

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**Modelos de regressão para identificação de  
marcadores preditivos de asma na descendência  
de mulheres com atopia**

**Tiago Miguel Dias Domingues**

Mestrado em Bioestatística

Trabalho de Projeto orientado por:  
Professora Doutora Maria Fernanda Nunes Diamantino  
Professor Doutor Luís Miguel Nabais Borrego

2016



# Resumo

A atopia pode definir-se como sendo a tendência pessoal e/ou familiar para a produção de Imunoglobulina E (IgE) contra alérgenos ambientais, traduzindo-se por eczema atópico, rinite alérgica e asma alérgica cujas manifestações se iniciam, normalmente, durante a infância e a adolescência. A possibilidade de identificar marcadores de risco para o desenvolvimento de atopia, precocemente, poderá ser o primeiro passo para o desenvolvimento de estratégias de prevenção para os indivíduos em risco.

Assim, inicialmente, com este trabalho pretendeu-se identificar diferenças no perfil imunológico de mulheres grávidas com e sem atopia. Posteriormente, foi estudada a possível relação entre as diferenças encontradas nos parâmetros avaliados e o desenvolvimento de atopia na descendência das mulheres grávidas atópicas. Pretendeu-se, assim, não só reconhecer de que forma se relacionam entre si as variações nas diferentes populações celulares avaliadas, mas também se as suas variações de alguma forma se encontravam associadas com o desenvolvimento de atopia nas crianças. Confirmando-se a sua existência, esta relação entre parâmetros imunes maternos e o desenvolvimento de atopia, poderá potenciar a utilização futura destes parâmetros como marcadores preditivos para o desenvolvimento de atopia.

De acordo com os objetivos definidos, procedeu-se à construção de modelos de regressão logística, considerando como covariáveis os marcadores de risco anteriormente identificados, bem como outras variáveis relevantes. A variável resposta consistiu na presença/ausência de atopia na descendência. Pela aplicação de modelos de regressão logística, verificou-se a influência dos linfócitos B de transição como fator de risco para a atopia na descendência de mulheres grávidas atópicas.

*Palavras-Chave:* atopia, gravidez, parâmetros imunológicos, análise de regressão

# Abstract

Atopy can be defined as the personal and/or familiar tendency to production of immunoglobulin E (IgE) against environmental allergens, resulting in atopic eczema, allergic rhinitis and allergic asthma whose manifestations begin, usually during childhood and adolescence. The possibility to identify risk factors for the development of atopy, early in life, could be the first step for the development of prevention strategies for individuals at risk.

So, initially, this work aimed to identify differences in the immune profile of pregnant women with and without atopy. Subsequently, the possible relationship between the differences in the evaluated parameters and the development of atopy in the offspring of atopic pregnant women was studied. Thus, the aim was not only to recognize how the variation in the different evaluated cell populations were related, but also if their variations were somehow associated with the development of atopy in children. Confirming its existence, the relationship between maternal immune parameters and the development of atopy, may enhance the future use of these parameters as predictive markers for the development of atopy.

According to the objectives set, the construction of regression models were made considering as covariates the previously identified risk markers, as well as other relevant variables. The dependent variable was the presence/absence of atopy in the offspring. Using logistic regression models, transitional B cells were identified as a factor risk for atopy event in the progeny of atopic pregnant women.

*Key-Words:* atopy, pregnancy, immunological parameters, regression analysis

Ao meu avô, *Fernando Jorge da Silva Dias (in memoriam)*

Tiago Dias Domingues

# Agradecimentos

A realização desta tese de mestrado não teria sido possível sem o apoio de algumas pessoas que me acompanharam durante este percurso.

À minha orientadora, Professora Doutora Fernanda Diamantino, agradeço o apoio incondicional, por acreditar comigo que este trabalho seria possível e sobretudo pela amizade demonstrada. Foi uma aprendizagem conjunta que permitiu que este trabalho chegasse a bom porto.

Ao meu orientador, Professor Doutor Luís Miguel Borrego, pelo apoio, disponibilidade e interesse demonstrado.

Um agradecimento muito especial à Professora Doutora Catarina Martins, pela forma como me acolheu, pelo apoio e interesse demonstrados. Por juntos conseguirmos fazer melhor.

À Professora Doutora Teresa Alpuim, por me ter incentivado a gostar da área onde hoje me encontro e não me ter deixado desistir.

Uma referência especial à Professora Doutora Lisete Sousa pelo honroso convite para ser seu assistente. Pela sua amizade, confiança e disponibilidade em me tornar numa pessoa melhor. É sem dúvida uma das pessoas que guardo com carinho e amizade.

Ao Professor Doutor Fernando Sequeira, a quem tenho o privilégio de chamar amigo. Pela sua humildade, espontaneidade e confiança que deposita em mim. Por ter sempre a porta aberta para me receber e ouvir nos bons e maus momentos. Por me ajudar sempre nas minhas conquistas.

Aos meus pais e irmã, José Domingues, Paula Domingues e Susana Domingues, por acreditarem e me darem força todos os dias, por me ajudarem sempre na concretização dos meus objetivos.

Aos meus avós paternos, Celeste de Jesus Miguel e António Domingues por estarem sempre comigo.

Aos meus padrinhos, Fernando Carlos Dias e Eulália Dias por me apoiarem

sempre que preciso.

Um agradecimento especial à Cláudia Bicho, minha amiga e colega de curso. Por me ter ajudado sempre nos momentos mais difíceis do meu percurso académico. Por termos sido muitas vezes o suporte um do outro.

À Maria João Santos, minha amiga e colega de trabalho da Datamedica, Consultoria e Serviços em Bioestatística, Lda, pelo apoio incondicional na realização deste trabalho, por todos os ensinamentos que me ajudaram a crescer como profissional e pessoa. A ela, o meu obrigado, bem como à restante equipa que testemunhou a concretização deste projeto, nas pessoas de Margarida Vicente, Rita Marçal e Rita Ribeiro.

Ao Dr. Luís Fatela, CEO da Datamedica, pelo seu apoio incondicional, bem como ao Sérgio Faria por todas as conversas que me ajudaram a crescer.

À Professora Doutora Ana Moita de Macedo, CEO da KeyPoint, Consultoria Científica, por me ter iniciado na vida profissional.

À Marta, por uma amizade de 18 anos que espero eterna. Não tenho palavras para lhe agradecer tudo o que já fez por mim.

À Sara e ao João, por serem dois grandes pilares. Por tudo o que já passámos, o meu sincero agradecimento.

Aos meus amigos de sempre, por toda a paciência e apoio demonstrados.

Por último, agradecer aos meus alunos, com os quais aprendi imenso.

*A todos o meu Obrigado!*

# Conteúdo

<b>Resumo</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Agradecimentos</b>	<b>iv</b>
<b>1 Introdução</b>	<b>2</b>
1.1 Estado da arte . . . . .	2
1.2 Objetivos do estudo . . . . .	8
1.3 Guia de leitura . . . . .	8
<b>2 Metodologia</b>	<b>10</b>
2.1 Caracterização da base de dados e definição das variáveis . . . . .	11
2.1.1 Variáveis sociodemográficas . . . . .	11
2.1.2 Variáveis antropométricas . . . . .	11
2.1.3 Variáveis clínicas . . . . .	12
2.2 Grupos de estudo . . . . .	13
2.3 Variável resposta . . . . .	13
2.4 Teste de Kolmogorov-Smirnov e teste de Shapiro-Wilk . . . . .	14
2.5 Teste t para amostras independentes . . . . .	16
2.5.1 Populações normais, variâncias desconhecidas e iguais . . . . .	16
2.5.2 Populacionais normais, variâncias desconhecidas e diferentes . . . . .	16
2.6 Teste de Mann-Whitney Wilcoxon . . . . .	18
2.7 Teste de independência do $\chi^2$ . . . . .	19
2.8 Modelos Lineares Generalizados . . . . .	21
2.8.1 O modelo linear clássico . . . . .	22
2.8.2 Família Exponencial . . . . .	24



2.9	Regressão Logística . . . . .	28
2.9.1	Conceitos básicos . . . . .	28
2.9.2	Estimação . . . . .	31
2.9.3	Teste de razão de verossimilhanças . . . . .	32
2.9.4	Interpretação dos coeficientes do modelo . . . . .	33
2.9.5	Seleção de modelos . . . . .	36
2.9.6	Confundimento e interação . . . . .	37
2.10	Qualidade de um modelo de regressão	
	logística . . . . .	38
2.10.1	Análise dos resíduos . . . . .	38
2.10.2	Multicolinearidade . . . . .	40
2.10.3	Observações influentes . . . . .	40
2.10.4	Estatística de Hosmer e Lemeshow . . . . .	41
2.10.5	Capacidade preditiva do modelo . . . . .	42
2.10.6	Curva ROC . . . . .	44
<b>3</b>	<b>Análise exploratória dos dados</b>	<b>46</b>
3.1	Caracterização da amostra . . . . .	46
3.2	Caracterização dos grupos de estudo . . . . .	52
3.3	Análise dos parâmetros imunológicos . . . . .	62
<b>4</b>	<b>Identificação de marcadores de risco para o desenvolvimento de atopia na descendência de mulheres atópicas</b>	<b>69</b>
4.1	Caracterização sociodemográfica e antropométrica das crianças, filhas de mulheres atópicas . . . . .	70
4.2	Identificação de marcadores de risco . . . . .	71
4.3	Análise de regressão . . . . .	83
	4.3.1 Modelos de regressão logística . . . . .	83
	4.3.2 Modelo de regressão logística tendo como covariáveis parâmetros imunológicos percentuais (modelo 1) . . . . .	84
	4.3.3 Modelo de regressão logística tendo como covariáveis parâmetros imunológicos em valor absoluto (modelo 2) . . . . .	95
<b>5</b>	<b>Discussão</b>	<b>103</b>
<b>6</b>	<b>Conclusões</b>	<b>106</b>
6.1	Divulgação de resultados . . . . .	107

<b>Bibliografia</b>	<b>107</b>
<b>Apêndice A Código R - análise de regressão</b>	<b>112</b>
<b>Apêndice B Parâmetros imunológicos analisados no 3.º trimestre de gestação</b>	<b>119</b>
<b>Apêndice C Classificação celular</b>	<b>122</b>

# Lista de siglas e abreviaturas

- ▷ **AIC** *Akaike Information Criterion*
- ▷ **AUC** *Area Under the Curve*
- ▷ **Breg** Células B reguladoras
- ▷ **CD** *Cluster of differentiation*
- ▷ **CTLA-4** *Cytotoxic T-lymphocyte-associated protein 4* / Molécula tipo 4 associada a linfócito T citotóxico
- ▷ **DN** Duplos Negativos
- ▷ **f.d.p** Função densidade de probabilidade
- ▷ **f.m.p** Função massa de probabilidade
- ▷ **FN** Falso Negativo
- ▷ **Foxp3** *Forkhead box protein 3*
- ▷ **FP** Falso Positivo
- ▷ **GA** Mulheres grávidas atópicas
- ▷ **GNA** Mulheres grávidas não atópicas
- ▷ **IC** Intervalo de Confiança
- ▷ **IFN** Interferão
- ▷ **Ig** Imunoglobulina

- ▷ **IgE** Imunoglobulina E
- ▷ **IL** Interleucina
- ▷ **IMC** Índice de Massa Corporal
- ▷ **MFI** *Mean Fluorescence Intensity* / Intensidade de fluorescência média
- ▷ **MLG** Modelos Lineares Generalizados
- ▷ **NGA** Mulheres não grávidas atópicas
- ▷ **NGNA** Mulheres não grávidas não atópicas
- ▷ **OR** *Odds Ratio*
- ▷ **Th** Células T *helper* / auxiliaadoras
- ▷ **TNF** *Tumor necrosis factor* / Fator de necrose tumoral
- ▷ **Treg** Células T reguladoras
- ▷ **ROC** *Receiver Operating Characteristic*
- ▷ **VIF** *Variance Inflation Factor*
- ▷ **VN** Verdadeiro Negativo
- ▷ **VP** Verdadeiro Positivo
- ▷ **VPN** Valor Preditivo Negativo
- ▷ **VPP** Valor Preditivo Positivo

# Lista de Tabelas

2.1	Distribuições (algumas) pertencentes à família exponencial . . .	27
3.1	Idade (anos) . . . . .	46
3.2	Classificação IMC . . . . .	48
3.3	Nível de escolaridade . . . . .	48
3.4	Pressões arteriais (mmHg) . . . . .	49
3.5	Classificação da pressão arterial . . . . .	49
3.6	Grupos de estudo . . . . .	52
3.7	Variáveis antropométricas (mulheres saudáveis) . . . . .	52
3.8	Nível de escolaridade (mulheres saudáveis) . . . . .	53
3.9	Pressões arteriais (mulheres saudáveis) . . . . .	54
3.10	História familiar de atopia (mulheres saudáveis) . . . . .	54
3.11	Número de gravidezes anteriores (mulheres saudáveis) . . . . .	55
3.12	Número de filhos (mulheres saudáveis) . . . . .	55
3.13	Parto de termo (mulheres saudáveis) . . . . .	55
3.14	Trabalho de parto (mulheres saudáveis) . . . . .	56
3.15	Tipo de parto (mulheres saudáveis) . . . . .	56
3.16	Variáveis antropométricas (mulheres atópicas) . . . . .	57
3.17	Nível de escolaridade (mulheres atópicas) . . . . .	58
3.18	Pressões arteriais (mulheres atópicas) . . . . .	59
3.19	História familiar de atopia (mulheres atópicas) . . . . .	59
3.20	Número de gravidezes anteriores (mulheres atópicas) . . . . .	60
3.21	Número de filhos (mulheres atópicas) . . . . .	60
3.22	Parto de termo (mulheres atópicas) . . . . .	60
3.23	Trabalho de parto (mulheres atópicas) . . . . .	61
3.24	Tipo de parto (mulheres atópicas) . . . . .	61
3.25	Análise de diferenças entre grávidas saudáveis e não grávidas saudáveis relativamente aos parâmetros imunológicos avaliados no 3.º trimestre . . . . .	64

3.26	Análise de diferenças entre grávidas saudáveis e não grávidas saudáveis relativamente aos parâmetros imunológicos avaliados no 3.º trimestre (continuação) . . . . .	65
3.27	Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre de gestação . . . . .	65
3.28	Análise de diferenças entre não grávidas saudáveis e não grávidas atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre . . . . .	66
3.29	Análise de diferenças entre não grávidas saudáveis e não grávidas atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre (continuação) . . . . .	67
3.30	Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no dia do parto . . . . .	68
3.31	Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no pós-parto . . . . .	68
4.1	Problemas neonatais de atopia . . . . .	70
4.2	Características antropométricas . . . . .	70
4.3	Peso (gramas) . . . . .	71
4.4	Comparação entre D. atópica e D. não atópica relativamente aos parâmetros imunológicos maternos avaliados no 3.º trimestre de gestação . . . . .	73
4.5	Análise de associação entre os parâmetros imunológicos e presença/ausência de atopia na descendência . . . . .	80
4.6	Resultados do modelo para a presença/ ausência de atopia na descendência (CD3+) % . . . . .	85
4.7	Resultados do modelo para a presença/ ausência de atopia na descendência (Breg 24+38%) . . . . .	87
4.8	Resultados do modelo para a presença/ ausência de atopia na descendência (B transição %) . . . . .	89
4.9	Avaliação da multicolinearidade . . . . .	91
4.10	Resultados do modelo para a presença/ ausência de atopia na descendência . . . . .	92
4.11	Resultados do modelo final para a presença/ ausência de atopia na descendência . . . . .	93

4.12	Resultados do modelo para a presença/ ausência de atopia na descendência (Bm1) . . . . .	96
4.13	Resultados do modelo para a presença/ ausência de atopia na descendência (B memória) . . . . .	98
4.14	Resultados do modelo para a presença/ ausência de atopia na descendência (B memória não <i>switched</i> ) . . . . .	100
4.15	Avaliação da multicolinearidade . . . . .	101
B.1	Parâmetros imunológicos analisados . . . . .	120
B.2	Parâmetros imunológicos analisados (continuação) . . . . .	121

# Lista de Figuras

1.1	Contacto com alérgenos. Imagem disponível em <i>gluten celi-aki.blogspot.com</i> . . . . .	2
1.2	A influência do ambiente pré-natal no desenvolvimento de doenças alérgicas e atopia. Retirado de Hawrylowicz <i>et al.</i> Treg - Células T reguladoras. . . . .	6
2.1	Curva ROC . . . . .	45
3.1	Diagrama em caixa de bigodes relativo à variável idade . . . . .	47
3.2	Diagrama em caixa de bigodes - P. sistólica . . . . .	50
3.3	Diagrama em caixa de bigodes - P. diastólica . . . . .	50
3.4	Diagrama em caixa de bigodes relativo à idade (mulheres saudáveis) . . . . .	53
3.5	Diagrama em caixa de bigodes relativo ao IMC (mulheres saudáveis) . . . . .	53
3.6	Diagrama em caixa de bigodes relativo à idade (mulheres atópicas) . . . . .	57
3.7	Diagrama em caixa de bigodes relativo ao IMC (mulheres atópicas) . . . . .	57
4.1	Curva ROC para linfócitos T (CD3+) % . . . . .	73
4.2	Curva ROC para linfócitos B transição % . . . . .	74
4.3	Curva ROC para linfócitos Breg 24 <sup>hi</sup> +38 <sup>hi</sup> % . . . . .	75
4.4	Curva ROC para linfócitos B (CD19+) % . . . . .	76
4.5	Curva ROC para linfócitos B memória . . . . .	77
4.6	Curva ROC para linfócitos Bm1 . . . . .	78
4.7	Curva ROC para linfócitos B memória não <i>switched</i> . . . . .	79
4.8	Diagrama em caixa de bigodes paralelas relativo à variável linfócitos T (CD3+) % . . . . .	81



4.9	Diagrama em caixa de bigodes paralelas relativo à variável Btransição % . . . . .	82
4.10	Diagrama em caixa de bigodes paralelas relativo à variável linfócitos Breg $24^{hi}+38^{hi}$ % . . . . .	82
4.11	Diagrama em caixa de bigodes relativo à variável Bm1 . . . . .	82
4.12	Diagrama em caixa de bigodes relativo à variável B memória . . . . .	83
4.13	Diagrama em caixa de bigodes relativo à variável B memória não switched . . . . .	83
4.14	Análise dos resíduos (CD3+ %) . . . . .	85
4.15	<i>Medida de leverage</i> (CD3+ %) . . . . .	85
4.16	Distância de <i>Cook</i> (CD3+ %) . . . . .	86
4.17	Análise dos resíduos (B reguladoras $24^{hi}+38^{hi}\%$ ) . . . . .	87
4.18	<i>Medida de leverage</i> (B reguladoras $24^{hi}+38^{hi}\%$ ) . . . . .	87
4.19	Distância de <i>Cook</i> (B reguladoras $24^{hi}+38^{hi}\%$ ) . . . . .	88
4.20	Análise dos resíduos (B de transição %) . . . . .	89
4.21	<i>Medida de leverage</i> (B de transição %) . . . . .	89
4.22	Distância de <i>Cook</i> (B de transição %) . . . . .	90
4.23	Análise dos resíduos (modelo 1) . . . . .	94
4.24	<i>Medida de leverage</i> (modelo 1) . . . . .	94
4.25	Distância de <i>Cook</i> (modelo 1) . . . . .	94
4.26	Curva ROC do modelo 1 . . . . .	95
4.27	Análise dos resíduos (Bm1) . . . . .	96
4.28	<i>Medida de leverage</i> (Bm1) . . . . .	96
4.29	Distância de <i>Cook</i> (Bm1) . . . . .	97
4.30	Análise dos resíduos (B memória) . . . . .	98
4.31	<i>Medida de leverage</i> (B memória) . . . . .	98
4.32	Distância de <i>Cook</i> (B memória) . . . . .	99
4.33	Análise dos resíduos (B memória não <i>switched</i> ) . . . . .	100
4.34	<i>Medida de leverage</i> (B memória não <i>switched</i> ) . . . . .	100
4.35	Distância de <i>Cook</i> (B memória não <i>switched</i> ) . . . . .	101

# Capítulo 1

## Introdução

### 1.1 Estado da arte

A atopia pode ser definida como uma tendência pessoal e/ou familiar para a produção de anticorpos da classe (IgE) em resposta a antigénios ambientais, normalmente proteicos, conhecidos como alergénios (Johansson, SG; *et al.*, 2004). A atopia manifesta-se inicialmente na infância e adolescência, com a marcha atópica a ser considerada quando há o desenvolvimento de eczema atópico na infância, acompanhado do subsequente estabelecimento de rinite alérgica e asma alérgica ao longo da infância e adolescência (Bantz, SK; *et al.*, 2014).



Figura 1.1: Contacto com alergénios. Imagem disponível em *gluten celi-aki.blogspot.com*

Nas últimas décadas, a prevalência das doenças atópicas (eczema atópico, rinite alérgica e asma alérgica) tem vindo a aumentar significativamente, em

particular nos países ocidentalizados, estimando-se que afetem, aproximadamente, 20% da população mundial (Bantz, SK; *et al.*, 2014). Na infância, as doenças atópicas manifestam-se com mais frequência em crianças do sexo masculino, mas a partir da adolescência, é no sexo feminino que se encontram mais frequentemente manifestações de doenças atópicas como a asma alérgica.

Um dos elementos-chave das doenças atópicas parece ser a orquestração da inflamação alérgica por células  $Th_2$  (Prussin, C; *et al.*, 2010). As células  $Th_2$ , ao serem ativadas induzem os linfócitos B a produzir anticorpos IgE específicos de antígeno (alergénio). A resposta alérgica tem três fases: a sensibilização, a reação de hipersensibilidade imediata e a reação tardia.

Atualmente sabe-se que distintos compartimentos celulares têm provado o seu papel nos processos inflamatórios alérgicos tais como linfócitos  $Th_{17}$ , células T reguladoras, mastócitos, e linfócitos B (Shah, A, 2012).

Em relação à importância das células B no contexto da atopia, durante muito tempo, considerou-se que a sua função se limitava à produção de IgE específicas para alergénio, mas na realidade pouco mais se sabe sobre o seu papel na atopia.

As recentemente descritas células B reguladoras (Breg) deram uma nova perspectiva à avaliação do compartimento B no contexto da atopia (Braza, F; *et al.*, 2014; Noh, G; Lee, JH, 2011; Jin, G; *et al.*, 2014). No entanto, a caracterização desta população celular, as células Breg, não é consensual, e, em humanos, funções reguladoras têm vindo a ser propostas para vários compartimentos celulares B distintos: as células B  $IgD^+CD27^-$  naive, as células B de transição  $CD24^{hi}CD38^{hi}$ , as células B10  $CD24^{hi}CD27^+$  ou as células B1 orc  $IgD^{hi}CD43^-CD27^+CD70^-CD11b^+$  (Sanz, I, 2014; Iwata, Y; *et al.*, 2011; Blair, PA; *et al.*, 2010; Griffin, DO; Rothstein, TL, 2012; Amu, S; *et al.*, 2010).

É indiscutível porém que são necessários mais estudos para uma melhor caracterização imunofenotípica e avaliação funcional das células Breg, particularmente nos humanos, uma vez que estas células podem ser fundamentais para o desenvolvimento de novas abordagens terapêuticas para patologias alérgicas, tal como as células Treg.

O reconhecimento do fator de transcrição Foxp3, bem como a caracterização de outros marcadores característicos (CD25 e CD127) nas células Treg foi crucial para a sua identificação (Hori, S; Nomura, T; Sakaguchi, S, 2003; Fontenot, JD; Gavin, MA; Rudensky, AY, 2003). Concretamente, permitiu o isolamento de células Treg, existindo hoje em dia um conhecimento substan-

cial acerca das suas funções na fisiopatologia de várias doenças. Mais ainda, estes avanços na identificação e caracterização das células Treg têm possibilitado a sua aplicação, concretamente como alvos terapêuticos a desenvolver em várias áreas.

- **Atopia e gravidez**

Na perspetiva da resposta imunitária, as semelhanças entre gravidez e atopia/alergia são consideráveis. Esta proximidade imunológica desde cedo animou a comunidade científica, no sentido de compreender o impacto das doenças atópicas na gestação e o impacto da gravidez na doente atópica. De facto, a fim de proteger o feto de respostas imunitárias maternas agressivas contra os antigénios paternos expressos nas células fetais, estabelece-se durante a gravidez um estadio de imunossupressão fisiológica. Os mecanismos de tolerância imunológica têm de ser regulados, no entanto, de modo a continuar a assegurar a normal proteção da mãe contra outros antigénios estranhos.

Tal como se observa na atopia, a gravidez parece assentar num desvio da resposta imune para um perfil  $Th_2$ , com um forte componente anti-inflamatória, para a qual contribuem também interações reguladas entre o sistema imune materno e o sistema imune fetal. Citocinas  $Th_2$  produzidas ao nível do interface maternofetal contribuem também para a manutenção da gravidez, interagindo com o sistema imune e o sistema endócrino em simultâneo (Sykes, L; *et al.*, 2012; Piccinni, MP; *et al.*, 2015). Como anteriormente referido, tem vindo a ser postulado que mais do que um desvio da resposta para o perfil  $Th_2$ , a gravidez assume mecanismos reguladores que equilibram a produção de citocinas (Halonen, M; *et al.*, 2009). Por um lado, não se reportaram eventos abortivos em ratinhos deficitários na secreção de citocinas  $Th_2$ , obtidos por manipulação genética. Em paralelo, citocinas tipicamente  $Th_1$ , como IFN- $\gamma$  ou TNF- $\alpha$ , têm um papel importante no desenvolvimento da placenta (Svensson-Arvelund, J; *et al.*, 2013). Mais ainda, a ocorrência de abortos espontâneos foi apontada em situações com predomínio de resposta  $Th_1$  (Makhseed, M; *et al.*, 2001; Raghupathy, R; *et al.*, 2000). São mais frequentemente relatadas contudo associações entre a resposta  $Th_1$  e pré-eclampsia ou ocorrência de partos pré-termo (Sykes, L; *et al.*, 2012). As células  $Th_{17}$ , bem como o tipo de respostas que gerem, também têm a sua importância na ocorrência de abortos espontâneos. Se associadas, as respos-

tas  $Th_{17}$  e  $Th_1$  podem representar uma combinação com efeitos ainda mais negativos no curso da gravidez. A manutenção da tolerância materno-fetal é recorrentemente associada aos efeitos reguladores da resposta  $Th_2$ , mas sobretudo a mecanismos impostos pelas células T reguladoras. Também os linfócitos Breg parecem implicados no processo de tolerância materno-fetal.

