwakku, P. (2004). Concurrent validity of the Bayley Scales of Infant Development II. Motor Scale and the Peabody Developmental Motor Scales-2 in children with developmental delays. *Pediatric Physical Therapy*, 16(3), 149-156.
- Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54(4), 587-599.
- Ribeiro, J.P. (2007). *Avaliação em Psicologia da Saúde*. Lisboa: Climepsi.
- Richards, D., Lawrence, A., & Sackett, D.L. (1997). Bringing an evidence-base to dentistry. *Community Dentist Health*, 14, 63-65.
- Robins, D.L. (2008). Screening for Autism Spectrum in primary care settings. *Autism*, 12(5), 537-556.
- Robins, D.L., Fein, D., Barton, M.L., & Green, J.A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive development disorders. *Journal of Autism and Development Disorders*, 31(2), 131-144.
- Rocha, S.R., Dornelas, L.F., Magalhães, L.C. (2013). Instrumentos utilizados para avaliação do desenvolvimento de recém-nascidos pré-termo no Brasil: revisão da literatura. *Cadernos de Terapia Ocupacional*, 21(1), 109-117.
- Rosenberg, M. (1962). Manual of the Rosenberg self-Esteem Scale.
www.bsos.umd.edu/socy/Research/rosenberg.htm

- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosenblith, J.F., & Lipsitt, L.P. (1959). Interscorer agreement for the Graham behavior test for neonates. *The Journal of Pediatrics*, 54(2), 200-205.
- Rosenblith, J.F. (1961). The Modified Graham Behavior Test for Neonates: Test-Retest Reliability, Normative Data, and Hypotheses for Future Work. *Biology of Neonate*, 3, 174-192.
- Russell, D.J., Rosenbaum, P.L., Avery, L.M., & Lane, M. (2002). *Gross Motor Function Measure (GMFM-66 & GMFM-88) user's manual*. Cambridge: Cambridge University Press.
- Sackett, D.L., Straus, S.E., Richardson, W.C., Rosenberg, W., & Haynes, R.M. (2000). *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingstone.
- Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., & Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 13(312), 71-72.
- Secção de Neonatologia e Sociedade de Pediatria do Neurodesenvolvimento da Sociedade Portuguesa de Pediatria (SNSPN-SPP) (2012). *Consenso do Seguimento das Crianças Nascidas de Muito Baixo Peso (<1500 G) e/ou Menos de 32 Semanas Inseridas na Base do Registo Nacional do MBP*. Lisboa: SPP.
- Sehgal, A., & Stack, J. (2006). Developmentally Supportive Care and NIDCAP. *Indian Journal of Pediatrics*, 73(11), 1007-1010.
- Sharma, A., & Cockerill, H. (2014). *Mary Sheridan's From Birth to Five Years: Children's Developmental Progress*, 4th Ed. London: Routledge.
- Sheridan, M.D. (1984). *From birth to five years: children development progress*. Hong Kong: The Nfer-Nelson Publishing Company, 1-74.
- Sheridan, M., Sharma, A., & Cockerill, H. (2007). *From Birth to Five Years: Children's Developmental Progress*, 3rd Ed. New York: Routledge.
- Simões, M.M.M. (2005). *Humanização de cuidados de saúde e educação reflexiva em ensino clínico com recurso ao debate em grupo*. Braga: Universidade Católica Portuguesa. Dissertação de Mestrado.
- Smith, T.D., Vezina, M.L., & Samost, M.E. (2009). Mediated Roles: Working Through Other People. In Lucille A. Joel. *Advance Practice Nursing: Essential for Role Development*. Philadelphia: F. A. Davis Company, 211-229.
- Spielberger, C.D. (1966). *Theory and research on anxiety*. In C.D. Spielberger (Ed.), *Anxiety and behavior*, 3-20. New York: Academic Press.
- Spielberger, C.D. (1983). *State-trait anxiety and sports psychology*. In E. Aptsch (Ed.), *Anxiety in sport*, 74-85. Magglingen, Sweden: FEPSAC.
- Spielberger, C.D., Gorsuch, R.L., & Lushere, R.E. (1970). *Manual for the State-Trait Anxiety Inventory – Form X*. Palo Alto: Consulting Psychologists Press.
- Stevens, B., Johnston, C., Petryshen, P., Taddio A. (1996). Premature infant pain profile: Development and initial validation. *Clinical Journal of Pain*, 12(1), 13-22.
- Swanson, J.R., Bradley-Johnson, S., Johnson, C.M., & O'Dell, A.R. (2009). The Cognitive Abilities Scale, Second Edition Preschool Form: Studies of concurrent criterion-related, construct, and predictive criterion-related validity. *Journal of Psychoeducational Assessment*, 27(1), 46-56.
- Thoyre, S.M., Shaker, C.S., & Pridham, K.F. (2007). *The Manual of Early Feeding Skills-EFS*, Chapel Hill, The University of North Carolina.
- Thoyre, S.M., Shaker, C.S., & Pridham, K.F. (2005). The early feeding skills assessment for preterm infants. *Neonatal Network*, 24(3), 7-16.

- Thurstone, L.L. (1928/1966). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Tomashek, K.M., Shapiro-Mendoza, C.K., Davidoff, M.J., & Petrini, J.R. (2007). Differences in mortality between late-preterm and term singleton infants in the United States, 1995-2002. *Journal of Pediatrics*, 151(5), 450-456.
- World Health Organization (WHO) (2008). *The World Health Report 2008: Primary Care Now More Than Ever*. Genève: World Health Organization.
- Worthington, R., & Whittaker, T. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, 34(6), 806-838.
- Zimmerman, I.L., & Castilleja, N.F. (2005). The role of language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities*, 11, 238-246.
- Zimmerman, I.L., Steiner, V.G., & Pond R E. (2002). *Preschool Language Scale*, 4th Ed. San Antonio, TX: The Psychological Corporation.
- Zimet, G.D., Dahlem, N.W., Zimet, S.G., & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52(1), 30-41.
- Zimet, G.D., Powell, S.S., Farley, G.K., Werkman, S., & Berkoff, K.A. (1990). Psychometric characteristics of the multidimensional scale of perceived social support. *Journal of Personality Assessment*, 55, 610-617.

Capítulo III. A Medida e as Técnicas Estatísticas

Nos dois subcapítulos que se seguem são abordados os aspectos relacionados com o conceito de medida e o formato de medida ordinal e as técnicas de análise de variância multivariada: MANOVA paramétrica e MANOVA não paramétrica.

3.1 O Conceito de Medida e as Escalas de Medida

Durante um processo de investigação e, mais especificamente, na fase de análise estatística, o investigador obrigatoriamente depara-se com o conceito de variável e com o pressuposto da sua medição. Marôco (2010a) reforça esta ideia dizendo que o objeto de estudo da estatística são as variáveis e a informação que estas podem fornecer.

A informação e a qualidade dessa informação irão depender da forma como as variáveis são quantificadas e da qualidade da medida e, mais especificamente, do erro experimental associado. Assim, podemos classificar as variáveis em quantitativas e qualitativas. As variáveis quantitativas cujo formato de medida permite a ordenação e quantificação de diferenças entre elas, podendo ser intervalares ou de razão. Como exemplo das variáveis intervalares temos Quociente de Inteligência, Psi 20 e como exemplo das variáveis de razão temos o comprimento do braço, a altura, o perímetro cefálico e o peso, que diferem das anteriores por possuírem zero absoluto (Marôco, 2010a; Hill & Hill, 2009).

As variáveis qualitativas são medidas em escalas que indicam a presença de categorias de classificação dos dados, exaustivas e mutuamente exclusivas, podendo ser nominais (e.g., género, estado civil) ou ordinais (e.g., escalão de risco, nível de satisfação). Nas escalas de medida ordinal, as variáveis (itens) são medidas em classes, entre as quais se pode estabelecer uma relação de ordem (Hill & Hill, 2009).

Na investigação na área das ciências sociais e da saúde, medir é uma das maiores preocupações dos investigadores. Frequentemente há a necessidade de avaliar as opiniões, os comportamentos, as atitudes, as angústias e as crenças dos indivíduos, através de um processo de auto-descrição ou de heteroavaliação, sendo os instrumentos usados compostos por variáveis ordinais não observáveis ou não medidas diretamente. Medir pode ser entendido como a forma de determinação do grau de manifestação de

um atributo mental, de uma faculdade perceptiva ou de uma face do comportamento social, de saúde e psicológico de um indivíduo. Estas manifestações são constructos medidos através de instrumentos (escalas de avaliação que permitem quantificar o não observável) representados por testes psicológicos, por escalas, por inventários, etc.

“O problema é que sem um fator de calibração consensual (gold standard), nas medidas de ciências sociais e humanas nunca conseguimos ter a certeza de que o instrumento está a medir o que é suposto medir, de forma válida, fiável e sensível” (Marôco, 2009, p. 5).

Esta situação é única pelo facto de não se conseguir observar diretamente a variável latente, apenas conseguimos observar as suas manifestações. Enquanto medir características físicas é fácil, visto termos instrumentos padronizados que nos permitem fazer essas medições, o mesmo não podemos dizer quando medimos características ligadas ao comportamento humano, em que temos sempre implícita a opinião do indivíduo, que se transforma no instrumento de medida e como tal pode levar a um erro de medida.

As escalas de avaliação compostas por variáveis qualitativas (itens) vêm a ser utilizadas há longa data com o objetivo de medir grandezas que não são mensuráveis diretamente. A utilização das escalas teve uma grande expansão ao longo do século XX e gerou muita discussão, sobretudo nas questões ligadas à forma de medir. Apesar da problemática da medida, estes instrumentos tiveram uma grande aceitação na área das ciências sociais, desenvolvendo-se durante a primeira metade do século passado e também, na saúde e têm sido amplamente utilizados para medir opiniões, atitudes e comportamentos (Thurstone, 1928, 1954; Likert, 1932; Guttman 1950). Ao analisarmos alguns dos trabalhos dos primeiros autores que estudaram estas questões, verificámos que fazer a análise de variáveis qualitativas não é fácil. Para Thurstone (1928, 1954), as medidas geradas pelo indivíduo estão limitadas pelo conhecimento que o próprio tem de si mesmo e das suas atitudes e pelo que está disposto a fazer transparecer ou a divulgar. Por isso pode existir uma discrepância entre a opinião e a ação e entre o indicador (opinião) e a atitude que dele se infere.

Os estudos de Likert (1932) preconizam um método específico de construção de escalas que utilizam afirmações que permitem que pessoas com diferentes opiniões e

diferentes pontos de vista respondam diferenciadamente. O autor considerou a utilização de cinco categorias (sendo a categoria 3 afetada à posição do indeciso em cada afirmação), considerando o grau de concordância do respondente com as afirmações (itens do instrumento). Os itens de Likert devem apresentar um formato de medida que contenha um número de classes que inclua um ponto neutro e cujos extremos são opostos e simétricos (e.g., Discordo Completamente-1, Discordo-2, Nem discordo nem concordo-3, Concordo-4, Concordo completamente-5). Após estudos de fiabilidade³ e análise dos diferentes itens, o autor sugeriu que a medida de atitude, comportamento ou outra, fosse o resultado da soma das valorações dos itens elegidos (escalas adicionadas). Apesar da sua ampla utilização e da sua aplicabilidade nos mais diferentes estudos, as escalas têm sido rodeadas de alguma polémica no que concerne ao seu tratamento estatístico pelo facto de variáveis com formato de medida ordinal serem tratadas como se fossem intervalares (Pell, 2005).

No que diz respeito ao tratamento estatístico destas variáveis existem duas escolas de pensamento: autores que consideram viável fazer uma análise destas variáveis com testes paramétricos, tratando estas variáveis ordinais como se fossem intervalares e outros que consideram que só deve ser feita uma análise não paramétrica, pois partem do pressuposto de que, medir consiste na afetação de números aos objetos ou aos acontecimentos tendo por base determinadas regras, das quais não fazem parte as escalas ou itens com um formato de medida de tipo Likert (escalas cujo formato de medida não apresenta um ponto neutro) ou de Likert (Stevens, 1946, 1951; Jamieson, 2004). Alguns autores consideram que se não utilizarmos itens individualmente, mas sim somatórios de grupos de itens, podemos dizer que as escalas de *Likert* (conjunto de itens), em oposição aos itens individuais de Likert, não são ordinais em característica e em natureza, mas sim intervalares. Deste modo, podem ser analisados através de estatística paramétrica desde que sejam verificados os seus pressupostos com todos os benefícios do uso de testes mais potentes (Pell, 2005; Carifio & Perla, 2008, 2007).

³ A escala mede o constructo de forma consistente e reprodutível, ou seja, sujeitos com a mesma característica, ou o mesmo sujeito em momentos equivalentes, apresentam o mesmo valor na medida.

Urdan (2005) diz que quando se usam variáveis ordinais, a distância entre as observações não é considerada. O autor reporta um exemplo, que passo a descrever: se quisermos saber quem são as dez pessoas mais ricas na América, o indivíduo mais rico poderá receber o *score* 1, o seguinte o *score* 2, até chegar a 10, sendo que este sistema nos diz qual a posição de cada indivíduo, no entanto não diz qual é a distância (em dólares) entre eles. Ao contrário, as variáveis com escalas intervalares ou de rácio contêm informação relativa aos dois pressupostos, ordem e distância. Num exemplo de altura de indivíduos, ao usar as “polegadas” como unidade de medida para quantificar esta característica, a consistência da medida dá-nos uma informação fidedigna acerca da posição relativa e da distância. Tanto as escalas intervalares como as de rácio usam medidas com iguais distâncias entre cada unidade e as escalas de rácio também incluem o zero absoluto.

Carifio (1976, 1978) diz que se usarmos uma linha com 100 mm com dois a sete pontos ancorados como formato de medida, para afirmações de atitude ou frases derivadas de um diferencial semântico, produzimos dados que empiricamente são conjuntos de itens sujeitos a transformação lineares e em carácter intervalares.

Carifio e Perla consideram ainda que:

“It is, therefore, as the intervalist contend, perfectly appropriate to summarize the ratings generated from Likert scales using means and standard deviations, and it is perfectly appropriate to use parametric techniques like Analysis of Variance to analyze Likert scales” (Carifio & Perla, 2008: 1151).

O que vai ao encontro do preconizado por Rensis Likert, cuja metodologia é uma das mais usadas em diversos campos da investigação e particularmente na área das ciências sociais e da saúde.

Relativamente a esta discussão e tendo por base a leitura dos autores mais conceituados nesta área de investigação, Marôco (2009) e Bollen (1989) referem que no que concerne às variáveis ordinais, resultantes destes instrumentos e desde que tenham pelo menos um formato de medida com cinco pontos, muitos autores as consideram como sendo intervalares ou quase-intervalares partindo do pressuposto que estas são a operacionalização de um constructo subjacente contínuo.

As escalas de avaliação, instrumentos muito usados na investigação e nas práticas de cuidados, aparecem na literatura científica com diferentes formatos, apresentando escalas de medida dos itens que podem oscilar entre dois e dez ou mais pontos. Podem apresentar-se como escalas unidimensionais, formadas por um único conjunto de itens ou dimensão, ou compostas por vários conjuntos de itens ou subescalas ou dimensões. As escalas de avaliação nem sempre apresentam uma construção homogénea. Podemos encontrar um instrumento com duas subescalas com vinte itens cada, um instrumento com cinco dimensões, com dois, três, quatro, cinco ou sete itens, ou outras variações.

A construção deste tipo de instrumentos deve seguir alguns fundamentos no que diz respeito à qualidade e quantidade dos itens (variáveis). Relativamente ao formato de medida podemos encontrar escalas com um número de pontos par ou ímpar: 2 pontos (e.g., 0=Não e 1=Sim), 3 pontos (e.g., 1=Insatisfeito, 2=Nem satisfeito nem insatisfeito, 3=Satisfeito), 4 pontos (e.g., 1=Discordo totalmente, 2=Discordo, 3=Concordo, 4=Concordo totalmente), 5 pontos (e.g., 1=Discordo totalmente, 2=Discordo, 3=Indeciso, 4=Concordo, 5=Concordo totalmente) e 7 pontos (e.g., 1=Nunca, 2=Quase nunca, 3=Algumas vezes, 4=Regularmente, 5=Bastantes vezes, 6=Quase Sempre, 7=Sempre). Alguns autores consideram que se deve ter em atenção o número de possibilidades que o respondente tem e a sensibilidade dos itens, considerando que quanto maior o número de classes de resposta ao item, maior a possibilidade dada ao respondente, o que se vai refletir na sensibilidade dos itens, isto é, na capacidade que o item apresenta para discriminar indivíduos estruturalmente diferentes (Marôco, 2009). Os itens com maior número de classes são, geralmente, mais sensíveis e dão maior garantia de credibilidade aos resultados da análise estatística (Hill & Hill, 2009; Carifio & Perla, 2007).

A pesquisa desenvolvida com a finalidade de relacionar o formato de medida dos itens com a consistência interna dos itens mostra diferentes interpretações desta questão. Há autores que consideram que esta relação não existe (Green & Rao 1970; Jacoby & Matell, 1971) e outros que, contrariamente dizem que esta relação existe

(Krosnick and Fabrigar, 1997), indicando como número ótimo, para obter a validade⁴ e fiabilidade dos itens, o formato de medida entre cinco e sete pontos. Pois a fiabilidade tende a aumentar até aos sete pontos, começando a diminuir depois disso, pelo que os itens com formato de medida de dez ou mais pontos podem não ser a alternativa mais fiável (Alwin & Krosnick, 1991; Dawes, 2008).

Os estudos de simulação, através do Método de Monte Carlo, desenvolvidos por Lozano, García-Cueto e Muñiz (2008) com o objetivo de determinar o número de alternativas de resposta aos itens que maximiza as propriedades psicométricas fundamentais de uma escala (fiabilidade e validade) mostram que o formato de medida entre quatro e sete pontos é uma boa opção.

Paralelamente existe uma discussão ligada ao facto da escala de medida ter ou não ponto central. Alguma literatura diz que este ponto central (ponto neutro) pode ser interpretado como ponto neutro, e como tal, mais facilmente identificado e/ou selecionado pelo respondente ou observador. Alguns autores consideram que, nos formatos com ponto central, este pode ser mais selecionado, mas não interfere mais do que outras características dos itens na relação com o respondente (Presser & Schuman, 1980). Os estudos de Moors (2007) confirmam este pressuposto, pois ao investigarem a utilização de um formato de medida par (seis pontos) e um formato de medida ímpar (cinco pontos) com o mesmo questionário, verificaram que não havia grandes diferenças nos resultados obtidos aquando da utilização do formato par ou ímpar.

No caso de as escalas serem compostas por subescalas ou dimensões, Marôco (2009) considera aconselhável um número entre cinco e vinte itens por subescala ou dimensão (no mínimo devem conter três itens). O não cumprimento deste pressuposto pode trazer problemas de fiabilidade e de validade (Marôco, 2009).

A validade inclui diferentes aspetos que podemos classificar em validade relacionada com o conteúdo⁵, validade relacionada com o critério⁶ e validade de constructo⁷ (Worthington & Whitaker, 2006; Anastasi, 1990).

⁴ A escala mede o que era presumível medir.

⁵ Os itens são representativos do domínio que a escala pretende avaliar e traduz o grau de concordância de um painel de especialistas.

⁶ A operacionalização do constructo está em concordância com critérios previamente definidos, traduz a qualidade da medida do constructo e demonstra-se através da validade concorrente (relação do constructo com outros concorrentes) e preditiva (capacidade para prever características, atitudes).

O número de participantes é muito importante quando se usam determinadas metodologias de análise de dados. Quando se trabalha com escalas de avaliação deve usar-se um número mínimo de observações por cada variável manifesta. A razão entre o número de itens e de sujeitos não deve ser inferior a 5:1, devendo ser superior à razão de 10:1 e idealmente estar na razão de 1:15; pois se as amostras tiverem dimensão muito pequena podem ocorrer erros de estimação (Kahn, 2006; Worthington & Whitaker, 2006).

A heterogeneidade no desenvolvimento destes instrumentos e a dimensão da amostra podem acrescentar problemas à análise de dados. A primeira traz problemas relativos à validade e à fiabilidade da medida que se está a tentar operacionalizar e a segunda pode influenciar a probabilidade de erro e a potência dos testes.

Os investigadores e profissionais que utilizam este tipo de instrumentos devem ter em atenção todos estes aspetos para poderem tomar de forma esclarecida e segura a melhor decisão quando escolhem instrumentos para usar nas suas práticas, quer de investigação, quer clínicas.

3.2 A Análise da Variância Multivariada

Durante o processo de investigação, o investigador tem que tomar decisões metodológicas. Toma decisões sobre instrumentos e as variáveis que os compõem e sobre o tipo de análise a efetuar (univariada ou multivariada) e os testes estatísticos a utilizar (paramétricos ou não paramétricos).

As técnicas multivariadas são extensões de análises univariadas, cumprindo o pressuposto de que (...) *all the variables must be random and interrelated in such ways that their different effects cannot meaningfully be interpreted separately* (...) (Hair et al., 2010, p. 309).

Relativamente à metodologia estatística a utilizar, a opção pode ser por testes paramétricos ou não paramétricos, sendo uma das exigências de muitos dos testes

⁷ A escala mede o que se propõe medir e é determinada pela validade convergente (os itens de determinado constructo devem estar correlacionados entre si) e discriminante (os itens que definem determinado constructo não estão correlacionados com itens que medem outro constructo).

paramétricos que a distribuição da população ou populações envolvidas seja normal ou normal multivariada (Marôco, 2010a). Para Reis (2001), a importância do estudo da distribuição normal passa pelo facto de muitos fenómenos reais seguirem uma distribuição normal e muitas distribuições amostrais de estatísticas de teste serem normais (ou aproximadamente normais segundo o teorema limite central).

A distribuição normal ou de *Gauss* é uma das distribuições de probabilidade mais importantes e utilizadas em estatística, sendo caracterizada pelos parâmetros fundamentais μ e σ (Marôco & Bispo, 2003; Timm, 2002).

Uma variável aleatória X tem distribuição normal de valor médio μ ($\mu \in \mathbb{R}$) e variância σ^2 ($\sigma > 0$), representa-se por $X \sim N(\mu, \sigma)$, e a sua função densidade de probabilidade é $f(x) = \left(1/\sqrt{2\pi\sigma^2}\right) e^{-[(x-\mu)/\sigma]^2/2}$, para $x \in \mathbb{R}$ (Johnson & Wichern, 2007, p. 149). Esta função densidade de probabilidade define uma curva simétrica em relação ao eixo $x = \mu$ com um máximo de coordenadas $x = \mu$ e $y = (1/\sigma\sqrt{2\pi})$, com pontos de inflexão $x = \mu \pm \sigma$ e com o eixo das abcissas como assíntota e em que a área sob a curva é igual à unidade (Johnson & Wichern, 2007; Marôco, 2010a; Timm, 2002; Reis, 2001).

Subtraindo à variável X o seu valor médio e dividindo pelo seu desvio padrão, obtém-se $Z = (X - \mu)/\sigma$ que tem distribuição normal com valor médio zero e desvio padrão 1, e escreve-se $Z \sim N(0,1)$, ou seja a variável Z tem distribuição normal padrão.

The z value (...) is a measure in standard deviation units of how far a score is from the mean, and is thus referred to as a standard deviation score (X is the specific score, μ is the value of the population mean and σ is the value of the population standard deviation). (Sheskin, 2004, p. 618)

Esta densidade pode ser generalizada para um vetor com p variáveis.

A densidade normal multivariada para o vetor aleatório $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ ou distribuição normal multivariada cujos parâmetros são dados pelo vetor $\boldsymbol{\mu} = E(\mathbf{X})$ e $\boldsymbol{\mu}$ representa o valor esperado do vetor \mathbf{X} ; e a matriz $\boldsymbol{\Sigma} = V(\mathbf{X})$ e representa a matriz de variância-covariância de \mathbf{X} ; e tem como função densidade conjunta:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \text{ com } \mathbf{x} \in \mathbb{R}^p.$$

Esta densidade p -dimensional tem uma notação análoga à densidade normal do caso univariado escrevendo-se $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Johnson & Wichern, 2007; Marôco, 2010a).

A distribuição normal multivariada apresenta como propriedades adicionais: 1) combinações lineares dos componentes de \mathbf{X} são normalmente distribuídas, 2) todos os subconjuntos de componentes de \mathbf{X} têm uma distribuição normal multivariada, 3) covariância nula entre duas componentes de \mathbf{X} implica que as suas distribuições são independentes, 4) as distribuições condicionais dos componentes de \mathbf{X} são normais multivariadas. Pelo facto de os testes paramétricos apresentarem exigências muito específicas quanto ao cumprimento dos seus pressupostos, um dos quais a normalidade, muito tem sido estudado sobre o efeito da violação do pressuposto de normalidade nos resultados destes testes. Marôco (2010a) salienta alguns resultados de investigação (Scheffé, 1959) e resultados de estudos de simulação (Harwell et al., 1992; Refinetti, 1996) que mostram que:

(...) os métodos paramétricos são robustos à violação do pressuposto da Normalidade desde que as distribuições não sejam extremamente enviesadas ou achatadas e que as dimensões das amostras não sejam extremamente pequenas (uma consequência direta do teorema do limite central) (Marôco, 2010a, p. 203)

Um teste diz-se robusto⁸ se não houver alteração da sua probabilidade de erro tipo I (e estiver próxima do nível de significância α fixado *à priori*) e se a sua potência for mantida em níveis adequados, mesmo que os seus pressupostos de aplicação não sejam válidos (Marôco, 2010a; Pontes, 2005, 2001; Rencher, 2002).

A investigação de Pearson (1930, 1931) sobre a potência dos testes paramétricos perante a violação dos seus pressupostos é um exemplo da importância atribuída ao estudo da robustez. Estes estudos começaram por analisar a probabilidade de erro de tipo I e só depois a probabilidade de erro de tipo II e a potência de teste. A simulação estatística é uma das ferramentas mais importantes para fazer este tipo de estudos, possibilitando a comparação da performance dos testes paramétricos e não paramétricos

⁸ Teste robusto produz resultados de “confiança” mesmo quando aplicado em condições diferentes daquelas para que foi deduzido.

(Blair & Higgins, 1985; Lehman, 1975) e facilitando a análise do comportamento da probabilidade do erro de tipo I e tipo II e conseqüentemente da potência do teste. A probabilidade do erro de tipo I (α) é controlada à partida pelo investigador, ao especificar o nível de significância do teste. É comum utilizar como valor critério para a probabilidade de rejeitar uma hipótese nula verdadeira os níveis de significância de 1% (0.01), 5% (0.05) e 10% (0.10).

A probabilidade de erro de tipo II (β) e a potência de teste ($1-\beta$) são mais difíceis de controlar uma vez que dependem do nível de significância, da dimensão da amostra e da dimensão do efeito. A potência de teste apresentará valores tanto mais elevados quanto maiores forem o nível de significância, a dimensão da amostra e a dimensão do efeito (Loureiro & Gameiro, 2011; Marôco, 2010a).

Os testes não paramétricos não exigem o conhecimento da distribuição populacional da variável sob estudo. Só devem ser usados se não se puderem aplicar os testes paramétricos, pois as suas estatísticas de teste são baseadas em contagens, se estamos perante variáveis nominais, ou em ordens, não sendo considerada a grandeza das diferenças entre observações se estamos perante variáveis quantitativas ou ordinais. Embora os testes não paramétricos pareçam menos exigentes e mais fáceis de usar, são mais conservadores, precisam de amostras de maior dimensão para que se detetem efeitos com diferenças significativas (Marôco, 2010a).

Uma das técnicas estatísticas mais utilizadas nas áreas das Ciências Sociais e Ciências da Saúde é a análise da variância multivariada (MANOVA). Esta técnica permite explorar simultaneamente as relações entre variáveis independentes categóricas que definem os grupos, os tratamentos ou os fatores) e duas ou mais variáveis dependentes (quantitativas).

A utilização da MANOVA paramétrica depende da não violação dos pressupostos (distribuição normal multivariada do vetor das variáveis dependentes em cada grupo e homogeneidade das matrizes de variância-covariância dos grupos). Caso não seja possível usar esta técnica paramétrica, podemos aplicar a técnica não paramétrica correspondente, a MANOVA não paramétrica.

3.2.1 MANOVA Paramétrica a Um Fator

Ao estudarmos os conceitos de MANOVA podemos dizer que esta é um extensão da ANOVA e que as duas diferem entre si pelo facto de a ANOVA avaliar as diferenças entre médias para uma só variável critério e a MANOVA fazer a comparação dos vetores de valores médios em situações em que se pretende comparar em simultâneo diversas variáveis critério (Marôco, 2010a; Timm, 2002; Reis, 2001; Jobson, 1992). Jobson (1992) refere ainda que: *ANOVA is a special case of multiple regression, (...) multivariate analysis of variance (MANOVA) can be viewed as a special case of multivariate linear regression* (Jobson, 1992, p. 213).

A utilização das estatísticas da MANOVA em detrimento da ANOVA tem implícita a diminuição da probabilidade de erro. Se o objetivo for comparar p variáveis dependentes em k grupos com base na mesma amostra e se recorrermos a p ANOVAs, a probabilidade do erro tipo I, fixado à partida, não será α mas sim $\alpha \times p$, pelo que a avaliação da potência passa pela estabilização das probabilidades do erro tipo I e de tipo II (Marôco, 2010a; Sheskin, 2007; Rencher, 2002).

Na MANOVA as variáveis dependentes estão a ser observadas em simultâneo, com organização composta e com os efeitos associados a cada uma das variáveis (ponderados pela correlação) no sentido de manter a probabilidade do erro de tipo I igual a α .

Ao efetuarmos várias ANOVAs, assumimos que as variáveis dependentes são ortogonais (i.e., independentes) e como tal que não existe relação entre elas. Ao utilizarmos as técnicas univariadas, se duas ou mais variáveis dependentes estiverem correlacionadas, as diferenças entre os grupos poderão ser subavaliadas ou sobreavaliadas (dependendo da magnitude das correlações e do seu sinal). A aplicação de diversas ANOVAs pode levar a identificar diferenças significativas, que na realidade não existem ou a não as identificar quando elas existem. Os aspetos referidos anteriormente têm feito com que os investigadores das ciências sociais e da saúde considerem a análise de variância multivariada (MANOVA) uma ferramenta poderosa para analisar dados, pois oferece várias vantagens sobre a ANOVA (Finch & Davenport, 2009; Finch, 2005; Tabachnick & Fidell, 2001).

Marôco (2010a) refere ainda que a revisão sistemática que Cochran (1947) fez sobre a investigação nesta área mostrou que a violação da normalidade tem um efeito reduzido na probabilidade de erro de tipo I. Por exemplo, que a ANOVA ou o teste *t de Student* são robustos no que diz respeito à taxa de erro de tipo I mesmo quando as variáveis apresentam enviesamento e/ou achatamento considerável.

Sheskin (2007) acrescenta que: *As a general rule, the multivariate analysis of variance is reasonably robust with respect to assumption violation, especially with a large simple size and an equal number of subjects in each group.* (Sheskin, 2007, p. 1500).

A heterocedasticidade ou violação do pressuposto da homogeneidade da variância interfere com a probabilidade de erro tipo I e de erro de tipo II, no entanto o efeito sobre a probabilidade de erro de tipo I é função das dimensões dos grupos e de estas serem iguais ou desiguais, ou seja, de estarmos perante um delineamento balanceado ou não (Myers & Well, 2003; Sahai & Ageel, 2000).

No caso de delineamentos balanceados, existe consenso na literatura sobre o facto de que os métodos paramétricos serem robustos à violação da homogeneidade da variância. Para que esse facto se observe, o rácio entre a menor e a maior variância na amostra tem que ser inferior na razão de 1:4 ou na razão de 1:9; 2), os grupos não tenham dimensão muito pequena e as médias dos grupos não estejam linearmente relacionadas com as suas respetivas variâncias (Lixe, Keselman & Keselman, 1996, Marôco, 2010a).

No caso de delineamentos não balanceados, (...) *o efeito da violação da homogeneidade de variâncias, depende da relação entre a dimensão dos grupos ou amostras da variância desses grupos. Se as dimensões dos grupos estiverem correlacionadas positivamente com a variância (i.e., quanto maior a dimensão dos grupos maior a variância) a probabilidade de erro tipo I é, geralmente, menor do que o a fixado a priori e o teste é conservador; se a dimensão dos grupos estiver correlacionada negativamente com as variâncias (i.e., quanto menor a dimensão maior a variância) A probabilidade de erro de tipo I é inflacionada e o teste é liberal* (Harwell et al., 1992; Sharma, 1996; cit. por Marôco, 2010a, p. 204-205).

Na MANOVA paramétrica a um fator, pretende-se testar se os vetores de médias, referentes a p variáveis medidas em g grupos, são iguais ou se existem pelo menos, dois vetores de médias que diferem entre si de forma significativa.

Na sua forma matricial temos como modelo da MANOVA a um fator: $\mathbf{X}_{li} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\varepsilon}_{li}$, onde \mathbf{X}_{li} é o vetor aleatório das observações do i -ésimo indivíduo com $i = 1, \dots, n_l$ e $l = 1, \dots, g$, e $\boldsymbol{\mu}$ é o vetor valor médio e $\boldsymbol{\tau}_l$ representa o efeito do nível l do fator

(tratamento) com $\sum_{l=1}^g n_l \boldsymbol{\tau}_l = \mathbf{0}$, e $\boldsymbol{\varepsilon}_{li}$ são os erros onde $\boldsymbol{\varepsilon}_{li} \sim \mathbb{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. A hipótese nula pode ser apresentada como $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \dots = \boldsymbol{\tau}_g = \mathbf{0}$ e a hipótese alternativa é

apresentada como $H_1 : \exists_l : \boldsymbol{\tau}_l \neq \mathbf{0} (l \in \{1, \dots, g\})$, onde $\boldsymbol{\tau}_l = \begin{bmatrix} \tau_{l1} \\ \tau_{l2} \\ \vdots \\ \tau_{lp} \end{bmatrix}$, com $l = 1, \dots, g$ (Johnson

& Wichern, 2007). Representando por $\boldsymbol{\mu}_l = \boldsymbol{\mu} + \boldsymbol{\tau}_l$, as hipóteses podem ser dadas por: $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g$ vs $H_1 : \exists_{l,k} : \boldsymbol{\mu}_l \neq \boldsymbol{\mu}_k (l, k \in \{1, \dots, g\})$.

Os pressupostos sobre a estrutura dos dados para a MANOVA One-Way ou a um fator exigem que cada população tenha uma distribuição normal multivariada e todas as populações tenham uma matriz de variância-covariância $\boldsymbol{\Sigma}$ comum.

De acordo com o modelo, cada componente do vetor de observações \mathbf{X}_{li} satisfaz o modelo univariado e embora os erros das componentes de \mathbf{X}_{li} estejam correlacionados, as matrizes de variância-covariância $\boldsymbol{\Sigma}$ são iguais para todas as populações.

Se particularizarmos o modelo populacional para as observações amostrais temos: $\mathbf{X}_{li} = \bar{\mathbf{X}} + (\bar{\mathbf{X}}_l - \bar{\mathbf{X}}) + (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)$, em que $\bar{\mathbf{X}}$ é o vetor médio amostral (estimativa de $\boldsymbol{\mu}$) e $(\bar{\mathbf{X}}_l - \bar{\mathbf{X}})$ é o efeito do nível l do tratamento (estimativa de $\boldsymbol{\tau}_l$) e $(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)$ são os resíduos (estimativa de $\boldsymbol{\varepsilon}_{li}$). A extensão matricial das somas dos quadrados da ANOVA, a soma dos quadrados devido ao fator ou tratamento (SQF) e a soma dos quadrados do erro (SQE) respetivamente, dá origem à matriz \mathbf{B} , designada por matriz da soma de quadrados e produtos cruzados entre grupos que pode ser apresentada como $\mathbf{B} = \sum_{l=1}^g n_l (\bar{\mathbf{X}}_l - \bar{\mathbf{X}})(\bar{\mathbf{X}}_l - \bar{\mathbf{X}})'$ e à matriz \mathbf{W} denominada como matriz da soma de quadrados e produtos cruzados dentro dos grupos que pode ser representada como $\mathbf{W} = \sum_{l=1}^g \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)' = (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \dots + (n_g - 1)\mathbf{S}_g$ onde

S_l é a estimativa da matriz de variância-covariância do nível l do fator definida por, $S_l = \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)' / (n_l - 1)$. A matriz da soma de quadrados e produtos cruzados totais que representa a variabilidade global representa-se por \mathbf{T} , sendo $\mathbf{T} = \mathbf{B} + \mathbf{W}$ (Marôco, 2010; Manly, 2005; Montgomery, 2001).

O número total de graus de liberdade, $n - 1$, também é particionado na soma de duas parcelas $g - 1$ e $n - g$. No caso da MANOVA não existem quadrados médios nem estatística F , em vez disso consideram-se os determinantes das matrizes \mathbf{T} , \mathbf{B} e \mathbf{W} (cf. Tabela 3. 1). As distribuições amostrais destas estatísticas são uma aproximação à distribuição F . Quanto menor for a estatística de teste, maior a evidência a favor da hipótese nula, ou seja, ausência de efeito do tratamento. As estatísticas de teste mais comuns da MANOVA são:

O lambda de Wilks, $\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}$, é obtido à custa dos determinantes referidos anteriormente e é a estatística mais utilizada,

O traço de Pillai, $P = tr(\mathbf{B}\mathbf{T}^{-1}) = tr(\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1})$, é a soma dos valores próprios (*eigenvalues*) da matriz $\mathbf{B}\mathbf{T}^{-1}$,

O traço de Hotelling, $H = tr(\mathbf{B}\mathbf{W}^{-1})$, é a soma dos valores próprios da matriz $\mathbf{B}\mathbf{W}^{-1}$,

A maior raiz de Roy (R) é o maior dos valores próprios da matriz $\mathbf{B}\mathbf{W}^{-1}$ (Marôco, 2010; Johnson & Wichern, 2007; Manly, 2005).

Tabela 3. 1 - MANOVA para comparação de vetores médios (Notação matricial)

Fonte de variação	Matriz da soma de quadrados	Graus de Liberdade	Determinantes	Λ
Tratamento Grupo/Fator (entre amostras)	$\mathbf{B} = \sum_{l=1}^g n_l (\bar{\mathbf{X}}_l - \bar{\mathbf{X}})(\bar{\mathbf{X}}_l - \bar{\mathbf{X}})'$	$g - 1$	$ \mathbf{B} $	$ \mathbf{W} / \mathbf{T} $
Erro	$\mathbf{W} = \sum_{l=1}^g \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)'$	$\sum_{l=1}^g n_l - g$	$ \mathbf{W} $	
Total	$\mathbf{T} = \sum_{l=1}^g \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}})(\mathbf{X}_{li} - \bar{\mathbf{X}})'$	$\sum_{l=1}^g n_l - 1$	$ \mathbf{T} $	

Adaptado de Johnson e Wichern (2007), Applied Multivariate Statistical Analysis, p. 302.

Johnson e Wichern (2007) referem que os quatro testes parecem ser quase equivalentes para amostras de grande dimensão. Para dimensões amostrais moderadas, verifica-se que os três primeiros testes têm potência semelhante, enquanto a maior raiz de Roy apresenta valores de potência mais elevados quando se verifica uma magnitude de correlação elevada entre as variáveis independentes. O traço de Pillai é mais potente em amostras de pequena dimensão e/ou de diferentes dimensões e variância-covariância heterogéneas (Hair et al., 2010). Se não forem violados os pressupostos da MANOVA, o Lambda de Wilks é o método que apresenta valores de potência mais elevados. Com variáveis fortemente correlacionadas, a maior raiz de Roy é um dos que apresenta maior potência (Newton & Rudestam, 1999; Marôco, 2010a; Johnson, 1998).

Para as situações em que não se deve aplicar a MANOVA paramétrica (e.g., amostras de dimensão muito pequena e violação dos pressupostos dos testes) pode recorrer-se a outras técnicas similares, não paramétricas, como as propostas por Puri e Sen (1974, 1969). Estas propostas serviram de suporte ao estudo de duas estatísticas da MANOVA não paramétricas homólogas das paramétricas; o traço de Pillai aplicado às ordens (Zwick, 1985; Katz & McSweeney, 1980) e o lambda de Wilks aplicado às ordens (Nath & Pavur, 1985). Os autores consideram que a MANOVA não paramétrica pode ser melhor solução relativamente à paramétrica, se forem violados os pressupostos de normalidade multivariada e a homogeneidade das matrizes de variância-covariância.

3.2.2 MANOVA não Paramétrica a um Fator

A revisão da bibliografia tem dado a conhecer o facto de a estatística não paramétrica ser utilizada em grande escala no caso univariado, mas não no caso multivariado (Pontes, 2005). Apesar de terem ocorrido tentativas isoladas, de alguns investigadores, nos anos setenta do século passado, só nos anos noventa se verificou algum desenvolvimento da investigação com testes paramétricos e não paramétricos no caso univariado, no que concerne à comparação e performance dos métodos, no entanto, no caso multivariado, pouco ou nada se investigou (Ittenbach, et al., 1993; Finch, 2005). No final dos anos noventa os testes não paramétricos multivariados começam a ser mais utilizados, sobretudo na investigação agrária, economia, psicologia e sociologia. Pontes e Corrente (2001) e Pontes (2005) referem a importância do uso de técnicas não

paramétricas pelo facto de estas não serem tão restritivas e não serem sensíveis à violação dos pressupostos da normalidade multivariada, simetria e homogeneidade das matrizes de variância-covariância.

Os autores consideram que a utilização destes métodos não exige que a distribuição da variável seja conhecida, uma vez que se baseiam na ordem das observações e não no seu valor. *A base dos testes não paramétricos está na ordenação (ranks) dos dados e não no seu valor intrínseco, e na aleatorização, onde se consideram todas as possíveis permutações (rearranjos) dos dados*” (Pontes e Corrente, 2001, p. 180).

Pontes (2005) considera que a perda na troca do verdadeiro valor da observação pela sua ordem faz aumentar a eficiência e a compreensão dos dados. No caso de amostras de pequena dimensão, o teste de Kruskal-Wallis é mais potente do que a ANOVA e o teste U de Man-Whitney mais potente do que o t de Student para amostras independentes, mas, se estamos perante violação dos pressupostos dos testes paramétricos, e as amostras são de grande dimensão, a ANOVA e o t de Student apresentam valores elevados de potência (Marôco, 2010a).

As simulações descritas por Zimmerman (2000, 2005) demonstram que a probabilidade de erro de tipo I de testes paramétricos, como o t de Student para amostras emparelhadas, é enviesada por variâncias heterogéneas especialmente quando as dimensões das amostras são muito desiguais.

Ao longo dos anos, diversos investigadores têm sugerido que a probabilidade de erro de tipo I em alguns testes não paramétricos como é o caso do U de Mann-Whitney e do teste de Kruskal-Wallis não é influenciada pela heterogeneidade de variâncias. Porém, mais recentemente tem-se verificado que estes testes quando aplicados em amostras de menor dimensão podem apresentar uma probabilidade de erro de tipo I mais elevada, sem igualar, no entanto, a obtida no caso paramétrico. Os testes não paramétricos são normalmente associados à hipótese de igualdade de distribuições nos grupos de tratamento, em contraste com a hipótese de valores médios iguais, que é uma característica dos testes paramétricos. Ao referirem a utilização dos testes não paramétricos ou paramétricos, a maioria dos autores coloca em primeiro plano as potências dos testes, considerando que esta é superior nos testes paramétricos e por isso a probabilidade de rejeitar corretamente H_0 é maior nestes testes do que nos não

paramétricos (Marôco, 2010a). Para o autor, os testes não paramétricos só devem ser usados quando não existe uma alternativa paramétrica, ou seja, quando não é possível homogeneizar a variância, ou normalizar a variável dependente, ou ainda quando a variável dependente é qualitativa.

Os métodos não paramétricos para delineamentos multivariados são aplicados às ordens das observações e, a partir das médias das ordens e das matrizes de covariâncias das ordens, é possível obter uma extensão multivariada da estatística de Kruskal-Wallis. Tal como nos testes paramétricos, para controlo do erro de tipo I deve começar-se por fazer uma análise multivariada e se esta mostrar efeitos significativos passa-se então à análise univariada. Isto porque numa fase inicial, quando se está a testar a igualdade de g distribuições em p variáveis dependentes, pode-se ser facilmente induzido a fazer p testes de Kruskal-Wallis, mas este facto poderá levar ao aumento do erro de tipo I.

A MANOVA não paramétrica a um fator apresenta-se como uma alternativa à correspondente metodologia paramétrica. Como já foi referido, esta técnica é considerada como uma extensão multivariada do teste de Kruskal-Wallis univariado e apenas requer que as variáveis dependentes tenham distribuição de probabilidade contínua $F(\mathbf{X}_1), \dots, F(\mathbf{X}_g)$ (Katz & McSweeney, 1980). A técnica é um procedimento de ordens em que os scores para cada uma das p variáveis são classificados separadamente de 1 para n , com a ordem 1 para a menor observação e a ordem de n para a maior observação. Se determinadas observações estão empatadas, a cada uma destas observações é atribuída a média das ordens das observações. Após este procedimento existem p conjuntos de ordens, um conjunto para cada variável sob estudo.

Para testarmos a igualdade das distribuições em todas as variáveis e grupos usamos a MANOVA não paramétrica e as hipóteses: $H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = \dots = F(\mathbf{X}_g)$ vs. $H_1 : \exists l, l^* : F(\mathbf{X}_l) \neq F(\mathbf{X}_{l^*})$, com $l, l^* \in \{1, \dots, g\}$ onde $F(\mathbf{X}_l)$ é a função de distribuição multivariada de $\mathbf{X}_l = (X_{l1}, X_{l2}, \dots, X_{lp})'$, com $l = 1, \dots, g$. No caso do traço de Pillai aplicado às ordens, a estatística de teste é $\chi^2 = \mathbf{V}'\Sigma_{\mathbf{V}}^{-1}\mathbf{V} \sim \chi_{p(g-1)}^2$ (Marôco, 2010; Katz & McSweeney, 1980). O vetor das médias das ordens \mathbf{V}_j é definido em termos

dos desvios dos primeiros $(g-1)$ grupos de médias das ordens dos respectivos valores esperados sob a validade de H_0 e pode escrever-se como:

$$\mathbf{V}_j = \left[\bar{R}_j^{(1)} - (n+1)/2, \dots, \bar{R}_j^{(g-1)} - (n+1)/2 \right]', \quad j=1, \dots, p; \text{ onde } \mathbf{V}_{(g-1)p \times 1} = (\mathbf{V}'_1, \mathbf{V}'_2, \dots, \mathbf{V}'_p)'$$

e Σ_v é a matriz de variância-covariância do vetor aleatório \mathbf{V} . Os desvios relativos ao grupo g são inferidos dos anteriores $(g-1)$ grupos. Para $\sum_{l=1}^g n_l = n$, o vetor aleatório \mathbf{V} é aproximadamente normal multivariado com matriz de covariâncias Σ_v , de característica máxima $p(g-1)$. Sob H_0 , $E(\mathbf{V}) = \mathbf{0}$, tornando χ^2 uma variável aleatória qui-quadrado central com $p(g-1)$ graus de liberdade. Valores elevados de χ^2 indicam que existem diferenças entre \mathbf{V} e o valor esperado para \mathbf{V} sob a validade de H_0 . Assim a hipótese nula é rejeitada, ao nível de significância α sempre que $\chi^2 \geq \chi^2_{1-\alpha; p(g-1)}$ (Katz & McSweeney, 1980, p. 284).

A metodologia utilizada por Katz e McSweeney (1980) é a estatística χ^2 que pode ser apresentada como $\chi^2 = (n-1)TP_R$, onde n é a dimensão da amostra global e TP_R é o traço de Pillai calculado a partir de ordens das observações (Marôco, 2010a; Zwick, 1985). Rejeita-se H_0 ao nível de significância de α se $\chi^2 \geq \chi^2_{1-\alpha; p(g-1)}$ ou se p -value do teste for menor ou igual a α (p -value $\leq \alpha$).

A outra metodologia multivariada é calculada a partir do Lambda de Wilks aplicado às ordens e é obtido por $LW_R = |\mathbf{W}_R|/|\mathbf{T}_R|$, em que as matrizes \mathbf{W}_R e \mathbf{T}_R são as matrizes derivadas de \mathbf{W} e \mathbf{T} sendo as observações substituídas pelas respectivas ordens. De acordo com a metodologia de “transformação de ordens”, o teste paramétrico é aplicado às ordens e a distribuição da estatística de teste LW_R é aproximada à do Lambda de Wilks paramétrico. *These testing procedure are known to perform better than their classical counterparts especially when the data are from a heavy-tailed distribution* (Nath & Pavur, 1985, p. 298).

Quando $n \rightarrow \infty$ a distribuição de $-\left[n_l - 1 - (p+g)/2 \right] \ln LW_R$ é aproximada à distribuição qui-quadrado com $p(g-1)$ graus de liberdade, em que $\ln LW$ é o logaritmo natural do lambda de Wilks. Os autores consideraram que X_{li} é substituído pela ordem correspondente. Desta forma cada $X_{li}^{(m)}$ é substituído pela ordem correspondente, $R_{li}^{(m)}$, sendo $R_{li}^{(m)}$ a ordem referente ao i -ésimo sujeito, no l -ésimo grupo, em que $l=1, \dots, g$, $m=1, \dots, p$ e $i=1, \dots, n_l$ e $1 \leq R_{li}^{(m)} \leq n$. A média das ordens por

grupo na i -ésima variável pode escrever-se como $\bar{R}_i^{(m)}$, $l=1,\dots,g$ e $i=1,\dots,n_l$ e $\sum_{l=1}^g \sum_{i=1}^{n_l} R_{ii}^{(m)} = n(n+1)/2$. O vetor da média das ordens \mathbf{V}_j nas $j=1,\dots,p$ variáveis dependentes é definido em termos de desvio do primeiro grupo de médias das ordens e respetivos valores esperados e pode escrever-se como: $\mathbf{V}_j^{(m)} = (1/n) \sum_{i=1}^{n_l} R_{ii}^{(m)}$, com $j=1,\dots,p$ e $i=1,\dots,n_l$; onde $\mathbf{V}_l = (\mathbf{V}_l^1, \mathbf{V}_l^2, \dots, \mathbf{V}_l^p)$, definindo-se a matriz de variância-covariância do vetor aleatório \mathbf{V} como $\mathbf{V} = (v_{ij})$, $i, j=1,2,\dots,p$.

A estatística χ^2 pode ser calculada através do lambda (Λ_R) de Wilks aplicado às ordens das observações e escreve-se como $\chi^2 = -(n-1-(p+g)/2) \ln LW_R$ onde n é a dimensão da amostra global (Nath & Pavur, 1985; Todorov & Filzmoser, 2009, 2010; Van Aelst & Willems, 2011). Rejeita-se H_0 ao nível de significância α se $\chi^2 \geq \chi_{1-\alpha; p(g-1)}^2$ ou se $p\text{-value} \leq \alpha$.

3.3 Referências Bibliográficas

- Alwin, D.F., & Krosnick, J.A (1991). The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods and Research*, 20(1), 139-181.
- Anastasi, A. (1990). *Psychological Testing* (6th ed.). New York: Macmillan Publishing Company.
- Blair, R.C., & Higgins, J.J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin*, 97, 119-128.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Series in Probability and Mathematical Statistics. New York: Wiley.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
- Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Carifio, J. (1978). Measuring vocational preferences: ranking versus categorical rating procedures. *Career Education Quarterly*, 3(1), 34-66.
- Carifio, J. (1976). Assigning students to career education programs by preference: scaling preference data for program assignments. *Career Education Quarterly*, 1(1), 7-26.
- Cochran, W.G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 3(1), 22-38.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61-77.
- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions are Violated. *Methodology*, 1(1), 27-38.
- Finch, W.H., & Davenport, T. (2009). Performance of Monte Carlo Permutation and Approximate Tests for Multivariate Means Comparisons with Small Sample Sizes when Parametric Assumptions are Violated. *Methodology*, 5(2), 60-70.
- Gutman, L. (1950). *The basis for Scalogram analysis*. In S.A. Stouffer et al. *Measurement and Prediction: The American Soldier*, Vol. IV. New York: Wiley.
- Green, P.E., & Rao, V.R. (1970). Rating Scales and Information Recovery: How Many Scales and Response Categories to Use? *Journal of Marketing*, 34, 33-39.
- Hair Jr., J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Harwell, M.R., Rubinstain, E.N., Hayes, W.S., & Olds, C.C. (1992). Summarizing Monte Carlo results in methodological research: The one and two-factor fixed effect ANOVA cases. *Journal of Educational Statistics*, 17(4), 315-339.
- Hill, M.M., & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Ittenbach, R.F., Chayer, D.E., Bruininks, R.H., Thrulow, M.L., & Beirne-Smith, M. (1993). Adjustment of young adults with mental retardation in community settings: Comparison of parametric and nonparametric statistical techniques. *American Journal of Mental Retardation*, 97, 607-615.
- Jobson, J.D. (1992). *Applied Multivariate Data Analysis*. New York: Springer.
- Jacoby, J., & Matell, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, VIII, 495-500.
- Johnson, D.E. (1998). *Applied Multivariate Methods for Data Analysts*. Pacific Grove, CA: Brooks/Cole Publishing.
- Johnson, R.A., & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Kahn, J. (2006). Factor Analysis in Counseling Psychology Research, Training, and Practice: Principles, Advances and Applications. *The Counseling Psychologist*, 34(5), 684-718.
- Katz B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Krosnick, J.A., & Fabrigar, L.R. (1997). *Designing rating scales for effective measurement in surveys*. In L.E. Lyberg, P. Biemer, M. Collins and E.D. Leeuw (eds.) *Survey*.
- Krosnick, J.A., & Presser, S. (2010). *Question and Questionnaire Design*. In P.V. Marsden & J.D. Wright, *Handbook of Survey Research*. Emerald Group Publishing, Ltd. Bingley.
- Lehmann, E.L. (1975). *Nonparametrics, statistical methods based on ranks*. San Francisco: Holden-Day.
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 55.
- Lixe, L.M., Keselman, J.C., & Keselman, H.J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of Educational Research*, 66, 579-619.

- Loureiro, L.M.J., & Gameiro, M.G.H. (2011). Interpretação crítica dos resultados estatísticos: para lá da significância estatística. *Revista de Enfermagem Referência*, 3, 151-162.
- Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73-79.
- Manly, B.F.J. (2005). *Multivariate Statistical Methods* (3rd Ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Marôco, J. (2010a). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Marôco, J. (2010b). *Análise de Equações Estruturais: Fundamentos teóricos, Software & Aplicações*. Pêro Pinheiro: ReportNumber.
- Marôco, J. (2009). *Avaliação das qualidades psicométricas de uma escala*. Lisboa: ISPA-IUL. Manuscrito não publicado.
- Marôco, J., & Bispo, R. (2003). *Estatística aplicada às ciências sociais e humanas* (1^a ed.). Lisboa: Climepsi Editores.
- Montgomery, D.C. (2001). *Design and Analysis of Experiments* (5th ed.). New York: John Wiley & Sons, Inc.
- Moors, G. (2007). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42(6), 779-794.
- Myers, J.L., & Well, A.D. (2003). *Research Design and Statistical Analysis*. London: Lawrence Erlbaum Associates.
- Nath, R. (1982). *Empirical investigation of an alternative to one-way multivariate analysis of variance (MANOVA), Technical Report*. Memphis: Department of Management, Memphis State University.
- Nath, R., & Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics and Data Analysis*, 2, 297-315.
- Newton, R.R., & Rudestam, K.E. (1999). *Your Statistical Consultant. Answers to your Data Analysis Questions*. Thousand Oaks, CA: Sage Publications.
- Olson, C.L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 69, 894-908.
- Pearson, E.S. (1930). A further development of tests of normality. *Biometrika*, 22, 239-249.
- Pearson, E.S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- Pell, G. (2005). Use and misuse of Likert Scales. *Medical Education*, 39, 970.
- Pontes, A.C.F., & Corrente, J.E. (2001). Comparações múltiplas não paramétricas para o delineamento com factor de classificação simples. *Revista de Matemática e Estatística*, 19, 179-197.
- Pontes, A.C.F. (2005). *Análise de Variância multivariada com utilização de testes não paramétricos e componentes principais baseados em matrizes de postos*. Tese de Doutoramento. S. Paulo: Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz.
- Presser, S., & Schuman H. (1980). *The Measurement of a Middle Position in Attitude Surveys*. *Public Opinion Quarterly*, 44(1), 70-85.
- Puri, M.L., & Sen, P.K. (1974). *Nonparametric methods in multivariate analysis*. New York, John Wiley & Sons.
- Puri, M.L., & Sen, P.K. (1969). Class of Rank Order Tests for a General Linear Hypothesis. *Annals of Mathematical Statistics*, 40(4), 1325-1343.

- Ramsey, P.H., & Ramsey, P.P. (2013). Pairwise testing of group mean vectors in MANOVA with small samples. *Journal of Statistical Computation and Simulation*, 83(6), 1037-1057.
- Ramsey, P.H., Ramsey, P.P., Hachimine, P., & Andiloro, N. (2011). Robustness, Power and Interpretability of Pairwise Tests of Discriminant Functions in MANOVA. *Journal of Modern Applied Statistical Methods*, 10(2), 403-417.
- R Development Core Team (2008). A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. (<http://www.R-project.org>).
- Reffinetti, R. (1996). Demonstrating the consequences of violations of assumptions in analysis of variance. *Teaching of Psychology*, 23, 51-54.
- Reis, E. (2001). *Estatística Multivariada Aplicada* (2^a ed.). Lisboa: Edições Sílabo.
- Rencher, A.C. (2002). *Methods of Multivariate Analysis* (2nd ed.). New York: John Wiley & Sons.
- Sahai, H., & Ageel, M.I. (2000). *The Analysis of Variance*. Boston: Birkhäuser.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley and Sons.
- Sheskin, D.J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures* (4th ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 667-680.
- Stevens, S.S. (1951). *Mathematics, measurement, and psychophysics*. In S.S. Stevens (ed.), *Handbook of experimental psychology*. New York: Wiley.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. New York: Pearson Education Inc.
- Timm, N.H. (2002). *Applied Multivariate Analysis*. New York: Springer-Verlag.
- Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistics and Data Analysis*, 54, 37-48.
- Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Thurstone, L.L. (1954). The measurement of values. *Psychological Review*, 61, 47-58.
- Urdan, T.C. (2005). *Statistics in Plain English*. London: Laurence Erlbaum Associates.
- Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 667-680.
- Van Aelst, S., & Willems, G. (2011). Robust and Efficient One-Way MANOVA Tests. *Journal of the American Statistical Association*, 106(494), 706-718.
- Worthington, R., & Whittaker, T. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, 34(6), 806-838.
- Zwick, R. (1985). Nonparametric One-Way Multivariate Analysis of Variance: A Computational Approach Based on the Pillai-Bartlett Trace. *Psychological Bulletin*, 47(1), 148-152.
- Zimmerman, D.W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127, 354-364.
- Zimmerman, I.L., & Castilleja, N.F. (2005). The role of language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities*, 11, 238-246.

Capítulo IV. A Simulação e a Geração Aleatória de dados

Neste capítulo o enfoque está na técnica de simulação pelo método de Monte Carlo e na geração aleatória de dados ordinais multivariados com uma determinada estrutura de correlação, utilizando duas metodologias: o método de conversão binária e o mapeamento de médias.

4.1 O Método de Monte Carlo

A simulação é uma técnica de pesquisa operacional que usa experiências numéricas com modelos lógico-matemáticos que envolvem um grande número de cálculos e um grande número de repetições, pelo que necessitam de uso intensivo de recursos computacionais (Saliby, 1989), o que inicialmente foi problemático para os investigadores. Este problema deixou de existir com o aumento da velocidade de processamento e da capacidade de armazenamento dos computadores, associada à evolução das próprias metodologias de simulação (Yoriyaz, 2009).

As práticas de simulação começam a ter alguma visibilidade em 1940, quando os cientistas de Los Alamos (área nuclear), Stanislaw Ulam, John Von Neumann e Nicholas Metropolis desenvolvem o Método de Simulação de Monte Carlo. Através deste método estatístico consegue-se determinar a probabilidade de ocorrência de uma situação experimental através de geração aleatória de amostras. Como se disse anteriormente, estas técnicas de amostragem eram pouco usadas por serem demasiado demoradas, porém com o aparecimento do primeiro computador, desenvolvido durante a segunda guerra mundial, o Electronic Numerical Integrator And Computer (ENIAC), o problema da demora foi ultrapassado e a sua utilização mais efetiva. A apresentação formal, à comunidade científica, foi feita por Nicholas Metropolis e Stanislaw Ulam aquando da publicação do artigo “*Monte Carlo Method*” no *American Statistical Association Journal* (Metropolis & Ulam, 1949). Os anos 40 do século passado, foram assim um marco na disseminação deste método em conferências, simpósios e seminários (Hurd, 1985).

O Método de Monte Carlo foi assim denominado em homenagem ao carácter aleatório dos jogos de azar do Casino de Monte Carlo, no Mónaco. As práticas de

simulação com este método permitem determinar a probabilidade de ocorrência de uma situação experimental através de geração aleatória de amostras e mediante um número suficientemente elevado de réplicas. Verificou-se também o interesse pela utilização deste método quer em experiências que tinham subjacente o uso de modelos determinísticos, quer em experiências às quais estavam associados modelos probabilísticos. Os modelos de simulação probabilística tiveram a sua origem no Método de Monte Carlo o qual pode ser descrito como um método estatístico que na simulação utiliza sequencialmente números pseudoaleatórios e que permite avaliar as distribuições de amostragem das estatísticas e estudar o efeito da violação de alguns pressupostos.

A dificuldade na observação direta de problemas relacionados com fenómenos reais e a crescente complexidade desses mesmos problemas fez com que aumentasse a utilização de métodos de simulação na resolução de problemas reais, pois a facilidade com que se trabalham os seus algoritmos e a eficiência com que se obtêm resultados, em condições muito difíceis, têm justificado o seu uso em diversas áreas do conhecimento.

The Monte Carlo method is a statistical sampling technique that over the years has been applied successfully to a vast number of scientific problems (Eckhardt, 1987, p. 131). Mediante o tipo de distribuição de probabilidade, são obtidas as amostras aleatórias, sendo o processo repetido, o número de vezes determinado pelo investigador e utilizadas as técnicas estatísticas também por ele demarcadas. Halton (1970, p. 1) define o método de Monte Carlo como (...) *the solution of a problem as a parameter of a hypothetical population, and using a random sequence of numbers to construct a sample of the population (...)* a partir da qual se podem obter as estimativas do parâmetro.

Podemos dizer que a metodologia de Monte Carlo tem sido usada ao longo dos anos para analisar distribuições de amostragem de diferentes estatísticas e avaliar em que medida a violação de pressupostos origina alteração na potência dos testes. Um dos exemplos mais comuns de utilização desta metodologia com este propósito é a análise de variância univariada e multivariada (Ramsey & Ramsey, 2013, 2011; Todorov & Filzmoser, 2010; Finch & Davenport, 2009; Bilodeau & Micheaux, 2005; Finch, 2005; Zwick, 1985; Hummel & Sligo, 1971).

4.2 A Geração Aleatória de Variáveis Ordinais

O recurso a computadores para gerar números aleatórios (pseudoaleatórios) é uma prática comum em muitas áreas do conhecimento, embora menos comum com dados ordinais. A simulação envolve componentes como: função densidade de probabilidade, técnicas de amostragem, variáveis aleatórias e geração de números aleatórios. A geração de números aleatórios baseia-se em algoritmos matemáticos, cuja ocorrência obedece a uma certa aleatoriedade semelhante à encontrada na natureza. Assim, se tivermos um conjunto de números definidos no intervalo $[0,1]$, se eles se distribuírem uniformemente nesse intervalo e não estiverem correlacionados, constituem-se como uma sequência de números aleatórios (Yoriyaz, 2009).

O método de Biswas apresenta algoritmos para gerar dados ordinais para algumas estruturas específicas de correlações. O autor considera que o modelo proposto é muito importante para aplicar na resolução de problemas com dados ordinais multivariados.

The immediate applicability of the present model is to the analysis of longitudinal data where covariates are not time dependent and also to the analysis of clustered data (Biswas, 2004, p. 25). Mas, segundo Demirtas (2006), a exigência de distribuições idênticas e independentes torna o método bastante restritivo. No que diz respeito às variáveis ordinais, Demirtas (2006, p. 1017) refere que: (...) *one common approach is generating latent continuous variables and converting them to the ordinal ones using pre-specified threshold points*. Este autor considera esta abordagem inapropriada e providência métodos específicos para gerar amostras de dados ordinais com determinadas estruturas de correlações. Kaiser, Traeger & Leisch (2011) apresentam algoritmos para gerar dados ordinais correlacionados através de dois métodos: a conversão binária e o mapeamento de médias. Na geração de dados ordinais correlacionados X_i com valores inteiros $\{1,2,3,\dots,k\}$, a distribuição correspondente é

definida pelo vetor de probabilidades, $p_{X_i} = \begin{pmatrix} P(X_i = 1) \\ P(X_i = 2) \\ \vdots \\ P(X_i = k) \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$.

Quando se geram números pseudoaleatórios de modo a obter variáveis ordinais X_1, \dots, X_p que tomem valores inteiros $\{1,2,\dots,k\}$, o utilizador necessita especificar as probabilidades marginais $\mathbf{p}_{X_i}, i = 1, \dots, p$ e pré-definir uma matriz de correlações semi-definida positiva:

$$\mathbf{cor} = \begin{pmatrix} r(x_1, x_1) & r(x_1, x_2) & \dots & r(x_1, x_p) \\ r(x_2, x_1) & r(x_2, x_2) & \dots & r(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_p, x_1) & r(x_p, x_2) & \dots & r(x_p, x_p) \end{pmatrix}$$

Kaiser, Traeger & Leisch (2011) usaram na geração de dados, a função do *package* “*rmvord*” (Kaiser and Leisch, 2010, R Development Core Team, 2010), que tem como função principal o retorno de n_l observações com determinadas probabilidades marginais e correlações e matriz de correlações **cor**. Os autores basearam o método de conversão binária nos estudos de Demirtas (2006).

The main idea is to draw binary random variables with the correct correlation structure, and conditional on the outcome of the binary variable convert an independent uniform random to an ordinal variable with the desired marginal and correlations (Kaiser, Traeger & Leisch; 2011, p.4).

O tempo de execução do algoritmo do método de conversão binário, função do *package* “*rmvord_b*”, é muito rápido, mesmo que o delineamento do estudo abranja um número elevado de variáveis ou de pontos da escala, mas não funciona com valores de correlação elevados (Demirtas, 2006). No sentido de ultrapassar esta questão, os autores criaram uma metodologia alternativa, o método de mapeamento de médias. Este método gera números aleatórios ordinais com uma determinada estrutura de correlação, tendo como suporte os estudos de Leisch, Weingressel & Hornik (1998). O método de mapeamento de médias, função do *package* “*rmvord*”, é mais estável, aceita valores de correlação mais elevados na matriz de simetria composta que define a estrutura de correlação, mas é computacionalmente muito mais demorado. O tempo de execução deste método depende do número de pontos da escala e do número de variáveis (Kaiser, Traeger & Leisch, 2011).

Ao compararmos os dois métodos, pode dizer-se que ambos apresentam vantagens e desvantagens. Por um lado, o método de conversão binária, apesar de mais rápido, limita o conjunto de matrizes de correlação exequíveis. Por outro lado, o método de mapeamento de médias supera esse problema e tem uma precisão idêntica ao método de conversão binária, mas aumenta drasticamente o tempo de execução das simulações.

4.3 Referências Bibliográficas

- Bilodeau, M., & Micheaux, P.L. (2005). A multivariate empirical characteristic function test of independence with normal marginal. *Journal of Multivariate Analysis*, 95, 345-369.
- Biswas, A. (2004). Generating correlated ordinal categorical random samples. *Statistics & Probability Letters*, 70, 25-35.
- Demirtas, H. (2006). A method for multivariate ordinal data generation given marginal distributions and correlations. *Journal of Statistical Computation and Simulation*, 76(11), 1017-1025.
- Eckhart, R. (1987). Stan Ulam, John Von Neumann and the Monte Carlo Method. *Los Alamos Science*, 15, 131-136.
- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions are Violated. *Methodology*, 1(1), 27-38.
- Finch, W.H., & Davenport, T. (2009). Performance of Monte Carlo Permutation and Approximate Tests for Multivariate Means Comparisons with Small Sample Sizes when Parametric Assumptions are Violated. *Methodology*, 5(2), 60-70.
- Halton, J.H. (1970). A retrospective and prospective survey of the Monte Carlo method. *SIAM Review* (Society for Industrial and Applied Mathematics), 12, 1-63.
- Hummel, T.J., & Sligo, J.R. (1971). Empirical comparison of univariate and multivariate analysis. *Psychological Bulletin*, 76, 49-57.
- Hurd, C.C. (1985). A note on early Monte Carlo Computations and Scientific Meetings. *Annals of the History of Computing*, 7(2), 141-145.
- Kaiser, S., Träger, D., & Leisch, F. (2011). Generating Correlated Ordinal Random Values. *Technical Report Number*, 94, Munich: Department of Statistics, University of Munich.
- Kaiser, S., & Leisch, F. (2010). Orddata: Generation of Artificial Ordinal and Binary Data, R package version 0.1.
- Metropolis, N., & Ulman, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), 335-341.
- Ramsey, P.H., & Ramsey, P.P. (2013). Pairwise testing of group mean vectors in MANOVA with small samples. *Journal of Statistical Computation and Simulation*, 83(6), 1037-1057.
- Ramsey, P.H., Ramsey, P.P., Hachimine, P., & Andiloro, N. (2011). Robustness, Power and Interpretability of Pairwise Tests of Discriminant Functions in MANOVA. *Journal of Modern Applied Statistical Methods*, 10(2), 403-417.
- R Development Core Team (2010). *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (<http://www.R-project.org>).
- Saliby, E. (1989). *Repensando a Simulação: A Amostragem Descritiva*. São Paulo: Atlas, 1989.
- Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistics and Data Analysis*, 54, 37-48.
- Yoriyaz, H. (2009). Método de Monte Carlo. Princípios e Aplicações em Física Médica. *Revista Brasileira de Física Médica*, 3(1), 141-149.
- Zwick, R. (1985). Nonparametric One-Way Multivariate Analysis of Variance: A Computational Approach Based on the Pillai-Bartlett Trace. *Psychological Bulletin*, 47(1), 148-152.

PARTE II. Aplicações da Teoria

Introdução

A terceira parte do trabalho inclui dois capítulos nos quais estão compilados dois artigos referentes às leituras prévias ao estudo de simulação e assim justificar parte da problemática inerente a este trabalho.

A pesquisa efetuada nas bases de dados permitiu selecionar alguns trabalhos que na sua origem têm instrumentos desenvolvidos com variáveis (itens) com formato de medida ordinal.

No primeiro artigo foi feita uma revisão sistemática da literatura no sentido de dar resposta à primeira questão de investigação: Quais os testes estatísticos mais frequentemente utilizados pelos investigadores da área da saúde infantil e pediatria, quando usam “escalas de avaliação” compostas por variáveis com um formato de medida ordinal?

Os artigos mostram que os investigadores recorrem tanto a testes paramétricos como a não paramétricos, em muitas situações sem apresentar ou justificar o porquê da sua opção. Também se encontram casos em que não se verificam os pressupostos dos testes paramétricos e a opção do autor por estes testes é a sua robustez mesmo quando as distribuições sob estudo apresentam enviesamento e/ou achatamento consideráveis (Rodrigues, 2008).

No segundo artigo apresenta-se uma reflexão teórica no sentido de salientar respostas dadas pelos autores a muitos problemas reais e de investigação, de quem utiliza nas suas práticas instrumentos construídos com variáveis ordinais.

Apesar de terem sido encontradas algumas respostas, que tentámos salientar nos dois artigos, verifica-se a necessidade de desenvolver mais estudos de simulação que confirmem alguma da teoria já existente e facilitem a tomada de decisão dos investigadores que utilizam variáveis com formato de medida ordinal.

Capítulo V. Análise Estatística de Escalas Ordinais: Aplicações na área da Saúde Infantil e Pediatria

Resumo

A investigação nas diferentes áreas do conhecimento envolve grande número de variáveis que podem ser avaliadas com técnicas univariadas ou multivariadas e com testes paramétricos ou não paramétricos. A decisão depende da validação dos pressupostos inerentes a cada um deles, do tipo de variáveis e das dimensões das amostras. Começámos com uma revisão sistemática da literatura, que incluía artigos científicos cujo instrumento de recolha de dados era uma escala de avaliação/psicométrica na área da saúde da criança e jovem. Definidos os critérios de inclusão, foram seleccionados 240 artigos científicos e após leitura de títulos e resumos reduziu-se para 189. A análise exploratória das metodologias de investigação utilizadas originou uma redução para 127 artigos. A análise exploratória de dados permitiu-nos identificar o tipo de técnicas estatísticas utilizadas pelos investigadores quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização das técnicas estatísticas com este tipo de variáveis e com amostra de pequena e grande dimensão.

Palavras-chave: escalas; amostra; análise estatística; saúde infantil; pediatria

Abstract

Research in different areas of knowledge involves a large number of variables that can be evaluated with univariate and multivariate techniques and parametric or nonparametric tests. The researcher decision depends on the validation of the assumptions inherent to each, the type of variables and sample sizes. We began with a systematic literature review, which included scientific articles that had, as instrument of data collection, assessment scales or psychometrics scales in the health of children and youth. Defined inclusion criteria, 240 articles were selected, after reading titles and abstracts reduced to 189 papers. Exploratory analysis of research methodologies used resulted in a reduction to 127 articles. The data analysis allowed us to identify the type of statistical techniques used by investigators when the instruments are composed of ordinal variables and analyze whether they met the assumptions inherent in the use of statistical techniques with this type of variables with small and large samples.

Keywords: scales, sample, statistical analysis, child health, pediatrics.

5.1 Introdução

A investigação nas diferentes áreas do conhecimento envolve análise de variáveis através de diferentes técnicas estatísticas dependendo da validação dos pressupostos inerentes a cada uma delas, do tipo de variáveis e da dimensão da amostra. O processo de investigação inclui o conceito de variável e a forma de medida que o investigador utiliza.

Em ciências sociais e humanas e nas ciências da saúde, temos muitas vezes que utilizar estratégias que nos permitam quantificar o não observável. É neste contexto que emergem as “escalas de avaliação” (Hill & Hill, 2009) compostas por variáveis qualitativas com uma escala de medida ordinal que permitem estimar um constructo e que o investigador, muitas vezes, trata como uma variável contínua (Marôco, 2010).

Os itens qualitativos que constituem estas escalas de medida podem apresentar-se com os mais diversos números de classes de pontuação (3, 5, 7), havendo uma grande variabilidade na utilização dos testes estatísticos. Como refere Marôco (2010), uma das necessidades mais frequentes em análise estatística passa pela comparação de parâmetros populacionais, tendo por base amostras aleatórias, e em que se pretende testar se o tratamento, intervenção ou manipulação teve um efeito significativo na variável resposta, no caso univariado, ou no vetor de variáveis resposta, no caso multivariado. As metodologias estatísticas inferenciais usadas para este tipo de análises podem ser de cariz paramétrico ou não paramétrico.

Se os seus requisitos forem cumpridos, o investigador pode usar a estatística paramétrica (mais potente do que a não paramétrica) (Marôco, 2009; Pontes, 2005). Em caso contrário, a tradição de análise recomenda o uso da estatística não paramétrica, que não impõe a especificação de uma família de distribuições para a distribuição amostral. Tradicionalmente, nas ciências sociais e humanas, a estatística não paramétrica tem sido utilizada em alternativa à estatística paramétrica quando os seus pressupostos de aplicação não se verificam. Contudo, estudos de simulação (Zimmerman, 2000) têm sugerido que alguns testes não paramétricos (e.g., teste U de Mann-Whitney e teste de Kruskal-Wallis) são tão sensíveis quanto os paramétricos à violação de algumas condições, nomeadamente a homocedasticidade, e que a não verificação destas condições conduz geralmente à inflação da probabilidade dos erros tipo I e tipo II. Assim sendo, a utilização da estatística não paramétrica em alternativa à paramétrica pode não ser a melhor estratégia de análise. Por outro lado, é frequente, nas ciências

sociais, ver variáveis qualitativas ordinais analisadas com recurso à estatística paramétrica. Urdan (2010) questiona a validade dos métodos paramétricos quando aplicados a variáveis ordinais e refere que as variáveis ordinais (...) *são variables measured with numerical values where the numbers are meaningful (e.g., 2 is larger than 1) but the distance between the number is not constant* (Urdan, 2010, p. 6). A ausência de recomendações claras relativamente ao uso da estatística paramétrica versus não paramétrica é ainda mais pronunciada no caso multivariado. Com frequência, nas ciências sociais e humanas, a natureza latente de um grande número de variáveis de interesse obriga à utilização de escalas sociométricas e psicométricas. Estes instrumentos de medida são geralmente constituídos por um conjunto de itens ordinais, assumidamente manifestações das variáveis latentes que, no domínio das ciências sociais e humanas não é possível observar diretamente. De uma forma geral, este conjunto de itens são sujeitos a transformações lineares de forma a operacionalizar a medida, mas esta prática tem sido questionada por uns e defendida por outros (Carifio & Perla, 2007, 2008).

A análise multivariada de dados engloba várias técnicas que passam pela análise de variância multivariada (MANOVA), de componentes principais, de agrupamentos, de correlação canónica, discriminante e fatorial. São vários os autores que consideram que a MANOVA tem vantagens relativamente à utilização de várias ANOVAs, das quais salientam a habilidade de medir os vários ângulos do problema, podendo em alguns casos melhorar a potência do teste e apresentar uma probabilidade de erro de tipo I mais reduzida (Tabachnick & Fidell, 2007).

Em muitos estudos, o uso de dados multivariados é relevante se a entidade medida tem uma variedade de componentes conceptuais, como a proficiência numa habilidade cognitiva, como a leitura ou a matemática, e se o investigador pretender comparar grupos simultaneamente nessas componentes. De notar que, em alguns casos onde respostas múltiplas são recolhidas para dois ou mais grupos de sujeitos, as hipóteses de interesse são univariadas por natureza e, nesse caso, cada variável de resposta deve ser vista isoladamente (Marôco, 2009).

Para que a MANOVA produza conclusões válidas, existem postulados que devem ser cumpridos e que são muito similares àqueles que são requeridos para o uso da ANOVA: (1) as observações devem ser independentes, (2) o vetor de variáveis resposta deve ter distribuição normal multivariada para as diferentes populações, e (3) as matrizes de covariâncias das populações devem ser homogéneas (Marôco, 2009;

Sheskin, 2007). Perceber a performance da MANOVA quando os pressupostos não são cumpridos tem sido o objetivo de muitos investigadores. Finch (2005) encontrou alguns casos em que a violação do pressuposto de normalidade tem um impacto limitado na taxa de erro de tipo I, referindo também alguns estudos que têm mostrado que a potência da MANOVA decresce substancialmente quando as variáveis do vetor de resposta são platicúrticas. Estes resultados são similares ao que foi encontrado em investigações sobre a potência da ANOVA.

Algumas pesquisas indicam que, quando os grupos não variam em tamanho, a heterogeneidade das matrizes de covariância tem um efeito muito pequeno na probabilidade de erro de tipo I e na potência do teste, mas quando as amostras não apresentam as mesmas dimensões, a heterogeneidade das matrizes de variância-covariância pode causar um aumento da taxa de erro de tipo I e da potência do teste estatístico pode diminuir. Estes resultados são mais evidentes à medida que as “desigualdades” nas matrizes de covariâncias e nas dimensões das amostras se vão acentuando (Finch, 2005).

Perante os aspetos apresentados e porque em ciências sociais e humanas e nas ciências da saúde temos muitas vezes que utilizar estratégias que nos permitam quantificar o não observável, emergem as escalas, compostas por variáveis qualitativas com uma escala de medida ordinal e que o investigador, muitas vezes, trata como se o constructo fosse de uma variável contínua (escalas, testes psicológicos, inventários). Porém o problema surge pelo facto de na ausência de um fator de calibração consensual (*gold standard*) nunca conseguirmos ter a certeza de que o instrumento está a medir o que é suposto medir, de forma válida, fiável e sensível (Marôco, 2010)

Ao analisarmos as variáveis qualitativas, cuja escala de medida é ordinal, não podemos deixar de falar de Rensis Likert por todo o trabalho desenvolvido por este investigador em torno das variáveis ordinais. As escalas de tipo Likert são das mais usadas em diversos campos da investigação e particularmente na área da saúde (enfermagem e educação médica), psicologia e outras ciências sociais. Quanto à classificação do tipo destas variáveis não verificamos a existência de dúvidas, são variáveis qualitativas ordinais, com escalas de medida que podem apresentar diferentes formas (as mais comuns 3, 4, 5, 6, 7, 10 pontos); contudo, o mesmo não se pode dizer quando, através de operações algébricas (soma ou média) com os conjuntos de itens, transformamos a variável ordinal a um nível de medida intervalar (*Summated scales*) obtendo *scores* referentes a constructos teóricos. Neste ponto existe uma grande

controvérsia entre os que são favoráveis a esta transformação e os que são desfavoráveis (Carifio & Perla, 2008, 2007; Jamieson, 2004; Wang, et al., 1999; Knapp, 1990, 1993; Mitchell, 1986; Weiss, 1986).

Nem todos os autores estão de acordo que se adicionem os itens, pois não conseguem compreender que, por exemplo, ao somar “concordo completamente” com «discordo totalmente», o que o investigador soma é o grau e não o conceito e consideram que as variáveis são ordinais e o seu tratamento estatístico tem de privilegiar essa medida e não outra (Urda, 2010).

Ainda existem outros autores que consideram que esta operacionalização pode ser feita se os itens das escalas tiverem pelo menos 5 pontos, na escala de medida, e forem consideradas como intervalares ou quase intervalares, tendo subjacente um constructo contínuo (Marôco, 2010; Bollen, 1989). Os investigadores que usam estes tipos de escala referem os estudos que valorizam a sua utilização sem prejuízo da investigação e recomendam que devemos estar atentos: à escala de medida, ao número de itens, se a escala é constituída por subescalas (se estas têm um número de itens iguais ou diferentes) e à dimensão da amostra.

No que concerne aos testes estatísticos, as opiniões voltam a divergir relativamente à sua aplicação, pois alguns autores consideram que com variáveis ordinais somente podem ser usados testes não paramétricos outros consideram que após a transformação das variáveis (*summated scales*) devem ser usados os paramétricos. Pell (2005) é um dos autores que concorda com este pressuposto, dizendo que a soma de itens pode ser analisada com estatística paramétrica desde que se verifiquem os pressupostos dos testes com todos os benefícios da sua robustez. Carifio e Perla (2008) reforçam esta ideia, dizendo que é perfeitamente adequado usar técnicas paramétricas com este tipo de variáveis.

Na literatura científica verificamos que muitos investigadores, no caso univariado, estudam a normalidade da distribuição das variáveis (usando por exemplo o teste de *Kolmogorov-Smirnov*) e verificam a homogeneidade das variâncias (usando por exemplo o teste de *Levene*) com o objetivo de decidirem se optam pelos testes paramétricos (e.g., teste *t* ou ANOVA) ou pelos não paramétricos (e.g., teste U de Mann-Whitney ou o teste de Kruskal-Wallis). Porém, também verificamos que alguns investigadores têm esta prática, mas perante uma situação em que as variáveis não possuem uma distribuição normal, consideram que os testes paramétricos (...) ANOVA, MANOVA, *t-Student* são robustos mesmo quando as distribuições sob estudo

apresentam enviesamento ou achatamento considerável (...) (Rodrigues, 2008, p.170) e a sua decisão passa pelos testes paramétricos em detrimento dos não paramétricos. Se por um lado os testes paramétricos exigem que a forma de distribuição amostral seja conhecida (e.g., Normal), os testes não paramétricos não têm esse tipo de exigência. Então, porque não se utilizam sempre os testes não paramétricos? Esta é uma questão que pode surgir. Alguns autores justificam este aspeto dizendo que a potência de teste e a probabilidade de rejeitar corretamente a hipótese nula é maior nos testes paramétricos do que nos testes não paramétricos.

Para Marôco (...) os testes não paramétricos devem apenas usar-se quando não existe alternativa, i. e., quando não é possível validar as condições de aplicação dos testes paramétricos ou quando as variáveis não são quantitativas (Marôco, 2010, p.199).

Da reflexão sobre as diferentes questões apresentadas e sobre a análise de dados com este tipo de variáveis e porque no nosso quotidiano somos confrontados diariamente com estes problemas, resolvemos analisar de forma mais sistematizada o que emerge na literatura científica.

Esta análise exploratória das metodologias de investigação vai no sentido de compreender qual o planeamento da análise dos dados, uma vez que na área das ciências sociais e humanas, da saúde e motricidade humana, áreas que utilizam este tipo de variáveis, os investigadores muitas vezes se interrogam se estão a seguir o melhor caminho.

Assim, este estudo tem como objetivos: identificar o tipo de análise estatística (testes paramétricos ou não paramétricos) utilizada pelos investigadores quando os instrumentos utilizados são compostos por variáveis ordinais (com 3, 4, 5, 7 e 10 pontos na escala de medida) e analisar se são cumpridos os pressupostos inerentes à utilização das diferentes técnicas estatísticas.

5.2 Metodologia

Este estudo começou com uma revisão sistemática da literatura (publicações periódicas em três idiomas, português, espanhol e inglês), onde foram incluídos os artigos científicos que tinham como instrumento de recolha de dados uma escala de avaliação/psicométrica. No contexto deste estudo definimos como critérios de inclusão que estas escalas: 1) fossem compostas por variáveis ordinais, 2) tivessem uma escala

de medida de 3, 4, 5, 7, 10 pontos; 3) fossem escalas usadas na área da saúde infantil e pediatria, podendo abranger desde o recém-nascido ao jovem adulto (com e sem deficiência motora), e 4) usassem na análise de dados testes paramétricos ou não paramétricos.

5.2.1 Estratégias de Pesquisa

A pesquisa realizou-se em Outubro de 2010, em bases eletrônicas, de estudos não publicados (Dissertations and Theses (ProQuest); Bibliografia Nacional Portuguesa “on line”) e de estudos publicados em bases de dados como: CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature; EMBASE, Medical Library Association; ISI Web of Knowledge; MEDLINE, U.S. National Library of Medicine; PsychINFO, abstract database of the psychological literature e SciELO, Scientific Electronic Library Online. As palavras utilizadas foram: *newborn, child, infant, adolescente, youth, scales, ordinal variables, parametric tests, non parametric tests*.

Foram selecionados 240 artigos científicos. Após uma leitura de títulos e resumos decidimos pela inclusão no estudo de 189 desses artigos. Com uma leitura mais minuciosa dos mesmos, foram excluídos 62 artigos que não cumpriam na íntegra os critérios de inclusão. Passaram assim, à análise exploratória das metodologias de investigação utilizadas 127 artigos.

5.3 Resultados

Tendo por base o objetivo de identificar se os investigadores utilizam técnicas estatísticas paramétricas ou não paramétricas quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização destas técnicas, será efetuada uma análise para salientar o que emerge relativamente ao instrumento utilizado e sua caracterização pelos autores, ao tamanho da amostra e à sua relação com o número de itens (variáveis) do instrumento e o tipo de testes estatísticos aplicados nos diferentes estudos. Os dados foram extraídos dos 127 artigos selecionados para uma matriz de dados, especialmente construída para este estudo e cuja grelha incluía todos os aspetos em análise, sendo depois apresentados graficamente para uma melhor visualização dos resultados. Estes artigos tinham como

instrumento de recolha de dados uma ou duas das dezassete escalas de avaliação apresentadas na

Figura 5. 1, as quais eram compostas por escalas únicas (um conjunto de itens) ou com várias subescalas (vários conjuntos de itens) constituídas por variáveis ordinais.

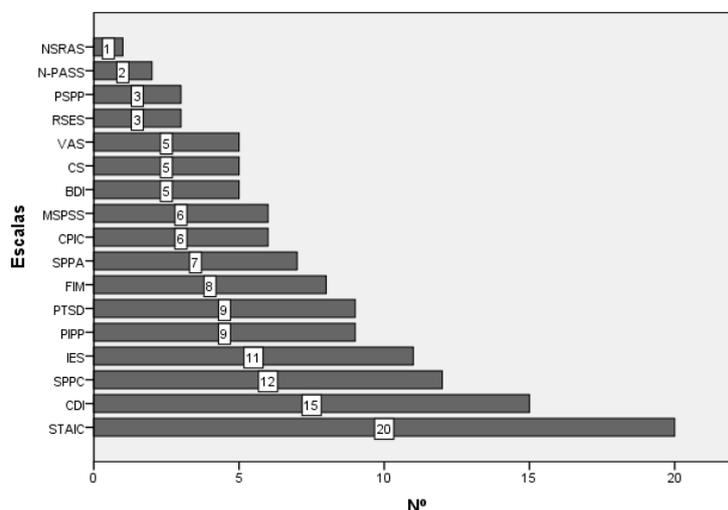


Figura 5. 1 - Número (Nº) de Artigos por escala de avaliação, nos 127 artigos, que cumpriam todos os critérios de inclusão.

Os instrumentos mais utilizados foram: a State-Trait Anxiety Inventory for Children, STAIC (20 artigos); Child Depression Inventory, CDI (15 artigos); Self-Perception Profile for Children, SPPC (12 artigos); Impact of Event Scale, IES (11 artigos); a Premature Infant Profile, PIPP e a Posttraumatic Stress Disorder scale, PTSD (9 artigos cada); a Functional Independence Measure, FIM (8 artigos); a Self-Perception Profile for Adolescents, SPPA (7 artigos); a Children's Perception of Interparental Conflict Scale, CPIC e a Multidimensional Scale of Perceived Social Support, MSPSS (6 artigos cada); a Beck Depression Inventory, BDI, a Comfort Scale, CS (5 artigos), a Visual Analogue Scale, VAS (5 artigos cada); a Rosenberg Self-Esteem Scale, RSES e a Self-Perception Profile for Children, PSPP (3 artigos cada); a Neonatal Pain, Agitation and Sedation Scale, N-PASS (2) e Neonatal Skin Risk Assessment Scale, NSRAS (1).

Ao fazermos uma análise mais pormenorizada dos artigos no que respeita ao tipo de instrumento, verificamos que os artigos apresentam de forma clara e explícita o tipo de instrumento (escalas) utilizado pelos investigadores, o número de itens por que são compostas as escalas e também se estamos perante uma escala única (um só conjunto de itens) ou se as escalas são composta por subescalas (vários conjuntos de itens), se estes conjuntos são equilibrados numericamente ou não e o número de pontos da escala de medida (3, 4, 5, 7, 10 pontos) (cf. Tabela 5. 1).

Tabela 5. 1 - Artigos científicos analisados por tipo de escala de avaliação/psicométrica

Escalas de Avaliação (Psicométricas)	Nº Itens	Subescalas (nº subescalas * nº itens)	Escala de Medida	Total de Artigos
N-PASS	5	-	3	2
STAIC	40	2(2*20)	3	20
CPIC	48	9(3*4; 1*5; 4*6; 1*7)	3	6
	40	7(1*4; 1*5; 4*6; 1*7)		
CDI	27	-	3	15
PSPP	36	-	4	3
SPPC	36	6*6	4	12
SPPA	51	10 (9*5; 1*6)	4	7
	45	9 (9*5)		
	36	6 (6*6)		
	20	4 (4*5)		
RSES	10	-	4	3
	42	2 (1*8; 1*34)		
IES	15	2 (1*7; 1*8)	4	11
	22	3(2*7; 1*8)		
BDI	21	-	4	5
PIPP	4	-	4	9
NSRAS	6	-	4	1
PTSD	20	-	5	9
MSPSS	12	3 (3*4)	5	6
CS	8	-	5	5
FIM	30	5(1*2; 1*3; 2*4; 2*5; 1*7)	7	8
	18	3(1*6; 2*3; 3*2)		
VAS	10	-	10	5
TOTAL				127

Na maioria dos casos consegue-se perceber, sem recorrer a outro tipo de consulta, a composição dos instrumentos, das subescalas, e assim saber se existem diferenças nos conjuntos de itens. No que diz respeito ao número de itens por escala, nos instrumentos analisados, existem instrumentos compostos por diferentes números de itens e estes oscilam entre 4 e 51 itens. Verifica-se também que podem ser compostos apenas por uma única escala ou até 10 subescalas, variando estas em número de itens em cada uma delas, o que nem sempre permite caracterizá-las com alguma uniformidade. Relativamente à escala de medida, os instrumentos analisados neste estudo apresentavam diferentes escalas de medida que variam entre 3 e 10 pontos.

Amostra e Teste Estatísticos

A escolha da população que vamos inquirir está intimamente ligada ao objeto sob estudo, aos objetivos e método de investigação. Os estudos que têm uma lógica extensiva associada (usando o inquérito por questionário como instrumento), e como tal uma estratégia de investigação quantitativa, exigem cuidados com a dimensão da amostra. Se alguma das amostras for de pequena dimensão (inferior a 30), devemos verificar a normalidade da variável dependente nessa amostra, através de um teste de normalidade (como exemplo o Teste de Shapiro-Wilk) já que no caso contrário o Teorema do Limite Central diz que a distribuição da média amostral é aproximadamente normal. Outros autores referem-se ao uso das “escalas”, instrumentos compostos por várias questões/afirmações ou itens que têm como finalidade avaliar uma determinada variável ou característica, como sendo um tipo de instrumento que, ao exigir que sejam feitas transformações de variáveis, e aplicados alguns testes estatísticos, obriga a que o número de inquiridos seja mais elevado, no mínimo de cinco indivíduos por item devendo o ideal ser 10 indivíduos por item da escala (por exemplo, para a aplicação da escala SPPC com 36 itens deveria ser considerada uma amostra entre 180 e 360 e para a STAIC com 40 itens deveria ser considerada uma dimensão de amostra de pelo menos 200 indivíduos, sendo o ideal 400), tendo também em conta o tipo de testes estatísticos que vão ser aplicados e as variáveis independentes utilizadas.

Ao analisarmos os testes estatísticos aplicados nos diferentes estudos, podemos dizer que, dos oito artigos publicados em língua portuguesa, três apresentavam simultaneamente testes paramétricos e não paramétricos (t de Student, ANOVA, MANOVA, U de Mann-Whitney) e cinco apresentavam apenas testes paramétricos. Os restantes 119 apresentam-se distribuídos da seguinte forma: seis usaram testes não paramétricos (teste U de Mann-Whitney) e 113 utilizam testes paramétricos (t-Student, ANOVA, MANOVA, ANCOVA), sendo que em 11 destes estudos foram utilizados simultaneamente testes não paramétricos.

Verificamos na literatura científica que alguns autores partem do princípio de que a análise de variância exige que sejam observados os pressupostos da normalidade (a variável dependente deve ter uma distribuição normal em todos os grupos) e da homocedasticidade (as variâncias devem ser iguais em todos os grupos), mas nem sempre os artigos científicos têm expresso a verificação deste pressuposto. Na realidade, o que constatamos é que na maioria dos artigos analisados não aparece explícito se

foram verificados os pressupostos das metodologias utilizadas e também não aparece qualquer justificção para a utilizaçõ dos testes paramétricos ou não paramétricos. Daí nem sempre se conseguir fazer uma avaliaçõ rigorosa dos pressupostos enunciados pelo facto de estes estarem omissos nos artigos.

Como não podemos analisar todos os instrumentos, seleccionámos a escala mais usada nos artigos revistos, a State-Trait Anxiety Inventory for Children (STAIC) que aparece em vinte dos artigos analisados (Rodrigues, 2008; Telze, et al., 2008; Bilgiç et al., 2006; Fernandez-Berrocal et al., 2006; Vidanovic & Andelkovic, 2006; Wolbeek, 2006; Huizinga et al., 2005; Bakib, 2004; Bas et al., 2004; O’Dea & Abraham, 2004, 2000; Arslan et al., 2003; Vila et al., 2003, 1999; Derryberry & Reed, 2002; Von Weiss, et al. 2002; Hobbie et al., 2000; Muris, et al., 1999a, 1999b, 1996).

Verificamos que esta é composta por duas subescalas com 20 itens cada, cumprindo o pressuposto de que as subescalas deveriam ter um número de itens semelhante. No entanto emerge na análise o incumprimento dos pressupostos relacionados com a dimensão da amostra (apenas em 30% dos estudos a amostra é superior a 200 observações) e com os testes estatísticos aplicados não emerge a justificção do uso de testes paramétricos ou não paramétricos (são aplicados testes paramétricos em todos os estudos sendo que em 15% se usa simultaneamente testes não paramétricos).

A STAIC foi usada em quatro estudos experimentais, sendo que num dos estudos foram usados testes paramétricos e não paramétricos e as dimensões das amostras sob estudo eram no grupo experimental $n=43$ e no grupo de controlo $n=20$. Dos quatro estudos, dois apresentavam amostras com dimensões iguais para o grupo experimental e de controlo ($n=30$ e $n=35$, respetivamente) e no quarto estudo as dimensões amostrais eram $n=284$ e $n=220$ (para o grupo experimental e grupo de controlo, respetivamente). Nos restantes dezasseis estudos verificou-se que tamanho da amostra variou entre $n=20$ e $n=3467$ e que na sua maioria foram usados testes paramétricos. Apenas em três estudos foram usados simultaneamente testes paramétricos e não paramétricos. O que nos permite dizer que os pressupostos das dimensões mínimas da amostra não foram cumpridos na maioria dos estudos, tendo por base o pressuposto da dimensão da amostra (amostra de grande dimensão), o pressuposto do número de participantes por item e o tipo de testes estatísticos aplicados. Como as amostras não apresentam as dimensões exigidas, nem têm as mesmas dimensões, o que pode levar a uma heterogeneidade das matrizes de covariância,

podendo refletir-se num aumento da probabilidade do erro de tipo I, a potência do teste estatístico pode diminuir. Quanto aos testes estatísticos aplicados, não é apresentada qualquer justificação.

5.4 Conclusões

Nesta análise exploratória de dados, pretendeu-se identificar o tipo de técnicas estatísticas utilizadas pelos investigadores (através dos artigos científicos ligados à área de saúde da criança, adolescente e jovem adulto com e sem deficiência) quando os instrumentos utilizados são compostos por variáveis ordinais e analisar se são cumpridos os pressupostos inerentes à utilização das técnicas estatísticas com este tipo de variáveis e com amostras de pequena e grande dimensão.

Foram analisados 127 artigos que cumpriam os critérios de inclusão, tendo como primeira intenção de análise o tipo de instrumento (escala única ou dividida em subescalas) com formato de medida ordinal com 3, 4, 5, 7 e 10 pontos. Como segunda intenção de análise a dimensão da amostra e como terceira intenção de análise os testes estatísticos aplicados.

Da análise exploratória dos artigos, concluímos que muitos não trazem expressos alguns dos pressupostos que gostaríamos de analisar em profundidade, ou porque os autores não os incluíram no artigo, ou porque poderão não ter sido trabalhados (e.g., distribuição, homogeneidade das variâncias, transformação das variáveis para poderem ser tratadas como contínuas).

No que diz respeito à normalidade da distribuição das variáveis, nem todos os artigos mostram se foram aplicados testes nem quais (alguns dizem que foi aplicado o teste de Kolmogorov-Smirnov), o mesmo acontecendo para a homogeneidade das variâncias (alguns referem o teste de *Levene*). Desta forma não fica claro para o leitor o que esteve na origem da tomada de decisão de se optar por testes paramétricos ou não paramétricos.

Pelo facto de estarmos, em alguns dos estudos analisados, perante amostras de pequena dimensão onde são usadas técnicas estatísticas multivariadas, poderia dizer-se que os investigadores deveriam ter usado a inferência não paramétrica, visto esta não exigir a especificação de um modelo populacional, ao contrário do que acontece na inferência paramétrica e porque em alguns dos casos não se verificam os pressupostos referidos anteriormente. Mas isso não se verifica também, nestes estudos, pois na sua maioria são usados testes paramétricos.

Os investigadores que optaram por aplicar testes paramétricos basearam-se em estudos de simulação que têm sugerido que alguns testes não paramétricos são tão sensíveis quanto os paramétricos à violação de algumas condições, como a homocedasticidade na inflação dos erros tipo I e tipo II (Rodrigues, 2008; Zimmerman, 2000). Mas esta opção nem sempre é apresentada ou justificada nos artigos científicos. Para alguns investigadores que trabalham nestas áreas onde nem sempre é fácil medir, pois muitas vezes não se consegue ter a certeza de que o instrumento está a medir o que é suposto medir, de forma válida, fiável e sensível, torna-se importante a valorização do que tem sido investigado com este tipo de variáveis para que seja dada consistência à análise de dados desta natureza e assim poderemos continuar a investigar, no sentido de contribuímos com evidência científica para que o rigor esteja em crescendo e os artigos científicos apresentem de forma expressa todos os pressupostos em análise.

5.5 Referências Bibliográficas

- Bakib, O., Erdogana, A., Kantarcic, O., Akisikd, G., Kayaalpa, L., & Yalcinkayab, C. (2004). Anxiety and depression in children with epilepsy and their mothers. *Epilepsy and Behavior*, 5, 958-964.
- Arslan, N., Büyükgebiz, B., Öztürk, Y., & Akay, A.P. (2003). Depression and anxiety in chronic hepatitis B: effect of hepatitis B virus infection on psychological state in childhood. *The Turkish Journal of Pediatrics*, 45, 26-28.
- Bas, M., Asçi, F., Karabudak, E., & Kiziltan, G. (2004). Eating Attitudes and their Psychological correlates among Turkish Adolescent. *Adolescence*, 39, 593-599.
- Bilgiç, A., Yılmaz, S., Tires, S., Seda, G., & Kiliç, E.Z. (2006). Depression and Anxiety Symptom Severity in a Group of Children with Epilepsy and Related Factors. *Turkish Journal of Psychiatry*, 17(3), 165-172.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Series in Probability and Mathematical Statistics. New York: Wiley.
- Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
- Derryberry, D., & Reed, M.A. (2002). Anxiety-Related Attentional Biases and Their Regulation by Attentional Control. *Journal of Abnormal Psychology*, 111(2), 225-236.
- Hill, M.M., & Hill, A. (2000). *Investigação por Questionário*. Lisboa: Edições Sílabo.
- Hobbie, W.L., Stuber, M., Meeske, K., Wissler, K., Rourke, M.T., & Ruccione, K. (2000). Symptoms of Posttraumatic Stress in Young Adult Survivors of Childhood Cancer. *Journal of Clinical Oncology*, 18(24), 4060-4066.
- Finch, H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumption are Violated. *Methodology*, 1(1), 27-38.

- Fernandez-Berrocal, P., Alcaide, R., Extremera, N., & Pizarro, D. (2006). The Role of Emotional Intelligence in Anxiety and Depression among Adolescents. *Individual Differences Research*, 4(1), 16-26.
- Huizinga, G.A., Visser, A., Van der Graaf, W.T.A., Hoekstra, H.J., Klip, E.C., & Pras, E. (2005). Stress response symptoms in adolescent and young adult children of parents diagnosed with cancer. *European Journal of Cancer*, 41, 288-295.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Knapp, T.R. (1990). Treating Ordinal Scales as Interval Scales: An Attempt to Resolve the Controversy. *Nursing Research*, March/Jun, 121-123.
- Knapp, T.R. (1993). Treating Ordinal Scales as Ordinal Scales. *Nursing Research*, May/Jun, 184-186.
- Marôco, J. (2010). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Marôco, J. (2009). Avaliação das qualidades psicométricas de uma escala. Lisboa: ISPA, Manuscrito não publicado.
- Mitchell, J. (1986). Measurement Scales and Statistics: A Clash of Paradigms. *Psychological Bulletin*, 100(3), 398-407.
- Muris, P., Merckelbach, H., Schmidt, H., & Tierney, S. (1999). Disgust sensitivity, trait anxiety and anxiety disorders symptoms in normal children. *Behaviour Research and Therapy*, 37, 953-961.
- Muris, P., Merckelbach, H., Van Brakel, A., Mayer, B., & Van Dongen, L. (1999). The Screen for Child Anxiety Related Emotional Disorders (Scared): Relationship with Anxiety and Depression in Normal Children. *Personality and Individual Differences*, 24(4), 451-456.
- Muris, P., Steerneman, M., Merckelbach, H., & Meesters, C. (1996). The role of parental fearfulness and modeling in children's fear. *Behaviour Research and Therapy*, 34(3), 265-268.
- O'Dea, J.A., Abraham, S. (2000). Improving the Body Image, Eating Attitudes, and Behaviors of Young Male and Female Adolescents: A New Educational Approach that Focuses on Self-Esteem. *International Journal of Eating Disorders*, 28(1), 43-57.
- O'Dea, J.A., & Abraham, S. (2004). Onset of Disordered eating attitudes and behavior in early adolescence: interplay of pubertal status, gender weight and age. *Adolescence*, 39, 155-165.
- Pell, G. (2005). Uses and misuses of Likert scales. *Medical Education*, 39-970.
- Pontes, A.C.F. (2005). Análise de Variância multivariada com utilização de testes não-paramétricos e componentes principais baseados em matrizes de postos. São Paulo, ESALQ, Universidade de São Paulo. Retrieved December 29, 2010, from:
http://www.lce.esalq.usp.br/tadeu/AntonioPontes_tese.pdf.
- Rodrigues, M.J.B. (2008). Reações Emocionais e Percepções da Criança ao Conflito Parental. Porto: Instituto Superior de Ciências Biomédicas de Abel Salazar, Universidade do Porto. Retrieved December 29, 2010, from:
<http://repositorioaberto.up.pt/bitstream/10216/19395/2/ReacesEmocionaisPercepoCrianaConflitoParental.pdf>
- Sheskin, D.J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures* (4th ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. New York: Pearson Education Inc.

- Telze, E.H., Mogg, K., Bradley, B.P., Mai, X., Ernst, M., & Pine, D.S. (2008). Relationship Between Trait Anxiety, Prefrontal Cortex, and Attention Bias to Angry Faces in Children and Adolescents. *Biological Psychology*, 79(2), 216-222.
- Urduan, T.C. (2010). *Statistics in Plain English* (3rd ed.). New York: Routledge, Taylor & Francis Group.
- Vidanovic, S., & Andelkovic, V. (2006). Ego development and the anxiety of gifted Adolescents. *Philosophy, Sociology and Psychology*, 5(1), 87-102.
- Vila, G., Hayder, R., Bertrand, C., Falissard, B., De Blic, J., & Mouren-Simeoni, M-C. (2003). Psychopathology and Quality of Life for Adolescents with Asthma and their Parents. *Psychosomatics*, 44(4), 319-328.
- Vila, G., Porche, M., Mouren-Simeoni, M-C. (1999). An 18-Month Longitudinal Study of Posttraumatic Disorders in Children who were Taken Hostage in their School. *Psychosomatic Medicine*, 61, 746-754.
- Von Weiss, R.T., Rapoff, M.A., Varni, J.W., Lindsley, C.B., Olson, N.Y., & Madson, K.L. (2002). Daily Hassles and social support as predictors of adjustment in children with pediatric rheumatic disease. *Journal of Pediatric Psychology*, 27(2), 155-165.
- Wang, S.-T., Wang, C.J., Yu, M.-L., & Huang, C.-C. (1999). Bridging the Gap between the Pros and Cons in Treating Ordinal Scales from an Analysis Point of View. *Nursing Research*, July/August, 226-229.
- Weiss, D.J. (1986). The Discriminating Power of Ordinal Data. *Journal of Social Behavior and Personality*, 1(3), 381-389.
- Wolbeek, M., Van Doornen, L.J.P., Kavelaars, A., & Heijnen, C.J. (2006). Severe fatigue in adolescents: A common phenomenon. *Pediatrics*, 117, 1078-1086.
- Zimmerman, D.W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127, 354-364.

Capítulo VI. Análise de Variáveis Não Diretamente Observáveis: Influência na Tomada de Decisão Durante o Processo de Investigação

Resumo

A dimensão da amostra, o tipo de variáveis, o seu formato de medida, a construção dos instrumentos de recolha de dados válidos e fiáveis, são aspetos a ter em consideração no processo de investigação. No âmbito das ciências sociais, da saúde e especificamente na área de enfermagem, os instrumentos de recolha de dados são muitas vezes compostos por variáveis componentes ou indicadores que dão origem a variáveis latentes ou não observáveis diretamente, daí a importância da decisão relativa à forma como são medidas (escala ordinal, Likert ou de tipo Likert). As escalas psicométricas são exemplos de instrumentos, pelo tipo de variáveis que as integram, que podem trazer problemas de medida e de análise estatística (testes paramétricos versus não paramétricos). Assim o investigador quando usa estas variáveis deve respeitar alguns pressupostos baseados em estudos de simulação ou em recomendações fundamentadas na evidência científica, de forma a tomar a melhor decisão.

Descritores: variável, medida, escala.

Abstract

The sample dimension, types of variables, format used for measurement, and construction of instruments to collect valid and reliable data must be considered in during the research process. In social and health sciences, and more specifically in nursing, data-collection instruments are usually composed of latent variables or variables that cannot be directly observed. Such facts emphasize the importance of deciding how to measure study variables (using an ordinal scale or a Likert or Likert-type scale). Psychometric scales are examples of instruments that are affected by the type of variables that comprise them, which could cause problems with measurement and statistical analysis (parametric tests versus non parametric tests). Hence, investigators using these variables must rely on suppositions based on simulation studies or recommendations based on scientific evidence in order to make the best decisions.

Descriptors: Variable, measure, scale.

6.1 Introdução

Ao longo do processo de investigação o investigador torna-se perito no seu objeto de estudo e nos métodos e técnicas que vai utilizar. Passa por vários estádios e obrigatoriamente depara-se com o conceito de variável e com o pressuposto da sua medição (Watty & López, 1997). Porém, medir em saúde pode apresentar duas facetas, uma mais ligada à medida global da população (índices, taxas, etc.) e outra mais ligada à medida individual (seja ela objetiva ou subjetiva). Este tipo de medida objetiva e subjetiva emerge, respetivamente, de dois tipos de variáveis: as diretamente observáveis, relacionadas com características físicas e biológicas (e.g., peso, altura, temperatura corporal, pH, hemoglobina), e como tal mais fáceis de medir; e as não diretamente observáveis (e.g., desenvolvimento motor oral, dor, satisfação, bem-estar e saúde, capacidades de desempenho de atividades de vida diária, stress, *burnout*) (Marôco, 2010a; Hill & Hill, 2009) cuja medição é feita através das suas manifestações, e por isso mais difíceis de medir. A ideia subjacente ao conceito de medida destas variáveis passa pelo facto de ser possível avaliar atitudes, comportamentos, angústias, opiniões sobre saúde e bem-estar físico e psicológico dos indivíduos através de um processo de autodescrição ou de heteroavaliação (Agresti, 2002; Anastasi, 1990), o que se tem revelado como uma mais-valia na avaliação da saúde individual. É neste contexto que emergem as “escalas de avaliação” (Hill & Hill, 2009) compostas por variáveis qualitativas com um formato de medida ordinal cuja transformação numérica (soma, média, etc...) permite estimar o constructo latente que está subjacente, que é uma manifestação das variáveis (itens) presentes na escala (Marôco, 2010a, Marôco, 2010b; Hill & Hill, 2009). Um dos maiores problemas associado a estes instrumentos (e.g., testes psicológicos, escalas, inventários) e aos constructos que eles procuram medir é que, não havendo um fator de calibração consensual (*gold standard*), não se consegue ter a certeza de que o instrumento está a medir de forma válida, fiável e sensível o que está no seu objetivo medir (Marôco, 2009). O dilema de não haver um fator de calibração da medida de constructos psicológicos, sociológicos, de saúde, etc., torna-se assim único pelo fato de não se conseguir observar diretamente a variável latente (Marôco, 2010b). Nas áreas em que se privilegia o uso de medidas quantitativas, medir características físicas torna-se mais acessível, pelo facto de existirem instrumentos padronizados que permitem fazer essas medições. Ao contrário, a medição

de características ligadas ao comportamento humano tem sempre implícita a opinião do indivíduo e, como tal, pode levar a um acréscimo no erro de medida.

Na área das ciências da saúde e especificamente em enfermagem, os investigadores e os profissionais confrontam-se, demasiadas vezes, com problemas desta natureza, fazendo a sua discussão a dois níveis. Um ao nível da investigação empírica em que se discutem metodologias e análise de dados ordinais (instrumentos e formato de medida das variáveis, amostras, testes estatísticos); e outro ao nível clínico, em que os profissionais de saúde têm dificuldade na escolha de instrumentos de observação e avaliação para usar nas suas práticas. Estes profissionais, ao pesquisarem a melhor evidência científica, no sentido de replicarem estudos ou de traduzirem e validarem estatisticamente “escalas de avaliação”, deparam-se com uma grande diversidade de instrumentos, com diferentes formatos de medida e com diferentes formas de análise estatística, o que dificulta a sua tomada de decisão.

Todos estes aspetos têm feito parte de um quotidiano e de uma vivência profissional e de investigação, em que muitas vezes é difícil determinar qual o melhor caminho, quer na construção do projeto de investigação, quer quando se selecionam ou constroem os instrumentos de recolha de dados e se planeia a sua análise. Acerca desta temática, as discussões e as questões são muitas e, por vezes, de difícil resposta.

Destacamos algumas das questões mais comuns com que nos deparamos: O formato de medida dos itens deve ter um número par ou ímpar? Se for ímpar, há o risco de os respondentes escolherem frequentemente o ponto neutral (ancoragem)? Quantas classes de medida ou qual o formato de medida por item (3,4,5,6,7,...,10)? O formato de medida tem implicações nos testes e análises estatísticas? Se o instrumento for constituído por subescalas, é necessário estar atento ao número de itens por subescala (equilíbrio numérico dos itens que constituem subescalas ou dimensões)? Podem ser feitos, por exemplo, cálculos de médias e desvios padrão (item a item ou somente nos itens somados)? Devemos usar estatística univariada ou multivariada? Testes paramétricos ou não paramétricos? Aquando da utilização dos testes devem ser tidos em conta os seus pressupostos? A robustez dos testes tem um papel importante na escolha do teste a utilizar? A dimensão da amostra influencia a opção por um determinado tipo de teste? Algumas respostas às questões enunciadas emergem de forma clara na literatura científica, outras continuam a levantar algumas dúvidas e necessitam de mais investigação (e.g., simulação estatística com variáveis ordinais). Assim, com este trabalho, pretendemos analisar alguns aspetos a ter em conta quando se utilizam

“escalas de avaliação” compostas por variáveis não diretamente observáveis (formato de medida ordinal) e alertar para a importância da dimensão da amostra e dos testes estatísticos a usar com este tipo de variáveis.

6.2 As Variáveis e a Medida

A variável é um recetáculo de características, qualidades ou quantidades que de alguma forma transporta a informação de um determinado fenómeno descritível. A informação que é fornecida pelas variáveis em estudo é fundamental para o investigador e para o analista de dados. Porém essa informação e a sua qualidade irão depender da forma como as variáveis são quantificadas e da qualidade da sua mensuração e, mais especificamente, do erro experimental que lhe esteja associado (Marôco, 2010a; Hill & Hill, 2009). São vários os autores que classificam as variáveis, em função da informação que transportam, como quantitativas ou qualitativas.

As variáveis quantitativas são aquelas, cujo formato de medida permite a ordenação e quantificação de diferenças entre elas, podendo ser intervalares ou de razão. As variáveis intervalares ou consideradas como tal (e.g., Quociente de Inteligência, Psi 20) assumem valores quantitativos, assim como as variáveis de razão (e.g., comprimento do braço, altura, perímetro cefálico, peso) que diferem das anteriores por possuírem zero absoluto (Marôco, 2010a, 2010b; Hill & Hill, 2009; Agresti, 2002; Anastasi, 1990). As variáveis qualitativas são medidas em escalas que indicam a presença de categorias de classificação discreta dos dados, exaustivas e mutuamente exclusivas. Podem ser nominais (e.g., género, estado civil) e ordinais (e.g., escalão de risco, nível de satisfação). Nas escalas de medida ordinal, as variáveis (itens) são medidas em classes discretas, entre as quais se verifica uma ordem (apresentam uma relação descritível, mas não quantificável) (Marôco, 2010a; Hill & Hill, 2009).

Medir é um processo de observação e registo de informação sobre a forma de atributos que reflete as qualidades ou as quantidades (Marôco, 2009), ou seja, atribuem-se números a objetos ou indivíduos seguindo um determinado conjunto de regras. Porém a mensuração de dimensões físicas e químicas em que se usam instrumentos com calibração consensual (e.g., balança, pipeta, termómetro) é diferente de medir sem padrão de calibração, como acontece muitas vezes em determinadas áreas do conhecimento. As ciências exatas trabalham com variáveis observáveis e manipuláveis diretamente ou acontecimentos manifestos, enquanto as ciências sociais e humanas e da

saúde usam, muitas vezes, variáveis que não são observáveis nem manipuláveis diretamente (Marôco, 2010b), as chamadas variáveis latentes.

A construção de instrumentos com variáveis ordinais tem sido usada com o objetivo de medir grandezas que não são mensuráveis diretamente. Uma das referências mais antigas é a escala de avaliação do “brilho das estrelas” (escala com seis pontos usada por Hiparco, 150 anos AC) (Lodge, 1981). A utilização destas escalas teve uma grande expansão ao longo do século XX, com grande aceitação nas ciências sociais e humanas e também em áreas onde, apesar de se usarem mais variáveis quantitativas, se usam também este tipo de instrumentos com variáveis qualitativas que funcionam independentemente na investigação, ou como complemento às variáveis quantitativas.

Algumas das escalas mais conhecidas são a Escalas de *Likert* (Likert, 1932), a Escala de *Thurstone* (Thurstone, 1928) e a Escala de *Guttman* (Guttman, 1950). Construir instrumentos com variáveis qualitativas, cujo objeto de mensuração não é diretamente observável, dificulta o trabalho do investigador e a dificuldade continua aquando da análise de dados. Rensis Likert foi um dos investigadores que trabalhou de forma sistematizada com este tipo de variáveis (Likert, 1932). A metodologia é uma das mais usadas nos mais diversos campos da investigação e particularmente na área da psicologia, saúde e educação médica. Os estudos deste autor preconizam um método específico de construção de escalas que utilizam afirmações que permitem que pessoas com diferentes opiniões e diferentes pontos de vista respondam diferenciadamente (Likert, 1932). Na construção deste tipo de escalas de medida por item, o autor considerou a utilização de um número ímpar de pontos por item, em que o ponto central é considerado neutro e os extremos são opostos e simétricos. Destas emergiram outro tipo de escalas de medida, chamadas de tipo *Likert*, em que apesar de os itens possuírem um formato de medida ordinal, ou não existe ponto neutro (ponto central) ou os extremos não são opostos nem simétricos (Marôco, 2009). Após estudos de fiabilidade e análise dos diferentes itens, o autor sugeriu que a medida de atitude, comportamento ou outra, fosse o resultado da soma das valorações dos itens elegidos (escalas adicionadas) (Likert, 1932).

A Escala de *Thurstone* é constituída por itens com diferentes pesos, em que o respondente deve manifestar o seu acordo ou desacordo e a atitude do sujeito é medida tendo por base a média ponderada dos itens em acordo (Thurstone, 1928). A Escala Cumulativa de *Guttman* é também constituída por itens, em que o respondente tem que dar o seu acordo ou desacordo, mas cuja organização é feita de forma hierárquica, pelo

que a construção do instrumento exige que a ordenação dos itens seja efetuada do menos favorável para o mais favorável. Assim, se o respondente concorda com um item, isso implica que está em concordância com os anteriores (Guttman, 1950). Estes dois tipos de escalas, apesar de apresentarem diferenças relativamente às de Likert, endereçam o mesmo problema, da mensuração de variáveis latentes, não observáveis diretamente.

A construção de escalas de avaliação, psicométricas ou sociométricas, começou a ser feita por investigadores da área das ciências sociais, sobretudo da psicologia, sendo depois também apropriada pelas ciências da saúde. Estes instrumentos (Guttman, 1950; Stevens, 1946; Likert, 1932; Thurstone, 1928) têm gerado grande discussão, sobretudo nas questões ligadas à forma de medir as variáveis. Assim, quando se pretende analisar constructos recorrendo a este tipo de escalas, é fundamental definir a base conceptual que descreve o constructo e o suporte empírico que a escala ou instrumento de medida dá ao constructo (Marôco, 2009). Portanto, a operacionalização do conceito passa pela especificação do mesmo nas suas várias componentes (indicadores), uma vez que o conceito corresponde mais a um conjunto complexo de fenómenos do que a um fenómeno simples e diretamente observável.

Os indicadores (variáveis ou itens) são chamados conceitos operacionais e classificatórios que assumem vários valores, passíveis de serem medidos (formato de medida ordinal) (e.g. conceito de *burnout*, avaliado pela *Maslach Burnout Inventory*, que permite observar a resposta prolongada, no tempo, ao stress e que é constituída por 15 itens ordinais que refletem estados emocionais e sentimentos, organizados em três dimensões, compostas por 5, 4, 6 itens, respetivamente, com um formato de medida ordinal (com 7 classes) e que somados permitem avaliar as dimensões: exaustão emocional, despersonalização, redução da realização pessoal).

Os investigadores que trabalham com variáveis com formato de medida ordinal, *Likert* ou de tipo *Likert*, muitas vezes analisadas como se fossem quantitativas, sabem da controvérsia que o seu tratamento estatístico suscita (Urdan, 2005). Estes autores partem do princípio que estas variáveis que compõem as escalas usadas na área da saúde e da enfermagem, na psicologia e sociologia originam muitas vezes *scores*, que são tratados como variáveis latentes do tipo intervalar (Marôco, 2010a, 2010b). Estas variáveis latentes, por vezes, apresentam problemas com as qualidades métricas da medida pois, por um lado, podem não medir o constructo que era suposto medir (isto é, não ter validade) e, por outro lado, o formato de medida pode ser pouco consistente

(pouco fiável), podendo levar a um aumento dos erros na análise dos dados. No entanto, desde que feita usando os itens adicionados, a sua análise possibilita tratá-las como intervalares (Hill & Hill, 2009), o que é justificado, empiricamente, por estudos de simulação (Carifio & Perla, 2008; Glass, Peckham & Sanders, 1972), e conceptualmente:

Likert scales (collections of items) as opposed to individual Likert items are not ordinal in character, but rather are interval in nature and, thus, may be analyzed parametrically with all the associated benefits and power of these higher levels of analyses. (Carifio & Perla, 2008, p. 1151)

Apesar da controvérsia relacionada com a análise destas variáveis como se fossem quantitativas, existem estudos que defendem a sua utilização, justificando que os resultados são de confiança desde que se usem escalas com classes de medida de pelo menos cinco pontos e que a distribuição das classes seja próxima da distribuição normal (Marôco, 2010b; Worthington & Whittaker, 2006), o que possibilita o uso de testes paramétricos e não paramétricos:

It is, therefore, as the intervalist contend, perfectly appropriate to summarize the ratings generated from Likert scales using means and standard deviations, and it is perfectly appropriate to use parametric techniques like Analysis of Variance to analyze Likert scales. (Carifio & Perla, 2008, p. 1151)

Podemos assim dizer que a transformação das variáveis manifestas (soma, média dos itens) que irão permitir estimar o constructo (variável latente) que a escala procura medir emerge no centro da discussão.

Os investigadores mais tradicionais não se reveem na transformação de itens ordinais para obter uma medida do constructo. Na falta de uma medida precisa (régua), pode-se operacionalizar uma medida contínua (linha com 100 mm) usando uma escala ordinal com sete pontos ancorados como formato de resposta para afirmações de atitude ou frases derivadas de um diferencial semântico (Carifio & Perla, 2007).

Produzem-se assim dados que empiricamente são conjuntos de itens sujeitos a transformação linear e, em carácter, intervalares, de forma a podermos operacionalizar a medida (Pell, 2005; Wang et al., 1999; Knapp, 1990, 1993; Gaito, 1980). Como tal podem ser analisados através de estatística paramétrica desde que sejam verificados os pressupostos deste tipo de testes, com todos os benefícios do uso destes pelo facto de serem considerados mais potentes (Carifio & Perla, 2008; Worthington & Whittaker, 2006). Em 2004, este assunto é abordado num artigo no *Journal of Social Sciences*. A

ênfase do mesmo centrava-se em abusos na utilização de instrumentos com escalas de Likert e na escolha de metodologia para análise de resultados.

The response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal. (Jamieson, 2004, p. 1217). Esta afirmação é, porém, contestada por outros autores (Carifio & Perla, 2007; Knapp, 1993) que se referem a estes aspetos no artigo e à literatura consultada como estando impregnados de mal-entendidos e erros grosseiros que advêm de uma informação baseada em mitos, não verdades e erros conceptuais acerca das escalas de *Likert*. Pois quando se parte de uma premissa errada e não se entende ou ainda não se está familiarizado com as fontes primárias dificilmente se chega a uma teoria (Carifio & Perla, 2007; Knapp, 1993).

Historically, there has been debate between those who maintain the ordinalist (rank order) and intervalist views in Likert scales. (Carifio & Perla, 2008, p. 1150).

No que concerne à construção dos instrumentos, as escalas de avaliação, aparecem na literatura científica de diferentes formas, apresentando um número de classes de medida (formato de medida dos itens) que oscilam entre dois e dez pontos. No caso dos instrumentos compostos por um conjunto de subescalas, estas nem sempre apresentam uma construção homogénea, no que concerne ao número de itens que as compõem, encontramos na literatura subescalas que agrupam mais ou menos o mesmo número de itens (e.g., um instrumento com duas subescalas com vinte itens cada) ou um instrumento com cinco dimensões, com 2, 3, 4, 5, 7 itens respetivamente) (Wilson, 2009; Gliem & Gliem, 2003).

A construção deste tipo de instrumentos deve seguir alguns fundamentos relativamente à qualidade e quantidade dos itens (variáveis). Quanto ao formato de medida, a preocupação não deve estar centrada na questão do número par ou ímpar, mas sim no número de classes por item. Se as opções forem entre dois (e.g., 0=Não e 1=Sim), três (e.g., 1=Insatisfeito, 2=Nem satisfeito nem insatisfeito, 3=Satisfeito), cinco (e.g., 1=Discordo totalmente, 2=Discordo, 3=Indeciso, 4=Concordo, 5=Concordo totalmente) e sete (e.g., 1=Nunca, 2=Quase nunca, 3=Algumas vezes, 4=Regularmente, 5=Bastantes vezes, 6=Quase Sempre, 7=Sempre), importa considerar dois aspetos. Um relacionado com o número de possibilidades que o respondente tem e o outro com a sensibilidade dos itens. Assim, quanto maior o número de classes do item, maior a possibilidade dada ao respondente, o que se vai refletir na sensibilidade dos itens, isto é, na capacidade que o item apresenta para discriminar indivíduos estruturalmente diferentes (Marôco, 2009; Gliem & Gliem, 2003). Os itens com maior número de

classes são, geralmente mais sensíveis, e dão maior garantia de credibilidade aos resultados da análise estatística (Hill & Hill, 2009; Marôco, 2009; Carifio & Perla, 2007). Como já foi referido, algumas escalas são compostas por subescalas ou dimensões. Nestas situações é aconselhável que cada subescala ou dimensão tenha entre cinco e dez itens, devendo no mínimo conter três (Marôco, 2009). O não cumprimento deste pressuposto pode levar a problemas relacionados com a fiabilidade⁹ e com a validade¹⁰ (Marôco, 2009; Pasquali, 2009; Anastasi, 1999). A validade reporta-se a diferentes aspetos como: 1) a validade de conteúdo¹¹; 2) a validade de constructo¹², que pode ser convergente¹³, discriminante¹⁴, de critério¹⁵ (Marôco, 2009; Worthington & Whitaker; Anastasi, 1999). Assim, a heterogeneidade na construção dos instrumentos, também pode trazer problemas relativos à validade e à fiabilidade da medida que se está a tentar operacionalizar.

O alfa de Cronbach é usado com alguma frequência pelos investigadores para estimar a fiabilidade (consistência interna dos itens). No entanto, pode apresentar valores elevados quando afetados por determinados fatores como: o número de itens da escala (quanto maior o seu número, em princípio, maior o valor do alfa); a variabilidade intra e inter-itens (quanto menor a variabilidade das respostas intra-sujeito e maior a variabilidade das respostas inter-sujeito, maior o valor do alfa); a homogeneidade de variâncias inter-itens (quando existe homogeneidade de variâncias inter-itens, maior o valor do alfa) (Hill & Hill, 2009; Marôco & Garcia-Marques, 2006); e a dimensão da amostra (quanto maior a dimensão da amostra maior o valor do alfa). Alguns desses aspetos exigem que antes de calcular o alfa, standardizem-se as observações usando as correlações médias entre itens (covariâncias standardizadas) para corrigir a sobrestimação do alfa provocada pelas covariâncias heterogéneas inter-itens (Marôco & Garcia-Marques, 2006).

⁹A escala mede o constructo de forma consistente e reprodutível; sujeitos com a mesma característica, ou o mesmo sujeito em momentos equivalentes, apresentam o mesmo valor na medida.

¹⁰Traduz o grau de concordância de um painel de especialistas e avalia se os itens são representativos do domínio que a escala vai avaliar.

¹¹Se a escala mede o que é presumível medir.

¹²Mede o que se pretende medir, com base numa teoria (constructo) subjacente.

¹³Os itens que fazem parte do constructo estão correlacionados.

¹⁴Os itens não se correlacionam com outros que medem outro constructo.

¹⁵A operacionalização de um constructo está de acordo com um critério previamente estabelecido.

6.3 Dimensão da Amostra

O processo de investigação passa por diferentes estádios, dos quais já salientámos a importância das variáveis e, neste caso específico, das variáveis não observáveis diretamente e da forma como são medidas. Importa ainda discutir a questão de “onde vão ser medidas”, ou seja, se vão ser medidas na população ou numa amostra dessa população e de que forma vão ser medidas. Trabalhar com a população teórica (todos os elementos, casos, eventos, objetos, indivíduos, agregáveis e sobre os quais queremos generalizar) é quase impossível em investigação, por isso se recorre a grupos mais restritos (população acessível, população de estudo), dos quais vamos determinar a amostra (subconjunto da população que vai ser usado no estudo e que deve ser representativa da população) (Marôco, 2010a). Em ciências da saúde e da enfermagem, é muito difícil (e em algumas situações impossível) aceder a populações, sendo mais fácil usar uma amostra, a qual pode ser representativa da população (se for utilizado algum tipo de amostragem probabilística) ou pode não representar convenientemente as características da população (se for utilizado uma amostragem não probabilística) (Marôco, 2010a). No domínio em análise, é de relevante importância a dimensão da amostra. Este é um dos problemas mais discutidos na literatura científica e em debates entre investigadores. Quando se quer fazer inferência estatística, dependendo do tipo de testes estatísticos que vão ser usados, existem pressupostos aos quais o investigador deve obedecer, de modo a não comprometer a validade dos resultados obtidos. Porém, a dimensão da amostra pode ser uma questão difícil de contornar sobretudo em áreas como a saúde, em que muitas vezes o número de indivíduos é muito pequeno (e.g., doenças raras, situações de saúde específicas). Neste contexto, há investigadores que optam por obter uma amostra, com uma dimensão tão grande quanto possível, dentro do número de indivíduos disponíveis, ficando muitas vezes, mesmo assim, perante uma amostra de dimensão tão pequena que não podem fazer uso de técnicas de análise multivariadas.

Alguns investigadores optam por determinar a dimensão mínima da amostra necessária para efetuar uma análise estatística adequada, usando para tal “regras de polegar” (*rules of thumb*) (Hill & Hill, 2009). Estas são baseadas na experiência da investigação e têm como pressuposto de base a complexidade das técnicas de análise (univariadas, bivariadas ou multivariadas). Há investigadores que fundamentam a sua decisão em estudos de análise estatística da potência dos testes de modo a garantir que a

amostra tem dimensão adequada (Hill & Hill, 2009). Quando estamos a trabalhar com instrumentos, como os que temos vindo a discutir, se as amostras são muito pequenas podem ocorrer erros de estimação (Stevens, 1946), sobretudo se a relação entre o número de sujeitos e o número de itens é baixa e também se a amostra não é representativa da população pretendida, o que se verifica na prática, em muitas situações. Assim, a dimensão da amostra deve garantir os objetivos e a qualidade da investigação, devendo o investigador respeitar algumas regras de polegar e adequar o tamanho da amostra segundo o teste estatístico que pretende aplicar, a potência pretendida para o teste, o tamanho do efeito (*effect size*) que se pretende detetar e o valor do nível de significância (Marôco, 2010a; Hill & Hill, 2009).

O número de participantes está muitas vezes relacionado com o tipo de estudo e com as metodologias de análise de dados. Se a análise passa por técnicas de análise estatística como a análise da regressão e a análise fatorial exploratória devem manter-se no mínimo cinco observações por variável no modelo (Marôco, 2010b; Hill & Hill, 2009; Gliem & Gliem, 2003; Curado, Teles & Marôco, 2013). Este pressuposto advém da necessidade de existência de variabilidade para estimar os parâmetros do modelo. No caso de estarmos perante um modelo de equações estruturais, onde os “dados” do modelo são as variâncias-covariâncias não redundantes entre as variáveis, a dimensão da amostra deve ser maior. Para que haja garantia de variabilidade para estimar os parâmetros do modelo, o número de observações por cada variável manifesta deve ser entre dez e quinze (Marôco, 2010b; Hill & Hill, 2009).

Nas ciências sociais e da saúde, os investigadores têm usado estatística não paramétrica em alternativa à estatística paramétrica, com variáveis ordinais, quando os pressupostos desta não se verificam (distribuição normal, homogeneidade de variâncias, independência das observações). Contudo, estudos de simulação (Zimmerman, 2000) têm sugerido que alguns testes não paramétricos (e.g., teste U de Mann-Whitney e teste de Kruskal-Wallis) são tão sensíveis quanto os paramétricos à violação de algumas condições de aplicação, o que se traduz na inflação das probabilidades dos erros de tipo I e II, facto esse que não favorece a opção de usar testes não paramétricos.

A ausência de recomendações claras relativamente ao uso da estatística paramétrica versus não paramétrica com variáveis ordinais, sobretudo no caso multivariado tem feito com que os investigadores invistam cada vez mais em estudos de simulação estatística (Finch, 2005; Katz & MacSweeney, 1980). Alguns estudos salientam a importância da análise multivariada, quando a entidade a medir tem várias

componentes conceituais e o investigador pretender comparar grupos simultaneamente nessas componentes (e.g., Análise da variância multivariada).

A MANOVA oferece vantagem sobre múltiplas ANOVAs, pois permite medir os vários ângulos do problema, aumentar a potência (em alguns casos), e diminuir a probabilidade do erro de tipo I (Hill & Hill, 2009; Finch, 2005), mas exige o cumprimento de postulados como: (i) as observações devem ser independentes, (ii) o vetor de variáveis resposta deve ter distribuição normal multivariada para as diferentes populações, e (iii) as matrizes de variância-covariância das populações devem ser homogêneas (Marôco, 2010a), o que nem sempre se verifica na prática.

6.4 Considerações Finais

As considerações finais que iremos apresentar reportam a importância de dar às variáveis e sua medida e à dimensão da amostra quando trabalhamos com variáveis que não são observáveis diretamente. Procuraremos também deixar algumas recomendações teóricas para serem usadas por jovens investigadores na área das ciências da saúde, especificamente na enfermagem, e nas ciências sociais. A análise dos aspetos teóricos e da investigação com variáveis não diretamente observáveis permite-nos dizer que as tomadas de decisão devem ser ponderadas no sentido de diminuir a margem de erro, seja ele de medida ou de análise de dados, pois a construção e aplicação de instrumentos compostos por estas variáveis é complexa e nem sempre compreendida.

Os estudos que seguem a teoria de *Likert* utilizam instrumentos cujas variáveis (itens) apresentam um número ímpar de classes de medida (cinco pontos) e a medida dos itens não é feita individualmente obtida pela soma das valorizações dos mesmos (*score*), justificando a denominação de “escalas adicionadas” (*summated scales*). Esses *scores* podem ser tratados estatisticamente como variáveis de tipo intervalar e, na sua análise, pode ser equacionada a utilização de testes paramétricos ou não paramétricos.

A discussão sobre o número de classes da variável (item) ser par (e.g., 4, 6) ou ímpar (e.g., 5, 7, 9) é considerada irrelevante. O mesmo não pode ser dito quanto se refere a sua classificação. Para que o formato de medida seja chamado de escala ou item de *Likert*, o seu número deve ser ímpar e o ponto central é um ponto neutro cujos extremos são opostos e simétricos (e.g., Discordo Completamente-1, Discordo-2, Nem discordo nem concordo-3, Concordo-4, Concordo completamente-5). O formato de medida que não cumprir este aspeto é chamado de tipo *Likert*, o que pode ser importante, sob o ponto de vista do rigor, quando se descreve a metodologia de um

trabalho de investigação. Relativamente às dimensões que compõem a escala, a sua heterogeneidade pode originar problemas de fiabilidade e de validade, pelo que as recomendações apontam para um mínimo de três itens por dimensão, sendo o ideal entre cinco e vinte itens. Quanto ao número de classes por item, seja ele, par ou ímpar, alguns autores valorizam os instrumentos cujas variáveis apresentam um maior número de classes (cinco ou mais) em detrimento dos que apresentam três ou quatro classes, pois consideram que, quanto maior for o número de classes, maiores as possibilidades de resposta dos participantes e melhor a qualidade da soma dos itens, o que vai refletir-se na sua sensibilidade e fiabilidade. Mas na área da saúde, possivelmente devido a questões de aplicação prática destes instrumentos, o formato com três e quatro classes é utilizado com frequência.

Relativamente ao tipo de testes estatísticos, verificámos que a frequente utilização de testes paramétricos, mesmo quando não se cumprem as condições de aplicação, deve-se ao facto de alguns investigadores considerarem que ambos os testes são sensíveis à violação dos pressupostos com este tipo de variáveis, mas os paramétricos são mais potentes do que os não paramétricos. Se a análise passa por técnicas usuais de análise estatística como a análise da regressão ou a análise fatorial exploratória devem manter-se no mínimo, cinco observações por variável (item) no modelo. Mas, se optarmos por uma análise de equações estruturais, em que os dados são as correlações (ou as variâncias-covariâncias) entre as variáveis manifestas, teremos que manter um número mínimo de dez a quinze observações por variável no modelo.