- **A atopia na descendência da mulher atópica**

Para além do impacto da gravidez no decurso de doenças atópicas maternas, a história materna de atopia representa também um fator de risco para o desenvolvimento de atopia na criança, ainda mais importante do que a história paterna de atopia. Assim, embora deva ser realçada a transmissão de um fundo genético com maior predisposição para o perfil de resposta  $Th_2$ , estes dados são também muito sugestivos de que a imunomodulação observada durante a gravidez possa igualmente afetar o sistema imunitário do feto em desenvolvimento. Assim, o impacto da história alérgica materna no desenvolvimento de atopia na criança, leva-nos a atentar na importância do período da vida intrauterino para o desenvolvimento do sistema imune fetal. Ainda que não compreendida na sua totalidade, a modulação materna das respostas imunes na criança estender-se-á para além da transmissão genética. Recorde-se, por exemplo, a proteção que confere a exposição materna a ambientes rurais para a sensibilização atópica da progénie. Dados estes que sustentam a importância do período pré-natal para o desenvolvimento futuro de doença na criança. Assim, ainda que a suscetibilidade para atopia seja uma herança genética, os fatores que levam ao aparecimento de uma doença atópica são multifatoriais.

Alterações no compartimento de células T circulante foram já relatados em mulheres grávidas saudáveis e atópicas, mas os dados sobre o compartimento da célula B durante a gestação, em mulheres saudáveis e atópicas, são escassos (Muzzio, DO; *et al.*, 2014; Toldi, G; *et al.*, 2011; Bohacs, A; *et al.*, 2010). Numa abordagem abrangente das principais subpopulações T e B circulantes na grávida do terceiro trimestre, este estudo visou colmatar a pouca informação acerca do sistema imune materno na gravidez. Com diferentes abordagens fenotípicas de caracterização de subpopulações T e B (células incluindo Treg e Breg) do sangue periférico de mulheres atópicas e não atópicas, grávidas e não grávidas, este trabalho pretendeu analisar os perfis de diferenciação de células T e B presentes na periferia. Desta forma,

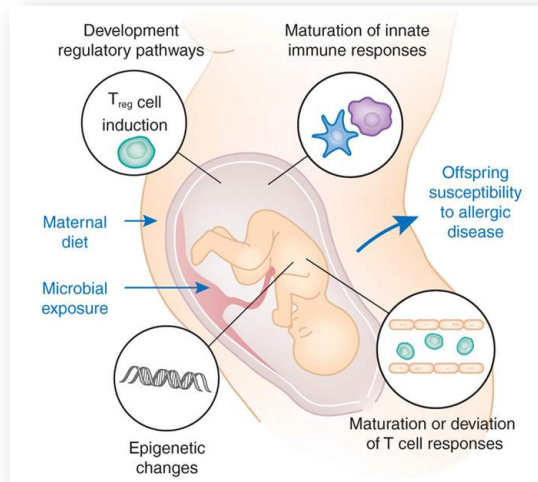


Figura 1.2: A influência do ambiente pré-natal no desenvolvimento de doenças alérgicas e atopia. Retirado de Hawrylowicz *et al.* Treg - Células T reguladoras.

a avaliação de alterações em distintos compartimentos T e B circulantes em mulheres atópicas durante a gravidez, permitirá reconhecer o seu impacto no desenvolvimento de atopia na sua descendência.

Em última análise, a possibilidade de identificar marcadores de risco precoces para atopia na descendência de mulheres atópicas, poderá ser o primeiro passo para o desenvolvimento de estratégias de prevenção para os indivíduos em risco.

• Estudos anteriores

Autor e Ano	Objetivo	Variáveis	Metodologia
Ganesa Wegienka <i>et al.</i> , 2010	Avaliar a trajetória das células T reguladoras durante a gravidez, nomeadamente durante três períodos após o puerpério (1, 6 e 12 meses).	Estatuto primogénito; primeira gravidez; sexo do bebé; presença de asma; atopia; contacto com animais; fumadora; idade gestacional; tipo de parto; pré-eclampsia/ gravidez induzida; hipertensão; raça	Construção de modelos multinível, ajustados através do método da máxima verosimilhança, incluindo modelos de efeitos fixos e modelos de efeitos aleatórios. A variável resposta foi definida como o número de meses entre a colheita de sangue e o nascimento da criança.
Anikó Bohács <i>et al.</i> , 2010	Avaliar os mecanismos imunológicos nas interações entre asma e gravidez.	Idade (em anos); idade gestacional no momento de avaliação (semanas); distribuição da idade gestacional em três trimestres); idade gestacional no parto (semanas); peso do bebé (gramas); sexo do bebé; peso das crianças do sexo feminino.	Utilização do teste de Kruskal-Wallis para comparações entre mais do que 2 grupos. No caso de existirem diferenças estatisticamente significativas, foi aplicado o teste de Dunn para comparações múltiplas. Utilização de regressão linear simples. Teste do $\chi^2$ para testar associações. Foi ainda utilizado o teste de Mann-Whitney Wilcoxon para comparações de dois grupos.

## 1.2 Objetivos do estudo

Para a realização deste projeto foram definidos os seguintes objetivos:

- Efetuar uma análise descritiva de todas as variáveis em estudo para a amostra total;
- Efetuar uma análise descritiva de todas as variáveis em estudo para as mulheres grávidas e mulheres não grávidas, com e sem atopia;
  - Caracterizar demográfica e antropométricamente as mulheres grávidas e mulheres não grávidas, com e sem atopia;
  - Identificar diferenças, nos parâmetros imunológicos, entre mulheres grávidas e mulheres não grávidas, com e sem atopia;
- Avaliar parâmetros imunológicos maternos durante o terceiro trimestre de gestação, durante o parto e no puerpério (6 semanas após parto);
- Estudar a possível relação entre as diferenças encontradas nos parâmetros avaliados e o desenvolvimento de atopia na descendência das mulheres grávidas acompanhadas, através da construção de modelos de regressão logística.

## 1.3 Guia de leitura

Antes de começar, façamos uma breve referência ao modo como este trabalho se encontra organizado.

No Capítulo 1 (Estado da arte), apresenta-se o essencial sobre o que é a atopia, como se manifesta, a sua evolução ao longo dos últimos anos e os objetivos do projeto. No fundo, um breve enquadramento sobre o que nos vai ocupar este trabalho.



No Capítulo 2 (Metodologia) é apresentado o tipo de estudo, a recolha de dados, a caracterização da base de dados e a definição das variáveis. São apresentados os grupos de estudo. Ao longo deste projeto trabalhámos com mulheres grávidas saudáveis, não grávidas saudáveis, grávidas atópicas e não grávidas atópicas. É também apresentada toda a metodologia estatística na qual assenta este trabalho, tendo em conta a definição dos objetivos apresentados no Capítulo 1.

No Capítulo 3 (Análise exploratória dos dados) apresenta-se uma caracterização da amostra recolhida, no que se refere a variáveis sociodemográficas e antropométricas, bem como uma caracterização do historial clínico das mulheres incluídas no estudo.

De forma análoga é feita uma caracterização dos grupos em estudo. Neste capítulo é ainda feita uma análise dos parâmetros imunológicos recolhidos, no sentido de perceber o comportamento das diferentes populações celulares nos diferentes grupos de estudo.

No Capítulo 4 (Identificação de marcadores de risco para o desenvolvimento de atopia na descendência de mulheres atópicas) são apresentados os resultados principais deste projeto. Começamos com uma breve caracterização das crianças filhas de mães atópicas, identificando quais destas são consideradas como descendência atópica ou não atópica.

Dos parâmetros imunológicos maternos recolhidos, são identificados aqueles que são considerados marcadores de risco para a descendência através da construção de curvas ROC para identificação de pontos de corte e do cálculo de medidas de risco, como seja o *odds ratio*.

A última secção deste capítulo dedica-se à análise de regressão logística, de modo a perceber-se, de entre os marcadores de risco identificados, quais aqueles que de facto influenciam a resposta à presença/ausência de atopia na descendência.

Os Capítulos 5 e 6 são reservados à discussão dos resultados obtidos e principais conclusões, respetivamente. Por último, apresenta-se a bibliografia e anexos.

# Capítulo 2

## Metodologia

Para este estudo foram recrutadas mulheres grávidas e mulheres não grávidas, com e sem atopia seguidas no Hospital CUF Descobertas, Lisboa, no período de julho de 2013 a dezembro de 2014. Todas as mulheres grávidas foram avaliadas durante o 3.º trimestre de gravidez, no dia do parto e após o puerpério (6 semanas após o parto). A todas as mulheres foi feita uma entrevista prévia com vista à recolha de dados para este estudo. Foram avaliados vários parâmetros imunológicos maternos durante os três momentos de avaliação definidos anteriormente. Além disso, foi efetuada uma avaliação clínica das crianças, filhas de mães atópicas, aos seis meses de idade. Este trabalho trata-se de um estudo transversal, com mulheres com idade igual ou superior a 18 anos.

Com estes dados pretendeu-se caracterizar demográfica e antropométricamente os grupos de mulheres grávidas e mulheres não grávidas, com e sem atopia, bem como identificar marcadores de risco para o desenvolvimento de atopia na descendência das mulheres grávidas atópicas.

A identificação de marcadores de risco para a descendência foi feita com base na construção de modelos de regressão logística, tendo como variável resposta a presença/ausência de atopia na descendência de mulheres grávidas atópicas.

A análise estatística foi realizada com recurso ao *software* SPSS (*Statistical Packages for the Social Sciences*) versão 22 e ao *software* R versão 3.2.2.

Todos os resultados com um valor de prova (*p-value*) inferior a 0,05 foram considerados estatisticamente significativos.

## 2.1 Caracterização da base de dados e definição das variáveis

A base de dados utilizada para este estudo contém registos de 134 mulheres com idade igual ou superior a 18 anos e que decidiram entrar no estudo, através do seu consentimento.

### 2.1.1 Variáveis sociodemográficas

Variável	Significado	Codificação
Escolaridade	Nível de escolaridade das mulheres em estudo	1- Ensino básico; 2- Ensino secundário; 3- Ensino superior
Profissional	Atividade profissional	1- Conta de outrém; 2- Conta própria; 3- Desempregada

### 2.1.2 Variáveis antropométricas

**Idade** - Variável que representa a idade, em anos.

**Peso** - Variável que representa o peso, em kg.

**Altura** - Variável que representa a altura, em cm.

**IMC** - Variável que representa o Índice de Massa Corporal, em  $kg/m^2$ .

### 2.1.3 Variáveis clínicas

Variável	Significado	Codificação
Motivo_Internamento	Motivo de internamento	1- Cesariana; 2- Trabalho de parto; 3- Indução do trabalho de parto; 4- Pré-eclampsia
Analgesia	Recebeu analgesia	1- Sim; 2- Não
Anestesia	Recebeu anestesia	1- Sim; 2- Não
História_atopia	História familiar de atopia	1- Não; 2- Pai; 3- Avós maternos; 4- Avós maternos
Parto_termo	Parto de termo	1- Sim; 2- Não
Trabalho_parto	Trabalho de parto	1- Cesarian; 2- Espontâneo; 3- Induzido
Tipo_parto	Tipo de parto	1- Cesariana eletiva; 2- Cesariana intraparto; 3- Eutócico; 4- Vaginal com ventosa
Gravidezes	Número de gravidezes anteriores	
Nfilhos	Número de filhos	
Habitos_tabagicos	Hábitos tabágicos	
Fumador_Passivo	Fumadoras passivas	
Sistolica	Pressão arterial sistólica, em mmHg	
Diastolica	Pressão arterial diastólica, em mmHg	

## 2.2 Grupos de estudo

<b>Grupo</b>	<b>Significado</b>	<b>Codificação</b>	<b>Dimensão</b>
GA	Grupo de mulheres grávidas atópicas	0- Não; 1- Sim	$n = 24$
GNA	Grupo de mulheres grávidas não atópicas	0- Não; 1- Sim	$n = 43$
NGA	Grupo de mulheres não grávidas atópicas	0- Não; 1- Sim	$n = 32$
NGNA	Grupo de mulheres não grávidas não atópicas	0- Não; 1- Sim	$n = 35$

## 2.3 Variável resposta

Para este trabalho de projeto foi considerada como variável de interesse a presença/ ausência de atopia na descendência de mulheres grávidas atópicas.

Nas secções seguintes apresenta-se toda a metodologia estatística aplicada neste trabalho. Perceber com que dados estamos a trabalhar é essencial para se definir a metodologia estatística a utilizar. Assim, começamos com uma abordagem ao teste de Kolmogorov-Smirnov e ao teste de Shapiro-Wilk, testes não paramétricos que nos permitem inferir sobre a distribuição de determinada população.

## 2.4 Teste de Kolmogorov-Smirnov e teste de Shapiro-Wilk

O teste de Kolmogorov-Smirnov destina-se a averiguar se uma amostra pode ser considerada como proveniente de uma população com uma determinada distribuição. Seja  $F(x)$  a função de distribuição de uma população  $X$ ,  $(X_1, \dots, X_n)$  uma amostra aleatória recolhida de  $X$ . Pretende-se então testar as seguintes hipóteses

$$H_0 : X \sim F(x)$$

*vs.*

$$H_1 : X \sim \text{outra distribuição diferente de } F(x)$$

Seja  $F_n(x)$  a função de distribuição empírica e consideremos as estatísticas unilaterais

$$D_n^+ = \max\{\max_{1 \leq i \leq n} [\frac{i}{n} - F_n(x_{(i)})], 0\},$$

e

$$D_n^- = \max\{\max_{1 \leq i \leq n} [F_n(x_{(i)}) - \frac{i-1}{n}], 0\}.$$

A estatística de teste é dada por

$$D_n = \max(D_n^+, D_n^-),$$

em que, sob  $H_0$ ,  $D_n$  tem distribuição de Kolmogorov-Smirnov, cujos quantis se encontram tabelados.

Face ao valor observado da estatística de teste,  $d_n$ , rejeita-se  $H_0$  se  $d_n \geq d_{n,1-\alpha}$ , em que  $d_{n,1-\alpha}$  é o quantil de probabilidade  $1 - \alpha$  da distribuição de Kolmogorov-Smirnov.

Contudo, a aplicação do teste de Kolmogorov-Smirnov só se deve fazer na presença de amostras suficientemente grandes ( $n \geq 30$ ). Caso contrário, a alternativa passa por utilizar o teste de Shapiro-Wilk.

Segundo Maroco (2010), o teste de Shapiro-Wilk, é também um teste não paramétrico que permite averiguar se uma determinada população segue uma determinada distribuição. Temos portanto que as hipóteses em teste são as mesmas do teste de Kolmogorov-Smirnov, sendo o valor observado da estatística de teste dado por

$$W_{obs} = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2},$$

onde  $x_{(i)}$  são os valores observados na amostra ordenada proveniente de  $X$ ,  $\bar{x}$  é a estimativa do valor médio de  $X$  e  $a_i$  são constantes geradas a partir do valor médio, variância e covariância de  $n$  estatísticas ordinais com distribuição normal padrão. Estes valores de  $a_i$  encontram-se tabelados.

Valores pequenos de  $W$  indicam que a distribuição da população não é  $F(x)$ . Os valores críticos de  $W$  também se encontram tabelados.

## 2.5 Teste t para amostras independentes

### 2.5.1 Populações normais, variâncias desconhecidas e iguais

Sejam  $\mu_X$  e  $\mu_Y$  os valores médios de duas populações  $X$  e  $Y$ , respetivamente. O teste t para amostras independentes é utilizado para averiguar se são iguais os valores médios populacionais, dadas duas amostras independentes  $X = (x_1, \dots, x_n)$  e  $Y = (y_1, \dots, y_m)$ , recolhidas de uma população  $X$  e  $Y$ , respetivamente. Pretende-se testar:

$$H_0 : \mu_X = \mu_Y \text{ vs. } H_1 : \mu_X \neq \mu_Y$$

Admitindo populações normais e variâncias desconhecidas e iguais, temos que a estatística de teste é dada por:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t_{(n_X+n_Y-2)} \text{ se } H_0 \text{ verdadeira,}$$

em que  $S_p^2$ , também chamada de *pooled variance* é dada por:

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}.$$

Fixado um nível de significância  $\alpha$ , rejeita-se a hipótese nula se  $|t| \geq t_{n_X+n_Y-2; 1-\alpha/2}$ , em que  $t_{1-\alpha/2}$  designa o quantil de probabilidade  $1 - \alpha/2$  de uma t-Student com  $(n_X + n_Y - 2)$  graus de liberdade.

### 2.5.2 Populacionais normais, variâncias desconhecidas e diferentes

Para investigar a igualdade das variâncias de uma forma rigorosa deve-se usar um teste como, por exemplo, o teste  $F$  (apenas no caso de populações



normais) ou o teste de Levene (bastante robusto para afastamentos do modelo normal) para testar

$$H_0 : \sigma_X^2 = \sigma_Y^2 \quad vs. \quad H_1 : \sigma_X^2 \neq \sigma_Y^2$$

Caso seja rejeitada a hipótese nula, no caso de populações normais, podemos usar a aproximação de Welch-Satterwaite para testar

$$H_0 : \mu_X - \mu_Y = \delta_0 \quad vs. \quad H_1 : \mu_X - \mu_Y \neq \delta_0$$

Então, para populações normais, variâncias desconhecidas e diferentes, a estatística de teste é

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Se  $H_0$  verdadeira, a estatística de teste tem distribuição aproximada t-Student com  $k$  graus de liberdade, em que

$$k = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X-1} + \frac{(s_Y^2/n_Y)^2}{n_Y-1}}$$

utilizando-se na prática o natural mais próximo do valor de  $k$ . Rejeita-se  $H_0$  se  $|t| > t_{1-\alpha/2}(k)$ .

## 2.6 Teste de Mann-Whitney Wilcoxon

O teste de Mann-Whitney Wilcoxon é apropriado para averiguar se são iguais as medianas ( $\chi_{1/2}(X)$  e  $\chi_{1/2}(Y)$ ) de duas populações  $X$  e  $Y$ . Sejam  $\mathbf{X} = (X_1, \dots, X_n)$  e  $\mathbf{Y} = (Y_1, \dots, Y_m)$  duas amostras aleatórias de  $X$  e  $Y$ , respetivamente, e suponha-se que  $n \leq m$ .

**Hipóteses em teste:**

$$H_0 : \chi_{1/2}(X) = \chi_{1/2}(Y) \quad vs. \quad H_1 : \chi_{1/2}(X) \neq \chi_{1/2}(Y)$$

**Procedimento:** juntam-se os elementos das duas amostras numa só (amostra combinada) identificando a origem ( $X$  ou  $Y$ ) de cada elemento desta nova amostra. Ordena-se a amostra combinada e observam-se os *ranks* ocupados pelos elementos de  $X$  e pelos elementos de  $Y$ .

O teste é desenvolvido com base na soma dos *ranks*, na amostra combinada, pelos elementos da amostra inicial de menor dimensão. No caso de haver uma ou mais observações iguais nas duas amostras iniciais (empates), utiliza-se a média das posições das observações empatadas.

**Estatística de teste:** seja  $R(X_i)$  o *rank* da observação  $X_i, i = 1, \dots, n$ . A estatística de teste é dada por

$$T = \sum_{i=1}^n R(X_i).$$

Sob  $H_0$ , os quantis da distribuição de  $T$  para valores pequenos de  $n$  e  $m$  ( $n \leq 10, m \leq 10$ ), encontram-se tabelados. Para valores de  $n$  e  $m$  suficientemente grandes, tem-se que

$$E(T) = m \frac{N + 1}{2},$$

com  $N = n + m$ .

Pestana e Velosa (2010), provam que no que respeita à variância de cada uma das parcelas, usando o teorema de König,

$$\frac{1}{N} \frac{N(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 = \frac{N^2-1}{12}.$$

Ora, estamos a extrair  $m$  *ranks*, sem reposição, de uma população de  $N = n + m$ . Então,

$$\text{Var}(T) = m \left(\frac{N^2-1}{12}\right) \left(\frac{N-m}{N-1}\right) = \frac{mn(N+1)}{12}.$$

Donde, pelo Teorema Limite Central, tem-se

$$\frac{T - m \frac{N+1}{2}}{\sqrt{\frac{mn(N+1)}{12}}} \xrightarrow{d}_N Z \sim N(0, 1).$$

## 2.7 Teste de independência do $\chi^2$

Quando estamos na presença de dados que resultam de contagens, estes podem ser organizados sob a forma de tabelas de frequências que representam um conjunto de dados classificados simultaneamente segundo várias características de acordo com classificações exaustivas e mutuamente exclusivas.

A estas tabelas dá-se o nome de tabelas de contingência.

Numa fase inicial, estamos interessados em averiguar se as variáveis aleatórias às quais se efetuou uma classificação cruzada e que deram origem a uma dada tabela de contingência são ou não independentes, ou seja, se é possível que exista alguma associação entre elas.

Por questões de notação, considere-se  $\mathbf{x}$  uma amostra de dimensão  $n$  recolhida de uma população  $X$  e  $A$  e  $B$  duas variáveis aleatórias com  $r$  e  $c$  categorias. A tabela de contingência resultante da classificação dos  $n$  elementos de  $\mathbf{x}$ , segundo estas variáveis, será da forma

	$B_1$	$B_2$	$\dots$	$B_c$	
$A_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1c}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2c}$	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$A_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rc}$	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.c}$	$n$

onde  $n_{ij}$  representa o número de elementos da amostra classificados simultaneamente nas categorias  $A_i$  de  $A$  e  $B_j$  de  $B$ ,  $i = 1, \dots, r; j = 1, \dots, c$ . A marginal de linha é representada por  $n_{i.}$  e a marginal de coluna é representada por  $n_{.j}$ . A dimensão da amostra é dada pela soma de uma das marginais, obtendo-se  $n$ .

Associada a esta tabela, tem-se uma tabela de probabilidades onde denotaremos  $p_{ij}$  como a probabilidade de um elemento ser classificado na célula  $(i, j)$ .

Depois de definida a notação a utilizar, tem-se como objetivo num teste de independência testar

$H_0$  : As variáveis são independentes

*vs.*

$H_1$  : As variáveis não são independentes

Considerando as hipóteses postuladas anteriormente, a estatística de teste terá distribuição de um  $\chi^2$  com  $(r - 1)(c - 1)$  graus de liberdade. O valor observado da estatística de teste é dada por

$$X_{obs}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}},$$

onde  $e_{ij}$  representa a frequência esperada na célula  $(i, j)$ . Sob a validade da hipótese nula, rejeita-se  $H_0$  se  $X_{obs}^2 \geq \chi_{1-\alpha; (r-1)(c-1)}^2$ , em que  $\chi_{1-\alpha; (r-1)(c-1)}^2$  representa o quantil de probabilidade  $1 - \alpha$  de um  $\chi^2$  com  $(r - 1)(c - 1)$  graus de liberdade.

## 2.8 Modelos Lineares Generalizados

A análise de regressão surge da necessidade de em muitas situações se pretender estudar a relação entre variáveis, ou mais propriamente, analisar a influência que uma ou mais variáveis, chamadas explicativas, têm sobre uma variável de interesse, chamada variável resposta.

Os Modelos Lineares Generalizados (MLG) que abordaremos nesta secção são uma generalização do modelo linear, onde se pretende que os modelos criados apresentem uma estrutura de regressão linear e que tenham em comum o facto de a variável resposta seguir uma distribuição pertencente a uma família de distribuições com propriedades específicas, a família exponencial. Nelder e Wedderburn (1972) propuseram uma síntese de vários modelos estatísticos a que deram o nome de MLG, como uma extensão dos modelos lineares clássicos.

São considerados como exemplos destes modelos:

- Modelo de regressão linear clássico;
- Modelo de análise de variância e covariância;
- Modelo de regressão logística;
- Modelo de regressão de Poisson;
- Modelo log-linear para tabelas de contingência multidimensionais;
- Modelo *probit* para estudos de proporções, etc.

Antes de introduzirmos o modelo linear clássico, por questões de notação, consideremos uma variável aleatória  $Y$  a que chamaremos variável resposta e um vetor  $\mathbf{x} = (x_1, \dots, x_k)^T$  constituído por  $k$  variáveis explicativas, ou covariáveis, que se acredita explicar parte da variabilidade inerente a  $Y$ . A variável resposta pode ser de qualquer natureza, contínua, discreta ou dicotómica, bem como as covariáveis que se considerem na construção do modelo.

### 2.8.1 O modelo linear clássico

Turkman e Silva (2000) propõem a seguinte caracterização para o modelo linear clássico:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

onde  $\mathbf{Z}$  é uma matriz de dimensão  $(n \times k)$  de covariáveis  $X$ , em que a primeira coluna desta matriz é um vetor unitário e está associada a um vetor  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$  de parâmetros e  $\boldsymbol{\varepsilon}$  é um vetor de erros aleatórios com distribuição  $N_n(\mathbf{0}, \sigma^2\mathbf{I})$ .

Além disso, com a aplicação do modelo linear clássico pretende-se conhecer o valor médio da variável resposta,  $Y$ , em função das covariáveis que se acredita que expliquem parte da variabilidade,  $\mathbf{Z}$ , sendo então definido o valor médio condicional  $E(\mathbf{Y}|\mathbf{Z}) = \boldsymbol{\mu}$ , com  $\boldsymbol{\mu} = \mathbf{Z}\boldsymbol{\beta}$ .

Através do modelo clássico, Nelder e Wedderburn (1972) mostraram que a maioria dos problemas estatísticos de diversas áreas do conhecimento podem ser tratados sob a forma de modelos de regressão. Os MLG propostos por Nelder e Wedderburn (1972), além de serem uma extensão do modelo linear clássico, envolvem além da variável resposta e das covariáveis que já foram faladas anteriormente, ainda:

- **componente aleatória**, em que  $Y$  é uma variável aleatória da qual se recolhem  $n$  observações independentes e cuja distribuição de probabilidades faz parte da família exponencial, que abordaremos mais à frente;
- **componente sistemática** que consiste numa combinação linear de variáveis preditoras (covariáveis).  
Consideremos que existem  $k$  variáveis preditoras e  $n$  observações

$$\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}, \forall i \in \{1, \dots, n\}.$$

Pode-se simplificar a notação construindo a matriz do modelo  $\mathbf{Z}_{n \times (k+1)}$ : uma coluna unitária associada à constante aditiva ( $\beta_0$ ) e  $k$  colunas adicionais dadas pelas observações de cada variável preditora,

$$\mathbf{Z} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}.$$

A componente sistemática é assim dada por

$$\boldsymbol{\eta} = \mathbf{z}^T \boldsymbol{\beta},$$

onde  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$  é o vetor de parâmetros das  $n$  combinações lineares das variáveis predictoras definidas pelas  $n$  observações.

- **função de ligação** que tem como objetivo ligar a componente aleatória e a componente sistemática, através da relação  $g(\mu_i) = \mathbf{z}_i^T \boldsymbol{\beta}$ , onde  $g(\mu_i)$  é a função de ligação, monótona e diferenciável, que pode ser outra que não a função de ligação canónica associada à distribuição Normal.

A relação é estabelecida então da seguinte forma,

$$\begin{aligned} g(\mu) &= g(E[Y]) = \mathbf{Z}\boldsymbol{\beta} \Leftrightarrow g(\mu_i) = g(E[Y_i]) = \mathbf{z}_i^T \boldsymbol{\beta} = \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{j(i)}, \forall i \in \{1, \dots, n\}. \end{aligned} \quad (2.8.1)$$

## 2.8.2 Família Exponencial

Tal como foi referido anteriormente, as variáveis resposta de um MLG têm como pressuposto o facto de a função de distribuição pertencer à família exponencial. Considere-se a seguinte definição para o caso uniparamétrico.