A evidência científica mostra que a dimensão da amostra influencia de forma determinante a análise de dados. Assim, ao reportar os resultados de uma investigação, em que foram usados testes de hipóteses, o investigador deve incluir, além do nível de significância, uma medida de dimensão de efeito, a potência do teste e, dependendo da análise efetuada, intervalos de confiança para as estimativas dos parâmetros. Como as amostras de pequena dimensão podem originar resultados estatisticamente não significativos em situações em que têm significância prática e as de grande dimensão podem, ao contrário, originar resultados estatisticamente significativos, mesmo que sem significância prática, importa por isso que estejam expressos os resultados referentes quer à significância estatística quer à significância prática. A presença de resultados opostos nestes dois tipos de significância pode dever-se à dimensão da amostra, ao erro de tipo I e à potência de teste. Assim, na fase de planeamento da investigação, os investigadores menos experientes devem ter em atenção, por um lado, a escolha e a

construção do instrumento (único ou com subescalas), o número de itens (variáveis) que o compõem e o formato de medida que lhe está associado. Por outro lado, devem estar atentos à dimensão da amostra (e.g., ter em atenção o número de observações por item) e decidir as técnicas estatísticas a utilizar com base nestes dois fatores. Importa realçar que na área das ciências da saúde e especificamente na disciplina de enfermagem deve haver uma atenção redobrada no que respeita à dimensão da amostra, sobretudo quando estamos perante amostras de pequena dimensão (doenças raras, etc.).

Algumas das dúvidas iniciais vão continuar por esclarecer, pois trabalhar com variáveis não diretamente observáveis é um problema real e ainda há muito trabalho a desenvolver para que mais recomendações possam emergir da pesquisa. Fica por isso a sugestão para que os profissionais de saúde e outros profissionais que usam este tipo de variáveis em investigação e na prática de cuidados se envolvam em projetos de investigação conjuntos. Isto possibilita, por um lado, que equipas multiprofissionais investiguem e discutam estas problemáticas sob diferentes pontos de vista e contribuam para o esclarecimento das dúvidas que ainda persistem e, por outro lado, que cada vez mais, os instrumentos usados nas práticas de cuidados sejam fundamentados na melhor evidência científica, pois só através do conhecimento que emerge da investigação e da sua apropriação pelos profissionais, se faz a otimização dos resultados em saúde e se favorecem as “boas práticas”.

6.6 Referências Bibliográficas

- Agresti, A. (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons, Inc.
- Anastasi, A. (1990). *Psychological Testing*. New York: Macmillan Publishing Company.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
- Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Curado, M.A.S., Teles, J., & Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, 446-457. <http://www.um.es/eglobal>.
- Curado, M.A.S., Teles, J. & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process. *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48, 1, 146-52.
- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions are Violated. *Methodology*, 1(1), 27-38.
- Gaito, J. (1980). Measurement Scales and Statistics: Resurgence of an Old Misconception. *Psychological Bulletin*, 87(3), 564-567.

- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237-88.
- Gliem, J.A., & Gliem, R.R. (2003). *Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales*. Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, OH.
- Gutman, L. (1950). The basis for Scalogram analysis. In Stouffer et al. *Measurement and Prediction. The American Soldier*, Vol. IV. New York: Wiley.
- Hill, M.M., & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Kahn, J. (2006). Factor Analysis in Counseling Psychology Research, Training, and Practice: Principles, Advances and Applications. *The Counseling Psychologist*, 34(5), 684-718.
- Katz, B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Knapp, T.R. (1990). Treating Ordinal Scales as Interval Scales: An Attempt to Resolve the Controversy. *Nursing Research*, 39(2), 121-123.
- Knapp, T.R. (1993). Treating Ordinal Scales as Ordinal Scales. *Nursing Research*, 42, 184-186.
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1-50.
- Lodge, M. (1981). *Magnitude Scaling: Quantitative Measurement of Opinions*. Beverly Hills: Sage.
- Marôco, J. (2010a). *Análise Estatística com o SPSS*. Pêro Pinheiro: Report Number, Lda.
- Marôco, J. (2010b). *Análise de Equações Estruturais. Fundamentos teóricos, Software e Aplicações*. Pêro Pinheiro: Report Number, Lda.
- Marôco, J. (2009). *Avaliação das qualidades psicométricas de uma escala*. Lisboa: Manuscripto, não publicado.
- Marôco, J., & Garcia-Marques, T. (2006). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Laboratório de Psicologia*, 4(1), 65-90.
- Pasquali, L. (2009). Psychometrics. *Revista da Escola de Enfermagem da USP*, 43, 992-999.
- Pell, G. (2005). Use and misuse of Likert Scales. *Medical Education*, 39, 970.
- Wang, S.-T., Yu, M.-L., Wang, C.-J., & Huang, C.-C. (1999). Bridging the Gap between the Pros and Cons in Treating Ordinal Scales as Interval Scales from an Analysis Point of View. *Nursing Research*, 48(4), 226-229.
- Watty, A.D., & López, J.L. (1997). La importancia de medir. *Veterinaria México*, 28(1), 69-72.
- Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 667-680.
- Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Urdan, T.C. (2005). *Statistics in Plain English*. London: Laurence Erlbaum Associates.
- Wilson, F.C. (2009). Analysis of intensive outpatient neuro-rehabilitation outcomes using FIM+FAM (UK). *Neurorehabilitation*, 24, 377-382.
- Worthington, R., & Whittaker, T. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, 34(6), 806-838.
- Zimmerman, D.W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127, 354-364.

PARTE III. Estudo de Simulação de Monte Carlo

Introdução

Na parte três apresenta-se o estudo de simulação pelo método de Monte Carlo, o qual inclui o suporte teórico da geração aleatória de dados ordinais e da MANOVA paramétrica e não paramétrica, faz-se o delineamento do estudo, são apresentados e discutidos os resultados e as respectivas conclusões.

Capítulo VII. MANOVA Paramétrica e não Paramétrica com Variáveis Ordinais: A Probabilidade de Erro de Tipo I e a Potência de Teste

Resumo

Estudo de simulação, através do Método de Monte Carlo, efetuado no Software R, com o objetivo de comparar a performance, em termos de potência e probabilidade de erro de tipo I, das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica quando se utilizam variáveis ordinais correlacionadas, geradas aleatoriamente. No caso paramétrico foram consideradas 4 estatísticas de teste, Traço de Pillai, lambda de Wilks, traço de Hotelling e maior raiz de Roy, e no caso não paramétrico 2 estatísticas de teste, o lambda de Wilks e o traço de Pillai aplicados às ordens.

O delineamento do estudo de simulação incluiu 3 variáveis dependentes, uma variável independente (fator com três grupos ou níveis), escalas de avaliação com um formato de medida com 3, 4, 5, e 7 pontos, diferentes probabilidades marginais (p_1 para distribuição simétrica, p_2 para distribuição assimétrica positiva, p_3 para distribuição assimétrica negativa e p_4 para distribuição uniforme) em cada um dos três grupos, correlações de baixa, média e elevada magnitude ($r=0.10$, $r=0.40$, $r=0.70$, respectivamente), e seis dimensões de amostras ($n=30$, 60, 90, 120, 240, 300).

A análise dos resultados permite dizer que a maior raiz de Roy foi a estatística que apresentou valores mais elevados de probabilidade de erro de tipo I, sendo também uma das estatísticas com potências mais elevadas. A potência dos testes apresenta comportamentos diferentes, dependendo da distribuição de frequência, da magnitude das correlações entre itens, da dimensão da amostra e do formato de medida da escala.

Podemos assim concluir que a probabilidade de erro de tipo I é muito elevada (superior a 0.14) na estatística maior raiz de Roy e, de um modo geral inferior ao nível de significância nominal ($\alpha = 0.05$) ou ligeiramente superior nas restantes estatísticas da MANOVA paramétrica e da não paramétrica.

Na análise dos resultados da potência dos testes e tendo por base distribuição de frequência, identificámos 3 situações distintas: a primeira (com probabilidades marginais p_1, p_1, p_4 e p_4, p_4, p_1) em que se verificaram potências muito baixas para qualquer um dos cenários de simulação, devido ao facto de os grupos serem muito semelhantes e a MANOVA não identificar as diferenças; uma segunda situação (com probabilidades marginais p_2, p_3, p_4 ; p_1, p_2, p_3 e p_2, p_2, p_3) em que a magnitude das potências é elevada, nas amostras com dimensão superior ou igual a 60 observações e nas escalas com 3, 4, 5 pontos e potências de magnitude menos elevada nas escalas com 7 pontos, mas com a mesma magnitude nas amostras com dimensão igual a 120 observações para todos os cenários simulados; uma terceira situação (com probabilidades marginais p_1, p_1, p_2 ; p_1, p_2, p_4 ; p_2, p_2, p_1 ; p_4, p_4, p_2 e p_2, p_2, p_4) em que quanto maiores, a intensidade das correlações entre itens e o número de pontos da escala, e menor a dimensão das amostras, menor a potência dos testes, sendo o lambda de Wilks aplicado às ordens mais potente do que todas as outras estatísticas da MANOVA, com valores imediatamente a seguir à maior raiz de Roy. No entanto, a magnitude das potências dos testes paramétricos e não paramétricos assemelha-se nas amostras com dimensão superior a 90 observações (com correlações de baixa e média magnitude), nas escalas com 3, 4 e 5 pontos; e superiores a 240 observações, (para correlações de baixa intensidade) nas escalas com 7 pontos.

No estudo de simulação e tendo por base a distribuição de frequência, concluímos que na primeira situação de simulação e para os diferentes cenários, as potências são de baixa magnitude devido ao facto de a MANOVA não detetar diferenças entre grupos pela sua similaridade. Na segunda situação de simulação e para os diferentes cenários, a magnitude das potências é elevada em todos os cenários cuja dimensão da amostra seja superior a 60 observações, pelo que é possível aplicar testes paramétricos.

Na terceira situação de simulação, e para os diferentes cenários quanto menor a dimensão da amostra e mais elevada a intensidade das correlações e o número de pontos da escala, menor a potência dos testes, sendo a magnitude das potências mais elevadas no teste de Wilks aplicado às ordens, seguido do traço de Pillai aplicado às ordens. No entanto, a magnitude das potências dos testes paramétricos e não paramétricos assemelha-se nas amostras com maior dimensão e correlações de baixa e média magnitude.

Palavras-chave: variáveis ordinais, simulação de Monte Carlo, probabilidade de erro de tipo I, potência, MANOVA paramétrica e não paramétrica.

Abstract

Using R software, a simulation study was developed to evaluate and compare the performance of four parametric (Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root) and two non-parametric tests (Pillai's Trace and Wilks' Lambda applied to the rank transformed data). In the data generation process, scales with 3, 4, 5 and 7 points were considered and a vector of three dependent variables, one independent variable with three levels, different marginal distributions in groups (symmetric, p_1 ; positive skew, p_2 ; negative skew, p_3 ; and uniform, p_4 ; different correlation structures (low, $r=0.10$; average, $r=0.40$ and high, $r=0.70$), and six sample sizes (30, 60, 90, 120, 240, 300). A Monte Carlo resampling simulation (1000 replications) was carried out to estimate the power and type I error rate from the proportion of rejected null hypotheses at a 5% significance level of parametric and non-parametric one-way MANOVA with ordinal data.

The analysis of the simulation results allows us to conclude that Roy's largest root was the statistic that presented the highest values of the probability of type I error, and was also one of the statistics with the biggest statistical power rate. The power presents different behaviours depending on the distribution, the magnitude of correlation between items, the sample dimension, and the points of the scale.

Based on the frequency distribution, the data analysis of the power results allows us to identify three distinct situations. In the first situation, (with marginal probabilities

p_1, p_1, p_4 e p_4, p_4, p_1) the power is of low magnitude due to the fact that MANOVA does not detect differences between groups due to their similarity. In the second situation, (with marginal probabilities p_2, p_3, p_4 ; p_1, p_2, p_3 e p_2, p_2, p_3) the magnitude of the power rate depends on the sample size and the number of scale points, and in different scenarios, the power has a higher magnitude if the sample size is larger, for statistics of parametric and non-parametric MANOVA. In the third situation, (with marginal probabilities p_1, p_1, p_2 ; p_1, p_2, p_4 ; p_2, p_2, p_1 ; p_4, p_4, p_2 e p_2, p_2, p_4) the magnitude of the power rate of MANOVA depends on a combination of the correlation between dependent variables, the sample size, and the number of scale points. As the correlations and number of scale points increased and the sample size decreased, the power rate of MANOVA decreased, and Wilks' lambda applied to the rank transformed data had a higher power rate than the other statistics. But the power magnitude is similar in all statistics, parametric and non-parametric, if the sample size is larger and the magnitude correlation decreased to low and average magnitude.

Considering the frequency distribution, in the second situation, we can use parametric statistics. In the third situation we can use non-parametric statistics with small sample sizes and, average correlation (scales with 3, 4, 5, points) and low correlation (scales with 7 points). With larger sample size and low and average magnitude of correlation we can use parametric statistics.

Key-words: ordinal data, Monte Carlo simulation, power, type I error, parametric and non-parametric MANOVA.

7.1 Introdução

A simulação é uma técnica de pesquisa operacional que usa experiências numéricas com modelos lógico-matemáticos que envolvem um grande número de cálculos e um grande número de repetições, pelo que necessitam de uso intensivo de recursos computacionais (Saliby, 1989). Este problema deixou de existir com o aumento da velocidade de processamento e com a capacidade de armazenamento dos computadores, associadas à evolução das próprias metodologias de simulação (Yoriyaz, 2009), das quais destacamos a simulação pelo método de Monte Carlo.

Um dos exemplos mais comuns, de utilização do método de Monte Carlo, com este propósito é a análise de variância univariada e multivariada, estando esta ligada à rejeição ou não da hipótese nula e ao erro na tomada desta decisão (Ramsey & Ramsey, 2013, 2011; Todorov & Filzmoser, 2010; Finch & Davenport, 2009; Bilodeau & Micheaux, 2005; Finch, 2005; Zwick, 1985; Hummel & Sligo, 1971).

Se rejeitamos a hipótese nula sendo ela verdadeira, estamos perante um erro de tipo I, cuja probabilidade é representada por (α) , se não rejeitamos a hipótese nula sendo ela falsa, estamos perante o erro de tipo II, cuja probabilidade é representada por (β) , sendo a potência de um teste o seu complemento $(1 - \beta)$ (Pearson, 1930, 1931). A metodologia de simulação tem sido usada ao longo dos anos para analisar distribuições de amostragem de diferentes estatísticas e avaliar em que medida a violação de pressupostos origina alteração na potência dos testes.

A simulação de Monte Carlo é amplamente utilizada na área dos testes estatísticos. A estatística de teste que permite a tomada de decisão relativamente à rejeição ou não da hipótese nula é representada por $T = T(X_1, X_2, \dots, X_n)$, o nível de significância é representado por α e a região de rejeição por R com $P[T \in R | H_0]$. O procedimento pode ser organizado em três passos:

- 1) De acordo com o modelo postulado, gerar N amostras independentes $(X_1^{(i)}, X_2^{(i)}, \dots, X_n^{(i)})$, para $i = 1, 2, \dots, N$, sendo n a dimensão da amostra e N o número de réplicas.
- 2) Calcular $T^{(i)} = T(X_1^{(i)}, X_2^{(i)}, \dots, X_n^{(i)})$, para $i = 1, 2, \dots, N$.
- 3) Determina-se qual a percentagem de rejeições de H_0 , quando esta é verdadeira, nas amostras geradas aleatoriamente pelo método de Monte Carlo.

A simulação inicia-se com a geração de amostras de tamanho n com determinados pressupostos pré-estabelecidos e em que se repetem os testes da MANOVA paramétrica e não paramétrica ($N=1000$) e se verifica em quantas dessas repetições do teste se rejeita ou não a hipótese nula.

Para estimar a probabilidade do erro de tipo I faz-se uma contagem do número de rejeições com recurso a um contador introduzido no programa que atribui $\delta = 1$ se se rejeita H_0 e $\delta = 0$ para a não rejeição de H_0 . A probabilidade do erro tipo I $\alpha = P[\text{Erro de Tipo I}] = P[T \in R | H_0]$ é obtida através do estimador da proporção

representado por $\hat{\alpha} = 1/N \sum_{i=1}^N \delta_i$ (Carrasco & Lemes, 2014; Lemes & Carrasco, 2010; Carrasco & Silva, 2009). A potência do teste $1 - \beta = P[T \in R | H_1]$ é obtida do mesmo modo.

Tendo em atenção a importância da análise multivariada, com variáveis ordinais, e a sua ampla utilização na área da saúde, o objetivo deste estudo é “comparar a performance, através de estudos de simulação, em termos de potência e probabilidade de erro de tipo I, das quatro estatísticas da MANOVA paramétrica com duas estatísticas da MANOVA não paramétrica (a um fator) com variáveis ordinais com um formato de medida com 3, 4, 5 e 7 pontos e diferentes magnitudes de correlação”.

MANOVA a um Fator

A análise da variância multivariada (MANOVA) permite fazer a comparação dos vetores de valores médios em situações em que se pretende comparar em simultâneo diversas variáveis critério (Marôco, 2010; Timm, 2002; Reis, 2001; Jobson, 1992). Na MANOVA paramétrica a um fator, pretende-se testar se os vetores de médias, referentes a p variáveis medidas em g grupos, são iguais ou se existem pelo menos dois vetores de médias que diferem entre si de forma significativa. Na sua forma matricial temos como modelo da MANOVA a um fator: $\mathbf{X}_{li} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\varepsilon}_{li}$, onde \mathbf{X}_{li} é o vetor aleatório das observações do i -ésimo indivíduo do l -ésimo grupo, com $i = 1, \dots, n_l$ e $l = 1, \dots, g$, $\boldsymbol{\mu}$ é o vetor valor médio, $\boldsymbol{\tau}_l$ representa o efeito do nível l do fator (tratamento) com $\sum_{l=1}^g n_l \boldsymbol{\tau}_l = \mathbf{0}$, e $\boldsymbol{\varepsilon}_{li}$ são os erros aleatórios, onde $\boldsymbol{\varepsilon}_{li} \sim \mathbb{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. A hipótese nula pode ser apresentada como $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \dots = \boldsymbol{\tau}_g = \mathbf{0}$ e a hipótese alternativa

é apresentada como $H_1 : \exists_l : \boldsymbol{\tau}_l \neq \mathbf{0} (l \in \{1, \dots, g\})$, onde $\boldsymbol{\tau}_l = \begin{bmatrix} \tau_{l1} \\ \tau_{l2} \\ \vdots \\ \tau_{lp} \end{bmatrix}$, com $l = 1, \dots, g$

(Johnson & Wichern, 2007). Representando por $\boldsymbol{\mu}_l = \boldsymbol{\mu} + \boldsymbol{\tau}_l$, as hipóteses podem ser dadas por: $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g$ vs $H_1 : \exists_{l,k} : \boldsymbol{\mu}_l \neq \boldsymbol{\mu}_k (l, k \in \{1, \dots, g\})$.

Os pressupostos sobre a estrutura dos dados para a MANOVA One-Way ou a um fator exigem que cada população tenha uma distribuição normal multivariada e todas as populações tenham uma matriz de variância-covariância Σ comum.

De acordo com o modelo, cada componente do vetor de observações \mathbf{X}_{li} satisfaz o modelo univariado e embora os erros das componentes de \mathbf{X}_{li} estejam correlacionados, as matrizes de variância-covariância Σ são iguais para todas as populações. Se particularizarmos o modelo populacional para as observações amostrais temos: $\mathbf{X}_{li} = \bar{\mathbf{X}} + (\bar{\mathbf{X}}_l - \bar{\mathbf{X}}) + (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)$, em que $\bar{\mathbf{X}}$ é o vetor médio amostral (estimativa de μ) e $(\bar{\mathbf{X}}_l - \bar{\mathbf{X}})$ é o efeito do nível l do tratamento (estimativa de τ_l) e $(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)$ são os resíduos (estimativa de ε_{li}). A extensão matricial das somas dos quadrados da ANOVA, a soma dos quadrados devido ao fator ou tratamento (SQF) e a soma dos quadrados do erro (SQE) respetivamente, dá origem à matriz \mathbf{B} , designada por matriz da soma de quadrados e produtos cruzados entre grupos que pode ser apresentada como $\mathbf{B} = \sum_{l=1}^g n_l (\bar{\mathbf{X}}_l - \bar{\mathbf{X}})(\bar{\mathbf{X}}_l - \bar{\mathbf{X}})'$ e à matriz \mathbf{W} denominada como matriz da soma de quadrados e produtos cruzados dentro dos grupos que pode ser representada como $\mathbf{W} = \sum_{l=1}^g \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)' = (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \dots + (n_g - 1)\mathbf{S}_g$, onde \mathbf{S}_l é a estimativa da matriz de variância-covariância do nível l do fator, definida por $\mathbf{S}_l = \sum_{i=1}^{n_l} (\mathbf{X}_{li} - \bar{\mathbf{X}}_l)(\mathbf{X}_{li} - \bar{\mathbf{X}}_l)' / (n_l - 1)$. A matriz da soma de quadrados e produtos cruzados totais que representa a variabilidade global representa-se por \mathbf{T} , sendo $\mathbf{T} = \mathbf{B} + \mathbf{W}$ (Marôco, 2010; Manly, 2005; Montgomery, 2001).

O número total de graus de liberdade, $n - 1$, também é particionado na soma de duas parcelas $g - 1$ e $n - g$. No caso da MANOVA não existem quadrados médios nem estatística F , em vez disso consideram-se os determinantes das matrizes \mathbf{T} , \mathbf{B} e \mathbf{W} .

As estatísticas de teste da MANOVA paramétrica são: 1) O lambda de Wilks,

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}; \quad 2) \text{ O traço de Pillai, } TP = tr(\mathbf{B}\mathbf{T}^{-1}) = tr(\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}); \quad 3) \text{ O traço}$$

de Hotelling $TH = tr(\mathbf{B}\mathbf{W}^{-1})$; 4) A maior raiz de Roy que é o maior dos valores

próprios da matriz $\mathbf{B}\mathbf{W}^{-1}$ (Marôco, 2010; Johnson & Wichern, 2007; Manly, 2005).

O teste não paramétrico correspondente à MANOVA a um fator é a MANOVA não paramétrica a um fator. Como já foi referido, esta técnica é considerada como uma extensão multivariada do teste de Kruskal-Wallis. Subjacente a esta técnica existe um procedimento de atribuição de ordens: as observações de cada uma das p variáveis são classificadas de 1 para n , sendo a ordem 1 atribuída à menor observação e a ordem n à maior observação. Se determinadas observações estão empatadas, a cada uma destas observações é atribuída a média das ordens que as observações teriam caso não estivessem empatadas. Após este procedimento existem p conjuntos de ordens, um conjunto para cada variável sob estudo. Para testarmos a igualdade das distribuições em todas as variáveis e grupos com a MANOVA não paramétrica, as hipóteses são: $H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = \dots = F(\mathbf{X}_g)$ vs. $H_1 : \exists l, l^* : F(\mathbf{X}_l) \neq F(\mathbf{X}_{l^*})$, com $l, l^* \in \{1, \dots, g\}$, onde $F(\mathbf{X}_l)$ é a função de distribuição multivariada de $\mathbf{X}_l = (X_{l1}, X_{l2}, \dots, X_{lp})'$, com $l = 1, \dots, g$. As duas estatísticas de teste da MANOVA não paramétrica que vamos considerar neste estudo de simulação são: o traço de Pillai e o lambda de Wilks aplicados às ordens.

O traço de Pillai aplicado às ordens tem como estatística de teste $\chi^2 = \mathbf{V}' \Sigma_{\mathbf{V}}^{-1} \mathbf{V} \sim \chi_{p(g-1)}^2$, e o vetor aleatório \mathbf{V} é representado por $\mathbf{V}_{(g-1)p \times 1} = (\mathbf{V}'_1, \mathbf{V}'_2, \dots, \mathbf{V}'_p)'$ e $\Sigma_{\mathbf{V}}$ é a matriz de variância-covariância, sendo o vetor \mathbf{V}_j definido em termos dos desvios dos primeiros $(g-1)$ grupos de médias das ordens aos respetivos valores esperados sob a validade de H_0 , sendo $\mathbf{V}_j = \left[\bar{R}_j^{(1)} - (n+1)/2, \dots, \bar{R}_j^{(g-1)} - (n+1)/2 \right]'$, $j = 1, \dots, p$ (Marôco, 2010; Katz & McSweeney, 1980). Os desvios relativos ao grupo g são inferidos dos anteriores $(g-1)$ grupos. O vetor valor médio \mathbf{V} é aproximadamente normal multivariado nas amostras de grande dimensão, com matriz de variância-covariâncias $\Sigma_{\mathbf{V}}$, de característica máxima $p(g-1)$. Sob H_0 , $E(\mathbf{V}) = \mathbf{0}$, tornando χ^2 uma variável aleatória qui-quadrado central com $p(g-1)$ graus de liberdade. Valores elevados de χ^2 indicam que existem diferenças entre \mathbf{V} e o valor esperado para \mathbf{V} sob a validade de H_0 . Assim a hipótese nula é rejeitada, ao nível de significância α , sempre que $\chi^2 \geq \chi^2_{1-\alpha; p(g-1)}$ ou se p -value do teste for menor ou igual a α (p -value $\leq \alpha$) (Katz & McSweeney, 1980, p. 284).

A estatística proposta por Katz e McSweeney (1980) pode ser apresentada como $\chi^2 = (n-1)TP_R$, onde n é a dimensão da amostra global e TP_R é o traço de Pillai calculado a partir de ordens das observações (Marôco, 2010a; Zwick, 1985).

A outra metodologia não paramétrica tem por base o Lambda de Wilks aplicado às ordens. A estatística de teste é $LW_R = |\mathbf{W}_R|/|\mathbf{T}_R|$, em que as matrizes \mathbf{W}_R e \mathbf{T}_R são as matrizes correspondentes a \mathbf{W} e \mathbf{T} , respetivamente, obtidas substituindo as observações pelas respetivas ordens. De acordo com a metodologia de “transformação de ordens”, o teste paramétrico é aplicado às ordens e a distribuição da estatística de teste LW_R é aproximada à do Lambda de Wilks paramétrico. Cada $X_{li}^{(m)}$ é substituído pela ordem correspondente, $R_{li}^{(m)}$, sendo $R_{li}^{(m)}$ a ordem associada ao i -ésimo sujeito, do l -ésimo grupo, na m -ésima variável dependente, com $l=1, \dots, g$, $m=1, \dots, p$ e $i=1, \dots, n_l$ e $1 \leq R_{li}^{(m)} \leq n$. A média das ordens por grupo na m -ésima variável pode escrever-se como $\bar{R}_{li}^{(m)}$, $l=1, \dots, g$ e $i=1, \dots, n_l$ e $\sum_{l=1}^g \sum_{i=1}^{n_l} R_{li}^{(m)} = n(n+1)/2$.

A estatística χ^2 pode ser calculada através do lambda (Λ_R) de Wilks aplicado às ordens das observações e escreve-se como $\chi^2 = -(n-1-(p+g)/2) \ln LW_R$ onde n é a dimensão da amostra global e $\ln LW_R$ é o logaritmo natural do lambda de Wilks aplicado às ordens (Nath & Pavur, 1985; Todorov & Filzmoser, 2009, 2010; Van Aelst & Willems, 2011). Rejeita-se H_0 ao nível de significância α se $\chi^2 \geq \chi_{1-\alpha; p(g-1)}^2$ ou se $p\text{-value} \leq \alpha$.

7.2 Metodologia

Para comparar a performance, em termos de potência e probabilidade de erro de tipo I, da MANOVA paramétrica e não paramétrica foi efetuado, um estudo de simulação, através do Método de Monte Carlo, no *software* R (R Development Core Team, 2012). A literatura apresenta vários estudos de comparação da performance dos testes paramétricos e não paramétricos (Blair & Higgins, 1985; Lehman, 1975), mas os estudos com variáveis ordinais são menos comuns (Finch, 2005; Nath & Pavur, 1985; Katz & MacSweeney, 1980). Considerámos o caso balanceado, vetor com 3 variáveis de pendentas e uma variável independente com $g=3$ grupos, logo as dimensões das amostras são iguais nos três grupos, isto é $n_l = n/3$ para $l=1, 2, 3$.

No delineamento do estudo de simulação foram consideradas diferentes distribuições: simétricas, (maior probabilidade de resposta localizada no ponto central da escala); assimétricas positivas, (maior probabilidade de resposta localizada no ponto inferior da escala), assimétricas negativas, (maior probabilidade de resposta localizada no ponto superior da escala) e uniformemente distribuídas, (probabilidade de resposta igualmente distribuída nos diversos pontos da escala); foram, também, considerados três tipos de correlações de diferente intensidade, na matriz de simetria composta, $r=0.10$, $r=0.40$ e $r=0.70$; seis diferentes dimensões de amostras ($n=30$, $n=60$, $n=90$, $n=120$, $n=240$, $n=300$) e um nível de significância de 5% ($\alpha = 0.05$).

No caso da MANOVA paramétrica, a veracidade da hipótese nula foi avaliada com utilização das estatísticas lambda de Wilks, traço de Pillai, traço de Hotelling, maior raiz de Roy, e para a mesma avaliação na MANOVA não paramétrica foram usadas as estatísticas lambda de Wilks e traço de Pillai aplicadas às ordens das observações (lambda de Wilks não paramétrico e traço de Pillai não paramétrico).

Para avaliar a probabilidade de erro tipo I e a probabilidade de erro tipo II da MANOVA não paramétrica, foram gerados três tipos de variáveis ordinais, com o mesmo tipo de distribuição, ou seja H_0 verdadeira ($H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = F(\mathbf{X}_3)$, verdadeira) e com distribuições diferentes, ou seja H_0 falsa ($H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) \neq F(\mathbf{X}_3)$, falsa).

Neste estudo determinou-se que o número de réplicas seriam de 1000 em cada condição, pelo facto de um dos métodos de simulação ser muito lento. As taxas de erro de tipo I e a potência do teste foram estimadas pela proporção de rejeição da hipótese nula, quando a hipótese nula é verdadeira ou quando ela é falsa, respetivamente (Ramsey & Ramsey, 2012; Van Aelst & Willems, 2011; Todorov & Filzmoser, 2010).

Assim, se verdadeira $H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = F(\mathbf{X}_3)$; com $P[\text{Erro de tipo I}] = P[\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}] = \#\{p\text{-value} \leq \alpha\} / N$, se falsa $H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = F(\mathbf{X}_3)$ e $P[\text{Erro de tipo II}] = P[\text{Não rejeitar } H_0 | H_0 \text{ é falsa}] = \#\{p\text{-value} > \alpha\} / N$, em que a potência de teste $= 1 - P[\text{Erro de tipo II}]$; com $N=1000$ e $\alpha = 0.05$. A geração aleatória de dados foi fundamentada nos algoritmos para gerar dados ordinais correlacionados segundo duas metodologias, a conversão binária e o mapeamento de médias (Kaiser, Traeger & Leisch, 2011). Na geração de dados ordinais correlacionados X_i com

valores inteiros $\{1, 2, 3, \dots, k\}$, a distribuição correspondente é definida pelo vetor de

probabilidades, $\mathbf{p}_{X_i} = \begin{pmatrix} P(X_i = 1) \\ P(X_i = 2) \\ \vdots \\ P(X_i = k) \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$.

Quando se geram números aleatórios (pseudoaleatórios) de modo a obter variáveis ordinais X_1, \dots, X_p que tomem valores inteiros $\{1, 2, \dots, k\}$, o utilizador necessita especificar as probabilidades marginais $\mathbf{p}_{X_i}, i = 1, \dots, p$ e pré-definir uma matriz de correlações semi-definida positiva:

$$\mathbf{cor} = \begin{pmatrix} r(x_1, x_1) & r(x_1, x_2) & \dots & r(x_1, x_p) \\ r(x_2, x_1) & r(x_2, x_2) & \dots & r(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_p, x_1) & r(x_p, x_2) & \dots & r(x_p, x_p) \end{pmatrix}$$

Os autores consideram que ao gerar variáveis ordinais com os possíveis valores inteiros $\{1, 2, 3, \dots, k\}$, os valores observados correspondem às ordens e como tal, a opção deve ser pela utilização de Correlações de Spearman, devido à natureza das variáveis envolvidas (as correlações subjacentes aos algoritmos desenvolvidos são as correlações de Spearman).

As bibliotecas do R utilizadas foram a *Modern Applied Statistics with S* ("MASS"); as *Applied Smoothing Techniques for Data Analysis* ("sm"); e a *Psychological, Psychometric, and Personality Research* ("psych"). Para fazer a geração aleatória de dados foi usada a biblioteca *Generation of Artificial Ordinal and Binary Data*, ("orddata", repos="http://R-Forge.R-project.org")>library("orddata"), com as funções principais *rmvord* e *rmvord_b*, as quais fazem o retorno de n observações das variáveis ordinais, com um formato de medida com diferente número de pontos, com diferentes probabilidades marginais e com diferentes correlações (Kaiser, Traeger & Leisch, 2011). No estudo de simulação, as distribuições de frequência são classificadas atendendo à simetria: distribuições simétricas (p_1), distribuições assimétricas positivas (p_2), distribuições assimétricas negativas (p_3), e distribuições uniformes (p_4) e as probabilidades marginais foram agrupadas por número de variáveis, neste caso três

variáveis (e.g., p_1, p_1, p_1 ou p_1, p_2, p_3) e têm subjacente o formato de medida das variáveis ordinais (3, 4, 5 e 7 pontos), (cf. Tabela 7. 1).

Tabela 7. 1 - Probabilidades marginais (*p.m.*) por formato de medida da escala ordinal usadas no estudo de simulação.

<i>p.m.</i> <i>simetria</i>	Formato de medida			
	3 pontos	4 pontos	5 pontos	7 pontos
<i>p</i> ₁	simétrica			
	(0.20, 0.60, 0.20)	(0.10, 0.40, 0.40, 0.10)	(0.10, 0.20, 0.40, 0.20, 0.10)	(0.10, 0.10, 0.10, 0.40, 0.10, 0.10, 0.10)
<i>p</i> ₂	assimétrica positiva			
	(0.70, 0.20, 0.10)	(0.60, 0.20, 0.10, 0.10)	(0.50, 0.20, 0.10, 0.10, 0.10)	(0.40, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10)
<i>p</i> ₃	assimétrica negativa			
	(0.10, 0.20, 0.70)	(0.10, 0.10, 0.20, 0.60)	(0.10, 0.10, 0.10, 0.20, 0.50)	(0.10, 0.10, 0.10, 0.10, 0.10, 0.10, 0.40)
<i>p</i> ₄	uniforme			
	(0.34, 0.33, 0.33)	(0.25, 0.25, 0.25, 0.25)	(0.20, 0.20, 0.20, 0.20, 0.20)	(0.14, 0.14, 0.14, 0.14, 0.14, 0.15, 0.15)

A matriz de correlações é uma matriz semi-definida positiva, como referido anteriormente, com três tipos diferentes de intensidade de correlação ($r=0.10$, $r=0.40$, $r=0.70$). Kaiser, Traeger e Leisch (2011) basearam o método de conversão binária nos estudos de Demirtas (2006). A execução do algoritmo do método de conversão binário, função *rmvord_b* da biblioteca "orddata", é rápida, mesmo que o delineamento do estudo abranja um número elevado de variáveis ou de pontos da escala, mas não funciona com valores de correlação elevados (Demirtas, 2006).

No sentido de ultrapassar este problema, Leisch, Weingressel e Hornik (1998) criaram uma metodologia alternativa, o método de mapeamento de médias. Este método gera números aleatórios ordinais com uma determinada estrutura de correlação. O método de mapeamento de médias (*rmvord*), função da biblioteca "orddata", é mais estável, aceita valores de correlação mais elevados na matriz de simetria composta que define a estrutura de correlação, mas é computacionalmente muito mais demorado. O tempo de execução deste método depende do número de pontos da escala e do número de variáveis (Kaiser, Traeger & Leisch, 2011).

Ao compararmos os dois métodos, pode dizer-se que ambos apresentam vantagens e desvantagens. Por um lado, o método de conversão binária, apesar de mais rápido, limita o conjunto de matrizes de correlação exequíveis. Por outro lado, o método de mapeamento de médias supera esse problema e tem uma precisão idêntica ao método de conversão binária, mas aumenta drasticamente o tempo de execução das simulações.

O facto de o delineamento experimental deste estudo incluir correlações de magnitude baixa, média e elevada, fez com que fosse necessário utilizar as duas metodologias, método de conversão binária e o método de mapeamento de médias para fazer a geração de dados ordinais, pois era importante perceber se o comportamento, dos testes multivariados sob estudo, era influenciado pela intensidade das correlações (Demirtas, 2006; Kaiser, Traeger & Leisch, 2011).

Ao analisarmos as correlações atingidas após simulação e comparando-as com os valores de *input*, verifica-se que a precisão aumenta à medida que o número de pontos da escala aumenta e também com o aumento das correlações (cf. Tabela 7. 2). O aumento das correlações de média e sobretudo de elevada intensidade faz com que o algoritmo limite o conjunto de matrizes de correlação exequíveis pelo método de conversão binária (Demirtas, 2006) pelo que se utilizou também o método de mapeamento de médias (Kaiser, Traeger & Leisch, 2011).

Tabela 7. 2 - Diferença máxima entre as correlações de *input* e as correlações amostrais.

Formato medida	Magnitude das Correlações								
	<i>r=0.10</i>			<i>r=0.40</i>			<i>r=0.70</i>		
	Probabilidades Marginais			Probabilidades Marginais			Probabilidades Marginais		
	<i>p</i> ₁	<i>p</i> _{2 e p} ₃	<i>p</i> ₄	<i>p</i> ₁	<i>p</i> _{2 e p} ₃	<i>p</i> ₄	<i>p</i> ₁	<i>p</i> _{2 e p} ₃	<i>p</i> ₄
3 Pontos	0.028	0.028	0.028	0.054	0.054	0.054	0.026	0.026	0.026
4 Pontos	0.016	0.016	0.016	0.038	0.038	0.038	0.017	0.017	0.017
5 Pontos	0.010	0.010	0.010	0.028	0.028	0.028	0.013	0.013	0.013
7 Pontos	0.009	0.009	0.009	0.019	0.019	0.019	0.009	0.009	0.009

Para verificar se as distribuições dos valores da probabilidade de erro de tipo I e potência de teste são iguais, para as diferentes estatísticas da MANOVA, ou apresentam diferenças estatisticamente significativas recorreu-se ao teste não paramétrico de Friedman e com recurso ao teste LSD de Fisher sobre as ordens, fizeram-se as comparações múltiplas de médias das ordens com o procedimento da ANOVA fatorial (IBM SPSS Statistic, v. 21, SPSS Inc, Chicago, IL).

7.3 Apresentação dos Resultados

A análise de dados que envolve o uso da inferência estatística, não oferece garantias de que os resultados obtidos na amostra são o reflexo dos resultados obtidos na população. O investigador ao tomar decisões baseadas na evidência experimental pode cometer alguns erros nos quais se incluem os erros estatísticos.

Apresentamos em seguida, os resultados da probabilidade de erro de tipo I e da potência (em valor percentual) dos quatro testes da MANOVA paramétrica e dos dois testes da MANOVA não paramétrica a um fator, referentes ao estudo de simulação de Monte Carlo com variáveis ordinais, com um formato de medida com 3, 4, 5, e 7 pontos e correlações de baixa, média e elevada intensidade entre os itens.

7.3.1 Probabilidade de Erro de Tipo I

Os resultados referentes à probabilidade de erro de tipo I advêm das diferentes situações simuladas para as 4 estatísticas da MANOVA paramétrica e 2 da MANOVA não paramétrica, com distribuições de frequência iguais em cada um dos 3 grupos, com a maior probabilidade de resposta aos itens localizada no ponto central da escala (distribuição de frequência simétrica), nos extremos superior ou inferior (distribuição de frequência assimétricas negativa e positiva, respetivamente) e uniformemente distribuídas nos pontos da escala de medida (distribuição uniforme); 6 amostras de diferentes dimensões, e 3 tipos de intensidade de correlação e escalas com diferentes formatos de medida (3, 4, 5 e 7 pontos) (cf. Tabela 7.3).

Tabela 7.3 - Condições testadas nas diferentes simulações (432 cenários) com quatro tipos de probabilidades marginais (p), seis dimensões amostrais (n), correlações com três tipos de magnitude (r), e escalas de avaliação com quatro formatos de medida (3, 4, 5, e 7 pontos).

Situações simuladas para a probabilidade de erro de tipo I												
Probabilidades marginais (Distribuição)	P ₁ ,P ₁ ,P ₁ simétrica			P ₂ ,P ₂ ,P ₂ assimétrica positiva			P ₃ ,P ₃ ,P ₃ assimétrica negativa			P ₄ ,P ₄ ,P ₄ uniforme		
Dimensão da amostra (n) Formato de medida	n=30, 60, 90, 120, 240, 300											
	Magnitude das correlações (r)											
3 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
4 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
5 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
7 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70

Os resultados que se destacam do estudo de simulação quando se consideram 3 variáveis ordinais, com formato de medida com 3, 4, 5 e 7 pontos, mostram que a estatística maior raiz de Roy apresenta valores de probabilidade de erro de tipo I superiores a 14% em todas as situações simuladas. As restantes estatísticas da MANOVA paramétrica e não paramétrica apresentam valores de probabilidade de erro de tipo I semelhantes e inferiores ou ligeiramente superiores ao nível de significância nominal ($\alpha=0.05$). As estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks e traço de Hotelling e as estatísticas da MANOVA não paramétrica, traço de Pillai e lambda de Wilks aplicadas às ordens, com diferentes distribuições de frequência, nos diferentes formatos de medida, apresentam maioritariamente valores de probabilidade de erro de tipo I inferiores a 5% e apenas em algumas dimensões amostrais ligeiramente superiores a 5% (7%). Para averiguar se as distribuições dos valores de probabilidade de erro de tipo I são iguais para as diferentes estatísticas da MANOVA ou se existe pelo menos uma distribuição significativamente diferente das outras usou-se o teste de Friedman nas 4 escalas com diferente formato de medida e nos três cenários com correlações entre variáveis de diferente intensidade ($r=0.10$, $r=0.40$ e $r=0.70$).

Formato de Medida com três pontos

Para a escala com 3 pontos (*Likert 3*) no caso de distribuições de frequência simétricas (p_1, p_1, p_1) (cf. Figura 7. 1) e com correlações de baixa magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas ($\chi^2_f(5) = 25.180$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.002$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai e o traço de Pillai aplicado às ordens ($p < 0.001$) e o lambda de Wilks ($p=0.011$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.002$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.011$). Com correlações de média magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica também apresenta diferenças

estatisticamente significativas nas diferentes dimensões amostrais ($\chi^2_F(5) = 26.969$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai, o traço de Pillai aplicado às ordens ($p < 0.001$); entre o traço de Pillai e o lambda de Wilks ($p = 0.002$) e o lambda de Wilks aplicado às ordens ($p < 0.001$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p < 0.001$).

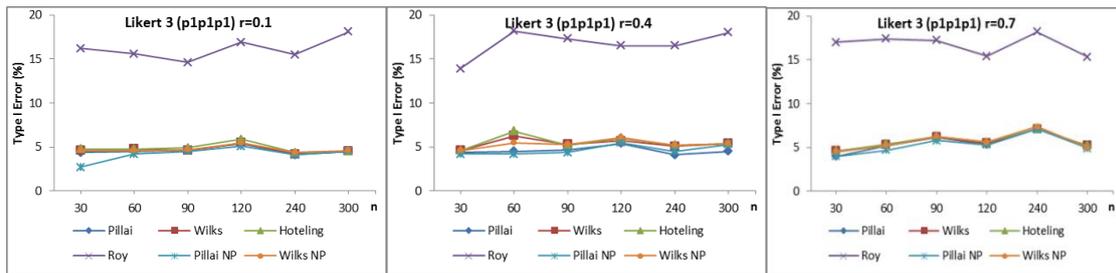


Figura 7. 1 - Probabilidade de erro de tipo I (em porcentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada no ponto central da escala (distribuições iguais (p_1, p_1, p_1), ou seja H_0 verdadeira).

Para as correlações de elevada magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica também apresenta diferenças estatisticamente significativas ($\chi^2_F(5) = 23.854$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling com $p=0.002$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.001$), o traço de Pillai aplicado às ordens ($p < 0.001$); entre o traço de Pillai e o lambda de Wilks ($p=0.022$) e o lambda de Wilks aplicado às ordens ($p=0.015$).

Com o mesmo formato de medida (Likert 3) e para distribuições de frequência assimétricas positivas (p_2, p_2, p_2) (cf. Figura 7.2) para as correlações de baixa magnitude, o teste de Friedman mostra diferenças estatisticamente significativas nas probabilidade de erro de tipo I dos testes da MANOVA ($\chi^2_F(5) = 25.180$; $p < 0.001$). Neste cenário, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.002$ e as restantes estatísticas com

$p < 0.001$); entre o traço de Hotelling e o traço de Pillai e o traço de Pillai aplicado às ordens ($p < 0.001$) e o lambda de Wilks ($p = 0.011$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p = 0.002$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.011$).

Para as correlações de média magnitude, a probabilidade de erro de tipo I também apresenta diferenças estatisticamente significativas ($\chi^2_F(5) = 26.969$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre o erro de tipo I da maior raiz de Roy e todas as outras estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai, o traço de Pillai aplicado às ordens ($p < 0.001$); entre o traço de Pillai e o lambda de Wilks ($p = 0.002$) e o lambda de Wilks aplicado às ordens ($p < 0.001$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Com correlações com elevada magnitude também há diferenças estatisticamente significativas nas diferentes estatísticas da MANOVA, ($\chi^2_F(5) = 23.854$; $p < 0.001$). As diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling com $p = 0.002$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.001$), o traço de Pillai aplicado às ordens ($p < 0.001$); entre o traço de Pillai e o lambda de Wilks ($p = 0.022$) e o lambda de Wilks aplicado às ordens ($p = 0.015$).

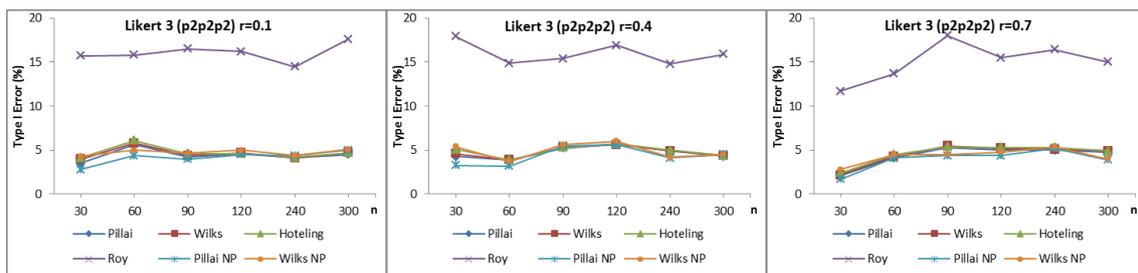


Figura 7.2 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1 = n_2 = n_3 = 10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r = 0.10$), média ($r = 0.40$) e elevada intensidade ($r = 0.70$) e com a maior probabilidade de resposta aos itens localizada no extremo inferior da escala (com distribuições iguais (p_2, p_2, p_2) , ou seja H_0 verdadeira).

Para distribuições de frequência assimétricas negativas (p_3, p_3, p_3) (cf. Figura 7.3), o teste de Friedman mostra que as estatísticas da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas para as correlações de baixa intensidade ($\chi_F^2(5) = 21.552; p < 0.001$), de média intensidade ($\chi_F^2(5) = 20.947; p < 0.001$), de elevada intensidade ($\chi_F^2(5) = 21.658; p < 0.001$). Para correlações de baixa intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.006$, lambda de Wilks aplicado às ordens com $p=0.002$ e as restantes com $p<0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.009$) e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o lambda de Wilks aplicados às ordens ($p=0.030$). Para as correlações com média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.028$ e as restantes com $p<0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.003$) e o traço de Pillai aplicado às ordens ($p<0.001$) e o lambda de Wilks aplicados às ordens ($p=0.037$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.016$); e entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.037$). Para as correlações de elevada intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.012$, entre o lambda de Wilks aplicado às ordens com $p=0.001$ e as restantes com $p<0.001$); entre o traço de Pillai e o lambda de Wilks ($p=0.029$) e o traço de Pillai aplicados às ordens e o lambda de Wilks ($p=0.016$) e o traço de Pillai e o lambda de Wilks aplicados às ordens ($p=0.004$).

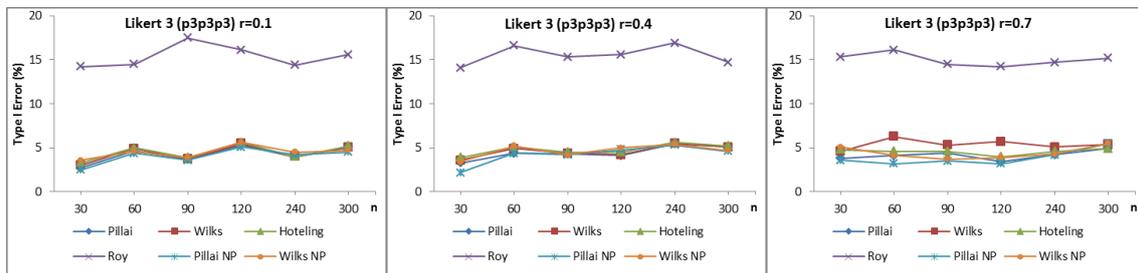


Figura 7.3 - Probabilidade de erro de tipo I (em percentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada no extremo superior da escala (com distribuições iguais (p_3, p_3, p_3), ou seja H_0 verdadeira).

Nas distribuições de frequência uniformes (p_4, p_4, p_4) (cf. Figura 7. 4) o teste de Friedman apresenta diferenças estatisticamente significativas para as correlações de baixa intensidade ($\chi_F^2(5)=27.325; p<0.001$; de média intensidade ($\chi_F^2(5)=24.084; p<0.001$; e de elevada intensidade ($\chi_F^2(5)=23.316; p<0.001$; para a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica.

Os testes *post-hoc* mostram que com correlações de baixa intensidade, as diferenças estatisticamente significativas na probabilidade de erro de tipo I, ocorreram entre a maior raiz de Roy e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling e o traço de Pillai ($p<0.001$), e o traço de Pillai aplicado às ordens ($p<0.001$) e o lambda de Wilks ($p=0.001$); entre o traço de Pillai e o traço de Pillai aplicados às ordens ($p<0.001$) e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens e, o lambda de Wilks e o lambda de Wilks aplicados às ordens ($p<0.001$).

Com correlações com média intensidade, as diferenças estatisticamente significativas na probabilidade de erro de tipo I ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$ e as restantes com $p<0.001$); entre o traço de Hotelling e, o traço de Pillai e o traço de Pillai aplicado às ordens ($p<0.001$); entre o lambda de Wilks ($p=0.006$); entre o traço de Pillai e o lambda de Wilks ($p=0.004$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.013$) e o lambda de Wilks aplicado às ordens ($p<0.001$).

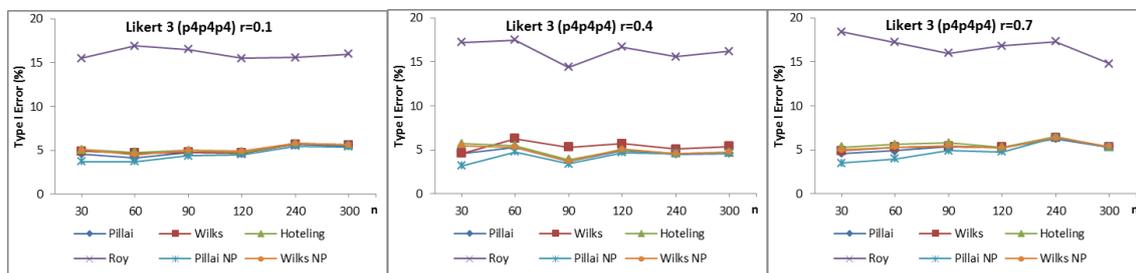


Figura 7. 4 - Probabilidade de erro de tipo I (em porcentagem) nos quatro testes da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de três pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens igualmente distribuída na escala (com distribuições iguais (p_4, p_4, p_4), ou seja H_0 verdadeira).

Para correlações de magnitude elevada, as diferenças estatisticamente significativas aconteceram na probabilidade de erro de tipo I da maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.006$ e as restantes com $p<0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.001$), e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o lambda de Wilks aplicados às ordens ($p=0.027$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.009$) e o lambda de Wilks aplicado às ordens ($p=0.001$).

Formato de Medida com quatro pontos

Para a escala com 4 pontos (tipo Likert 4) com distribuições de frequência simétricas e com correlações com baixa magnitude, o teste de Friedman mostra que existem diferenças estatisticamente significativas. No cenário com correlações de baixa intensidade ($\chi^2_F(5) = 20.498$; $p < 0.001$); as diferenças estatisticamente significativas ocorreram entre o erro de tipo I da maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$; traço de Pillai, lambda de Wilks e traço de Pillai aplicado às ordens ($p<0.001$; e com o lambda de Wilks aplicado às ordens ($p=0.003$); entre o traço de Hotelling, e o traço de Pillai ($p=0.030$, e o traço de Pillai aplicado às ordens ($p=0.001$), e o lambda de Wilks ($p=0.011$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.039$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.017$), e o lambda de Wilks aplicado às ordens ($p=0.002$).

Para as correlações de média magnitude, também há diferenças estatisticamente significativas nos diferentes testes da MANOVA ($\chi^2_F(5) = 25.443$; $p < 0.001$).

As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.001$) e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.001$), o traço de Pillai aplicado às ordens ($p<0.001$), o lambda de Wilks ($p=0.011$), o traço de Pillai aplicado às ordens ($p=0.004$), e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$); e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.016$).

Para as correlações de elevada magnitude, há diferenças estatisticamente significativas ($\chi^2_F(5) = 21.146$; $p < 0.001$). As diferenças estatisticamente significativas

ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling com $p=0.011$), o traço de Pillai aplicado às ordens e o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.003$), o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.026$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p<0.020$), e o lambda de Wilks aplicado às ordens ($p=0.004$).

Com o mesmo formato de medida (tipo Likert 4) e para distribuições de frequência assimétricas positivas (p_2, p_2, p_2), para as correlações com baixa magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas ($\chi_F^2(5) = 20.076$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.002$, lambda de Wilks aplicado às ordens ($p=0.010$), e as restantes estatísticas com $p<0.001$; entre o traço de Hotelling e o traço de Pillai ($p=0.008$), e o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.032$); entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.042$).

Para as correlações com média magnitude, a probabilidade de erro de tipo I também apresenta diferenças ($\chi_F^2(5) = 17.164$; $p = 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e o traço de Pillai, e o traço de Pillai aplicado às ordens ($p<0.001$); o traço de Hotelling e o lambda de Wilks ($p=0.001$), e o lambda de Wilks aplicado às ordens ($p=0.014$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.018$).

No cenário com correlações com elevada magnitude mantêm-se as diferenças estatisticamente significativas ($\chi_F^2(5) = 17.279$; $p = 0.001$). As diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling ($p=0.002$), o lambda de Wilks aplicado às ordens ($p=0.009$) e com as restantes estatísticas ($p<0.001$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p<0.047$).

Para distribuições de frequência assimétricas negativas (p_3, p_3, p_3), o teste de Friedman mostra diferenças estatisticamente significativas no cenário com correlações

de baixa intensidade ($\chi_F^2(5)=19.112$; $p < 0.001$); de média intensidade ($\chi_F^2(5)=19.577$; $p < 0.001$); e de elevada intensidade ($\chi_F^2(5)=21.468$; $p < 0.001$), para a probabilidade de erro de tipo I das estatísticas da MANOVA.

No cenário com correlações de baixa intensidade, os testes *post-hoc* apresentam diferenças estatisticamente significativas entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$, lambda de Wilks ($p=0.004$) e lambda de Wilks aplicado às ordens com $p=0.002$ e as restantes com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.006$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.031$), e o lambda de Wilks aplicado às ordens ($p=0.014$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling ($p=0.007$), lambda de Wilks aplicado às ordens ($p=0.005$) e as restantes estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.007$), e o traço de Pillai aplicado às ordens ($p=0.017$); entre o traço de Pillai e o lambda de Wilks aplicados às ordens ($p=0.010$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.022$).

Para as correlações de elevada intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling ($p=0.023$) e entre as restantes estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.002$), e o traço de Pillai aplicados às ordens ($p < 0.001$), e o lambda de Wilks aplicados às ordens ($p=0.030$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.012$) e o lambda de Wilks aplicado às ordens ($p=0.030$).

Para distribuições de frequência uniformes (p_4, p_4, p_4), o teste de Friedman mostra que a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas para as correlações de baixa intensidade ($\chi_F^2(5)=25.746$; $p < 0.001$), de média intensidade ($\chi_F^2(5)=26.133$; $p < 0.001$) e de elevada intensidade ($\chi_F^2(5)=27.273$; $p < 0.001$). Os testes *post-hoc* mostram que com correlações de baixa intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy, e o traço de Hotelling ($p=0.002$) e

todas as outras estatísticas ($p < 0.001$); entre o traço de Hotelling, e o traço de Pillai, e o traço de Pillai aplicado às ordens ($p < 0.001$), e o lambda de Wilks aplicado às ordens ($p = 0.002$); entre o traço de Pillai e o traço de Pillai aplicados às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.030$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Com as correlações de média intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p = 0.004$ e as restantes com $p < 0.001$); entre o traço de Hotelling, o traço de Pillai, e o traço de Pillai aplicado às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.004$), e o lambda de Wilks aplicado às ordens ($p = 0.024$); entre o traço de Pillai, o lambda de Wilks ($p = 0.006$), e o lambda de Wilks aplicado às ordens ($p = 0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Para as correlações de elevada intensidade, as diferenças estatisticamente significativas aconteceram entre a maior raiz de Roy e todas as outras estatísticas ($p < 0.001$); entre o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens e o lambda de Wilks ($p < 0.001$); entre o traço de Pillai, o lambda de Wilks ($p = 0.024$) e o lambda de Wilks aplicado às ordens ($p < 0.001$); e entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p = 0.001$), e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Formato de Medida com cinco pontos

Para a escala com 5 pontos (Likert 5) no caso de distribuições de frequência simétricas (p_1, p_1, p_1), a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas no cenário com correlações de baixa intensidade ($\chi^2_F(5) = 27.437$; $p < 0.001$), as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p = 0.001$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p < 0.001$) e o lambda de Wilks ($p = 0.012$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.001$); entre o traço de Pillai

aplicado às ordens, o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p < 0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.004$).

Para as correlações de média magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica também apresentam diferenças estatisticamente significativas ($\chi^2_F(5) = 22.935$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p = 0.005$), o lambda de Wilks aplicado às ordens ($p = 0.002$) e as outras estatísticas ($p < 0.001$); entre o traço de Hotelling, e o traço de Pillai, e o traço de Pillai aplicado às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.026$); entre o traço de Pillai, o lambda de Wilks ($p = 0.048$), e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.001$); e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.048$).

Para as correlações de elevada magnitude, a probabilidade de erro de tipo I também apresenta diferenças estatisticamente significativas ($\chi^2_F(5) = 18.706$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling com $p = 0.016$), o lambda de Wilks aplicado às ordens com $p = 0.002$, e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.007$), o traço de Pillai aplicado às ordens ($p = 0.021$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p = 0.045$).

Com o mesmo formato de medida (Likert 5) e para distribuições de frequência assimétricas positivas (p_2, p_2, p_2), com correlações com baixa magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas ($\chi^2_F(5) = 20.783$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre o erro de tipo I da maior raiz de Roy e todas as outras estatísticas (traço de Hotelling ($p = 0.008$), o lambda de Wilks aplicado às ordens ($p = 0.002$), e as restantes estatísticas ($p < 0.001$)); entre o traço de Hotelling e o traço de Pillai, e o traço de Pillai aplicado às ordens ($p = 0.002$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p < 0.011$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.011$).

Para as correlações de média magnitude, a probabilidade de erro de tipo I também apresenta diferenças ($\chi_F^2(5) = 21.232$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e o traço de Pillai, e o traço de Pillai aplicado às ordens ($p < 0.001$), o traço de Hotelling ($p = 0.011$), o lambda de Wilks ($p = 0.001$) e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.006$), e o traço de Pillai aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p = 0.003$), e o lambda de Wilks aplicado às ordens ($p = 0.008$).

Nas correlações de elevada magnitude também há diferenças estatisticamente significativas ($\chi_F^2(5) = 20.169$; $p < 0.001$). As diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling e com o lambda de Wilks aplicado às ordens ($p = 0.001$), e com as restantes estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.013$), e o traço de Pillai aplicado às ordens ($p = 0.024$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p = 0.024$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.042$).

Para distribuições de frequência assimétricas negativas (p_3, p_3, p_3) a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas quer no cenário com correlações de baixa intensidade ($\chi_F^2(5) = 20.672$; $p < 0.001$), de média intensidade ($\chi_F^2(5) = 20.275$; $p < 0.001$), e de elevada intensidade ($\chi_F^2(5) = 22.337$; $p < 0.001$).

Para as correlações de baixa intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p = 0.009$, lambda de Wilks com $p = 0.001$, e as restantes com $p < 0.001$); entre o traço de Hotelling, o traço de Pillai ($p = 0.005$), e o traço de Pillai aplicado às ordens ($p = 0.001$); entre o traço de Pillai e o lambda de Wilks ($p = 0.049$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p = 0.012$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicados às ordens ($p = 0.016$).

Com as correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p = 0.007$, lambda de Wilks aplicado às ordens com $p = 0.002$ e as restantes com $p < 0.001$); entre o traço de Hotelling, o traço de Pillai, e o traço de Pillai aplicado às

ordens ($p=0.004$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.013$).

Para as correlações com elevada intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.002$, lambda de Wilks aplicados às ordens $p=0.006$ e entre as restantes estatísticas com $p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.001$), e o traço de Pillai aplicados às ordens ($p=0.004$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.001$); e o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.022$).

Para distribuições de frequência uniformes (p_4, p_4, p_4) o teste de Friedman mostra que a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas com correlações de baixa intensidade ($\chi_F^2(5) = 22.143$; $p < 0.001$), de média intensidade ($\chi_F^2(5) = 22.399$; $p < 0.001$), e de elevada intensidade ($\chi_F^2(5) = 24.577$; $p < 0.001$).

Os teste *post-hoc* apontam que para correlações de baixa intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.005$ e as restantes com $p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.019$), e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.025$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.002$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicados às ordens ($p=0.001$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.003$ e as restantes com $p<0.001$); entre o traço de Hotelling, e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.004$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks, e o lambda de Wilks aplicados às ordens ($p<0.001$).

No cenário com correlações de elevada intensidade, as diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$ e entre as restantes estatísticas com $p<0.001$); entre o traço de Hotelling, o traço de

Pillai, e o traço de Pillai aplicados às ordens ($p < 0.001$); entre e o traço de Pillai aplicado às ordens ($p = 0.018$) e o lambda de Wilks ($p = 0.026$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Formato de Medida com sete pontos

A escala com 7 pontos (Likert 7) com distribuições simétricas e com correlações de baixa magnitude, a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas. O teste de Friedman mostra essas diferenças, no cenário com correlações de baixa intensidade ($\chi^2_F(5) = 19.706$; $p < 0.001$), entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p = 0.017$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling, o traço de Pillai $p = 0.049$, e o traço de Pillai aplicado às ordens ($p = 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p = 0.022$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.038$).

Com correlações com média magnitude, a probabilidade de erro de tipo I também apresenta diferenças estatisticamente significativas ($\chi^2_F(5) = 19.534$; $p < 0.001$). As diferenças ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p = 0.005$), o lambda de Wilks aplicado às ordens ($p = 0.010$), e as outras estatísticas ($p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.023$), e o traço de Pillai aplicado às ordens ($p = 0.013$); entre o traço de Pillai e o lambda de Wilks ($p = 0.010$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.006$). Para as correlações de elevada magnitude verificam-se também diferenças ($\chi^2_F(5) = 20.419$; $p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (com o traço de Hotelling com $p = 0.006$), o lambda de Wilks aplicado às ordens com $p = 0.005$, e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai ($p = 0.007$), o traço de Pillai aplicado às ordens ($p = 0.004$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p = 0.007$); e entre o traço de Pillai e o lambda de Wilks aplicados às ordens ($p = 0.004$).

Com o mesmo formato de medida (Likert 7) e para distribuições de frequência assimétricas positivas (p_2, p_2, p_2) com correlações com baixa magnitude, a probabilidade

de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas ($\chi_F^2(5) = 20.813; p < 0.001$). As diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.002$, lambda de Wilks aplicado às ordens com $p=0.006$ e as restantes estatísticas com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.037$), e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Para as correlações de média magnitude, a probabilidade de erro de tipo I também apresenta diferenças nas diferentes estatísticas da MANOVA ($\chi_F^2(5) = 18.653; p < 0.001$). As diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy, o traço de Pillai, o traço de Pillai aplicado às ordens ($p < 0.001$), e o traço de Hotelling ($p=0.003$), entre o lambda de Wilks ($p=0.002$) e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens, o traço de Hotelling ($p=0.011$), e o lambda de Wilks ($p=0.019$).

Para as correlações de elevada magnitude também se registam diferenças estatisticamente significativas ($\chi_F^2(5) = 18.719; p < 0.001$). As diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (o lambda de Wilks aplicado às ordens com $p=0.026$ e com as restantes estatísticas $p < 0.001$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.004$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.020$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.044$).

O teste de Friedman mostra que para distribuições assimétricas negativas (p_3, p_3, p_3) a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas no cenário com correlações de baixa intensidade ($\chi_F^2(5) = 20.025; p < 0.001$), de média intensidade ($\chi_F^2(5) = 19.600; p < 0.001$), e de elevada intensidade ($\chi_F^2(5) = 21.513; p < 0.001$).

Para as correlações de baixa intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$, lambda de Wilks com $p=0.002$, e as restantes com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.001$); entre o traço de Pillai

aplicado às ordens e o lambda de Wilks ($p=0.003$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicados às ordens ($p=0.019$).

Nas correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.004$, lambda de Wilks aplicados às ordens com $p=0.002$ e lambda de Wilks com $p=0.001$, e as restantes estatísticas com $p<0.001$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.003$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.012$), e o lambda de Wilks aplicados às ordens ($p=0.007$). Para as correlações com elevada intensidade, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.001$, lambda de Wilks aplicados às ordens $p=0.008$ e entre as restantes estatísticas com $p<0.001$); entre o traço de Hotelling e o traço de Pillai aplicados às ordens ($p=0.002$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.015$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.028$).

O teste de Friedman mostra que, para distribuições de frequência uniformes (p_4, p_4, p_4) a probabilidade de erro de tipo I das estatísticas da MANOVA paramétrica e não paramétrica apresenta diferenças estatisticamente significativas com correlações de baixa intensidade ($\chi_F^2(5) = 24.409$, $p < 0.001$), de média intensidade ($\chi_F^2(5) = 23.425$; $p < 0.001$) e de elevada intensidade ($\chi_F^2(5) = 25.767$; $p < 0.001$).

Para as correlações de baixa intensidade, as diferenças ocorreram entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling e lambda de Wilks com $p=0.001$, traço de Pillai, traço de Pillai aplicado às ordens e lambda de Wilks aplicado às ordens com $p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.004$) e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.014$); entre o traço de Pillai e o lambda de Wilks ($p=0.004$), e o lambda de Wilks aplicado às ordens ($p=0.041$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks, e o lambda de Wilks aplicados às ordens ($p<0.001$).

Com correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre o erro de tipo I da maior raiz de Roy e todas as outras

estatísticas (traço de Hotelling com $p=0.015$ e as restantes com $p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.001$), o traço de Pillai aplicado às ordens ($p<0.001$) e o lambda de Wilks aplicados às ordens ($p<0.007$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.001$), e o lambda de Wilks aplicados às ordens ($p=0.007$). Com correlações com elevada intensidade, as diferenças estatisticamente significativas ocorreram entre o erro de tipo I da maior raiz de Roy e todas as outras estatísticas (traço de Hotelling com $p=0.001$ e entre as restantes estatísticas com $p<0.001$); entre o traço de Hotelling, o traço de Pillai, e o traço de Pillai aplicados às ordens ($p<0.001$), e o lambda de Wilks aplicado às ordens ($p=0.020$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.008$), o lambda de Wilks ($p=0.001$), e o lambda de Wilks aplicado às ordens ($p=0.046$); e entre o traço de Pillai aplicado às ordens, o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

7.3.2 Potência dos Testes

Os resultados das simulações mostram diferenças nos valores da potência dos testes da MANOVA paramétrica e da MANOVA não paramétrica. Tendo por base a distribuição de frequência, apresentamos 3 situações distintas, associadas ao formato de medida (número de pontos da escala), à dimensão das amostras e à intensidade de correlação entre os itens.

Na primeira situação, com as probabilidades marginais (p_4, p_4, p_1) em que a distribuição é uniforme em 2 dos grupos e simétrica no terceiro grupo e com probabilidades marginais (p_1, p_1, p_4) em que a distribuição é simétrica em 2 dos grupos e no terceiro é uniforme (cf. Tabela 7.4), seja qual for o formato de medida da escala, a magnitude das correlações ou a dimensão das amostras, observam-se resultados de potência de teste muito baixos, em qualquer das estatísticas da MANOVA paramétrica ou da MANOVA não paramétrica, nos diferentes cenários simulados.

Tabela 7.4 - Condições testadas nas diferentes simulações (216 cenários) com dois tipos de probabilidades marginais (p), seis dimensões amostrais (n), correlações com três tipos de magnitude (r), e escalas de avaliação com quatro formatos de medida (3, 4, 5, e 7 pontos).

Situações simuladas para a potência de teste						
Probabilidades marginais (p)	$p_{12}p_{13}p_{14}$		$p_{43}p_{44}p_{41}$			
Distribuições	Simétrica/uniforme		Simétrica/uniforme			
Dimensão das amostras (n)	$n=30, 60, 90, 120, 240, 300$					
Formato de Medida	Magnitude das correlações (r)					
3 pontos	0.10	0.40	0.70	0.10	0.40	0.70
4 pontos	0.10	0.40	0.70	0.10	0.40	0.70
5 pontos	0.10	0.40	0.70	0.10	0.40	0.70
7 pontos	0.10	0.40	0.70	0.10	0.40	0.70

Os resultados mostram que a maior raiz de Roy apresenta valores mais elevados de potência do que as outras estatísticas da MANOVA paramétrica, mantendo valores de potência baixos (entre 14% e 18%). As restantes estatísticas retratam potências de teste ainda mais baixas (entre 3% e 7%) em todos os formatos de medida (3, 4, 5 e 7 pontos), em todas as dimensões das amostras, com qualquer intensidade das correlações (baixas, médias e altas) quer na MANOVA, paramétrica quer na não paramétrica (cf. Figuras 7.5, 7.6, 7.7, 7.8).

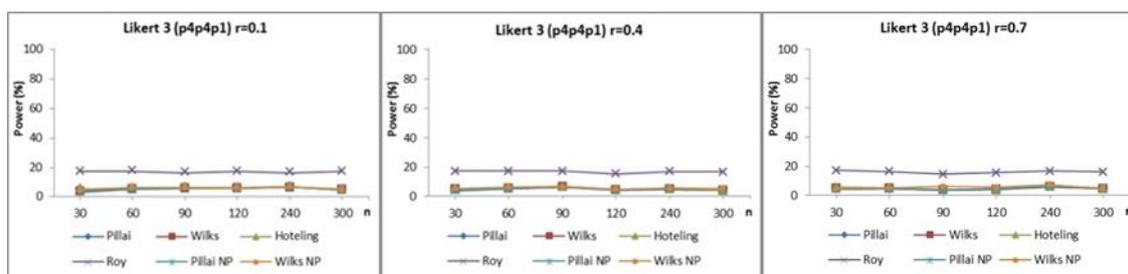


Figura 7.5 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada uniformemente e simétrica na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_{43}, p_{44}, p_{41}) ou seja H_0 falsa).

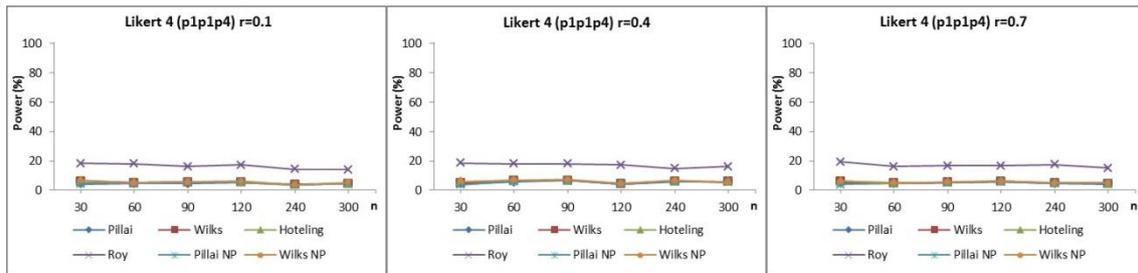


Figura 7.6 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100, 30, 60, 90, 120, 240,$ e 300 com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa magnitude ($r=0.10$), média ($r=0.40$) e elevada ($r=0.70$) e com a maior probabilidade de resposta aos itens com localização simétrica e uniforme na escala (distribuições iguais em dois grupos e diferentes do terceiro grupo (p_1, p_1, p_4), ou seja H_0 falsa).

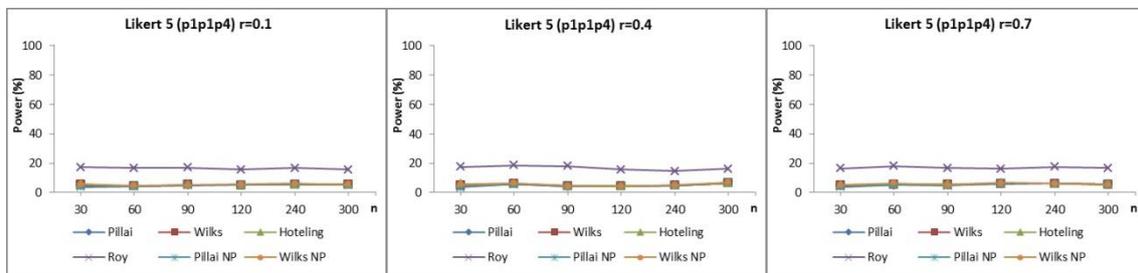


Figura 7.7 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100,$ com um formato de medida com 5 pontos (Likert 5 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens com localização simétrica e uniforme na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_1, p_1, p_4), ou seja H_0 falsa).

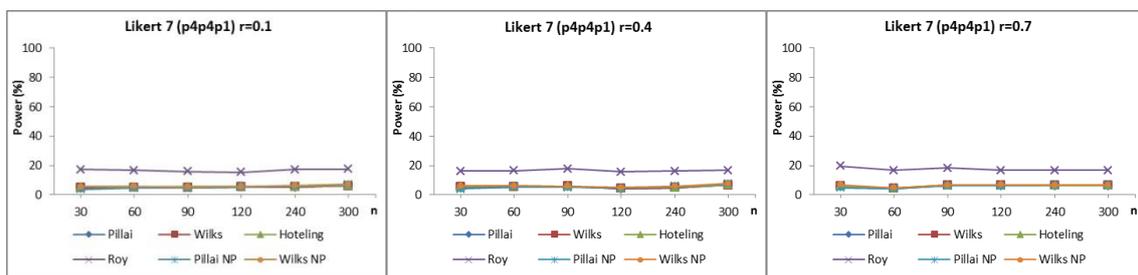


Figura 7.8 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100,$ com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com a maior probabilidade de resposta aos itens localizada uniformemente e simétrica na escala (distribuições iguais em dois grupos e diferentes no terceiro grupo (p_4, p_4, p_1) ou seja H_0 falsa).

Na **segunda situação**, os resultados referentes à potência de teste provêm dos diferentes cenários simulados para as 4 estatísticas da MANOVA paramétrica e para as 2 da MANOVA não paramétrica, com 3 tipos de probabilidades marginais: distribuição assimétrica positiva, assimétrica negativa, e distribuição uniforme (p_2, p_3, p_4) nos 3 grupos respectivamente; distribuição simétrica, distribuição assimétrica positiva e distribuição assimétrica negativa (p_1, p_2, p_3) nos três grupos respectivamente; e distribuição assimétrica positiva em 2 dos grupos e assimétrica negativa no terceiro grupo (p_2, p_2, p_3) (cf. Tabela 7. 5).

Tabela 7. 5 – Condições testadas nas diferentes simulações (324 cenários) com três tipos de probabilidades marginais (p), seis dimensões amostrais (n), correlações com três tipos de magnitude (r), e quatro escalas de avaliação com um formato de medida com 3, 4, 5, e 7 pontos.

Situações simuladas para a potência de teste									
Probabilidades marginais Distribuição	p_2, p_2, p_3 Assimétrica			p_2, p_3, p_4 Assimétrica			p_1, p_2, p_3 Simétrica, assimétrica		
	Positiva e negativa			positiva, negativa e uniforme			Positiva e negativa		
Dimensão da Amostra (n)	n=30, 60, 90, 120, 240, 300								
Formato de Medida	Magnitude das Correlações (r)								
3 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
4 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
5 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70
7 pontos	0.10	0.40	0.70	0.10	0.40	0.70	0.10	0.40	0.70

Nesta segunda situação, a maior raiz de Roy apresenta valores mais elevados de potência, do que qualquer um dos outros testes. Porém todas as outras estatísticas apresentam resultados de potência de teste elevados (superiores a 90%) nas escalas com formato de medida com 3, 4 e 5 pontos e nas amostras superiores ou iguais a 60 observações. Nas escalas com 7 pontos as potências são de baixa magnitude nas situações simuladas com amostras com 30 e 60 observações. Nas amostras iguais a 90 ou mais observações as potências são elevadas em todas os cenários simulados. As figuras 7.9 a 7.15 apresentam exemplos associados aos diferentes cenários usados nas simulações.

Probabilidades Marginais (p_1, p_2, p_3)

Com distribuições de frequência diferentes em cada um dos 3 grupos (simétricas e assimétrica positiva e assimétrica negativa), as potências apresentam-se superiores a

75% em todos os testes, nas escalas com formato de medida com 3 e 4 pontos; 60% nas escalas com 5 pontos, baixando para aproximadamente 35% ($r=0.40$) e 27% ($r=0.70$) nas escalas com 7 pontos e nas amostras com 30 observações, mantendo nas restantes dimensões amostrais, potências superiores a 55%. Com este tipo de distribuição e dando como exemplo as variáveis ordinais com um formato de medida com 3 e 4 pontos, tanto os testes paramétricos como os não paramétricos apresentam valores de potência elevados (superiores ou aproximados a 90%), com exceção para as amostras com $n=30$ e correlações de elevada magnitude ($r=0.70$) em que as potências oscilam entre 73% e 92.2 % e os valores da potência diminuem cerca de dois pontos percentuais da escala com um formato de medida de 3 pontos para a de 4 pontos. O mesmo caso com variáveis ordinais com um formato de medida com 5 e 7 pontos, a maior raiz de Roy apresenta potências mais elevadas para $n=30$ e com elevada intensidade na correlação entre as variáveis ($r=0.70$). Nas variáveis com um formato de medida com 5 pontos e uma correlação de baixa intensidade ($r=0.10$) e amostras de pequena dimensão ($n=30$), o menor valor da potência de teste é de 85.5% para o traço de Pillai aplicado às ordens, seguido do lambda de Wilks aplicado às ordens e do traço de Pillai com 87.5%. O traço de Hotelling apresenta uma potência de 91.5% e a maior raiz de Roy 97.5%; com correlações de média intensidade ($r=0.40$). Com correlações de elevada intensidade ($r=0.70$) acentua-se a diminuição da potência dos testes. Relativamente às variáveis com escala de medida com sete pontos, os valores da potência diminuem consideravelmente à medida que a intensidade das correlações aumenta, nas amostras de menor dimensão ($n=30, 60$), embora se mantenham aceitáveis (superiores a 55%).

Nas escalas com formato de medida com 3 e 4 pontos, à medida que a magnitude das correlações diminui as potências aproximam-se de 100% e na escala com formato de medida com cinco pontos, as potências atingem os 80%. Na escala com formato de medida de sete pontos, verificou-se uma diminuição da potência dos testes nas dimensões amostrais com menos de 240 observações, o que se acentua com o aumento da intensidade das correlações (de $r=0.40$ para $r=0.70$) na amostra com trinta observações (cf. Figura 7.9, Figura 7.10).

Para este tipo de distribuições nas escalas com formato de medida com 3 e 4 pontos, o teste de Friedman mostra que não existem diferenças estatisticamente significativas entre as potências dos testes da MANOVA paramétrica e não paramétrica. Para escalas com formato de medida com 5 pontos, também não existem diferenças estatisticamente significativas entre as potências, nos cenários com correlações de baixa e média intensidade, apenas com correlações de elevada intensidade se verificam diferenças ($\chi^2_F(5) = 14.600$; $p < 0.001$).

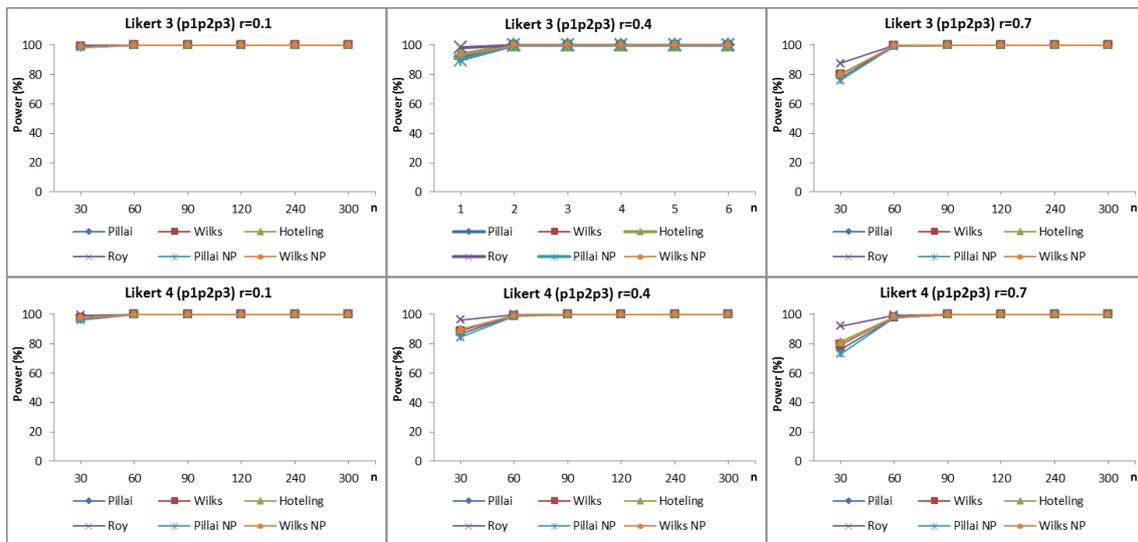


Figura 7.9 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 e 4 pontos (Likert 3 e tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada no ponto central e nos extremos da escala de medida (com distribuições diferentes em cada um dos três grupos (p_1, p_2, p_3), ou seja H_0 falsa).

As diferenças estatisticamente significativas nas potências confirmam-se entre a maior raiz de Roy e o traço de Pillai ($p=0.004$), e o traço de Pillai aplicado às ordens ($p < 0.001$) e o lambda de Wilks ($p=0.017$); e entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.009$) e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.009$). Para escalas com formato de medida com sete pontos (cf. Figura 7.10) o teste de Friedman mostra que os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas entre as potências, nos três cenários correlacionais; magnitude baixa ($\chi^2_F(5) = 12.054$;

$p=0.015$), média ($\chi_F^2(5)=19.000$; $p<0.001$), e elevada ($\chi_F^2(5)=18.714$; $p<0.001$). Para as correlações de baixa magnitude, as diferenças estatisticamente significativas verificaram-se entre a potência da maior raiz de Roy e o traço de Pillai ($p=0.001$), e o traço de Pillai aplicado às ordens ($p=0.004$); e entre o traço de Hotelling e o traço de Pillai ($p=0.029$).

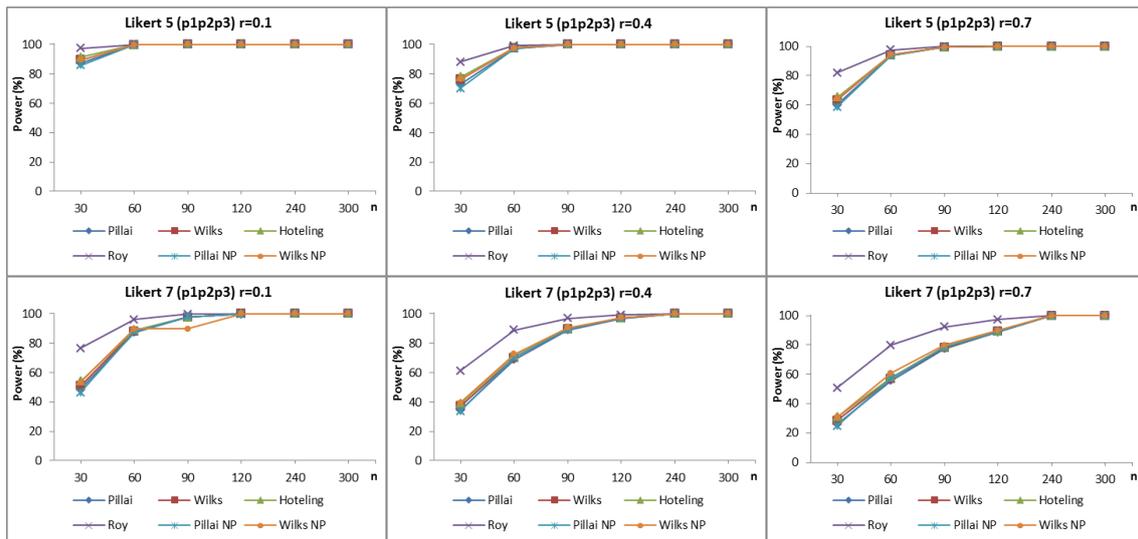


Figura 7.10 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 e 7 pontos (Likert 5 e 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada no ponto central e nos extremos da escala de medida (com distribuições diferentes em cada um dos três grupos (p_1, p_2, p_3), ou seja H_0 falsa).

Para as correlações de média magnitude, as diferenças estatisticamente significativas verificaram-se entre a potência da maior raiz de Roy e o traço de Hotelling ($p=0.031$); o traço de Pillai e o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.004$), e o traço de Pillai aplicado às ordens ($p=0.031$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.002$). Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o traço de Hotelling ($p=0.033$); o traço de Pillai e o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.005$); entre o traço de Pillai

e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.005$) e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.009$).

Probabilidades Marginais (p_2, p_2, p_3)

Com distribuições assimétricas positivas nos 2 grupos e assimétrica negativa no terceiro grupo (p_2, p_2, p_3), as potências apresentam-se superiores a 90% (para $r = 0.10$), 80% (para $r = 0.40$) e 70% (para $r = 0.70$), em todos os testes, nas escalas com formato de medida com 3, 4 e 5 pontos, seja qual for a dimensão da amostra o formato de medida e a intensidade das correlações. Nas escalas com formato de medida com 7 pontos as potências são mais baixas nas amostras de menor dimensão (30 e 60 observações com potências de 30% e 60% respectivamente).

Para este tipo de distribuições de frequência e com escalas com formato de medida com 3 e 4 pontos (cf. Figura 7.11), os testes da MANOVA paramétrica e não paramétrica não apresentam diferenças estatisticamente significativas entre as potências, seja no cenário com correlações de baixa, de média ou de elevada intensidade. Nas escalas com formato de medida com 5 pontos (cf. figura 7.11), o teste de Friedman mostra que as potências dos testes da MANOVA paramétrica e não paramétrica não apresentam diferenças estatisticamente significativas nos cenários com correlações de baixa ou média intensidade.

Para as correlações de elevada intensidade, as diferenças são estatisticamente significativas ($\chi^2_F(5) = 14.826$; $p = 0.001$). Estas diferenças sucederam entre a maior raiz de Roy e todas as outras estatísticas (traço de Hotelling, com $p = 0.042$, Lambda de Wilks, com $p = 0.042$, lambda de Wilks aplicado às ordens, com $p = 0.009$, traço de Pillai e traço de Pillai aplicado às ordens, com $p < 0.001$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p = 0.042$).

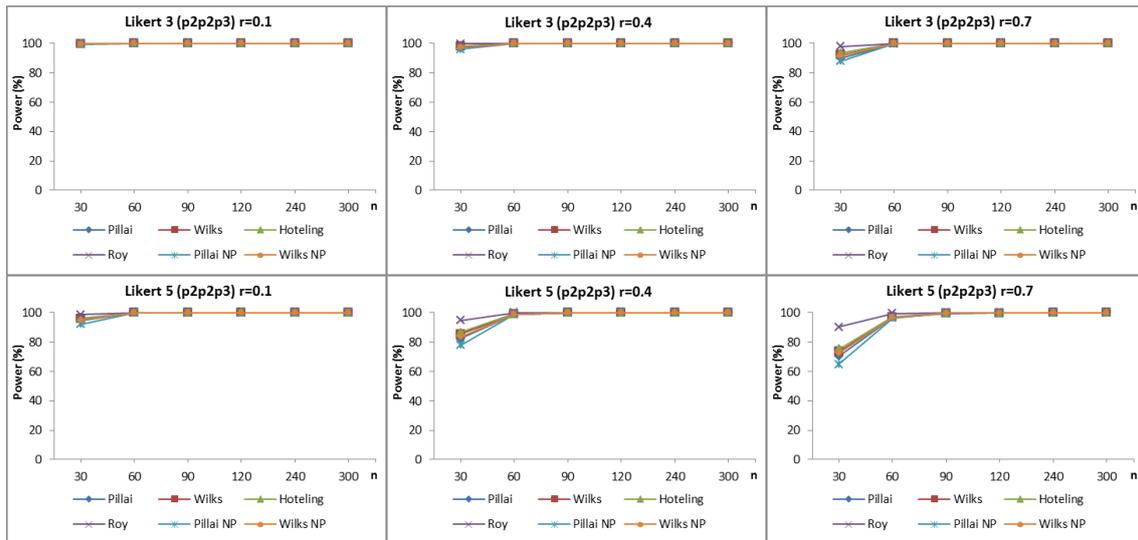


Figura 7.11 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida de 3 e 5 pontos (Likert 3 e 5 pontos), com correlações de baixa ($r=0.10$ média), ($r=0.40$) e elevada intensidade ($r=0.70$) e com maior probabilidade de resposta aos itens localizada nos extremos da escala de medida (com distribuições iguais para dois grupos e diferentes para o terceiro grupo (p_2, p_2, p_3) , ou seja H_0 falsa).

Nas escalas com formato de medida com 7 pontos (cf. Figura 7.12), o teste de Friedman mostra diferenças estatisticamente significativas nas potências dos testes da MANOVA paramétrica e não paramétrica para as correlações de baixa ($\chi^2_F(5) = 15.763$; $p < 0.001$), de média ($\chi^2_F(5) = 17.571$; $p < 0.001$), e de elevada intensidade, ($\chi^2_F(5) = 17.086$; $p < 0.001$). Com correlações de baixa magnitude, as diferenças ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.016$), o traço de Pillai ($p < 0.001$), o lambda de Wilks ($p=0.001$), o lambda de Wilks aplicado às ordens ($p=0.008$) e o traço de Pillai aplicado às ordens ($p < 0.001$). Com correlações de média magnitude, as diferenças nas potências ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.013$), o traço de Pillai ($p < 0.001$), o lambda de Wilks ($p < 0.001$), e o traço de Pillai aplicado às ordens ($p=0.004$); entre o traço de Hotelling e o traço de Pillai ($p=0.024$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p < 0.001$) e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.042$).

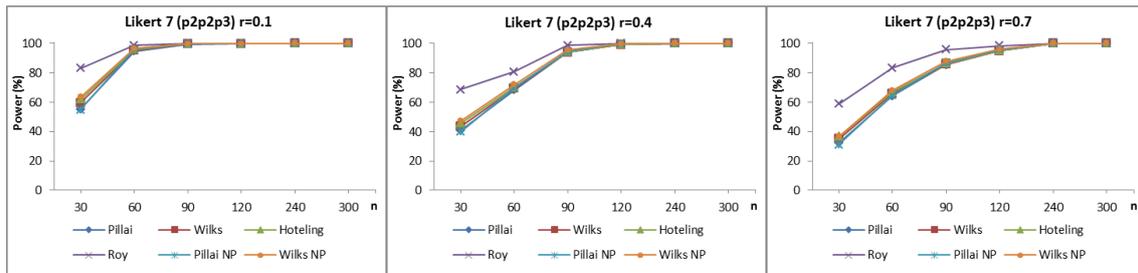


Figura 7.12 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos da escala de medida (com distribuições iguais para dois grupos e diferentes para o terceiro grupo (p_2, p_2, p_3), ou seja H_0 falsa).

Com correlações de elevada magnitude, as diferenças nas potências ocorreram entre a potência da maior raiz de Roy e o traço de Hotelling ($p=0.014$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.002$) e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.034$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.010$) e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.026$).

Probabilidades Marginais (p_2, p_3, p_4)

Para distribuições de frequência diferentes em cada um dos 3 grupos (p_2, p_3, p_4), assimétrica positiva no primeiro grupo, assimétrica negativa no segundo e uniforme no terceiro, nas escalas de medida com 3 pontos a magnitude das potências são superiores a 75%, (cf. Figura 7.13) e na escala com 4 pontos são superiores a 70%, em todos os testes, seja qual for a dimensão da amostra e a intensidade das correlações.

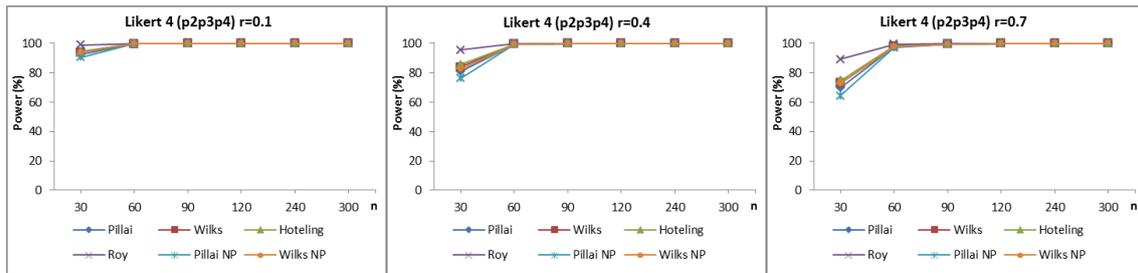


Figura 7.13 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa).

Com um formato de medida com 5 pontos (cf. Figura 7.14), na amostra com $n=30$ observações, verificam-se potências de teste aproximadas a 50%, para as correlações de elevada magnitude, e 65% nas de magnitude média, tendo os restantes cenários potências superiores a 80%. Para as escalas com 7 pontos os valores das potências diminuem nas amostras com 30 observações (aproximadamente 50% para $r=0.10$, 30% para $r=0.40$, e 25% para $r=0.70$).

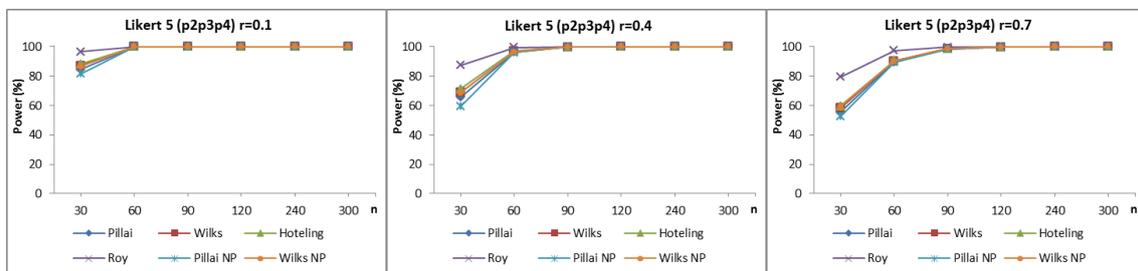


Figura 7.14 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 pontos (Likert 5 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa).

Para este tipo de distribuições de frequência e com escalas com formato de medida com 3 pontos, os testes da MANOVA paramétrica e não paramétrica não apresentam diferenças estatisticamente significativas nas potências, seja no cenário com correlações de baixa, de média ou de elevada intensidade. Na escala com 4 pontos,

apenas no cenário com correlações de elevada magnitude existem diferenças significativas nas potências dos testes ($\chi_F^2(5)=12.647$; $p < 0.001$). Estas diferenças verificaram-se entre a maior raiz de Roy, e o traço de Pillai ($p=0.004$), o traço de Pillai aplicado às ordens ($p=0.001$); e entre o traço de Pillai aplicado às ordens e o traço de Hotelling ($p=0.028$).

O teste de Friedman mostra que com escalas com formato de medida com 5 pontos, não existem diferenças estatisticamente significativas nas potências dos testes da MANOVA paramétrica e não paramétrica no cenário com correlações de baixa intensidade. As diferenças surgem nos cenários com correlações de média ($\chi_F^2(5)=12.647$; $p < 0.001$), e elevada intensidade ($\chi_F^2(5)=16.783$; $p < 0.001$).

Os testes *post-hoc* mostram que no cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências verificaram-se entre a maior raiz de Roy e o traço de Pillai ($p=0.004$), o traço de Pillai aplicado às ordens ($p=0.001$) e o lambda de Wilks aplicado às ordens ($p=0.015$); e entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.028$). No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o traço de Pillai ($p < 0.001$), o traço de Pillai aplicado às ordens ($p < 0.001$), o lambda de Wilks ($p=0.006$), e o lambda de Wilks aplicado às ordens ($p=0.040$); entre o traço de Hotelling e o traço de Pillai ($p=0.030$), e o traço de Pillai aplicado às ordens ($p=0.005$); e entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p=0.009$).

Para escalas com formato de medida com 7 pontos (Figura 7.15), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos 3 cenários correlacionais, baixa, média e elevada, respetivamente ($\chi_F^2(5)=16.333$; $p < 0.001$), ($\chi_F^2(5)=17.014$; $p < 0.001$), ($\chi_F^2(5)=17.286$; $p < 0.001$). Para as correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências verificaram-se entre a maior raiz de Roy e o traço de Hotelling ($p=0.037$), o traço de Pillai ($p < 0.001$), o traço de Pillai aplicado às ordens ($p < 0.001$) e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.020$); e entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.026$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências confirmaram-se entre a maior raiz de Roy e o traço de

Hotelling ($p=0.034$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.002$) e o lambda de Wilks ($p=0.001$); e entre o traço de Hotelling e o traço de Pillai ($p=0.011$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o lambda de Wilks aplicado às ordens, o traço de Pillai aplicado às ordens ($p=0.034$), e o lambda de Wilks ($p=0.026$).

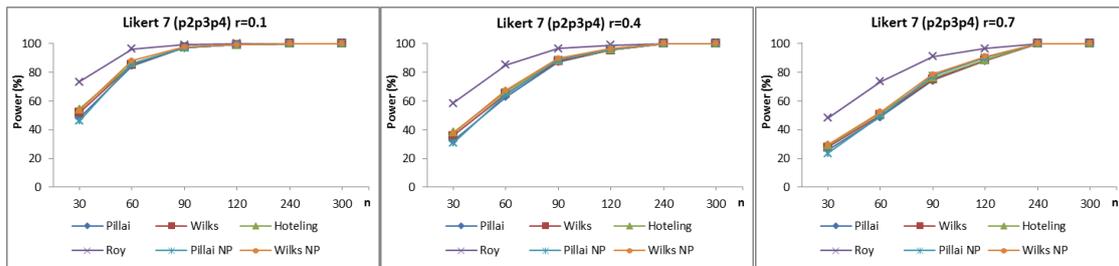


Figura 7.15 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada nos extremos e uniformemente distribuída na escala de medida (com distribuições diferentes em cada um dos três grupos (p_2, p_3, p_4), ou seja H_0 falsa).

No cenário com correlações de elevada magnitude as diferenças nas potências são entre a maior raiz de Roy e o traço de Hotelling ($p=0.014$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.002$), o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.025$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.001$); entre o lambda de Wilks aplicado às ordens e o traço de Pillai aplicado às ordens ($p=0.025$), e o lambda de Wilks ($p=0.014$).

Na terceira situação os resultados referentes à potência de teste procedem das diferentes situações simuladas para as 4 estatísticas da MANOVA paramétrica e 2 da MANOVA não paramétrica, com 5 tipos de probabilidades: com distribuições simétricas em 2 dos grupos e assimétrica positiva no terceiro (p_1, p_1, p_2); com distribuições simétrica, assimétrica positiva e uniforme (p_1, p_2, p_4) nos 3 grupos respetivamente; com distribuição de frequência assimétrica positiva em dois dos grupos e simétrica no terceiro (p_2, p_2, p_1), com distribuição de frequência uniforme em dois dos grupos e assimétrica positiva no terceiro (p_4, p_4, p_2); e com distribuição de frequência assimétrica positiva em dois dos grupos e uniforme no terceiro (p_2, p_2, p_4).

Nestes casos a maior raiz de Roy apresenta valores mais elevados de potência do que qualquer outro dos testes da MANOVA, seguido do lambda de Wilks aplicado às ordens. Os resultados da potência dos testes permitem ainda dizer que, relativamente as situações anteriores, estes parecem ainda mais influenciados pela dimensão da amostra, pela intensidade da correlação entre itens e pelo número de pontos da escala de medida. As estimativas das potências diminuem, de forma acentuada, à medida que aumenta o número de pontos do formato de medida (de 3 para 7 pontos), com a magnitude das correlações (da baixa intensidade para a média intensidade, e desta para a elevada intensidade).

Os resultados mostram que quanto menor a dimensão das amostras ($n=30$ e $n=60$) e maior a intensidade da correlação entre variáveis ($r=0.70$), menor a potência dos testes, sendo que, uma das estatísticas da MANOVA não paramétrica, lambda de Wilks aplicado às ordens apresenta valores de potência mais elevados do que a maioria dos testes paramétricos e mais elevados do que o traço de Pillai aplicado às ordens. O teste de Pillai aplicado às ordens também apresenta valores de potência elevados, aproximando-se dos resultados os testes paramétricos ou superando-os em algumas casos, o que se verifica com distribuições de frequência (p_1, p_2, p_4). A diminuição da potência dos testes acentua-se nas amostras de menor dimensão, à medida que aumenta a intensidade de correlação. Para verificar se existem estas diferenças nas distribuições dos valores das potências usou-se o teste de Friedman.

Probabilidades Marginais (p_1, p_1, p_2)

Nas distribuições simétricas nos dois primeiros grupos e assimétrica positiva no terceiro grupo (p_1, p_1, p_2), para escalas com formato de medida com **3 pontos** (cf. Figura 7.16), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: correlações de baixa intensidade, ($\chi_F^2(5) = 12.320$; $p < 0.005$); com correlações de média intensidade, ($\chi_F^2(5) = 17.815$; $p < 0.001$); e com correlações de elevada intensidade, ($\chi_F^2(5) = 18.453$; $p < 0.001$).

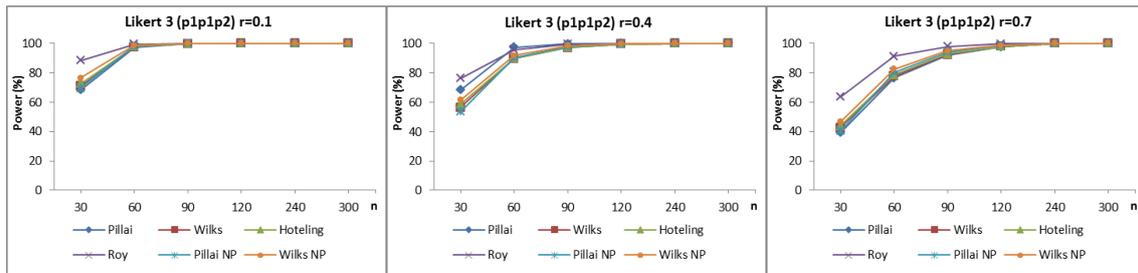


Figura 7.16 - Potência de teste (em percentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens-Pillai NP, lambda de Wilks aplicado às ordens-Wilks NP e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 3 pontos (Likert 3 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no ponto central e no extremo inferior na escala de medida (com distribuições iguais em dois dos grupos e diferentes no terceiro grupo, (p_1, p_1, p_2) , ou seja H_0 falsa).

Os testes *post-hoc* mostram que no cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências verificaram-se entre a maior raiz de Roy e o traço de Hotelling ($p=0.041$), o traço de Pillai ($p=0.002$), o lambda de Wilks ($p=0.012$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.006$); e entre o lambda de Wilks e o o lambda de Wilks aplicado às ordens ($p=0.041$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.003$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.015$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.049$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.015$), o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.008$).

No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o traço de Hotelling ($p=0.005$), o traço de Pillai e o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.014$); entre o traço de Hotelling e o traço de Pillai ($p=0.019$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.007$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.007$).

Com o mesmo tipo de distribuição, para escalas com formato de medida com **4 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças

estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi_F^2(5)=12.292$; $p=0.006$), com correlações de média intensidade ($\chi_F^2(5)=18.037$; $p<0.001$), e com correlações de elevada intensidade ($\chi_F^2(5)=18.261$; $p<0.001$).

Para as correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy e o traço de Pillai ($p=0.002$), e o lambda de Wilks ($p=0.012$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.006$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.040$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.002$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.010$); entre o traço de Hotelling e o traço de Pillai aplicado às ordens ($p=0.026$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.019$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.007$).

Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o traço de Hotelling ($p=0.005$); o traço de Pillai e o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.014$); entre o traço de Hotelling e o traço de Pillai ($p=0.026$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.010$) e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.005$).

Para escalas com formato de medida com **5 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi_F^2(5)=17.086$; $p<0.001$), com correlações de média intensidade ($\chi_F^2(5)=17.302$; $p<0.001$), e com correlações de elevada intensidade ($\chi_F^2(5)=24.840$; $p<0.001$).

No cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências se atestam entre a maior raiz de Roy e o traço de Pillai ($p=0.008$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.010$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling e o traço de Pillai ($p=0.045$);

entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.034$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.010$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy e o traço de Hotelling ($p=0.005$), o traço de Pillai ($p<0.001$), o lambda de Wilks ($p=0.001$), e o traço de Pillai aplicado às ordens ($p=0.007$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.043$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.010$).

Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o lambda de Wilks aplicado às ordens ($p=0.008$), e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.001$), o lambda de Wilks ($p=0.041$), e o lambda de Wilks aplicado às ordens ($p=0.005$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.001$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.041$), e o lambda de Wilks aplicado às ordens ($p=0.005$).

Nas escalas com formato de medida com **7 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5) = 22.661$; $p < 0.006$), com correlações de média intensidade ($\chi^2_F(5) = 22.884$; $p < 0.001$), e com correlações de elevada intensidade ($\chi^2_F(5) = 25.957$; $p < 0.001$).

Os post-hoc do teste de Friedman, para as correlações de baixa magnitude, mostram que, as diferenças estatisticamente significativas nas potências ocorrem entre a maior raiz de Roy e o traço de Pillai, o lambda de Wilks, o lambda de Wilks aplicado às ordens ($p<0.001$), e o traço de Hotelling ($p=0.009$), e o lambda de Wilks aplicado às ordens ($p=0.037$); entre o traço de Hotelling e o traço de Pillai ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.001$); entre o traço de Pillai, o lambda de Wilks ($p=0.018$), e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.018$).

No cenário com correlações de intensidade média, as diferenças estatisticamente significativas nas potências verificam-se entre a maior raiz de Roy, o traço de Hotelling ($p=0.001$), o lambda de Wilks aplicado às ordens ($p=0.004$), e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.001$), e o traço de Pillai aplicado às ordens ($p=0.005$); entre o traço de Pillai, o lambda de Wilks ($p=0.031$), e o traço de Pillai aplicado às ordens ($p<0.001$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.031$).

No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o lambda de Wilks aplicado às ordens ($p=0.012$) e todas as outras estatísticas ($p<0.001$); o traço de Hotelling com o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.012$), o lambda de Wilks ($p=0.003$), o lambda de Wilks aplicado às ordens ($p=0.030$); entre o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.003$); o lambda de Wilks ($p=0.012$), e o lambda de Wilks aplicado às ordens ($p<0.001$).

Probabilidades Marginais (p_1, p_2, p_4)

Com distribuições de frequência diferentes, em cada um dos três grupos, (p_1, p_2, p_4), simétricas, assimétricas positivas e uniformes na escala de medida, para escalas com **3 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi_F^2(5)=11.941$; $p=0.003$); com correlações de média intensidade ($\chi_F^2(5)=18.111$; $p<0.001$); e com correlações de elevada intensidade ($\chi_F^2(5)=18.022$; $p<0.001$).

No cenário com correlações de intensidade baixa, os *post-hoc* do teste de Friedman mostram diferenças estatisticamente significativas nas potências, entre a maior raiz de Roy, o traço de Hotelling ($p=0.031$), o traço de Pillai ($p=0.001$), o traço de Pillai aplicado às ordens ($p=0.017$), e o lambda de Wilks ($p=0.009$); entre o traço de Pillai e o lambda de Wilks aplicado às ordens ($p=0.031$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling

($p=0.004$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.007$); entre o traço de Hotelling, o traço de Pillai ($p=0.047$), e o lambda de Wilks aplicado às ordens ($p=0.047$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.026$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.005$).

No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o traço de Hotelling ($p=0.003$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.011$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.050$), e o lambda de Wilks aplicado às ordens ($p=0.038$); entre o traço de Pillai, e o traço de Pillai aplicado às ordens ($p=0.015$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.008$).

O teste de Friedman mostra que nas escalas com formato de medida com **4 pontos** (cf. Figura 7.17), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos cenários com correlações de baixa ($\chi_F^2(5) = 16.371$; $p < 0.001$); de média ($\chi_F^2(5) = 18.088$; $p < 0.001$); e de elevada intensidade ($\chi_F^2(5) = 22.246$; $p < 0.001$).

No cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.029$), o traço de Pillai ($p<0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, e o traço de Pillai ($p=0.039$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.022$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.004$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas nas potências surgiram entre a maior raiz de Roy, o traço de Hotelling ($p=0.003$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.010$); entre o traço de Hotelling, e o lambda de Wilks aplicado às ordens ($p=0.036$); entre o traço de Pillai, e o traço de Pillai aplicado às ordens ($p=0.019$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.005$).

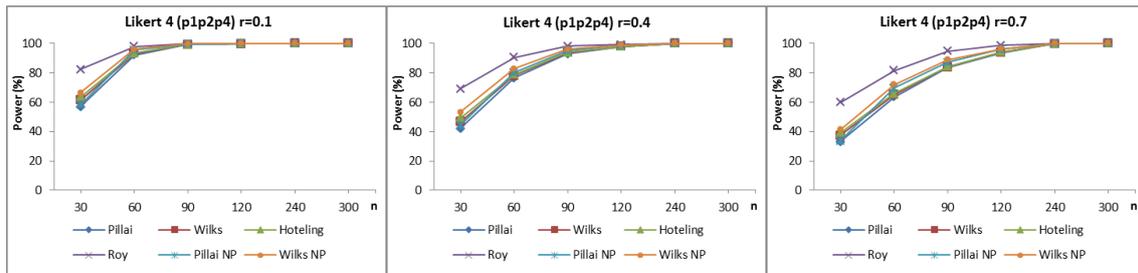


Figura 7.17 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 4 pontos (tipo Likert 4 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$), e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no ponto central e uniformemente distribuída na escala de medida (com distribuições de frequência diferentes em cada um dos três grupos, (p_1, p_2, p_4) , ou seja H_0 falsa).

Para correlações de intensidade elevada as diferenças são entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.005$); entre o traço de Hotelling, o traço de Pillai ($p=0.010$), e o lambda de Wilks aplicado às ordens ($p=0.005$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.020$), e o lambda de Wilks aplicado às ordens ($p=0.001$).

Para escalas com formato de medida com **5 pontos**, o teste de Friedman mostra que as estatísticas da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5) = 17.086$; $p < 0.001$); com correlações de média intensidade ($\chi^2_F(5) = 21.258$; $p < 0.001$); e com correlações de elevada intensidade ($\chi^2_F(5) = 26.231$; $p < 0.001$).

Os *post-hoc* do teste de Friedman mostram que no cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências se confirmam entre a maior raiz de Roy, o traço de Hotelling ($p=0.008$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.010$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, e o traço de Pillai ($p=0.045$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.034$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.010$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, e o lambda de Wilks ($p < 0.001$), o traço de Pillai aplicado às ordens ($p = 0.001$), e o lambda de Wilks aplicado às ordens ($p = 0.044$); entre o traço de Hotelling, o traço de Pillai ($p = 0.011$), e o lambda de Wilks aplicado às ordens ($p = 0.044$); entre o o traço de Pillai, e o traço de Pillai aplicado às ordens ($p < 0.006$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.003$).

No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy e o lambda de Wilks aplicado às ordens ($p = 0.008$) e todas as outras estatísticas ($p < 0.001$); entre o traço de Hotelling, e o traço de Pillai e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks aplicado às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.020$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p = 0.041$), e o lambda de Wilks aplicado às ordens ($p < 0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p < 0.001$).

Nas escalas com formato de medida com **7 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, para as correlações de baixa ($\chi^2_F(5) = 20.706$; $p < 0.006$); de média ($\chi^2_F(5) = 26.476$; $p < 0.001$); e de elevada intensidade ($\chi^2_F(5) = 27.010$; $p < 0.001$).

Os testes *post-hoc* de Friedman mostram que para as correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências se atestam entre a maior raiz de Roy e o traço de Hotelling ($p = 0.001$), o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p < 0.001$); entre o traço de Hotelling, o traço de Pillai ($p = 0.018$), e o lambda de Wilks aplicado às ordens ($p = 0.025$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p = 0.025$), e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks aplicado às ordens ($p = 0.018$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p = 0.002$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks ($p < 0.001$), e o

lambda de Wilks aplicado às ordens ($p=0.021$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.021$); entre o traço de Pillai, e o traço de Pillai aplicado às ordens ($p=0.008$); entre o lambda de Wilks ($p=0.008$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks aplicado às ordens, e o lambda de Wilks ($p<0.001$).

No cenário com correlações de intensidade elevada as diferenças são entre a maior raiz de Roy, e o lambda de Wilks aplicado às ordens ($p=0.013$), e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai, e o lambda de Wilks ($p=0.035$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p<0.001$).

Probabilidades Marginais (p_2, p_2, p_1)

Com distribuições de frequência iguais em dois grupos, assimétricas positivas e simétrica no terceiro grupo (p_2, p_2, p_1), para escalas com **3 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5)=10.000$; $p=0.001$); com correlações de média intensidade ($\chi^2_F(5)=18.282$; $p<0.001$); e com correlações de elevada intensidade ($\chi^2_F(5)=19.388$; $p<0.001$).

Os *post-hoc* do teste de Friedman mostram que no cenário com correlações de intensidade baixa, as diferenças estatisticamente significativas nas potências confirmaram-se entre a maior raiz de Roy, o traço de Hotelling ($p=0.006$), e o lambda de Wilks ($p=0.025$); e entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.025$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling ($p=0.004$), o traço de Pillai, e o lambda de Wilks ($p<0.001$); entre o traço de Hotelling, e o lambda de Wilks aplicado às ordens ($p=0.031$); entre o traço de Pillai, o traço de Pillai

aplicado às ordens ($p=0.006$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p=0.023$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.001$).

No cenário com correlações de elevada magnitude as diferenças nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.003$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.015$); entre o traço de Hotelling, o traço de Pillai ($p=0.020$), e o lambda de Wilks aplicado às ordens ($p=0.039$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.004$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks aplicado às ordens, e o lambda de Wilks ($p=0.003$).

Para escalas com formato de medida com **4 pontos**, o teste de Friedman mostra que as estatísticas da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, para correlações de baixa ($\chi^2_F(5) = 10.000$; $p = 0.001$); de média intensidade ($\chi^2_F(5) = 17.664$; $p < 0.001$); e de elevada intensidade ($\chi^2_F(5) = 18.597$; $p < 0.001$).

Os *post-hoc* do teste de Friedman mostram que no cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências se confirmam entre a maior raiz de Roy, o traço de Pillai ($p=0.006$), e o lambda de Wilks ($p=0.025$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.025$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.022$), o traço de Pillai ($p<0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, e o lambda de Wilks aplicado às ordens (0.012); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.002$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.006$); entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.001$).

Nas correlações de intensidade elevada as diferenças são entre a maior raiz de Roy, o traço de Hotelling ($p=0.010$), o traço de Pillai ($p<0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, e o traço de Pillai ($p=0.034$), e o lambda de Wilks aplicado às ordens ($p=0.025$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.001$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai

aplicado às ordens, e o lambda de Wilks ($p=0.013$); entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.002$).

Para escalas com formato de medida com **5 pontos** (cf. Figura 7.18), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nas correlações de baixa intensidade ($\chi^2_F(5)=19.046$; $p < 0.001$); de média intensidade ($\chi^2_F(5)=18.597$; $p < 0.001$); e de elevada intensidade ($\chi^2_F(5)=23.795$; $p < 0.001$).

Os *post-hoc* do teste de Friedman mostram que para correlações de baixa magnitude, as diferenças estatisticamente significativas se atestam entre a potência da maior raiz de Roy e o traço de Hotelling ($p=0.002$), o traço de Pillai e o lambda de Wilks ($p < 0.001$); entre o traço de Hotelling e o lambda de Wilks aplicado às ordens ($p=0.014$); o traço de Pillai e o traço de Pillai aplicado às ordens ($p=0.002$), lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.010$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.001$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a potência da maior raiz de Roy, o traço de Hotelling ($p=0.010$), o traço de Pillai ($p < 0.001$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.034$), e o lambda de Wilks aplicado às ordens ($p=0.025$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.001$), o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens e o lambda de Wilks ($p=0.001$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p=0.002$).

Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o lambda de Wilks ($p < 0.001$), e o traço de Pillai aplicado às ordens ($p=0.007$); entre o traço de Hotelling, o traço de Pillai ($p=0.007$), o traço de Pillai aplicado às ordens ($p=0.047$), e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p < 0.001$).

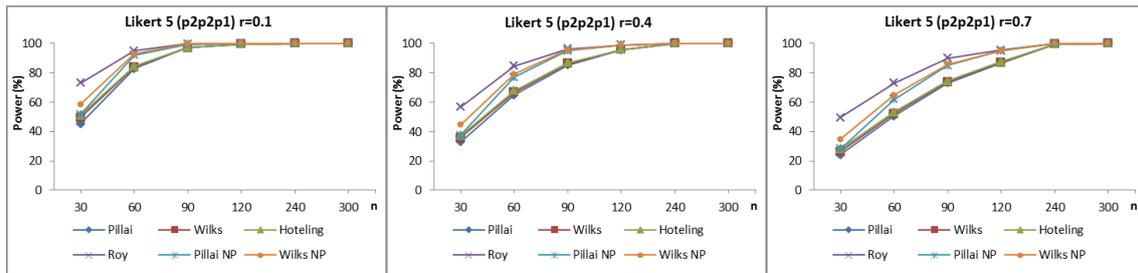


Figura 7.18 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks-W, traço de Hotelling-H, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 5 pontos (Likert 5 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$), e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada no extremo inferior e no ponto central da escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_2, p_2, p_1) , ou seja H_0 falsa).

Nas escalas com formato de medida com **7 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5) = 25.102$; $p < 0.006$); com correlações de média intensidade ($\chi^2_F(5) = 26.602$; $p < 0.001$); e com correlações de elevada intensidade ($\chi^2_F(5) = 27.813$; $p < 0.001$).

O *post-hoc* do teste de Friedman mostra que no cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências surgem entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p < 0.001$); entre o traço de Hotelling, o traço de Pillai ($p = 0.008$), e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p = 0.002$), e o lambda de Wilks aplicado às ordens ($p = 0.019$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks ($p < 0.001$), e o lambda de Wilks aplicado às ordens ($p = 0.019$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p < 0.001$), e o lambda de Wilks ($p = 0.045$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p < 0.001$), o lambda de Wilks ($p = 0.045$), e o lambda de Wilks aplicado às ordens ($p < 0.001$); entre o traço de Pillai

aplicado às ordens, e o lambda de Wilks ($p=0.002$); e entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$).

Nas correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.004$), e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.002$); entre o traço de Pillai, e o lambda de Wilks ($p=0.008$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

Probabilidades Marginais (p_4, p_4, p_2)

O teste de Friedman mostra que com distribuições de frequência uniformes em 2 grupos e assimétrica positiva no terceiro grupo, (p_4, p_4, p_2), com escalas com **3 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5)=15.639$; $p=0.001$); com correlações de média intensidade ($\chi^2_F(5)=17.806$; $p<0.001$); e com correlações de elevada intensidade ($\chi^2_F(5)=20.327$; $p<0.001$).

Os *post-hoc* do teste Friedman mostram que para correlações de intensidade baixa, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.023$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.001$), e o lambda de Wilks ($p=0.004$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.005$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.023$).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling ($p=0.030$), o traço de Pillai, o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.002$); entre o traço de Hotelling, e o traço de Pillai ($p=0.009$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.009$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.030$).

Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o traço de Hotelling ($p=0.002$), o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks ($p<0.001$), e o lambda de Wilks aplicado às ordens ($p=0.039$); entre o traço de Hotelling, e o traço de Pillai ($p=0.008$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.005$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.028$).

Nas escalas com formato de medida com **quatro pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5)=17.713$; $p<0.001$); com correlações de média intensidade ($\chi^2_F(5)=17.571$; $p<0.001$); e com correlações de elevada intensidade ($\chi^2_F(5)=21.140$; $p<0.001$).

Os *post-hoc* do teste de Friedman mostram que no cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.010$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.001$); entre o traço de Hotelling, e o traço de Pillai ($p=0.045$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.010$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.007$).

No cenário com correlações de média intensidade, as diferenças estatisticamente significativas nas potências surgiram entre a maior raiz de Roy, o traço de Hotelling ($p=0.013$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.004$); entre o traço de Hotelling, e o traço de Pillai ($p=0.024$); o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.042$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.007$).

Nas correlações de intensidade elevada as diferenças são entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p<0.001$); entre o traço de Hotelling, o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.016$); entre o traço de Pillai, o traço de Pillai aplicado às

ordens ($p=0.030$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.008$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.001$).

Para escalas com formato de medida com **5 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários com correlações de baixa ($\chi_F^2(5)=17.355$; $p < 0.001$); de média ($\chi_F^2(5)=23.324$; $p < 0.001$); e de elevada intensidade ($\chi_F^2(5)=25.421$; $p < 0.001$). No cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.013$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.003$); entre o traço de Hotelling, e o traço de Pillai ($p=0.032$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.032$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.007$). Com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.016$) e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.003$), e o lambda de Wilks aplicado às ordens ($p=0.033$); entre o o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.033$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.003$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p<0.001$). Para as correlações de elevada intensidade, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.016$) e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.001$), e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o o traço de Pillai, o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.037$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.025$), e o lambda de Wilks aplicado às ordens ($p=0.004$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p<0.001$).

Nas escalas com formato de medida com **7 pontos** (cf. Figura 7.19), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi^2_F(5) = 20.353$; $p < 0.006$), com correlações de média intensidade ($\chi^2_F(5) = 25.288$; $p < 0.001$); e com correlações de elevada intensidade ($\chi^2_F(5) = 27.048$; $p < 0.001$). Os *post-hoc* do teste de Friedman mostram que com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai aplicado às ordens ($p=0.001$), o traço de Pillai, e o lambda de Wilks ($p<0.001$); entre o traço de Hotelling, e o traço de Pillai ($p=0.011$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.021$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.002$). No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks ($p<0.001$), e o lambda de Wilks aplicado às ordens ($p=0.019$); entre o traço de Hotelling, o traço de Pillai ($p<0.001$), e o lambda de Wilks aplicado às ordens ($p=0.004$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.008$); entre o traço de Pillai aplicado e o lambda de Wilks aplicados às ordens ($p=0.001$); e o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

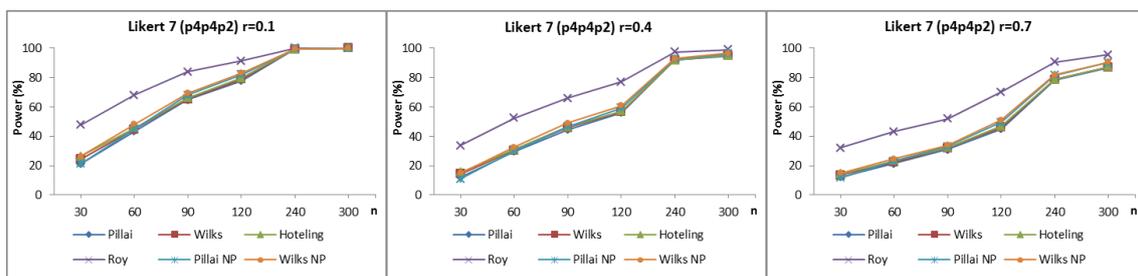


Figura 7.19 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$), média ($r=0.40$) e elevada intensidade ($r=0.70$), e com maior probabilidade de resposta aos itens localizada uniformemente e no extremo inferior da escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_4, p_4, p_2) , ou seja H_0 falsa).

Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.012$), e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.004$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.012$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

Probabilidades Marginais (p_2, p_2, p_4)

O teste de Friedman mostra que com distribuições de frequência diferentes, em dois dos três grupos, (p_2, p_2, p_4), assimétricas positivas e uniformemente distribuídas na escala de medida, com escalas com **3 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais: com correlações de baixa intensidade ($\chi_F^2(5)=16.271$; $p<0.001$); com correlações de média intensidade ($\chi_F^2(5)=16.971$; $p<0.001$); e com correlações de elevada intensidade ($\chi_F^2(5)=20.844$; $p<0.001$).

Os *post-hoc* do teste de Friedman exibem diferenças estatisticamente significativas nas potências, no cenário com correlações de intensidade baixa, entre a maior raiz de Roy, o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.002$), e o lambda de Wilks ($p=0.014$); entre o traço de Hotelling, o traço de Pillai ($p=0.002$), e o traço de Pillai aplicado às ordens ($p=0.036$); entre o traço de Pillai, o lambda de Wilks ($p=0.048$), e o lambda de Wilks aplicado às ordens ($p=0.002$).

Para as correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling ($p=0.025$), o traço de Pillai, o traço de Pillai aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.004$); entre o traço de Hotelling, e o traço de Pillai ($p=0.019$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.010$).

No cenário com correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o traço de Hotelling ($p=0.001$), o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks ($p<0.001$); entre o traço de Hotelling, o traço

de Pillai ($p=0.028$), e lambda de Wilks aplicado às ordens ($p=0.028$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.004$).

Nas escalas com formato de medida com **4 pontos**, os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos cenários com correlações de baixa ($\chi^2_F(5) = 14.580$; $p = 0.001$); de média intensidade ($\chi^2_F(5) = 17.158$; $p < 0.001$); e de elevada intensidade ($\chi^2_F(5) = 20.808$; $p < 0.001$). No cenário com correlações de baixa magnitude, as diferenças estatisticamente significativas nas potências verificam-se entre a maior raiz de Roy, o traço de Hotelling ($p=0.015$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.006$), e o lambda de Wilks ($p=0.001$); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.020$). Para as correlações de média intensidade, as diferenças estatisticamente significativas nas potências surgiram entre da maior raiz de Roy, o traço de Hotelling ($p=0.010$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.002$), e o lambda de Wilks ($p=0.001$); entre o traço de Hotelling, e o traço de Pillai (0.034); entre o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p=0.001$); o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.025$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.019$). Com correlações de intensidade elevada as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Hotelling ($p=0.001$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.004$); entre o traço de Hotelling, o traço de Pillai ($p=0.031$), e o lambda de Wilks aplicado às ordens ($p=0.009$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.004$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.001$).

Para escalas com formato de medida com **5 pontos**, o teste de Friedman indica que as estatísticas da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos três cenários correlacionais; com correlações de baixa intensidade ($\chi^2_F(5) = 16.898$; $p < 0.001$); com correlações de média intensidade ($\chi^2_F(5) = 19.731$; $p < 0.001$); e correlações de elevada intensidade

($\chi_F^2(5) = 25.245$; $p < 0.001$). O *post-hoc* do teste de Friedman mostra que no cenário com correlações de intensidade baixa, as diferenças estatisticamente significativas das potências se comprovam entre a maior raiz de Roy, o traço de Hotelling ($p=0.004$), o traço de Pillai ($p<0.001$), o traço de Pillai aplicado às ordens ($p=0.010$), e o lambda de Wilks ($p=0.001$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.034$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.019$). No cenário com correlações de média magnitude, as diferenças estatisticamente significativas ocorreram entre a maior raiz de Roy, o traço de Hotelling ($p=0.002$), o traço de Pillai, o lambda de Wilks ($p<0.001$), e o traço de Pillai aplicado às ordens ($p=0.006$); entre o traço de Hotelling, o traço de Pillai ($p=0.023$), e o lambda de Wilks aplicado às ordens ($p=0.041$); entre o traço de Pillai, o traço de Pillai aplicado às ordens ($p=0.006$), e o lambda de Wilks aplicado às ordens ($p<0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.002$). Para as correlações de intensidade elevada as diferenças são entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.028$) e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.003$), e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.003$); e entre o lambda de Wilks e o lambda de Wilks aplicado às ordens ($p<0.001$).

Nas escalas com formato de medida com **7 pontos** (cf. Figura 7.20), os testes da MANOVA paramétrica e não paramétrica apresentam diferenças estatisticamente significativas nas potências, nos cenários com correlações de baixa ($\chi_F^2(5) = 24.314$; $p < 0.001$); de média ($\chi_F^2(5) = 26.731$; $p < 0.001$); e de elevada intensidade ($\chi_F^2(5) = 26.292$; $p < 0.001$). O *post-hoc* do teste de Friedman mostra que para correlações de intensidade baixa, as diferenças estatisticamente significativas nas potências são entre a maior raiz de Roy, o traço de Pillai aplicado às ordens ($p=0.001$), o traço de Hotelling, o traço de Pillai, e o lambda de Wilks ($p<0.001$); entre o traço de Hotelling, o traço de Pillai ($p=0.003$), e o lambda de Wilks aplicado às ordens ($p=0.001$); entre o traço de Pillai, e o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, e o

lambda de Wilks ($p=0.003$); entre o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p=0.030$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

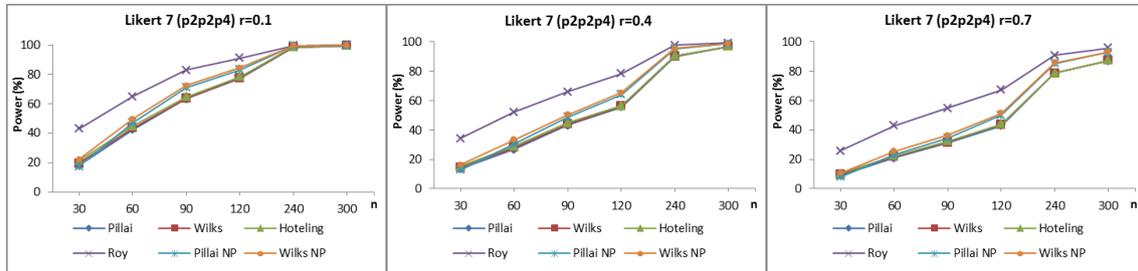


Figura 7.20 - Potência de teste (em porcentagem) nas quatro estatísticas da MANOVA paramétrica, traço de Pillai-P, lambda de Wilks, traço de Hotelling, maior raiz de Roy e nos dois da MANOVA não paramétrica, traço de Pillai aplicado às ordens (Pillai NP), lambda de Wilks aplicado às ordens (Wilks NP) e para amostras de dimensão $n_1=n_2=n_3=10, 20, 30, 40, 80, 100$, com formato de medida com 7 pontos (Likert 7 pontos), com correlações de baixa ($r=0.10$) média ($r=0.40$) e elevada intensidade ($r=0.70$), com maior probabilidade de resposta aos itens localizada no extremo inferior e uniformemente distribuída na escala de medida (com distribuições de frequência diferentes em dois dos três grupos, (p_2, p_2, p_4) , ou seja H_0 falsa).

No cenário com correlações de média magnitude, as diferenças estatisticamente significativas nas potências ocorreram entre a maior raiz de Roy, o traço de Hotelling, o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks ($p<0.001$), e o lambda de Wilks aplicado às ordens ($p=0.002$); entre o traço de Hotelling, o traço de Pillai, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.042$); entre o traço de Pillai, o traço de Pillai aplicado às ordens, o lambda de Wilks aplicado às ordens ($p<0.001$), e o lambda de Wilks ($p=0.010$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p=0.002$); e entre o lambda de Wilks aplicado às ordens, e o lambda de Wilks ($p<0.001$). Para as correlações de elevada magnitude as diferenças são entre a maior raiz de Roy, o lambda de Wilks aplicado às ordens ($p=0.024$) e todas as outras estatísticas ($p<0.001$); entre o traço de Hotelling, o traço de Pillai, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai, o lambda de Wilks ($p=0.015$), o traço de Pillai aplicado às ordens, e o lambda de Wilks aplicado às ordens ($p<0.001$); entre o traço de Pillai aplicado às ordens, o lambda de Wilks ($p=0.009$), e o lambda de Wilks aplicado às ordens ($p=0.001$); e entre o lambda de Wilks, e o lambda de Wilks aplicado às ordens ($p<0.001$).

7.4 Discussão dos Resultados

O presente estudo de simulação teve como objeto a análise do desempenho das estatísticas da MANOVA paramétrica e não paramétrica, ou seja, avaliar a probabilidade de erro de tipo I e a potência de teste com variáveis ordinais, formatos de medida com pontos diferentes, diferentes distribuições e magnitude de correlação entre variáveis. A importância desta avaliação deve-se ao facto de o investigador poder tomar decisões incorretas quando rejeita uma hipótese nula verdadeira (erro de Tipo I) ou quando não rejeita uma hipótese nula falsa (erro de Tipo II), o que interfere com a potência dos testes.

Para atingirmos o objetivo, procurámos construir diversos cenários, que incluíam quatro formatos de medida (3, 4, 5, e 7 pontos), distribuições de frequência simétricas, assimétricas positivas e assimétricas negativas e uniformemente distribuídas, correlações de diferente magnitude (baixa, média e elevada), três variáveis dependentes, uma variável independente com três níveis e seis dimensões de amostras, esperando abranger um número alargado das situações com que o investigador poderá deparar-se na vida real e na prática da investigação.

Os resultados deste estudo de simulação mostram valores de probabilidade de erro de tipo I muito elevados para a maior raiz de Roy (0.14) e, de um modo geral inferior ao nível de significância nominal ou ligeiramente superior nas restantes estatísticas da MANOVA nos diferentes cenários delineados para este estudo.

No que concerne à potência dos testes, das estatísticas da MANOVA verificam-se três situações distintas, uma situação em que os valores da potência dos testes são muito baixos, em qualquer dos cenários delineados, outra em que os valores das potências são elevados com uma exceção para amostras com 30 observações em que nos cenários com correlações de elevada intensidade as potências são baixas, e uma terceira situação em que o valor das potências é influenciado pela dimensão da amostra, pela distribuição de frequência, pelo número de pontos do formato de medida e pela intensidade das correlações.

A **primeira situação**, em que temos um cenário em que a distribuição é uniforme em dois dos grupos e simétrica no terceiro e outro cenário em que a distribuição é simétrica em dois dos grupos e uniforme no terceiro verificou-se que seja

qual for o número de pontos da escala, a intensidade da correlação entre os itens, ou a dimensão da amostra, a magnitude da potência das estatísticas da MANOVA é muito baixa. No sentido de justificar a existência de potências de baixa magnitude, foi determinada a assimetria e a curtose médias, nos três grupos, nas mil réplicas, utilizando a biblioteca “*rela*” (<http://cran.r-project.org/web/packages/rela>) (cf. Apêndice I) e calculadas as médias dos valores observados da estatística de teste de Kruskal-Wallis das três variáveis nos diferentes cenários simulados e a sua correlação (coeficiente de correlação de Pearson) com os valores das potências dos testes da MANOVA paramétrica e da MANOVA não paramétrica.

Os resultados da assimetria (g_1) e curtose (g_2) mostram valores próximos de zero (entre -0.5 e 0.5) nos cenários com potências de baixa e elevada magnitude e escalas com formato de medida com 3 e 4 pontos. Nas escalas com formato de medida com 5 e 7 pontos os resultados são próximos ou superiores a 1. Porém, na 1ª situação com potências de baixa magnitude e em todos os cenários simulados os resultados de g_1 são muito semelhantes nos três grupos, sejam ou não os valores de g_1 próximos de zero, o que permite dizer que as distribuições de frequência são tão idênticas nos 3 níveis que, a MANOVA não deteta as diferenças entre grupos. Os resultados das médias das estatísticas de Kruskal-Wallis são pequenos o que reforça a ideia de o foco de atenção ser o erro de tipo I e não a potência de teste (cf. Apêndice II).

Nas outras duas situações destaca-se a maior raiz de Roy com valores mais elevados de potência do que qualquer uma das outras estatísticas, o que vai ao encontro do referido na literatura científica. A maior raiz de Roy é mais potente do que as outras estatísticas, sobretudo se estivermos perante correlações de elevada intensidade entre as variáveis dependentes (Ramsey & Ramsey, 2012; Marôco, 2010; Hair et al., 2010; Johnson & Wichern, 2007).

Contrariamente ao verificado nas potências de baixa magnitude (1ª situação), no caso das potências de elevada magnitude (2ª e 3ª situações), a distribuição de frequência nos grupos é diferente em pelo menos um dos 3 grupos, mediante o tipo de probabilidade marginal e os resultados das médias das estatísticas de Kruskal-Wallis são maiores do que na situação anterior (cf. Apêndice III).

A **segunda situação**, em que os testes da MANOVA reportam a 3 cenários nas diferentes situações simuladas: distribuição assimétrica positiva e assimétrica negativa, e uniformemente distribuída; distribuição assimétrica negativa e assimétrica positiva; e a maior probabilidade de resposta aos itens localizada nos extremos da escala, distribuição assimétrica positiva e assimétrica negativa; as estatísticas apresentam valores de potência aceitáveis ou elevados, superiores a 60%, nas amostras com dimensão igual ou superior a 60 observações. Na escala com formato de medida com 7 pontos e nas situações simuladas com amostras com 60 observações, as potências apresentam resultados mais baixos. Assim, podemos dizer que, nas situações simuladas, com um com um número de observações superior a 30 com distribuições de resposta aos itens do tipo anteriormente descrito, para as escalas com 3, 4 e 5 pontos, os testes apresentam valores que oscilam entre 60% e 100%, sendo o valor mais baixo considerado como aceitável (Land & Hui, 2010; Hill & Hill, 2009), não se verificando diferenças entre as estatísticas da MANOVA paramétrica e da MANOVA não paramétrica. Importa salientar que para as 3 primeiras escalas e com amostras com dimensão superior a 60 observações, a magnitude das potências se eleva para 90% e nas escalas com 7 pontos para 80% nas amostras com dimensão superior a 90 observações. Estes resultados estão em consonância com os resultados dos estudos de Johnson e Wichern (2007) que consideram que em amostras de grande dimensão, os resultados dos testes paramétricos e não paramétricos são equivalentes.