**Definição 1.** *Seja  $Y$  uma variável aleatória. Diz-se que  $Y$  tem distribuição pertencente à família exponencial se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) se puder escrever na forma*

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (2.8.2)$$

onde  $\theta$  é a forma canónica do parâmetro de localização,  $\phi$  é o parâmetro de dispersão e  $a(\cdot), b(\cdot)$  e  $c(\cdot, \cdot)$  são funções reais conhecidas.

### • Valor médio e variância

Seja  $\ell(\theta; \phi, y) = \ln(f(y|\theta, \phi)) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$  e a função *score*

$$S(\theta) = \frac{\partial \ell(\theta; \phi, Y)}{\partial \theta}.$$

No caso de famílias regulares tem-se

$$E(S(\theta)) = 0;$$

$$E(S^2(\theta)) = E\left[\left(\frac{\partial \ell(\theta; \phi, Y)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ell(\theta; \phi, Y)}{\partial \theta^2}\right] \quad (2.8.3)$$

e portanto como, no caso em que  $f(y|\theta, \phi)$  é dado por (2.8.2),

$$\ell(\theta; \phi, y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi),$$



obtém-se,

$$S(\theta) = \frac{Y - b'(\theta)}{a(\phi)} \quad e \quad \frac{\partial S(\theta)}{\partial \theta} = -\frac{b''(\theta)}{a(\phi)}, \quad (2.8.4)$$

onde  $b'(\theta) = \frac{\partial b(\theta)}{\partial \theta}$  e  $b''(\theta) = \frac{\partial^2 b(\theta)}{\partial \theta^2}$ .

Assim de (2.8.3) e (2.8.4), resulta que

$$E(Y) = \mu = a(\phi)E(S(\theta)) + b'(\theta) = b'(\theta) \quad e,$$

$$var(Y) = a^2(\phi)var(S(\theta)) = a^2(\phi)\frac{b''(\theta)}{a(\phi)} = a(\phi)b''(\theta).$$

Temos então que a variância de  $Y$  depende da função  $b''(\theta)$ , isto é, depende do valor médio. A esta função dá-se o nome de *função de variância* que se costuma representar por  $V(\mu)$ .

Em muitas situações de interesse, observa-se que a função  $a(\phi)$  toma a forma

$$a(\phi) = \frac{\phi}{\omega},$$

onde  $\omega$  é uma constante conhecida, obtendo-se portanto a variância de  $Y$  como produto do parâmetro de dispersão por uma função apenas do valor médio.

Neste caso a função definida em (2.8.2) escreve-se na forma

$$f(y|\theta, \phi, \omega) = \exp\left\{\frac{\omega}{\phi}(y\theta - b(\theta)) + c(y, \phi, \omega)\right\}. \quad (2.8.5)$$

Turkman e Silva (2000) apresentam exemplos de algumas distribuições da família exponencial - que não faremos a prova por se afastar do âmbito deste trabalho - a distribuição normal, binomial, Poisson, gama e gaussiana inversa (Tabela 2.1).

Assim, os MLG são obtidos através da extensão das hipóteses subjacentes aos modelos lineares no que diz respeito à distribuição de  $Y_i$  poder ser qualquer uma que pertença à família exponencial; permite a existência de outras formas de ligação entre o preditor linear e o valor médio.

Referimos atrás que uma das distribuições de probabilidade que pertence à classe de distribuições da família exponencial é a distribuição binomial. Ao longo das próximas secções iremos dar especial importância às variáveis resposta que seguem distribuição binomial e um conjunto de variáveis explicativas que podem ser de qualquer natureza, por ser um dos temas a abordar neste trabalho.

A metodologia que assenta no que atrás foi referido é a regressão logística que abordamos na secção seguinte.

Tabela 2.1: Distribuições (algumas) pertencentes à família exponencial

Distribuição	normal	binomial	Poisson		gama	gaussiana inversa
Notação	$N(\mu, \sigma^2)$	$B(m, p)/m$	$P(\lambda)$		$Ga(v, \frac{v}{\mu})$	$IG(\mu, \sigma^2)$
Suporte	$(-\infty, +\infty)$	$\{0, \frac{1}{m}, \dots, 1\}$	$\{0, 1, \dots\}$		$(0, +\infty)$	$(0, +\infty)$
$\theta$	$\mu$	$\ln(\frac{p}{1-p})$	$\ln \lambda$		$-\frac{1}{\mu}$	$-\frac{1}{2\mu^2}$
$a(\phi)$	$\sigma^2$	$\frac{1}{m}$	1		$\frac{1}{v}$	$\sigma^2$
$\phi$	$\sigma^2$	1	1		$\frac{1}{v}$	$\sigma^2$
$\omega$	1	$m$	1		1	1
$c(y; \phi)$	$-\frac{1}{2}(\frac{y^2}{\phi} + \ln(2\pi\phi))$	$\ln \binom{n}{k}$	$-\ln y!$	$v \ln v - \ln \Gamma(v) + (v-1) \ln y$	$-\frac{1}{2}\{\ln(2\pi\phi y^3) + \frac{1}{y\phi}\}$	
$b(\theta)$	$\frac{\theta^2}{2}$	$\ln(1 + e^\theta)$	$e^\theta$		$-\ln(-\theta)$	$-(-2\theta)^{-1/2}$
$b'(\theta) = E(Y)$	$\theta$	$p = \frac{e^\theta}{1+e^\theta}$	$\lambda = e^\theta$		$\mu = -\frac{1}{\theta}$	$\mu = (-2\theta)^{-1/2}$
$b''(\theta) = V(\mu)$	1	$p(1-p)$	$\lambda$		$\mu^2$	$\mu^3$
$var(Y)$	$\sigma^2$	$\frac{p(1-p)}{m}$	$\lambda$		$\frac{\mu^2}{v}$	$\mu^3\sigma^2$

27

Reprodução da tabela apresentada em Turkman e Silva (2000).

## 2.9 Regressão Logística

### 2.9.1 Conceitos básicos

A regressão logística é talvez dos métodos de regressão mais usado em investigação clínica, muito pela fácil interpretação dos parâmetros do modelo, em comparação com, por exemplo, a regressão linear múltipla.

Consideremos  $Y_i$  uma variável resposta codificada em 0 ou 1, em que ( $Y = 0$ ) representa a ausência e ( $Y = 1$ ) a presença de determinada característica que se pretende estudar. Como  $Y_i$  apenas toma dois valores, 0 ou 1, pode definir-se

$$Y_i = \begin{cases} 0 & , 1-p_i \\ 1 & , p_i \end{cases}$$

em que  $p_i$  designa a probabilidade de "sucesso" e  $1 - p_i$  a probabilidade de "insucesso", donde se conclui que  $Y_i \sim \text{Bernoulli}(p_i)$ .

A função massa de probabilidade é dada por

$$f(y_i|p_i) = (p_i)^{y_i}(1 - p_i)^{1-y_i}, y_i = 0, 1, \quad i = 1, \dots, n. \quad (2.9.1)$$

Das propriedades da distribuição de Bernoulli, facilmente se verifica que

$$E[Y_i] = \mu = p_i \quad e \quad \text{var}[Y_i] = \sigma^2 = p_i(1 - p_i).$$

Tal como foi visto na secção anterior, para que possamos aplicar o modelo logístico, teremos que ver que a *f.m.p.* descrita em (2.9.1) pode ser escrita na forma apresentada em (2.8.2).

Temos então que,

$$\begin{aligned}
 f(y_i|p_i) &= \exp\{\ln(p_i^{y_i}(1-p_i)^{1-y_i})\} & (2.9.2) \\
 &= \exp\{\ln(p_i^{y_i}) + \ln(1-p_i)^{1-y_i}\} \\
 &= \exp\{y_i \ln(p_i) + (1-y_i) \ln(1-p_i)\} \\
 &= \exp\{y_i \ln(p_i) + \ln(1-p_i) - y_i \ln(1-p_i)\} \\
 &= \exp\{y_i[\ln(p_i) - \ln(1-p_i)] + \ln(1-p_i)\} \\
 &= \exp\left\{y_i \ln\left(\frac{p_i}{1-p_i}\right) + \ln(1-p_i)\right\} \\
 &= \exp\left\{y_i \ln\left(\frac{p_i}{1-p_i}\right) - (-\ln(1-p_i))\right\}
 \end{aligned}$$

donde resulta que,

- $\theta = \ln\left(\frac{p_i}{1-p_i}\right)$
- $a(\phi) = 1$
- $\phi = 1$
- $b(\theta) = -\ln(1-p_i)$
- $c(y, \phi) = 0$

e portanto podemos concluir que pertence à família exponencial.

Vimos na secção anterior como estabelecer a ligação entre a componente aleatória e a componente sistemática através da função de ligação  $g(\mu)$  para o caso do modelo linear clássico.

Provou-se que a forma canónica do parâmetro de localização para uma variável aleatória dicotómica é dado por

$$\theta = \ln\left(\frac{p_i}{1-p_i}\right),$$

pelo que será natural considerar como função de ligação canónica para variáveis resposta com distribuição de probabilidade Bernoulli( $p_i$ ),

$$g(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right),$$

que se designa por função *logit*.

A função de ligação *logit* não é mais do que o logaritmo do quociente entre a probabilidade de sucesso e a probabilidade de insucesso, que não é mais do que o *log-odds ratio*. Abordaremos o conceito de *log-odds ratio* mais à frente.

A utilização desta função em detrimento de outras que se podem considerar tem que ver com o facto de, do ponto de vista matemático, o logaritmo ser uma função flexível e de fácil utilização.

Definida a função de ligação, estamos então em condições de definir o modelo de regressão logística múltiplo

$$\ln \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \quad (2.9.3)$$

em que a probabilidade de sucesso é dada por

$$p_i = \frac{\exp\{\mathbf{z}_i^T \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{z}_i^T \boldsymbol{\beta}\}}. \quad (2.9.4)$$

Depois de definido o modelo de regressão logístico múltiplo, é necessário estimar os parâmetros do modelo.

## 2.9.2 Estimação

Nesta subsecção faremos uma abordagem ao método de estimação para os parâmetros do modelo,  $\beta_0, \dots, \beta_k$ .

O método de estimação que abordaremos é o da máxima verosimilhança, método de inferência utilizado nos MLG, em que se pretende encontrar os valores de  $(\beta_0, \dots, \beta_k)$  que maximizam a expressão da log-verosimilhança, quando o logaritmo do *odds* é linear com  $X$ .

A função de verosimilhança é dada por

$$L(\boldsymbol{\beta}|x, y, n) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}. \quad (2.9.5)$$

Aplicando o logaritmo à expressão anterior, obtém-se,

$$\ln(L(\boldsymbol{\beta}|x, y, n)) = \sum_{i=1}^n y_i \ln p_i + (1 - y_i) \ln(1 - p_i). \quad (2.9.6)$$

Substituindo a expressão de  $p_i$  obtida em (2.9.4) na expressão (2.9.6), obtém-se

$$\ln(L(\boldsymbol{\beta}|x, y, n)) = \sum_{i=1}^n \left( y_i \left( \beta_0 + \sum_{k=1}^n \beta_k x_{ik} \right) + \ln \left( 1 + \exp \left\{ \beta_0 + \sum_{k=1}^n \beta_k x_{ik} \right\} \right) \right). \quad (2.9.7)$$

De seguida, teríamos que obter as derivadas parciais da expressão (2.9.7) em ordem a cada coeficiente  $\beta$  do modelo, obtendo-se  $S(\beta)$ , ou seja, a função *score*. As estimativas de máxima verosimilhança,  $\hat{\beta}$ , seriam obtidas igualando o sistema de equações a zero. No caso de estarmos na presença de um número elevado de covariáveis, obter as estimativas de máxima verosimilhança de forma analítica pode tornar-se num trabalho moroso, de modo que existem métodos computacionais que permitem o cálculo das estimativas de máxima verosimilhança como o método de *Newton-Raphson*. Obtidas as estimativas, estamos em condições de calcular as estimativas das probabilidades ajustadas,  $\hat{y}_i$ .

Tal como vimos no início desta secção, os métodos de inferência dos MLG

baseiam-se na máxima verosimilhança. Deste modo, para averiguar se as covariáveis incluídas são estatisticamente significativas, são utilizadas três estatísticas, nomeadamente a estatística de Wald, a estatística de razão de verosimilhanças e a estatística *score*. Durante este trabalho apenas abordaremos a estatística de razão de verosimilhanças.

### 2.9.3 Teste de razão de verosimilhanças

Antes de se introduzir o teste de razão de verosimilhanças, é conveniente definir alguns conceitos que serão utilizados ao longo desta subsecção.

Considera-se,

- **modelo completo ou saturado** ( $M_s$ ), o modelo com todas as covariáveis em estudo. Trata-se de um modelo que se ajusta perfeitamente aos dados, mas cuja extração de informação é pouco interessante uma vez que não se trata de um modelo simplificado.
- **modelo nulo** ( $M_0$ ), como sendo o modelo que contém apenas uma covariável, ou seja, em que todas as variáveis  $Y_i$  têm o mesmo valor médio,  $\mu$ . Trata-se de um modelo pouco interessante, uma vez que são raros os casos em que se explica uma variável dita de interesse através de uma única covariável.
- **modelo maximal**, é modelo com o maior número de parâmetros, sendo a sua interpretação complexa.
- **modelo minimal**, que corresponde ao modelo com o menor número de parâmetros.
- **modelo corrente** ( $M_1$ ) é o modelo que se encontra em avaliação e que se pode situar entre o modelo maximal e o modelo minimal.

A regressão logística tem a particularidade de permitir a comparação do modelo corrente ( $M_1$ ) com o modelo nulo ( $M_0$ ) ou com o modelo saturado



( $M_s$ ). Estas comparações são feitas através do teste de razão de verosimilhanças.

Considerando que estamos na presença de dois modelos com  $v$  e  $t$  parâmetros ( $v < t$ ), podemos testar a nulidade de  $(t - v)$  parâmetros.

Assim,

$$H_0 : \beta_j = 0 \quad vs. \quad H_1 : \beta_j \neq 0, \text{ para algum } j, j = 1, \dots, (t - v)$$

em que a estatística de teste é dada por

$$\lambda = -2 \ln \left( \frac{L_v}{L_t} \right) \sim \chi_{t-v}^2, \text{ se } H_0 \text{ verdadeira} \quad (2.9.8)$$

onde  $L_v$  e  $L_t$  designam as funções de verosimilhança dos modelos com  $v$  e  $t$  parâmetros, respetivamente.

A hipótese nula é rejeitada a um nível de significância  $\alpha$ , se o valor observado da estatística de teste for superior ao quantil de probabilidade  $(1 - \alpha)$  de um  $\chi_{(t-v)}^2$ .

## 2.9.4 Interpretação dos coeficientes do modelo

Depois de obtidas as estimativas dos parâmetros do modelo, torna-se necessário introduzir algumas noções de como interpretar tais estimativas. Para tal serão introduzidos os conceitos de *odds* e *odds ratio*.

**Definição 2.** *Seja  $p_i$  a probabilidade de sucesso e  $1 - p_i$  a probabilidade de insucesso. Define-se odds de um determinado acontecimento como o quociente entre a probabilidade de sucesso e a probabilidade de insucesso, isto é, a chance de se observar o sucesso ( $Y = 1$ ) relativamente ao insucesso ( $Y = 0$ ). Ou seja,*

$$\frac{p_i}{1 - p_i}.$$

**Exemplo 1.** Suponhamos que foi levado a cabo um estudo que tem como objetivo perceber a influência da diabetes *mellitus* tipo 2 nos problemas de visão de indivíduos com mais de 45 anos.

Seja  $Y$  a variável que representa a presença/ausência de problemas de visão em indivíduos com mais de 45 anos e  $X$  a variável que representa os níveis de diabetes *mellitus* tipo 2.

Estamos portanto no caso mais simples, que é o de considerar apenas uma covariável. Este exemplo é facilmente extensível para um conjunto de  $k$  covariáveis, contudo, ficaremos pelo caso mais simples.

A variável  $Y$  foi codificada em 0 ou 1, em que  $Y = 0$  representa a ausência (*insucesso*) de problemas de visão e  $Y = 1$  a presença (*sucesso*) de problemas de visão. A variável  $X$  foi codificada também em 0 ou 1, em que ( $X = 0$ ) indica níveis baixos de diabetes e ( $X = 1$ ) indica níveis elevados de diabetes.

Seja  $p_0$  a probabilidade de um indivíduo ter problemas de visão nos indivíduos que têm níveis baixos de diabetes e  $p_1$  a probabilidade de ter problemas de visão nos indivíduos que têm níveis elevados de diabetes.

Ao quociente

$$\frac{p_1/(1-p_1)}{p_0/(1-p_0)} \quad (2.9.9)$$

dá-se o nome de *odds ratio* ( $OR$ ) (ou razão de chances).

Tendo em conta a expressão (2.9.4), temos então as seguintes expressões para os coeficientes de um modelo de regressão logística:

$Y \backslash X$	$x = 0$	$x = 1$
$y = 0$	$1 - p_0 = \frac{1}{1+e^{\beta_0}}$	$1 - p_1 = \frac{1}{1+e^{\beta_0+\beta_1}}$
$y = 1$	$p_0 = \frac{e^{\beta_0}}{1+e^{\beta_0}}$	$p_1 = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$

O *odds ratio* vem então dado por:

$$\begin{aligned}
OR &= \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)/\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)/\left(\frac{1}{1+e^{\beta_0}}\right)} = \\
&= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{(\beta_0+\beta_1)-\beta_0} = \\
&= e^{\beta_1}.
\end{aligned}$$

• **Propriedades:**

- (a) Quando os *odds* nos dois grupos são iguais, o *OR* é 1. Tal acontece quando as probabilidades de sucesso dadas pelos *odds* nos dois grupos são iguais;
- (b) Valores de *OR* menores que 1 sugerem que o *odds* no primeiro grupo é menor do que no segundo. Alternativamente, valores de *OR* maiores do que 1 sugerem que o *odds* é superior no segundo grupo do que no primeiro;
- (c) Aplicando o logaritmo à expressão (2.9.9), ficamos com

$$\ln OR = \ln \left( \frac{p_1}{1-p_1} \right) - \ln \left( \frac{p_0}{1-p_0} \right),$$

que não é mais do que a diferença dos *logit* nos dois grupos.

Retomemos o exemplo 1 de modo a clarificar os conceitos abordados anteriormente. Considerando os grupos: níveis baixos de diabetes ( $X = 0$ ) e níveis elevados de diabetes ( $X = 1$ ) e aplicando o logaritmo do *odds* de cada um dos grupos, obtém-se:

- $\ln \frac{P(Y=1|X)}{P(Y=0|X)} = \beta_0 + 0$  se  $X = 0$ ,
- $\ln \frac{P(Y=1|X)}{P(Y=0|X)} = \beta_0 + \beta_1$  se  $X = 1$ .

Assim,

- Se  $\beta_1 > 0$ , o *sucesso* tem mais chances no grupo com níveis elevados de diabetes;
- Se  $\beta_1 < 0$ , o *sucesso* tem mais chances no grupo com níveis baixos de diabetes.

### 2.9.5 Seleção de modelos

Uma vez definido o conjunto de covariáveis a ser incluído num modelo, resta saber qual a melhor maneira de encontrarmos um modelo simplificado que inclua apenas as covariáveis mais importantes para explicar a probabilidade de sucesso,  $p_i$ .

Na literatura são propostos três métodos de seleção de modelos como sejam o método *stepwise*, *backward* e *forward*. Para este trabalho será utilizado o método *stepwise* que é uma combinação do método *backward* e *forward*.

Este método consiste em incluir todas as variáveis no modelo inicial e vão-se removendo (*backward*) e adicionando (*forward*) variáveis sucessivamente ou ao contrário.

A significância das covariáveis é avaliada pelo teste de razão de verosimilhanças. Em geral é usado o critério AIC (*Akaike's Information Criterion*) como critério de seleção. Assim, após a aplicação deste método, o modelo que apresentar um menor AIC será selecionado. O critério AIC é definido da seguinte forma:

$$AIC_q = -2 \ln L(\hat{\beta}) + 2q,$$

em que  $\ln L(\hat{\beta})$  representa a função log-verosimilhança para um modelo com  $q$  variáveis explicativas (logo  $q + 1$  parâmetros) e  $\hat{\beta}$  o estimador de máxima verosimilhança de  $\beta$ .

## 2.9.6 Confundimento e interação

- **Confundimento**

Um dos problemas que surge frequentemente quando se tratam vários fatores de risco em conjunto é a sua não independência. Pode haver fatores de risco que estejam ligados de tal forma que seja muito difícil impedir que os seus efeitos não se confundam entre si potenciando-se ou anulando-se uns aos outros. Este problema é conhecido pelo nome de confusão de efeitos ou simplesmente confundimento (*confounding*).

Qualquer variável exterior que seja também um fator de risco para determinada doença em estudo e, além disso, esteja relacionada com a variável de exposição chama-se um fator de confundimento ou variável de confundimento desde que altere substancialmente a força da relação entre a doença e a variável de exposição.

Tais variáveis podem ser parcial ou completamente responsáveis por qualquer aparente associação entre a variável de exposição e a doença enviesando as estimativas da verdadeira ligação subjacente que possa existir entre elas.

- **Interação**

Qualquer variável exterior que, consoante os seus níveis, produz forças de relação entre determinada doença e a variável de exposição significativamente diferentes, é conhecida como um fator modificador de efeito ou variável de interação.

Mais especificamente: o confundimento existe quando a força da associação entre a variável de exposição e a doença é alterada depois de se descontar ou controlar o efeito de um segundo fator de risco; a interação existe quando a força da associação entre a variável de exposição e a doença é diferente para diferentes níveis do segundo fator de risco.

## 2.10 Qualidade de um modelo de regressão logística

Depois de selecionado o modelo, é necessário avaliar a qualidade do mesmo através da análise dos resíduos, análise de multicolinearidade, análise de observações influentes, estatística de Hosmer e Lemeshow e avaliação da capacidade preditiva do modelo.

### 2.10.1 Análise dos resíduos

Os resíduos representam a discrepância dos valores observados ( $y_i$ ) e os valores ajustados ( $\hat{y}_i$ ). Ou seja,

$$r_i = y_i - \hat{y}_i, i = 1, \dots, n.$$

Uma vez que a variável resposta é binária, os valores ajustados variam no intervalo  $[0, 1]$ , os resíduos do modelo variam no intervalo  $[-1, 1]$ . Quando  $r_i > 0$ ,  $y_i = 1$  e, para  $r_i < 0$ ,  $y_i = 0$ . Para  $r_i = 0$ , o ajustamento é perfeito, isto é,  $y_i = \hat{y}_i$ , os valores observados são iguais aos valores ajustados. Contudo, os resíduos tal como foram descritos apresentam a desvantagem de apresentar grande variabilidade.

Antes de avançarmos, é importante introduzir um conceito que nos irá acompanhar de seguida, o conceito de *deviance* ou função de desvio.

A *deviance* é uma medida baseada na estatística de razão de verosimilhanças e avalia a discrepância entre o modelo saturado ( $M_s$ ) e o modelo corrente ( $M_1$ ).

Seja  $L(\cdot)$  a função de verosimilhança; o teste de razão de verosimilhanças permite afirmar que, com  $\lambda_1 = \frac{L(M_1)}{L(M_s)}$ , sob a hipótese de  $M_1$  ser o melhor modelo,  $-2 \ln(\lambda_1) \sim \chi_{n-2}^2$ .

Temos assim que o desvio do modelo corrente ( $M_1$ ) ao modelo saturado ( $M_s$ ), ou seja, a então chamada *deviance*, vem dada por

$$G_1 = -2 \ln(\lambda_1),$$

ou seja,

$$G_1 = -2\{\ln L(M_1) - \ln L(M_s)\}.$$

Na análise de modelos de regressão logística os resíduos da *deviance* padronizados são bem utilizados para a análise de qualidade do modelo, sendo dados por (Hosmer e Lemeshow, 2000)

$$R_i^D = \frac{r_D}{\sqrt{\hat{\phi}(1 - h_{ii})}},$$

onde,

$$r_D = \text{sinal}(y_i - \hat{y}_i) \sqrt{d_i},$$

em que  $d_i$  representa a contribuição da  $i$ -ésima observação para a função de desvio e  $h_{ii}$  o  $i$ -ésimo elemento da diagonal principal da matriz de projeção generalizada (*hat matrix*) dada por

$$\mathbf{H} = \mathbf{V}^{\frac{1}{2}} \mathbf{Z} (\mathbf{Z}^T \mathbf{V} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{\frac{1}{2}}$$

em que  $\mathbf{V}$  é a matriz diagonal ( $n \times n$ ) com elemento geral dado por  $v_i = \hat{y}_i(1 - \hat{y}_i)$ . Representando graficamente os valores ajustados *vs* resíduos da *deviance* padronizados, espera-se que 95% dos resíduos estejam compreendidos entre -2 e 2 (Dupont W, 2002)

- Interpretação da *deviance*:

- **Deviance pequena:** Significa que a explicação do modelo ajustado (corrente) é praticamente igual à do modelo completo, pelo que se pode concluir que se trata de um bom ajustamento. Nestes casos, deve-se optar pelo modelo ajustado, uma vez que contém menos preditores, tornando-se assim num modelo parcimonioso.

- **Deviance elevada:** Indica uma fraca explicação do modelo ajustado. Deve evitar-se a sua utilização.

### 2.10.2 Multicolinearidade

Uma outra questão de enorme importância no diagnóstico de um modelo, tem a ver com a multicolinearidade. O facto de covariáveis serem colineares, leva a que exista correlação, logo que a interpretação de quais as variáveis a incluir no modelo possa ficar comprometida. Assim sendo, para a análise de multicolinearidade deve considerar-se o cálculo do VIF- *Variance Inflation Factor*, que indica a existência ou não de multicolinearidade, sendo dado por

$$VIF_k = \frac{1}{1 - R_k^2},$$

onde  $R_k^2$  é o coeficiente de determinação não ajustado da regressão de  $x_i, i = 1, \dots, k$  contra as restantes covariáveis do modelo. No caso de existirem covariáveis com  $VIF$  elevado ( $VIF > 10$ ), a solução passa por retirar uma dessas covariáveis e refazer a análise.

### 2.10.3 Observações influentes

Uma observação é considerada influente se a sua exclusão produz alterações significativas nas estimativas dos parâmetros do modelo. A sua presença pode assim originar conclusões que não correspondem de todo à realidade. Assim, apresentam-se de seguida duas medidas de diagnóstico: a medida de *leverage* e a distância de *Cook*.

- **Medida de *leverage***

Os valores de *leverage* ou valores de alavanca são os valores correspondentes aos valores da matriz *hat*.

Consiste em representar  $h_{ii}$  (distância entre a  $i$ -ésima observação em relação às restantes observações) *versus* o seu número de ordem. Uma observação é considerada influente se  $h_{ii} > 2$ .