A **terceira situação**, em que os testes da MANOVA reportam a 5 tipos diferentes de probabilidades marginais, nas diferentes situações simuladas: (p_1, p_1, p_2) , em que a maior probabilidade de resposta aos itens está localizada no ponto central e no extremo inferior da escala (distribuições diferentes em dois dos grupos, simétricas e assimétrica positiva); (p_1, p_2, p_4) , em que a maior probabilidade de resposta aos itens está localizada no ponto central, no extremo inferior e uniformemente distribuída (simétrica, assimétrica positiva e uniforme); (p_2, p_2, p_1) , em que a maior probabilidade de resposta aos itens está localizada no extremo inferior e no ponto central da escala (distribuição assimétricas positiva e simétrica); (p_4, p_4, p_2) , em que a maior probabilidade de resposta aos itens está localizada uniformemente na escala e no extremo inferior da escala (distribuição uniforme e assimétrica positiva); e (p_2, p_2, p_4) , em que a maior

probabilidade de resposta aos itens está localizada no extremo inferior e uniformemente na escala (distribuição assimétrica positiva e uniforme); a maior raiz de Roy apresenta valores mais elevados de potência do que qualquer outro dos testes da MANOVA, seguido do lambda de Wilks aplicado às ordens.

Da análise dos resultados podemos dizer que a potência dos testes da MANOVA é influenciada pela distribuição de frequência, pela dimensão da amostra, pelo número de pontos da escala, pela magnitude das correlações e pelo tipo de teste.

Quanto menor a dimensão da amostra e maior a intensidade da correlação entre variáveis, menor a potência dos testes, e os testes não paramétricos apresentam valores de potência mais elevados nas amostras de menor dimensão, sendo o lambda de Wilks aplicado às ordens mais potente do que o traço de Pillai aplicado às ordens.

Van Aelst e Willems (2011) mostraram no seu estudo de simulação que o lambda de Wilks, aplicado às ordens, apresenta valores de potência elevados confirmando o referido anteriormente por Nath e Pavur (1985) e Finch (2005) e que também se confirma neste estudo de simulação. Os estudos de Katz e MacSweeney (1980) incentivam o uso do teste de Pillai aplicado às ordens sempre que o investigador tem amostras de pequena ou moderada dimensão. Verifica-se ainda que as estimativas das potências diminuem, de forma acentuada, à medida que aumenta o número de pontos do formato de medida (de 3 para 7 pontos), à medida que aumenta a intensidade da correlação entre os itens (da baixa intensidade para a média intensidade, e desta para a elevada intensidade); este resultado relacionado com a magnitude das correlações está em concordância com os resultados obtidos por Ramsey (1982) na sua investigação. O autor considerou que à medida que a magnitude das correlações entre as variáveis dependentes aumentam, a potência da MANOVA diminui, o que se verifica na terceira situação simulada neste estudo e para os diferentes cenários. Assim, podemos dizer que para os 5 cenários simulados nesta situação, para além da valorização da dimensão da amostra e do número de pontos da escala, o investigador deve valorizar também a correlação entre variáveis independentes (itens).

A magnitude das potências dos testes paramétricos e não paramétricos, assemelham-se nos diferentes cenários simulados e com escalas com 3, 4, e 5 pontos, nas amostras com dimensão superior a 90, 120 e 240 observações e com correlações de baixa, média e

elevada magnitude (0.10, 0.40, 0.70 respetivamente) e nas escalas com 7 pontos a apenas nas amostras com dimensão superior ou igual a 120, para correlações de baixa e média intensidade e, 240 observações, para correlações de elevada intensidade.

O facto de se usarem escalas com um número par ou ímpar de pontos no formato de medida não parece interferir com os resultados. Apesar do estudo de simulação desenvolvido neste trabalho ser completamente diferente dos estudos desenvolvidos por Lozano, García-Cueto e Muñiz (2008) e Moors (2007), que desenvolveram estudos para obter a fiabilidade e validade dos itens das escalas, os resultados referentes à escala de medida mostram que escalas com um número de pontos superior a quatro e inferior a sete, independentemente de possuírem ponto central ou não, dão melhores resultados, o que está em consonância com os resultados deste estudo.

Ao relacionarmos (através do coeficiente de correlação de Pearson) as médias dos valores observados da estatística de teste de Kruskal-Wallis das três variáveis nas diferentes situações simuladas com os valores das potências dos testes da MANOVA paramétrica e da MANOVA não paramétrica, verificámos, na maioria dos casos, correlações positivas e de elevada magnitude (>0.70); com dados dos diferentes formatos de medida (3, 4 e 5 pontos); ou média (entre 0.40 e 0.70) com dados da escala com formato de medida com 7 pontos (cf.

Tabela 7. 6, 7.7, 7.8), o que explica o facto de na primeira situação as potências serem de baixa magnitude.

Tabela 7. 6 - Correlações (e respetivo *p-value*) entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para as situações de simulação com $n=30, 60, 120, 300$, nas escalas com 3 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica.

Formato de medida com 3 pontos							
Kruskal-Wallis		Pillai	Wilks	Hoteling	Roy	PillaNP WilksNP	
Média para cada n							
KW ₃₀	Pearson Correlation	0.867**	0.868**	0.867**	0.854**	0.872**	0.868**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₆₀	Pearson Correlation	0.866**	0.865**	0.864**	0.852**	0.861**	0.859**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₁₂₀	Pearson Correlation	0.796**	0.796**	0.796**	0.793**	0.795**	0.794**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₃₀₀	Pearson Correlation	0.821**	0.821**	0.821**	0.821**	0.821**	0.821**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 7. 7- Correlações (e respetivo *p-value*) entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para as situações de simulação com $n=30, 60, 120, 300$, nas escalas com 4 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica.

Formato de medida com 4 pontos

Kruskal-Wallis Média para cada n		Pillai	Wilks	Hoteling	Roy	PillaNP	WilksNP
KW ₃₀	Pearson Correlation	0.776**	0.787**	0.790**	0.822**	0.777**	0.792**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₆₀	Pearson Correlation	0.866**	0.866**	0.866**	0.860**	0.865**	0.864**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₁₂₀	Pearson Correlation	0.840**	0.840**	0.840**	0.834**	0.837**	0.837**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₃₀₀	Pearson Correlation	0.831**	0.831**	0.831**	0.831**	0.831**	0.831**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 7. 8 - Correlações (e respetivo *p-value*) entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para as situações de simulação com $n=30, 60, 120, 300$, nas escalas com 5 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica.

Formato de medida com 5 pontos

Kruskal-Wallis Média para cada n		Pillai	Wilks	Hoteling	Roy	PillaNP	WilksNP
KW ₃₀	Pearson Correlation	0.772**	0.783**	0.789**	0.858**	0.773**	0.798**
	Sig. (2-tailed)	0.001	0.001	<0.001	<0.001	0.001	<0.001
KW ₆₀	Pearson Correlation	0.816**	0.814**	0.814**	0.804**	0.810**	0.808**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₁₂₀	Pearson Correlation	0.870**	0.869**	0.869**	0.860**	0.865**	0.865**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
KW ₃₀₀	Pearson Correlation	0.851**	0.851**	0.851**	0.541*	0.851**	0.851**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Nas escalas com formato de medida com 7 pontos e na situação de simulação com amostras de menor dimensão ($n=30$) as correlações são positivas, mas não são significativas. Nas situações simuladas com amostras com 60 e 120 observações as correlações são positivas e de magnitude média (entre 0.40 e 70) e para 300 observações as correlações são positivas e de magnitude elevada (>0.70) (cf. Tabela 7. 9).

Tabela 7. 9 - Correlações (e respetivo *p-value*) entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis, nas mil réplicas, para as situações simuladas com $n=30, 60, 120, 300$, nas escalas com 7 pontos, e a potência de teste das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica.

Formato de medida com 7 pontos							
Kruskal-Wallis		Pillai	Wilks	Hotelling	Roy	PillaNP	WilksNP
Média para cada n							
KW ₃₀	Pearson Correlation	0.395	0.378	0.384	0.424	0.386	0.377
	Sig. (2-tailed)	0.145	0.164	0.158	0.116	0.155	0.167
KW ₆₀	Pearson Correlation	0.567*	0.570*	0.573*	0.621*	0.568*	0.577*
	Sig. (2-tailed)	0.027	0.026	0.026	0.014	0.027	0.024
KW ₁₂₀	Pearson Correlation	0.567*	0.570*	0.573*	0.621*	0.568*	0.577*
	Sig. (2-tailed)	0.027	0.026	0.026	0.014	0.027	0.024
KW ₃₀₀	Pearson Correlation	0.849**	0.848**	0.848**	0.836**	0.846**	0.846**
	Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

** . Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

7.5 Conclusões

A análise multivariada de dados ordinais que resulta da utilização de escalas de avaliação, tem levantado alguma discussão pelo facto de se aplicarem testes estatísticos sem se verificarem os seus pressupostos de aplicação, usando indiscriminadamente testes paramétricos e não paramétricos e também não ser levada em conta a dimensão da amostra, o número de pontos da escala, a magnitude das correlações e as probabilidades marginais. Os estudos de simulação têm sido fundamentais na tomada de decisão do investigador sobre os testes a aplicar e sobre a importância do cumprimento dos pressupostos da sua utilização, o que tem vindo a ser de muita utilidade aos profissionais que na sua prática clínica desenvolvem, utilizam, adaptam e validam instrumentos compostos por variáveis ordinais.

O desenvolvimento deste estudo de simulação através do método de Monte Carlo teve como objetivo comparar a potência e a probabilidade de erro de tipo I, de quatro estatísticas da MANOVA paramétrica com duas estatísticas da MANOVA não paramétrica, quando se utilizam variáveis ordinais correlacionadas.

Da análise dos resultados concluímos que a probabilidade do erro de tipo I e a potência de teste são mais elevados na maior raiz de Roy do que em qualquer uma das

outras estatísticas da MANOVA. A probabilidade de erro de tipo I é muito superior ao nível de significância nominal na maior raiz de Roy e, de um modo geral, inferior ao nível de significância nominal ($\alpha = 0.05$) ou ligeiramente superior nas restantes estatísticas da MANOVA paramétrica e da não paramétrica (traço de Pillai, lambda de Wilks, traço de Hotelling e traço de Pillai e Lambda de Wilks aplicados às ordens). O traço de Pillai aplicado às ordens é o teste que apresenta, de um modo geral, probabilidade de erro de tipo I menos elevada.

Os resultados da potência de teste apresentam um comportamento diferente. Podem verificar-se potências de muito baixa magnitude nos 2 cenários simulados (primeira situação); potências de elevada magnitude nos 3 cenários simulados para amostras com dimensão superior a 60 observações e escalas com 3, 4 e 5 pontos; e para amostras com dimensão superior a 90 observações, para escalas com 7 pontos (segunda situação); e potências de magnitude mais ou menos elevada (terceira situação) dependendo do número de pontos da escala de medida, da intensidade das correlações entre as variáveis, da dimensão da amostra, ou seja quanto menor a dimensão da amostra, maior o número de pontos da escala e maior a correlação entre variáveis dependentes, menor a potência dos testes e o lambda de Wilks aplicado às ordens é o teste com resultados mais elevados seguido do traço de Pillai aplicado às ordens. No entanto, a magnitude das potências dos testes paramétricos e não paramétricos assemelha-se nas amostras com maior dimensão e correlações de baixa e média magnitude.

Concluimos assim que tendo em atenção a distribuição de frequência das respostas aos itens nos diferentes grupos, na primeira situação, com um cenário com distribuições de frequência simétricas em 2 grupos e uniforme no terceiro grupo e distribuição uniforme em 2 grupos e simétrica no terceiro, as distribuições são muito semelhantes nos 3 grupos e as médias das estatísticas de Kruskal-Wallis das 3 variáveis são pequenas o que explica o facto de as potências serem de baixa magnitude (pois a leitura é feita como se do erro de tipo I se tratasse). De facto, verificámos que a existência de potências baixas se deve ao facto de a magnitude da diferença entre grupos ser pequena, o que é explicado pelas correlações entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis e, a potência dos testes da

MANOVA paramétrica e da MANOVA não paramétrica, que apresentam na sua maioria correlações positivas e de elevada ou média magnitude (>0.70 , entre 0.40 e 0.70) nos diferentes cenários.

Na segunda situação e, tendo a em atenção a distribuição de frequência das respostas aos itens nos diferentes grupos, as potências das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica são elevadas ou aceitáveis mesmo nos cenários de simulação com amostras de menor dimensão (superiores 30 e iguais a 60 observações) nas escalas com 3, 4, 5 pontos, seja qual for a correlação entre variáveis. Nas escalas com 7 pontos a magnitude das potências diminui consideravelmente nas amostras de menor dimensão e apresenta-se elevada nas amostras de dimensão igual ou superior a 120 observações; o que pode confirmar a utilização de testes paramétricos por muitos investigadores na área das ciências sociais e ciências da saúde, em detrimento dos testes não paramétricos, no entanto, salienta-se a importância da dimensão da amostra e o número de pontos da escala de medida.

Na terceira situação, as estimativas das potências diminuem, de forma acentuada, à medida que a dimensão da amostra diminui e aumenta o número de pontos da escala e a intensidade da correlação entre os itens. Assim, podemos dizer que para os 5 cenários simulados nesta situação, para além da valorização da dimensão da amostra e do número de pontos da escala, o investigador deve valorizar também a correlação entre variáveis independentes (itens).

O lambda de Wilks aplicado às ordens e traço de Pillai aplicados às ordens, apresentam valores de potência mais elevados do que a maioria dos testes paramétricos, podendo por isso, ser uma boa solução para este tipo de distribuições, sobretudo se estivermos a trabalhar com amostras de menor dimensão e com correlações elevadas entre os itens.

Importa ainda referir, que nos diferentes cenários simulados e com escalas com 3, 4, e 5 pontos, a magnitude da potência dos testes paramétricos e não paramétricos se assemelha nas amostras com dimensão superior a 90, 120 e 240 observações e com correlações de baixa, média e elevada magnitude (0.10 , 0.40 , 0.70 respetivamente) entre as variáveis dependentes. Nas escalas com 7 pontos a magnitude da potência dos testes paramétricos e não paramétricos também se assemelha, mas apenas nas amostras com

dimensão superior ou igual a 120, para correlações de baixa e média intensidade e, 240 observações, para correlações de elevada intensidade.

Para além do estudo do comportamento dos resultados da potência dos testes da MANOVA e, tendo presente que os determinantes da potência de um teste estatístico são o nível de significância, a dimensão da amostra e a dimensão do efeito, teria sido uma mais-valia para este estudo determinar uma medida da dimensão do efeito e fazer a sua comparação nas estatísticas paramétricas e não paramétricas. Porém como não estão ainda feitas as deduções estatísticas deste cálculo para os testes não paramétricos, não foi viável a sua execução neste estudo de simulação. Deixando aqui uma sugestão para próximos trabalhos.

7.6 Referências Bibliográficas

Alwin, D.F. (1992). Information Transmission in the Survey Interview: Number of response categories and the reliability of attitude measurement. In P.V. Marsden (Ed.). *Sociological Methodology*. Washington: American Sociological Association, 83-118.

Alwin, D.F., & Krosnick, J.A. (1991). The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods and Research*, 20(1), 139-181.

Bilodeau, M., & Micheaux, P.L. (2005). A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, 95, 345-369.

Blair, R.C., & Higgins, J.J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon's signed-rank test under various population shapes. *Psychological Bulletin*, 97, 119-128.

Carrasco, C.G., & Lemes, T.S. (2014). Uma Avaliação do Erro Tipo II no Uso do Teste *t-student*. *Revista Eletrônica Paulista de Matemática*, 3(1), 7-15.

Carrasco, C.G., & Silva, L.A. (2009). Avaliação do erro tipo I na aplicação de um teste de hipóteses para a média. *Revista Mosaicum*, 9, 63-68.

Carrasco, C.G., & Silva, L.A. (2013). Um estudo do erro tipo II em um teste de hipóteses para a média. *Revista Nucleus*, 10(2), 7-12.

Cardinal, R.N. & Aitken, M.R.F. (2006). *ANOVA for the behavioural sciences researcher*. New Jersey: Lawrence Erlbaum Associates.

Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3 (3), 106-116.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale: Lawrence Erlbaum.
- Contractor, S.H. & Fox R.J. (2011). An investigation of the relationship between the number of response categories and scale sensitivity. *Journal of Targeting, Measurement and Analysis for Marketing*, 19, 23-35.
- Curado, M.A.S., Teles, J., & Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, 446-457. <http://www.um.es/eglobal>.
- Curado, M.A.S., Teles, J. & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process, *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48, 1, 146-152.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales, *International Journal of Market Research*, 50(1), 61-77.
- Demirtas, H. (2006). A method for multivariate ordinal data generation given marginal distributions and correlations. *Journal of Statistical Computation and Simulation*, 76(11), 1017-1025.
- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions are Violated. *Methodology*, 1(1), 27-38.
- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42, (3), 237-288.
- Green, P.E. & Rao, V.R. (1970). Rating Scales and Information Recovery: How Many Scales and Response Categories to Use? *Journal of Marketing*, 34, 33-39.
- Hair Jr., J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hill, M.M., & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Hummel, T.J., & Sligo, J.R. (1971). Empirical comparison of univariate and multivariate analysis. *Psychological Bulletin*, 76, 49-57.
- Jobson, J.D. (1992). *Applied Multivariate Data Analysis*. New York: Springer.
- Jacoby, J. & Mattel, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, VIII, 495-500.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217-1218.
- Johnson, D.E. (1998). *Applied Multivariate Methods for Data Analysts*. Pacific Grove, CA: Brooks/Cole Publishing.
- Johnson, R.A., & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Kaiser, S., Träger, D., Leisch, F. (2011). Generating Correlated Ordinal Random Values. *Technical Report Number*, 94, Munich: Department of Statistics, University of Munich.
- Katz B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Krosnick, J.A. & Fabrigar, L.R. (1997). Designing rating scales for effective measurement in surveys. In L.E. Lyberg, P. Biemer, M. Collins and E.D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin. *Survey measurement and process quality*, Hoboken, NJ. Wiley, 14–164.
- Krosnick, J.A. & Presser, S. (2010). Question and Questionnaire Design. In P.V. Marsden & J.D. Wright, *Handbook of Survey Research*, Vol. 2. Bingley, U.K. Emerald Group Publishing, Ltd., 263-314.

- Land, K.C., & Hui, Z. (2010). Sample Size, Optimum Allocation, and Power Analysis. In P.V. Marsden & J.D. Wright, *Handbook of Survey Research*. Emerald Group Publishing, Ltd. Bingley, 199-220.
- Laureano, R.M.S. (2011). *Testes de Hipóteses com o SPSS. O meu manual de consulta rápida*. Lisboa: Edições Sílabo.
- Lehmann, E.L. (1975). *Nonparametrics, statistical methods based on ranks*. San Francisco: Holden-Day.
- Leisch, F., Weingessel, A., & Hornik, K. (1998). On the generation of correlated artificial binary data. *Technical Report 13*, SFB Adaptive Information Systems and Modelling in Economics and Management Science, Wirtschaftsuniversität, Vienna-Austria, 2-6, A-1090 Wien, Austria.
- Loureiro, L.M.J., & Gameiro, M.G.H. (2011). Interpretação crítica dos resultados estatísticos: para lá da significância estatística. *Revista de Enfermagem Referência*, 3, 151-162.
- Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *European Journal of Research Methods for the Behavioral and Social Sciences*, 4 (2), 73-79.
- Manly, B.F.J. (2005). *Multivariate Statistical Methods*, (3rd Ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Marôco, J. (2010). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Montgomery, D.C. (2001). *Design and Analysis of Experiments*, (5th Ed.). New York: John Wiley & Sons, Inc.
- Moors, G. (2007). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42(6), 779-794.
- Myers, J.L., Well, A.D., & Lorch Jr., R.F. (2010). *Research Design and Statistical Analysis* (3rd Ed.). New York: Routledge.
- Nath, R. (1982). An investigation of the rank transform in the multivariate one-sample location problem. *Journal of Statistical Computation and Simulation*, 16, 139-155.
- Nath, R., & Pavur, R. (1985). A new statistic in the One-way multivariate analysis of variance. *Computational Statistics and Data Analysis*, 2, 297-315.
- Pearson, E.S. (1930). A further development of tests of normality. *Biometrika*, 22, 239-249.
- Pearson, E.S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- Presser, S. & Schuman H. (1980). The Measurement of a Middle Position in Attitude Surveys. *Public Opinion Quarterly*, 44(1), 70-85.
- Puri, M.L., & Sen, P.K. (1971). *Nonparametric methods in multivariate analysis*. New York: Wiley.
- Puri, M.L., & Sen, P.K. (1969). Class of Rank Order Tests for a General Linear Hypothesis. *Annals of Mathematical Statistics*, 40(4), 1325-1343.
- Ramsey, P.H. (1982). Empirical power of procedure for comparing two groups on p variables. *Journal of Educational Statistics*, 7, 139-156.
- Ramsey, P.H., & Ramsey, P.P. (2012). Pairwise testing of group mean vectors in MANOVA with small samples. *Journal of Statistical Computation and Simulation*, 83(6), 1037-1057.
- Ramsey, P.H., Ramsey, P.P., Hachimine, P., & Andiloro, N. (2011). Robustness, Power and Interpretability of Pairwise Tests of Discriminant Functions in MANOVA. *Journal of Modern Applied Statistical Methods*, 10(2), 403-417.
- Reis, E. (2001). *Estatística Multivariada Aplicada* (2^a ed.). Lisboa: Edições Sílabo.

- Revilla, M.A., Saris, W.E., & Krosnick, J.A. (2014). Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*, 43(1), 73-97.
- R Development Core Team (2012). *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (<http://www.R-project.org>).
- Sheskin, D.J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures* (4th Ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. New York: Pearson Education Inc.
- Temme, D. (2006). Assessing measurement invariance of ordinal indicators in cross-national research. In S. Diehl & R. Terlutter (Eds.), *International advertising and communication* (pp. 452-472). Wiesbaden, Germany: DUV.
- Timm, N.H. (2002). *Applied Multivariate Analysis*. New York: Springer-Verlag.
- Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistic and Data Analysis*, 54, 37-48.
- Todorov, V., & Filzmoser, P. (2009). An Object-Oriented Framework for Robust Multivariate Analysis. *Journal of Statistical Software*, 32, 1-47.
- Urdan, T.C. (2005). *Statistics in Plain English*. London: Laurence Erlbaum Associates.
- Van Aelst, S., & Willems, G. (2011). Robust and Efficient One-Way MANOVA Tests. *Journal of the American Statistical Association*, 106(494), 706-718.
- Yoriyaz, H. (2009). Método de Monte Carlo: princípios e aplicações em Física Médica. *Revista Brasileira de Física Médica*, 3 (1), 141-149.
- Zimmerman, D.W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127, 354-364.
- Zwick, R. (1985). Nonparametric One-Way Multivariate Analysis of Variance: A Computational Approach Based on the Pillai-Bartlett Trace. *Psychological Bulletin*, 47(1), 148-152.

PARTE IV. Aplicações Práticas

Introdução

Os padrões de saúde da criança e do jovem são influenciados pelo progresso na área da saúde e pelas tendências sociais, sendo a promoção da saúde e a prevenção da doença e da incapacidade essenciais para reduzir as diferenças no *status* de saúde, assegurar a igualdade de oportunidades e os recursos a fim de capacitar a criança e o jovem para alcançar o seu potencial máximo de desenvolvimento e saúde (Hockenberry & Barrera, 2014; WHO, 2007).

A investigação na área da saúde da criança e do jovem tem funcionado como base para a melhoria da qualidade de cuidados, exigindo dos profissionais de saúde conhecimentos específicos na área pediátrica e metodológica, com incidência nas técnicas de observação, recolha, análise de dados e implementação dos resultados da pesquisa nas práticas de cuidados.

Os profissionais de saúde são observadores privilegiados das respostas humanas à saúde e à doença, podendo contribuir para o desenvolvimento e bem-estar dos indivíduos especificamente numa área tão singular como a pediatria. Atualmente a ênfase está nos cuidados centrados no desenvolvimento e nos resultados mensuráveis que permitam determinar a eficácia das intervenções em saúde infantil e juvenil (Hockenberry & Barrera, 2014). Neste contexto realça-se a importância atribuída às práticas baseadas na evidência e à aplicação dos resultados da investigação nas intervenções em saúde e ao aperfeiçoamento de medidas de avaliação, seja ela mais ou menos subjetiva. A observação através de instrumentos de mensuração padronizados, como as escalas de avaliação, é essencial no dia-dia dos profissionais de saúde pois facilita a apreciação da saúde da criança (e.g. desenvolvimento, risco de lesão da pele, dor, independência funcional, bem-estar) e a observação sistematizada.

Os cuidados à criança estão suportados por modelos teóricos como o modelo do universo dos cuidados para o desenvolvimento, o qual foca a sua avaliação nos cuidados centrados na família, no ambiente seguro, sono e repouso, na avaliação da dor e do stress e nas atividades de vida diária promotoras do desenvolvimento harmonioso da criança e jovem. O modelo dos cuidados centrados na família ao colocar a família no centro dos cuidados exige que os profissionais a valorizem como objeto de cuidados e façam a avaliação do seu nível emocional, bem-estar e competências para cuidar da

criança (Kaaresen et al., 2008; Doucette & Pinelli, 2004). Neste modelo, a família deve ter acesso aos serviços de saúde, sem qualquer tipo de restrições e com a possibilidade de cuidar da sua criança desde o nascimento e em todos os contextos, sejam eles hospitalares ou comunitários, promovendo a segurança da criança (Nibert & Ondrejka, 2005; Jonhson, Abraham & Parrish, 2004). Os cuidados promotores da segurança da criança passam por um ambiente seguro (luz, ruído) (Wachman & Lahav, 2010, 2011; White, 2010, 2007) adequado à sua situação de saúde e por um sono tranquilo e protetor em que as medidas como o método de canguru, o aconchego, o embalo, o apoio na escola e com os pares, são fundamentais ao bem-estar da criança (Ludington-Hoe, Hosseini & Torowicks, 2005) e, como tal, funcionam como elementos favoráveis ao seu crescimento e desenvolvimento. Este tipo de cuidados inclui a avaliação da dor como outra medida fundamental, pelo facto de esta interferir com todos os sistemas fisiológicos e comportamentais. A dor pode provocar danos irreversíveis na saúde da criança pelo que a sua avaliação deve ser valorizada desde o período neonatal até à adolescência (Anand & Hall, 2008, 2006; Stevens et al., 2005).

A otimização de cuidados que facilitem o desenvolvimento neuro-motor, a estabilidade fisiológica e o conforto estão intrinsecamente ligados aos aspetos focados anteriormente e também às atividades de vida diária. Dentro destas, salientamos, no caso do recém-nascido, o posicionamento (manter a criança contida e alinhada e com os membros em flexão) (Vaivre-Douret & Golse, 2007; Vaivre-Douret et al., 2004; Chizawsky & Scott-Findlay, 2005); a alimentação oral (avaliação das competências oro-motoras e intervenção na alimentação oral) (Figueiredo et al., 2007a, 2007b; Thoyre, Shaker & Pridham, 2007, 2005; Cunha, 2006; Pinelli e Symington, 2005); os cuidados à pele (avaliar a integridade do sistema tegumentar e o risco de lesão da pele) (Curley et al., 2003; Lund et al., 2001; Huffines & Lodgson, 1997); e no caso da criança e jovem com espinha bífida a independência funcional nas suas atividades de vida diária e, consequentemente, na sua autonomia (Caseiro, Gonçalves & Malheiro, 2013; Malheiro, 2005, Laíns, 1991). São estes aspetos, de entre muitos outros, que exigem dos profissionais de saúde medidas de avaliação cada vez mais rigorosas, estruturadas e validadas às respetivas populações no sentido da melhoria efetiva da qualidade de cuidados à criança e ao jovem, o que se vai traduzir em ganhos em saúde.

Capítulo VIII. Escalas de Avaliação da criança e jovem

Neste capítulo vamos fazer aplicações práticas com três escalas de avaliação usadas para fazer a observação da criança, dando assim resposta ao terceiro objetivo deste trabalho. As escalas selecionadas foram a Early Feeding Skills (EFS), com um formato de medida com 3 pontos; a Neonatal Skin Risk Assessment Scale (NSRAS), com um formato de medida com 4 pontos; e a Functional Independence Measure (FIM), com formatos de medida com 5 e 7 pontos. A opção pela EFS, NSRAS e FIM fundamenta-se no facto de estas apresentarem alguns dos formatos de medida (número de pontos) mais utilizados no desenvolvimento de escalas de avaliação especificamente na área da saúde da criança e jovem (Curado, Teles & Marôco, 2013, 2014; Pereira da Silva & Justo da Silva, 2010) e, também porque foram os formatos de medida utilizados no estudo de simulação (capítulo VII).

A EFS com um formato de medida com 3 pontos avalia o desenvolvimento motor oral do recém-nascido e os comportamentos de regulação da criança em três domínios fundamentais: capacidade para organizar o funcionamento motor oral, capacidade para coordenar a deglutição e capacidade para manter a estabilidade fisiológica, podendo ser utilizada até às 52 semanas de vida (Thoyre, Shaker & Pridham, 2007, 2005). A seleção desta escala está implicitamente ligada ao facto de a alimentação ser um dos problemas mais comuns na população neonatal e especificamente nos recém-nascidos prematuros.

De um modo geral, depois do quadro respiratório estabilizado, a prioridade para os recém-nascidos pré-termo passa a ser o aspeto nutricional. (...) que favoreça o seu crescimento e desenvolvimento, assim como a redução dos déficits motores orais, (...) (Figueiredo et al., 2007b, p. 3).

A Neonatal Skin Risk Assessment Scale (NSRAS) (Huffines & Lodgson, 1997), com um formato de medida com 4 pontos, avalia o risco de úlceras de pressão, no recém-nascido, nos primeiros 28 dias de vida. Esta escala avalia a condição física geral, o estado mental, a mobilidade, a atividade, a nutrição e a humidade da pele. A utilização da NSRAS tem subjacente o quanto é importante a avaliação da pele e a manutenção da sua integridade. Manter a integridade cutânea, por um lado é indicador da homeostasia da criança e da qualidade de cuidados prestados, e por outro lado, dá visibilidade à

importância da prevenção de úlceras de pressão. Estas são consideradas pela Organização Mundial de Saúde (OMS, 2007) e pela Direção Geral de Saúde (DGS, 2011) como um grave problema de saúde pública e como um fator causal, relevante, de dor e sofrimento. No caso do recém-nascido a observação da pele é fundamental para fazer a prevenção das úlceras de pressão, pois para além de evitar a dor e o sofrimento pode também evitar a morte em recém-nascidos de risco (EPUAP-NPUAP, 2009).

A NSRAS surge na literatura como um dos melhores instrumentos para avaliar o risco de lesão da pele, pelo que foi feita a sua adaptação para a língua portuguesa e está em fase de finalização a colheita de dados para fazer a sua validação estatística para a população neonatal.

The Functional Independence Measure (FIM) (Sperle et al., 1997; Keith et al., 1984, 1987; Granger et al., 1986) apresenta-se na versão portuguesa como Medida de Independência Funcional (MIF) (Laíns, 1991) e foi selecionada por ser um instrumento que permite fazer o acompanhamento de pessoas em processo de reabilitação, e porque não é focado na capacidade de realizar tarefas, mas sim na realização efetiva e de forma independente das atividades de vida diárias. A MIF permite avaliar o desempenho da criança e jovem na realização de um conjunto de dezoito tarefas incluídas em dois domínios, o motor e o cognitivo, e com um formato de medida com 7 pontos.

Malheiro, Gaspar e Barros (2013) usaram os dados do domínio motor da MIF para comparar com os dados de outra escala de avaliação e, assim, fazerem a sua validação. Para tal, ajustaram o formato de medida de 7 pontos a 5 pontos, pelo que neste estudo iremos usar os dois formatos de medida.

Para fazer a análise dos resultados que emergem destas escalas foram realizadas quatro aplicações práticas recorrendo a dados fornecidos por três grupos de investigadores com projetos de investigação a nível nacional cujos sujeitos empíricos são recém-nascidos e crianças e jovens com espinha bífida.

Este estudo teve como objetivo “enquadrar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados reais provenientes de escalas de avaliação, com quatro formatos de medida (3, 4, 5 e 7 pontos) nos resultados do estudo de simulação estatística” (Capítulo VII).

8.1 Metodologia

Os resultados da investigação são o indicador que permite ao profissional reconhecer os instrumentos de observação mais adequados, quer teoricamente, quer metodologicamente, determinando a viabilidade da sua aplicação nas práticas de cuidados e na investigação. Seja qual for a área da saúde, a idade do cliente ou a sua condição clínica, a tomada de decisão na escolha dos instrumentos de observação deve ser centrada no cliente, baseada na teoria e nos métodos e procedimentos utilizados no seu desenvolvimento, identificando pontos fortes e fracos na análise estatística e na leitura dos resultados que emergem da sua aplicação prática.

Para enquadrar os resultados da utilização das 3 escalas selecionadas, nos resultados do estudo de simulação (Capítulo VII), criaram-se cenários metodológicos semelhantes no que concerne ao número de variáveis, ao número de pontos do formato de medida, às distribuições de frequência por cada um dos 3 grupos, à magnitude das correlações e às dimensões das amostras. Foi identificada a forma de distribuição de frequência (simétrica, assimétrica positiva, assimétrica negativa e uniformemente distribuída na escala) e feita a sua representação gráfica por grupo e também a identificação da intensidade das correlações entre itens e em seguida selecionados os três itens (variáveis) que cumpriam os pressupostos do delineamento do estudo de simulação.

Não fazendo sentido avaliar os pressupostos da MANOVA (distribuição normal multivariada e a homogeneidade das matrizes de variância-covariância), por se tratar de variáveis ordinais, determinou-se o coeficiente de assimetria ou skewness (g_1) e o coeficiente de achatamento ou Kurtosis (g_2) e a mediana (Me), fundamentando-se que uma distribuição se considera simétrica se o coeficiente de assimetria for inferior ao erro padrão de assimetria ($|g_1| < 2 \times SE$) (Field, 2009).

Para a análise dos dados procedentes da aplicação da EFS a variável independente foi a “idade gestacional” organizada em 3 classes, das 24-27, das 28-32 e das 33-37 semanas de gestação; para a análise dos dados provenientes da aplicação da NSRAS foi considerada a variável “peso” organizada em 3 classes, inferior ou igual a 1500g, entre 1501 e 2500 e superior a 2500g; e para a análise dos dados provenientes da aplicação da MIF motora modificada (formato de medida com 5 pontos), a variável

independente foi o nível da lesão medular, organizado em três níveis, lesão ao nível da região sacrada, da região lombar e da região dorsal da coluna vertebral. Para a MIF motora (formato de medida com 7 pontos) a variável independente foi a idade das crianças e jovens com espinha bífida, organizada em 3 classes etárias, dos 8-12, dos 13-15 e dos 16-23 anos.

Os testes estatísticos aplicados foram as 4 estatísticas da Manova paramétrica: lambda de Wilks, traço de Pillai, traço de Hotelling, maior raiz de Roy e 2 da MANOVA não paramétrica, lambda de Wilks e traço de Pillai aplicados às ordens, para averiguar se existem diferenças significativas entre grupos.

No caso da MANOVA paramétrica e depois de identificados os efeitos estatisticamente significativos, poderá prosseguir-se com ANOVAs para cada uma das variáveis dependentes, seguida do teste de Post-hoc HSD de Tukey ou de uma análise discriminante, de modo a identificar em que variáveis e grupos estão as diferenças.

Para a MANOVA não paramétrica pode recorrer-se ao teste de Kruskal-Wallis, seguido das comparações múltiplas de médias das ordens, com recurso ao teste *LSD* de Fisher para as ordens (apropriado para 4 ou menos grupos) para identificar em que variáveis e grupos estão as diferenças.

Os resultados enfatizam a potência dos testes, a dimensão da amostra, a *p-value* e como medida de dimensão do efeito, o eta quadrado (η_p^2). O eta quadrado representa a fração da variabilidade total da variável dependente que é explicada pelo fator *p* sob estudo (Marôco, 2010) e pode classificar-se como baixa (≤ 0.05), média ($]0.05; 0.25]$), elevada ($]0.25; 0.50]$) e muito elevada (>0.5) (Marôco, 2010; Cohen, 1998). Quanto mais elevado o valor da magnitude do efeito maior será a manifestação do fenómeno na população (Loureiro & Gameiro, 2011).

Para testarmos a igualdade das distribuições em todas as variáveis e grupos usamos a MANOVA não paramétrica e as hipóteses: $H_0 : F(\mathbf{X}_1) = F(\mathbf{X}_2) = \dots = F(\mathbf{X}_g)$ vs. $H_1 : \exists l, l^* : F(\mathbf{X}_l) \neq F(\mathbf{X}_{l^*})$, com $l, l^* \in \{1, \dots, g\}$ e onde $F(\mathbf{X}_l)$ é a função de distribuição multivariada de $\mathbf{X}_l = (X_{l1}, X_{l2}, \dots, X_{lp})'$, com $l = 1, \dots, g$.

Os testes da MANOVA não paramétrica não estão implementados no IBM-SPSS (v.21, IBM Corporation, 1989, 2012), assim, para o traço de Pillai aplicado às

ordens usamos a fórmula $\chi^2 = (n-1)TP_R$, onde n é a dimensão da amostra global, e TP_R é o traço de Pillai calculado a partir das ordens das observações (Marôco, 2010; Zwick, 1985; Katz & McSweeney, 1980); e para o lambda de Wilks aplicado às ordens usamos a fórmula $\chi^2 = -(n-1-(p+g)/2)\ln LW_R$, em que n é a dimensão da amostra, p é o número de variáveis e g o número de grupos e $\ln LW_R$, o logaritmo natural do lambda de Wilks aplicado às ordens (Van Aelst & Willems, 2011; Nath & Pavur, 1985).

Em qualquer um destes testes, a decisão, ao nível de significância alfa, consiste em rejeitar H_0 se $\chi^2 \geq \chi^2_{1-\alpha;p(g-1)}$ ou se $p\text{-value} \leq \alpha$, sendo p o número de variáveis e g o número de grupos.

O $p\text{-value}$ das duas estatísticas foi calculado no IBM-SPSS *Statistics* usando a função CDF.CHISQ, $p\text{-value} = P[\chi^2 \geq x^2]$. A função CDF.CHISQ(x^2, gl) que devolve a probabilidade da distribuição χ^2 ser inferior a x^2 . Como o $p\text{-value}$ é a probabilidade de se observar o valor da estatística de teste ou um valor superior, usamos a probabilidade complementar, $1 - \text{CDF.CHISQ}(x^2, gl)$ (Marôco, 2010, p. 367).

8.2 Avaliação das Competências na Alimentação Oral do Neonato

O desenvolvimento das capacidades do recém-nascido pré-termo é influenciado por vários fatores, sendo um dos mais importantes a idade gestacional. A imaturidade subjacente a uma baixa idade gestacional reflete-se de forma significativa na coordenação motora, sendo esta influenciada pelo peso, capacidade corporal, crescimento, maturidade do sistema nervoso e meio ambiente. Apesar da evolução em cuidados neonatais ter permitido uma maior sobrevivência aos prematuros, ainda não está resolvido o problema da morbidade que lhe está associada. Muitas destas crianças apresentam problemas de saúde que se refletem no seu comportamento e no seu desenvolvimento, sendo um dos problemas mais importantes a dificuldade na alimentação oral (Fucile et al., 2011).

Castro et al. (2007) avaliaram a associação entre a idade gestacional, do recém-nascido pré-termo, o desenvolvimento motor global e os sinais precoces de alteração do desenvolvimento do sistema sensorio motor oral. Os resultados do seu estudo mostraram que a idade gestacional influencia o desenvolvimento motor global e, mais especificamente, o motor oral. Os autores consideram que, quanto menor a idade gestacional, maior a probabilidade do recém-nascido apresentar problemas oro-motores e conseqüentemente dificuldades na alimentação oral.

A alimentação oral é um dos problemas com que se confrontam os profissionais de saúde e famílias que têm criança prematuras. A preparação para a alimentação destas crianças exige estabilidade fisiológica, neurológica e comportamental, posicionamento correto, ambiente seguro, escolha do momento adequado para iniciar a alimentação e respectivas medidas de suporte, especificamente sensoriomotoras.

Sensorimotor intervention is used to improve oral feeding. It is described as the provision of developmentally appropriate sensory inputs including oral, tactile, kinaesthetic, vestibular, and auditory to facilitate the development of existing rudimentary skills and to prevent/minimize the deleterious environmental effects.
(Fucile et al., 2011, p. 829)

A introdução da alimentação oral nos prematuros tem tido como suporte na decisão as variáveis peso, idade gestacional, estabilidade fisiológica e estado de saúde. Apesar de esta apreciação ter sido uma mais-valia ao longo de muitos anos, atualmente é considerada ineficiente, para fazer uma avaliação global da situação alimentar da criança, por se basear apenas na experiência e na observação clínica. Os profissionais de saúde ao fazerem a avaliação da criança devem, por um lado, fundamentar as suas decisões na evidência científica disponível em neonatologia, no reconhecimento da preferência da família da criança, no contexto de cuidados e na perícia clínica, promovendo a prática baseada na evidência; por outro lado, devem orientar a sua intervenção numa prática de cuidados de apoio ao desenvolvimento, usando a informação fisiológica associada à informação comportamental para melhor compreender os “estados” da criança e assim cuidar efetivamente tendo como suporte o “cuidar para desenvolvimento”. Os estados comportamentais são individualizados e formados pela experiência, maturação, ritmos circadianos e genética, sendo que:

A organização dos estados tem sido descrita como um mecanismo protetor da hiperestimulação do córtex cerebral que coordena os sistemas corticais sensoriais de atenção e execução. (Askin & Wilson, 2014, p. 353).

Para alguns autores, os estados comportamentais podem ser organizados em padrões de sono e atividade: sono profundo, sono ativo, sonolência, alerta calmo, alerta ativo e choro (Wheeler, 2014). Os recém-nascidos pré-termo despendem mais de 70% do seu tempo no nível de sono ativo, sendo a sua maturação relacionada com o decréscimo do sono ativo e o aumento de sono calmo (Holditch-Davis & Blackburn, 2007; Holditch-Davis, 1998), pelo que é fundamental fazer a sua avaliação e promover o repouso. Thoyre (2007) considera que, associada a esta apreciação deve avaliar-se o desenvolvimento oro-motor, pois a partir do momento em que as estruturas motoras envolvidas na alimentação estão ativas e funcionais, o mais importante é manter a criança em estado de alerta calmo e com boa tonicidade para manter a coordenação entre a sucção, a deglutição e a respiração. Importa também salientar que no caso do recém-nascido pré-termo, os estados comportamentais estão alterados devido à imaturidade da criança e, por isso, as suas respostas acontecem, normalmente, pela via das alterações fisiológicas e não comportamentais, sendo fundamental a observação dessas alterações fisiológicas. Se não se verificarem alterações fisiológicas durante a alimentação, os cuidados devem estar centrados nos estados comportamentais, tais como, não deixar adormecer a criança, estimulando-a suavemente para evitar a depressão do sistema respiratório de forma a facilitar o processo de alimentação. Como a regulação respiratória durante o alerta calmo estabiliza com o aumento da idade gestacional, a criança pré-termo tem mais problemas quer esteja em alerta calmo, quer sonolenta (Thoyre, 2007; Holditch-Davis, 1998), o que se manifesta na coordenação dos reflexos orais com a respiração. Pode-se assim dizer que, a alimentação envolve aspetos gerais, relacionados com as questões nutricionais e outros mais específicos, referentes ao desenvolvimento, e especificamente ao desenvolvimento oro-motor e à maturação neuro-comportamental (Fucile et al., 2012).

O desenvolvimento oro-motor, a maturidade neuro-comportamental e do sistema gastrointestinal, que incluem a presença de reflexos orais e a sua coordenação, são essenciais em todo o processo de alimentação.

Os reflexos orais como a sucção e deglutição devidamente coordenados com a respiração permitem uma alimentação oral segura, sobretudo se associados: ao reflexo de engasgamento (GAG) e ao reflexo de morder e tossir, como reflexos protetores; ao reflexo dos quatro pontos cardiais e palmares, que facilitam a estimulação da criança; ao reflexo de preensão e busca, em que a criança vira a cabeça, abre a boca quando a tetina ou a mama toca na bochecha ou nos lábios, colocação da língua para fora da boca e para baixo, para receber a tetina ou o mamilo para iniciar a alimentação e também a funcionalidade do músculo transverso da língua. O envolvimento destas estruturas na alimentação são a base da capacitação da criança para se alimentar e são essenciais para que se efetive a sucção nutritiva rítmica e sem riscos para a criança.

Como temos vindo a afirmar, a idade gestacional influencia a maturação do sistema gastrointestinal, pelo que o recém-nascido com idade gestacional inferior a 26 semanas apresenta um sistema imaturo, o que faz com que seja incapaz de se alimentar por tetina. Mas, apesar da imaturidade do sistema gastrointestinal, a criança apresenta os recetores de odores e sabores funcionais, os quais podem ser usados para estimular os sentidos e fazer a sua preparação para a alimentação oral, através do contacto com o leite materno. Semana a semana pode observar-se a evolução do desenvolvimento e comportamento destas crianças. Entre a 26^a e a 28^a semanas a criança pode tentar a sucção se a boca, a língua ou o palato forem estimulados (através de chupeta ou dedo), iniciando em seguida a sucção não nutritiva, pois a sua imaturidade e desorganização ao nível da sucção, deglutição e respiração não permite a alimentação por tetina e/ou mama. Entre a 28^a e a 30^a semana podem estar presentes, alguns reflexos arcaicos fundamentais para a alimentação, embora nem sempre estejam funcionais.

O reflexo dos pontos cardeais pode estar presente, embora possa ser imperfeito e lento, e os reflexos de sucção e deglutição podem estar presentes, mas imaturos e descoordenados entre si, descoordenadas com a respiração, colocando em risco uma alimentação oral segura. Entre a 30^a e a 33^a semana o reflexo de sucção está mais maturo, o que permite observar uma sucção rítmica, embora à 32^a semana algumas crianças ainda possam apresentar movimentos descoordenados da língua, não estando este movimento sincronizado nem coordenado com a respiração (Askin & Wilson, 2014). Entre a 33^a e a 37^a semana a sucção inicial pode não ser acompanhada pela

deglutição e as contrações esofágicas podem estar descoordenadas, o que pode levar a criança a estados de exaustão (Askin & Wilson, 2014, p. 353). Apesar de se verificar alguma inconsistência na coordenação destes reflexos, algumas crianças prematuras conseguem iniciar a alimentação oral nesta fase.

A avaliação dos reflexos de sucção e deglutição é fundamental por serem dois reflexos essenciais ao processo de alimentação. A sucção apresenta-se como uma atividade bifásica composta pela contração rítmica da língua associada ao movimento do osso hioide, da mandíbula e do lábio inferior. Esta atividade inicia-se com a estimulação do palato (através da tetina ou mamilo), a qual desencadeia uma ação rítmica da língua (onde o transversa da língua tem uma função primordial) e do maxilar que facilita a retirada de leite da mama/tetina, sendo a retirada do leite auxiliada pela pressão negativa intraoral, provocada pelo rebaixamento da língua e da mandíbula quando a boca do recém-nascido se enche com leite (Fucile et al., 2005; Lau et al., 2003, 2000).

A deglutição apresenta-se como uma atividade motora coordenada e complexa caracterizada por três níveis, o oral, o faríngeo e o esofágico. O nível oral permite o transporte do leite da parte posterior da cavidade oral para a faringe (consequência do reflexo de sucção); o faríngeo permite o transporte do leite para a faringe e o esofágico permite o transporte do leite do esfíncter esofágico superior ao inferior e de seguida para o estômago através dos movimentos peristálticos (Kenner & McGrath, 2004).

Entre a 36^a/37^a e a 40^a semana a criança apresenta uma melhor organização comportamental, está mais capacitada para se manter alerta e os reflexos oro-motores estão mais maduros, estando os reflexos de sucção e deglutição presentes e coordenados com a respiração (Hadley et al., 1999). Outro aspeto importante da observação é a avaliação dos surtos de sucção (conjunto de sucções seguidas entre cada respiração), pois estes são um bom indicador para avaliar a maturidade da criança para iniciar a alimentação oral. Os surtos de sucção variam segundo a idade gestacional, podendo apresentar 3 a 5 sucções por surto na sucção imatura; 6 a 9 sucções por surto, na sucção transitória e 10 a 30 sucções por surto na sucção madura. A importância da avaliação deste amadurecimento estrutural suporta uma alimentação oral segura e eficiente, minimiza os gastos de energia e facilita o crescimento e desenvolvimento da criança.

A avaliação das competências do recém-nascido prematuro para iniciar a alimentação oral depende da sua observação rigorosa com instrumentos de medida adaptados à população neonatal e validados estatisticamente.

Na literatura científica encontra-se uma grande variedade de instrumentos que avaliam as competências do recém-nascido pré-termo durante a alimentação oral, mas um dos mais utilizados, internacionalmente, na prática de cuidados e na investigação é a EFS, por ser uma escala que permite a intervenção ao longo de todo o processo de alimentação. A EFS (Thoyre, 2007, Thoyre, Shaker & Pridham, 2005; Shaker, 1990, 1992, 1997, 1999; Pridham, et al., 1994, 1998, 1999), adaptada para português como Escala de Observação das Competências Precoces na Alimentação Oral e validada estatisticamente (Curado et al., 2013), é muito usada nas unidades de neonatologia por permitir uma abordagem multidimensional das competências da criança para a alimentação oral facilitando a avaliação de características fisiológicas e neuro-comportamentais e também porque permite a intervenção (e.g., estimulação) da família ou dos profissionais durante a alimentação, o que vai ao encontro das necessidades da criança e da clínica.

8.2.1 Escala de Observação das Competências Precoces na Alimentação Oral

Os investigadores que desenvolveram a EFS basearam-se na observação da complexidade do processo de alimentação oral e na análise dos múltiplos sistemas fisiológicos e comportamentais envolvidos, observando também a interação entre esses sistemas em períodos problemáticos para o recém-nascido.

Thoyre (2003a, 2003b) fez uma extensão do conteúdo do instrumento para fazer sessões de ensino às famílias acerca de como poderiam fazer uma observação sistematizada do recém-nascido durante a alimentação, facilitando o envolvimento dos pais nas competências alimentares da criança e na sua preparação para a alta (Thoyre, 2007; Thoyre & Brown, 2004; Thoyre & Carlson, 2003a, 2003b). A escala permite observar o recém-nascido ao longo de todo o processo de alimentação. Imediatamente antes de iniciar a alimentação avalia-se a postura, o estado de alerta, a tonicidade e a linha basal de saturação de oxigénio (>93%). Cinco minutos após a alimentação,

observa-se o comportamento da criança e a tolerância alimentar. Durante a alimentação são avaliadas de forma sistematizada as competências na alimentação oral em três grandes áreas: capacidade para organizar o funcionamento oro-motor (COFO), capacidade para coordenar a deglutição (CCD) e capacidade para manter a estabilidade fisiológica (CMEF). A EFS adaptada para a população neonatal portuguesa é constituída por 22 itens com um formato de medida ordinal com 3 pontos (0, 1, 2). Estes itens estão distribuídos pelas três dimensões: COFO com 7 itens, CCD com 5 itens e CMEF com 10 itens. A dimensão COFO reflete a maturação das estruturas envolvidas na alimentação e a capacidade neurológica para as coordenar (Curado et al., 2013) (cf. Quadro 8.2. 1).

Quadro 8.2. 1 - Descritores da dimensão Capacidade em organizar o funcionamento Oro-motor

Capacidade para Organizar o Funcionamento Oro-Motor			
Item	Opções de resposta		
Abre a boca prontamente quando a tetina toca os lábios no início da mamada	2. Abre a boca prontamente todas as vezes	1. Abre a boca prontamente algumas vezes	0. Nunca abre a boca prontamente
A língua descai para receber a tetina do biberão no início da alimentação	2. A língua descai todas as vezes	1. A língua descai algumas vezes	0. A língua nunca descai
Imediatamente após a introdução da tetina, a sucção da criança está organizada, rítmica e suave. (A sucção organizada tem um padrão de uma sucção – pausa; na sucção desorganizada a criança trinca ou cerra o maxilar, retrai a língua ou empurra a tetina com a língua)	2. A sucção está sempre organizada	1. Algumas vezes a sucção está organizada	0. A sucção nunca está organizada
Assim que a alimentação está a decorrer, mantém um padrão de sucção suave e rítmico (ver descrição em cima)	2. Observa-se estabilidade e consistência	1. Alguma desorganização da sucção no decurso da mamada	0. Incapacidade em manter a sucção organizada
A pressão da sucção é constante e forte (i.e., chupa com vigor e a um ritmo constante)	2. Observa-se estabilidade e consistência	1. Algumas sucções débeis	0. Sucção frequentemente débil
Empenha-se em fazer longas sucções (7 a 10) sem apresentar sinais de stresse ou instabilidade fisiológica e/ou uma resposta cardiorrespiratória adversa ou negativa. (sinais de stress incluem o franzir das sobrancelhas, olhar preocupado, afastamento da tetina, etc.)	2. Faz sempre surtos longos de sucção sem sinais de stresse	1. Alguns surtos longos de sucção sem sinais de stresse	0. Sem surtos longos de sucção ou todos os surtos acompanhados de sinais de stresse
A língua mantém contacto constante com a tetina – não desliza da tetina – ao fazer a sucção ouve-se um som de um clique	2. Sem som de clique	1. Alguns sons de clique	0. Sons frequentes de clique

A dimensão CCD permite verificar a coordenação entre deglutição, sucção e respiração e se esta é feita de forma tranquila e segura permitindo a passagem de ar suficiente para manter a homeostasia fisiológica entre as deglutições (momento de apneia e as cordas vocais encerram até à sua conclusão (cf. Quadro 8.2.2).

Quadro 8.2.2 - Descritores da dimensão Capacidade em Coordenar a Deglutição

Capacidade para Coordenar a Deglutição			
Item	Opções de resposta		
Gere os fluidos durante a deglutição sem os perder pelos cantos da boca. (i.e. não se baba)	2. Sem perda de líquido	1. Alguma perda de líquido	0. Frequente perda de líquido
Os sons faríngeos são limpos – não se ouve o gorgolejo criado pelo líquido no nariz ou faringe	2. Sem sons de gorgolejo	1. Alguns sons de gorgolejo	0. Frequentes sons de gorgolejo
A deglutição é suave – sem esforço para engolir	2. Suaves deglutições	1. Algumas deglutições com dificuldade	0. Frequentes deglutições com dificuldade
Uma simples deglutição assimila o bólus da sucção – múltiplas deglutições não são necessárias para esvaziar os fluidos para a garganta	2. Deglutições únicas para o mesmo bólus	1. Algumas deglutições múltiplas para o mesmo bólus	0. Frequentes múltiplas deglutições para o mesmo bólus
Tosse ou sons de engasgamento	2. Nenhum evento observado	1. Pelo menos um evento observado	0. Dois ou mais eventos observados

A dimensão CMEF permite avaliar sinais fisiológicos como: a coloração da pele, a saturação de oxigénio, a frequência respiratória e cardíaca e também os surtos de sucção (conjunto de sucções seguidas entre cada respiração) (cf. Quadro 8.2.3).

Quadro 8.2.3 - Descritores da dimensão Capacidade em Manter a Estabilidade Fisiológica

Capacidade para Manter a Estabilidade Fisiológica			
Item	Opções de resposta		
Cerca de 30s depois de iniciar a mamada a saturação de O ₂ estabiliza e não há sinais de stresse.	2. Observa-se estabilidade e consistência	1. Começam a tornar-se visíveis os sinais de stresse	0. Não se observa
Interrompe a sucção para respirar. O prestador de cuidados não tem que interromper a mamada para a criança respirar.	2. Interrompe a sucção para respirar de forma consistente	1. Começam a emergir algumas capacidades: interrompe na maioria das vezes	0. Não interrompe por si para respirar
Quando interrompe a sucção para respirar, observam-se vários ciclos respiratórios completos	2. Faz ciclos respiratórios completos para se equilibrar	1. Faz ciclos respiratórios completos para se equilibrar na maioria das vezes	0. Não interrompe a sucção para respirar
A criança pára para respirar antes de começar a demonstrar sinais de stresse. (sinais de stress observados incluem: franzir das sobrancelhas, olhar preocupado, afastamento da tetina, etc.).	2. Pára para respirar antes de surgirem sinais de stresse	1. Frequentemente pára para respirar antes de surgirem sinais de stresse	0. Não pára para respirar antes de surgirem sinais de stresse
Sons respiratórios limpos – sem roncos (expiração prolongada, glote parcialmente encerrada na expiração).	2. Sem roncos	1. Roncos esporádicos	0. Roncos frequentes
Sons respiratórios limpos – sem sinais de estridor que sugerem uma passagem de ar restrita no canal.	2. Sem estridor	1. Estridor esporádico	0. Estridor frequente
19. Adejo nasal e/ou alteração da cor (palidez).	2. Sem adejo nasal	1. Adejo nasal esporádico	0. Adejo nasal frequente
20. Utilização dos músculos acessórios na respiração (elevação do queixo, cabeça para trás, retração e tiragem).	2. Nunca há utilização dos músculos acessórios	1. Utilização esporádica dos músculos acessórios	0. Utilização frequente dos músculos acessórios
21. Alteração da cor durante a alimentação (palidez e cianose peri-bocal e/ou periorbital).	2. Nunca há alteração da cor	1. Alteração esporádica da cor	0. Alteração frequente da cor
22. Descida da saturação de O ₂ abaixo dos 90%.	2. Nunca	1. Ocasional	0. Frequente

8.2.2 Procedimentos

Os dados utilizados para fazer as aplicações práticas foram cedidos pela equipa do projeto (Anexo I) que se encontra a fazer a tradução da EFS e a sua validação estatística. Das observações cedidas, pela equipa de investigação, que resultaram da observação de recém-nascidos com a EFS, foram selecionadas aleatoriamente 100 observações ($n_i=100$) de cada um dos grupos originados pela categorização da variável idade gestacional (24-27, 28-32 e 33-37 semanas de gestação) e depois selecionadas aleatoriamente as restantes amostras ($n_i=10, 20, 30, 60, 80$).

8.2.3 Apresentação dos Resultados

Os testes da MANOVA paramétrica e da MANOVA não paramétrica foram aplicados com 3 variáveis da escala EFS, 2 da dimensão COFO, itens 4 (Assim que a alimentação está a decorrer, mantém um padrão de sucção suave e rítmico) e 5 (A pressão da sucção é constante e forte, i.e., chupa com vigor e a um ritmo constante) e uma da dimensão CCD, item 11 (Uma simples deglutição assimila o “bólus” da sucção, múltiplas deglutições não são necessárias para esvaziar os fluidos para a garganta).

Para fazer o enquadramento dos dados resultantes da observação dos recém-nascidos nos resultados do estudo de simulação (Capítulo VII), verificou-se a localização da maior probabilidade das respostas aos itens em cada um dos três grupos. As observações efetuadas com a escala EFS apresentam distribuição simétrica no primeiro grupo e assimétrica negativa nos outros 2 grupos (p_1, p_3, p_3), com correlações de baixa intensidade ($r_{45}=0.355$, $r_{4,11}=0.282$, $r_{5,11}=0.226$), cujos resultados podem ser enquadrados no caso (p_2, p_2, p_1) do estudo de simulação (cf. Figura 8.2.1 e Tabela 8.2.2).

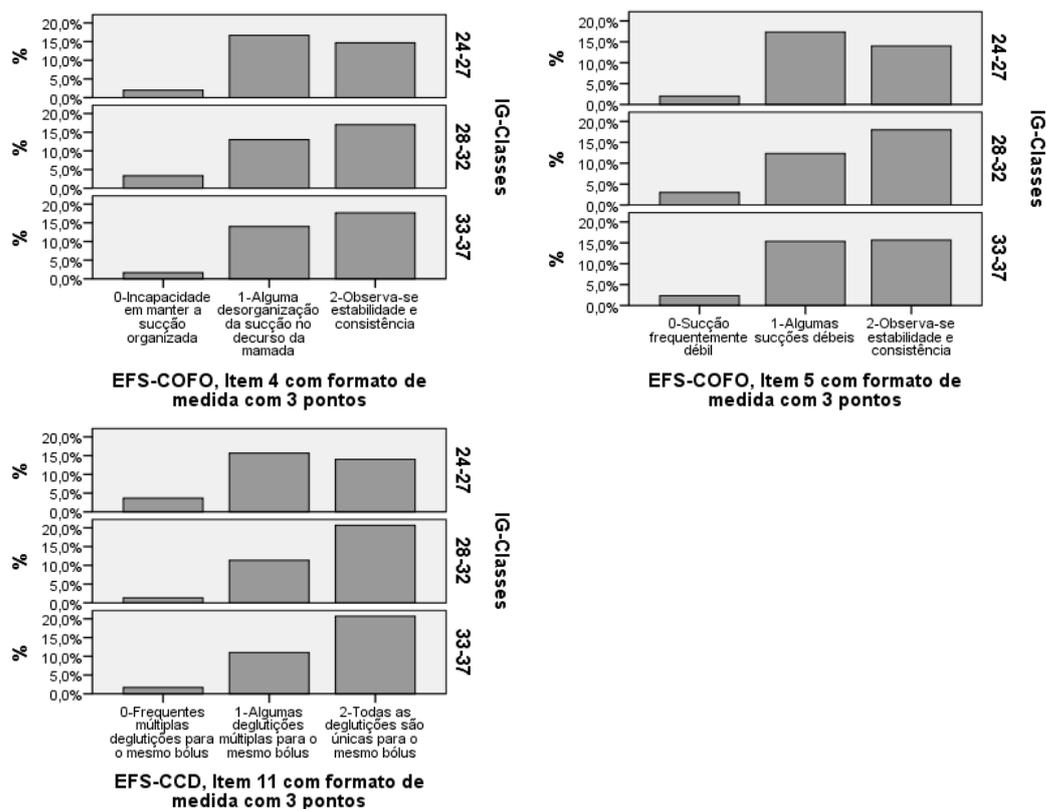


Figura 8.2.1 - Distribuições de frequência em cada um dos três grupos dos itens 4, 5 e 11 da EFS para $n_i=100$

Nas Tabela 8.2. 1 e Tabela 8.2. 2 apresentam-se os valores da mediana (Me), do coeficiente de assimetria (g_1) e coeficiente de achatamento (g_2) para os três itens (4, 5 e 11) da EFS por diferentes classes de idades gestacionais (IG-Classes).

Para $n_i=100$, os resultados da mediana encontram-se entre 1 e 2. Para o primeiro grupo, o coeficiente de assimetria apresenta resultados próximos de zero, em módulo, e inferiores ao dobro do erro padrão (SE) de assimetria (0.482). Os outros apresentam valores superiores ao dobro do erro padrão do coeficiente de assimetria, mas com sinal negativo (cf. Tabela 8.2.2). Deste modo o caso em estudo enquadra-se na distribuição simétrica para o primeiro grupo e assimétrica negativa para os outros dois grupos.

Tabela 8.2. 1 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2), nos itens 4, 5, e 11 da EFS para $n_i=10, 20$ e 30

EFS – Formato de medida com 3 pontos										
n_i		$n_i=10$			$n_i=20$			$n_i=30$		
Itens	IG-Classes	Me	g_1	g_2	Me	g_1	g_2	Me	g_1	g_2
Item 4	24-27	1.5	0.000	-2.571	1.0	-0.292	-0.734	1.0	0.294	-0.550
	28-32	1.0	-0.434	-0.283	1.0	-0.393	-0.570	1.0	-0.499	-0.781
	33-37	1.5	-0.780	-0.146	1.0	-0.212	-0.552	2.0	-0.758	-0.321
Item 5	24-27	1.0	0.484	-2.277	1.0	0.000	-0.279	1.0	0.294	-0.550
	28-32	1.0	0.484	-2.277	1.0	-0.393	-0.570	1.0	-0.586	-0.589
	33-37	2.0	-1.658	2.045	1.0	0.294	0.457	1.0	-0.517	-0.534
Item 11	24-27	1.0	-0.434	-0.283	1.0	-0.296	-1.399	1.0	0.294	-0.550
	28-32	1.5	0.000	-2.571	2.0	-0.218	-2.183	2.0	-0.591	0.620
	33-37	2.0	-1.658	2.045	1.5	-0.583	0.459	2.0	-1.172	0.431

Tabela 8.2. 2 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) com um erro padrão de assimetria de 0.241, nos itens 4, 5, e 11 da EFS para $n_i=40, 80$ e 100

EFS – Formato de medida com 3 pontos										
n_i		$n_i=40$			$n_i=80$			$n_i=100$ $SE_{g1}=0.241$		
Itens	IG-Classes	Me	g_1	g_2	Me	g_1	g_2	Me	g_1	g_2
Item 4	24-27	1.0	-0.171	-0.467	1.0	-0.388	-0.642	1.0	-0.384	-0.655
	28-32	2.0	-0.013	-0.275	2.0	-0.753	-0.577	2.0	-0.700	-0.572
	33-37	2.0	-1.240	0.504	1.5	-0.549	-0.601	2.0	-0.658	-0.503
Item 5	24-27	1.0	-0.099	-0.367	1.0	-0.150	-0.552	1.0	-0.319	-0.656
	28-32	2.0	-0.926	-0.408	2.0	-0.793	-0.472	2.0	-0.793	-0.432
	33-37	1.0	-0.604	-0.603	1.0	-0.392	-0.632	1.0	-0.520	-0.605
Item 11	24-27	1.0	-0.164	-0.786	1.0	-0.491	-0.729	1.0	-0.438	-0.727
	28-32	2.0	-1.339	0.567	2.0	-0.944	-0.071	2.0	-0.983	-0.009
	33-37	2.0	-1.657	2.018	2.0	-1.036	0.074	2.0	-1.028	-0.085

No estudo de simulação (Capítulo VII) com estas distribuições de frequência nos três grupos (p_2, p_2, p_1), os resultados são influenciados pela dimensão da amostra (amostras de menor dimensão apresentam potências mais baixas), pela intensidade da correlação entre os itens (quanto maior a intensidade de correlação entre itens menor a potência dos testes) e pelo tipo de testes, se são paramétricos ou não paramétricos (a maior raiz de Roy é o teste com potências mais elevadas seguida do lambda Wilks aplicado às ordens).

Os resultados das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais, com correlações de baixa intensidade entre os três itens da EFS (4, 5, 11) estão sumariados, por teste, nas Tabela 8.2.3, Tabela 8.2.4, Tabela

8.2.5, Tabela 8.2.6, Tabela 8.2.7. As diferentes estatísticas da MANOVA paramétrica mostram potências de teste superiores a 80% nas amostras com 120 e 300 observações. Para $n=240$ as potências oscilam entre 50% e 60%, para $n=60$ as potências estão entre 40% e 60% e as restantes iguais ou inferiores a 20%.

Tabela 8.2.3 - Traço de Pillai nas diferentes dimensões amostrais na escala EFS

EFS - traço de Pillai						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.117	0.540	(6,52)	0.775	0.059	0.198
60	0.128	1.279	(6,112)	0.273	0.064	0.485
90	0.039	0.577	(6,172)	0.749	0.020	0.227
120	0.166	3.508	(6,232)	0.002	0.083	0.946
240	0.038	1.540	(6,472)	0.163	0.019	0.596
300	0.049	2.474	(6,592)	0.023	0.024	0.834

Tabela 8.2.4 - Lambda de Wilks nas diferentes dimensões amostrais na escala EFS

EFS – lambda de Wilks						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.884	1.528	(6,50)	0.785	0.160	0.194
60	0.872	1.295	(6,110)	0.266	0.066	0.490
90	0.961	0.571	(6,170)	0.753	0.020	0.225
120	0.836	3.597	(6,230)	0.002	0.086	0.952
240	0.962	1.538	(6,470)	0.164	0.019	0.596
300	0.951	2.480	(6,590)	0.022	0.025	0.835

Tabela 8.2.5 - Traço de Hotelling nas diferentes dimensões amostrais na escala EFS

EFS – traço de Hotelling						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.129	0.515	(6,48)	0.794	0.060	0.189
60	0.145	1.309	(6,108)	0.259	0.068	0.494
90	0.040	0.565	(6,168)	0.758	0.020	0.222
120	0.194	3.685	(6,228)	0.002	0.088	0.956
240	0.039	1.537	(6,468)	0.164	0.019	0.595
300	0.051	2.486	(6,588)	0.022	0.025	0.836

Tabela 8.2.6 - Maior raiz de Roy nas diferentes dimensões amostrais na escala EFS

EFS – maior raiz de Roy						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power(π)
30	0.112	0.967	(3,26)	0.423	0.100	0.233
60	0.140	2.608	(3,56)	0.061	0.123	0.609
90	0.028	0.808	(3,86)	0.493	0.027	0.218
120	0.180	6.953	(3,116)	<0.001	0.152	0.976
240	0.031	2.462	(3,236)	0.063	0.030	0.607
300	0.043	4.218	(3,296)	0.006	0.041	0.856

Tabela 8.2.7 - Traço de Pillai e Lambda de Wilks aplicado às ordens na escala EFS

EFS – MANOVA não paramétrica								
n	Traço de Pillai NP				Lambda de Wilks NP			
	TP_{NP}	χ^2	df	p-value	LW_{NP}	χ^2	df	p-value
30	0.137	3.973	6	0.680	0.865	3.770	6	0.610
60	0.107	6.313	6	0.390	0.894	6.272	6	0.390
90	0.042	3.738	6	0.710	0.958	3.698	6	0.720
120	0.181	21.539	6	<0.001	0.821	22.852	6	<0.001
240	0.039	9.321	6	0.160	0.961	9.204	6	0.160
300	0.047	14.053	6	0.030	0.953	14.208	6	0.030

Os resultados mostram ainda que nas duas amostras com potências mais elevadas os resultados dos testes são estatisticamente significativos ($p\text{-value} < \alpha = 0.05$), pelo que se rejeita a hipótese H_0 de igualdade dos vetores de valores médios e conclui-se que nestes dois casos existe efeito da idade gestacional nos itens da escala EFS, ou seja há diferença significativa entre pelo menos 2 dos 3 grupos com diferentes classes de idade gestacional (Cf. Tabela 8.2.8).

Tabela 8.2.8 - p-value das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais na escala EFS

p-value - EFS						
MANOVA	n					
	30	60	90	120	240	300
TP	0.775	0.273	0.749	0.002	0.163	0.023
LW	0.785	0.266	0.753	0.002	0.164	0.022
TH	0.794	0.259	0.758	0.002	0.164	0.022
MRR	0.423	0.061	0.493	<0.001	0.063	0.006
TPNP	0.680	0.390	0.710	<0.001	0.160	0.030
LWNP	0.610	0.390	0.720	<0.001	0.160	0.030

8.2.4 Discussão dos Resultados

A discussão dos resultados foi organizada para dar resposta ao objetivo do estudo “Analisar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados ordinais provenientes da utilização da Escala de Observação das Competências Precoces na Alimentação Oral (3 itens das dimensões COFO e CCD) e fazer o seu enquadramento nos resultados do estudo de simulação de Monte Carlo”.

Os itens 4, 5 e 11 da EFS apresentam correlações de baixa magnitude, com valores da mediana enviesados para pontuações elevadas (Me entre 1 e 2) e coeficiente de assimetria (g_1) próximo de zero para o primeiro grupo e os outros dois grupos apresentam valores superiores ao dobro do erro padrão do coeficiente de assimetria com sinal negativo. No estudo de simulação com este tipo de distribuição (p_2, p_2, p_1), ou seja com distribuições assimétricas negativas e simétrica, a potência nas amostras de menor dimensão é influenciada pelas correlações entre os itens e pelo tipo de teste, sendo a maior raiz de Roy a estatística com potências mais elevadas, seguida do lambda de Wilks aplicado às ordens. Quando as correlações são de baixa magnitude e $n=30$ a potência da maior raiz de Roy é superior a 90% e do lambda de Wilks aplicado às ordens, superior a 80%, os restantes testes apresentam potências aceitáveis (superior a 69%) (Hill & Hill, 2009).

Relativamente aos resultados com dados reais da escala EFS e com correlações de baixa magnitude, verifica-se tal como no estudo de simulação, se a dimensão da amostra diminui a potência dos testes paramétricos diminui, embora com valores de potências mais baixos relativamente ao estudo de simulação. Com os dados da EFS apenas se registam potências elevadas na amostra com 120 e 300 observações. Para as amostras com dimensão de 60 e 240 observações as potências podem considerar-se aceitáveis (superiores a 50%) (Hill & Hill, 2009) e nas dimensões com 60 e 240 observações a magnitude das potências é mais baixa, o que não se verificou no estudo de simulação e pode estar relacionado com o facto de apenas termos 3 pontos na escala de medida.

No caso paramétrico a dimensão do efeito mantém-se entre valores baixos (≤ 0.05) e médios $]0.05; 0.25]$.

Quando se aplica a MANOVA paramétrica e depois de identificados os efeitos estatisticamente significativos, poderá prosseguir-se com ANOVAs para cada uma das variáveis dependentes, seguida do teste de Post-hoc HSD de Tukey ou efetuar uma análise discriminante. No entanto, perante o formato da escala de medida, a opção deveria ser pela MANOVA não paramétrica com recurso ao teste de Kruskal-Wallis, seguido das comparações múltiplas de médias das ordens, com recurso ao teste *LSD* de Fisher para as ordens para identificar em que variáveis e grupos estão as diferenças.

Embora sem termo de comparação com o estudo de simulação, no que concerne à potência dos testes e à magnitude dos efeitos da MANOVA não paramétrica aplicada aos dados reais da EFS, o comportamento dos *p-value* é semelhante entre os testes paramétricos e os não paramétricos; os testes não são estatisticamente significativos na maioria das amostras. Apenas na amostra com 120 e 300 observações os testes são estatisticamente significativos com $p\text{-value} < \alpha = 0.05$, quer nas estatísticas paramétricas quer nas não paramétricas.

8.2.5 Considerações Finais

As observações efetuadas com a escala EFS enquadraram-se em distribuições simétricas e assimétricas positivas, que no estudo de simulação originaram resultados de potência dos testes da MANOVA que são influenciados pela dimensão da amostra, pela intensidade da correlação entre os itens e pelo tipo de teste. A maior raiz de Roy é a estatística com valor de potência mais elevado seguida do lambda de Wilks aplicado às ordens, todos os outros teste da MANOVA paramétrica e traço de Pillai aplicado às ordens apresentam potências mais baixas (entre 10 e 20%) o que não foi muito evidente no exemplo com os dados reais da EFS pelo facto de a correlação, entre os itens, ser de baixa magnitude.

O enquadramento no estudo de simulação é razoável, pois embora o comportamento dos resultados das potências seja semelhante, a magnitude das potências é mais baixa do que no estudo de simulação e não existe uma uniformidade relativamente à dimensão da amostra. Em alguns dos testes as potências são mais elevadas na amostra com menor dimensão e ligeiramente mais baixas, nas de maior dimensão, o que poderá estar relacionado com o número de pontos da escala de medida.

Neste exemplo e considerando os estudos de Katz & McSweeney (1980) em que o autor considera que as variáveis com formato de medida com 3 pontos violam claramente o pressuposto de continuidade, tanto a aplicação dos testes paramétricos como a aplicação dos testes não paramétricos pode ser questionável. Em escalas com um formato de medida com 3 pontos os investigadores devem ser cautelosos na tomada de decisão aquando da aplicação dos testes.

Na área da saúde muitas escalas apresentam este formato de medida com 3 pontos. Esta evidência associada ao facto de as amostras, muitas vezes, serem de pequena dimensão (doenças raras, situações específicas) exige dos profissionais e dos investigadores uma atenção redobrada a estes aspetos e só depois equacionar o tipo de análise de dados a usar ou optar por escalas com um formato de medida com um maior número de pontos na escala.

8.3 Avaliação do Risco de Lesão da Pele em Neonatos

A pele como maior órgão sensitivo do corpo humano representa 16% do peso corporal e está anatomicamente estratificada em três camadas independentes e funcionalmente relacionadas, a epiderme, a derme e a hipoderme (Barata, 2002).

A epiderme é basicamente constituída por queratinócitos, formando o epitélio pavimentoso estratificado córneo, camada mais externa da pele que forma a barreira epidérmica e a camada basal que separa a epiderme da derme (em constante renovação). A derme é a camada sob a epiderme onde se encontram a rede venosa e linfática, os fibroblastos responsáveis por sintetizar a elastina e o colagénio (suportam a elasticidade da pele) e as células imunes responsáveis pelas defesas do organismo. Neste folheto da pele tem origem as glândulas sudoríparas, as glândulas sebáceas e os folículos no tecido conjuntivo denso e a hipoderme no tecido celular subcutâneo (tecido adiposo depositado no feto durante o último trimestre de gestação). A hipoderme ou tecido celular subcutâneo é composta por tecido adiposo e tem como funções o isolamento térmico, a reserva de calor e a proteção de choques mecânicos (Barel, Paye & Maibach, 2009; Barbosa, Campos & Chaves, 2006). A pele como órgão fundamental à manutenção da vida é uma estrutura que atua como interface entre o meio interno (o indivíduo) e o

meio externo (meio ambiente) e, tem como funções principais a proteção contra a perda de água, a proteção contra a absorção de substâncias nocivas, a prevenção de traumatismos e a defesa dos micro-organismos (Darmstadt & Dinulos, 2000).

No primeiro trimestre da gravidez inicia-se o desenvolvimento da pele, o qual começa com a formação da epiderme entre a 4^a e a 6^a semana de gestação e a diferenciação dos queratinócitos no estrato córneo entre as 22 e as 24 semanas de gestação. O estrato córneo é composto por 2 a 3 camadas de células, formadas à 28^a semana de gestação, evoluindo para 15 camadas às 32 semanas de gestação (Leite & Mendes, 2008; Barata, 2002). Esta estrutura atinge a maturação entre a 2^a e a 4^a semanas após o nascimento, podendo esta maturação ser prolongada até à 8^a semana no caso dos recém-nascidos pré-termo (mais especificamente com idade gestacional inferior a 30 semanas) (Martins & Tapia, 2009).

A maturidade da pele é basilar na proteção de lesões cutâneas, as quais podem surgir devido a causas mecânicas, químicas, térmicas ou infecciosas, daí a importância em determinar a causa da lesão da pele e estabelecer as intervenções apropriadas, baseadas na evidência científica, pois a reparação de feridas é um processo fisiológico extremamente lento, complexo e doloroso (AWHONN, 2013); este processo é agravado no recém-nascido pré-termo, devido à sua imaturidade e, como tal, maior vulnerabilidade.

A infeção é um dos principais problemas das crianças prematuras e a pele o órgão protetor fundamental à sua prevenção, por isso o profissional de saúde deve ser um cuidador da manutenção da integridade da pele, o que obriga a uma avaliação constante das alterações da coloração da pele, da sua sensibilidade e do seu pH.

Ao nascer a criança tem um pH alcalino (6.34) que vai acidificando (4.95) nos primeiros dias de vida num processo que tem como finalidade a proteção contra a infeção (Lund, 1999). Este processo de acidificação pode ser adiado no “grande prematuro” (Fox, Nelson & Wareham, 1998) devido ao facto de a função de proteção da pele estar menos ativa porque o estrato córneo apresenta um atraso de desenvolvimento. A imaturidade do estrato córneo pode causar instabilidade térmica e consequentemente uma maior perda de calor por evaporação, um incremento das necessidades de água e um aumento do risco de absorção, de substâncias aplicadas na pele o que pode levar ao

aumento do risco de lesão da pele e, naturalmente aumentar o risco de infecção. O facto do estrato córneo se apresentar mais fino faz com que a função de barreira da pele esteja reduzida em recém-nascidos com menos de 34 semanas de gestação (Rutter, 1987; Sedin et al., 1985; Irving, 2006a, 2006b), o que se traduz na diminuição da sua protecção e no agravamento do risco de lesão da pele e consequentemente de infecção.

A pele da criança prematura é caracteristicamente imatura, por isso mais fina, sensível, frágil, transparente e com rede capilar visível devido à escassez de tecido adiposo. A falta dos apêndices que ancoram a epiderme à derme (molas epiteliais) faz com que haja menos coesão entre as camadas mais finas da pele, o que facilita a sua lesão (irritações, escoriações, queimaduras químicas) (Askin & Wilson, 2014). Todos estes aspetos funcionam como elementos facilitadores de lesão da pele em recém-nascidos e, mais especificamente, nos grandes prematuros.

Um dos grandes problemas destas crianças é a aplicação e remoção de dispositivos aderentes à pele (adesivos, elétrodos, etc.), porque a ligação entre esses dispositivos e a pele é mais forte do que a ligação entre a epiderme e a derme ou entre as camadas da epiderme (Hardman, 2006; Inving, 2006a) o que faz aumentar o risco de lesão. Os fatores de risco mais relevantes relacionados com o desenvolvimento de úlceras de pressão em populações neonatais são a imobilidade, a fricção, o cisalhamento (deformação de um corpo quando sujeito à ação combinada da gravidade e fricção), a desnutrição, a deficiente perfusão tecidual, o comprometimento da oxigenação, e a utilização de dispositivos fundamentais à prestação de cuidados em neonatologia (WOCNS, 2011; Irving et al., 2006, Irving, 2001). As úlceras de pressão aparecem normalmente localizadas em proeminências ósseas, constituindo uma área de lesão devido a mecanismos de pressão, fricção ou da associação entre os dois (Fujii et al., 2010; EPUAP & NPUAP, 2009; Sanada et al., 2008).

A Organização Mundial de Saúde (OMS, 2007) considera fundamental a avaliação do risco de lesão da pele e a prevenção de úlceras de pressão, incentivando o uso de mecanismos protetores. Esta organização tem vindo a utilizar as taxas de incidência e de prevalência das úlceras de pressão como um dos indicadores para determinar a qualidade dos cuidados em saúde e a Direção Geral de Saúde (DGS) menciona este tipo de lesão como um dos grandes problemas de saúde pública do nosso

país devido ao sofrimento causado aos clientes e devido à diminuição da qualidade de vida dos doentes e dos cuidadores.

A DGS alerta para a importância da avaliação dos fatores de risco do desenvolvimento de úlceras de pressão, pois a sua associação ao conhecimento da etiologia das lesões é fundamental ao eficaz planeamento e implementação de medidas para a prevenção e tratamento de lesões da pele. As estimativas sobre as úlceras de pressão, apresentadas pela DGS (2011), indicam que 95% das situações em que aparece uma lesão cutânea, esta podia ter sido evitada desde que precocemente identificada e avaliado o seu grau de risco (DGS, 2011; EPUAP & NPUAP, 2009).

Para que os profissionais façam a identificação do risco de lesão da pele é fundamental uma observação rigorosa, com instrumentos sustentados pela teoria e pelos resultados da investigação. Este tipo de instrumentos permite aos profissionais de saúde uma linguagem comum na observação da criança e na deteção precoce de fatores de risco de úlceras de pressão (Ferreira et al., 2007). A observação da pele dos recém-nascidos, pela sua especificidade, exige a utilização de instrumentos de medida validados para a população neonatal (do zero aos 28 dias). Instrumentos válidos e fiáveis permitem ao profissional de saúde, identificar facilmente a suscetibilidade de um recém-nascido desenvolver úlceras de pressão e estar alerta para o risco de lesão, facilitam a avaliação dos cuidados, conferem objetividade aos planos de cuidados e permitem medir os resultados sensíveis aos cuidados para o desenvolvimento da criança e para a saúde individual e coletiva.

8.3.1 Escala de Observação do Risco de Lesão da Pele

As escalas de observação da pele em pediatria foram baseadas nas escalas desenvolvidas para observar adultos. A *Braden Scale* para a população adulta (Fernandes & Caliri, 2008; Braden & Bergstrom, 1987) serviu de suporte ao desenvolvimento da *Neonatal Skin Risk Assessment Scale* (NSRAS), a qual utiliza os fatores de risco específicos para o desenvolvimento de úlceras de pressão em recém-nascidos (Huffines & Logsdon, 1997).

A NSRAS encontra-se em fase de validação estatística, com recém-nascidos portugueses, como Escala de Observação do Risco de Lesão da Pele em Neonatos sendo

constituída por 6 itens (Condição Física Geral, Estado Mental, Mobilidade, Atividade, Nutrição e Humidade), com um formato de medida com 4 pontos (Martins, Curado & Marçal, 2014).

O primeiro item faz a caracterização do recém-nascido relativamente à sua “condição física” sendo suportado pela variável idade gestacional, tendo em consideração o facto de a imaturidade ser um fator de risco cutâneo. O item 2 avalia o “estado mental” baseado no comportamento do recém-nascido perante os estímulos dolorosos e o desconforto, oscilando entre a incapacidade de resposta (estado de sedação ou incapacidade sensorial) e a resposta sensorial “normal” (cf. Quadro 8.3. 1).

Quadro 8.3. 1 - Descritores da Escala de Observação do Risco de Lesão da Pele de Neonatos

Escala de Observação do Risco de lesão da Pele de Neonatos				
Item	Opções de resposta			
Condição Física Geral	4. Idade Gestacional <28 semanas	3. Idade Gestacional > 28 semanas, mas <33 semanas	2. Idade gestacional > 33 semanas, mas <38 semanas	1. Idade gestacional > 38 semanas
Estado Mental	4. Completamente limitado. Não responde a estímulos dolorosos, devido à redução do nível de consciência ou sedação ou ao desenvolvimento motor esperado para a idade gestacional (não se estremece, agarra ou geme, não há aumento da pressão arterial ou da frequência cardíaca)	3. Muito limitado. Responde apenas a estímulos dolorosos (estremece, agarra, geme, aumento da pressão arterial e frequência cardíaca)	2. Ligeiramente limitado. Letárgico/Hipotónico	1. Nenhuma limitação. Desperto e Ativo
Mobilidade	4. Completamente imóvel. Não faz qualquer alteração ligeira de posição do corpo ou das extremidades sem ajuda (com ou sem sedação)	3. Muito limitada. Capaz de fazer ligeiras alterações ocasionais do corpo ou das extremidades, mas incapaz de fazer alterações frequentes de forma independente	2. Ligeiramente limitada. Capaz de fazer alterações frequentes, embora ligeiras, de posição do corpo ou extremidades de forma independente	1. Nenhuma limitação. Capaz de fazer alterações frequentes e significativas de posição sem ajuda (por exemplo, virar a cabeça)
Atividade	4. Completamente limitado à incubadora. Está confinado à incubadora aquecida com humidade sem poder sair dela	3. Limitado à incubadora. Está confinado à incubadora só saindo dela excepcionalmente	2. Ligeiramente limitada. Numa incubadora, mas pode ir ao colo, canguru, etc.	1. Nenhuma limitação. Num berço aberto
Nutrição	4. Muito pobre. Nada “Per os” (fluidos endovenosos exclusivos)	3. Inadequada. Não recebe a quantidade ideal de dieta líquida para as necessidades (fórmula / leite materno) tendo de ser suplementada com fluidos endovenosos	2. Adequada. Alimentação por sonda gástrica que permite satisfazer as necessidades nutricionais para o crescimento	1. Excelente. Biberon/Amamentação em todas as refeições que satisfazem as necessidades nutricionais para o crescimento
Humidade	4. Constantemente húmida. A pele está húmida sempre que a criança é manipulada ou posicionada	3. Húmida. A pele está frequentemente húmida. Os lençóis têm que ser trocados pelo menos uma vez de 8/8 horas, devido a frequentes dejeções semilíquidas e/ou vômitos, bolsados	2. Ocasionalmente húmida. A pele está ocasionalmente húmida. Requer uma troca de lençóis extra aproximadamente uma vez por dia, devido a algumas dejeções e/ou vômitos, bolsados	1. Raramente húmida. A pele está geralmente seca, os lençóis apenas requerem ser mudados a cada 24 horas

O item 3 pontua a “mobilidade” e a sua relação com o posicionamento da criança e como tal o alívio de pressão em regiões mais suscetíveis como as regiões ósseas. O item 4, avalia a “atividade” da criança tendo em atenção as alternâncias entre deitado no leito e “método canguru”, posição que alivia a pressão nas regiões ósseas. O item 5 “nutrição” é influenciado pelo tipo de alimentação da criança e pela desnutrição que apresenta. O item 6 avalia a “humidade” comparativamente à exposição da pele a fluidos nocivos e a relação desta com a pressão em determinadas zonas (Martins, Curado & Marçal, 2014; Baharestani & Ratliff, 2007; Curley et al., 2003).

A lesão da pele é um dos problemas recorrentes nas unidades de internamento, sendo considerado como um dos fatores causais de dor e mal-estar dos clientes que afeta as famílias e os profissionais. Este problema atinge todo o tipo de clientes, adultos, crianças ou recém-nascidos com enfoque nos neonatos de risco e é um problema que pode ser prevenido (DGS, 2011). Relativamente ao neonato, a investigação tem vindo a provar que a observação da pele com instrumentos apropriados permite fazer uma prevenção efetiva das úlceras de pressão, evitando as lesões da pele o que vai repercutir-se no bem-estar e na saúde do recém-nascido (EPUAP-NPUAP, 2009).

8.3.2 Procedimentos

Os dados utilizados para fazer a aplicação prática foram cedidos pela equipa de projeto (Anexo I) que se encontra a fazer a adaptação para a língua portuguesa da *Neonatal Skin Risk Assessment Scale* (NSRAS), como Escala de Observação do Risco de Lesão da Pele de Neonatos (Martins, Curado & Marçal, 2014) e a validação estatística para os recém-nascidos portugueses.

Das observações cedidas, pela equipa de investigação, que resultaram da avaliação de recém-nascidos com a NSRAS, foram selecionadas aleatoriamente 40 observações ($n_i=40$) de cada um dos grupos: com peso inferior ou igual a 1500g, com peso entre 1501 e 2500g e com peso superior a 2500g e depois selecionadas aleatoriamente as restantes amostras ($n_i=10, 20, 30$).

8.3.3 Apresentação dos Resultados

A MANOVA paramétrica e a MANOVA não paramétrica foram aplicadas com 3 variáveis da escala NSRAS. Para fazer o enquadramento dos dados resultantes da avaliação dos recém-nascidos, verificou-se a distribuição de frequência das respostas aos itens em cada um dos grupos, para enquadrar os dados reais nos resultados do estudo de simulação (Capítulo VII).

Da escala NSRAS selecionaram-se os itens 1 (condição física), 3 (mobilidade) e 6 (humidade) que se enquadraram nas distribuições de frequência simétricas em 2 classes de peso e assimétrica positiva na terceira classe de pesos (p_1, p_1, p_2) (cf. Figura 8.3.1 e Tabela 8.3.1) e apresentam correlações de baixa intensidade ($r_{13}=0.374$, $r_{16}=0.385$, $r_{36}=0.421$).

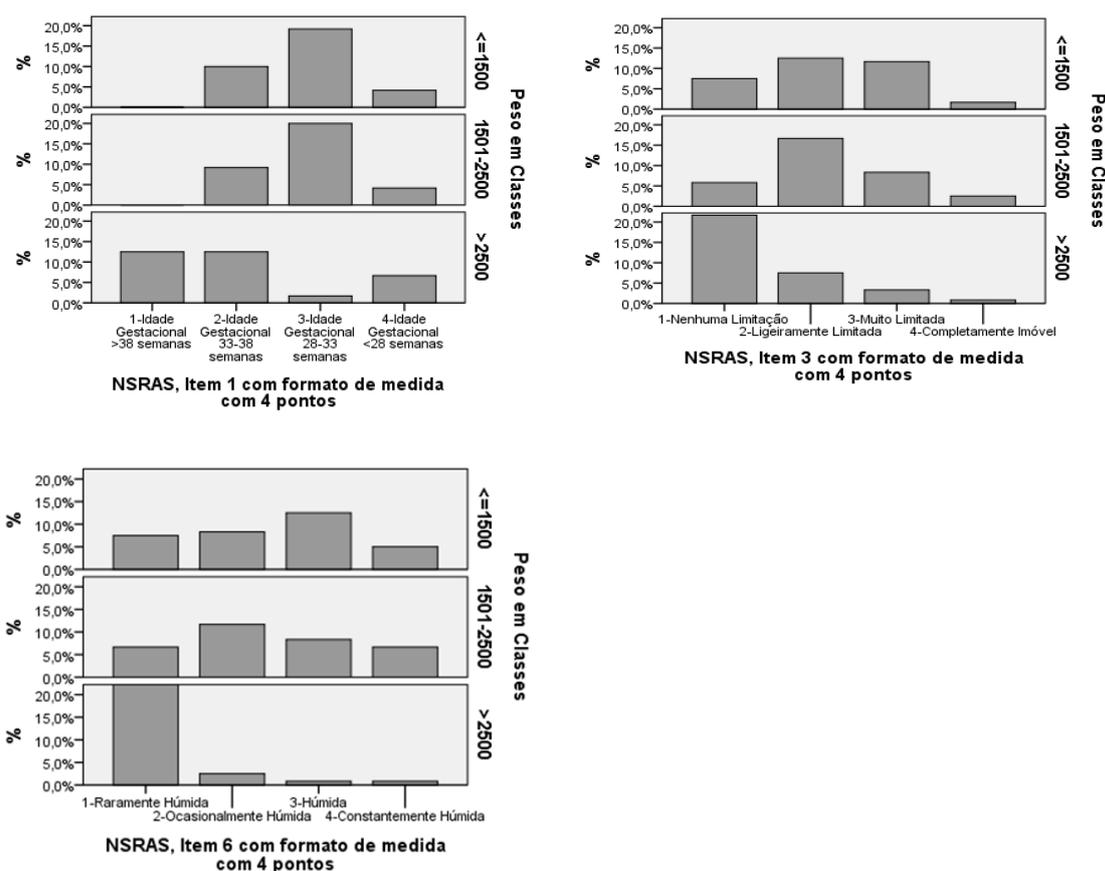


Figura 8.3.1 - Distribuições de frequência em cada um dos três grupos (p_1, p_1, p_2) dos itens 1, 3 e 6 da NSRAS para $n_i=40$

Na Tabela 8.3. 1 apresentam-se os valores da mediana (Me), do coeficiente de assimetria (g_1) e do coeficiente de achatamento (g_2) para os três itens da escala NSRAS, por classes de peso (inferior ou igual a 1500g, de 1501-2500g, superior a 2500).

Para $n_i=40$, os resultados da mediana encontram-se entre 1 e 3. Para o primeiro e segundo grupos, o módulo do coeficiente de assimetria apresenta resultados próximos de zero e inferiores ao dobro do erro padrão (SE) de assimetria (0.748), e o terceiro grupo com resultados do coeficiente de assimetria superior a 0.772 com sinal positivo.

Deste modo o caso em estudo enquadra-se na distribuição simétrica para os dois primeiros grupos e assimétrica positiva para o terceiro grupo (p_1, p_1, p_2).

Tabela 8.3. 1 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 1, 3, e 6 da NSRAS para $n_i=10, 20, 30$ e 40

NSRAS – Formato de medida com 4 pontos													
Itens	n_i Peso	$n_i=10$			$n_i=20$			$n_i=30$			$n_i=40$ $SE_{g_1}=0.374$		
		Me	g_1	g_2	Me	g_1	g_2	Me	g_1	g_2	Me	g_1	g_2
Item 1	<=1500	3.0	0.407	-1.074	3.0	0.080	-0.250	3.0	0.166	-0.502	3.0	0.157	-0.483
	1501-2500	3.0	0.132	0.179	3.0	0.000	-0.279	3.0	0.107	-0.557	3.0	0.104	-0.353
	>2500	2.0	1.085	0.914	2.0	1.147	2.362	2.0	0.703	-0.771	2.0	0.772	-0.734
Item 3	<=1500	2.5	0.687	-1.104	2.0	-0.549	-0.548	2.0	0.109	-0.618	2.0	0.042	-0.796
	1501-2500	2.0	-0.091	1.498	2.0	0.225	-0.018	2.0	0.415	-0.501	2.0	0.391	-0.169
	>2500	1.0	1.658	2.045	1.0	1.514	1.170	1.0	1.608	1.905	1.0	1.510	1.612
Item 6	<=1500	2.5	-0.041	-1.457	2.0	0.136	-1.379	2.0	0.170	-1.064	2.0	0.363	-1.032
	1501-2500	2.0	0.738	-0.878	2.0	0.273	-1.205	2.0	0.190	-1.068	2.0	0.140	-1.094
	>2500	1.0	1.984	2.816	1.0	2.782	7.302	1.0	4.232	19.246	1.0	3.494	12.824

As estatísticas da MANOVA paramétrica e da não paramétrica, nas diferentes dimensões amostrais, referentes aos itens 1, 3 e 6 correspondentes à condição física geral, mobilidade e humidade, respetivamente, da escala NSRAS estão sumariados, por estatística, nas Tabela 8.3.2, Tabela 8.3.3, Tabela 8.3.5, Tabela 8.3.6.

Tabela 8.3.2 - Traço de Pillai nas diferentes dimensões amostrais na escala NSRAS

Traço de Pillai						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.378	2.018	(6,52)	0.080	0.189	0.682
60	0.478	5.859	(6,112)	<0.001	0.239	0.997
90	0.348	6.042	(6,172)	<0.001	0.174	0.998
120	0.337	7.837	(6,232)	<0.001	0.169	1.000

Tabela 8.3.3 - Lambda de Wilks nas diferentes dimensões amostrais na escala NSRAS

Lambda de Wilks						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.639	2.091	(6,50)	0.071	0.201	0.698
60	0.526	6.937	(6,110)	<0.001	0.275	0.999
90	0.653	6.719	(6,170)	<0.001	0.192	0.999
120	0.665	8.679	(6,230)	<0.001	0.185	1.000

Tabela 8.3.4 - Traço de Hotelling nas diferentes dimensões amostrais na escala NSRAS

Traço de Hotelling						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.538	2.153	(6,48)	0.064	0.212	0.710
60	0.892	8.029	(6,108)	<0.001	0.253	0.998
90	0.528	7.394	(6,168)	<0.001	0.308	1.000
120	0.501	9.524	(6,228)	<0.001	0.200	1.000

Tabela 8.3.5 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala NSRAS

Maior raiz de Roy						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.484	4.192	(3,26)	0.015	0.326	0.796
60	0.883	16.486	(3,56)	<0.001	0.469	1.000
90	0.524	15.011	(3,86)	<0.001	0.344	1.000
120	0.495	19.159	(3,116)	<0.001	0.331	1.000

Tabela 8.3.6 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens com a NSRAS

NSRAS – MANOVA não paramétrica								
n	Traço de Pillai - NP				Lambda de Wilks - NP			
	TP_{NP}	χ^2	df	p-value	LW_{NP}	χ^2	df	p-value
30	0.383	11.107	6	0.090	0.633	11.882	6	0.060
60	0.502	29.618	6	<0.001	0.504	38.360	6	<0.001
90	0.400	35.600	6	<0.001	0.602	43.602	6	<0.001
120	0.379	45.101	6	<0.001	0.623	54.868	6	<0.001

As diferentes estatísticas da MANOVA paramétrica mostram potências de teste superiores a 66% na amostra de menor dimensão ($n=30$) e superiores a 99% em todas as outras dimensões amostrais, o que se apresenta em consonância com os resultados obtidos no estudo de simulação.

Os resultados mostram ainda que os testes são estatisticamente significativos com $p\text{-value} < \alpha = 0.05$ em todas as amostras com dimensão igual ou superior a 60 observações em qualquer dos testes, paramétricos ou não paramétricos, pelo que rejeitamos a hipótese H_0 de igualdade dos vetores de valores médios e concluimos que existe efeito do peso dos recém-nascidos nos três itens da NSRAS, ou seja, há diferenças estatisticamente significativas entre pelo menos 2 dos 3 grupos com diferentes classes de peso. Para $n=30$ apenas a maior raiz de Roy apresenta resultados estatisticamente significativos (cf. Tabela 8.3.7).

Tabela 8.3.7 - Valores do $p\text{-value}$ das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais na escala NSRAS

$p\text{-value} - NSRAS$				
MANOVA	n			
	30	60	90	120
TP	0.080	<0.001	<0.001	<0.001
LW	0.071	<0.001	<0.001	<0.001
TH	0.064	<0.001	<0.001	<0.001
MRR	0.015	<0.001	<0.001	<0.001
TPNP	0.090	<0.001	<0.001	<0.001
LWNP	0.060	<0.001	<0.001	<0.001

8.3.4 Discussão dos Resultados

A discussão dos resultados foi organizada para dar resposta ao objetivo do estudo “Analisar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados ordinais provenientes da utilização da Escala de Observação do Risco de Lesão da Pele de Neonatos (NSRAS) e fazer o seu enquadramento nos resultados do estudo de simulação de Monte Carlo” (Capítulo VII).

No estudo de simulação, na 3ª situação, no cenário com escalas com 4 pontos, com distribuições simétricas em 2 grupos e assimétricas positivas no terceiro grupo, a potência nas amostras de menor dimensão é influenciada pelas correlações entre os itens e pelo tipo de teste, sendo a maior raiz de Roy a estatística com potências mais elevadas seguida do lambda de Wilks aplicado às ordens. Porém, quando as correlações são de baixa magnitude, que é caso da NSRAS aqui apresentado, mesmo para $n=30$ as potências são aceitáveis (superiores a 66%), a maior raiz de Roy apresenta uma potência superior a 80% e o lambda de Wilks aplicado às ordens, superior a 75% (Hill & Hill, 2009). Nas amostras com 60 observações ou mais, as potências são superiores a 95%. A maior raiz de Roy apresenta uma potência superior a 99% e o lambda de Wilks aplicado às ordens, superior a 98%.

Relativamente aos resultados que emergiram da análise dos dados reais da escala NSRAS, verifica-se o mesmo comportamento que o obtido no estudo de simulação para os resultados da MANOVA paramétrica, potências de teste elevadas nas amostras com 60 ou mais observações e dimensão do efeito com valores médios]0.05; 0.25] e elevados ([0.25; 0.50]).

Embora sem termo de comparação com o estudo de simulação, no que concerne à potência dos testes e à magnitude dos efeitos da MANOVA não paramétrica aplicada aos dados reais da NSRAS, o comportamento dos *p-value* é semelhante entre os testes paramétricos e os não paramétricos; os testes são estatisticamente significativos com *p-value* $< \alpha = 0.05$ nas dimensões amostrais iguais e superiores a 60 e na maior raiz de Roy para $n=30$ (*p-value*=0.015).

8.3.5 Considerações Finais

A análise do instrumento NSRAS com um formato de medida com quatro pontos permite-nos salientar a simplicidade da construção do instrumento, o que o torna muito fácil de aplicar na prática clínica. O formato de medida com 4 pontos (número par na escala de medida) não parece interferir com as potências dos testes. Os resultados verificados no estudo de simulação e os resultados da aplicação a dados reais com a escala NSRAS são muito semelhantes fazendo um bom enquadramento dos resultados.

Na prática a utilização dos dados da NSRAS com distribuições de frequência simétricas e assimétricas positivas, revelou potências de teste aceitáveis e elevadas nos diferentes testes com magnitude de correlação entre os itens de baixa intensidade. Nas amostras com dimensão inferior a 60 observações os resultados dos dois testes não paramétricos aproximam-se dos obtidos com a maior raiz de Roy, embora o lambda de Wilks seja o que apresenta potências mais elevadas e superiores a todos os outros, o que faz com que possa ser uma boa opção de análise de dados. Com as restantes dimensões amostrais e porque as correlações de baixa magnitude, poderão ser utilizados os testes paramétricos e os não paramétricos pois qualquer deles apresenta potências elevadas.

8.4 Avaliação da Independência Funcional da Criança/Jovem com Espinha Bífida

O movimento é essencial à independência do ser humano, funciona como elemento facilitador do conhecimento, da exploração e do domínio do meio ambiente mais próximo e alargado, e conseqüentemente facilita um desenvolvimento mais harmonioso da criança.

A aquisição de competências motoras advém da experiência sensorial e exploração de si próprio e do meio envolvente, o que pode estar limitado em crianças com problemas motores, pelo facto de todo o processo de aprendizagem estar condicionado por padrões de postura e movimento diferentes, que se vão estabelecendo gradualmente e de forma diferente do padrão dito “normal” (Zonta, Júnior & Santos, 2011). A obtenção constante e gradual de capacidades motoras é fundamental na criança com desenvolvimento “normal” ou “atípico”, pois confere à criança o domínio da independência e capacita-a para uma melhor integração e adaptação social, pelo que a

avaliação da saúde e do desenvolvimento destas crianças são decisivos para a deteção precoce de problemas que podem ser cruciais para o seu futuro.

A criança nasce com determinadas características genéticas que, associadas a fatores ambientais como, a alimentação, as condições de vida e os cuidados de saúde, influenciam o crescimento e o desenvolvimento. A doença é um dos fatores que influencia o crescimento e o desenvolvimento da criança e interfere com a vida da criança e da sua família, nomeadamente se for um processo de doença crónica.

As doenças crónicas que mais afetam as crianças e jovens e provocam incapacidade são as do foro respiratório, alterações da fala, alteração da função sensorial, perturbações mentais e perturbações do sistema nervoso central, as quais incluem as malformações congénitas (McElfresh & Merck, 2014; Arango, 2005) sendo estas responsáveis por alterações da função sensorial, do movimento e consequentemente, do comportamento motor.

Em Portugal, as malformações congénitas têm vindo a diminuir ao longo das últimas décadas, particularmente devido à vigilância pré-natal, à melhoria das condições nutricionais da grávida e à administração pré-natal de ácido fólico. Esta melhoria nos cuidados foi desenvolvida e apoiada pela Comissão Nacional de Saúde Materna e Infantil (CNSMI) com todo o trabalho desenvolvido nos anos oitenta e noventa do século passado (CNSMI, 1989).

O Relatório Nacional de Anomalias Congénitas (2002-2007), divulgado pelo Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA, IP), reporta a prevalência dos defeitos do tubo neural e dos subgrupos neles inseridos, verificando-se que:

Em todos eles se observa uma diminuição das prevalências observadas, em especial em relação à Espinha Bífida quando comparado com as publicações dos anos anteriores (3.0 por 10 000 nados vivos) (INSA, IP, 2010, p. 22).

No entanto, a diminuição da prevalência não implica um menor número de crianças com espinha bífida, pois a sobrevivência até à idade adulta destas crianças aumentou de 10% nos anos cinquenta, para 90% nos anos oitenta, devido aos avanços tecnológicos e ao apoio diferenciado facultado a estas crianças (Caseiro, Gonçalves, & Malheiro, 2013; Antolovich & Wray, 2008). Deste modo, o número de crianças e jovens

adultos com espinha bífida tem aumentado e esta situação crónica continua a ser uma preocupação na área da saúde.

As causas desta malformação não estão claramente identificadas, sendo, no entanto, associadas a fatores de risco genéticos e ambientais, abrangendo um conjunto complexo de anomalias congénitas, que afetam o desenvolvimento do tubo neural e estruturas relacionadas. A carência de ácido fólico tem sido sinalizada como fator de risco, e como tal amplamente estudada, pelo facto de intervir na biossíntese de purinas e pirimidinas e, conseqüentemente, na biossíntese do ácido ribonucleico e desoxirribonucleico, fundamentais para a rápida divisão celular que ocorre durante o desenvolvimento fetal precoce. O encerramento do tubo neural ocorre à quarta semana de gestação e se este encerramento não ocorrer e não for completada a fase de neurulação, verifica-se a ocorrência da malformação, a qual pode apresentar-se de diferentes formas, dependendo da localização da lesão (Wald, et al., 2001; Ray, Singh, & Burrow, 2004).

As formas mais comuns de espinha bífida são a espinha bífida oculta e a espinha bífida cística (meningocelo e mielomeningocelo). Estas malformações têm um grande impacto quer na vida das crianças quer na vida das famílias pelo facto de apresentarem uma grande variabilidade de problemas relacionados com o nível da lesão medular. Salientamos alguns dos problemas mais comuns que afetam: a área motora, podendo verificar-se paralisia flácida, a qual poderá condicionar a mobilidade e a marcha, o que vai determinar o uso de meios auxiliares de marcha (cadeira de rodas, canadianas e outros); a área neuro-sensorial, que podem favorecer a diminuição ou perda de sensibilidade, o que aumenta o risco de lesão da pele; o sistema urológico, podendo surgir bexiga espástica ou flácida, o que aumenta o risco de infeção urinária com conseqüentes lesões renais (necessidade de várias algaliações ao longo do dia) e também a função intestinal, podendo aparecer incontinência por flacidez do esfíncter anal o que obriga ao treino intestinal e muitas vezes ao uso de fraldas (Caseiro, Gonçalves, & Malheiro, 2013; Malheiro, 2005).

O mielomeningocelo é considerado como a forma mais grave e mais comum de espinha bífida, no entanto, 75% das crianças nascidas com este problema apresentam uma boa expectativa de vida e facilmente atingem a idade adulta (Bowman et al., 2001).

Apesar da expectativa de vida, as crianças com situações crónicas apresentam limitações nas suas competências, que se manifestam nas suas atividades de vida diária tornando-as mais ou menos dependentes de cuidados de outros, sobretudo da família. Este tipo de cuidados implica o desenvolvimento de parcerias entre as famílias e os profissionais de saúde no sentido de facilitar a adaptação da criança às múltiplas mudanças no seu desenvolvimento e consequentemente na sua vida (McElfresh & Merck, 2014).

Os portadores de espinha bífida estão incluídos no grupo de clientes com doença crónica. Segundo a Organização Mundial de Saúde (WHO, 2002) a doença crónica inclui pelo menos uma das seguintes especificidades: ser permanente, originar incapacidade, provocar situações irreversíveis e exigir do cliente formação ou supervisão para a sua reabilitação. A espinha bífida passou a ser considerada como doença crónica por despacho conjunto (n.º 861/99 de 10 de setembro) dos Ministérios da Saúde e da Segurança Social e do Trabalho, sendo incluída no grupo das doenças com longa duração. Como situação que envolve aspetos multidimensionais, de evolução gradual dos sintomas, potencialmente incapacitante, de gravidade elevada pelas limitações nas possibilidades de tratamento médico e aceitação pelo doente é uma situação clínica que deve ser considerada no contexto da vida familiar, escolar e laboral, para que todo o apoio seja coordenado e efetivo. Assim, as crianças portadoras de espinha bífida, pela sua incapacidade, estão incluídas no grupo de indivíduos com doença crónica e que dependem especificamente da família a qual pode funcionar como elemento facilitador da sua adaptação.

Os profissionais que trabalham com estes jovens devem estar atentos para verificar se a família funciona como elemento facilitador ou não, pois muitos pais na tentativa de ajudar os filhos acabam por resolver todos os seus problemas, substituindo-os em muitas das suas atividades de vida diária, impedindo-os de progredir no seu desenvolvimento e autonomia (Caseiro, Gonçalves, & Malheiro, 2013; Easton, Rens, & Alexander, 1998). A substituição da criança e a perda de autonomia podem ser agravadas se a criança está a iniciar a fase da adolescência, pois a adolescência, como período de transição da infância para a idade adulta caracteriza-se por mudanças importantes no crescimento e desenvolvimento a nível intelectual, físico, social e

psicológico e é a fase por excelência onde a aquisição da autonomia é primordial (McElfresh & Merck, 2014). Neste estágio, os desafios estão relacionados com a independência das crianças relativamente aos pais, numa procura de identidade, autonomia emocional e económica e ao mesmo tempo na procura de relações interpessoais. No entanto, o tipo de limitação destas crianças e jovens e a tomada de consciência da sua diferença e da sua incapacidade, podem levar ao isolamento social e como tal diminuir a procura desta autonomia (Sandler, 2004). Por todos os aspetos apresentados importa, na reabilitação destas crianças e jovens, reforçar o seu desenvolvimento, capacitando-os no sentido da sua independência, promovendo a sua saúde e o seu bem-estar e habilitando-os a desenvolver as suas potencialidades e capacidades, pois:

O processo de habilitação para além de promover a independência do jovem, possibilita e estimula a sua autonomia nas atividades de vida diária (AVD) e conseqüentemente melhora a sua autoestima e autoconceito. (Caseiro, Gonçalves, & Malheiro, 2013, p. 124).

O conceito de atividades de vida diária (AVDs) surgiu em 1954 para exprimir a independência do indivíduo para as atividades do dia-a-dia (Itamini, 2008; Neri, 2001), seguido dos conceitos de capacitação para a independência funcional e avaliação funcional. O primeiro alerta para a capacidade de o indivíduo realizar atividades básicas de vida diária ou de autocuidado e o segundo para a medida do desempenho do indivíduo nessas atividades (Kawasaki, Cruz & Diogo, 2004).

O objetivo dos profissionais de saúde e das famílias deve ser o de capacitar estas crianças para que atinjam a sua autonomia, pelo que a intervenção deve começar, o mais cedo possível, após uma avaliação prévia da sua situação (Caseiro, Gonçalves, & Malheiro, 2013).

O início da intervenção precoce com crianças com espinha bífida exige a mensuração das suas competências e da sua independência funcional com instrumentos válidos e fiáveis no sentido de planear a intervenção de forma individualizada e assim promover a autonomia nas atividades de vida diária, adequando a reabilitação, no sentido de manter capacidades, trabalhar as potencialidades e ainda fazer a avaliação da evolução e da resposta ao tratamento implementado (Zonta, Júnior & Santos, 2011).

Sabe-se que, dependendo do tipo de lesão medular, o nível de compromisso motor, cognitivo e sensorial será também diferente, pelo que importa usar instrumentos eficazes para avaliar, de forma objetiva, o desenvolvimento e a funcionalidade das crianças com espinha bífida. Ao medir a independência funcional destas crianças e jovens, seja pela via da heteroavaliação seja pela via da autoavaliação, o profissional de saúde verifica a evolução do cliente no seu processo de reabilitação e faz o ajustamento das intervenções terapêuticas com o objetivo da melhoria da qualidade de vida e da qualidade de cuidados (Greve, 2007).

Um dos instrumentos mais utilizados, quer internacionalmente quer a nível nacional, tem sido a Medida de Independência Funcional (MIF), a qual permite fazer o acompanhamento de pessoas em processo de reabilitação e cuja ênfase se situa na realização das atividades de vida diárias de forma independente. Como instrumento sustentado pela teoria e pelos resultados da investigação, a MIF tem permitido o estabelecimento de uma linguagem comum na determinação do grau de independência de crianças e jovens com espinha bífida (Malheiro, Gaspar & Barros, 2013; Laíns, 1991).

8.4.1 Medida de Independência Funcional

A reabilitação de crianças e jovens com espinha bífida exige a utilização de espaços que promovam a sua autonomia e o autoconceito pela via do ensino para a independência. A intervenção na reabilitação exige o recurso a instrumentos de medição apropriados, validados para as populações sob estudo e, no caso da autoavaliação, facilmente compreensíveis pelos jovens que os vão aplicar.

The Functional Independence Measure foi desenvolvida pela Academia de Medicina Física de Reabilitação e pelo Congresso Americano de Medicina de Reabilitação (Sperle et al., 1997; Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM), 1993; Guide for the Functional Independence Measure for Children (WeeFIM), 1993; Keith, et al., 1984, 1987; Granger et al., 1986). O objetivo do seu desenvolvimento foi padronizar conceitos e definições sobre incapacidade e obter um único instrumento de avaliação da reabilitação, capaz de medir o grau de solicitação de cuidados de terceiros, que o indivíduo portador de deficiência exige para

a realização de tarefas motoras e cognitivas (Riberto et al., 2004). Em termos clínicos, este instrumento possibilita a determinação da severidade da incapacidade, a avaliação e acompanhamento dos ganhos funcionais do doente e a avaliação dos resultados da intervenção.

Laíns (1991) fez a tradução e adaptação deste instrumento, para a língua portuguesa, como Medida de Independência Funcional (MIF) e a sua validação estatística, justificando a importância da utilização do instrumento, no facto de este permitir fazer o acompanhamento de pessoas em processo de reabilitação, não sendo focado na capacidade de realizar tarefas, mas sim na realização efetiva e de forma independente das atividades de vida diárias.

A MIF permite avaliar o desempenho da criança e jovem na realização de um conjunto de 18 tarefas incluídas em 2 domínios, o motor (4 dimensões com 13 itens) e o cognitivo (2 dimensões com 5 itens). Cada item apresenta um formato de medida com 7 pontos que vai desde a dependência total até à independência total (cf. Quadro 8.4. 1).

Quadro 8.4. 1- Formato de medida dos itens da escala MIF

N Í V E I S	7. Independência Completa (em segurança e tempo normal)	SEM AJUDA
	6. Independência Modificada (ajuda técnica, tempo extra)	
	Dependência Modificada	
	5. Supervisão (preparação do material/colocação de uma ortótese)	COM AJUDA
	4. Ajuda mínima (o indivíduo realiza mais de 75% da tarefa)	
	3. Ajuda Moderada (o indivíduo realiza entre 50% a 75% da tarefa)	
2. Ajuda Máxima (o indivíduo realiza 25% a 50% da tarefa)		
1 Ajuda Total (o indivíduo tem uma realização inferior a 25% da tarefa)		

O domínio motor é composto pelas subescalas: autocuidado (6 itens), controlo esfinteriano (2 itens), mobilidade e transferências (3 itens), locomoção (2 itens). O domínio cognitivo é composto pelas subescalas: comunicação (2 itens) e cognição social (3 itens) (cf. Quadro 8.4.2).

O score total máximo da escala é de 126 pontos classificados da seguinte forma: a) 18 pontos, dependência completa, b) 19-60 pontos, dependência modificada, c) 61-103 pontos, dependência modificada, d) 104-126 pontos, independência completa/ e/ou modificada (Riberto et al., 2004).

A MIF também permite a avaliação da independência funcional dos clientes, independentemente das sequelas de ordem física, dos problemas de comunicação, dos

problemas funcionais e emocionais, sendo muito usada em situações de doenças do foro neurológico.

Quadro 8.4.2 - Medida de Independência Funcional (MIF)

Medida de Independência Funcional (MIF)		
Domínio Motor	Autocuidado	Alimentação
		Higiene pessoal (cara, dentes, pentear)
		Banho (lavar o corpo, duche ou banheira)
		Vestir metade superior (t-shirt, camisa, <i>top</i>)
		Vestir metade inferior (cuecas, saia ou calças, meias, sapatos)
		Utilização da sanita
	Controlo dos Esfínteres	Bexiga (preparação do material, manobra, limpeza local)
		Intestino (administração de clister, supositório, limpeza períneo)
	Mobilidade Transferências	Leito, cadeira, cadeira de rodas
		Sanita
Banheira ou duche		
Locomoção	Marcha, cadeira de rodas	
	Escadas	
Domínio Cognitivo	Comunicação	Compreensão (visual e auditiva)
		Expressão (verbal e não verbal)
	Cognição Social	Interação social
		Resolução de problemas
		Memória

Malheiro, Gaspar & Barros (2013) usaram a MIF adaptada por Laíns (1991) na sua totalidade e também fizeram a adaptação do Domínio Motor desta escala com um novo formato de medida com cinco pontos para compararem os resultados desta subescala com dados de outra escala, pelo que transformaram o formato de medida da MIF-domínio motor, de sete para cinco pontos. Esta modificação da escala de medida teve em consideração as necessidades de ajuda das crianças com espinha bífida, considerando os pontos 7 (Independência completa), 6 (Independência modificada) e 5 (Supervisão) como pontos em que a criança ou jovem não precisa de ajuda, passando estes três pontos a fazer parte do ponto 5 do novo formato de medida, mantendo os outros quatro pontos como estão na escala original. A MIF motora com 5 pontos apresenta a seguinte escala de medida: 5 - Sem ajuda ou supervisão, 4 - Ajuda mínima, 3 - Ajuda moderada, 2 - Ajuda máxima, 1 - Ajuda total.

Na aplicação prática foram utilizadas 90 observações de crianças/jovens com espinha bífida, com idades compreendidas entre os 8 e os 23 anos e com três diferentes tipos de lesão medular.

8.4.2 Procedimentos

Os dados utilizados para fazer as aplicações práticas foram cedidos pela equipa de investigação (Anexo I) que se encontra a desenvolver um trabalho sobre os “Efeitos de um Programa de Educação para a Autogestão da Incapacidade na Independência Funcional, Autoconceito e Autoestima em Jovens com espinha bífida (Malheiro, Gaspar & Barros, 2013), sendo um dos instrumentos utilizados, a MIF.

Das observações cedidas, pela equipa de investigação, que resultaram da autoavaliação, dos jovens com espinha bífida, com a MIF com uma escala com 5 pontos foram selecionadas aleatoriamente 20 observações ($n_i=20$) de cada um dos grupos com diferentes níveis de lesão medular (sagrada, lombar e dorsal) e depois selecionada aleatoriamente 1 amostras ($n_i=10$), cumprindo 2 exemplos das mesmas dimensões das amostras utilizadas no estudo de simulação ($n=30, 60$).

Para a MIF com uma escala com 7 pontos, foram selecionadas aleatoriamente 30 observações ($n_i=30$) de cada um dos grupos etários (8-12, 13-15 e 16-23 anos) e depois selecionadas aleatoriamente as restantes amostras ($n_i=10, 20$), cumprindo 3 dimensões das amostras utilizadas no estudo de simulação ($n=30, 60, 90$).

8.4.3 Apresentação dos Resultados

A MANOVA paramétrica e a MANOVA não paramétrica foram aplicadas com três variáveis da MIF motora modificada e da MIF motora. Para fazer o enquadramento dos dados resultantes da autoavaliação efetuada pelas crianças e jovens com espinha bífida, nos dados simulados (capítulo VII), verificou-se a distribuição de frequência e a intensidade das correlações entre os itens, os quais passamos a apresentar em seguida.

MIF Dimensão Motora Modificada

Das avaliações efetuadas com a MIF dimensão motora modificada, com formato de medida com 5 pontos, selecionaram-se o item 6 (Autocuidado-utilização da sanita), o item 10 (Mobilidade-transferência para a sanita) e o item 13 (Locomoção-subir e descer escadas) que apresentam distribuições de frequência assimétrica negativa no primeiro grupo e assimétrica positiva nos outros dois (p_3, p_2, p_2) (cf. Figura 8.4.1 e Tabela 8.4.1), com correlações de média magnitude ($r_{6,10}=0.648, r_{6,13}=0.495, r_{10,13}=0.598$).

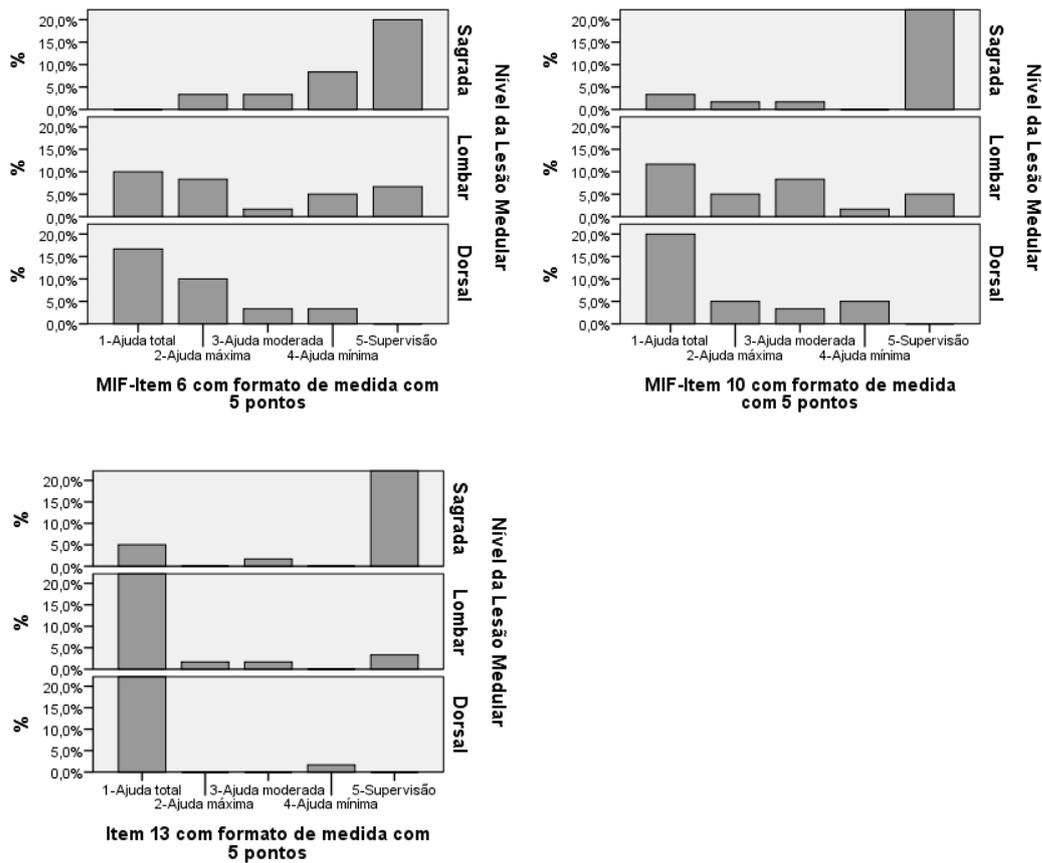


Figura 8.4.1 - Distribuições de frequência em cada um dos três grupos (p_1, p_2, p_3) dos itens 6, 10 e 13 da MIF motora modificada

Na Tabela 8.4. 1 apresentam-se os valores da mediana (Me), do coeficiente de assimetria (g_1) e do coeficiente de achatamento (g_2) para os três itens 6, 10 e 13 da MIF motora modificada, com formato de medida com 5 pontos, pelos três níveis da lesão medular (dorsal, lombar e sagrada).

Para $n_i=20$, os resultados da mediana encontram-se entre 1 e 5. Para o primeiro grupo o módulo do coeficiente de assimetria apresenta resultados superiores ao dobro do erro padrão (SE) de assimetria (1.002) e com sinal negativo, o segundo grupo com valores inferiores nos itens 6 e 10 e, o terceiro grupo com resultados do coeficiente de assimetria superior ao erro padrão de assimetria e com sinal positivo.

Deste modo o cenário em estudo enquadra-se na segunda situação simulada (Capítulo VII) com uma distribuição assimétrica negativa para o primeiro grupo e assimétrica positiva para os outros dois grupo (p_3, p_2, p_2).

Tabela 8.4. 1 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 6, 10, e 13 da MIF motora modificada para $n_i=10$ e 20.

MIF modificada – Formato de medida com 5 pontos							
	n_i		$n_i=10$		$n_i=20$		$SE_{g_1}=0.501$
Itens	Nível da lesão medular	Me	g_1	g_2	Me	g_1	g_2
Item 6	Sagrada	5.0	-1.179	0.571	5.0	-1.527	1.448
	Lombar	20	0.583	-0.756	2.0	0.372	-1.440
	Dorsal	2.0	0.687	-1.346	1.5	1.133	0.359
Item 10	Sagrada	5.0	-2.277	4.765	5.0	-1.923	2.085
	Lombar	2.0	1.328	2.209	2.0	0.649	-0.717
	Dorsal	2.0	0.815	0.254	1.0	1.121	-0.281
Item 13	Sagrada	5.0	-1.779	1.406	5.0	-1.540	0.641
	Lombar	1.0	1.779	1.406	2.0	1.104	1.067
	Dorsal	1.0	-	-	2.0	1.018	1.080

No estudo de simulação, com este tipo de distribuição, com variáveis ordinais com um formato de medida com 5 pontos e correlações de média magnitude tanto os testes paramétricos como os não paramétricos apresentam valores de potência elevados (Hill & Hill, 2009), superiores a 70% para $n=30$ e nas restantes dimensões das amostras, as potências são superiores a 97% em todos os testes.

Os resultados das estatísticas da MANOVA paramétrica e não paramétrica, nas diferentes dimensões amostrais e nas 3 variáveis, referentes à MIF, dimensão motora: itens 6 (Autocuidado-utilização da sanita), 10 (Mobilidade-transferência para a sanita) e 13 (Locomoção-subir e descer escadas) com um formato de medida com 5 pontos e com distribuição de frequência assimétrica negativa no primeiro grupo e assimétrica positiva nos outros 2 grupos (p_3, p_2, p_2) e com correlações de intensidade média estão sumariados, por teste, nas Tabela 8.4.2, Tabela 8.4.3, Tabela 8.4.4, Tabela 8.4.5, Tabela 8.4.6.

As potências de teste são superiores a 99% nas duas dimensões amostrais, seja qual for o teste, resultados que se enquadram nos resultados obtidos no estudo de simulação.

Tabela 8.4.2 - Traço de Pillai nas diferentes dimensões amostrais na escala MIF Motora modificada (itens 6, 10, 13)

Traço de Pillai						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.782	5.563	(6,52)	<0.001	0.391	0.993
60	0.737	10.892	(6,112)	<0.001	0.368	1.000

Tabela 8.4.3 - Lambda de Wilks nas diferentes dimensões amostrais na escala MIF Motora modificada (itens 6, 10, 13)

Lambda de Wilks						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.315	6.511	(6,50)	<0.001	0.439	0.998
60	0.323	13.929	(6,110)	<0.001	0.432	1.000

Tabela 8.4.4 - Traço de Hotelling nas diferentes dimensões amostrais na escala MIF Motora modificada (itens 6, 10, 13)

Traço de Hotelling						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	1.865	7.462	(6,48)	<0.001	0.483	0.999
60	1.911	17.201	(6,108)	<0.001	0.489	1.000

Tabela 8.4.5 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala MIF Motora modificada (itens 6, 10, 13)

Maior Raiz de Roy						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	1.682	14.581	(3,26)	<0.001	0.627	1.000
60	1.809	33.760	(3,56)	<0.001	0.644	1.000

Tabela 8.4.6 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens na escala MIF dimensão Motora modificada (itens 6, 10, 13)

MIF-Motora – MANOVA não paramétrica								
Traço de Pillai NP					Lambda de Wilks NP			
n	TP_{NP}	χ^2	df	p-value	LW_{NP}	χ^2	df	p-value
30	0.777	22.533	6	<0.001	0.322	29.458	6	<0.001
60	0.653	38.527	6	<0.001	0.0393	52.304	6	<0.001

Os resultados dos testes da MANOVA paramétrica e da MANOVA não paramétrica são estatisticamente significativos com $p\text{-value} < \alpha = 0.05$, para $n=30$ e $n=60$, pelo que rejeitamos a hipótese H_0 de igualdade dos vetores de valores médios e concluímos que existe efeito do nível de lesão da coluna nos três itens da MIF motora modificada, ou seja, há diferenças estatisticamente significativas entre pelo menos 2 dos 3 grupos com diferentes níveis de lesão medular (cf. Tabela 8.4.7).

Tabela 8.4.7 - Valores do *p-value* das estatísticas da MANOVA paramétrica e não paramétrica nas duas dimensões amostrais na MIF-Motora modificada (itens 6, 10, 13)

<i>p-value</i> – MIF Motora com formato de medida com 5 pontos		
	<i>n</i>	
MANOVA	30	60
TP	<0.001	<0.001
LW	<0.001	<0.001
TH	<0.001	<0.001
MRR	<0.001	<0.001
TPNP	<0.001	<0.001
LWNP	<0.001	<0.001

MIF Dimensão Motora

Os testes da MANOVA paramétrica e não paramétrica foram aplicados com 3 variáveis da MIF dimensão motora, com formato de medida com 7 pontos selecionaram-se o item 1 (Autocuidado-alimentação), o item 2 (Autocuidado-higiene pessoal) e o item 4 (Autocuidado-vestir a metade superior (t-shirt, camisa, top), Para fazer o enquadramento dos dados resultantes da observação dos jovens com espinha bífida nos resultados do estudo de simulação (Capítulo VII), verificou-se a distribuição de frequência em cada um dos três grupos. As observações efetuadas com a MIF motora apresentavam correlações de magnitude média ($r_{12}=0.525$, $r_{14}=0.535$, $r_{24}=0.568$) e distribuição simétrica no primeiro grupo e assimétrica negativa nos outros 2 grupos (p_1, p_3, p_3) (cf. Figura 8.4.2 e Tabela 8.4.8). Estes resultados permitem fazer o enquadramento no caso (p_2, p_2, p_1) do estudo de simulação.

No estudo de simulação com este tipo de distribuição, os resultados são influenciados pelo número de pontos da escala, pela dimensão da amostra (e.g. amostras de menor dimensão apresentam potências mais baixas), pela intensidade da correlação entre os itens (e.g. quanto maior a intensidade de correlação entre itens menor a potência dos testes). O teste com potência mais elevada é a maior raiz de Roy seguido do lambda Wilks e traço de Pillai aplicados às ordens. Neste cenário com dados que emergem de uma escala com 7 pontos e com um a correlação média entre os itens, as potências diminuem à medida que a dimensão das amostras também diminui.

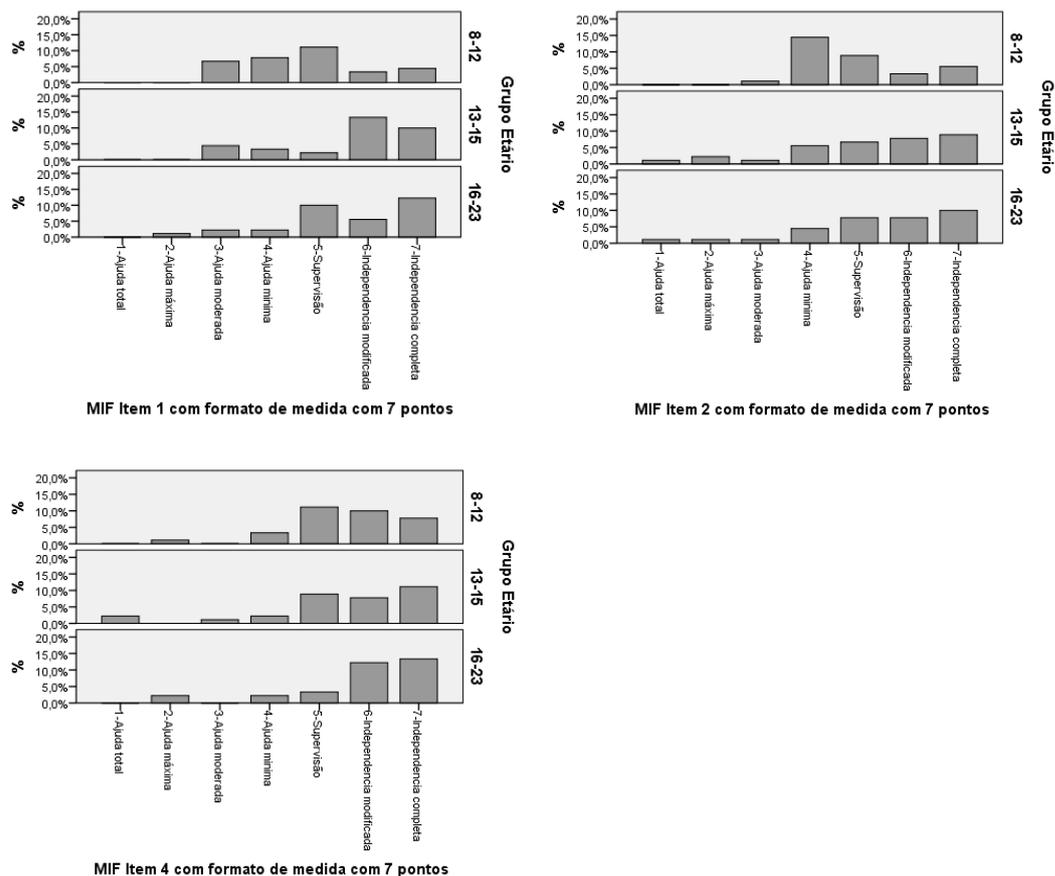


Figura 8.4.2 - Distribuições de frequência em cada um dos três grupos (p_1, p_3, p_3) dos itens 1, 2 e 4 da MIF dimensão motora

Na Tabela 8.4.8 apresentam-se os valores da mediana (Me), do coeficiente de assimetria (g_1) e do coeficiente de achatamento (g_2) por grupos etários nos três itens (1, 2 e 4) da MIF dimensão motora com formato de medida com 7.

Tabela 8.4.8 - Mediana (Me), coeficiente de assimetria (g_1) e coeficiente de curtose (g_2) nos itens 1, 2 e 4 da MIF dimensão motora para $n_i=10, 20$ e 30 .

MIF Motora – Formato de medida com 7 pontos										
Itens	n_i Grupos etários	$n_i=10$			$n_i=20$			$n_i=30$ $SE_{g_1}=0.427$		
		Me	g_1	g_2	Me	g_1	g_2	Me	g_1	g_2
Item 1	8-12	5.0	0.659	-0.406	5.0	0.256	-0.387	5.0	0.329	-0.697
	13-15	6.0	-0.371	-1.589	6.0	-0.973	-0.085	6.0	-0.893	-0.437
	16-23	5.0	-0.390	-0.782	6.0	-0.999	0.445	6.0	-0.820	0.201
Item 2	8-12	5.0	0.431	-1.461	5.0	1.073	0.067	5.0	0.687	-0.668
	13-15	5.5	-0.739	-0.310	6.0	-1.758	4.159	5.5	-0.864	0.168
	16-23	5.5	-1.246	1.815	5.5	-0.911	0.258	6.0	-1.069	1.005
Item 4	8-12	6.0	0.000	-1.393	6.0	-1.042	1.238	6.0	-0.873	1.515
	13-15	6.0	-1.944	4.528	6.0	-0.786	0.468	6.0	-1.446	2.132
	16-23	6.0	-1.758	2.876	6.0	-1.362	1.118	6.0	-1.691	2.723

Para $n_i=30$, os resultados da mediana encontram-se entre 5 e 6. Para o primeiro grupo, o módulo do coeficiente de assimetria apresenta resultados ligeiramente superiores ao dobro do erro padrão (SE) de assimetria (0.427) no item 4 e ligeiramente inferiores no item 1.

Os resultados das estatísticas da MANOVA paramétrica e não paramétrica, nas diferentes dimensões amostrais e nas três variáveis (itens 1, 2 e 4) referentes à MIF dimensão motora com um formato de medida com 7 pontos e com distribuição de frequência simétrica e assimétrica negativa (p_1, p_3, p_3), que se enquadram no caso (p_2, p_2, p_1) do estudo de simulação; estão sumariados, por estatística, nas Tabela 8.4.9, 8.4.10, 8.4.11, 8.4.12, 8.4.13).

Tabela 8.4.9 - Traço de Pillai nas diferentes dimensões amostrais na escala MIF Motora (itens 1, 2, 4).

Traço de Pillai						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.031	0.137	(6,52)	0.991	0.016	0.081
60	0.148	1.489	(6,112)	0.188	0.074	0.558
90	0.132	2.022	(6,172)	0.065	0.066	0.725

Tabela 8.4.10 - Lambda de Wilks nas diferentes dimensões amostrais na escala MIF Motora (itens 1, 2, 4).

Lambda de Wilks						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.969	0.133	(6,50)	0.991	0.016	0.080
60	0.855	1.496	(6,110)	0.186	0.075	0.560
90	0.870	2.042	(6,170)	0.063	0.067	0.730

Tabela 8.4.11 - Traço de Hotelling nas diferentes dimensões amostrais na escala MIF Motora itens (itens 1, 2, 4).