- **Distância de *Cook***

A distância de *Cook* é uma medida de influência baseada no valor estimado dos coeficientes da regressão. Para o seu cálculo deve considerar-se a amostra sem a observação que se pensa ser influente e a amostra com esta observação. Se o valor estimado se alterar, considera-se que a observação pode ser influente.

O cálculo da distância de *Cook* é dado por:

$$\Delta\hat{\beta}_i = (\hat{\beta} - \hat{\beta}_{(-i)})^T (\mathbf{X}^T \mathbf{V} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(-i)}), \quad (2.10.1)$$

em que  $\hat{\beta}$  representa as estimativas de máxima verosimilhança dos parâmetros do modelo e  $\hat{\beta}_{(-i)}$  as estimativas de máxima verosimilhança dos parâmetros do modelo retirando a observação que se considera influente,  $i$ . Consideram-se assim influentes as observações em que  $\Delta\hat{\beta}_i > 4/n$ , em que  $n$  representa o número total de observações.

#### 2.10.4 Estatística de Hosmer e Lemeshow

Hosmer e Lemeshow (2000) propuseram uma estatística que permite avaliar o ajustamento do modelo aos dados, baseada na comparação do número observado com o número esperado de sucessos de  $g$  grupos formados, geralmente 10, em função do número de células distintas resultantes do cruzamento das classes das variáveis independentes.

Neste caso as hipóteses em teste são:

$$\begin{aligned} H_0 &: \text{o modelo ajusta-se aos dados} \\ &vs. \\ H_1 &: \text{o modelo não se ajusta aos dados} \end{aligned}$$

A estatística de teste é dada por

$$X_{HL}^2 = \sum_{i=1}^g \frac{(z_i - \hat{y}_i)^2}{\hat{y}_i},$$

onde,  $z_i = \sum_{j=1}^{n_i} y_{ij}$ ,  $n_i$  é a dimensão do grupo  $i$  e  $\hat{y}_i = \sum_{j=1}^{n_i} p_{ij}$ . Para amostras de dimensão suficientemente grande, a distribuição de amostragem da estatística de teste é de um  $\chi^2$  com  $g - 2$  graus de liberdade.

### 2.10.5 Capacidade preditiva do modelo

Após ter sido feita a análise de adequabilidade do modelo, é necessário avaliar a capacidade preditiva do mesmo.

Muitas vezes estamos interessados em realizar análises que consistem na medição ou quantificação de parâmetros bioquímicos ou em contagens de células de determinado tipo. Nas análises clínicas, por exemplo, este é um procedimento bastante comum, sendo o diagnóstico do doente determinado através da comparação dos valores obtidos - através das ditas medições ou contagens - com um valor de referência.

Perante determinada doença ( $D$ ), se o parâmetro bioquímico tender a apresentar-se superior ao valor de referência, o indivíduo é classificado como doente, sendo diagnosticado como não doente, caso contrário. Analogamente, se o parâmetro bioquímico tender a apresentar-se inferior ao valor de referência, então o indivíduo é classificado como doente. Designaremos por teste de diagnóstico o duo (análise clínica, regra de classificação).

Tanto o estado de saúde como o teste de diagnóstico induzem partições no espaço de resultados associado à situação de um qualquer indivíduo da população. No que se refere à doença, o espaço de resultados reparte-se entre *doentes* e *não doentes*. Relativamente ao teste de diagnóstico, reparte-se entre *positivos* e *negativos*.

Qualquer indivíduo que se submeta ao teste de diagnóstico pertence, portanto, a uma de quatro categorias segundo duas partições do espaço de resultados, como se apresenta na tabela seguinte:

Resultado da análise/Estado	Doente	Não doente
Positivo	VP	FP
Negativo	FN	VN

As categorias representadas são:

- VP - verdadeiro positivo - corresponde a um indivíduo que **é doente** e tem **resultado positivo** na análise ( $D \cap +$ );
- FP - falso positivo - corresponde a um indivíduo que **não é doente** e tem **resultado positivo** na análise ( $\bar{D} \cap +$ );
- VN - verdadeiro negativo - corresponde a um indivíduo que **não é doente** e tem **resultado negativo** na análise ( $\bar{D} \cap -$ );
- FN - falso negativo - corresponde a um indivíduo que **é doente** e tem **resultado negativo** na análise ( $D \cap -$ ).

A qualidade do desempenho de um teste de diagnóstico é medida pela sua capacidade de detetar a doença quando esta está presente, mas também pela capacidade que tenha de não reter um indivíduo não doente. Estas características de um teste de diagnóstico são medidas em termos de probabilidade e designam-se por sensibilidade e especificidade, respetivamente.

Define-se **sensibilidade** como a proporção de verdadeiros positivos entre os doentes,

$$sensibilidade = \frac{VP}{VP + FN} = \frac{VP}{D} = P(+|D),$$

e **especificidade** como a proporção de verdadeiros negativos entre os não doentes,

$$especificidade = \frac{VN}{FP + VN} = \frac{VN}{\bar{D}} = P(-|\bar{D}).$$

Idealmente, tanto a sensibilidade como a especificidade deveriam ser iguais a 1. Tal implicaria a inexistência de falsos positivos e de falsos negativos.

Segundo Maroco (2010), pode-se classificar a capacidade preditiva de acordo com os valores de sensibilidade e especificidade apresentados na tabela seguinte:

Porcentagem	Capacidade preditiva
> 80	Boas capacidades preditivas
[50; 80]	Capacidades preditivas razoáveis
< 50	Capacidades preditivas medíocres

De referir ainda duas medidas de desempenho dos testes de diagnóstico, o valor preditivo positivo ( $VPP$ ) e o valor preditivo negativo ( $VPN$ ). Correspondem, respetivamente, à proporção de diagnósticos corretos entre os que testaram positivo e à proporção de diagnósticos corretos entre os que testaram negativo:

- $VPP = \frac{VP}{VP+FP} = P(D|+)$ ;

- $VPN = \frac{VN}{VN+FN} = P(\bar{D}|-)$ .

Vimos anteriormente que muitas vezes o nosso interesse reside na comparação entre parâmetros bioquímicos recolhidos e um valor de referência. É sobre a determinação deste valor de referência que falaremos a seguir, introduzindo desde já a ideia subjacente a este procedimento, a construção de Curvas ROC.

### 2.10.6 Curva ROC

A Curva ROC é uma ferramenta poderosa para medir e especificar problemas no desempenho de testes de diagnóstico.

Esta análise é feita por meio de um método gráfico simples e robusto e permite estudar a variação da sensibilidade e da especificidade para diferentes valores de corte. Assim, define-se a curva ROC através da representação gráfica da sensibilidade, no eixo das ordenadas e, no eixo das abcissas, 1-especificidade.

Os valores de corte referidos anteriormente, são aqueles que maximizam o desempenho do modelo, denominado usualmente por *cut-off* ótimo. Este *cut-off* traduz o compromisso existente entre o número de verdadeiros positivos e o número de falsos positivos.

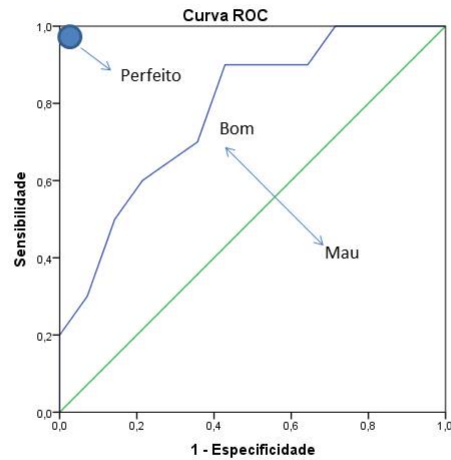


Figura 2.1: Curva ROC

A qualidade de um teste de diagnóstico é avaliada através da área abaixo da curva (*AUC*), que varia entre 0 e 1. Hosmer e Lemeshow (2000) apresentam valores indicativos da área abaixo da curva que podem ser utilizados para avaliar a qualidade da mesma, ou seja, o seu poder discriminativo.

<b>Área abaixo da curva</b>	<b>Poder discriminativo</b>
$[0,5; 0,7[$	Discriminação fraca
$[0,7; 0,8[$	Discriminação razoável
$[0,8; 0,9[$	Discriminação boa
$\geq 0,9$	Discriminação ótima

# Capítulo 3

## Análise exploratória dos dados

A amostra recolhida para este estudo é constituída por 134 mulheres. Para efeitos de notação, neste e nos restantes capítulos, sempre que necessário, considera-se  $\bar{x}$  a média amostral,  $m$  a mediana amostral,  $P_{25}$  o percentil 25 e  $P_{75}$  o percentil 75. Sem perigo de confusão, ao longo do texto,  $\bar{x}$  será apenas mencionado como média e  $m$  como mediana.

### 3.1 Caracterização da amostra

- **Caracterização antropométrica e sociodemográfica**

A tabela 3.1 diz respeito aos resultados relativos à variável idade das mulheres em estudo. De acordo com os resultados obtidos, a média registada foi de 33,64 anos, com um desvio padrão de 3,95 anos. A idade mediana foi de 34 anos com uma idade mínima de 20 anos e máxima de 44 anos.

Tabela 3.1: Idade (anos)

Média	Mediana	Desvio padrão	Mínimo	Máximo	$n$
33,64	34,00	3,95	20	44	134

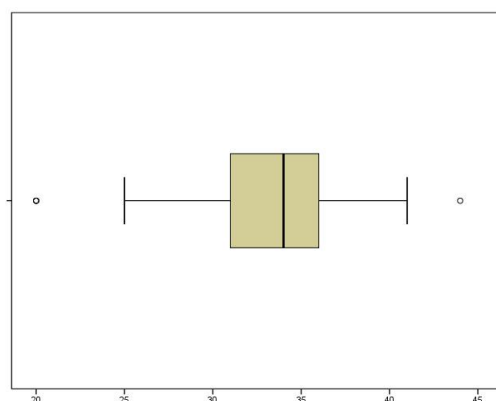


Figura 3.1: Diagrama em caixa de bigodes relativo à variável idade

Analisando o diagrama em caixa de bigodes e os resultados apresentados na tabela 3.1, verifica-se que não existe simetria da distribuição subjacente aos dados, uma vez que  $\bar{x} \neq m$  ( $\bar{x} = 33,64$  e  $m = 34,00$ ), ainda que sejam valores bastante próximos.

Observa-se a existência de dois candidatos a *outlier* (1 inferior e 1 superior), o que significa que duas mulheres registam idades que se afastam do padrão das restantes da amostra.

Verifica-se que 25% das mulheres registam idades menores ou iguais a 31 anos e que 75% das mulheres incluídas no estudo registam idades menores ou iguais a 36 anos ( $P_{25} = 31$  e  $P_{75} = 36$ ).

Como resultado das variáveis peso e altura, procedeu-se a uma análise do Índice de Massa Corporal (IMC), ou seja,

$$IMC = \frac{peso(kg)}{[altura(m)]^2}$$

Dado que a amostra recolhida é constituída por mulheres grávidas e não grávidas, os resultados subjacentes a esta variável serão apresentados mais à frente, na caracterização dos grupos de estudo, uma vez que qualquer observação desta variável para a amostra total estará mascarada pelas condições em que as mulheres se encontram.

Tabela 3.2: Classificação IMC

<b>Classificação adultos</b>	<b>IMC</b>
Baixo peso	$< 18$
Eutrofia	18,5 - 24,9
Pré-obesidade	25 - 29,9
Obesidade, grau I	30 - 34,9
Obesidade, grau II	35 - 39,9
Obesidade mórbida	$\geq 40$

Tabela 3.3: Nível de escolaridade

<b>Escolaridade</b>	<b><i>n</i></b>	<b>%</b>
Ensino básico	3	2,3
Ensino secundário	16	12,0
Ensino superior	114	85,7
<b>Total</b>	<b>133</b>	<b>100</b>

De acordo com dados obtidos, a grande maioria das mulheres em estudo tem o ensino superior (85,7%).

No que diz respeito à atividade profissional, apenas foram obtidos registos para 21 mulheres, constatando-se que a maioria trabalha por conta de outrém.

Relativamente a hábitos tabágicos, a grande maioria das mulheres referiu não ter tais hábitos (97,0%), bem como não serem fumadoras passivas (78,3%).

#### • Caracterização clínica

Foram recolhidos dados relativos à pressão arterial sistólica e diastólica. Os resultados encontram-se na tabela 3.4.

Analisando os resultados obtidos na tabela 3.4, observa-se que a pressão arterial sistólica média foi de 114,98 mmHg, com um desvio padrão de 12,92



Tabela 3.4: Pressões arteriais (mmHg)

	<b>Média</b>	<b>Mediana</b>	<b>Desvio padrão</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>n</b>
Sistólica	114,98	114,00	12,92	70	160	129
Diastólica	69,78	70,00	10,16	40	120	129

mmHg. A mediana foi de 114 mmHg, atingindo-se uma pressão arterial sistólica mínima de 70 mmHg e uma pressão arterial sistólica máxima de 160 mmHg.

Já para a pressão arterial diastólica, a média foi de 69,78 mmHg, com um desvio padrão de 10,16 mmHg. A mediana foi de 70 mmHg, atingindo-se uma pressão arterial diastólica mínima de 40 mmHg e uma pressão arterial diastólica máxima de 120 mmHg.

Segundo a classificação da Sociedade Portuguesa de Cardiologia, em média as mulheres em estudo apresentam níveis de pressão arterial normais.<sup>1</sup>

Tabela 3.5: Classificação da pressão arterial

<b>Classificação</b>	<b>Sistólica</b>		<b>Diastólica</b>
Ótima	< 120	e	<80
Normal	120-129	e/ou	80-84
Normal alta	130-139	e/ou	85-89
Hipertensão, grau I	140-159	e/ou	90-99
Hipertensão, grau II	160-179	e/ou	100-109
Hipertensão, grau III	≥ 180	e/ou	≥ 110
Hipertensão sistólica Isolada	≥ 140	e	<90

<sup>1</sup>[http://www.spc.pt/FS/AreaCientifica/recomendacoes/hipertensao\\_arterial\\_v2013.pdf](http://www.spc.pt/FS/AreaCientifica/recomendacoes/hipertensao_arterial_v2013.pdf)

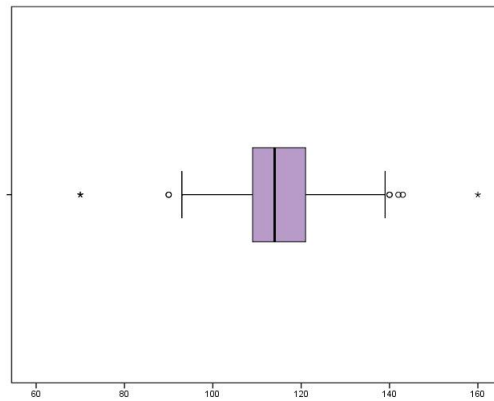


Figura 3.2: Diagrama em caixa de bigodes - P. sistólica

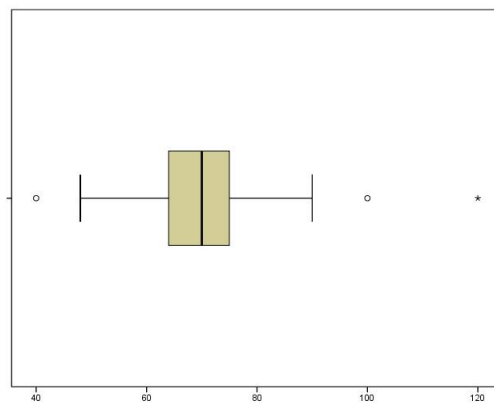


Figura 3.3: Diagrama em caixa de bigodes - P. diastólica

Através da análise do diagrama em caixa de bigodes apresentada para a pressão arterial sistólica (Figura 3.2) e com os resultados da tabela 3.4, verifica-se que não existe simetria da distribuição subjacente aos dados, uma vez que  $\bar{x} \neq m$  ( $\bar{x} = 114,98$  e  $m = 114,00$ ), ainda que sejam valores próximos.

Observa-se a existência de seis candidatos a *outlier* (2 inferiores e 4 superiores).

Verifica-se que 25% das mulheres registam valores de pressão arterial sistólica menores ou iguais a 109 mmHg e que 75% das mulheres registam valores de pressão arterial sistólica menores ou iguais a 121,5 mmHg ( $P_{25} = 109$  e  $P_{75} = 121,5$ ).

Já para os valores de pressão arterial diastólica (Figura 3.3), não se verifica simetria da distribuição subjacente aos dados dado que  $\bar{x} \neq m$  ( $\bar{x} = 69,78$  e  $m = 70,00$ ).

A registrar três candidatos a *outlier* (1 inferior e 2 superiores).

Observa-se que 25% das mulheres registam valores de pressão arterial diastólica menores ou iguais a 64 mmHg e que 75% das mulheres registam valores de pressão arterial diastólica menores ou iguais a 75 mmHg ( $P_{25} = 64$  e  $P_{75} = 75$ ).

**Em resumo:**

- A idade média das mulheres incluídas no estudo é de 33,64 anos;
- De acordo com os resultados obtidos para as variáveis pressão arterial sistólica e diastólica, as mulheres revelaram pressões arteriais normais de acordo com a classificação em vigor;
- Verificou-se que a maioria das mulheres não revelava hábitos tabágicos nem exposição ao tabaco;
- No que diz respeito ao nível de escolaridade, observou-se que 85,7% das mulheres tem o ensino superior, sendo que 21/133 encontra-se empregada por conta de outrém.

## 3.2 Caracterização dos grupos de estudo

Tabela 3.6: Grupos de estudo

	<i>n</i>
GA	24
GNA	43
NGA	32
NGNA	35
<b>Total</b>	<b>134</b>

### • Caracterização antropométrica e sociodemográfica (mulheres saudáveis)

A tabela seguinte diz respeito às variáveis idade e IMC.

Tabela 3.7: Variáveis antropométricas (mulheres saudáveis)

Idade (anos)	Média	Mediana	Desvio padrão	Mínimo	Máximo	<i>n</i>
GNA	32,74	32,00	3,38	25	41	43
NGNA	34,17	35,00	3,82	20	40	35
IMC (kg/m <sup>2</sup> )						
GNA	25,75	25,64	2,98	20,31	33,79	43
NGNA	21,56	21,09	2,76	16,94	31,77	35

De modo a analisar diferenças entre os dois grupos relativamente às variáveis analisadas, foi aplicado o teste à normalidade de Kolmogorov-Smirnov, concluindo-se que o modelo normal não se ajusta aos dados no grupo NGNA para a variável idade, pelo que se rejeitou a hipótese de normalidade (Idade:  $p = 0,2$  (GNA) e  $p = 0,028$  (NGNA)). Estamos portanto em condições de aplicar métodos de inferência estatística não paramétrica para a análise de diferenças entre os dois grupos em estudo relativamente à variável idade. Uma vez que se trata de amostras independentes e estando na presença de dois grupos, aplicou-se o teste de Mann-Whitney Wilcoxon para analisar diferenças entre GNA e NGNA, concluindo-se que existe evidência para afirmar

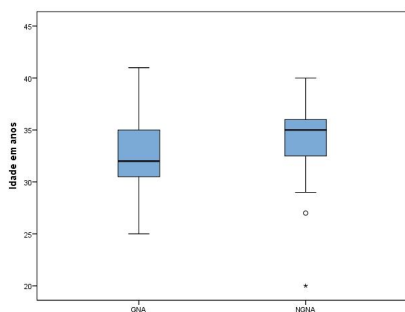


Figura 3.4: Diagrama em caixa de bigodes relativo à idade (mulheres saudáveis)

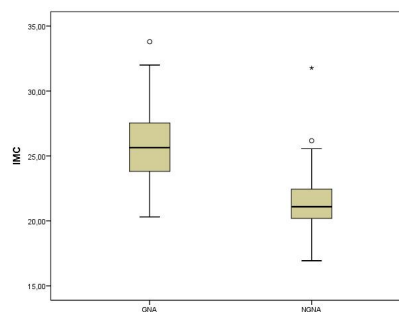


Figura 3.5: Diagrama em caixa de bigodes relativo ao IMC (mulheres saudáveis)

que existem diferenças estatisticamente significativas entre GNA e NGNA relativamente à variável idade ( $p = 0,024$ ). Relativamente à variável IMC, uma vez que não se rejeitou a hipótese de normalidade subjacente aos dados (IMC:  $p = 0,2$  (GNA) e  $p = 0,055$  (NGNA)), foi aplicado o teste t para amostras independentes, concluindo-se que existe evidência para afirmar que existem diferenças estatisticamente significativas entre GNA e NGNA relativamente ao IMC ( $p < 0,001$ ), como seria de esperar.

Tabela 3.8: Nível de escolaridade (mulheres saudáveis)

Escolaridade	GNA $n(\%)$	NGNA $n(\%)$
Ensino básico	1 (2,3%)	2 (5,7%)
Ensino secundário	3 (7,0%)	6 (17,1%)
Ensino superior	39 (90,7%)	27 (77,1%)
<b>Total</b>	<b>43 (100%)</b>	<b>35 (100%)</b>

De acordo com dados obtidos, a grande maioria das mulheres em estudo tinha o ensino superior.

Não se registaram resultados para a atividade profissional. Relativamente a hábitos tabágicos, não se registaram casos de tais hábitos, bem como a maioria não é fumador passivo.

• Caracterização clínica

Tabela 3.9: Pressões arteriais (mulheres saudáveis)

<b>P. Sistólica (mmHg)</b>	<b>Média</b>	<b>Mediana</b>	<b>Desvio padrão</b>	<b>Mínimo</b>	<b>Máximo</b>	<b><i>n</i></b>
GNA	115,51	114,00	10,04	93	138	43
NGNA	119,77	118,00	10,49	102	143	35
<b>P. Diastólica (mmHg)</b>						
GNA	66,60	66,00	8,18	48	85	43
NGNA	74,66	75,00	7,39	56	90	35

Por aplicação do teste à normalidade de Kolmogorov-Smirnov, concluiu-se que o modelo normal não se ajusta aos dados para a variável P. Sistólica, uma vez que se rejeitou a hipótese de normalidade no grupo GNA (P. Sistólica:  $p = 0,011$  (GNA) e  $p = 0,2$  (NGNA)). Assim, foi aplicado o teste de Mann-Whitney Wilcoxon para analisar a existência de diferenças entre os dois grupos de estudo relativamente à variável P. Sistólica, concluindo-se que não existe evidência para afirmar que existam diferenças estatisticamente significativas entre GNA e NGNA ( $p = 0,083$ ). Já para a pressão arterial diastólica, uma vez que não se rejeitou a hipótese de normalidade subjacente aos dados (P.Diastólica:  $p = 0,2$  (GNA) e  $p = 0,2$  (NGNA)), foi aplicado o teste t para amostras independentes, concluindo-se que existe evidência para afirmar que a pressão arterial diastólica é diferente nos dois grupos de estudo ( $p < 0,001$ ).

Tabela 3.10: História familiar de atopia (mulheres saudáveis)

	<b>GNA <i>n</i>(%)</b>
Não	20 (46,5%)
Pai	13 (30,2%)
Avós maternos	9 (20,9%)
Avós paternos	1 (2,3%)
<b>Total</b>	<b>43 (100%)</b>

No que diz respeito ao número de gravidezes anteriores, os resultados encontram-se na tabela seguinte:

Tabela 3.11: Número de gravidezes anteriores (mulheres saudáveis)

<b>Grav. anteriores</b>	<b>GNA <math>n(\%)</math></b>	<b>NGNA <math>n(\%)</math></b>
0	18 (41,9%)	4 (11,4%)
1	14 (32,6%)	12 (34,3%)
2	9 (20,9%)	15 (42,9%)
3	1 (2,3%)	3 (8,6%)
4	1 (2,3%)	0 (0,0%)
5	0 (0,0%)	1 (2,9%)
<b>Total</b>	<b>43 (100%)</b>	<b>35 (100%)</b>

Tabela 3.12: Número de filhos (mulheres saudáveis)

<b>N.º de filhos</b>	<b>GNA <math>n(\%)</math></b>	<b>NGNA <math>n(\%)</math></b>
0	24 (55,8%)	5 (14,3%)
1	18 (41,9%)	13 (37,1%)
2	1 (2,3%)	16 (45,7%)
3	0 (0,0%)	1 (2,9%)
<b>Total</b>	<b>43 (100%)</b>	<b>35 (100%)</b>

Tabela 3.13: Parto de termo (mulheres saudáveis)

	<b>GNA <math>n(\%)</math></b>
Sim	3 (7,0%)
Não	40 (93,0%)
<b>Total</b>	<b>43 (100%)</b>

Tabela 3.14: Trabalho de parto (mulheres saudáveis)

	<b>GNA</b> <i>n</i> (%)
Cesariana	14 (33,3%)
Espontâneo	20 (47,6%)
Induzido	8 (19,0%)
<b>Total</b>	<b>42 (100%)</b>

Tabela 3.15: Tipo de parto (mulheres saudáveis)

	<b>GNA</b> <i>n</i> (%)
Cesariana eletiva	14 (33,3%)
Cesariana intraparto	11 (26,2%)
Eutócico	7 (16,7%)
Vaginal com ventosa	10 (23,8%)
<b>Total</b>	<b>42 (100%)</b>

No que se refere a mulheres grávidas, quando questionadas sobre o motivo de internamento, para 47,6% das mulheres tal foi devido a trabalho de parto; 35,7% por cesariana e 16,7% por indução do trabalho de parto.

Mais de metade das mulheres grávidas recebeu analgesia (59,5%), bem como 61,9% receberam anestesia.



- **Caracterização antropométrica e sociodemográfica (mulheres atópicas)**

A tabela seguinte diz respeito às variáveis idade e IMC.

Tabela 3.16: Variáveis antropométricas (mulheres atópicas)

Idade (anos)	Média	Mediana	Desvio padrão	Mínimo	Máximo	<i>n</i>
GA	33,83	33,50	2,81	29	40	24
NGA	34,13	35,50	5,29	20	44	32
IMC (kg/m <sup>2</sup> )						
GA	27,12	26,24	3,18	22,76	34,89	24
NGA	23,97	21,94	4,83	19,13	36,98	31

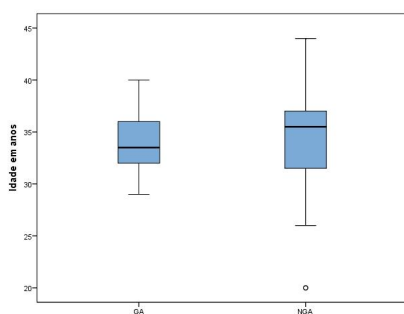


Figura 3.6: Diagrama em caixa de bigodes relativo à idade (mulheres atópicas)

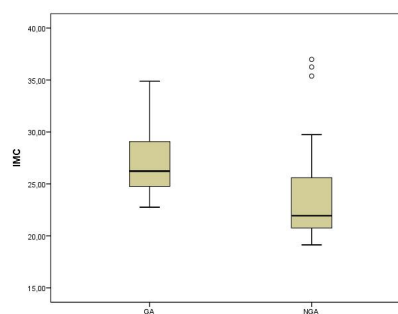


Figura 3.7: Diagrama em caixa de bigodes relativo ao IMC (mulheres atópicas)

De modo a analisar diferenças entre os dois grupos relativamente às variáveis analisadas na tabela 3.16, foi aplicado o teste à normalidade de Shapiro-Wilk, concluindo-se que o modelo normal não se ajusta aos dados para a variável idade, uma vez que se rejeitou a hipótese de normalidade no grupo NGA (Idade:  $p = 0,236$  (GA) e  $p = 0,010$  (NGA)). Estamos portanto em condições de aplicar métodos de inferência estatística não paramétrica para a análise de diferenças entre os dois grupos em estudo relativamente à variável idade.