Traço de Hotelling						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.032	0.128	(6,48)	0.992	0.016	0.079
60	0.167	1.503	(6,108)	0.184	0.077	0.561
90	0.147	2.061	(6,168)	0.060	0.069	0.734

Tabela 8.4.12 - Maior Raiz de Roy nas diferentes dimensões amostrais na escala MIF Motora itens (itens 1, 2, 4)

Maior Raiz de Roy						
n	Value	F	df	p-value	Eta Squared (η_p^2)	Observed Power (π)
30	0.031	0.268	(3,26)	0.848	0.030	0.094
60	0.147	2.745	(3,56)	0.051	0.128	0.634
90	0.131	3.763	(3,86)	0.014	0.116	0.795

Tabela 8.4.13 - Traço de Pillai e Lambda de Wilks calculado a partir das ordens na escala MIF dimensão motora itens (itens 1, 2, 4)

MIF-Motora – MANOVA não paramétrica								
n	Traço de Pillai NP				Lambda de Wilks NP			
	TP_{NP}	χ^2	df	p-value	LW_{NP}	χ^2	df	p-value
30	0.013	0.377	6	1.000	0.987	0.338	6	1.000
60	0.180	10.620	6	0.100	0.823	10.92	6	0.090
90	0.129	11.481	6	0.070	0.874	11.610	6	0.070

Tal como no estudo de simulação, a aplicação prática com a MIF motora com correlações de intensidade média entre os itens apresenta potências de teste baixa e média intensidade nas três dimensões amostrais (entre 10% e 70%), embora com valores ligeiramente superiores aos resultados do estudo de simulação nas amostras com 60 e 90 observações.

Os resultados dos testes não são estatisticamente significativos ($p\text{-value} > \alpha = 0.05$) nas três dimensões amostrais e em qualquer dos testes, paramétricos ou não paramétricos, pelo que não rejeitamos H_0 da igualdade dos vetores de valores médios e concluímos que não existe efeito dos grupos etários nos três itens da MIF motora (cf. Tabela 8.4.14).

Tabela 8.4.14 - Valores do $p\text{-value}$ das estatísticas da MANOVA paramétrica e não paramétrica nas diferentes dimensões amostrais na escala MIF motora (itens 1, 2, 4)

p-value - MIF			
	n		
MANOVA	30	60	90
TP	0.991	0.188	0.065
LW	0.991	0.186	0.063
TH	0.992	0.184	0.060
MRR	0.848	0.051	0.014
TPNP	1.000	0.100	0.070
LWNP	1.000	0.090	0.070

8.4.4 Discussão dos Resultados

A discussão dos resultados foi organizada para dar resposta ao objetivo do estudo “Analisar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados ordinais provenientes da utilização da Medida de Independência Funcional (MIF) com 2 formatos de medida com 5 e 7 pontos (Laíns, 1991; Malheiro, Gaspar & Barros, 2013) e fazer o seu enquadramento nos resultados do estudo de simulação de Monte Carlo” (Capítulo VII).

No estudo de simulação com variáveis ordinais com um formato de medida com 5 pontos, correlações entre itens de média intensidade e distribuições de frequência assimétrica negativa no primeiro grupo e assimétricas positivas nos outros dois (p_3, p_2, p_2), a potência dos testes é elevada em todos os testes, paramétricos ou não paramétricos, superiores a 83% (Hill & Hill, 2009).

Na aplicação com dados reais da MIF motora modificada, as potências são superiores a 90% em todos as estatísticas da MANOVA paramétrica, o que se enquadra nos resultados do estudo de simulação, embora com resultados de potências ligeiramente superiores. Os testes não são estatisticamente significativos, com *p-value* $< \alpha = 0.05$ e a dimensão do efeito apresenta valores baixos (< 0.05) e médios ($[0.05; 0.25]$) (Marôco, 2010; Cohen, 1988).

No estudo de simulação com variáveis ordinais com um formato de medida com 7 pontos, com correlações de média magnitude e com distribuições de frequência simétricas num dos grupos e assimétricas positivas nos outros dois grupos (p_1, p_2, p_2); a potência dos testes é influenciada pela dimensão da amostra, pelo número de pontos da escala e pela magnitude da correlação entre itens. Nesta situação específica as potências são baixas nas amostras de menor dimensão e diminuem com o aumento da magnitude das correlações. A maior raiz de Roy apresenta valores mais elevados, seguida do lambda de Wilks e do traço de Pillai aplicados às ordens em todas as amostras, mas também baixas, nas amostras com dimensão inferior a 90 observações.

Na aplicação com dados reais da MIF motora (7 pontos) as potências dos testes apresentam o mesmo comportamento observado no estudo de simulação quanto menor a dimensão da amostra, maior o número de pontos da escala e maior a correlação entre itens, menor a potência, mas com resultados superiores relativamente ao estudo de

simulação, quer para $n=60$, quer para $n=90$. Embora sem termo de comparação, no que concerne à potência dos testes e à magnitude do efeito da MANOVA não paramétrica, o comportamento dos resultados dos p -value é semelhante entre os testes paramétricos e os não paramétricos; apenas a maior raiz de Roy apresenta resultados estatisticamente significativos com p -value $< \alpha = 0.05$ para $n=60$ e 90.

8.4.5 Considerações Finais

A análise dos instrumentos MIF motora modificada com um formato de medida com 5 pontos e MIF motora com um formato de medida com 7 pontos permite-nos salientar alguns aspetos que consideramos fundamentais. Começamos por justificar a utilização dos dois formatos de medida para comparar os resultados que emergem nas duas situações e com o facto de a escala original ser muito utilizada na prática clínica.

No caso da MIF motora com um formato de medida com 7 pontos, verificamos resultados de potências muito baixas nas amostras de menor dimensão. No estudo de simulação a maior raiz de Roy e o lambda de Wilks aplicado às ordens são os testes com potências mais elevadas, o que nos permite dizer que nesta situação específica estes testes são uma boa opção de escolha para análise de dados. Acrescentamos ainda que tendo em conta o cenário enunciado, os testes não paramétricos apresentam potências mais elevadas nas amostras de menor dimensão, logo poderão ser uma mais-valia para análise de dados sempre que as amostras sejam de pequena dimensão.

Na escala modificada com um formato de medida com 5 pontos verificam-se potências elevadas em todos os testes o que permite considerar a aplicação quer dos testes paramétricos quer dos testes não paramétricos seja qual for o tipo de correlação entre itens e a dimensão da amostra.

A dimensão do efeito apresenta-se com valores elevados na MIF com 5 pontos e médios na MIF com 7 pontos, o que pode ser mais um aspeto a ter em atenção na escolha do número de pontos do formato de medida de uma escala. No caso em análise e, tendo em conta o resultado das potências dos testes, podemos dizer que o formato de medida com 5 pontos revela melhores resultados do que o formato com 7 pontos.

Os níveis de significância, os p -values nas diferentes estatísticas da MANOVA paramétrica e da MANOVA não paramétrica são inferiores ao nível de significância

nominal ($\alpha = 0.05$) nas diferentes dimensões amostrais na MIF motora modificada enquanto na MIF motora com formato de medida com 7 pontos os resultados apenas são estatisticamente significativos para a maior raiz de Roy ($n=30$ e $n=90$), os restantes testes não são estatisticamente significativos ($p\text{-value} > \alpha = 0.05$) para as amostras com dimensão entre 30 e 90 observações.

8.5 Conclusões

Em investigação e especificamente em análise de dados, a análise da potência dos testes e da probabilidade do erro de tipo I é fundamental à validação dos resultados. Em diferentes áreas do conhecimento, o investigador estabelece antecipadamente a probabilidade de erro de tipo I que normalmente está ao nível de significância nominal ($\alpha = 0.05$) e tem a pretensão de que o erro de tipo II também seja reduzido ($\beta \leq 0.20$), ou seja atingir uma potência de teste ≥ 0.80 (Marôco, 2010), havendo em algumas áreas aceitação de valores de potência superiores a 0.50 como satisfatórios. A potência de teste não é fácil controlar pelo facto de ser influenciada pela probabilidade de erro de tipo I (quanto maior o alfa maior a potência), pelo tipo de testes estatísticos, pela dimensão da amostra e pela dimensão do efeito observado.

Nesta aplicação prática, em que foram usados dados que emergem de três escalas de avaliação, EFS, NSRAS e MIF motora e MIF motora modificada e, à semelhança dos resultados encontrados no estudo de simulação, concluímos que com instrumentos com itens ordinais, o investigador deve ter em consideração, a distribuição de frequência da resposta aos itens, a dimensão da amostra, a intensidade de correlação entre itens e o número de pontos na escala de medida para poder tomar decisões relativamente ao tipo de teste a aplicar, o que vai ao encontro da investigação desenvolvida nesta área e também é confirmado no estudo de simulação apresentado anteriormente. Importa reforçar que com estas escalas, tendencialmente as distribuições de frequência foram do tipo simétricas, assimétricas negativas e assimétricas positivas, não se verificando nenhuma distribuição uniforme nas variáveis analisadas.

No estudo de simulação a potência dos testes da MANOVA é influenciada por todos os fatores enunciados, sendo a maior raiz de Roy o teste que apresenta valores de potência mais elevados. Da análise dos resultados que emergiram da aplicação das 4

escalas, podemos dizer que os provenientes das escalas com um formato de medida com 4 e 5 pontos se enquadraram melhor nos resultados do estudo de simulação, enquanto os resultados que procedentes das escalas com 3 e 7 pontos foram mais difíceis de enquadrar nos resultados do estudo de simulação, apresentando algumas discrepâncias.

8.6 Referências Bibliográficas

- Anand, K.J., & Hall, R.W. (2008). Love, pain, and intensive care. *Pediatrics*, 121, 825-827.
- Anand, K.J., & Hall, R.W. (2006). Pharmacological therapy for analgesia and sedation in the newborn. *Archives of Disease in Childhood, Fetal and Neonatal Edition*, 91, 448-453.
- Antolovich, G.C., & Wray, A.C. (2008). *Habilitation of Children and Young Adults with Spina Bifida*. In: Özek, M.M., Cinalli, G., & Maixner, W.J. (coord.). *The Spina Bifida – Management and Outcome*. Italia: Springer-Verlag, 341-348.
- Arango, P. (2005). *Special health care needs*. In Cosby, A.G. (Ed.), *About children: an authoritative resource on the state of childhood today*. Elk Grove, American Academy of Pediatrics.
- Askin, D.F., & Wilson, D. (2014). Recém-Nascido de Alto Risco e a Família. In M.J. Hockenberry & D. Wilson. *Wong: Enfermagem da Criança e do Adolescente*, 9ª Ed. Vol. I (Cap. 1-25). Loures. Lusociência, 331-411.
- Association of Women's Health, Obstetric and Neonatal Nurses (AWHONN, 2013). *Neonatal skin care: evidence based clinical practice guideline* (3rd Ed.). Washington, DC, AWHONN.
- Baharestani, M., & Ratliff, C. (2007). Pressure Ulcers in Neonates and Children: an NPUAP, White Paper. *Advances in Skin Wound Care*, 20 (4), 208-220.
- Barbosa, A.L., Campos, A.C.S., & Chaves, E.M.C. (2006). Complicações não clínicas da ventilação mecânica: Ênfase no cuidado de enfermagem neonatal. *Acta Paulista de Enfermagem*, 19(4), 439-443.
- Barata, E. (2002). *Cosméticos: Arte e Ciência*. Lidel, Ed. Técnicas, 1-27.
- Barel, A.O., Paye, M., & Maibach, H.I. (2009). Cosmetic Science and Technology. *Informa Healthcare*, 3, 881-887.
- Bowman, R., McLone, D., Grant, J., Tomita, T., & Ito, J. (2001). Spina bifida outcome: a 25-year prospective. *Pediatric Neurosurgery*, 34(3), 114-120.
- Braden, B. & Bergstrom, N. (1987). A conceptual scheme for the study of etiology of pressure sores. *Rehabilitation Nursing*, 12, 8-16.
- Caseiro, J., Gonçalves, T., & Malheiro, M.I.C. (2013). Construção da Autonomia dos Jovens Portadores de Spina Bífida – O que dizem os jovens e seus pais/cuidadores. *Enfermería Global*, 30, 121-134.
- Castro, A.G., Lima, M.C., Aquino, R.R., & Eickmann, S.H. (2007). Desenvolvimento do sensorio motor oral global em lactentes pré-termo. *Pró-fono Revista de Atualização Científica*, 19(1), 29-38.
- Comissão Nacional de Saúde Materno Infantil – Ministério da Saúde (CNSMI – MS) (1989). *Relatório da Comissão Nacional de Saúde Materno Infantil*. Lisboa, MS.

- Chizawsky, L.L., & Scott-Findlay, S. (2005). Tummy time! Preventing unwanted effects of the "Back to Sleep" campaign. *AWHONN Lifelines*, 9(5), 382-387.
- Curado, M.A.S., Mâroco, J., Vasconcellos, T., Marques, L.C., Oliveira, S.R., & Nicolau, C.S. (2013). Versão Reduzida da Escala de Observação de Competências Precoces na Alimentação Oral. Um estudo de validação estatística com crianças prematuras Portuguesas. In *Actas do 1º Encontro de Enfermeiros de Neonatologia da Área de Lisboa: "Cuidar para o Desenvolvimento"*, 15-16 novembro, resumo 1.
- Curado, M.A.S., Teles, J., & Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, 446-457. <http://www.um.es/eglobal>.
- Curado, M.A.S., Teles, J. & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process, *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48(1), 146-152.
- Cunha, M. (2006). *Avaliação do padrão de sucção nutritiva do recém-nascido prematuro de muito baixo peso ao recém-nascido de termo*. Cruz Quebrada, Ed. FMH, UTL.
- Curley, M.A.Q., Rasmus, I.S., Roberts, K.E., & Wypij, D. (2003). Predicting pressure ulcer risk in pediatric patients: The Braden Q Scale. *Nursing Research*, 52(1), 22-33.
- Darmstadt, G.L., & Dinulos, J.G. (2000). Neonatal skin care. *Pediatric Clinics of North America*, 47(4), 757-782.
- Direção Geral de Saúde (DGS) (2011). *Orientações sobre Úlceras de Pressão* (017). Escala de Braden: Versão Adulto e Pediátrica (Braden Q). Consultado em 22 de outubro, 2012: Disponível em: <http://www.dgs.pt/upload/membro.id/ficheiros/i015800.pdf>
- Doucette, J., & Pinelli, J. (2004). The effects of family resources, coping, and strains on family adjustment 18 to 24 months after the NICU experience. *Advances in Neonatal Care*, 4(2), 92-104.
- Easton, J.K., Rens, B., & Alexander, M.A. (1998). *Psychosocial Aspects of Childhood Disabilities*. In Molnar, G.E., Alexander, M.A. (Coord.), *Pediatric Rehabilitation*. Elsevier Health Sciences (Hanley and Belfus), 111-124.
- European Pressure Ulcer Advisory Panel and National Pressure Ulcer Advisory Panel (EPUAP-NPUAP) (2009). *Prevention and Treatment of Pressure Ulcers*. Washington DC: National Pressure Ulcer Advisory Panel.
- European Pressure Ulcer Advisory Panel and National Pressure Ulcer Advisory Panel (EPUAP & NPUAP) (2009). *Prevention and treatment of pressure ulcers: quick reference guide*. Washington DC: National Pressure Ulcer Advisory Panel.
- Fernandes, L.M., & Caliri, M.H.L. (2008). Uso da escala de Braden e de Glasgow para identificação do risco para úlceras de pressão em pacientes internados em centro de terapia intensiva. *Revista Latino-Americana Enfermagem*. 16(6), 1-7.
- Ferreira, P., Miguéns, C., Gouveia, J., & Furtado, K. (2007). *Risco de Desenvolvimento de Úlceras de Pressão: Implementação Nacional da Escala de Braden*. Coimbra, Lusociência.
- Figueiredo, H., Barreiros, J.M.P., Gonçalves, I.F.B., & Cunha, M. (2007a). A intervenção oral na alimentação do recém-nascido prétermo de muito baixo peso. *Desenvolvimento Motor da Criança*. FMH edições, 239-249.
- Figueiredo, H., Barreiros, J.M.P., Gonçalves, I.F.B., & Cunha, M. (2007b). Avaliação de um Programa de Intervenção Oral em Recém-nascidos Pré-termo de Muito Baixo Peso. *ESSFISIONLINE*, 3(2), 3-12.

- Finch, W.H. (2005). Comparison of the Performance on Nonparametric and Parametric MANOVA Test Statistics when Assumptions are Violated. *Methodology*, 1(1), 27-38.
- Fox, C., Nelson, D., & Wareham, J. (1998). The timing of skin acidification in very low birth weight infants. *Journal of Perinatology*, 18(4), 272-275.
- Fucile, S., McFarland, D.H., Gisel, E.G., & Lau C. (2012). Oral and non-oral sensorimotor interventions facilitate suck-swallow-respiration functions and their co-ordination in pre-term infants. *Early Human Development*, 88(6), 345-350.
- Fucile, S., Gisel, E.G., McFarland, D.H., & Lau, C. (2011). Oral and non-oral sensorimotor interventions enhance oral feeding performance in preterm infants. *Developmental Medicine and Child Neurology*, 53(9), 9-35.
- Fucile, S., Gisel, E.G., & Lau, C. (2005). Effect of an oral stimulation program on sucking skill maturation of preterm infants. *Developmental Medicine & Child Neurology*, 47(3), 158-162.
- Fujii K., Sugama J., Okuwa M., Sanada H., & Mizokami Y. (2010). Incidence and risk factors of pressure ulcers in seven neonatal intensive care units in Japan: a multisite prospective cohort study. *International Wound Journal*, 7, 323-328.
- Granger, C.V., Ottenbacher, K.J., & Fiedler, R. (1995). The Uniform Data System for Medical Rehabilitation: Report of First Admissions for 1993. *American Journal of Physical Medicine & Rehabilitation*, 74(1), 62-66.
- Granger, C.V., Hamilton, B.B., Keith, R.A., Zielezny, M., & Sherwin, F.S. (1986) Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation*, 1, 59-74.
- Greve, J. (2007). *Tratado de medicina de reabilitação*. São Paulo. Roca.
- Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM) (1993). Uniform Data System for Medical Rehabilitation: Procedures for scoring the functional independence measure (FIM), version 4.0, Buffalo (NY), State University of New York at Buffalo.
- Guide for the Functional Independence Measure for Children (WeeFIM) (1993). Uniform Data System for Medical Rehabilitation, version 4.0, *Community/Outpatient*. Buffalo (NY), State University of New York at Buffalo.
- Hadley, L.B., West, D., Turner, A., & Santangelo, S. (1999). *Developmental and Behavioral Characteristics of Preterm Infants*. Santa Rosa, NICU Ink Book Publishers.
- Hardman, M. (2006). In utero skin development. In R. White & J. Denyer (Eds.), *Paediatric skin and wound care*. Aberdeen: Wounds UK.
- Hockenberry, M.J., & Barrera, P. (2014). Perspetivas de Enfermagem Pediátrica. In M.J. Hockenberry & D. Wilson. *Wong: Enfermagem da Criança e do Adolescente*, 9ª Ed. Vol. I (Cap. 1-25). Loures, Lusociência, 1-20.
- Holditch-Davis, D., & Blackburn, S.T. (2007). *Newborn and infant neurobehavioral development*. In Kenner C., Lott J., (Ed.), *Comprehensive neonatal nursing: a physiologic perspective*. Philadelphia, Saunders.
- Holditch-Davis, D. (1998). *Neonatal sleep-wake states*. In Kenner C., Lott J.W., Flandermeyer A.A. (ed.), *Comprehensive neonatal care: a physiologic perspective*. Philadelphia, Saunders.
- Huffines, B., & Logsdon, M.C. (1997). The Neonatal Skin Risk Assessment Scale for predicting skin breakdown in neonates. *Issues Comprehensive Pediatrics Nursing*, 20, 103-114.
- Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA, IP) (2010). *Registo Nacional de Anomalias Congénitas. Relatório 2002-2007*. Lisboa, INSA, IP, Departamento de Epidemiologia.
- Irving V. (2001). Caring for and protecting the skin of pre-term neonates. *Journal of Wound Care*, 10(7), 253-256.

- Irving, V. (2006a). Wound care for a preterm neonate. *Journal Wound Care*, 2(3), 102-106.
- Irving, V. (2006b). *Principles of neonatal skin care*. In R. White & J. Denyer (Eds.), *Paediatric skin and wound care*, Aberdeen: Wounds UK, 42-57.
- Irving, V., Bethell, E., Burton, F., & On behalf of a Neonatal Advisory Group (2006). Neonatal wound care: minimizing trauma and pain. *Journal of Wound Care*, 2(1), 33-41.
- Itamini, L.T., Mancussi Faro, A.C., Meneghin, P., Oliveira Leite, R.C.B., & Silveira, C.T. (2009). Adultos com fraturas: das implicações funcionais e cirurgias à educação em saúde. *Revista da Escola de Enfermagem USP*, 43(2), 1238-1243.
- Johnson, B.H., Abraham, M.R., & Parrish, R.N. (2004). Designing the neonatal intensive care unit for optimal family involvement. *Clinics in Perinatology*, 31(2), 353-382.
- Kaarensen, P.I., Ronning, J.A., Tunby, J., Nordhov, S.M., Ulvund, S.E., & Dahl, L.B. (2008). A randomized controlled trial of an early intervention program in low birth weight children: Outcome at 2 years. *Early Human Development*, 84(3), 201-209.
- Katz B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Kawasaki, K., Cruz, K.C.T., & Diogo, M.J.D. (2004). A utilização da medida de independência funcional (MIF) em idosos: uma revisão bibliográfica. *Medicina de Reabilitação*, 23(3), 57-60.
- Keith, R., Granger, C., Hamilton, B., & Sherwin, F. (1984). *Functional Independence Measure, FIM. Uniform Data System for Medical Rehabilitation*. Buffalo, State University of New York.
- Keith, R.A., Granger, C.V., Hamilton, B.B., & Sherwin, F.S. (1987). The Functional Independence Measure: A new tool for rehabilitation. *Advanced Clinical Rehabilitation*, 1, 6-18.
- Kenner, C. & McGrath, J. (2004). *Developmental care of newborns & infants – A guide for health professionals*. St Louis, Elsevier.
- Kline, R.B. (1998). *Principles and Practices of Structural Equation Modelling*. New York, The Guilford Press.
- Laíns, J. (1991). *Guia para o uso do Banco Uniformizado de Dados para Reabilitação Médica – MIF. Versão Portuguesa de Guide for the Uniform Data System for Medical Rehabilitation, FIM (1984)*. Lisboa, Produção Sandoz.
- Land, K.C., & Hui, Z. (2010). *Sample Size, Optimum Allocation, and Power Analysis*. In P.V. Marsden & J.D. Wright, *Handbook of Survey Research*. Bingley, Emerald Group Publishing, Ltd.
- Lau, C., Alagugurusamy, R., Schanler, R.J., Smith, E.O., & Shulman, R.J. (2000). Characterization of the developmental stages of sucking in preterm infants during bottle feeding. *Acta Paediatrica*, 89, 846-852.
- Lau, C., Smith, E.O., & Schanler, R.J. (2003). Coordination of suck-swallow and swallow respiration in preterm infants. *Acta Paediatrica*, 92, 721-727.
- Leite, A.N., & Mendes, V. (2008). Cuidados à pele do recém-nascido pré-termo. Consultado em 29 de dezembro, 2013: Disponível em <http://www.nursingporuguesa.com/>.
- Ludington-Hoe, S.M., Hosseini, R., & Torowicz, D.L.(2005). Skin-to-skin contact (Kangaroo Care) analgesia for preterm infant heel stick. *AACN Clinical Issues*, 16(3), 373-387.
- Lund, C.H. (1999). Prevention and Management of Infant Skin Breakdown. *Nursing Clinics of North America*, 34(4), 907-920.
- Lund, C.H., Osborne, J.W., Kuller, J., Lane, A.T., Lott, J.W., & Raines, D.A. (2001). Neonatal skin care: clinical outcomes of the AWHONN/NANN evidence-based clinical practice guideline. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 30(1), 41-51.

- Nath, R., & Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics and Data Analysis*, 2, 297-315.
- Malheiro, M.I.C. (2005). *Integração/Inclusão de Jovens com Spina Bífida no Ensino Regular... A Voz da Experiência*. Lisboa: Faculdade de Motricidade Humana, Dissertação com vista à obtenção do grau de Mestre em Educação Especial.
- Malheiro, M.I.C., Gaspar, M.F.M., & Barros, L. (2013). The effects of a therapeutic camp: program for promotion of self-management of disability on functionality and self-concept of children with spina bífida. In *Abstracts Book of 57th Annual Meeting of Society for Research into Hydrocephalus and Spina Bífida*, 26-29 June, paper 11.
- Marôco, J. (2010). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Martins, C.O.A., Curado, M.A.S., & Marçal, T.O. (2014). *Adaptation and Validation of the Observation Scale of Skin Injury Risk in New-borns: A Statistical Study of Prévalidation with Portuguese New-Borns*. Abstracts of 13th International Conference of Nursing Research: Person Centered Culture. Fundação Calouste Gulbenkian, Lisboa: APE, resumo 25.
- Martins, C., & Tapia, C. (2009). A pele do recém-nascido prematuro sob avaliação do enfermeiro: cuidado norteando a manutenção da integridade cutânea. *Revista Brasileira de Enfermagem*, 62(5), 778-783.
- McElfresh, P.B., & Merck, T.T. (2014). Cuidados Centrados na Família em situações de Doença Crónica ou incapacidade. In M.J. Hockenberry & D. Wilson. *Wong: Enfermagem da Criança e do Adolescente*, 9^a Ed. Vol. I (Cap. 1-25). Loures. Lusociência, 897-930.
- Neri, A.L. (2001). *Desenvolvimento e envelhecimento*. Campinas, Papirus.
- Nibert, L., & Ondrejka, D. (2005). Family presence during pediatric resuscitation: An integrative review for evidence based practice. *Journal of pediatric nursing*, 20(2), 145-147.
- Organização Mundial de Saúde (OMS) (2007). *International Statistical Classification of Diseases and Related Health Problems*, 10th Revision. Consultado em 22 de outubro, 2013: Disponível em www.who.int/classifications/apps/icd/icd10online/
- Pereira da Silva, T., & Justo da Silva, L. (2010). Escalas de avaliação da dor utilizadas no recém-nascido: Revisão Sistemática. *Acta Médica Portuguesa*, 23(3), 437-454.
- Pinelli, J., & Symington, A. (2005). Non-nutritive sucking for promoting physiologic stability and nutrition in preterm infants. *Cochrane Database of Systematic Reviews*, 4.
- Pridham, K.F., Sondel, S., Clark, R., Green, C., & Brown, R. (1994). *Correlates of preterm and term infants feeding outcomes*. Final Report to the National Institute for Nursing Research.
- Pridham, K.F., Limbo, R., Schroeder, M., Thoyre, S., & Van Riper, M. (1998). Guided participation and development of care-giving competencies for families of low birth-weight infants. *Journal of Advances in Nursing*, 28, 948-958.
- Pridham, K.F., Sondel, S., Clark, R., Green, C., & Brown, R. (1999). Clinician help for mothers of infants with lung disease. Final Report to the Maternal Child Health Bureau.
- Ray, J.G., Singh, G., & Burrow, R.F. (2004). Evidence for suboptimal use of periconceptional folic acid supplements globally. *BJOG: An International Journal of Obstetrics and Gynaecology*, 111(5), 399-408.
- Riberto, M., Miyazaki, M.H., Jucá, S.S.H., Sakamoto, H., Pinto, P.P.N., & Battistella, L.R. (2004). Validação da versão brasileira da Medida de Independência Funcional. *Acta Fisiátrica*, 11(2), 72-76.
- Rutter, N. (1987). Percutaneous drug absorption in the newborn: Hazards and uses. *Clinics in Perinatology*, 14(4), 911-930.

- Sanada, H., Miyachi, Y., Ohura, T., Moriguchi, T., Tokunaga, K., Shido, K., & Nakagami, G. (2008). The Japanese pressure ulcer surveillance study: a retrospective cohort study to determine prevalence of pressure ulcers in Japanese hospitals. *Wounds*, 20, 176-182.
- Sandler, A. (2004). *Living with Spina Bifida. A Guide for Families and Professionals*. North Carolina, USA: The University of North Carolina Press.
- Sedin, G., Hammarlund, K., Nilsson, G.E., Strömberg, B., & Oberg, P.A. (1985). Measurements of transepidermal water loss in newborn infants. *Clinics in Perinatology*, 12(1), 79-99.
- Shaker, C.S. (1992). *Enhancing safety during nipple feeding: An individualized approach*. Paper presented at the conference. Developmental Interventions in Neonatal Care, Chicago, IL, October 21, 1992.
- Shaker, C.S. (1997). Behavioral state activity during nipple feedings for preterm infants. *Neonatal Network*, 16, 38.
- Shaker, C.S. (1999). Nipple feeding preterm infants: an individualized, developmentally supportive approach. *Neonatal Network*, 18, 15-22.
- Shaker, C., & Woida, A.M. (2007). An evidence-based approach to nipple feeding in a level III NICU: nurse autonomy, developmental care, and teamwork. *Neonatal Network*, 26, 77-83.
- Sperle, P.A., Orrenbacher, K.J., Braun, S.L., Lane, S.J., & Nochajski, S. (1997). Equivalence Reliability of the Functional Independence Measure for Children (WeeFIM®) Administration Methods. *The American Journal of Occupational Therapy*, 51(1), 35-41.
- Stevens, B., Yamada, J., Beyene, J., Gibbins, S., Petryshen, P., Stinson, J., & Narciso, J. (2005). Consistent management of repeated procedural pain with sucrose in preterm neonates: Is it effective and safe for repeated use over time? *Clinical Journal of Pain*, 21, 543-548.
- Stevens, D.L., Bisno, A.L., Chambers, H.F., Everett, E.D., Dellinger, P., Goldstein, E.J.C., Gorbach, S.L., Hirschmann, J.V., Kaplan, E.L., Montoya, J.G., & Wade, J.C. (2005). Practice Guidelines for the Diagnosis and Management of Skin and Soft-Tissue Infections. *Clinical Infectious Disease*, 41(10), 1373-1406.
- Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistic and Data Analysis*, 54, 37-48.
- Todorov, V., & Filzmoser, P. (2009). An Object-Oriented Framework for Robust Multivariate Analysis. *Journal of Statistical Software*, 32, 1-47.
- Vaivre-Douret, L., Ennouri, K., Jrad, I., Garrec, C., & Papiernik, E. (2004). Effect of positioning on the incidence of abnormalities of muscle tone in low-risk, preterm infants. *European Journal of Paediatric Neurology*, 8(1), 21-34.
- Vaivre-Douret, L., & Golse, B. (2007). Comparative effects of 2 positional supports on neurobehavioral and postural development in preterm neonates. *Journal of Perinatal & Neonatal Nursing*, 21 (4), 323-330.
- Van Aelst, S., & Willems, G. (2011). Robust and Efficient One-Way MANOVA Tests. *Journal of the American Statistical Association*, 106(494),706-718.
- Wachman, E.M., & Lahav, A. (2010). The effects of noise on preterm infants in the NICU. *Archives of Disease in Childhood, Fetal and Neonatal Edition*. 96(4), 306-309.
- Retrieved from: http://brighamandwomens.org/Research/depts/newbornmedicine/researchers/lahav/documents/Wachman&Lahav_2010_Effects_of_noise_on_premature_infants.pdf.
- Wachman, E.M., & Lahav, A. (2011). The effects of noise on preterm infants in the NICU. *Archives of Disease in Childhood Fetal and Neonatal Edition*, 96(4), 305-309.

- Wald, N.J., Law, M.R., Morris, J.K., & Wald, D.S. (2001). Quantifying the effect of folic acid. *Lancet*, 358(9298), 2069-2073.
- Wheeler, B.J. (2014). Promoção da Saúde do Recém-Nascido e da Família. In M.J. Hockenberry & D. Wilson. *Wong: Enfermagem da Criança e do Adolescente*, 9ª Ed. Vol. I (Cap. 1-25). Loures. Lusociência, 240-287.
- White, R.D. (2007). Recommended Standards for Newborn ICU Design. *Report of the Seventh Census Conference on Newborn ICU Design*. Retrieved from: <http://www.nd.edu/nicudes/index.html>.
- White, R. D. (2010). Single-Family Room Design in the Neonatal Intensive Care Unit- Challenges and Opportunities. *Newborn and Infant Nursing Reviews*, 10 (2), 83-86.
- World Health Organization (WHO) (2002). *Innovative Care for Chronic Conditions: Building Blocks for Action*. Geneve, WHO.
- World Health Organization (WHO) (2005). Preventing Chronic Disease, a vital investment. Geneve, WHO. Consultado em 23 de dezembro, 2013: Disponível em http://whqlibdoc.who.int/publications/2005/9241563001_eng.pdf.
- World Health Organization (WHO) (2007). WHO: School health and youth health promotion. Consultado em 23 de dezembro, 2013. Disponível em www.who.int/school_youth_health/en.
- Wound Ostomy and Continence Nursing Society (WOCNS) (2011). Society Position Statement: Pressure Ulcer Staging. Consultado em 26 de março, 2014: Disponível em [http://c.ymcdn.com/sites/www.wocn.org/resource/collection/E3050C1AFBF0-44ED-B28B-C41E24551CCC/Position_Statement_Pressure_Ulcer_Staging_\(2011\).pdf](http://c.ymcdn.com/sites/www.wocn.org/resource/collection/E3050C1AFBF0-44ED-B28B-C41E24551CCC/Position_Statement_Pressure_Ulcer_Staging_(2011).pdf).
- Zonta, M.B., Júnior, A.R., & Santos, L.H.C. (2011). Avaliação Funcional na Paralisia Cerebral. *Acta Pediátrica Portuguesa*, 42(1), 27-32.
- Zwick, R. (1985). Nonparametric One-Way Multivariate Analysis of Variance: A Computational Approach Based on the Pillai-Bartlett Trace. *Psychological Bulletin*, 47(1), 148-152.
- Wachman, E.M., & Lahav, A. (2011). The effects of noise on preterm infants in the NICU. *Archives of Disease in Childhood Fetal and Neonatal Edition*, 96(4), 305-309.

PARTE V: Discussão Geral

Introdução

Neste último capítulo da discussão geral dos resultados iremos dar relevo à importância da medida na prática de cuidados e na investigação em saúde. Neste âmbito salientamos as práticas baseadas na evidência, considerando que a investigação e a utilização dos seus resultados é o pilar da melhoria efetiva da qualidade de cuidados.

O estudo das populações pediátricas e neonatais e a observação subjetiva da saúde destas populações teve como finalidade mostrar a pertinência da avaliação padronizada de comportamentos, atitudes ou traços de personalidade com instrumentos denominados escalas de avaliação com formato de medida ordinal.

Na área da saúde, os investigadores e os profissionais confrontam-se, demasiadas vezes, com problemas relacionados com este tipo de instrumentos. Os investigadores, por um lado, apropriam-se da discussão ao nível da investigação empírica e da revisão sistemática da literatura, em que se discutem metodologias de investigação e análise de dados ordinais, com o objetivo de avaliar a adequação das técnicas estatísticas, aos problemas a investigar, às variáveis e à dimensão das amostras. Os profissionais de saúde, por outro lado, apropriam-se dos resultados da investigação no sentido de melhorar a intervenção clínica, colaborando nos grupos de investigação e selecionando os instrumentos mais adequados aos seus contextos e aos seus clientes.

Tendo como suporte o estudo de simulação desenvolvido no capítulo VII, evidenciamos a análise da probabilidade de erro de tipo I e da potência de teste das 4 estatísticas da MANOVA paramétrica e 2 estatísticas da MANOVA não paramétrica com dados ordinais e deixamos algumas sugestões que confirmam o trabalho desenvolvido por outros investigadores e apresentam outros aspetos que podem ser uma mais-valia para quem trabalha com este tipo de instrumentos. A análise da potência levanta sempre problemas do foro teórico e do foro prático, o que pode ser agravado pelo facto de não se calcular uma medida de efeito para as estatísticas da MANOVA não paramétrica.

Faremos ainda a análise crítica de alguns resultados que se encontram na literatura científica e sobre os resultados que emergiram da utilização dos dados das escalas de avaliação (EFS, NSRAS, MIF) que permitiram o enquadramento dos mesmos no estudo de simulação.

Iniciamos a discussão geral com as implicações teóricas seguidas das implicações práticas.

Capítulo IX. Discussão Geral dos Resultados

9.1 Implicações Teóricas

O recurso à estatística como ferramenta facilitadora da tomada de decisão na investigação faz com que o investigador se depare com a problemática do erro, mesmo antes de iniciar a realização dos testes estatísticos. Este problema faz com que cada vez mais os investigadores sintam a necessidade de adquirir competências em algumas áreas da estatística para mais facilmente analisarem e interpretarem os resultados da investigação.

A utilização de técnicas multivariadas é muito frequente na área das ciências sociais e ciências da saúde, sendo uma das mais utilizadas a MANOVA paramétrica. Esta técnica permite explorar simultaneamente as relações entre variáveis independentes categóricas e 2 ou mais variáveis dependentes, porém a sua aplicação exige a não violação dos pressupostos (distribuição normal multivariada das variáveis dependentes em todos os grupos e homogeneidade de matrizes de variância-covariância dos grupos), o que nem sempre é cumprido pelas mais diferentes razões, e nem sempre justificado. Apesar de menos utilizadas, encontram-se na literatura científica duas estatísticas da MANOVA não paramétrica (lambda de Wilks e traço de Pillai aplicados às ordens), que utilizamos neste trabalho, as quais podem ser uma solução no caso de violação dos pressupostos dos testes da MANOVA paramétrica e que facilmente os investigadores podem aplicar.

O estudo da performance da MANOVA paramétrica e da MANOVA não paramétrica está associado a 2 dos erros mais comuns em inferência estatística e que estão relacionados com tomadas de decisão incorretas na rejeição ou não da hipótese nula. Se, o investigador rejeita a hipótese nula sendo ela verdadeira, estamos perante um erro de tipo I ou se não rejeita a hipótese nula sendo ela falsa, estamos perante o erro de tipo II.

No que concerne ao controlo destes dois erros, o erro de tipo I é controlado à partida, através do nível de significância determinado pelo investigador. Este especifica o risco que pode correr na rejeição da hipótese nula, podendo considerar como valor critério para a sua rejeição, diferentes valores, dependendo do tipo de estudo. Na área da saúde utiliza-se muitas vezes como valor critério 1% e/ou 5%.

O erro de tipo II é mais difícil de controlar devido à sua relação com a potência de teste. Se considerarmos que os determinantes da potência de um teste estatístico são o nível de significância, a dimensão da amostra e a dimensão do efeito, a potência de teste apresentará valores tanto mais elevados quanto maiores forem o nível de significância, a dimensão da amostra e a dimensão do efeito (Loureiro & Gameiro, 2011; Marôco, 2010), o que faz com que o seu controlo seja mais difícil. A estimação da potência de teste está intimamente ligada a estes dois erros e começou a ser estudada nos anos trinta do século passado por Pearson (1930, 1931), tendo a sua introdução na área das ciências sociais e humanas ocorrido 3 décadas depois (Cohen, 1962, 1988).

Os estudos sobre probabilidade de erro de tipo I e II e estimativa da potência de teste enfatizavam a importância da simulação estatística como ferramenta fundamental (Blair & Higgins, 1985; Lehman, 1975) para estudar este tipo de problemas, pois não havendo possibilidade de obter analiticamente determinados resultados a solução passa pela simulação (ferramenta essencial aos investigadores e que se foi consolidando à medida que os computadores foram evoluindo).

No estudo de simulação através do Método de Monte Carlo, com o objetivo de “comparar a performance, em termos de potência e probabilidade de erro de tipo I, da MANOVA paramétrica e não paramétrica no caso em que o vetor resposta é constituído por dados ordinais”, desenvolvido neste trabalho, foram incluídos no delineamento do estudo, escalas ordinais com 3, 4, 5 e 7 pontos; diferentes distribuições de frequência (simétrica, assimétrica positiva e negativa e uniforme); correlações de diferente magnitude (baixa, média e elevada); 3 variáveis dependentes, uma variável independente, com 3 níveis ou grupos; amostras com diferentes dimensões ($n=30, 60, 90, 120, 240, 300$ e $n_i=10, 20, 30, 60, 80, 100$), 4 estatísticas da MANOVA paramétrica e 2 da MANOVA não paramétrica com a finalidade de construir diferentes cenários que permitissem mais facilmente o enquadramento de dados reais, e assim responder a algumas dúvidas que emergem nos estudos quando o investigador usa variáveis ordinais correlacionadas.

Os resultados permitiram observar que a maior raiz de Roy apresenta uma probabilidade de erro de tipo I muito superior ao nível de significância nominal ($\alpha = 0.05$) enquanto as restantes estatísticas da MANOVA paramétrica e da não paramétrica (traço de Pillai, lambda de Wilks, traço de Hotelling e traço de Pillai e

Lambda de Wilks aplicados às ordens) apresentam de forma geral, valores inferiores ou, em alguns casos, marginalmente superiores ao nível de significância nominal.

No que concerne aos resultados das potências de teste da MANOVA paramétrica e da MANOVA não paramétrica, o estudo de simulação mostra três situações distintas: na primeira situação, seja qual for o cenário as potências são baixas para todos os testes; na segunda situação as potências são elevadas na maioria dos cenários em que as amostras são iguais ou superiores a 60 e nas escalas com 3, 4 e 5 pontos e, nas escalas com 7 pontos as potências são elevadas para amostras com $n=120$ observações; na terceira situação, as potências podem apresentar uma maior ou menor magnitude. Tendo por base a distribuição de frequência, podemos dizer que o número de pontos na escala, a intensidade das correlações entre os itens e a dimensão da amostra influenciam a potência dos testes, ou seja, quanto menor a dimensão da amostra, maior o número de pontos da escala e maior a magnitude de correlação entre itens, menor a potência dos testes, emergindo os testes não paramétricos com potências mais elevadas.

Na primeira situação, o facto de as potências serem baixas para todos os testes da MANOVA deve-se ao facto de a magnitude das diferenças entre grupos ser tão pequena que a MANOVA não identifica as diferenças, o que é reforçado pelo sinal positivo e pela elevada magnitude das correlações entre o valor observado da estatística de teste de Kruskal-Wallis média, das três variáveis e, a potência dos testes da MANOVA. Na segunda situação, em que as potências das estatísticas da MANOVA paramétrica e da MANOVA não paramétrica são elevadas ou aceitáveis a decisão do investigador depende especificamente da dimensão da amostra e do número de pontos da escala, pelo que podem ser usados testes paramétricos e não paramétricos, o que valida a opção de muitos investigadores, pela utilização de testes paramétricos, em detrimento dos testes não paramétricos. Na terceira situação com magnitudes de potência mais ou menos elevadas dependendo do número de pontos da escala, da intensidade das correlações entre os itens, da dimensão da amostra e do tipo de testes. Os testes não paramétricos apresentam potências de magnitude mais elevada com amostras de dimensão inferior a 90 observações e correlações de média magnitude (nas escalas com 3, 4 e 5 pontos) e inferiores a 240 observações e correlações de média magnitude (nas escalas com 7 pontos) tornando-se por isso uma boa opção do investigador. Com amostras de maior dimensão, os testes paramétricos assemelham-se aos não paramétricos em termos de magnitude de potência, logo passam a ser também uma boa opção para o investigador.

9.2 Implicações Práticas

O processo de investigação inicia-se com uma fase exploratória em que o investigador questiona aspetos relativos ao objeto de estudo, aos pressupostos da investigação, às teorias principais e auxiliares de pesquisa, às metodologias a utilizar e às técnicas de recolha e análise de dados. Ao iniciar a análise dos dados o investigador pode fazer o confronto da teoria com os resultados da investigação e interpretar os resultados, faz ou não a confirmação dos pressupostos da pesquisa, e verifica se as questões formuladas obtiveram resposta e se os objetivos foram atingidos na totalidade ou em parte.

Ao longo do processo de investigação, a estatística está envolvida no planeamento dos estudos, na colheita de informação e no processamento e análise da informação. Podemos por isso dizer, que a estatística é essencial ao processo de pesquisa, seja na definição das técnicas de recolha de dados, na análise dos dados, na descrição de fenómenos ou na contestação ou validação das hipóteses sob estudo, pelo que investigadores e profissionais devem possuir o conhecimento necessário e suficiente para poderem analisar dados e interpretar resultados da investigação.

A medição dos constructos que constituem as escalas de avaliação (Hill & Hill, 2009) tem levantado alguns problemas do ponto de vista da análise estatística, pelo facto de a sua mensuração ser feita por intermédio de variáveis manifestas (não diretamente observáveis) cujo formato de medida ordinal pode apresentar diferentes números de pontos na escala de medida. Apesar da problemática enunciada, os instrumentos desenvolvidos com este tipo de variáveis permitem observar comportamentos num processo de heteroavaliação ou de autoavaliação que se tem revelado como uma mais-valia na avaliação da saúde individual de clientes, e especificamente de crianças e jovens, em contexto social e hospitalar. As escalas de avaliação/observação (Hill & Hill, 2009) são um exemplo muito comum dos instrumentos utilizados em neonatologia e em pediatria. Estes instrumentos apresentam-se com diferentes objetivos de observação, os quais estão relacionados com a população alvo, tipo de escalas, número de itens, formato de medida e número de pontos na escala de medida. O desenvolvimento destes instrumentos e/ou a sua adaptação para a língua portuguesa e validação estatística são fundamentais para aplicação nas práticas de cuidados, pois só depois de validadas estatisticamente devem ser aplicadas nas diferentes populações.

Ao analisarmos os dados que emergem destes instrumentos sabemos que a dimensão da amostra influencia de forma determinante essa análise. Assim, o investigador ao reportar os resultados em que foram usados testes de hipóteses deve estar atento à dimensão da amostra, ao nível de significância, a medidas de dimensão de efeito e à potência do teste (Curado, Teles e Marôco, 2013, 2014; Marôco 2010).

Na prática todos estes aspetos são valorizados, mas importa também valorizar de forma efetiva o número de pontos da escala de medida dos itens, a magnitude da correlação entre os itens e a distribuição de frequência nos grupos, para que se tome a melhor decisão sobre o tipo de testes a aplicar, pois os resultados das potências dependem de todos estes aspetos, conforme verificámos no estudo de simulação (capítulo VII). Considerando alguns autores que potências de nível elevado são ≥ 0.80 (Marôco, 2010; Hill & Hill, 2009) e potências de nível aceitável são superiores a 0.50 (Land & Hui, 2010; Hill & Hill, 2009), deixa-se ao critério do investigador a aceitação ou não níveis menos elevados, dependendo do tipo de estudo.

Na aplicação prática com dados que surgem de três escalas de avaliação, EFS (formato de medida com 3 pontos) (Thoyre, 2007; Thoyre, Shaker, & Pridham, 2005; Thoyre & Brown, 2004), NSRAS (formato de medida com 4 pontos) (Huffines & Martins, Curado & Marçal, 2014; Logsdon, 1997) e MIF (formato de medida com 5 e 7 pontos) (Malheiro, Gaspar & Barradas, 2014; Malheiro, 2005; Laíns, 1991) foi feito o enquadramento dos resultados no estudo de simulação. Os resultados obtidos com a EFS foram analisados tendo em atenção o número de pontos da escala, pois alguns autores consideram que escalas com 3 pontos violam o pressuposto de continuidade (Katz & McSweeney, 1980). Mas como esta escala e outras com o mesmo número de pontos têm grande aplicação nas práticas de cuidados, pelo facto de serem de fácil utilização, a sua análise pode servir para alertar os investigadores para os problemas relacionados com escalas com este número de pontos. De acordo com Katz & McSweeney (1980), as escalas com formato de medida com 4, 5 e 7 pontos cumprem o pressuposto de continuidade e são escalas também amplamente aplicadas nas práticas de cuidados. No entanto, o enquadramento dos dados que emergiram da aplicação, das 4 escalas, no estudo de simulação permite dizer que, houve menor dificuldade na integração dos resultados das escalas com 4 e 5 pontos do que na integração das escalas com 3 e 7 pontos.

9.3 Referências Bibliográficas

- Blair, R.C., & Higgins, J.J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon' signed-rank test under various population shapes. *Psychological Bulletin*, 97, 119-128.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum.
- Curado, M.A.S., Teles, J., & Marôco, J. (2013). Análisis estadístico de escalas ordinales. Aplicaciones en el Área de Salud infantil y Pediatría. *Enfermería Global*, 30, 446-457. <http://www.um.es/eglobal>.
- Curado, M.A.S., Teles, J. & Marôco, J. (2014). Analysis of variables that are not directly observable: influence on decision-making during the research process, *Revista da Escola de Enfermagem da Universidade de S. Paulo*, 48, 1, 146-52.
- Hill, M.M., & Hill, A. (2009). *Investigação por Questionário*. Lisboa: Edições Sílabo Lda.
- Huffines, B. & Lodgson M.C. (1997). The Neonatal Skin Risk Assessment Scale for predicting skin breakdown in neonates. *Issues Comprehensive Pediatrics Nursing*, 20, 103-114.
- Katz B.M., & McSweeney, M. (1980). A Multivariate Kruskal-Wallis Test with Post Hoc Procedures. *Multivariate Behavioral Research*, 15, 281-297.
- Laíns, J. (1991). Guia para o uso do Banco Uniformizado de Dados para Reabilitação Médica – MIF. Versão Portuguesa de Guide for the Uniform Data System for Medical Rehabilitation, FIM (1984). Produção Sandoz.
- Lehmann, E.L. (1975). *Nonparametrics, statistical methods based on ranks*. San Francisco: Holden-Day.
- Loureiro, L.M.J., & Gameiro, M.G.H. (2011). Interpretação crítica dos resultados estatísticos: para lá da significância estatística. *Revista de Enfermagem Referência*, 3, 151-162.
- Malheiro, M.I.D.C. (2005). *Integração/Inclusão de Jovens com Spina Bífida no Ensino Regular... A Voz da Experiência*. Lisboa: Faculdade de Motricidade Humana, Dissertação com vista à obtenção do grau de Mestre em Educação Especial.
- Malheiro, M.I.C., Gaspar, M.F.M., & Barros, L. (2013). The effects of a therapeutic camp: program for promotion of self-management of disability on functionality and self-concept of children with spina bífida. In *Abstracts Book of 57th Annual Meeting of Society for Research into Hydrocephalus and Spina Bífida*, 26-29 June, paper 11.
- Marôco, J. (2010). *Análise Estatística com SPSS*. Pêro Pinheiro: Report Number, Lda.
- Martins, C.O.A., Curado; M.A.S., Marçal, T. O. (2013). Adaptação transcultural e validação da “Neonatal Skin Risk Assessment Scale” (NSRAS) com recém-nascidos portugueses (no prelo).
- Pearson, E.S. (1930). A further development of tests of normality. *Biometrika*, 22, 239-249.
- Pearson, E.S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- R Development Core Team (2012). *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (<http://www.R-project.org>).

Thoyre, S.M. (2007). Feeding outcomes of extremely premature infants after neonatal care. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 36, 366-76.

Thoyre, S.M., Shaker, C., & Pridham, K.F. (2005). The Early Feeding Skills Assessment for preterm infants. *Neonatal Network*, 24, 2-10.

Thoyre, S.M., & Brown, R. (2004). Factors contributing to preterm infant engagement during bottle feeding. *Nursing Research*, 53, 304-313.

APÊNDICES

Apêndice I: Programas em R

Geração aleatória de dados ordinais: *Generation of Artificial Ordinal and Binary Data*, ("orddata", repos="http://R-Forge.R-project.org")>library("orddata"), com as funções principais *rmvord* e *rmvord_b*.

###POTÊNCIA DOS TESTES###

```
library("orddata")
library("MASS")
library("sm")
library("psych")
#####
#####
### Parametrização das simulações ###
k<-3          # N° de grupos
p<-3          # n° de variáveis dependentes
ndim<-300
r=.1          # Correlação
Lik=3
nrep<-1000
Y<-matrix(nrow=ndim,ncol=3)
RY<-matrix(nrow=ndim,ncol=3)
Pres<-matrix(nrow=nrep,ncol=2)
Wres<-matrix(nrow=nrep,ncol=2)
Hres<-matrix(nrow=nrep,ncol=2)
Rres<-matrix(nrow=nrep,ncol=2)
PNPres<-matrix(nrow=nrep,ncol=3)
WNPres<-matrix(nrow=nrep,ncol=3)
ResPar<-matrix(nrow=3,ncol=4)
countP<-numeric(nrep)
countW<-numeric(nrep)
countH<-numeric(nrep)
countR<-numeric(nrep)
countPNP<-numeric(nrep)
countWNP<-numeric(nrep)
###probabilidades marginais###
p1<-c(20, 60, 20)/100
p2<-c(70, 20, 10)/100
p3<-c(10, 20, 70)/100
p4<-c(34, 33, 33)/100
### correlações ###
matcor1<-c(1, r, r, r,1, r, r, r,1)
matcor2<-c(1, r, r, r,1, r, r, r,1)
matcor3<-c(1, r, r, r,1, r, r, r,1)
###0 X fora do ciclo###
x1<-rep(1,ndim/k)
x2<-rep(2,ndim/k)
x3<-rep(3,ndim/k)
X<-c(x1,x2,x3)
k1<-1
for(k1 in 1:nrep){
Cor1 <- matrix(matcor1, 3, 3)
###Geração de vectores resposta###
i<-1
for (i in 1:ndim){
if (X[i]==1){
probs1 <- list(p1, p1, p1)
Cor1 <- matrix(matcor1,3,3)
Y[i,]<-rmvord_b(n=1, probs = probs1, Cor = Cor1)
}
if (X[i]==2){
probs2 <- list(p2, p2, p2)
```

```

Cor2 <- matrix(matcor1,3,3)
Y[i,]<-rmvord_b(n=1, probs = probs2, Cor = Cor2)
}
if (X[i]==3){
probs3 <- list(p3, p3, p3)
Cor3 <- matrix(matcor1,3,3)
Y[i,]<-rmvord_b(n=1, probs = probs3, Cor = Cor3)
}
}
nome=paste("dadosY",k1)
write.matrix(Y,nome)

###manova parametrica###
# parametricos
fit<-manova(Y~as.factor(X))
P<-summary(fit, test="Pillai",intercept=TRUE)
Pres[k1,1]<-P$stats[5]
Pres[k1,2]<-P$stats[17]
W<-summary(fit, test="Wilks",intercept=TRUE)
Wres[k1,1]<-W$stats[5]
Wres[k1,2]<-W$stats[17]
H<-summary(fit, test="Hotelling-Lawley",intercept=TRUE)
Hres[k1,1]<-H$stats[5]
Hres[k1,2]<-H$stats[17]
R<-summary(fit, test="Roy",intercept=TRUE)
Rres[k1,1]<-R$stats[5]
Rres[k1,2]<-R$stats[17]
ResPar[1,]<-c(P$stats[5],W$stats[5],H$stats[5],R$stats[5])
ResPar[2,]<-c(P$stats[17],W$stats[17],H$stats[17],R$stats[17])
eta2pP=(P$stats[11]*P$stats[8])/((P$stats[11]*P$stats[8])+P$stats[14])
eta2pW=(W$stats[11]*W$stats[8])/((W$stats[11]*W$stats[8])+W$stats[14])
eta2pH=(H$stats[11]*H$stats[8])/((H$stats[11]*H$stats[8])+H$stats[14])
eta2pR=(R$stats[11]*R$stats[8])/((R$stats[11]*R$stats[8])+R$stats[14])
ResPar[3,]<-c(eta2pP,eta2pW,eta2pH,eta2pR)
ResPar
# Rankings #
j<-1
for (j in 1:3){
  RY[,j]<-rank(Y[,j])
}
##Manova Não parametrica McSweeney - Pillai##
fitNP<-manova(RY~as.factor(X))
PNP<-summary(fitNP, test="Pillai",intercept=TRUE)
x2PNP=(ndim-1)*PNP$stats[5]
pvaluePNP<-pchisq(x2PNP,p*(k-1),lower.tail=FALSE)
PNPres[k1,1]<-PNP$stats[5]
PNPres[k1,2]<-x2PNP
PNPres[k1,3]<-pvaluePNP
##Manova Não parametrica Nath e Pavur - Wilks##
WNP<-summary(fitNP, test="Wilks",intercept=TRUE)
x2WNP=- (ndim-1- (p+k) /2) *log(WNP$stats[5])
pvalueWNP<-pchisq(x2WNP,p*(k-1),lower.tail=FALSE)
WNPres[k1,1]<-WNP$stats[5]
WNPres[k1,2]<-x2WNP
WNPres[k1,3]<-pvalueWNP
# cálculo da potência
if (Pres[k1,2]<0.05) countP[k1]=countP[k1]+1
if (Wres[k1,2]<0.05) countW[k1]=countW[k1]+1
if (Hres[k1,2]<0.05) countH[k1]=countH[k1]+1
if (Rres[k1,2]<0.05) countR[k1]=countR[k1]+1
if (PNPres[k1,3]<0.05) countPNP[k1]=countPNP[k1]+1

```

```

if (WNPres[k1,3]<0.05) countWNP[k1]=countWNP[k1]+1
k1<-k1+1
}
powerP<-round(sum(countP)/nrep,5)*100
powerW<-round(sum(countW)/nrep,5)*100
powerH<-round(sum(countH)/nrep,5)*100
powerR<-round(sum(countR)/nrep,5)*100
powerPNP<-round(sum(countPNP)/nrep,5)*100
powerWNP<-round(sum(countWNP)/nrep,5)*100
powers<-c(powerP,powerW,powerH,powerR,powerPNP,powerWNP)
#Matrizes dos Testes#
write.matrix(Pres,"ficheiro_1.txt")
write.matrix(Wres,"ficheiro_2.txt")
write.matrix(Hres,"ficheiro_3.txt")
write.matrix(Rres,"ficheiro_4.txt")
write.matrix(PNPres,"ficheiro_5.txt")
write.matrix(WNPres,"ficheiro_6.txt")
write.matrix(powers,"ficheiro_7.tx
#Count p<.05
respP<-data.frame(countP,countW,countH,countR,countPNP,countWNP)
colnames(respP) <- c('P','W','H','R','PNP','WNP')
write.matrix(respP,"ficheiro_8.txt")
##### Gráficos #####
jpeg(filename = "plp2p3_n300_cor01_lik3_1.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
error.bars(respP,alpha=.05,main= paste("Mean with 95%
CI","n=",ndim/k,"r=",r,"Lik=",3), ylab="Power (%)", xlab="Method",
ylim=c(0,1),bars=TRUE)
dev.off()
jpeg(filename = "plp2p3_n300_cor01_lik3_2.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
error.bars(respP,alpha=.05,main= paste("Mean with 95%
CI","n=",ndim/k,"r=",r,"Lik=",3), ylab = "Power (%)", xlab="Method",
ylim=c(0,1))
dev.off()
###create value labels###
jpeg(filename = "plp2p3_n300_cor01_lik3_3.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
mdat<-
data.frame(Pres[,2],Wres[,2],Hres[,2],Rres[,2],PNPres[,3],WNPres[,3])
names(mdat)[1]<-paste("Pillai")
names(mdat)[2]<-paste("Wilks")
names(mdat)[3]<-paste("Hotelling")
names(mdat)[4]<-paste("Roy")
names(mdat)[5]<-paste("PNP")
names(mdat)[6]<-paste("WNP")
mdat=stack(mdat)
###plot densities###
sm.density.compare(mdat$values,mdat$ind,xlab="p", main="Density
plots")
#legenda no canto superior direito do grafico
colfill<-c(0:(2+length(levels(mdat$ind))))
legend("topright", levels(mdat$ind),fill=colfill)
dev.off()

```

###Probabilidade de Erro de Tipo I###

```
library("orddata")
library("MASS")
library("sm")
library("psych")
#####
### Parametrização das simulações ###
k<-3          # N° de grupos
p<-3          # n° de variáveis dependentes
ndim<-30
r=.1         # Correlação
Lik=3
nrep<-1000
Y<-matrix(nrow=ndim,ncol=3)
RY<-matrix(nrow=ndim,ncol=3)
Pres<-matrix(nrow=nrep,ncol=2)
Wres<-matrix(nrow=nrep,ncol=2)
Hres<-matrix(nrow=nrep,ncol=2)
Rres<-matrix(nrow=nrep,ncol=2)
PNPres<-matrix(nrow=nrep,ncol=3)
WNPres<-matrix(nrow=nrep,ncol=3)
ResPar<-matrix(nrow=3,ncol=4)
countP<-numeric(nrep)
countW<-numeric(nrep)
countH<-numeric(nrep)
countR<-numeric(nrep)
countPNP<-numeric(nrep)
countWNP<-numeric(nrep)
###probabiliddades marginais###
p1<-c(20, 60, 20)/100
p2<-c(70, 20, 10)/100
p3<-c(10, 20, 70)/100
p4<-c(34, 33, 33)/100
### correlações ###
matcor1<-c(1, r, r, r,1, r, r, r,1)
matcor2<-c(1, r, r, r,1, r, r, r,1)
matcor3<-c(1, r, r, r,1, r, r, r,1)
###0 X fora do ciclo###
x1<-rep(1,ndim/k)
x2<-rep(2,ndim/k)
x3<-rep(3,ndim/k)
X<-c(x1,x2,x3)
k1<-1
for(k1 in 1:nrep){
  Cor1 <- matrix(matcor1, 3, 3)
  ###Geração de vectores resposta###
  i<-1
  for (i in 1:ndim){
    if (X[i]==1){
      probs1 <- list(p1, p1, p1)
      Cor1 <- matrix(matcor1,3,3)
      Y[i,]<-rmvord(n=1, probs = probs1, Cor = Cor1)
    }
    if (X[i]==2){
      probs2 <- list(p1, p1, p1)
      Cor2 <- matrix(matcor1,3,3)
      Y[i,]<-rmvord(n=1, probs = probs2, Cor = Cor2)
    }
    if (X[i]==3){
      probs3 <- list(p1, p1, p1)
    }
  }
}
```

```

Cor3 <- matrix(matcor1,3,3)
Y[i,]<-rmvord(n=1, probs = probs3, Cor = Cor3)
}
}
nome=paste("dadosY",k1)
write.matrix(Y,nome)
###manova parametrica###
# parametricos
fit<-manova(Y~as.factor(X))
P<-summary(fit, test="Pillai",intercept=TRUE)
Pres[k1,1]<-P$stats[5]
Pres[k1,2]<-P$stats[17]
W<-summary(fit, test="Wilks",intercept=TRUE)
Wres[k1,1]<-W$stats[5]
Wres[k1,2]<-W$stats[17]
H<-summary(fit, test="Hotelling-Lawley",intercept=TRUE)
Hres[k1,1]<-H$stats[5]
Hres[k1,2]<-H$stats[17]
R<-summary(fit, test="Roy",intercept=TRUE)
Rres[k1,1]<-R$stats[5]
Rres[k1,2]<-R$stats[17]
ResPar[1,]<-c(P$stats[5],W$stats[5],H$stats[5],R$stats[5])
ResPar[2,]<-c(P$stats[17],W$stats[17],H$stats[17],R$stats[17])
eta2pP=(P$stats[11]*P$stats[8])/((P$stats[11]*P$stats[8])+P$stats[14])
eta2pW=(W$stats[11]*W$stats[8])/((W$stats[11]*W$stats[8])+W$stats[14])
eta2pH=(H$stats[11]*H$stats[8])/((H$stats[11]*H$stats[8])+H$stats[14])
eta2pR=(R$stats[11]*R$stats[8])/((R$stats[11]*R$stats[8])+R$stats[14])
ResPar[3,]<-c(eta2pP,eta2pW,eta2pH,eta2pR)
ResPar
# Rankings #
j<-1
for (j in 1:3){
  RY[,j]<-rank(Y[,j])
}
##Manova Não parametrica McSweeney - Pillai##
fitNP<-manova(RY~as.factor(X))
PNP<-summary(fitNP,test="Pillai",intercept=TRUE)
x2PNP=(ndim-1)*PNP$stats[5]
pvaluePNP<-pchisq(x2PNP,p*(k-1),lower.tail=FALSE)
PNPres[k1,1]<-PNP$stats[5]
PNPres[k1,2]<-x2PNP
PNPres[k1,3]<-pvaluePNP
##Manova Não parametrica Nath e Pavur - Wilks##
WNP<-summary(fitNP,test="Wilks",intercept=TRUE)
x2WNP=- (ndim-1- (p+k)/2) *log(WNP$stats[5])
pvalueWNP<-pchisq(x2WNP,p*(k-1),lower.tail=FALSE)
WNPres[k1,1]<-WNP$stats[5]
WNPres[k1,2]<-x2WNP
WNPres[k1,3]<-pvalueWNP
# cálculo da potência
if (Pres[k1,2]<0.05) countP[k1]=countP[k1]+1
if (Wres[k1,2]<0.05) countW[k1]=countW[k1]+1
if (Hres[k1,2]<0.05) countH[k1]=countH[k1]+1
if (Rres[k1,2]<0.05) countR[k1]=countR[k1]+1
if (PNPres[k1,3]<0.05) countPNP[k1]=countPNP[k1]+1
if (WNPres[k1,3]<0.05) countWNP[k1]=countWNP[k1]+1
k1<-k1+1
}
powerP<-round(sum(countP)/nrep,5)*100
powerW<-round(sum(countW)/nrep,5)*100
powerH<-round(sum(countH)/nrep,5)*100

```

```

powerR<-round(sum(countR)/nrep,5)*100
powerPNP<-round(sum(countPNP)/nrep,5)*100
powerWNP<-round(sum(countWNP)/nrep,5)*100
powers<-c(powerP,powerW,powerH,powerR,powerPNP,powerWNP)
#Matrizes dos Testes#
write.matrix(Pres,"ficheiro_1.txt")
write.matrix(Wres,"ficheiro_2.txt")
write.matrix(Hres,"ficheiro_3.txt")
write.matrix(Rres,"ficheiro_4.txt")
write.matrix(PNPres,"ficheiro_5.txt")
write.matrix(WNPres,"ficheiro_6.txt")
write.matrix(powers,"ficheiro_7.txt")
#Count p<.05
respP<-data.frame(countP,countW,counth,countR,countPNP,countWNP)
colnames(respP) <- c('P','W','H','R','PNP','WNP')
write.matrix(respP,"ficheiro_8.txt")
##### Gráficos #####
jpeg(filename = "plplpl_n30_cor01_lik3_1.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
error.bars(respP,alpha=.05,main= paste("Mean with 95%
CI","n=",ndim/k,"r=",r,"Lik=",3), ylab="Type I Error (%)",
xlab="Method", ylim=c(0,1),bars=TRUE)
dev.off()
jpeg(filename = "plplpl_n30_cor01_lik3_2.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
error.bars(respP,alpha=.05,main= paste("Mean with 95%
CI","n=",ndim/k,"r=",r,"Lik=",3), ylab = "Type I Error (%)",
xlab="Method", ylim=c(0,1))
dev.off()
###create value labels###
jpeg(filename = "plplpl_n30_cor01_lik3_3.jpg", width = 480, height =
480,
units = "px", pointsize = 12, quality = 100,
bg = "white", res = NA)
mdat<-
data.frame(Pres[,2],Wres[,2],Hres[,2],Rres[,2],PNPres[,3],WNPres[,3])
names(mdat)[1]<-paste("Pillai")
names(mdat)[2]<-paste("Wilks")
names(mdat)[3]<-paste("Hotelling")
names(mdat)[4]<-paste("Roy")
names(mdat)[5]<-paste("PNP")
names(mdat)[6]<-paste("WNP")
mdat=stack(mdat)
###plot densities###
sm.density.compare(mdat$values,mdat$ind,xlab="p", main="Density
plots")
#legenda no canto superior direito do grafico
colfill<-c(0:(2+length(levels(mdat$ind))))
legend("topright", levels(mdat$ind),fill=colfill)
dev.off()

```

“rela” (<http://cran.r-project.org/web/packages/rela>) calculo das medidas de forma e da estatística de teste de Kruskal-Wallis.

```
resultx1_media1<-0
resultx1_media2<-0
resultx1_media3<-0
resultx2_media1<-0
resultx2_media2<-0
resultx2_media3<-0
resultx3_media1<-0
resultx3_media2<-0
resultx3_media3<-0
```

```
resultx1_sk1<-0
resultx1_sk2<-0
resultx1_sk3<-0
resultx2_sk1<-0
resultx2_sk2<-0
resultx2_sk3<-0
resultx3_sk1<-0
resultx3_sk2<-0
resultx3_sk3<-0
```

```
resultx1_se1<-0
resultx1_se2<-0
resultx1_se3<-0
resultx2_se1<-0
resultx2_se2<-0
resultx2_se3<-0
resultx3_se1<-0
resultx3_se2<-0
resultx3_se3<-0
```

```
resultx1_ku1<-0
resultx1_ku2<-0
resultx1_ku3<-0
resultx2_ku1<-0
resultx2_ku2<-0
resultx2_ku3<-0
resultx3_ku1<-0
resultx3_ku2<-0
resultx3_ku3<-0
```

```
resultx1_kse1<-0
resultx1_kse2<-0
resultx1_kse3<-0
resultx2_kse1<-0
resultx2_kse2<-0
resultx2_kse3<-0
resultx3_kse1<-0
resultx3_kse2<-0
resultx3_kse3<-0
```

```
k1<-0
k2<-0
k3<-0
```

```
for (i in 2:5){
```

```

p<-paste(getwd(),"/dadosY ",i,sep="")
p
dados<-read.table(paste(p))
xx<-as.matrix(dados)
xx
k<-3
n<-120

x1<-xx[1:(n/3),]
x2<-xx[(n/3+1):(2*n/3),]
x3<-xx[(2*n/3+1):n,]
k1[i]<-kruskal.test(list(x1[,1],x2[,1],x3[,1]))$statistic
k2[i]<-kruskal.test(list(x1[,2],x2[,2],x3[,2]))$statistic
k3[i]<-kruskal.test(list(x1[,3],x2[,3],x3[,3]))$statistic

resultx1_medial[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Tendency[1]
resultx1_media2[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Tendency[2]
resultx1_media3[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Tendency[3]

resultx2_medial[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Tendency[1]
resultx2_media2[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Tendency[2]
resultx2_media3[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Tendency[3]

resultx3_medial[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Tendency[1]
resultx3_media2[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Tendency[2]
resultx3_media3[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Tendency[3]

resultx1_sk1[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Skewness[1]
resultx1_sk2[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Skewness[2]
resultx1_sk3[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Skewness[3]

resultx2_sk1[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Skewness[1]
resultx2_sk2[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Skewness[2]
resultx2_sk3[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Skewness[3]

resultx3_sk1[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Skewness[1]
resultx3_sk2[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Skewness[2]
resultx3_sk3[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Skewness[3]

resultx1_se1[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Skewness[,2][1]
resultx1_se2[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Skewness[,2][2]
resultx1_se3[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Skewness[,2][3]

resultx2_se1[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Skewness[,2][1]
resultx2_se2[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Skewness[,2][2]
resultx2_se3[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Skewness[,2][3]

```

```

resultx3_se1[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Skewness[,2][3]
resultx3_se2[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Skewness[,2][3]
resultx3_se3[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Skewness[,2][3]

resultx1_ku1[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Kurtosis[1]
resultx1_ku2[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Kurtosis[2]
resultx1_ku3[i]<-itemanal(x1, SE.par = 1.96, boots = 1000)$Kurtosis[3]

resultx2_ku1[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Kurtosis[1]
resultx2_ku2[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Kurtosis[2]
resultx2_ku3[i]<-itemanal(x2, SE.par = 1.96, boots = 1000)$Kurtosis[3]

resultx3_ku1[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Kurtosis[1]
resultx3_ku2[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Kurtosis[2]
resultx3_ku3[i]<-itemanal(x3, SE.par = 1.96, boots = 1000)$Kurtosis[3]

resultx1_kse1[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Kurtosis[,2][1]
resultx1_kse2[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Kurtosis[,2][2]
resultx1_kse3[i]<-itemanal(x1, SE.par = 1.96, boots =
1000)$Kurtosis[,2][3]

resultx2_kse1[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Kurtosis[,2][1]
resultx2_kse2[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Kurtosis[,2][2]
resultx2_kse3[i]<-itemanal(x2, SE.par = 1.96, boots =
1000)$Kurtosis[,2][3]

resultx3_kse1[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Kurtosis[,2][3]
resultx3_kse2[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Kurtosis[,2][3]
resultx3_kse3[i]<-itemanal(x3, SE.par = 1.96, boots =
1000)$Kurtosis[,2][3]
}
mean(resultx1_media1)
mean(resultx1_media2)
mean(resultx1_media3)
mean(resultx2_media1)
mean(resultx2_media2)
mean(resultx2_media3)
mean(resultx3_media1)
mean(resultx3_media2)
mean(resultx3_media3)

mean(resultx1_sk1)
mean(resultx1_sk2)
mean(resultx1_sk3)
mean(resultx2_sk1)
mean(resultx2_sk2)
mean(resultx2_sk3)
mean(resultx3_sk1)
mean(resultx3_sk2)
mean(resultx3_sk3)

```

```
mean(resultx1_se1)
mean(resultx1_se2)
mean(resultx1_se3)
mean(resultx2_se1)
mean(resultx2_se2)
mean(resultx2_se3)
mean(resultx3_se1)
mean(resultx3_se2)
mean(resultx3_se3)
```

```
mean(resultx1_ku1)
mean(resultx1_ku2)
mean(resultx1_ku3)
mean(resultx2_ku1)
mean(resultx2_ku2)
mean(resultx2_ku3)
mean(resultx3_ku1)
mean(resultx3_ku2)
mean(resultx3_ku3)
```

```
mean(resultx1_kse1)
mean(resultx1_kse2)
mean(resultx1_kse3)
mean(resultx2_kse1)
mean(resultx2_kse2)
mean(resultx2_kse3)
mean(resultx3_kse1)
mean(resultx3_kse2)
mean(resultx3_kse3)
```

```
mean(k1)
mean(k2)
mean(k3)
```

```
#####
```


**Apêndice II: Tabelas com resultados das médias da
Assimetria e Curtose**

Tabela 1 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa ($r=0.10$) intensidade, distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.3873$	$SE_{g_2}=0.7746$	$SE_{g_1}=0.0000$	$SE_{g_2}=0.0000$	$SE_{g_1}=0.2905$	$SE_{g_2}=0.5809$	$SE_{g_1}=0.1959$	$SE_{g_2}=0.3919$
1	1	0.00045	0.00516	0.00000	0.00000	-0.00012	0.00681	-0.00008	0.00406
	2	0.00059	0.00737	0.00000	0.00000	-0.00008	0.00575	-0.00010	0.00534
	3	0.00059	0.00737	0.00000	0.00000	-0.00009	0.00423	-0.00006	0.00515
2	1	0.00063	0.00171	0.00000	0.00000	-0.00008	0.00443	0.00012	0.00499
	2	0.00291	0.01301	0.00000	0.00000	-0.00014	0.00621	0.00013	0.00709
	3	0.00032	0.00069	0.00000	0.00000	-0.00020	0.00554	0.00011	0.00597
3	1	-0.01293	0.01441	0.00000	0.00000	0.00627	0.02669	0.00017	0.03306
	2	-0.01556	0.01844	0.00000	0.00000	0.00529	0.02203	0.00092	0.02726
	3	-0.00319	0.00223	0.00000	0.00000	0.00555	0.02057	0.00034	0.02793

Tabela 2 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.0000$	g_2 $SE_{g_2}=0.0000$	g_1 $SE_{g_1}=0.3652$	g_2 $SE_{g_2}=0.7303$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	0.00000	0.00000	-0.00170	0.00572	-0.00058	0.00682	-0.00014	0.00573
	2	0.00000	0.00000	-0.00233	0.00637	-0.00039	0.00571	-0.00010	0.00489
	3	0.00000	0.00000	-0.00117	0.00389	-0.00036	0.00575	-0.00021	0.00650
2	1	0.00000	0.00000	0.00050	0.00449	0.00011	0.00557	-0.00018	0.00558
	2	0.00000	0.00000	0.00049	0.00597	0.00029	0.00654	-0.00019	0.00628
	3	0.00000	0.00000	0.00072	0.00817	-0.00014	0.00403	-0.00019	0.00623
3	1	0.00000	0.00000	0.00257	0.02014	-0.00550	0.02357	-0.00022	0.02684
	2	0.00000	0.00000	0.00256	0.02228	-0.00509	0.02471	-0.00046	0.03083
	3	0.00000	0.00000	0.00329	0.02391	-0.00615	0.02391	-0.00021	0.03060

Tabela 3 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.5164$	$SE_{g_2}=1.0328$	$SE_{g_1}=0.4382$	$SE_{g_2}=$	$SE_{g_1}=0.3098$	$SE_{g_2}=0.6197$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	0.00056	0.00472	-0.00137	0.00528	-0.00018	0.00628	-0.00048	0.00575
	2	0.00236	0.00851	-0.00133	0.00639	-0.00015	0.00597	-0.00048	0.00619
	3	0.00070	0.00417	-0.00095	0.00569	-0.00014	0.00410	-0.00051	0.0062
2	1	-0.00068	0.00154	-0.00138	0.00748	0.00196	0.00436	-0.00011	0.00627
	2	-0.00068	0.00154	-0.00073	0.00678	0.00249	0.00619	0.00006	0.00556
	3	-0.00028	0.00047	-0.00103	0.00713	0.00245	0.00613	0.00056	0.00517
3	1	-0.01661	0.03474	-0.01986	0.03513	0.00349	0.02979	-0.00224	0.02616
	2	-0.01590	0.03465	-0.02297	0.04055	0.00333	0.02629	-0.00234	0.02647
	3	-0.01919	0.04254	-0.02405	0.04221	0.00306	0.02529	-0.00215	0.02232

Tabela 4 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 3 pontos

Formato de medida com 3 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.0000$	$SE_{g_2}=0.0000$	$SE_{g_1}=0.2739$	$SE_{g_2}=0.5477$	$SE_{g_1}=0.3098$	$SE_{g_2}=0.6197$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	0.00000	0.00000	0.00107	0.00792	0.00183	0.02597	0.00464	0.02529
	2	0.00000	0.00000	0.00166	0.01424	0.00052	0.02550	0.00563	0.02635
	3	0.00000	0.00000	0.00292	0.03026	0.00163	0.02756	0.00561	0.03002
2	1	0.00000	0.00000	-0.00279	0.00878	-0.00299	0.03036	-0.00296	0.02569
	2	0.00000	0.00000	-0.00390	0.01374	-0.00249	0.02829	-0.00279	0.02410
	3	0.00000	0.00000	-0.00184	0.00505	-0.00289	0.02689	-0.00259	0.02318
3	1	0.00000	0.00000	-0.00089	0.00422	0.00043	0.00812	-0.00022	0.00502
	2	0.00000	0.00000	-0.00124	0.00663	0.00023	0.00548	-0.00017	0.00492
	3	0.00000	0.00000	-0.00125	0.00667	0.00029	0.00654	-0.00019	0.00532

Tabela 5 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 3 pontos

Formato de medida com 3 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.0000$	$SE_{g_2}=0.0000$	$SE_{g_1}=0.3652$	$SE_{g_2}=0.7303$	$SE_{g_1}=0.3098$	$SE_{g_2}=0.6197$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	0.00000	0.00000	0.01426	0.02593	0.00586	0.03292	0.00069	0.02365
	2	0.00000	0.00000	0.01244	0.02250	0.00442	0.02477	0.00032	0.02322
	3	0.00000	0.00000	0.01321	0.02335	0.00541	0.02448	0.00071	0.02849
2	1	0.00000	0.00000	0.01270	0.03365	-0.00597	0.02079	-0.00152	0.02618
	2	0.00000	0.00000	0.00562	0.01422	-0.00704	0.02178	-0.00155	0.02409
	3	0.00000	0.00000	0.00984	0.02802	-0.00752	0.02149	-0.00158	0.02586
3	1	0.00000	0.00000	0.00028	0.00392	0.00029	0.00616	-0.00016	0.00518
	2	0.00000	0.00000	0.00033	0.00289	0.00031	0.00512	-0.00019	0.00545
	3	0.00000	0.00000	0.00043	0.00311	0.00026	0.00460	-0.00020	0.00642

Tabela 6 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 3 pontos

Formato de medida com 3 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.00446	0.01637	0.01634	0.03778	0.00512	0.02098	-0.00404	0.02852
	2	0.00417	0.01750	0.01536	0.03627	0.00362	0.02137	-0.00407	0.02283
	3	0.00663	0.01889	0.01401	0.03091	0.00419	0.02339	-0.00408	0.02846
2	1	0.00098	0.03727	0.01211	0.02588	0.00198	0.02093	0.00734	0.02280
	2	-0.00107	0.01943	0.00848	0.02048	0.00229	0.02204	0.00791	0.02528
	3	-0.00057	0.03833	0.00992	0.02252	0.00239	0.02307	0.00818	0.02661
3	1	0.00165	0.00607	0.00046	0.00835	-0.00078	0.00864	0.00045	0.00693
	2	0.00196	0.00861	0.00046	0.00815	-0.00062	0.00640	0.00046	0.00522
	3	0.00296	0.01385	0.00043	0.00717	-0.00061	0.00658	0.00042	0.00553

Tabela 7 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.3873$	g_2 $SEg_2=0.7746$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$
1	1	-0.00755	0.04540	0.00198	0.04461	0.00472	0.03472	0.00000	0.00000
	2	-0.00448	0.02262	0.00293	0.04420	0.00688	0.04941	0.00000	0.00000
	3	-0.00208	0.00814	0.00012	0.02975	0.00547	0.03868	0.00000	0.00000
2	1	-0.00121	0.00670	-0.00255	0.04222	0.00120	0.04919	0.00000	0.00000
	2	-0.00072	0.00335	-0.00086	0.03932	0.00050	0.03583	0.00000	0.00000
	3	-0.00242	0.01685	0.00054	0.03577	0.00108	0.03861	0.00000	0.00000
3	1	0.03089	0.17228	-0.02004	0.29743	0.03874	0.43287	0.00000	0.00000
	2	0.03821	0.22879	-0.02000	0.34984	0.03315	0.35070	0.00000	0.00000
	3	0.04553	0.28903	-0.02753	0.24287	0.03754	0.40419	0.00000	0.00000

Tabela 8 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.6197$	$SE_{g_2}=1.2394$	$SE_{g_1}=0.4382$	$SE_{g_2}=0.8764$	$SE_{g_1}=0.3098$	$SE_{g_2}=0.6197$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	-0.00350	0.06933	0.00402	0.03169	-0.00035	0.02244	-0.00213	0.03495
	2	0.00783	0.09739	0.00703	0.04509	-0.00083	0.02933	-0.00257	0.04037
	3	0.00726	0.05113	0.00841	0.04540	0.00046	0.02512	-0.00288	0.04355
2	1	-0.00663	0.06653	-0.00221	0.09985	0.00231	0.03721	-0.00232	0.04523
	2	0.00169	0.03091	-0.00201	0.07073	0.00243	0.03183	-0.00249	0.04947
	3	-0.00205	0.01581	-0.00196	0.08577	0.00370	0.03149	-0.00095	0.04569
3	1	-0.00438	0.49530	-0.01542	0.29848	0.02841	0.33037	0.03161	0.38100
	2	-0.01448	0.37969	-0.00435	0.29181	0.03092	0.34044	0.02925	0.36360
	3	-0.01979	0.49704	-0.00866	0.43168	0.03538	0.35755	0.03139	0.40541

Tabela 9 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1959$	g_2 $SE_{g_2}=0.3919$
1	1	-0.00078	0.04846	0.00481	0.01696	0.00106	0.29405	-0.00316	0.02847
	2	0.00325	0.04809	0.00342	0.03400	0.00122	0.02753	-0.00228	0.02253
	3	0.00084	0.06063	0.00567	0.02891	-0.00048	0.03147	-0.00248	0.02278
2	1	0.00129	0.02832	0.00600	0.06224	-0.00056	0.03688	-0.00130	0.03755
	2	-0.00423	0.02972	0.00130	0.03992	-0.00054	0.02158	-0.00125	0.04120
	3	-0.00169	0.02295	0.00014	0.02692	0.00019	0.03244	-0.00124	0.02905
3	1	0.02279	0.30705	-0.02100	0.03631	-0.00903	0.36273	0.00626	0.38273
	2	0.02164	0.18798	-0.02428	0.36969	-0.01090	0.31849	0.00494	0.38484
	3	0.01488	0.23632	-0.03478	0.41206	-0.00803	0.33997	0.00390	0.35845

Tabela 10 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 4 pontos

Formato de medida com 4 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.3873$	g_2 $SEg_2=0.7746$	g_1 $SEg_1=0.2739$	g_2 $SEg_2=0.5477$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.07767	0.14014	-0.01206	0.31854	0.02032	0.40534	0.02289	0.35027
	2	0.16825	0.39279	-0.00599	0.12542	0.02036	0.41354	0.02593	0.42327
	3	0.09390	0.18048	-0.06991	0.15403	0.01804	0.32251	0.02245	0.35403
2	1	-0.07275	0.14359	-0.00754	0.17074	-0.01632	0.39853	0.02378	0.39687
	2	-0.09189	0.19607	-0.00838	0.19664	-0.02446	0.37072	0.02349	0.40247
	3	-0.15759	0.40246	-0.00656	0.14180	-0.01732	0.39781	0.02373	0.39583
3	1	-0.00435	0.02339	-0.00413	0.06522	-0.00189	0.03152	-0.00255	0.03867
	2	-0.00703	0.04435	-0.00344	0.05107	-0.00288	0.03912	-0.00266	0.03634
	3	-0.00267	0.01217	-0.00442	0.07152	-0.00244	0.04294	-0.00255	0.04491

Tabela 11 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 4 pontos

Formato de medida com 4 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=1.2394$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	-0.05744	0.31832	-0.09359	0.55386	0.00100	0.02891	-0.03801	0.32502
	2	-0.04442	0.43689	-0.08859	0.61470	-0.00366	0.32719	-0.03334	0.28190
	3	-0.05449	0.30579	-0.07011	0.45575	-0.00477	0.35954	-0.03419	0.29247
2	1	0.03499	0.14779	0.00948	0.48642	0.00399	0.29384	-0.00666	0.35588
	2	0.06237	0.33297	0.00699	0.50505	0.00661	0.39752	-0.00715	0.35877
	3	0.08059	0.48738	-0.00403	0.43479	0.00492	0.36993	-0.00686	0.31682
3	1	0.00906	0.08443	0.00922	0.04197	-0.00018	0.03255	0.00141	0.03399
	2	-0.00158	0.01516	0.00418	0.02718	-0.00054	0.05037	0.00211	0.03893
	3	0.00213	0.02504	0.00656	0.03102	-0.00037	0.03818	0.00159	0.03280

Tabela 12 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.3873$	$SE_{g_2}=0.7746$	$SE_{g_1}=0.4382$	$SE_{g_2}=0.8764$	$SE_{g_1}=0.3098$	$SE_{g_2}=0.6197$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	-0.04216	0.28143	0.04284	0.29800	0.03016	0.40345	0.00173	0.3468
	2	-0.05877	0.43833	0.05276	0.37813	0.01747	0.34999	0.00145	0.28871
	3	-0.04216	0.28143	0.03530	0.23018	0.02686	0.36381	0.00102	0.31135
2	1	0.02859	0.06907	0.13406	0.56566	0.00019	0.41817	0.01476	0.37313
	2	0.03623	0.09472	0.12286	0.48122	-0.00306	0.49432	0.01444	0.34359
	3	0.05391	0.16091	0.14596	0.51531	-0.00556	0.46787	0.01385	0.38782
3	1	-0.02573	0.08839	-0.00176	0.06280	-0.00027	0.33126	-0.00204	0.04892
	2	-0.01284	0.03499	0.00119	0.07052	-0.00054	0.02071	-0.00374	0.05672
	3	-0.04860	0.20639	0.00155	0.03519	0.00143	0.02713	-0.00256	0.0503

Tabela 13 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$	g_1 $SEg_1=0.3652$	g_2 $SEg_2=0.7303$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.00000	0.00000	0.05413	0.55613	0.01011	0.53281	-0.01098	0.45747
	2	0.00000	0.00000	0.02319	0.22527	0.00778	0.46459	-0.01435	0.57608
	3	0.00000	0.00000	0.03288	0.38486	0.01230	0.46853	-0.01118	0.54057
2	1	0.00000	0.00000	-0.05367	0.48321	0.02211	0.48406	0.00280	0.42773
	2	0.00000	0.00000	-0.02471	0.19898	0.02007	0.38000	0.00165	0.46988
	3	0.00000	0.00000	-0.02277	0.14164	0.02207	0.44859	0.00257	0.54714
3	1	0.00000	0.00000	0.00672	2.08720	0.15725	2.65740	-0.08836	2.95900
	2	0.00000	0.00000	0.02013	1.39770	0.17590	2.89230	-0.08156	2.55860
	3	0.00000	0.00000	-0.05803	1.87940	0.17684	2.9289	-0.09155	2.73110

Tabela 14 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.3873$	g_2 $SEg_2=0.7746$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	-0.05389	1.4039	-0.07419	0.55960	0.01459	0.62115	-0.00209	0.43996
	2	-0.03190	0.69777	-0.05889	0.42753	0.01689	0.75889	-0.00155	0.41665
	3	-0.04951	1.25370	-0.07088	0.62897	0.00996	0.38488	-0.00212	0.42383
2	1	0.01025	0.16168	0.02232	0.56597	-0.03255	0.57902	-0.01526	0.39107
	2	0.01432	0.25262	0.03225	1.02040	-0.03442	0.52484	-0.01816	0.44002
	3	0.00794	0.11510	0.04166	1.25980	-0.02901	0.59341	-0.01436	0.43800
3	1	-0.01353	0.23778	0.02154	2.07000	0.02011	2.63180	-0.06787	2.61980
	2	-0.01254	0.21482	0.03497	1.94800	-0.00074	2.88710	-0.06406	2.33550
	3	-0.04231	1.08750	-0.00092	2.11000	-0.03459	2.49610	-0.07748	2.28120

Tabela 15 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.00862	0.19961	0.03693	0.67034	-0.00255	0.64352	0.01127	0.01127
	2	-0.00258	0.16382	0.04069	0.55865	-0.00036	0.54860	0.01263	0.01262
	3	-0.01119	0.16689	0.04305	0.90991	0.01089	0.82985	0.01129	0.01130
2	1	0.00468	0.07182	0.00582	0.25721	-0.00543	0.30905	-0.01686	-0.01686
	2	0.00317	0.04618	0.00871	0.32149	-0.00408	0.33167	-0.01599	-0.01599
	3	0.00162	0.01776	0.00495	0.23822	-0.00578	0.42217	-0.01691	-0.01692
3	1	-0.01272	1.26930	-0.24978	2.35900	0.21623	2.12790	-0.03731	-0.03731
	2	0.03283	1.91740	-0.27325	3.74210	0.24544	2.50360	-0.07155	-0.07155
	3	0.04594	1.81010	-0.32660	2.61880	0.20110	1.94910	-0.05638	-0.05638

Tabela 16 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.5164$	g_2 $SEg_2=1.0328$	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1959$	g_2 $SEg_2=0.3919$
1	1	-0.41909	4.80560	0.00000	0.00000	-0.06823	2.33920	-0.03903	2.22770
	2	-0.29326	2.45310	0.00000	0.00000	-0.09521	2.77970	-0.04447	2.55370
	3	-0.18117	1.20350	0.00000	0.00000	-0.05340	2.57140	-0.05026	2.44340
2	1	0.11591	1.75000	0.00000	0.00000	0.09715	1.83430	0.05072	2.03020
	2	0.17403	2.5208	0.00000	0.00000	0.11225	1.97030	0.04865	1.86750
	3	0.00790	1.46910	0.00000	0.00000	0.11327	2.7393	0.05892	1.99150
3	1	0.00069	0.46044	0.00000	0.00000	0.00230	0.40562	0.03715	0.56056
	2	0.00048	0.52453	0.00000	0.00000	0.00231	0.59879	0.02895	0.48327
	3	0.00039	0.39203	0.00000	0.00000	-0.00184	0.44223	0.02754	0.46167

Tabela 17 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	0.05984	3.03000	0.05345	1.36910	0.00421	2.47480	0.03068	2.59040
	2	-0.00825	4.07690	-0.00708	1.28970	0.01266	2.55190	0.03464	2.76430
	3	-0.16432	3.64950	0.00511	1.94620	0.00143	1.89910	0.03102	2.77010
2	1	0.15853	1.57510	-0.08474	1.8027	-0.07654	2.28600	0.03641	2.65720
	2	0.18359	2.40440	-0.12201	2.71620	-0.10181	2.1531	0.03693	2.36430
	3	-0.06014	1.77540	-0.12039	3.37280	-0.04452	1.90210	0.03291	2.20080
3	1	-0.07271	0.86354	-0.00107	0.52625	-0.01657	0.23789	-0.00735	0.40429
	2	-0.01220	0.28187	-0.01512	0.56639	-0.02100	0.34888	-0.00446	0.44509
	3	-0.03853	0.43518	0.00049	0.37123	-0.01384	0.28564	-0.00401	0.40871

Tabela 18 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=$
1	1	-0.32627	3.14470	0.05199	1.91910	-0.02786	3.24230	0.06326	2.76240
	2	-0.16307	2.44340	0.13332	2.62880	-0.02002	2.62140	0.07747	2.54840
	3	-0.15716	2.50880	0.06499	2.18610	-0.00534	2.66530	0.07211	2.21550
2	1	0.14564	1.59660	0.40264	2.97080	-0.10134	2.96100	-0.07652	2.41520
	2	0.20830	3.79720	0.35391	2.97720	-0.10359	2.92580	-0.06056	2.32390
	3	0.24281	3.75850	0.40781	3.17770	-0.09838	2.92740	-0.07774	2.52520
3	1	-0.26819	1.43350	-0.00347	0.61288	0.00095	0.60970	0.00539	0.39855
	2	-0.23270	1.15290	-0.00138	0.44940	-0.01508	0.71487	0.00491	0.34707
	3	-0.16454	0.73506	0.00323	0.50858	-0.00270	0.56390	0.00402	0.33550

Tabela 19 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
		$SE_{g_1}=0.6197$	$SE_{g_2}=1.2394$	$SE_{g_1}=0.4382$	$SE_{g_2}=0.8763$	$SE_{g_1}=0.1937$	$SE_{g_2}=0.3873$	$SE_{g_1}=0.1960$	$SE_{g_2}=0.3919$
1	1	0.17291	13.70800	-0.59238	14.13800	0.08986	5.49800	-0.05074	11.05600
	2	-0.30936	12.30300	-0.55654	17.59500	0.08431	5.04920	-0.03356	10.72700
	3	0.16528	9.21860	-0.48298	10.92900	0.11389	7.54080	-0.03938	10.34000
2	1	0.34441	15.08400	-1.02860	33.15700	-0.70326	9.93650	0.06310	16.15100
	2	0.12097	7.64130	-0.38824	25.05100	-0.30133	3.20960	0.08756	11.70100
	3	0.54762	11.95400	-1.22550	29.09500	-0.51605	6.57660	0.07595	14.40800
3	1	-1.30160	66.95000	0.54000	35.00600	-1.68200	43.44800	0.24169	41.48600
	2	-2.22670	75.40000	0.00515	41.48500	-0.81575	16.55600	0.18582	40.63800
	3	-1.50120	80.75500	0.46758	44.08400	-1.04170	22.93600	0.17838	49.21100

Tabela 20 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=1.2394$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.44811	9.68270	-0.50555	7.23390	0.37536	14.58200	-0.01555	14.46000
	2	0.81172	18.07600	-0.65954	9.68130	0.31065	10.79400	-0.02981	13.48200
	3	0.15904	17.16900	-0.89351	14.84800	0.35164	13.08500	-0.01045	16.37400
2	1	0.14024	13.62800	0.60883	17.43200	-0.25517	9.77970	-0.10051	12.17600
	2	-0.05611	19.34500	0.41557	11.99300	-0.31762	12.21200	-0.08639	11.52600
	3	-0.28710	24.18300	0.29507	7.81450	-0.34395	16.52200	-0.10268	16.39200
3	1	2.45980	35.03700	-1.45810	42.55700	1.07550	34.17600	-0.28482	45.68700
	2	0.65245	18.61000	-2.11040	54.91800	1.13680	35.51800	-0.25940	35.26700
	3	1.67330	20.05400	-1.60990	40.44900	1.38530	46.83300	-0.26187	38.83900

Tabela 21 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=1.2394$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.45670	25.22700	-0.02484	9.44880	0.42505	18.83400	0.00817	11.57000
	2	0.54183	25.68700	-0.03555	9.96630	0.40526	18.15100	0.02422	12.02600
	3	0.14305	18.67100	-0.02485	7.32840	0.38087	16.62400	0.01656	11.66900
2	1	0.16561	19.64900	0.15632	11.48300	0.73513	18.80000	-0.22416	11.37300
	2	-0.01760	13.90600	0.19704	7.90280	0.62729	13.35000	-0.20039	12.77200
	3	-0.06714	12.36300	0.17500	8.89320	0.72482	16.10800	-0.22262	12.18300
3	1	-4.56470	96.46100	1.65730	40.06900	1.36500	34.88500	-0.80890	45.62000
	2	-3.40020	54.83500	1.86700	37.74800	1.40880	37.17900	-0.90041	52.18100
	3	-2.91720	51.58000	1.83090	40.43600	1.52030	43.56600	-0.79308	45.75900

Tabela 22 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa intensidade ($r=0.10$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.10$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-1.66810	76.07900	-0.66691	49.75600	0.36869	44.83300	0.06464	36.75000
	2	-1.25700	44.13800	-0.67953	40.66800	0.17469	44.18900	0.11211	38.33600
	3	-1.47040	74.71700	-0.81421	47.63600	0.26394	46.50900	0.09793	31.85500
2	1	-2.26460	47.69600	-1.93060	30.14000	-0.23484	41.38900	0.07846	48.07100
	2	-0.25820	15.98500	-2.21930	36.91300	-0.11074	40.68800	0.04235	43.62000
	3	-2.52570	70.01400	-1.57440	27.16900	-0.08133	42.06400	0.03069	38.29200
3	1	-0.23022	11.33400	0.28548	8.73070	-0.12228	10.06000	-0.10417	10.04600
	2	-0.22467	23.53200	0.36858	14.70300	-0.09532	10.04200	-0.10601	12.45200
	3	-0.05141	5.87450	0.37048	12.83600	-0.09159	8.60320	-0.13220	15.17700

Tabela 23 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=1.2394$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3099$	g_2 $SEg_2=0.61968$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	2.12090	45.89300	-0.85011	29.19800	-0.48475	34.94900	0.03412	38.98700
	2	1.01520	27.52200	-0.80073	33.27100	-0.18782	59.12000	0.24687	38.68600
	3	1.72250	35.61300	-1.11190	30.80400	-0.53030	50.89800	0.09363	38.82200
2	1	0.83346	33.69200	0.13173	61.59500	0.82490	43.17400	-0.58965	37.12400
	2	1.33790	57.62500	-0.26429	54.51900	0.90907	53.26700	-0.53494	37.25700
	3	0.28154	25.99800	-0.59028	44.40300	0.68833	34.88900	-0.53044	37.35600
3	1	1.17350	15.32300	-0.15453	7.60550	-0.27512	10.69700	0.06365	13.25700
	2	1.48760	18.12400	-0.24466	9.26800	-0.25573	14.02600	0.04851	11.18000
	3	0.63158	7.17020	-0.23682	8.23620	-0.37662	14.928	0.04804	12.42700

Tabela 24 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de elevada intensidade ($r=0.70$), com distribuição uniforme em dois dos grupos e simétrica no terceiro ($p_4p_4p_1$) e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição uniforme em dois grupos e simétrica no terceiro ($p_4p_4p_1$), $r=0.70$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	0.08153	32.30100	-1.20280	36.24600	-2.29780	49.588	-0.87862	38.55400
	2	-0.32388	17.63400	-1.61400	41.02500	-2.21570	44.49600	-0.94253	39.85600
	3	0.45293	21.79300	-1.08940	39.78000	-2.03800	41.17300	-0.90037	38.33500
2	1	0.48896	69.52600	0.64666	35.22600	0.95571	31.96400	-0.30126	37.25500
	2	0.69610	72.03100	0.50235	28.83800	1.21740	36.77100	-0.33926	44.81400
	3	0.84913	112.57000	0.68883	39.31300	1.23450	38.03300	-0.31984	37.51200
3	1	-0.69605	9.11880	-0.20263	9.73970	0.29911	10.06700	0.20160	16.59500
	2	-0.65034	6.32930	-0.42406	22.00700	0.29006	10.46500	0.22999	16.24000
	3	-1.19790	14.37600	-0.45541	23.40100	0.29695	9.74480	0.17429	14.13600

Tabela 25 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica num grupo, assimétrica positiva no outro grupo e assimétrica negativa no outro grupo ($p_1p_2p_3$), e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição simétrica, assimétrica negativa e assimétrica positiva em cada grupo ($p_1p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.00000	0.00000	0.00012	0.00433	0.00012	0.00732	-0.00021	0.00674
	2	0.00000	0.00000	0.00020	0.00345	-0.00011	0.00486	-0.00017	0.00786
	3	0.00000	0.00000	0.00008	0.00241	0.00007	0.00557	-0.00021	0.00612
2	1	0.00000	0.00000	0.01726	0.01340	0.01822	0.01224	0.01939	0.01343
	2	0.00000	0.00000	0.02920	0.02669	0.01621	0.01049	0.01670	0.01107
	3	0.00000	0.00000	0.02294	0.01995	0.01608	0.01060	0.01784	0.01203
3	1	0.00000	0.00000	-0.01843	0.01314	-0.02191	0.01558	-0.01656	0.01060
	2	0.00000	0.00000	-0.00838	0.00462	-0.02245	0.01577	-0.01552	0.00971
	3	0.00000	0.00000	-0.01320	0.00883	-0.01807	0.01184	-0.01760	0.01140

Tabela 26 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica num grupo, assimétrica positiva no outro grupo e assimétrica negativa no outro grupo ($p_1p_2p_3$), e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição simétrica, assimétrica negativa e assimétrica positiva em cada grupo ($p_1p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.5810$	g_2 $SEg_2=1.1619$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	-0.00798	0.01110	-0.00425	0.03321	0.00351	0.03773	-0.00302	0.05220
	2	-0.02018	0.04493	-0.00282	0.02449	0.00328	0.03131	-0.00262	0.04461
	3	-0.01338	0.02264	-0.00680	0.05010	0.00620	0.04682	-0.00248	0.04133
2	1	0.27160	0.55063	0.16120	0.24752	0.17366	0.2548	0.15447	0.22247
	2	0.40269	0.89190	0.23917	0.39723	0.14980	0.21523	0.19042	0.28839
	3	0.22268	0.43825	0.18192	0.28506	0.17308	0.25421	0.17498	0.25875
3	1	-0.05977	0.09796	-0.22359	0.37196	-0.23584	0.41247	-0.17267	0.25893
	2	-0.08316	0.17590	-0.18512	0.27883	-0.18749	0.30352	-0.22519	0.36947
	3	-0.06651	0.12215	-0.19627	0.30277	-0.18446	0.29415	-0.17082	0.25523

Tabela 27 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica num grupo, assimétrica positiva no outro grupo e assimétrica negativa no outro grupo ($p_1p_2p_3$), e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição simétrica, assimétrica negativa e assimétrica positiva em cada grupo ($p_1p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.02490	0.50408	0.01798	0.69001	0.02717	0.30408	-0.01893	0.42093
	2	0.00041	0.20470	0.01572	0.69449	0.03097	0.35597	-0.02057	0.42148
	3	-0.02613	0.86035	0.00656	0.35333	0.02795	0.33191	-0.02139	0.45416
2	1	1.35060	5.08080	1.19060	4.76480	0.05470	3.21690	0.94329	2.77070
	2	0.91703	2.73960	0.95935	3.52720	1.09380	3.29060	0.91165	2.64230
	3	0.93129	3.63830	1.27200	5.17980	1.01450	3.06520	0.99725	2.98920
3	1	-1.22290	5.05950	-1.01990	3.80240	-1.14950	3.88930	-0.89050	2.54900
	2	-0.71966	3.23080	-1.18470	3.97920	-0.94477	2.89650	-0.80057	2.17960
	3	-0.98239	4.22250	-0.72429	2.13810	-1.09540	3.51170	-0.81803	2.24230

Tabela 28 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica num grupo, assimétrica positiva no outro grupo e assimétrica negativa no outro grupo ($p_1p_2p_3$), e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição simétrica, assimétrica negativa e assimétrica positiva em cada grupo ($p_1p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.72573	13.50100	0.20638	19.41900	0.15480	15.69500	0.15363	14.80000
	2	-0.24030	21.63400	0.15962	13.52400	0.16654	12.43500	0.16904	12.02800
	3	-0.47131	12.90300	0.4204	35.41000	0.17822	13.17800	0.13520	13.40600
2	1	9.27290	126.76000	7.58840	67.61000	7.37100	83.23100	8.55440	83.31400
	2	5.63930	66.21600	9.69670	83.86300	7.46230	81.97900	8.53210	85.27600
	3	3.56800	34.59300	8.91110	75.29100	6.60730	79.92300	8.37870	82.54800
3	1	-6.29470	133.66000	-10.40500	96.25300	-9.16300	79.06100	-8.60170	75.49300
	2	-4.59740	88.12900	-9.26860	82.32000	-7.06090	55.05400	-8.70190	76.90500
	3	-2.42520	62.66700	-9.06340	82.58200	-6.94180	54.39800	-9.30900	84.15400

Tabela 29 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.0000$	g_2 $SE_{g_2}=0.0000$	g_1 $SE_{g_1}=0.2739$	g_2 $SE_{g_2}=0.5477$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1959$	g_2 $SE_{g_2}=0.3919$
1	1	0.00000	0.00000	0.01285	0.00968	0.01509	0.00978	0.01513	0.00961
	2	0.00000	0.00000	0.00650	0.00390	0.01402	0.00881	0.01618	0.01028
	3	0.00000	0.00000	0.00629	0.00373	0.01296	0.00798	0.01541	0.00977
2	1	0.00000	0.00000	0.01462	0.01242	0.01566	0.01026	0.01724	0.01125
	2	0.00000	0.00000	0.01452	0.01231	0.01271	0.00784	0.01852	0.01245
	3	0.00000	0.00000	0.01374	0.01142	0.01641	0.01133	0.02069	0.01474
3	1	0.00000	0.00000	-0.01413	0.01068	-0.02284	0.01715	-0.01401	0.00868
	2	0.00000	0.00000	-0.00706	0.00423	-0.02111	0.01493	-0.01442	0.00899
	3	0.00000	0.00000	-0.00995	0.00669	-0.01949	0.01348	-0.01526	0.00976

Tabela 30 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.00000	0.00000	0.20915	0.36029	0.14747	0.24747	0.20649	0.33180
	2	0.00000	0.00000	0.16470	0.24779	0.15667	0.27151	0.16682	0.25312
	3	0.00000	0.00000	0.15311	0.23703	0.16560	0.28611	0.22956	0.38000
2	1	0.00000	0.00000	0.15506	0.25903	0.21688	0.34753	0.16187	0.23369
	2	0.00000	0.00000	0.16587	0.28329	0.26029	0.43919	0.18979	0.29112
	3	0.00000	0.00000	0.12094	0.19574	0.16252	0.23724	0.17081	0.25053
3	1	0.00000	0.00000	-0.17716	0.29350	-0.15776	0.23410	-0.18378	0.27132
	2	0.00000	0.00000	-0.15591	0.25171	-0.23761	0.40135	-0.19512	0.29368
	3	0.00000	0.00000	-0.21148	0.39026	-0.17076	0.26959	-0.22038	0.34883

Tabela 31 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	0.81659	2.46200	1.26480	3.59880	1.18720	3.69300	1.12370	3.23910
	2	1.05990	3.46790	1.41620	4.63740	0.89097	2.49700	0.87038	2.32550
	3	0.98077	3.53030	1.43640	4.59540	1.01110	3.00150	0.94904	2.59860
2	1	1.04090	2.89710	1.08610	3.56110	0.87366	2.47350	0.91489	2.63960
	2	1.07290	3.11980	0.59432	1.68480	0.91379	2.54100	1.02350	3.04980
	3	1.63470	5.17720	1.07280	3.41250	0.69983	1.83110	1.00890	2.98750
3	1	-0.98225	3.30390	-1.02140	2.77570	-1.20950	3.98010	-0.97258	2.88280
	2	-1.07560	3.48800	-0.83360	2.14790	-1.10340	3.51000	-0.98405	2.98950
	3	-0.94493	2.8368	-1.14870	3.33370	-1.27700	4.2654	-0.99716	2.92650

Tabela 32 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.6197$	g_2 $SEg_2=1.2394$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6197$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.39192$
1	1	11.59500	88.17100	8.00150	80.68200	9.87200	79.60000	7.64960	72.51600
	2	10.93300	93.58700	8.65940	87.03500	10.88400	86.92900	8.28630	79.77200
	3	3.92940	37.05400	11.03300	113.54000	10.87500	87.16100	7.53210	70.36500
2	1	4.10770	125.51000	7.67940	82.75900	7.53610	59.23100	9.22700	79.99900
	2	1.40320	74.15200	6.40370	64.90100	4.89100	36.47000	10.03800	86.65500
	3	3.59560	83.41900	5.83150	58.93800	8.43900	66.05800	11.18800	103.12000
3	1	-4.95710	90.16600	-7.99370	89.84100	-11.19400	97.31700	-6.50900	81.68800
	2	-6.19720	99.57000	-5.76970	88.39700	-10.00300	82.07700	-6.27710	79.57200
	3	-8.03060	148.79000	-9.00400	102.37000	-12.75300	110.50000	-6.02850	71.93800

Tabela 33 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), e escala com formato de medida com 3 pontos.

Formato de medida com 3 pontos									
Distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.5810$	g_2 $SE_{g_2}=1.1619$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.2582$	g_2 $SE_{g_2}=0.5164$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.00035	0.00342	-0.00166	0.00837	0.00085	0.00297	-0.00099	0.00649
	2	-0.00013	0.00138	-0.00041	0.00373	0.00076	0.00294	-0.00089	0.00501
	3	-0.00033	0.00510	-0.00064	0.00445	0.00010	0.00458	-0.00063	0.00438
2	1	0.01304	0.00805	0.01878	0.01443	0.01522	0.01024	0.01669	0.01069
	2	0.01229	0.00770	0.01566	0.01103	0.01924	0.01417	0.01661	0.01071
	3	0.01809	0.01304	0.01241	0.00897	0.01304	0.00857	0.01775	0.01168
3	1	0.00430	0.02049	-0.00060	0.03555	0.01042	0.02331	-0.00336	0.02664
	2	0.004112	0.02591	-0.00012	0.02878	0.00957	0.02381	-0.00351	0.02677
	3	0.00494	0.02593	-0.00091	0.02725	0.00914	0.02101	-0.00369	0.02688

Tabela 34 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), e escala com formato de medida com 4 pontos.

Formato de medida com 4 pontos									
Distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	-0.00051	0.09074	0.00098	0.03626	-0.00650	0.04892	0.00033	0.036124
	2	-0.01341	0.06080	0.00313	0.02838	-0.00526	0.03758	0.00024	0.03433
	3	-0.01068	0.12336	0.00078	0.02287	-0.00649	0.06086	0.00025	0.04078
2	1	0.20783	0.48169	0.17358	0.27074	0.23167	0.43215	0.21105	0.33420
	2	0.15325	0.36651	0.13187	0.19533	0.17235	0.28055	0.20653	0.32647
	3	0.14810	0.36188	0.18152	0.28986	0.20379	0.34318	0.17598	0.26273
3	1	0.00086	0.35372	-0.03145	0.42515	0.02724	0.29212	-0.00349	0.27578
	2	0.00662	0.17891	-0.00282	0.37902	0.02756	0.25644	-0.00254	0.31176
	3	0.00857	0.32683	-0.02405	0.40221	0.02803	0.26217	-0.00508	0.32746

Tabela 35 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), e escala com formato de medida com 5 pontos.

Formato de medida com 5 pontos									
Distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SEg_1=0.0000$	g_2 $SEg_2=0.0000$	g_1 $SEg_1=0.4382$	g_2 $SEg_2=0.8764$	g_1 $SEg_1=0.3098$	g_2 $SEg_2=0.6968$	g_1 $SEg_1=0.1960$	g_2 $SEg_2=0.3919$
1	1	0.00000	0.00000	-0.03804	0.49246	-0.01025	0.51021	0.02020	0.42485
	2	0.00000	0.00000	-0.04369	0.69346	-0.01624	0.46585	0.02123	0.48106
	3	0.00000	0.00000	-0.02622	0.42058	-0.01533	0.44954	0.02547	0.52573
2	1	0.00000	0.00000	0.61470	1.97070	1.02220	3.02630	0.77090	2.10870
	2	0.00000	0.00000	0.59613	1.81810	1.13190	3.49130	0.83295	2.29490
	3	0.00000	0.00000	0.53241	1.66150	0.95207	2.79110	0.85770	2.46690
3	1	0.00000	0.00000	0.14982	2.74660	-0.02324	1.89180	0.04230	2.25520
	2	0.00000	0.00000	0.15817	2.98600	-0.01876	2.05850	0.03925	2.21680
	3	0.00000	0.00000	0.07240	2.48850	-0.04033	1.97560	0.04402	2.25250

Tabela 36 – Medidas de forma, assimetria (g_1) e curtose (g_2), das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de média intensidade ($r=0.40$), com distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), e escala com formato de medida com 7 pontos.

Formato de medida com 7 pontos									
Distribuição simétrica, assimétrica positiva e uniforme em cada um dos grupos ($p_1p_2p_4$), $r=0.40$									
n_i		$n_i=10$		$n_i=20$		$n_i=40$		$n_i=100$	
Variáveis	Grupos	g_1 $SE_{g_1}=0.6197$	g_2 $SE_{g_2}=1.2394$	g_1 $SE_{g_1}=0.4382$	g_2 $SE_{g_2}=0.8764$	g_1 $SE_{g_1}=0.3098$	g_2 $SE_{g_2}=0.6197$	g_1 $SE_{g_1}=0.1960$	g_2 $SE_{g_2}=0.3919$
1	1	0.08598	6.39220	0.27774	15.51400	-0.05339	13.92100	0.08759	18.28500
	2	0.04906	18.01500	0.18153	9.97880	-0.00565	13.06600	0.07883	15.47200
	3	-0.08284	9.880300	0.33464	19.66100	-0.02985	12.35300	0.12817	17.16000
2	1	10.81300	90.76500	4.61730	49.02400	9.28200	84.71000	8.32530	67.53300
	2	8.74780	70.09400	6.45140	73.40300	8.45670	75.12400	8.53740	69.59100
	3	11.41400	96.62700	6.72240	74.71900	7.06740	59.31800	8.64260	69.99100
3	1	-0.61867	27.79500	-1.95960	37.76400	-0.04449	33.97500	-0.43561	40.72700
	2	-1.84360	42.06800	-1.52090	36.79300	0.10958	33.67200	-0.38902	39.52900
	3	-1.17300	38.32600	-0.62601	19.72300	-0.00172	33.87900	-0.38592	40.75300

Apêndice III: Tabelas com resultados das médias das estatísticas de Kruskal-Wallis

Tabela 1 – Valor observado da estatística de teste de Kruskal-Wallis das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa, média e elevada intensidade, distribuição simétrica em dois grupos e uniforme no terceiro e escalas com 3, 4, 5 e 7 pontos.

Distribuição simétrica em dois grupos e uniforme no terceiro ($p_1p_1p_4$)													
		$r=0.10$				$r=0.40$				$r=0.70$			
n_i		$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$
Formato de Medida	Variável	Estatística de Kruskal-Wallis nas três variáveis Kruskal- Wallis médio (KWME)											
3 Pontos	1	3.2540	0.2066	2.6322	0.5885	2.6820	1.3029	1.7851	1.5570	0.7765	2.1477	1.8811	1.8825
	2	0.2453	0.3740	1.5764	1.1814	0.7370	1.8873	0.8072	1.4131	2.1277	1.9671	2.3106	0.5646
	3	2.0344	3.3522	2.0202	0.7753	5.5193	2.1503	1.9284	1.2671	1.4516	1.6088	2.5667	1.0250
KWME		1.8446	1.3097	2.0763	0.8484	2.9794	1.7802	1.5069	1.4124	1.4519	1.9079	2.2528	1.1574
4 Pontos	1	0.9268	3.2878	0.6648	1.4008	0.6242	0.5302	0.4399	3.0543	1.5994	1.8915	0.7366	1.8927
	2	1.0640	1.3266	1.1536	0.6757	0.7356	2.0230	2.7065	2.0193	3.3664	1.9574	1.2309	2.6194
	3	2.5443	1.0522	1.9169	0.0434	0.1994	0.8875	2.0815	1.4080	2.1203	2.1522	2.3326	1.1850
KWME		1.5117	1.8875	1.2451	0.7066	0.5201	1.1469	1.7426	2.1605	2.3620	2.0003	1.4334	1.8990
5 Pontos	1	0.5058	1.4149	2.1968	1.5330	3.7873	3.6273	1.4738	0.6427	1.7474	1.0776	3.2552	2.4285
	2	1.3449	3.4784	0.8147	2.1384	0.9639	1.7139	0.9011	1.1033	2.3579	1.6703	1.6425	3.9585
	3	1.2250	3.7142	1.5920	1.6019	1.7156	1.7089	1.3205	1.1400	1.2746	1.6610	0.7778	3.0476
KWME		1.0252	2.8692	1.5345	1.7578	2.1556	2.3500	1.2318	0.9620	1.7933	1.4696	1.8918	3.1449
7 Pontos	1	2.6733	2.7453	1.4767	1.5586	2.5604	1.7733	1.9711	0.6132	2.7777	2.8298	1.7504	1.8654
	2	2.9481	1.2388	0.8787	2.7837	3.1955	0.3693	0.8502	1.1467	0.7911	3.8729	1.5334	0.4888
	3	1.2068	1.4824	1.7396	0.9713	1.6763	1.9960	1.9663	1.3225	1.8451	4.0216	1.1957	2.1176
KWME		2.2761	1.8222	1.3649	1.7712	2.4774	1.3795	1.5959	1.0275	1.8046	3.5748	1.4932	1.4906

Tabela 2 – Valor observado da estatística de teste de Kruskal-Wallis das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa, média e elevada intensidade, distribuição uniforme em 2 grupos e simétrica no terceiro ($p_4p_4p_1$) e escalas com 3, 4, 5 e 7 pontos.

Distribuição uniforme em 2 grupos e simétrica no terceiro grupo ($p_4p_4p_1$)													
Correlação (r)		$r=0.10$				$r=0.40$				$r=0.70$			
n_i		$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$
Formato de Medida	Variável	Kruskal-Wallis nas três variáveis Kruskal- Wallis médio (KWME)											
3 Pontos	1	0.0395	3.1936	1.3434	3.0071	0.0439	0.9178	2.2648	0.6763	2.8220	1.9160	2.5832	2.5943
	2	3.5914	0.1434	0.9592	1.6852	0.2417	1.7859	2.4032	1.8827	2.1978	1.4064	2.5075	2.7992
	3	1.8170	0.8025	2.3695	2.8973	0.7285	0.7374	2.6206	1.3879	3.3865	1.4018	2.4918	1.0598
KWME		1.8159	1.3798	1.5574	2.5299	0.3380	1.1470	2.4295	1.3156	2.4688	1.5747	2.5275	2.1511
4 Pontos	1	2.3610	1.0071	0.8259	3.3676	0.7875	1.1909	0.5634	2.9228	1.0318	1.3482	1.6651	0.9730
	2	1.5428	2.1070	1.7078	1.2767	0.6098	0.8808	0.2988	2.0513	2.0552	1.0361	2.8794	1.1865
	3	3.0814	2.3574	1.2693	2.5731	0.8499	2.0219	1.7766	1.7390	0.6158	1.4178	1.7683	1.2862
KWME		2.3284	1.8238	1.2677	2.4058	0.7491	1.3645	0.8796	2.2377	1.2343	1.2674	2.1043	1.1486
5 Pontos	1	1.4638	0.2379	1.1002	0.4152	1.8361	0.8889	2.7904	0.8266	1.4718	0.7780	0.4099	1.6418
	2	1.7688	1.1379	2.4764	2.8150	1.0707	1.3722	0.4335	1.8949	1.5395	1.0621	1.4309	1.2932
	3	1.9476	3.3838	0.6369	1.5810	1.4662	1.0698	2.7476	2.3260	1.4446	1.1453	1.1267	2.2924
KWME		1.7267	1.5866	1.4045	1.6037	1.4577	1.1107	1.9905	1.6825	1.4853	0.9951	0.9892	1.7425
7 Pontos	1	2.1336	2.2644	2.7603	1.2901	1.7350	1.4090	1.5534	3.0215	2.1959	2.6573	4.3977	2.3239
	2	1.0930	0.9649	2.1883	1.7691	0.7732	1.1759	0.8135	1.0537	2.9385	2.3120	2.1748	1.7431
	3	1.1230	2.0907	3.0204	2.7383	0.9444	1.7447	1.1218	2.4567	2.3140	2.6536	2.0063	2.6856
KWME		1.4499	1.7733	2.6563	1.9325	1.1509	1.4432	1.1629	2.1773	2.4828	2.5409	2.8596	2.2509

Tabela 3 – Valor observado da estatística de teste de Kruskal-Wallis das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa, média e elevada intensidade, distribuição simétrica, assimétrica positiva e assimétrica negativa em cada grupo ($p_1p_2p_3$) e escalas com 3, 4, 5 e 7 pontos.

Distribuição simétrica, assimétrica positiva e assimétrica negativa em cada um dos grupos ($p_1p_2p_3$)													
Correlação (r)		$r=0.10$				$r=0.40$				$r=0.70$			
n_i		$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$
Formato de Medida	Variável	Kruskal-Wallis nas três variáveis Kruskal- Wallis médio (KWME)											
3 Pontos	1	7.1674	11.7695	19.6133	50.2403	6.0773	17.0410	32.0560	81.4280	7.4389	16.3470	28.0930	69.9680
	2	9.9827	13.2243	20.2796	45.6043	9.3916	17.2070	34.7480	92.5700	4.4956	15.3480	23.9190	67.1090
	3	8.5724	17.4279	28.2189	57.4705	5.9035	16.0020	35.7170	87.3540	4.1659	15.9280	26.0290	70.8510
KWME		8.5742	14.1406	22.7039	53.7717	7.1241	16.7500	34.1737	87.1173	5.3668	15.8743	26.0137	69.3093
4 Pontos	1	8.4795	7.1010	29.0805	52.8076	9.5283	18.7160	31.4280	83.9730	6.2105	17.9370	26.0010	73.3320
	2	8.7852	6.0640	21.7241	55.6007	6.3347	16.0700	36.4220	70.4000	7.8529	19.9210	27.7460	70.9390
	3	4.4100	14.6160	10.7641	58.1203	8.4665	17.4160	34.6710	82.9600	6.6130	18.6720	26.8050	75.1140
KWME		7.2249	9.2603	20.5229	55.5095	8.1098	17.4007	34.1737	79.1110	6.8921	18.8433	26.8507	73.1283
5 Pontos	1	6.3312	9.3316	19.6460	34.4988	6.1795	8.7042	21.0880	59.3000	7.7776	13.7880	26.1930	58.6940
	2	8.5717	4.3342	14.2496	29.6514	7.2282	9.0249	25.0520	61.478	8.5813	13.3920	29.5670	67.9460
	3	4.8968	8.1211	18.3845	35.2748	7.0838	8.6602	23.1350	64.2650	6.2148	12.5220	28.2180	57.2270
KWME		6.5999	7.2623	17.4267	33.1417	6.8305	8.7964	23.0917	61.6810	7.5246	13.2340	27.9927	61.289
7 Pontos	1	3.2208	3.2823	8.9842	22.1768	1.6147	7.3417	9.4793	30.7550	1.8347	8.1979	13.0260	29.8820
	2	6.9064	3.3280	4.7756	18.4424	3.1754	8.0606	14.9720	30.7770	3.1102	10.1560	18.4100	30.7130
	3	3.2414	3.5577	8.7025	16.9815	3.7607	10.2760	20.5400	29.9830	3.1764	5.9831	16.0440	36.5280
KWME		4.4562	3.3894	7.4874	19.2002	2.9503	8.5594	14.9971	30.5050	2.7071	8.1123	15.8267	32.3743

Tabela 3 – Valor observado da estatística de teste de Kruskal-Wallis das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa, média e elevada intensidade, distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$) e escalas com 3, 4, 5 e 7 pontos.

Distribuição assimétrica positiva em dois grupos e assimétrica negativa no terceiro grupo ($p_2p_2p_3$)													
		$r=0.10$				$r=0.40$				$r=0.70$			
n_i		$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$
Formato de Medida	Variável	Kruskal-Wallis nas três variáveis Kruskal- Wallis médio (KWME)											
3 Pontos	1	5.3805	15.2662	33.4467	51.7229	7.1861	16.9430	41.1800	111.3700	7.4021	19.9040	43.5580	99.0610
	2	4.6763	9.2392	30.4608	73.2477	9.1640	14.9470	40.6180	101.0700	9.6450	20.1280	43.5010	102.7200
	3	8.6379	14.1653	28.3204	62.2987	6.6430	16.1470	43.8810	106.5200	8.0928	17.2980	39.4060	103.3700
KWME		6.2316	12.8902	30.7426	62.4231	7.6647	16.0123	41.8930	106.3200	8.3799	19.1100	42.1550	101.7170
4 Pontos	1	4.2483	14.4675	21.7900	49.2119	9.9640	15.5050	34.4510	91.3050	8.8110	18.0340	41.2500	92.9540
	2	3.1511	15.2409	19.0754	45.0502	9.9844	18.0540	27.0780	89.0460	9.93130	20.1040	38.8720	87.0970
	3	3.3233	12.4068	22.0899	42.1715	11.346	16.9250	37.1410	75.1700	9.0853	12.8210	33.7920	88.9800
KWME		3.5742	14.0384	20.9851	45.4778	10.4315	16.8280	32.8900	85.1737	9.2759	16.9863	37.9713	89.6770
5 Pontos	1	5.6753	5.6019	11.9456	56.7517	7.4904	14.1890	20.0250	64.2640	5.4031	15.3380	23.9610	62.4010
	2	4.3268	4.6125	15.1400	46.3240	6.4548	18.1740	25.3680	67.8530	6.1674	13.6060	26.1080	61.5820
	3	6.5647	12.4602	21.5261	38.0588	6.1268	12.2390	25.7470	65.6040	5.8717	15.9100	28.1880	59.4950
KWME		5.5223	7.5582	16.2039	47.0448	6.6906	14.8673	23.7133	65.9070	5.8141	14.9513	26.0857	61.1593
7 Pontos	1	0.0254	4.8047	10.0440	29.3513	4.4788	5.9246	19.2630	30.7910	4.5348	5.7362	15.6280	33.9940
	2	0.5719	3.8687	15.0341	19.9927	1.9343	10.6660	24.4530	29.3510	2.5038	7.0237	11.8990	37.0160
	3	0.0137	5.1438	7.5513	26.3575	5.4001	6.5660	17.2730	30.3010	2.0936	7.2332	11.3130	36.4430
KWME		1.2037	4.6057	10.8765	25.2338	3.9378	7.7189	20.3297	30.1477	3.0441	6.6644	12.9470	35.8177

Tabela 3 – Valor observado da estatística de teste de Kruskal-Wallis das três variáveis nas mil réplicas e respetiva média, para as situações simuladas com $n=30, 60, 120, 300$, correlações de baixa, média e elevada intensidade, distribuição simétrica, assimétrica positiva e uniforme em cada grupo ($p_1p_2p_4$) e escalas com 3, 4, 5 e 7 pontos.

Distribuição simétrica, assimétrica positiva e uniforme em cada grupo ($p_1p_2p_4$)													
		$r=0.10$				$r=0.40$				$r=0.70$			
n_i		$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$	$n_i=10$	$n_i=20$	$n_i=40$	$n_i=100$
Formato de Medida	Variável	Kruskal-Wallis nas três variáveis Kruskal-Wallis médio (<i>KWME</i>)											
3 Pontos	1	0.6643	1.5084	9.9356	21.2196	4.4599	7.3031	9.3351	40.127	5.8316	12.0590	16.3210	37.2380
	2	1.9656	4.6301	11.1701	19.8703	8.5516	7.8340	8.6338	40.3820	5.5491	12.2140	20.4230	37.6040
	3	3.9472	1.4426	3.7123	21.9370	6.6391	10.8920	12.6460	46.516	5.0828	13.6720	19.1650	36.4060
	<i>KWME</i>	2.1924	2.5270	8.2727	21.0089	6.5502	8.6764	10.2049	42.3417	5.4878	12.6480	18.6363	37.0827
4 Pontos	1	6.0185	5.0675	10.6615	27.4493	0.8883	6.5516	13.6440	30.9800	4.1096	9.6621	12.3410	36.2920
	2	7.8747	5.6818	14.7873	17.8922	3.7144	11.6670	18.3130	39.4170	5.7277	7.7057	14.7150	30.7810
	3	3.2971	5.8389	5.9450	20.8858	3.1942	6.5065	13.1760	32.9120	5.2371	9.5412	12.7960	34.1060
	<i>KWME</i>	5.7301	5.5294	10.4646	22.0757	2.5989	8.2417	15.0443	34.4363	5.0248	8.9697	13.2840	33.7263
5 Pontos	1	5.9337	3.6950	1.8458	13.9318	1.8027	4.3645	14.0630	31.7770	4.1096	9.6621	12.341	36.2920
	2	1.2523	1.8968	7.9327	22.7491	6.2564	4.3461	15.7370	26.8130	5.7277	7.7057	14.7150	30.7810
	3	2.8939	4.4546	7.9167	9.8090	2.7648	7.0329	16.376	30.9370	5.2371	9.5412	12.796	34.1060
	<i>KWME</i>	3.3599	3.3488	5.8984	15.4966	3.6080	5.2478	15.3920	29.8423	5.0248	8.9697	13.2840	33.7263
7 Pontos	1	0.5726	1.3123	4.4338	4.7501	3.2515	1.9962	4.8417	18.7880	2.8002	3.1885	4.1953	11.707
	2	1.1578	0.2809	3.6625	14.9287	1.7431	4.6414	5.2385	13.1020	2.0258	2.4815	4.0285	12.0600
	3	1.4170	1.7716	4.9169	8.1015	3.7385	4.7121	10.5510	16.8540	4.2158	3.0914	5.5987	10.7120
	<i>KWME</i>	1.0491	1.1216	4.3378	9.2601	2.9110	3.7832	6.8771	16.2480	3.0139	2.9205	4.6075	11.4930

Anexo I: Pedidos de Autorização

Exma. Sra. Coordenadora da Equipa de Projeto Early Feeding Skills

Sra. Enfermeira Thereza Vasconcellos

Chefe da Unidade de Cuidados Intensivos Neonatais

Hospital de S. Francisco Xavier

Centro Hospitalar de Lisboa Ocidental, EPE

Maria Alice dos Santos Curado, Professora Coordenadora na Escola Superior de Enfermagem de Lisboa, Especialista em Enfermagem de Saúde Infantil e Pediatria, Mestre em Desenvolvimento da Criança, Doutoranda em Ciências da Motricidade na Especialidade de Comportamento Motor, na Faculdade de Motricidade Humana, Universidade de Lisboa, encontra-se a desenvolver o trabalho de investigação "A Medida e as Escalas de Avaliação da Saúde das Populações Neonatais: Estudo de Simulação Estatística pelo Método de Monte Carlo com Variáveis Ordinais", sob Orientação dos Professores Doutores João Barreiros, Júlia Teles e João Marôco.

O referido trabalho tem como objetivos:

Comparar a performance, em termos de potência e erro de tipo I, da MANOVA paramétrica com diferentes propostas de MANOVA não paramétrica quando se utilizam variáveis ordinais correlacionadas, geradas aleatoriamente.

Comparar os resultados da aplicação dos testes da MANOVA paramétrica e da MANOVA não-paramétrica (dados reais provenientes de "escalas de avaliação" com um formato de medida 3, 4, 5 pontos) com os resultados das simulações estatísticas.

As escalas selecionadas para fazer as aplicações práticas foram: a 1) Early Feeding Skills (EFS); Neonatal Skin Risk Assessment Scale (NSRAS); 3) Neonatal Pain, Agitation, &

Sedation Scale (N-PASS). A primeira com um formato de medida de três classes, a segunda com quatro classes e a terceira com cinco classes, usando como variável independente a "idade gestacional" por considerarmos que é uma das variáveis mais importantes à sobrevivência da criança prematura e que influencia todos os sistemas e especificamente os que vamos analisar neste trabalho: a alimentação, a integridade da pele e a dor do recém-nascido prematuro.

Assim, vem por este meio solicitar cedência de dados referentes a trezentas observações relativas à aplicação da escala de Observação das Competências Precoces na Alimentação Oral – EFS, e da variável idade gestacional de forma a dar resposta a segundo objetivo da sua investigação, comprometendo-se a cumprir os pressupostos legais e éticos relacionados com a utilização dos dados.

Agradece antecipadamente

Com os melhores cumprimentos,
grasias

Maria Alice dos Santos Curado

Maria Alice dos Santos Curado

Maria Alice dos Santos Curado
Prof. Coordenadora

Lisboa, 04 de Março de 2014

*Eu, Maria Thereza Valle e Vasconcellos
Enf. Chefe do Serviço de Neonatologia
do HSFV, não vejo qualquer inconveniente em ceder os dados acima referidos, autorizando a sua utilização*

Lisboa, 7 de Março de 2014

Thereza Vasconcellos
Enf.ª Chefe
Neonatologia

Maria Alice dos Santos Curado
Prof. Coordenadora

Exma. Sra Enfermeira Chefe
Da Unidade de Cuidados Intensivos Neonatais
Exma. Sra. Coordenadora da Equipa de Projeto de Validação da
Neonatal Skin Risk Assessment Scale - NSRAS
Sra. Enfermeira Cláudia Olho Azul
Unidade de Cuidados Intensivos Neonatais
HPP - Hospital de Cascais
Dr. José de Almeida

Maria Alice dos Santos Curado, Professora Coordenadora na Escola Superior de Enfermagem de Lisboa, Especialista em Enfermagem de Saúde Infantil e Pediatria, Mestre em Desenvolvimento da Criança, Doutoranda em Ciências da Motricidade na Especialidade de Comportamento Motor, na Faculdade de Motricidade Humana, Universidade de Lisboa, encontra-se a desenvolver o trabalho de investigação "A Medida e as Escalas de Avaliação da Saúde das Populações Neonatais: Estudo de Simulação Estatística pelo Método de Monte Carlo com Variáveis Ordinais", sob Orientação dos Professores Doutores João Barreiros, Júlia Teles e João Marôco.

O referido trabalho tem como objetivos:

- 1) Identificar na literatura científica os testes estatísticos mais frequentemente utilizados pelos investigadores da área da saúde infantil e pediatria quando usam "escalas de avaliação" (instrumentos construídos com variáveis com um formato de medida ordinal).

- 2) Comparar a performance, através de estudos de simulação, em termos de potência e erro de tipo I, da MANOVA paramétrica com diferentes propostas de MANOVA não paramétrica quando se utilizam estes instrumentos.
- 3) Enquadrar os resultados da aplicação da MANOVA paramétrica e da MANOVA não paramétrica a dados reais provenientes de "escalas de avaliação" com um formato de medida com três, quatro e cinco pontos, nos resultados das simulações estatísticas.

As escalas selecionadas para fazer as aplicações práticas foram: a 1) Early Feeding Skills (EFS); Neonatal Skin Risk Assessment Scale (NSRAS); 3) Neonatal Pain, Agitation, & Sedation Scale (N-PASS). A primeira com um formato de medida de três classes, a segunda com quatro classes e a terceira com cinco classes, usando como variável independente a "idade gestacional" por considerarmos que é uma das variáveis mais importantes à sobrevivência da criança prematura e que influencia todos os sistemas e especificamente os que são avaliados com estas escalas: Competências na alimentação, a integridade da pele e a dor do recém-nascido.

Assim, vem por este meio solicitar a Vossas Exas. A cedência de dados referentes à aplicação da escala NSRAS (noventa), e da variável idade gestacional de forma a dar resposta ao terceiro objetivo da sua investigação, Comparar dados reais com dados simulados aleatoriamente, comprometendo-se a cumprir os pressupostos legais e éticos relacionados com a utilização dos dados.

Agradece antecipadamente

Com os melhores cumprimentos, pessoais

Tavel Casheira
Autorizo a utilização
da Escala NSRAS do doctor
da coordenadora

Maria Alice dos Santos Curado
Maria Alice dos Santos Curado

Maria Alice dos Santos Curado
Prof. Coordenadora

Lisboa, 07 de Abril de 2014

Maria Alice dos Santos Curado
Prof. Coordenadora

Exma. Sra. Professora Maria Isabel Costa Malheiro

Escola Superior de Enfermagem de Lisboa

Maria Alice dos Santos Curado, Professora Coordenadora na Escola Superior de Enfermagem de Lisboa, Especialista em Enfermagem de Saúde Infantil e Pediatria, Mestre em Desenvolvimento da Criança, Doutoranda em Ciências da Motricidade na Especialidade de Comportamento Motor, na Faculdade de Motricidade Humana, Universidade de Lisboa, encontra-se a desenvolver o trabalho de investigação "A Medida e as Escalas de Avaliação da Saúde das Populações Neonatais e Pediátricas: Estudo de Simulação de Monte Carlo com Variáveis Ordinais", sob Orientação dos Professores Doutores João Barreiros, Júlia Teles e João Marôco.

O referido trabalho tem como objetivos:

Comparar a performance, em termos de potência e erro de tipo I, da MANOVA paramétrica com diferentes propostas de MANOVA não paramétrica quando se utilizam variáveis ordinais correlacionadas, geradas aleatoriamente.

Comparar os resultados da aplicação dos testes da MANOVA paramétrica e da MANOVA não-paramétrica (dados reais provenientes de "escalas de avaliação" com um formato de medida com 3, 4, 5 e 7 pontos, com os resultados das simulações estatísticas.

As escalas selecionadas para fazer as aplicações práticas foram: a 1) Early Feeding Skills (EFS); Neonatal Skin Risk Assessment Scale (NSRAS); 3) Funcional Independence

Measure (FIM). A primeira com um formato de medida de três pontos, a segunda com quatro pontos e a terceira com cinco e sete pontos, usando como variável independente a "idade gestacional" por considerarmos que é uma das variáveis mais importantes à sobrevivência da criança prematura e que influencia todos os sistemas e especificamente os que vamos analisar neste trabalho: a alimentação, a integridade da pele no recém-nascido prematuro e independência funcional em crianças e jovens com espinha bífida..

Assim, vem por este meio solicitar cedência de dados referentes à aplicação da escala FIM e da variável idade gestacional de forma a dar resposta a segundo objetivo da sua investigação, comprometendo-se a cumprir os pressupostos legais e éticos relacionados com a utilização dos dados.

Agradece antecipadamente

Com os melhores cumprimentos, *passadas*

Maria Alice dos Santos Curado
Maria Alice dos Santos Curado

Maria Alice dos Santos Curado
Prof. Coordenadora

Lisboa, 04 de Março de 2014

Detorzo Lisboa 5 de Março 2014
Maria Isabel Dias de São Pedro