Uma vez que se trata de amostras independentes e estando na presença de dois grupos, aplicou-se o teste de Mann-Whitney Wilcoxon, concluindo-se que não existe evidência para afirmar que existem diferenças estatisticamente significativas relativamente à idade para os dois grupos de estudo ( $p = 0,287$ ). Relativamente à variável IMC, uma vez que se rejeitou a hipótese de normalidade subjacente aos dados nos dois grupos (IMC:  $p = 0,038$  (GA) e  $p < 0,001$  (NGA)), foi também aplicado o teste de Mann-Whitney Wilcoxon, concluindo-se que existe evidência para afirmar que existem diferenças estatisticamente significativas relativamente ao IMC para os dois grupos de estudo ( $p = 0,001$ ).

Tabela 3.17: Nível de escolaridade (mulheres atópicas)

<b>Escolaridade</b>	<b>GA</b> $n(\%)$	<b>NGA</b> $n(\%)$
Ensino secundário	2 (8,3%)	5 (16,1%)
Ensino superior	22 (91,7%)	26 (83,9%)
<b>Total</b>	<b>24 (100%)</b>	<b>31 (100%)</b>

De acordo com dados obtidos, a grande maioria das mulheres em estudo tinha o ensino superior.

No que diz respeito à atividade profissional, apenas foram obtidos registos para as 21 mulheres, grávidas atópicas, constatando-se que a maioria trabalha por conta de outrém (81,0%).

Relativamente a hábitos tabágicos, verificou-se que a grande maioria das mulheres dos dois grupos não fumavam. No que concerne à exposição ao fumo do tabaco, para o grupo das mulheres grávidas atópicas verificou-se que a maioria das mulheres não eram expostas ao fumo do tabaco. Já 45,2% das mulheres não grávidas atópicas afirmaram que sim.

- **Caracterização clínica**

Tabela 3.18: Pressões arteriais (mulheres atópicas)

<b>P. Sistólica</b>	<b>Média</b>	<b>Mediana</b>	<b>Desvio padrão</b>	<b>Mínimo</b>	<b>Máximo</b>	<b><i>n</i></b>
GA	113,54	110,00	12,13	100	139	24
NGA	109,19	110,00	17,79	70	160	27
<b>P. Diastólica</b>						
GA	66,50	65,00	7,84	50	88	24
NGA	71,41	70,00	14,59	40	120	27

Por aplicação do teste à normalidade de Kolmogorov-Smirnov, concluiu-se que o modelo normal não se ajusta aos dados da variável P. Diastólica, uma vez que se rejeitou a hipótese de normalidade no grupo NGA (P. Diastólica:  $p = 0,370$  (GA) e  $p < 0,001$  (NGA)). Assim, foi aplicado o teste de Mann-Whitney Wilcoxon para analisar a existência de diferenças entre os dois grupos de estudo relativamente à variável pressão arterial diastólica, concluindo-se que não existe evidência para afirmar que existam diferenças estatisticamente significativas entre GA e NGA relativamente a esta variável ( $p = 0,143$ ). Já para a pressão arterial sistólica, uma vez que se rejeitou a hipótese de normalidade subjacente aos dados nos dois grupos (P. Sistólica:  $p = 0,024$  (GA) e  $p = 0,010$  (NGA)), foi também aplicado o teste de Mann-Whitney Wilcoxon, concluindo-se que não existe evidência para afirmar que a pressão arterial sistólica seja diferente nos dois grupos de estudo ( $p = 0,477$ ).

Tabela 3.19: História familiar de atopia (mulheres atópicas)

	<b>GA <i>n</i>(%)</b>
Não	7 (29,2%)
Pai	10 (41,7%)
Avós maternos	6 (25,0%)
Avós paternos	1 (4,2%)
<b>Total</b>	<b>24 (100%)</b>

No que diz respeito ao número de gravidezes anteriores, os resultados

encontram-se na tabela seguinte:

Tabela 3.20: Número de gravidezes anteriores (mulheres atópicas)

<b>Grav. anteriores</b>	<b>GA <math>n(\%)</math></b>	<b>NGA <math>n(\%)</math></b>
0	9 (37,5%)	13 (40,6%)
1	8 (33,3%)	9 (28,1%)
2	6 (25,0%)	6 (18,8%)
3	0 (0,0%)	3 (9,4%)
4	1 (4,2%)	1 (3,1%)
<b>Total</b>	<b>24 (100%)</b>	<b>32 (100%)</b>

Tabela 3.21: Número de filhos (mulheres atópicas)

<b>N.º de filhos</b>	<b>GA <math>n(\%)</math></b>	<b>NGA <math>n(\%)</math></b>
0	12 (50,0%)	13 (40,6%)
1	11 (45,8%)	9 (28,1%)
2	1 (4,2%)	8 (25,0%)
3	0 (0,0%)	2 (6,3%)
<b>Total</b>	<b>24 (100%)</b>	<b>32 (100%)</b>

Tabela 3.22: Parto de termo (mulheres atópicas)

	<b>GA <math>n(\%)</math></b>
Sim	1 (4,2%)
Não	23 (95,8%)
<b>Total</b>	<b>24 (100%)</b>

No que se refere a mulheres grávidas, quando questionadas sobre o motivo de internamento, para 33,3% das mulheres grávidas atópicas tal foi devido a trabalho de parto; 33,3% por cesariana; 29,2% por indução do trabalho de parto e 4,2% por pré-eclâmpsia.

Tabela 3.23: Trabalho de parto (mulheres atópicas)

	<b>GA</b> <i>n</i> (%)
Cesariana	9 (37,5%)
Espontâneo	7 (29,2%)
Induzido	8 (33,3%)
<b>Total</b>	<b>24 (100%)</b>

Tabela 3.24: Tipo de parto (mulheres atópicas)

	<b>GA</b> <i>n</i> (%)
Cesariana eletiva	8 (33,3%)
Cesariana intraparto	3 (12,5%)
Eutócico	7 (29,2%)
Vaginal com ventosa	6 (25,0%)
<b>Total</b>	<b>24 (100%)</b>

Mais de metade das mulheres grávidas atópicas não recebeu analgesia (52,4%). Já 81,0% das mulheres grávidas atópicas receberam anestesia.

### 3.3 Análise dos parâmetros imunológicos

Após se descrever algumas características das mulheres incluídas no estudo, tanto a nível sociodemográfico como clínico, é o momento de se proceder à análise dos parâmetros imunológicos recolhidos com vista a chegar-se ao objetivo principal deste projeto, a identificação de marcadores de risco para a descendência. Todos os parâmetros imunológicos recolhidos encontram-se sob a forma de valor absoluto e percentual, sendo que neste último caso será interpretado como a proporção de células recolhidas de determinado compartimento celular.

Numa primeira parte serão analisados os parâmetros imunológicos de mulheres grávidas e não grávidas saudáveis, de modo a perceber-se de que forma a gravidez influencia o comportamento destes parâmetros.

Numa segunda parte irá proceder-se à análise dos parâmetros imunológicos recolhidos para o grupo de mulheres grávidas atópicas e grávidas não atópicas no 3.<sup>o</sup> trimestre de gestação, no parto e no pós-parto, de forma a perceber a influência da atopia na gravidez.

Os resultados serão apresentados como mediana ( $P_{25}, P_{75}$ ), em que  $P_{25}$  e  $P_{75}$  designam o percentil 25 e o percentil 75, respetivamente, uma vez que não se verificou normalidade subjacente aos dados. Por este facto, todas as comparações foram realizadas por aplicação do teste de Mann-Whitney Wilcoxon.

Dada a extensão dos parâmetros laboratoriais analisados, nesta secção serão apenas apresentados os resultados estatisticamente significativos.

Repare-se que o nosso grande objetivo consiste em estudar o compartimento B durante a gestação, em mulheres saudáveis e atópicas, já que a informação sobre este compartimento celular é escassa (Muzzio, DO; *et al.*, 2014; Toldi, G; *et al.*, 2011; Bohacs, A; *et al.*, 2010).

De acordo com os resultados obtidos nas tabelas seguintes, quando se procedeu à comparação de mulheres grávidas e não grávidas saudáveis (tabelas 3.25 e 3.26), observaram-se alterações importantes no compartimento de linfócitos B, nomeadamente, B reguladoras (Breg) e B de transição tanto em valor absoluto como percentual. Em termos medianos, valores mais elevados destas células são verificados em mulheres não grávidas saudáveis quando comparados com mulheres grávidas saudáveis. Aparentemente parece existir uma influência da gravidez na distribuição destas populações celulares.

Quando procedemos às comparações entre grávidas atópicas e grávidas não atópicas (tabela 3.27), observam-se valores mais elevados das células B de transição, em termos medianos, nas mulheres grávidas atópicas. A atopia parece assim influenciar o compartimento B, destacando-se valores destas células mais elevados tanto em termos percentuais como em valor absoluto.

De acordo com as análises efetuadas, as diferenças decorrentes da presença de atopia diluem-se durante a gravidez até ao momento do parto e estendem-se ainda assim até ao fim do puerpério.

Considerando que se observaram alterações no compartimento de linfócitos B quer em mulheres grávidas saudáveis, quer em mulheres grávidas atópicas, impôs-se estudar este compartimento do ponto de vista da descendência.

O período de vida pré-natal tem influência no desenvolvimento do sistema imunitário da criança onde se destacam as próprias interações com o sistema imune-materno. Neste sentido, as alterações identificadas no 3.º trimestre de gravidez ao nível do sistema imunitário poderão ter impacto para a criança, relacionando-se eventualmente com o desenvolvimento de atopia.

Tabela 3.25: Análise de diferenças entre grávidas saudáveis e não grávidas saudáveis relativamente aos parâmetros imunológicos avaliados no 3.º trimestre

<b>Parâmetro</b>	<b>GNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>NGNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>valor-p</b>
Linfócitos	1770,0 (1467,0; 2148,0)	2060,0 (1818,0; 2661,0)	0,001
CD3+ %	80,0 (77,0; 83,0)	75,0 (71,0; 78,0)	< 0,001
CD3+ abs	1440,0 (1127,0; 1740,0)	1571,0 (1305,0; 1990,0)	0,027
CD19+ %	10,0 (8,0; 12,0)	12,0 (10,0; 14,0)	0,023
CD19+ abs	173,0 (144,0; 219,0)	259,0 (176,0; 305,0)	< 0,001
CD16/56 %	9,0 (6,0; 11,0)	11,0 (10,0; 13,0)	0,001
CD16/56 abs	149,0 (100,0; 205,0)	271,0 (178,0; 327,0)	< 0,001
CD4+ %	48,0 (45,0; 53,0)	45,0 (42,0; 50,0)	0,015
CD4+ abs	829,0 (713,0; 1046,0)	987,0 (830,0; 1199,0)	0,043
CD4+ DR+ %	5,33 (3,66; 7,42)	6,14 (5,5; 8,16)	0,029
CD4+ DR+ abs	48,9 (32,16; 64,12)	69,7 (47,6; 87,96)	0,006
Neutrófilos %	71,7 (69,57; 75,65)	56,0 (50,52; 61,53)	< 0,001
Neutrófilos abs	7056,0 (5599,0; 8377,0)	4043,0 (2661,0; 4694,0)	< 0,001
Basófilos %	0,4 (0,27; 0,47)	0,9 (0,65; 1,08)	< 0,001
Basófilos abs	36,0 (27,0; 48,0)	54,0 (43,0; 83,0)	< 0,001
Eosinófilos %	1,0 (0,75; 1,53)	1,8 (1,3; 2,6)	< 0,001
Monócitos abs	663,0 (538,0; 799,0)	524,0 (400,0; 649,0)	< 0,001
3+8-IFNG+ abs	280,4 (230,94; 358,74)	357,4 (274,48; 430,12)	0,011
3+8-IL17+ abs	17,3 (13,48; 21,78)	20,5 (16,63; 27,68)	0,005
3+8-IFN Total abs	286,6 (234,6; 366,9)	369,1 (280,1; 439,9)	0,011
3+8-IL17+Total abs	26,8 (19,6; 32,4)	28,4 (24,8; 42,1)	0,012
3+8+IFNG+abs	242,3 (190,9; 338,6)	333,9 (212,9; 394,2)	0,027
3+8+IFN Total abs	256,7 (190,9; 338,6)	342,5 (219,4; 415,4)	0,03
19+IL10+ abs	1,6 (1,1; 2,3)	2,0 (1,5; 3,3)	0,013
BM1 %	7,9 (5,8; 11,4)	10,4 (8,4; 13,8)	0,038
BM1 abs	12,3 (9,4; 22,3)	25,8 (16,3; 34,1)	< 0,001
BM2 %	67,7 (60,3; 72,7)	57,4 (51,04; 64,16)	< 0,001
BM2' %	2,6 (1,6; 4,4)	6,2 (4,3; 10,7)	< 0,001
BM2' abs	4,2 (2,6; 7,8)	13,4 (9,7; 26,8)	< 0,001
BM3+BM4 %	1,5 (1,0; 2,4)	2,9 (1,6; 3,8)	0,002
BM3+BM4 abs	2,3 (1,7; 3,9)	6,8 (3,9; 9,8)	< 0,001
BM5 %	4,7 (3,7; 6,6)	6,0 (4,9; 8,4)	0,01
BM5 abs	8,1 (5,9; 11,9)	15,8 (9,9; 20,6)	< 0,001
eBM5 abs	18,9 (12,7; 26,1)	27,3 (16,1; 39,2)	0,009
DN %	2,5 (1,7; 3,6)	3,9 (2,9; 4,8)	0,001
DN abs	4,1 (2,6; 6,2)	9,5 (5,5; 13,4)	< 0,001
B transição %	1,9 (1,1; 2,6)	3,7 (2,4; 4,9)	< 0,001
B transição abs	3,3 (1,5; 4,8)	8,8 (6,1; 13,1)	< 0,001
Plasmablastos %	1,4 (0,9; 2,4)	2,9 (1,5; 4,1)	0,001
Plasmablastos abs	2,3 (1,7; 3,4)	6,7 (4,1; 10,9)	< 0,001
Bnaive abs	99,6 (71,5; 148,6)	149,4 (96,26; 188,47)	0,002



Tabela 3.26: Análise de diferenças entre grávidas saudáveis e não grávidas saudáveis relativamente aos parâmetros imunológicos avaliados no 3.º trimestre (continuação)

<b>Parâmetro</b>	<b>GNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>NGNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>valor-p</b>
Bmemória abs	63,8 (43,78; 85,15)	78,7 (63,1; 108,9)	0,006
Bmemória switched abs	24,4 (17,6; 30,9)	36,8 (25,8; 52,9)	<0,001
Breg 24+38+ %	1,4 (0,94; 2,32)	3,6 (2,17; 4,65)	<0,001
Breg 24+38+ abs	2,2 (1,7; 4,2)	8,4 (5,9; 13,8)	<0,001
Breg 24+27++ abs	1,1 (0,6; 1,9)	2,5 (1,3; 3,9)	0,001
Breg 24+27+ abs	44,3 (29,8; 70,7)	53,9 (44,3; 84,4)	0,037
FOX GeoMean	17,5 (14,3; 20,1)	21,1 (18,9; 23,7)	<0,001

Tabela 3.27: Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre de gestação

<b>Parâmetro</b>	<b>GA - m(<math>P_{25}, P_{75}</math>)</b>	<b>GNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>valor-p</b>
DN abs	5,6 (2,6; 7,2)	4,1 (2,6; 6,2)	0,004
CD8+DR+ %	23,2 (17,8; 35,6)	18,2 (12,8; 24,9)	0,039
Eosinófilos %	1,6 (1,0; 2,8)	1,0 (0,8; 1,5)	0,024
3+8-IFNG+IL17+ %	1,2 (0,9; 1,6)	0,9 (0,7; 1,1)	0,012
3+8-IL17+ Total %	3,5 (2,8; 4,0)	2,9 (2,3; 3,6)	0,039
3+8+IFNG+IL17+ %	1,0 (0,8; 1,4)	0,7 (0,5; 1,0)	0,028
3+8+IL17+Total %	1,8 (1,5; 2,6)	1,3 (0,9; 2,0)	0,02
3+8+IFNG+IL17+ abs	5,2 (3,2; 7,4)	3,5 (2,4; 5,4)	0,039
3+8+IL17+ Total abs	8,4 (6,4; 13,9)	6,6 (4,4; 9,5)	0,044
19+IL10+ %	1,3 (0,9; 1,6)	1,0 (0,6; 1,5)	0,022
19+IL10+ abs	2,4 (1,3; 3,5)	1,6 (1,1; 2,3)	0,011
B transição %	3,2 (1,9; 4,3)	1,9 (1,1; 2,6)	0,002
B transição abs	5,6 (3,1; 8,4)	3,3 (1,5; 4,8)	0,007

Tabela 3.28: Análise de diferenças entre não grávidas saudáveis e não grávidas atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre

<b>Parâmetro</b>	<b>NGA - m(<math>P_{25}, P_{75}</math>)</b>	<b>NGNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>valor-p</b>
Linfócitos	2752,5 (2113,25; 3464,0)	2060,0 (1818,0; 2661,0)	0,003
CD3+ abs	2098,5 (1581,5; 2588,25)	1571,0 (1305,0; 1990,0)	0,003
CD19+ abs	327,5 (249,25; 436,5)	259,0 (176,0; 305,0)	0,001
CD16/56 %	9,0 (5,25; 13,0)	11,0 (10,0; 13,0)	0,01
CD4+ abs	1207,0 (980,75; 1522,5)	987,0 (830,0; 1199,0)	0,011
CD8+ abs	781,0 (569,0; 999,0)	558,0 (415,0; 689,0)	0,001
CD3+CTLA4+ %	1,825 (1,053; 3,098)	3,28 (1,59; 9,17)	0,006
CD3+DR+ abs	241,21 (163,49; 323,12)	174,07 (114,06; 251,96)	0,031
CD3+CD25+ abs	418,48 (261,68; 640,18)	308,61 (214,14; 428,99)	0,009
CD4+CTLA+ %	1,56 (1,0; 2,88)	3,01 (1,38; 8,53)	0,004
CD4+ DR+ abs	86,19 (66,46; 111,58)	69,72 (47,61; 87,96)	0,009
CD4+CD25+ abs	387,98 (250,01; 557,08)	275,22 (199,92; 389,62)	0,01
CD8+CTLA+ %	1,92 (1,17; 3,42)	3,07 (1,79; 10,09)	0,005
CD8+CD25+4+ abs	22,07 (14,63; 35,89)	13,93 (7,90; 29,08)	0,04
Eosinófilos %	3,39 (2,13; 4,46)	1,78 (1,3; 2,61)	0,001
Eosinófilos abs	238,0 (163,5; 384,25)	118,0 (71,0; 189,0)	<0,001
Neutrófilos abs	4702,5 (3639,5; 5712,25)	4043,0 (2661,0; 4694,0)	0,023
Basófilos abs	77,0 (53,5; 96,75)	54,0 (43,0; 83,0)	0,025
CD4+CD127+ %	94,48 (92,74; 95,21)	95,96 (94,64; 96,66)	<0,001
CD4+CD127+ abs	1148,56 (932,21; 1413,79)	937,57 (772,94; 1167,11)	0,018
4+25++127- abs	96,48 (75,04; 126,28)	72,23 (59,03; 94,28)	0,003
CD4+CD25+Foxp3+ %	6,88 (5,45; 7,61)	5,48 (4,61; 6,30)	0,002
CD4+CD25+Foxp3+ abs	83,2 (68,88; 101,71)	56,05 (38,07; 68,75)	<0,001
CD25++CD4+ %	6,55 (5,94; 7,14)	5,98 (5,29; 6,71)	0,029
CD25++CD4+ abs	82,98 (63,65; 103,26)	55,77 (42,15; 76,28)	<0,001
3+8-IFNG+ abs	534,99 (338,37; 624,92)	357,39 (274,48; 430,13)	0,001
3+8-IFN Total abs	551,86 (353,47; 643,07)	369,05 (280,14; 439,92)	0,001
3+8-IL17+Total abs	40,79 (28,2; 58,09)	28,42 (24,76; 42,09)	0,032
3+8+IFNG+abs	401,79 (289,27; 670,68)	333,91 (212,98; 394,21)	0,016
3+8+IL17+ abs	4,35 (3,1; 7,9)	3,3 (2,5; 4,8)	0,032
3+8+IFN+ Total abs	407,44 (291,98; 674,02)	342,45 (219,36; 415,42)	0,014
3+IL10+ %	0,94 (0,73; 1,19)	0,72 (0,51; 1,08)	0,025
3+IL10+ abs	19,11 (14,11; 26,63)	11,65 (7,64; 18,51)	<0,001
3+8-IL10+ %	1,03 (0,84; 1,54)	0,84 (0,65; 1,20)	0,036
3+8-IL10+ abs	13,63 (10,17; 16,84)	8,49 (5,87; 11,15)	<0,001
3+8+IL10+ %	0,72 (0,48; 0,99)	0,39 (0,21; 0,80)	0,01
3+8+IL10+ abs	4,34 (2,94; 8,62)	1,94 (1,14; 4,03)	<0,001
19+IL10+ abs	3,65 (2,22; 4,75)	2,02 (1,47; 3,28)	0,001
BM1 %	13,76 (9,96; 20,26)	10,35 (8,41; 13,82)	0,013
BM1 abs	45,48 (33,22; 82,39)	25,79 (16,34; 34,12)	<0,001
BM2 abs	188,47 (124,48; 253,51)	149,07 (93,09; 184,44)	0,005

Tabela 3.29: Análise de diferenças entre não grávidas saudáveis e não grávidas atópicas relativamente aos parâmetros imunológicos avaliados no 3.º trimestre (continuação)

<b>Parâmetro</b>	<b>NGA - m(<math>P_{25}, P_{75}</math>)</b>	<b>NGNA - m(<math>P_{25}, P_{75}</math>)</b>	<b>valor-p</b>
BM2' %	3,54 (2,89; 5,56)	6,22 (4,33; 10,70)	0,002
BM3+BM4 %	1,06 (0,71; 1,56)	2,89 (1,59; 3,81)	<0,001
BM3+BM4 abs	3,78 (2,20; 6,01)	6,76 (3,91; 9,81)	0,001
BM5 %	8,83 (5,69; 12,85)	6,0 (4,93; 8,40)	0,006
BM5 abs	30,18 (18,37; 47,62)	15,79 (9,87; 20,56)	<0,001
eBM5 abs	41,99 (28,34; 57,14)	27,33 (16,05; 39,19)	0,002
DN %	2,31 (1,59; 4,38)	3,87 (2,96; 4,82)	0,026
Plasmablastos %	1,11 (0,66; 1,47)	2,85 (1,46; 4,09)	<0,001
Plasmablastos abs	3,5 (2,4; 5,3)	6,7 (4,1; 10,9)	0,001
Bnaive abs	187,65 (127,28; 267,84)	149,37 (96,08; 189,46)	0,015
B memoria abs	140,83 (91,90; 206,00)	78,68 (63,06; 108,85)	0,001
B memoria n switched abs	66,82 (52,02; 113,04)	45,44 (33,07; 67,89)	<0,001
B memoria switched abs	63,96 (43,86; 83,14)	36,82 (25,84; 52,93)	0,004
Breg 24+27+ abs	101,88 (68,45; 165,16)	53,98 (44,33; 84,39)	<0,001
Breg 24+38+ abs	12,16 (7,88; 15,47)	8,35 (5,95; 13,81)	0,022
CD3+CTLA4+H %	1,31 (0,84; 2,28)	0,9 (0,55; 1,50)	0,03
CD3+CTLA4+H abs	27,29 (14,27; 58,16)	14,30 (9,0; 23,6)	0,002
CD4+CTLA4+H abs	17,48 (9,33; 30,67)	7,44 (5,15; 13,49)	0,002
CD8+CTLA4+H abs	10,63 (4,47; 20,01)	5,10 (3,80; 9,02)	0,012
CD4+GEO MFI	10,38 (9,47; 11,24)	11,43 (10,04; 12,50)	0,005
Bm1 memoria +	9,28 (6,43; 12,31)	6,15 (4,88; 7,75)	0,004
Bm1 memoria -	28,40 (20,63; 57,37)	14,80 (10,45; 21,23)	<0,001
Bm1 naive	5,15 (3,11; 8,03)	3,92 (2,93; 5,90)	0,001
Bm5 CD27+	84,50 (74,29; 88,29)	73,79 (68,12; 77,24)	<0,001
Bm5 CD27-	15,49 (11,70; 25,70)	26,21 (22,76; 31,88)	<0,001
eBm5 CD27+	86,17 (79,02; 89,33)	78,51 (75,9; 84,1)	0,005
eBm5 CD27-	13,83 (10,66; 20,97)	21,49 (15,86; 24,10)	0,005
Bm3+4 CD27+	81,25 (71,23; 90,18)	88,39 (83,33; 92,26)	0,009
Bm3+4 CD27-	18,75 (9,78; 28,76)	11,61 (7,74; 16,67)	0,009
IFN1	10,73 (0,0; 36,2)	54,0 (12,1; 77,6)	0,001
IFN2	10,73 (2,8; 36,2)	54,0 (12,1; 77,6)	0,002

Tabela 3.30: Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no dia do parto

<b>Parâmetro</b>	<b>GA - <math>m(P_{25}, P_{75})</math></b>	<b>GNA - <math>m(P_{25}, P_{75})</math></b>	<b>valor-p</b>
DN %	1,0 (1,0; 3,0)	3,0 (2,0; 5,0)	0,016
3+8-IL17+ %	2,0 (1,6; 2,4)	1,6 (1,3; 2,2)	0,011
3+8-IFNG+IL17+ %	0,9 (0,7; 1,4)	0,7 (0,5; 0,9)	0,004
3+8-IL17+ Total %	2,9 (2,5; 3,7)	2,3 (1,7; 2,9)	0,004
3+8-IFNG+IL17+ abs	7,8 (6,4; 10,1)	5,6 (4,4; 8,1)	0,006
3+8-IL17+Total abs	23,9 (19,7; 32,5)	20,1 (15,5; 25,7)	0,033

Tabela 3.31: Comparações entre grávidas atópicas e grávidas não atópicas relativamente aos parâmetros imunológicos avaliados no pós-parto

<b>Parâmetro</b>	<b>GA - <math>m(P_{25}, P_{75})</math></b>	<b>GNA - <math>m(P_{25}, P_{75})</math></b>	<b>valor-p</b>
CD4+CD127+ %	94,9 (93,6; 96,5)	96,3 (94,9; 97,3)	0,029
3+8+IL17+ %	0,9 (0,4; 1,3)	0,5 (0,4; 0,7)	0,022
3+8+IL17+ abs	4,3 (2,8; 7,2)	2,9 (2,1; 4,5)	0,014
Bm1 abs	26,7 (16,4; 43,6)	17,7 (13,8; 22,1)	0,031
Plasmablastos %	2,0 (1,0; 3,8)	3,1 (2,0; 4,9)	0,022
MFI geo CD4+	10,8 (8,9; 12,3)	12,6 (10,7; 14,1)	0,022

## Capítulo 4

# Identificação de marcadores de risco para o desenvolvimento de atopia na descendência de mulheres atópicas

A partir deste momento começamos a aproximarmo-nos daquele que é o objetivo principal deste projeto: identificação de marcadores de risco para o desenvolvimento de atopia na descendência. Esta secção inicia-se com uma breve caracterização das crianças, filhas de mães atópicas, tanto a nível antropométrico como clínico, passando-se posteriormente para a identificação de marcadores de risco através do cálculo de *odds-ratio*. Para efeitos de notação, neste capítulo, designaremos descendência com manifestações de risco para a atopia como "D. atópica" e descendência não atópica como "D. não atópica".

## 4.1 Caracterização sociodemográfica e antropométrica das crianças, filhas de mulheres atópicas

Tabela 4.1: Problemas neonatais de atopia

	<i>n</i> (%)
Sim	10 (41,7%)
Não	14 (58,3%)
<b>Total</b>	<b>24 (100%)</b>

Tabela 4.2: Características antropométricas

	D. não atópica		D. atópica	
		<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Sexo	Masculino	4 (28,6%)	8 (80,0%)	
	Feminino	10 (71,4%)	2 (20,0%)	
Total		14 (100%)	10 (100%)	
Mês do nascimento outubro/março	Sim	7 (50,0%)	6 (60,0%)	
	Não	7 (50,0%)	4 (40,0%)	
Total		14 (100%)	10 (100%)	
Freq. infantário	Sim	3 (21,4%)	4 (44,4%)	
	Não	11 (78,6%)	5 (55,6%)	
Total		14 (100%)	9 (100%)	

Repare-se que os resultados obtidos na tabela 4.2, nomeadamente na análise de descendência atópica por sexo, são consistentes com a teoria de (Shah A, 2012; Braza F, *et al.*, 2014), que indica que a presença de atopia é superior nas crianças do sexo masculino. Observou-se que 80% das crianças atópicas são do sexo masculino.

Mais ainda, 60% das crianças atópicas nasceram no período de outubro/março, período onde se regista um maior número de alergias.

Tabela 4.3: Peso (gramas)

Peso (gramas)	Média	Mediana	Desvio padrão	Min.	Máx.	<i>n</i>
D. não atópica	3249,29	3295,00	410,89	2330	3940	14
D. atópica	3279,00	3287,50	177,23	3040	3605	10

De modo a analisar possíveis diferenças entre crianças atópicas e não atópicas no que respeita ao peso, foi aplicado o teste à normalidade de Shapiro-Wilk. De acordo com os resultados obtidos com este teste, concluiu-se que não se rejeita a hipótese de normalidade subjacente aos dados nos dois grupos ( $p = 0,881$  (D. não atópica) e  $p = 0,856$  (D. atópica)). Assim, foi aplicado o teste t para amostras independentes, concluindo-se que não existe evidência para afirmar que existam diferenças entre crianças atópicas e crianças não atópicas no que respeita ao peso ( $p = 0,833$ ).

## 4.2 Identificação de marcadores de risco

Após uma análise exploratória das variáveis respeitantes às crianças, filhas de mães atópicas, pretende-se agora identificar possíveis marcadores de risco para o desenvolvimento de atopia nestas crianças.

Serão tidas em conta para esta análise medidas de risco como o *odds-ratio* (*OR*) e respetivos intervalos a 95% de confiança.

Nestes casos, verificando-se o pressuposto de que não mais do que 20% das células tenham frequências esperadas inferiores a 5 será aplicado o teste de independência do qui-quadrado, uma vez que o principal objetivo é o de testar a independência das variáveis em estudo, ou seja, verificar se existe associação entre marcador e desenvolvimento de atopia. Caso o pressuposto de utilização do teste de independência do  $\chi^2$  não seja verificado será utilizado o teste exato de Fisher.

Tal como vimos anteriormente, a variável de interesse a considerar será a presença/ausência de atopia na descendência.

Para cada uma das variáveis pretende-se então testar:

$H_0$ : as variáveis são independentes

*vs.*

$H_1$ : as variáveis não são independentes

Contudo, como podemos constatar, a variável de interesse é categórica e os parâmetros imunológicos são variáveis contínuas, pelo que será necessário categorizar estes parâmetros. Assim sendo, num primeiro passo, irão analisar-se diferenças entre D. atópica e D. não atópica relativamente aos parâmetros imunológicos maternos avaliados no 3.º trimestre de gravidez. Depois desta análise, para aqueles onde se verificou a existência de diferenças estatisticamente significativas, serão construídas curvas ROC de modo a identificar pontos de corte que permitam categorizar os parâmetros imunológicos e então posteriormente testar-se possíveis associações com a variável resposta. Os pontos de corte foram determinados de acordo com o seu significado clínico e estatístico, no sentido de garantir um compromisso entre o número de verdadeiros positivos e o número de falsos positivos.

**Passo 1** - Comparação entre D. atópica e D. não atópica relativamente aos parâmetros imunológicos maternos avaliados no 3.º trimestre de gestação

Com o objetivo de analisar possíveis diferenças nos parâmetros imunológicos maternos, avaliados no 3.º trimestre de gravidez, entre mulheres com e sem manifestações atópicas na descendência, foi aplicado o teste à normalidade de Shapiro-Wilk, concluindo-se que o modelo normal não se ajusta aos dados. Estamos portanto em condições de aplicar métodos de inferência estatística não paramétrica para a análise de diferenças entre os dois grupos em estudo relativamente aos parâmetros imunológicos.

Uma vez que se trata de amostras independentes e estando na presença de dois grupos, aplicou-se o teste de Mann-Whitney Wilcoxon, concluindo-se que existe evidência para afirmar que existem diferenças estatisticamente significativas para os seguintes parâmetros:



Tabela 4.4: Comparação entre D. atópica e D. não atópica relativamente aos parâmetros imunológicos maternos avaliados no 3.º trimestre de gestação

Parâmetro	D. não atópica m ( $P_{0,25}, P_{0,75}$ )	D. atópica m ( $P_{0,25}, P_{0,75}$ )	valor-p
CD3+ %	76,0 (72,0; 78,5)	81,0 (77,0; 83,3)	0,02
CD19+ %	11,5 (9,8; 15,0)	8,5 (7,0; 10,8)	0,025
BM1 abs	26,8 (13,9; 39,1)	11,3 (7,1; 23,9)	0,035
B transição %	2,3 (1,4; 3,1)	4,3 (3,9; 5,2)	0,002
B memória abs	85,9 (49,4; 127,8)	50,3 (35,4; 66,1)	0,046
B memória não switched abs	55,5 (28,9; 68,9)	27,3 (21,1; 42,5)	0,03
Breg 24+38+ %	1,6 (0,7; 2,4)	2,7 (1,8; 4,6)	0,03

**Passo 2** - Construção das curvas ROC para obtenção dos pontos de corte

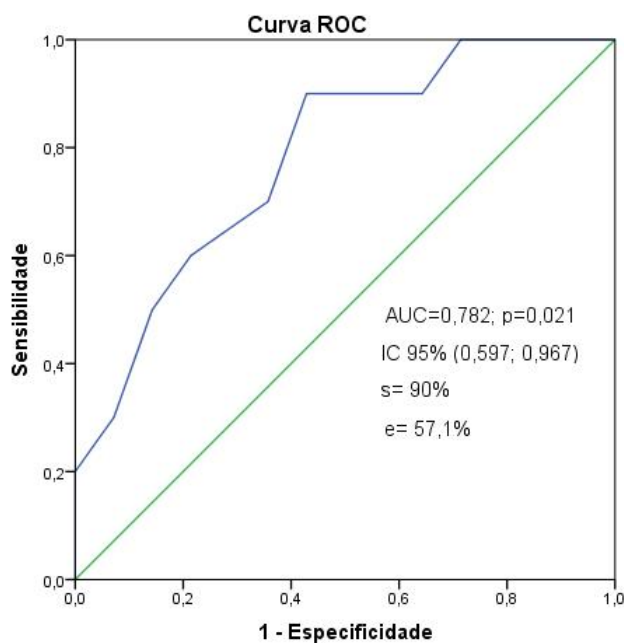


Figura 4.1: Curva ROC para linfócitos T (CD3+) %

Analisando a figura 4.1, que corresponde à curva ROC para o parâmetro imunológico linfócitos T (CD3+) %, obtém-se uma área abaixo da curva de 0,782, sendo esta estatisticamente significativa,  $p=0,021$  ( $H_0 : AUC = 0,5$  vs  $H_1 : AUC \neq 0,5$ ). Obteve-se ainda uma sensibilidade de 90 % e uma especificidade de 57,1 %. Sendo a sensibilidade, por definição, a probabilidade de se obter um resultado positivo, sabendo que tem problemas neonatais de atopia, ou seja, a proporção de verdadeiros positivos, temos então que se conseguem prever corretamente 90% dos casos. Já para a proporção de verdadeiros negativos, consegue-se prever corretamente 57% dos casos. O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 76,5 %.

A situação ideal seria obter valores de sensibilidade de 100%, contudo, isso implicaria a inexistência de falsos positivos e falsos negativos, o que é raro acontecer.

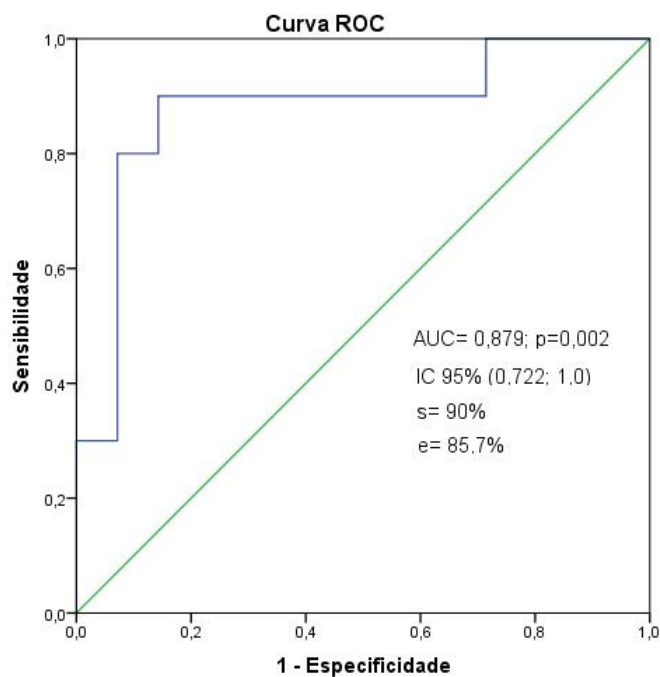


Figura 4.2: Curva ROC para linfócitos B transição %

Considerando as células B de transição (%), a curva ROC encontra-se representada na figura 4.2. Esta é desde já a curva com maior área relativamente às restantes ( $AUC = 0,879$ ), sendo esta estatisticamente significativa,  $p = 0,002$  ( $H_0 : AUC = 0,5$  vs  $H_1 : AUC \neq 0,5$ ). Obteve-se uma sensibilidade de 90 % e uma especificidade de 85,7 %, significando isto que com este parâmetro imunológico consegue-se prever corretamente 90% dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 86% dos verdadeiros casos negativos para problemas neonatais de atopia. Este foi sem dúvida o parâmetro para o qual se obteve melhores resultados para a predição de atopia na descendência. O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 3,585 %.

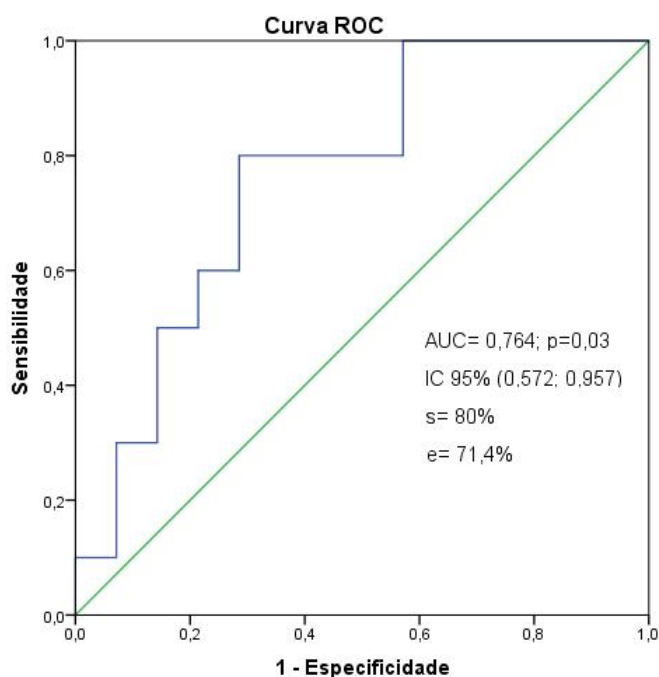


Figura 4.3: Curva ROC para linfócitos Breg 24<sup>hi</sup>+38<sup>hi</sup> %

Para as células B reguladoras CD24<sup>hi</sup>CD38<sup>hi</sup> %, obteve-se uma área abaixo da curva de 0,764, sendo esta estatisticamente significativa,  $p=0,03$  ( $H_0 : AUC = 0,5$  vs  $H_1 : AUC \neq 0,5$ ). Obtiveram-se valores de sensibilidade e especificidade de 80 % e 71,4 %, respetivamente. Querem estes valores dizer que com este parâmetro imunológico conseguem-se prever corretamente 80 % dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 71 % dos verdadeiros casos negativos para problemas neonatais de atopia.

O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 1,865 %.

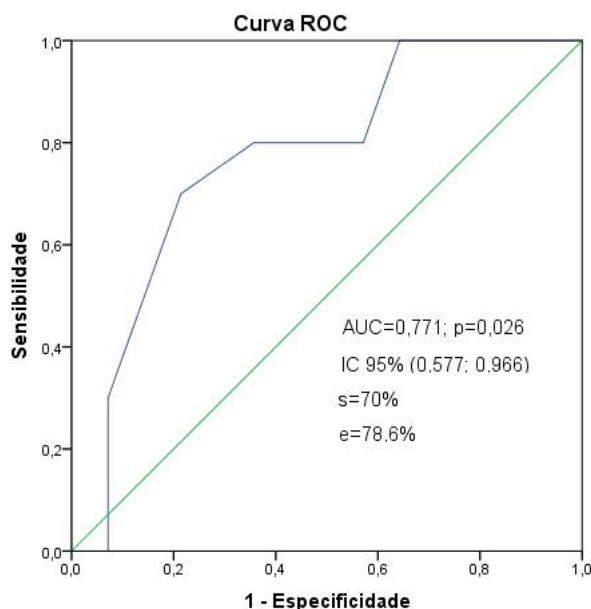


Figura 4.4: Curva ROC para linfócitos B (CD19+)%

Relativamente aos linfócitos B (CD19+)%, obteve-se uma área abaixo da curva de 0,771, sendo esta estatisticamente significativa ( $p = 0,026$ ). Obtiveram-se valores de sensibilidade e especificidade de 70 % e 78,6 %, respetivamente. Querem estes valores dizer que com este parâmetro imunológico conseguem-se prever corretamente 70 % dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 79 % dos verdadeiros

casos negativos para problemas neonatais de atopia.

O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 9,5 %.

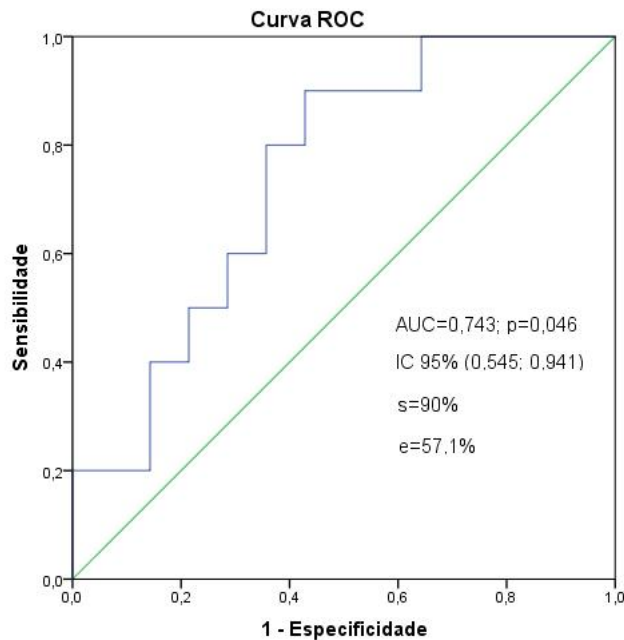


Figura 4.5: Curva ROC para linfócitos B memória

Para as células B de memória obteve-se uma área abaixo da curva de 0,743, sendo esta estatisticamente significativa ( $p = 0,046$ ). Obtiveram-se valores de sensibilidade e especificidade de 90 % e 57,1 %, respetivamente. Querem estes valores dizer que com este parâmetro imunológico conseguem-se prever corretamente 90 % dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 58 % dos verdadeiros casos negativos para problemas neonatais de atopia.

O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 85,5 células/ $\mu L$ .

Considerando os linfócitos Bm1, obteve-se uma área abaixo da curva de 0,757, sendo esta estatisticamente significativa ( $p = 0,035$ ). Obtiveram-se valores de sensibilidade e especificidade de 80 % e 64,3 %, respetivamente.

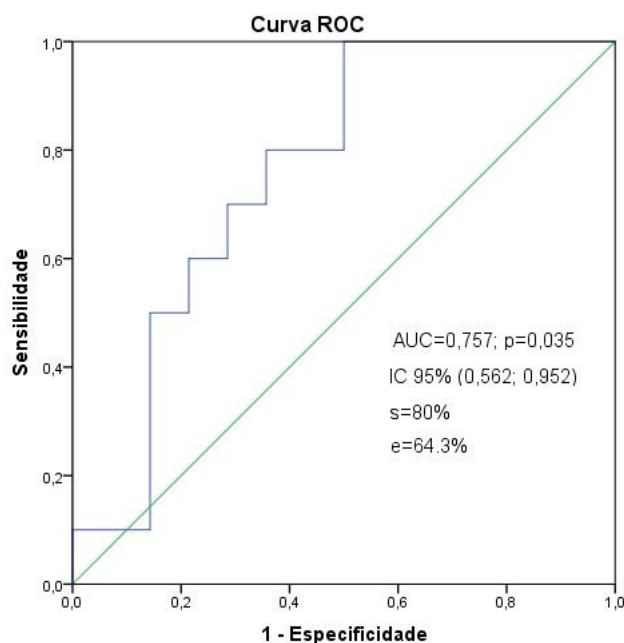


Figura 4.6: Curva ROC para linfócitos Bm1

Querem estes valores dizer que com este parâmetro imunológico conseguem-se prever corretamente 80 % dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 64 % dos verdadeiros casos negativos para problemas neonatais de atopia.

O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 24 células/ $\mu L$ .

Por último, no que diz respeito aos linfócitos B de memória não *switched*, obteve-se uma área abaixo da curva de 0,764, sendo esta estatisticamente significativa ( $p = 0,030$ ). Obtiveram-se valores de sensibilidade e especificidade de 90 % e 57,1 %, respetivamente. Querem estes valores dizer que com este parâmetro imunológico conseguem-se prever corretamente 90 % dos verdadeiros casos positivos para problemas neonatais de atopia e aproximadamente 57 % dos verdadeiros casos negativos para problemas neonatais de atopia.

O ponto de corte que resulta nos valores de sensibilidade e especificidade é de 46,5 células/ $\mu L$ .

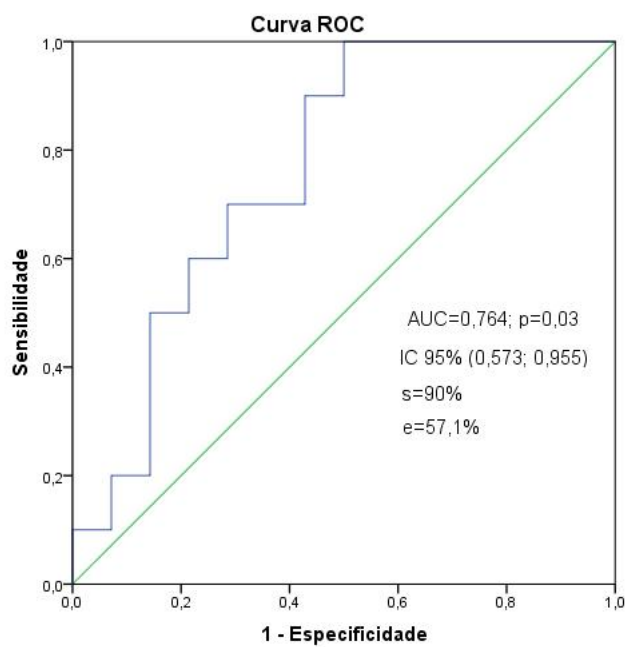


Figura 4.7: Curva ROC para linfócitos B memória não *switched*

### Passo 3 - Análise de associações

Na tabela seguinte encontram-se os resultados obtidos com a análise de associações entre os parâmetros imunológicos e a presença/ausência de atopia na descendência. Os valores de prova obtidos com \* resultaram da aplicação do teste exato de Fisher. Os restantes foram obtidos por aplicação do teste de independência do  $\chi^2$ .

Tabela 4.5: Análise de associação entre os parâmetros imunológicos e presença/ausência de atopia na descendência

	<b>D. não atópica</b>	<b>D. atópica</b>	<b>OR (IC 95%)</b>	<b>valor-p</b>
CD3+%	8/14 (57,1%)	1/10 (10,0%)	12 (1,18; 122,27)	0,033 *
CD19+%	11/14 (78,6%)	3/10 (30,0%)	8,56 (1,33; 54,95)	0,035 *
Bm1 abs	9/14 (64,3%)	2/10 (20,0%)	7,2 (1,08; 47,96)	0,047 *
B transição %	12/14 (85,7%)	1/10 (10,0%)	54 (4,2; 692,5)	0,001 *
B memória abs	8/14 (57,1%)	1/10 (10,0%)	12 (1,18; 122,27)	0,033 *
B não switched abs	8/14 (57,1%)	1/10 (10,0%)	12 (1,18; 122,27)	0,033 *
B reg 24+38+%	10/14 (71,4%)	2/10 (20,0%)	10 (1,44; 69,26)	0,013

De acordo com os resultados obtidos, todos os parâmetros analisados são marcadores de risco para o desenvolvimento de atopia na descendência. De realçar que as células B de transição foram o marcador com maior *odds ratio* ( $OR = 54$ ), o que significa que a chance de uma criança vir a desenvolver atopia é 54 vezes maior nas mulheres que tenham valores de células B de transição superiores a 3,585%.

Resta agora saber, quais destes marcadores de risco é que são de facto importantes para explicar a probabilidade de ocorrência de manifestações atópicas na descendência. Para tal, no capítulo seguinte irá proceder-se à construção de modelos de regressão logística, para tentar dar resposta a esta questão.

Uma vez que as células T (CD3+)% e as células B (CD19+) % fazem parte da mesma população de linfócitos, em termos clínicos e estatísticos não é relevante estar a repetir a mesma informação no modelo que se irá construir na secção seguinte. Assim, destas duas variáveis, a que será alvo



de análise serão as células T (CD3+) %.

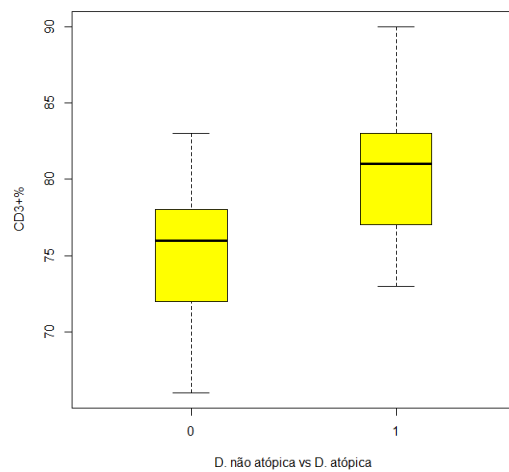


Figura 4.8: Diagrama em caixa de bigodes paralelas relativo à variável linfócitos T (CD3+) %

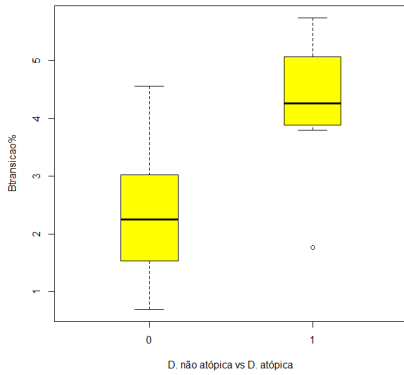


Figura 4.9: Diagrama em caixa de bigodes paralelas relativo à variável Btransição %

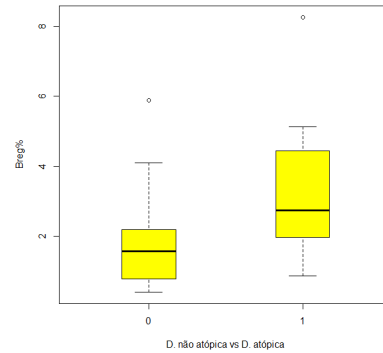


Figura 4.10: Diagrama em caixa de bigodes paralelas relativo à variável linfócitos Breg  $24^{hi}+38^{hi}$  %

De acordo com a análise dos diagramas em caixa de bigodes apresentados, observa-se que existe uma tendência para que os parâmetros imunológicos maternos analisados estejam mais elevados nas mulheres com descendência atópica. De seguida, vamos analisar como é que estes parâmetros influenciam a resposta à presença/ausência de atopia.

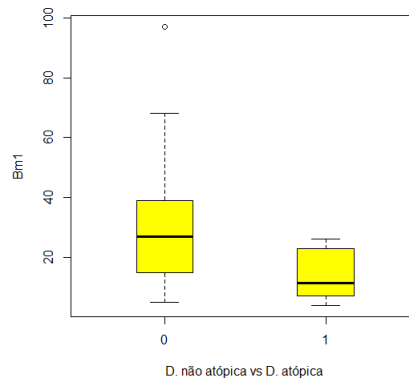


Figura 4.11: Diagrama em caixa de bigodes relativo à variável Bm1

Comparativamente aos parâmetros imunológicos percentuais, repare-se agora que estas populações celulares maternas apresentam contagens celulares mais elevadas nas mulheres sem manifestações atópicas na descendência.

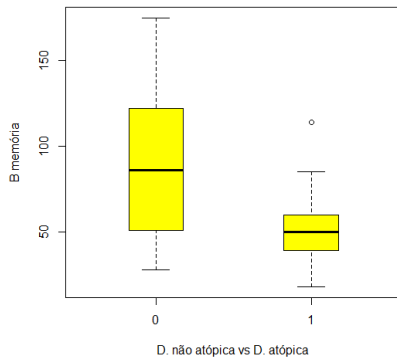


Figura 4.12: Diagrama em caixa de bigodes relativo à variável B memória

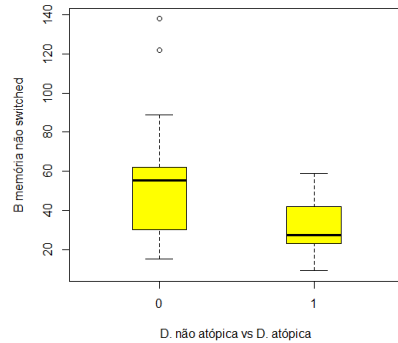


Figura 4.13: Diagrama em caixa de bigodes relativo à variável B memória não switched

## 4.3 Análise de regressão

Nesta secção será estudada a relação entre a presença/ausência de atopia na descendência e um conjunto de covariáveis que se consideraram relevantes. Toda a análise de regressão será efetuada com recurso ao *software R* versão 3.2.2.

### 4.3.1 Modelos de regressão logística

Dado que a variável resposta é binária, e sabendo que as covariáveis podem ser de qualquer natureza, procedeu-se à construção de um modelo de regressão logística para tentar dar resposta à questão da relação entre a presença/ausência de atopia na descendência e um conjunto de covariáveis.

As covariáveis a incluir no modelo inicial foram escolhidas pela sua significância clínica e estatística.

Uma vez que foram recolhidos parâmetros imunológicos (covariáveis) percentuais e em valor absoluto, esta informação será tida em conta aquando

da construção do modelo, ou seja, será construído um modelo de regressão tendo como variável resposta a presença/ausência de atopia na descendência e como covariáveis os parâmetros imunológicos em unidades percentuais e um modelo de regressão logística tendo exatamente a mesma variável resposta do anterior e como covariáveis os parâmetros imunológicos em valor absoluto.

Para a construção dos modelos foram tidos alguns critérios em consideração, como:

- (a) Avaliação da multicolinearidade através do cálculo do VIF (considera-se que não existe multicolinearidade se o valor do VIF for inferior a 10);
- (b) Seleção das covariáveis a incluir nos modelos através do método de seleção *stepwise*, utilizando o teste de razão de verossimilhanças como critério de inclusão e exclusão de covariáveis;
- (c) Análise de confundimento e interação;
- (d) Análise dos resíduos;
- (e) Análise de observações influentes;
- (f) Avaliação da adequabilidade dos modelos aos dados, através do teste de Hosmer e Lemeshow.

### **4.3.2 Modelo de regressão logística tendo como covariáveis parâmetros imunológicos percentuais (modelo 1)**

Nesta subsecção são apresentados os resultados relativos à construção do modelo de regressão logística tendo como variável resposta a presença/ausência de atopia na descendência e como covariáveis os parâmetros imunológicos células T (CD3+) %, células B de transição % e células B reguladoras CD24<sup>hi</sup>CD38<sup>hi</sup> %.

## Análise univariada

- **Covariável:** linfócitos T (CD3+) %

Considerando como covariável os linfócitos T (CD3+) %, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.6: Resultados do modelo para a presença/ ausência de atopia na descendência (CD3+)%

Covariável	Coefficiente	Erro padrão	Valor-p
constante	-20,6738	9,8572	0,0359
CD3%	0,2603	0,1256	0,0382

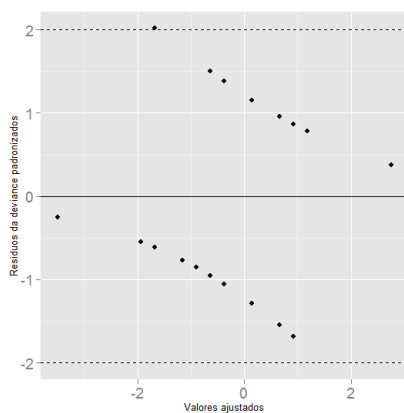


Figura 4.14: Análise dos resíduos (CD3+ %)

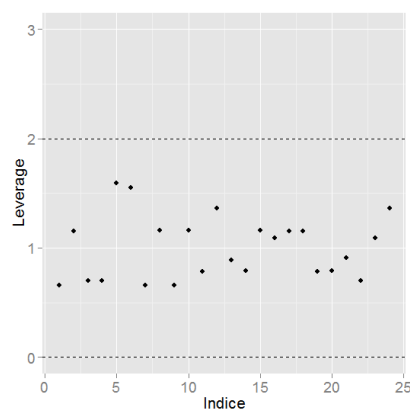


Figura 4.15: Medida de leverage (CD3+ %)

Para a avaliação do modelo obtido foi feita uma análise gráfica dos resíduos da *deviance* padronizados (Figura 4.14) bem como uma análise de observações influentes. Para tal, foram utilizadas as medidas da *leverage* (Figura 4.15) e distância de *Cook* (Figura 4.16). A análise gráfica permite perceber que os pressupostos foram cumpridos.

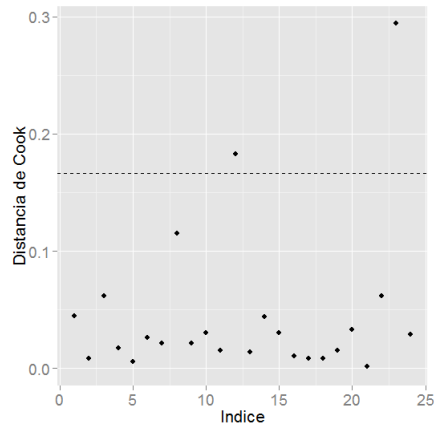


Figura 4.16: Distância de *Cook* (CD3+ %)

De acordo com a aplicação do teste de Hosmer e Lemeshow, não se rejeitou a hipótese de que o modelo se ajusta aos dados ( $p = 0,8926$ ).

Através do cálculo do *odds ratio*, a chance de desenvolvimento de atopia na descendência é 1,297 vezes maior nas mulheres que tenham valores de linfócitos T (CD3+) % superiores a 76,5%.

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.

- **Covariável:** células B reguladoras 24<sup>hi</sup>+38<sup>hi</sup>%

Considerando como covariável as células B reguladoras 24<sup>hi</sup>+38<sup>hi</sup>%, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.7: Resultados do modelo para a presença/ ausência de atopia na descendência (Breg 24+38%)

Covariável	Coefficiente	Erro padrão	Valor-p
constante	-1,5071	0,7937	0,0576
Breg 24+38%	0,4765	0,2761	0,0844

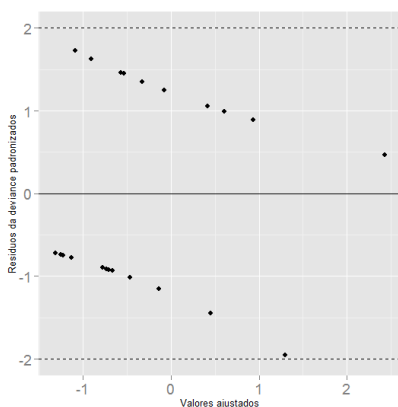


Figura 4.17: Análise dos resíduos (B reguladoras 24<sup>hi</sup>+38<sup>hi</sup>%)

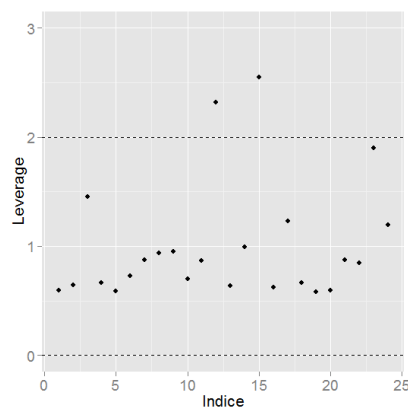


Figura 4.18: Medida de leverage (B reguladoras 24<sup>hi</sup>+38<sup>hi</sup>%)

Através da análise dos resultados obtidos, verifica-se que os linfócitos B reguladores 24<sup>hi</sup>+38<sup>hi</sup>% não se revelaram importantes para explicar a variável resposta nesta análise univariada. Apesar disso, segundo o teste de Hosmer e Lemeshow o modelo adequa-se aos dados ( $p = 0,4618$ ). De acordo com as representações gráficas de qualidade de ajustamento, verifica-se que os pressupostos foram cumpridos.

Através do cálculo do *odds ratio*, a chance de desenvolvimento de atopia na

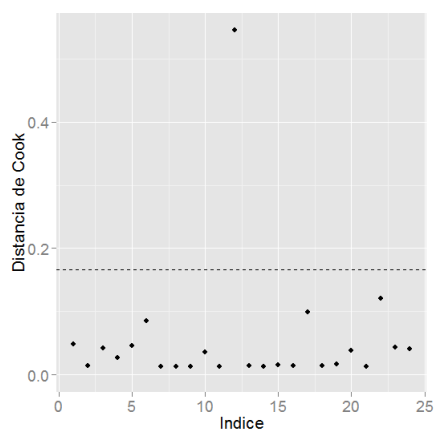


Figura 4.19: Distância de *Cook* (B reguladoras  $24^{hi}+38^{hi}\%$ )

descendência é 1,61 vezes maior nas mulheres que tenham valores de linfócitos B reguladores  $24^{hi}+38^{hi}\%$  superiores a 1,865%.

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.



- **Covariável:** células B de transição %

Considerando como covariável as células B de transição %, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.8: Resultados do modelo para a presença/ ausência de atopia na descendência (B transição %)

Covariável	Coefficiente	Erro padrão	Valor-p
constante	-5,3213	2,1145	0,0119
Btransição %	1,4799	0,5776	0,0104

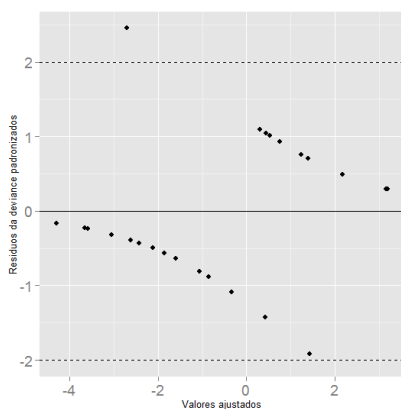


Figura 4.20: Análise dos resíduos (B de transição %)

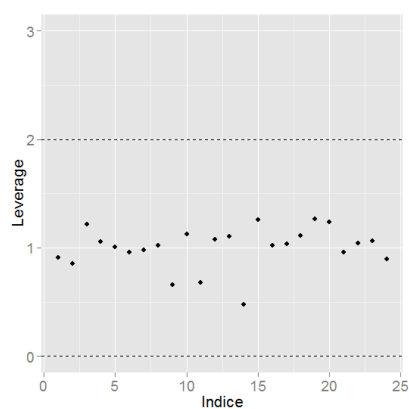


Figura 4.21: Medida de leverage (B de transição %)

Para a avaliação do modelo obtido foi feita uma análise gráfica dos resíduos (Figura 4.14), como uma análise de observações influentes. Para tal, foram utilizadas as medidas da *leverage* (Figura 4.15) e distância de *Cook* (Figura 4.16). A análise gráfica permite perceber que os pressupostos foram cumpridos.

De acordo com a aplicação do teste de Hosmer Lemeshow, não se rejeitou a hipótese de que o modelo se ajusta aos dados ( $p = 0,8198$ ).

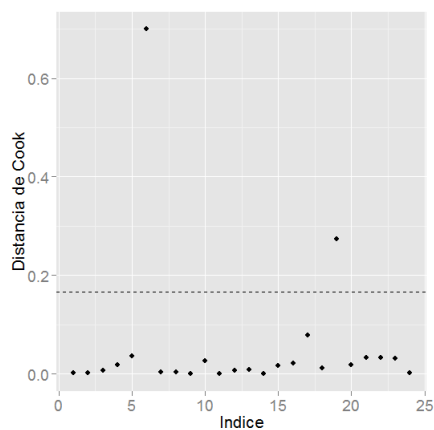


Figura 4.22: Distância de *Cook* (B de transição %)

Através do cálculo do *odds ratio*, a chance de desenvolvimento de atopia na descendência é 4 vezes maior nas mulheres que tenham valores de linfócitos B de transição % superiores a 3,585%.

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.

## Análise multivariada

Nesta subsecção são apresentados os resultados relativos à análise multivariada.

- **Avaliação da multicolinearidade**

A multicolinearidade será avaliada através do cálculo do VIF, que numa situação desejável não deverá ser superior a 10 para cada uma das covariáveis.

Tabela 4.9: Avaliação da multicolinearidade

Covariáveis	VIF
CD3+%	1,036
B transição %	1,391
Breg 24+38+ %	1,352

De acordo com os resultados obtidos para o cálculo do VIF, não se verifica existência de multicolinearidade ( $VIF < 10$ ).

Seguidamente aplicou-se o teste de razão de verossimilhanças para testar a significância de pelo menos um dos  $\beta's$ , ou seja, a significância de pelo menos uma das covariáveis candidatas a entrar no modelo. Para tal, procedeu-se à comparação do modelo nulo com o modelo saturado, concluindo-se que existe evidência para afirmar que pelo menos um dos  $\beta's$  é significativo ( $p < 0,001$ ).

O passo seguinte consiste em aplicar o método de seleção de variáveis *stepwise*, ou seja, partindo do modelo completo, chegar ao modelo que contém apenas as variáveis que são importantes para explicar a variável resposta.

## Seleção das variáveis a entrar no modelo

De acordo com o método *stepwise*, partimos do modelo com todas as variáveis incluídas no modelo, e vai-se removendo em etapas as variáveis que não são importantes no modelo.

Com a aplicação deste método, obteve-se o modelo ótimo com um  $AIC = 18,217$ , contendo as seguintes covariáveis:

Tabela 4.10: Resultados do modelo para a presença/ ausência de atopia na descendência

Covariáveis	Coefficiente	Erro padrão	Valor-p
constante	-40,7559	19,9837	0,0414
CD3%	0,4372	0,2314	0,0588
B transicao %	1,9455	0,8603	0,0237

Repare-se que na análise univariada, a variável linfócitos T (CD3+) % era estatisticamente significativa e agora deixou de o ser. Surge a questão se a variável linfócitos B de transição poderá ser considerada de confundimento ou de interação, ou ambos.

Procedeu-se à análise de associações da variável linfócitos B de transição com a variável linfócitos T (CD3+) %. De acordo com os pontos de corte determinados para cada uma das variáveis aquando da construção das respetivas curvas ROC, testou-se a possível associação através do teste de independência do  $\chi^2$ , concluindo-se que não existe evidência para afirmar que exista associação entre as duas variáveis ( $p = 0,423$ ). Concluímos assim que a variável linfócitos B de transição não é variável de confundimento.

Relativamente à análise de interação, procedeu-se à comparação do modelo com e sem interação.

De acordo com os resultados obtidos, conclui-se que não existe evidência para afirmar que existe interação ( $p = 0,559$ ).

Após verificado que não existe confundimento nem interação e uma vez que um dos critérios de seleção de modelos é a parcimónia, ou seja, selecionar

o modelo que contenha apenas as variáveis que são importantes para explicar a variável resposta, retirou-se a covariável linfócitos T (CD3+)%. Assim, o modelo final ( $M_1$ ) tem um AIC de 23,615, contendo a seguinte covariável:

Tabela 4.11: Resultados do modelo final para a presença/ ausência de atopia na descendência

Covariáveis	Coefficiente	Erro padrão	Valor-p	Odds Ratio
constante	-5,3213	2,1145	0,0119	0,00488
B transicao %	1,4799	0,5776	0,0104	4,393

De acordo com a *deviance*, o modelo corrente ( $M_1$ ) é melhor que o modelo saturado ( $M_s$ ), pelo que consideramos este como sendo o melhor modelo (*Deviance* = 7,3985,  $p$  = 0,02474).

O modelo final é então dado pela seguinte expressão:

$$\text{logit}(p_i) = -5,3213 + 1,4799.Btransicao.$$

### Avaliação e interpretação do modelo

Para a avaliação do modelo obtido foi feita uma análise gráfica dos resíduos (Figura 4.26), bem como uma análise de observações influentes. Para tal, foram utilizadas as medidas da *leverage* (Figura 4.27) e distância de *Cook* (Figura 4.28). A análise gráfica permite perceber que os pressupostos foram cumpridos.

De acordo com a aplicação do teste de Hosmer e Lemeshow, não se rejeitou a hipótese de que o modelo se ajusta aos dados ( $p$  = 0,8198).

Através do cálculo do *odds ratio*, a chance de desenvolvimento de atopia na descendência é 4 vezes maior nas mulheres que tenham valores de células B de transição superiores a 3,585%.

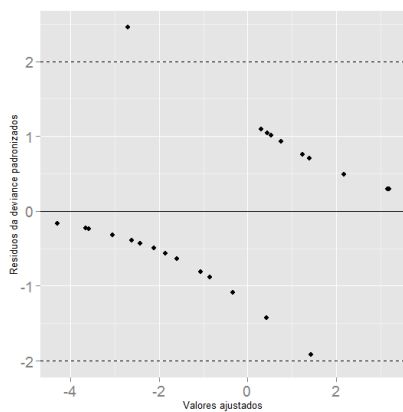


Figura 4.23: Análise dos resíduos (modelo 1)

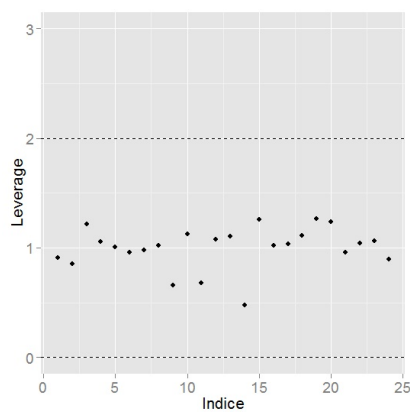


Figura 4.24: Medida de leverage (modelo 1)

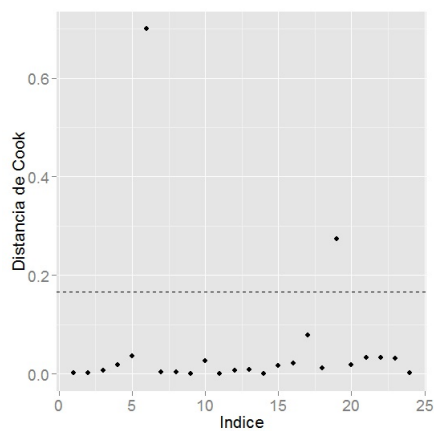


Figura 4.25: Distância de *Cook* (modelo 1)

## Capacidade preditiva do modelo

Para a análise da capacidade preditiva do modelo, procedeu-se à construção da curva ROC, obtendo-se:

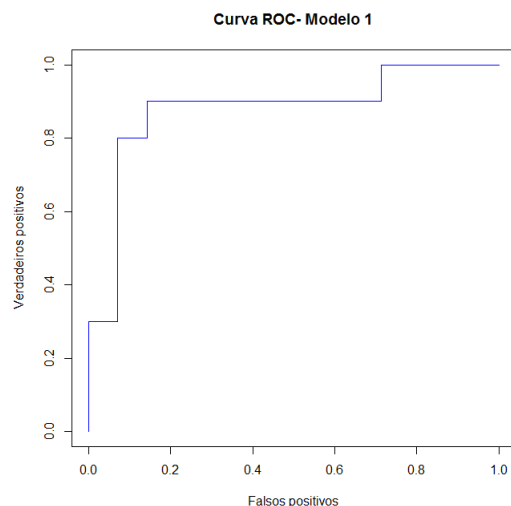


Figura 4.26: Curva ROC do modelo 1

De acordo com a curva ROC obtida, obteve-se uma sensibilidade de 90% e uma especificidade de 85,7% para um ponto de corte de 3,585%. A área abaixo da curva foi de 0,87, o que significa que existe uma discriminação boa por parte deste marcador para manifestações de risco para a atopia.

### 4.3.3 Modelo de regressão logística tendo como covariáveis parâmetros imunológicos em valor absoluto (modelo 2)

Nesta subsecção são apresentados os resultados relativos à construção de um modelo de regressão logística tendo como variável resposta a presença/ausência de atopia na descendência e como covariáveis os parâmetros imunológicos Bm1, B memória não switched e B memória.

## Análise univariada

- **Covariável:** linfócitos Bm1

Considerando como covariável linfócitos Bm1, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.12: Resultados do modelo para a presença/ ausência de atopia na descendência (Bm1)

Covariável	Coefficiente	Erro padrão	Valor-p
constante	1,385	0,914	0,1299
Bm1	-0,084	0,043	0,0545

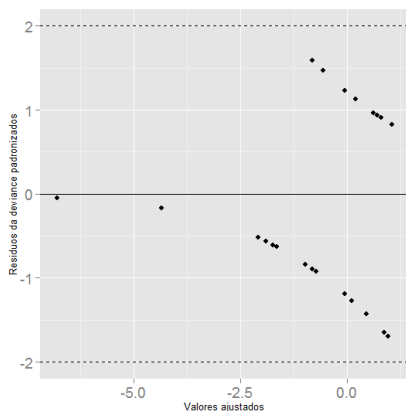


Figura 4.27: Análise dos resíduos (Bm1)

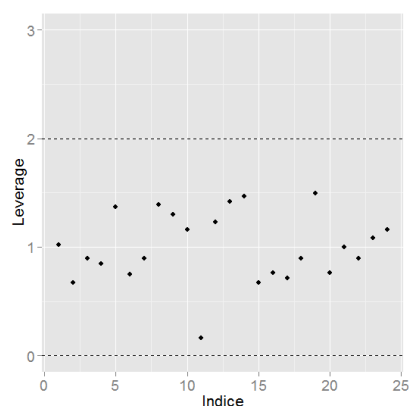


Figura 4.28: Medida de leverage (Bm1)

Através da análise dos resultados obtidos, verifica-se que os linfócitos Bm1 não se revelaram importantes para explicar a variável resposta nesta análise univariada. Apesar disso, segundo o teste de Hosmer e Lemeshow o modelo adequa-se aos dados ( $p = 0,9354$ ). De acordo com as representações gráficas de qualidade de ajustamento, verifica-se que os pressupostos foram



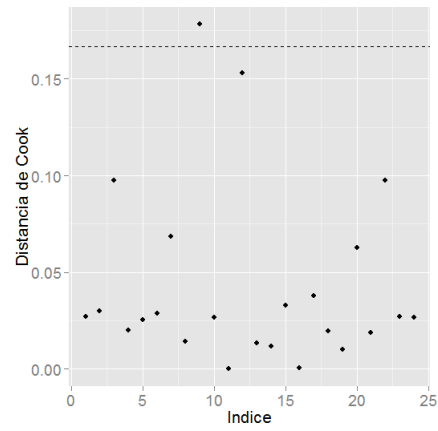


Figura 4.29: Distância de *Cook* (Bm1)

cumpridos.

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.

- **Covariável:** linfócitos B de memória

Considerando como covariável linfócitos B de memória, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.13: Resultados do modelo para a presença/ ausência de atopia na descendência (B memória)

Covariável	Coefficiente	Erro padrão	Valor-p
constante	1,485	1,015	0,1436
Bmemoria	-0,026	0,014	0,0647

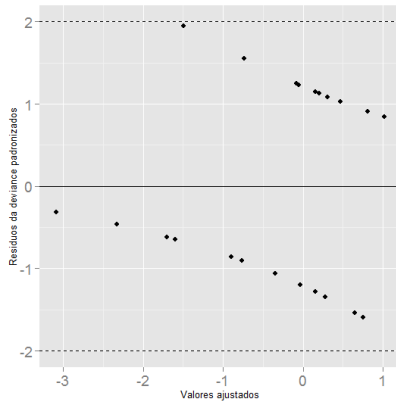


Figura 4.30: Análise dos resíduos (B memória)

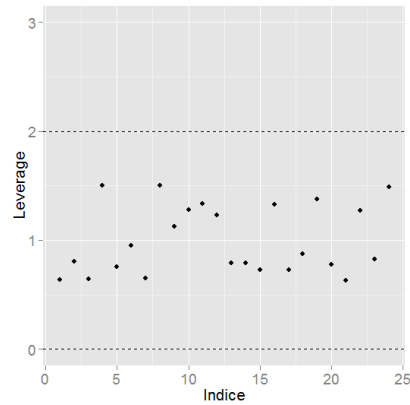


Figura 4.31: Medida de leverage (B memória)

Já para a variável linfócitos B de memória, através da análise dos resultados obtidos, verifica-se que esta variável não se revelou importante para explicar a variável resposta nesta análise univariada. Apesar disso, segundo o teste de Hosmer e Lemeshow o modelo adequa-se aos dados ( $p = 0,2774$ ). De acordo com as representações gráficas de qualidade de ajustamento, verifica-se que os pressupostos foram cumpridos.

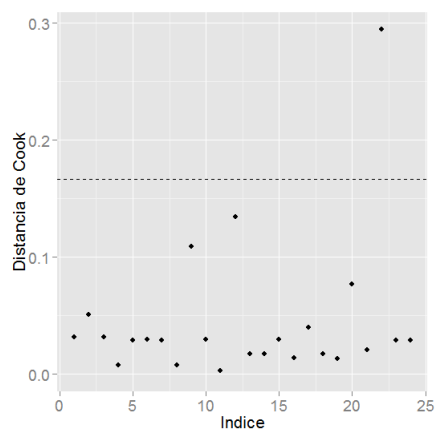


Figura 4.32: Distância de *Cook* (B memória)

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.

- **Covariável:** linfócitos B de memória não *switched*

Considerando como covariável linfócitos B de memória não *switched*, procedeu-se à construção de um modelo de regressão logística, tendo-se obtido os seguintes resultados:

Tabela 4.14: Resultados do modelo para a presença/ ausência de atopia na descendência (B memória não *switched*)

Covariável	Coefficiente	Erro padrão	Valor-p
constante	1,617	1,051	0,1237
B memória não <i>switched</i>	-0,048	0,026	0,0644

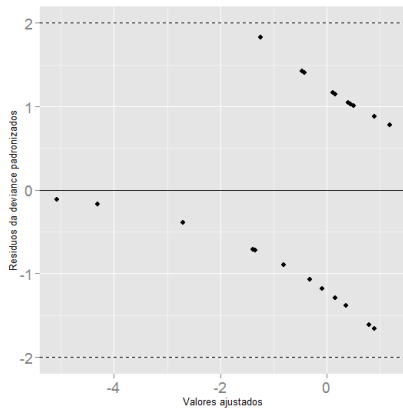


Figura 4.33: Análise dos resíduos (B memória não *switched*)

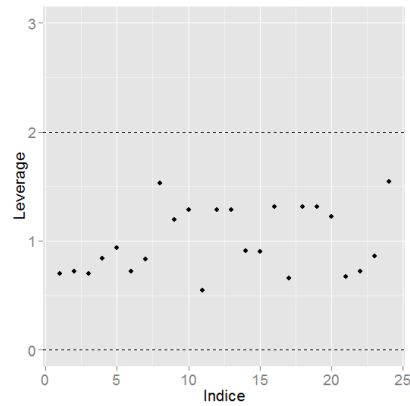


Figura 4.34: Medida de leverage (B memória não *switched*)

Considerando os linfócitos B de memória não *switched*, estes não se revelaram importantes para explicar a variável resposta nesta análise univariada. Apesar disso, segundo o teste de Hosmer e Lemeshow o modelo adequa-se aos dados ( $p = 0,9625$ ). De acordo com as representações gráficas de qualidade de ajustamento, verifica-se que os pressupostos foram cumpridos.

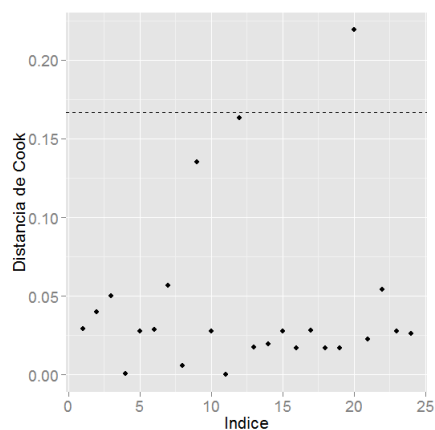


Figura 4.35: Distância de *Cook* (B memória não *switched*)

Relativamente à capacidade preditiva do modelo, dado que estamos na situação de modelo de regressão logística simples, a curva ROC obtida coincide com a apresentada na secção 4.2.

## Análise multivariada

- **Avaliação da multicolinearidade**

A multicolinearidade será avaliada através do cálculo do VIF, que numa situação desejável não deverá ser superior a 10 para cada uma das covariáveis.

Tabela 4.15: Avaliação da multicolinearidade

Covariáveis	VIF
B memória	8,98
B memória não <i>switched</i>	7,92
Bm1	2,97

De acordo com os resultados obtidos para o cálculo do VIF, não se verifica existência de multicolinearidade ( $VIF < 10$ ).

Foi aplicado o teste de razão de verossimilhanças para testar a significância de pelo menos um dos  $\beta$ 's, concluindo-se que nenhuma das covariáveis é candidata a entrar no modelo, dado que não existem diferenças estatisticamente significativas entre o modelo nulo e o modelo saturado ( $p = 0,06223$ ).

Assim sendo, apenas consideramos o modelo de regressão logística para parâmetros imunológicos percentuais (modelo 1).

# Capítulo 5

## Discussão

O impacto crescente das doenças atópicas tem despertado o interesse da comunidade científica para uma melhor compreensão dos seus mecanismos no sentido de se estabelecerem principais estratégias preventivas.

Reconhecendo o impacto da história familiar para o desenvolvimento de doenças atópicas em crianças, impõe-se uma melhor compreensão dos mecanismos imunológicos subjacentes, concretamente no período fetal já que a história materna de atopia parece ser um importante fator de risco.

De acordo com os resultados obtidos com este estudo, envolvendo 134 mulheres: 24 grávidas atópicas, 43 grávidas não atópicas, 32 não grávidas atópicas e 35 não grávidas não atópicas, verificou-se a existência de diferenças estatisticamente significativas relativamente à variável idade, sendo em média, as mulheres não grávidas mais velhas relativamente às mulheres grávidas.

No que se refere à análise das pressões arteriais, destacam-se as mulheres não grávidas saudáveis com valores mais elevados, em média, de pressão arterial sistólica e diastólica. Contudo, os valores revelaram-se dentro da normalidade de acordo com a classificação em vigor.

Relativamente ao IMC, como seria de esperar, este encontra-se mais elevado nas mulheres grávidas do que nas mulheres não grávidas.

Dado o principal objetivo deste projeto (identificação de marcadores de risco para a descendência de mulheres grávidas atópicas), a caracterização do perfil imunológico materno no 3.º trimestre de gestação, no dia do parto

e após o puerpério revelou-se essencial.

De acordo com as comparações efetuadas, ao nível de mulheres grávidas e não grávidas saudáveis, o compartimento B constituído por linfócitos B mereceu especial atenção dadas as alterações observadas ao nível das células B reguladoras, B de transição e B memória, tanto em valor absoluto como percentual. Há portanto uma influência da gravidez na distribuição destas populações celulares.

Relativamente à presença de atopia, em mulheres grávidas no 3.º trimestre de gestação, volta a observar-se uma presença destacada de células B, nomeadamente células B reguladoras.

Repare-se que as variações observadas desde o 3.º trimestre de gestação têm vindo a diluir-se ao longo do parto e após o puerpério, mantendo-se contudo ainda neste último momento.

Dadas as variações observadas em parâmetros imunológicos maternos avaliados tanto ao nível da gravidez como da presença de atopia, impôs-se estudar como é que estas variações, nomeadamente ao nível do compartimento B poderão influenciar a presença de atopia na descendência.

Foram observadas 24 crianças, filhas de mães atópicas, registando-se uma prevalência de atopia em 41%. Destas, 80% eram do sexo masculino. Tal vai de encontro com a literatura existente, referindo que a atopia é mais comum nos primeiros meses de vida nos rapazes.

Procedeu-se à análise de diferenças entre descendência com manifestações atópicas e descendência sem manifestações atópicas relativamente aos parâmetros maternos avaliados no 3.º trimestre de gestação, observando-se alterações ao nível do compartimento B, nomeadamente nas células B de transição, B de memória e B reguladoras.

É no compartimento B, mais negligenciado na literatura nesta área que se documentaram variações significativas entre grávidas com e sem descendência atópica. Contudo, a contagem das células B tende a ser mais baixa em mulheres com descendência atópica, em relação às mulheres sem descendência atópica, revelando-se assim mais imaturo em mulheres grávidas com descendência atópica.

Através da construção de curvas ROC para determinação dos pontos de corte que permitissem identificar marcadores de risco, constatou-se que as



células B de transição constituem um importante fator de risco para o desenvolvimento de atopia, bem como as células B de memória, linfócitos B (CD19+) %, linfócitos T (CD3+) % , Bm1 e B reguladoras. Foram mesmo as células B de transição o melhor marcador identificado, com elevados valores de sensibilidade e especificidade.

Para averiguar de entre os marcadores identificados como sendo fatores de risco para o desenvolvimento de atopia, qual ou quais seriam importantes para prever corretamente o desenvolvimento de doenças atópicas, procedeu-se à construção de modelos de regressão logística. Para tal, foram construídos dois modelos, dadas as diferentes unidades de recolha dos parâmetros imunológicos maternos avaliados. Assim, considerou-se um modelo para parâmetros imunológicos percentuais e outro para parâmetros imunológicos em valor absoluto.

Da análise resultante do primeiro caso, o modelo mais parcimonioso foi considerado aquele que tinha como covariável as células B de transição. Esta covariável apresentou um *odds ratio* de 4,393, significando que a chance de desenvolvimento de doenças atópicas é 4 vezes superior em mulheres que tenham valores elevados de células B de transição ( $> 3,685$  %). De referir que segundo o método *stepwise*, o melhor modelo continha as covariáveis células B de transição e os linfócitos T (CD3+)%. Contudo, como esta última não se revelou significativa no modelo, foi retirada de modo a obter-se o modelo mais parcimonioso possível.

Relativamente à capacidade preditiva do modelo, este demonstrou boas capacidades discriminatórias, com uma sensibilidade de 90% e especificidade de 85,7%. As representações gráficas apresentadas, nomeadamente a análise de resíduos, medida de *leverage* e distância de *Cook*, permitiram averiguar que os pressupostos do modelo foram cumpridos.

No que concerne ao modelo para parâmetros imunológicos em valor absoluto, nenhuma das variáveis se revelou candidata a entrar no modelo, pelo que não se procedeu à sua construção. De notar que para estas variáveis (valor absoluto), em termos medianos, as contagens apresentavam-se mais elevadas em mulheres com descendência não atópica.

# Capítulo 6

## Conclusões

Com este estudo, concluiu-se que a atopia na descendência de mulheres grávidas atópicas pode estar relacionada com alterações nas populações celulares B de transição, em que valores destas células superiores a 3,585% levam a que a descendência de mulheres grávidas atópicas fique mais suscetível ao desenvolvimento de atopia.

De acordo com os resultados finais obtidos, a avaliação de alterações em distintos compartimentos  $T$  e  $B$  circulantes em mulheres atópicas durante a gravidez, permitiu reconhecer o seu impacto no desenvolvimento de manifestações de risco para a atopia.

Pode deste modo vir a tornar-se possível estabelecer estratégias de identificação de marcadores de risco para o desenvolvimento de atopia na descendência de mulheres atópicas.

Este trabalho tornou-se assim pioneiro no que à avaliação do compartimento de células B diz respeito, dada a escassez de informação existente até ao momento na literatura, sendo o primeiro trabalho a avaliar as subpopulações de células B reguladoras em mulheres grávidas atópicas e na sua descendência.

## 6.1 Divulgação de resultados

A análise exploratória destes dados, bem como a identificação de marcadores de risco foi apresentada no XXII Congresso da Sociedade Portuguesa de Estatística em Olhão, Algarve, que se realizou nos dias 7, 8, 9 e 10 de outubro de 2015.

De referir também que grande parte da análise exploratória dos dados e identificação de marcadores de risco serviram também para o trabalho de doutoramento da Professora Doutora Catarina Martins da Faculdade de Ciências Médicas da Universidade Nova de Lisboa, cuja exposição ocorreu no dia 11 de dezembro de 2015 na sala de atos da Faculdade de Ciências Médicas da Universidade Nova de Lisboa.

Pretende-se apresentar os presentes resultados na Revista Portuguesa de Imunoalergologia (RPIA).

# Bibliografía

- Amu, S; Saunders, SP; Kronenberg, M; Mangan, NE; *et al.* Regulatory B cells prevent and reverse allergic airway inflammation via FoxP3-positive T regulatory cells in a murine model. *J Allergy Clin Immunol.* 125(5): p. 1114-1124 e8, 2010.
- Bantz, SK; Zhu, Z; Zheng, T. The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. *Journal of clinical and cellular immunology.* 5(2), 2014.
- Braza, F; Chesne, J; Castagnet, S; Magnan, A; *et al.* Regulatory functions of B cells in allergic diseases. *Allergy.* 69(11): p. 1454-63, 2014.
- Blair, PA; Noreña, LY; Flores-Borja, F; Rawlings, DJ; *et al.* CD19(+) CD24(hi)CD38(hi) B cells exhibit regulatory capacity in healthy individuals but are functionally impaired in systemic Lupus Erythematosus patients. *Immunity.* 32(1): p. 129-40, 2010.
- Bohacs, A; Pallinger, E; Tamasi, L; Jr, JR; *et al.* Surface markers of lymphocyte activation in pregnant asthmatics. *Inflamm Res.* 59(1):63-70, 2010.
- Christensen, R. *Log-Linear Models and Logistic Regression.* Springer, 1997.
- Dupont, W. *Statistical modeling for biomedical researchers: A simple introduction to the analysis of complex data.* Cambridge University Press, Cambridge, 2002.
- Fontenot, JD; Gavin, MA; Rudensky, AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol.* 4(4):330-6, 2003.

- Frank, E; Harrel, Jr. *Regression Modeling Strategies*. Springer, 2001.
- Griffin, DO; Rothstein, TL. Human "orchestrator" CD11b(+) B1 cells spontaneously secrete interleukin-10 and regulate T-cell activity. *Mol Med*. 18: p. 1003-8, 2012.
- Halonen, M; Lohman, IC; Stern, DA; Spangenberg, A; *et al.* Th1/Th2 patterns and balance in cytokine production in the parents and infants of a large birth cohort. *J Immunol*. 182(5):3285-93, 2009.
- Hawrylowicz, C; Ryanna, K. Asthma and allergy: the early beginnings. *Nat Med*. 16(3):274-5, 2010.
- Hori, S; Nomura, T; Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science*. 299(5609):1057-61, 2003.
- Hosmer, DW; Lemeshow, S. *Applied Logistic Regression*. A Wiley-Interscience Publication, Second Edition, 2000.
- Iwata, Y; Matsushita, T; Horikawa, M; DiLillo, DJ; *et al.*, Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells. *Blood*. 117(2): p. 530-41, 2011.
- Jin, G; Hamaguchi, Y; Matsushita, T; Hasegawa, M; *et al.*, B-cell linker protein expression contributes to controlling allergic and autoimmune diseases by mediating IL-10 production in regulatory B cells. *J Allergy Clin Immunol*. 131(6): p. 1674-82, 2013.
- Johansson, SG; Bieber, T; Dahl, R; Friedmann, PS; *et al.* Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol*. 113(5):832-6, 2004.
- Maroco, J. *Análise Estatística- com utilização do SPSS*. Edições Sílabo, 3ª Edição, 2007.
- Martins, Ana Catarina Silva Gregório da Costa. Perfil Imunológico de Grávidas Atópicas versus Grávidas Saudáveis. Tese de Doutoramento. Faculdade de Ciências Médicas da Universidade Nova de Lisboa, 2015.

- Makhseed, M; Raghupathy, R; Azizieh, F; Omu, A; *et al.* Th1 and Th2 cytokine profiles in recurrent aborters with successful pregnancy and with subsequent abortions. *Hum Reprod.* 16(10):2219-26, 2001.
- Murteira, B; Antunes, M. *Probabilidades e Estatística.* Escolar Editora, Volume II, 2012.
- Muzzio, DO; Soldati, R; Ehrhardt, J; Utpatel, K; *et al.* B cell development undergoes profound modifications and adaptations during pregnancy in mice. *Biol Reprod.* 91(5):115, 2014.
- Noh, G; Lee, JH. Regulatory B cells and allergic diseases. *Allergy Asthma Immunol Res.* 3(3): p. 168-77, 2011.
- Pestana, D; Velosa, S. *Introdução à Probabilidade e à Estatística.* Fundação Calouste Gulbenkian, Volume I, 4ª Edição, 2010.
- Piccinni, MP; Lombardelli, L; Logiodice, F; Kullolli, O; *et al.* T helper cell mediated-tolerance towards fetal allograft in successful pregnancy. *Clinical and molecular allergy : CMA.* 13(1):9, 2015.
- Prussin, C; Yin, Y; Upadhyaya, B. Th2 heterogeneity: Does function follow form? *The Journal of Allergy and Clinical Immunology.* 126(6):1094-8, 2010.
- Raghupathy, R; Makhseed, M; Azizieh, F; Omu, A; *et al.* Cytokine production by maternal lymphocytes during normal human pregnancy and in unexplained recurrent spontaneous abortion. *Hum Reprod.* 15(3):713-8, 2000.
- Sanz, I. Rationale for B cell targeting in SLE. *Semin Immunopathol.* 36(3): p. 365-75, 2014.
- Shah, A. The pathologic and clinical intersection of atopic and autoimmune disease. *Current allergy and asthma reports.* 12(6):520-9, 2012.
- Sykes, L; MacIntyre, DA; Yap, XJ; Teoh, TG; *et al.* The Th1:th2 dichotomy of pregnancy and preterm labour. *Mediators of inflammation.* 2012:967629, 2012.

- Svensson-Arvelund, J; Ernerudh, J; Buse, E; Cline, JM; *et al.* The placenta in toxicology. Part II: Systemic and local immune adaptations in pregnancy. *Toxicologic pathology*. 42(2):327-38, 2014.
- Toldi, G; Molvarec, A; Stenczer, B; Müller, V; *et al.* Peripheral T(h)1/T(h)2/T(h)17/ regulatory T-cell balance in asthmatic pregnancy. *Int Immunol*. 23(11):669-77, 2011.
- Turkman, M; Silva, G. *Modelos Lineares Generalizados - da teoria à prática*. Lisboa, 2000.

# Apêndice A

## Código R - análise de regressão

```
data2=read.table("Base_R_b.txt",head=T)
names(data2)
summary(data2)

####Análise univariada

###Covariável CD3+/Breg2438/Btransicao/Bm1/Bmemoria/Bmemoria n switched

modelo<-glm(data2$Problemas~data2$CD3p, family=binomial, data=data2)
summary(modelo)

library(car)

###Resíduos

res=rstandard(modelo,type="deviance",fitted=FALSE)
plot(res)

library(ggplot2)

ggplot(modelo,aes(.fitted,res))+geom_hline(yintercept=0)+
geom_point(size=2)+
geom_hline(y=c(-2,2),lty=2,col="black")+
scale_y_continuous(name="Resíduos desviancia padronizados")+
scale_x_continuous(name="Valores ajustados")+
theme(axis.title.x=element_text(size=10),
text=element_text(size=17))+
theme(axis.title.y=element_text(size=10))
```



```

#Leverage

lev=hat(model.matrix(modelo))
data2[lev >0.2,]

hat<-hatvalues(modelo) #matriz hat
p <- dim(model.matrix(modelo))[[2]] #numero de parametros
n <- dim(model.matrix(modelo))[[1]] #numero de observacoes
ggplot(modelo, aes(1:length(hat/(p/n)), hat/(p/n))) +
  geom_point(size=2) +
  geom_hline(linetype=2, size=.2) +
  scale_y_continuous(name="Leverage", limits=c(0,3)) +
  scale_x_continuous(name="Indice") +
  geom_hline(y=2, lty=2, col="black") +
  theme(axis.title.x=element_text(size=16),
        text=element_text(size=17)) +
  theme(axis.title.y=element_text(size=16))

#Distancia de COOK

ggplot(modelo, aes(1:length(.cooks), .cooks, ymin=0, ymax=.cooks)) +
  geom_point(size=2) +
  scale_y_continuous(name="Distancia de Cook") +
  scale_x_continuous(name="Indice") +
  geom_hline(y=4/n, lty=2, col="black") +
  theme(axis.title.x=element_text(size=16),
        text=element_text(size=17)) +
  theme(axis.title.y=element_text(size=16))

```

```

library(ROCR)
library(gplots)
library(reshape2)

library(MKmisc)

f=fitted(modelo)
summary(f)
X=cbind(data2$Problemas)
HLgof.test(fit=fitted(modelo),data2$Problemas, ngr=10,X,verbose=FALSE)

#Curva ROC

predA=prediction(f, data2$Problemas)
perfA=performance(predA,"tpr", "fpr")
plot(perfA,col="blue", xlab = "Falsos positivos",ylab = "Verdadeiros positivos",
,main = "Curva ROC- Modelo para CD3+ %")

#Área abaixo da curva

library(Hmisc)

y=data2$Problemas
somers2(f, y, weights=NULL, normwt=FALSE, na.rm=TRUE)

par(mfrow=c(2,2))
plot(f,wich=list(1,2,3,4)) ##gráfico dos residuos

#ODDS RATIO para os coeficientes do modelo

OR=exp(modelo$coefficients)
OR
IC_OR=exp(confint(modelo))
IC_OR

#Calculo da sensibilidade e especificidade

fit=modelo$fitted
cut=0.5
cfit=rep(0,nrow(data2))
cfit[fit=0.5]=0
cfit[fit>=0.5]=1
(tab=table(cfit,data2$Problemas,dnn=c("class","obs")))
sens=100*tab[2,2]/sum(tab[,2])
sens
esp=100*tab[1,1]/sum(tab[,1])
esp

```

```

#####Análise multivariada

#Cálculo do VIF

library(car)

vif(glm(data2$Problemas ~.,data=data2))

boxplot(data2$Btransicao ~ data2$Problemas, data2, boxwex = 0.35,
,col="yellow",main="",xlab = "Presença vs Ausência",ylab = "Btransicao")
boxplot(data2$Breg2438 ~ data2$Problemas, data2, boxwex = 0.35,
,col="yellow",main="",xlab = "Presença vs Ausência",ylab = "Breg")
boxplot(data2$CD3p ~ data2$Problemas, data2, boxwex = 0.35,
,col="yellow",main="",xlab = "Presença vs Ausência",ylab = "CD3")

g1<-glm(data2$Problemas ~.,binomial,data2)
summary(g1)

# teste da razão de verossimilhanças para testar a significância
# de pelo menos um dos betas.
g2<-glm(data2$Problemas~1,binomial,data2)
(d2<-deviance(g2))
(d1<-deviance(g1))
(x2<-(d2-d1))
(pval<-pchisq(x2,ncol(model.matrix(g1))-1,lower.tail=FALSE))
anova(g2,g1,test="Chisq")

```

```

# stepwise backward
g<-glm(data2$Problemas~.,binomial,data2)
summary(g)
library(MASS)
gback<-stepAIC(g,direction="both")
summary(gback)

#Melhor modelo

mmm<-glm(data2$Problemas~data2$Btransicao+data2$CD3p, binomial, data2)
summary(mmm)

###Sem CD3

mm<-glm(data2$Problemas~data2$Btransicao, binomial, data2)

###Residuos
res=rstandard(mm,type="deviance",fitted=FALSE)
plot(res)

library(ggplot2)

ggplot(mm,aes(.fitted,res))+geom_hline(yintercept=0)+
geom_point(size=2)+
geom_hline(y=c(-2,2),lty=2,col="black")+
scale_y_continuous(name="Residuos desviância padronizados")+
scale_x_continuous(name="Valores ajustados")+
theme(axis.title.x=element_text(size=10),
text=element_text(size=17))+
theme(axis.title.y=element_text(size=10))

```

```

###Hosmer e Lemeshow
library(MKmisc)

f=fitted(mm)
X=cbind(data2$Problemas)
HLgof.test(fit=fitted(mm),data2$Problemas, ngr=10,X,verbose=FALSE)

###Interação
int<-glm(data2$Problemas~data2$CD3p+data2$Btransicao, family=binomial,
,data=data2)
int2<-glm(data2$Problemas~data2$CD3p*data2$Btransicao, family=binomial,
,data=data2)

anova(int,int2,test="Chisq")

library(ROCR)
library(gplots)
library(reshape2)

#Curva ROC

predA=prediction(f, data2$Problemas)
perfA=performance(predA,"tpr", "fpr")
plot(perfA,col="blue", xlab = "Falsos positivos",ylab = "Verdadeiros positivos",
,main = "Curva ROC- Modelo 1")

#Área abaixo da curva

library(Hmisc)

y=data2$Problemas
somers2(f, y, weights=NULL, normwt=FALSE, na.rm=TRUE)

```

```

par(mfrow=c(2,2))
plot(mm,wich=list(1,2,3,4)) ##gráfico dos residuos

#ODDS RATIO para os coeficientes do modelo

OR=exp(mm$coefficients)
OR
IC_OR=exp(confint(mm))
IC_OR

#Calculo da sensibilidade e especificidade

fit=mm$fitted
cut=0.5
cfit=rep(0,nrow(data2))
cfit[fit=0.5]=0
cfit[fit>=0.5]=1
(tab=table(cfit,data2$Problemas,dnn=c("class","obs")))
sens=100*tab[2,2]/sum(tab[,2])
sens
esp=100*tab[1,1]/sum(tab[,1])
esp

#Leverage

lev=hat(model.matrix(mm))
data2[lev >0.2,]

hat<-hatvalues(mm) #matriz hat
p <- dim(model.matrix(mm))[[2]] #numero de parametros
n <- dim(model.matrix(mm))[[1]] #numero de observacoes
ggplot(mm,aes(1:length(hat/(p/n)),hat/(p/n)))+
geom_point(size=2)+
geom_hline(linetype=2,size=.2)+
scale_y_continuous(name="Leverage", limits=c(0,3))+
scale_x_continuous(name="Indice")+
geom_hline(y=2,lty=2,col="black")+
theme(axis.title.x=element_text(size=16),
text=element_text(size=17))+
theme(axis.title.y=element_text(size=16))

#Distancia de COOK

ggplot(mm,aes(1:length(.cooks), .cooks,ymin=0,ymax=.cooks))+
geom_point(size=2)+
scale_y_continuous(name="Distancia de Cook")+
scale_x_continuous(name="Indice")+
geom_hline(y=4/n,lty=2,col="black")+
theme(axis.title.x=element_text(size=16),
text=element_text(size=17))+
theme(axis.title.y=element_text(size=16))

```

## Apêndice B

### Parâmetros imunológicos analisados no 3.<sup>o</sup> trimestre de gestação

Na tabela seguinte encontram-se os parâmetros analisados ao longo deste trabalho.

Tabela B.1: Parâmetros imunológicos analisados

<b>Parâmetros</b>	
Linfócitos	BM1 %
CD3+ %	BM1 abs
CD3+ abs	BM2 %
CD19+ %	BM2' %
CD19+ abs	BM2' abs
CD16/56 %	BM3+BM4 %
CD16/56 abs	BM3+BM4 abs
CD4+ %	BM5 %
CD4+ abs	BM5 abs
CD4+ DR+ %	eBM5 abs
CD4+ DR+ abs	DN %
CD8+ abs	DN abs
CD3+CTLA4+ %	B transição %
CD3+DR+ abs	B transição abs
CD3+CD25+ abs	Plasmablastos %
CD4+CTLA+ %	Plasmablastos abs
CD4+CD25+ abs	Bnaive abs
CD8+CTLA+ %	B memoria abs
CD8+CD25+4+ abs	B memoria n switched abs
Neutrófilos %	B memoria switched abs
Neutrófilos abs	Breg 24+27+ %
Basófilos %	Breg 24+27+ abs
Basófilos abs	Breg24+38+ %
Eosinófilos abs	Breg 24+38+ abs
Eosinófilos %	CD3+CTLA4+H %
Monócitos abs	CD3+CTLA4+H abs
CD4+CD127+ %	CD4+CTLA4+H abs
CD4+CD127+ abs	CD8+CTLA4+H abs
4+25++127- abs	CD4+GEO MFI
CD4+CD25+Foxp3+ %	Bm1 memoria +
CD4+CD25+Foxp3+ abs	Bm1 memoria -
CD25++CD4+ %	Bm1 naive
CD25++CD4+ abs	Bm5 CD27+
3+8-IFNG+ abs	Bm5 CD27-
3+8-IL17+ abs	eBm5 CD27+
3+8-IFN Total abs	eBm5 CD27-



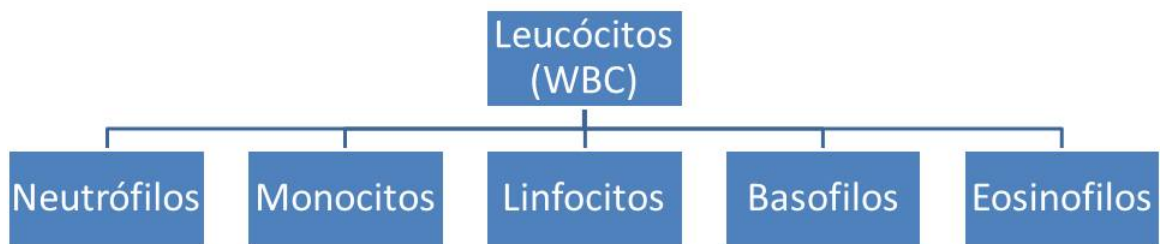
Tabela B.2: Parâmetros imunológicos analisados (continuação)

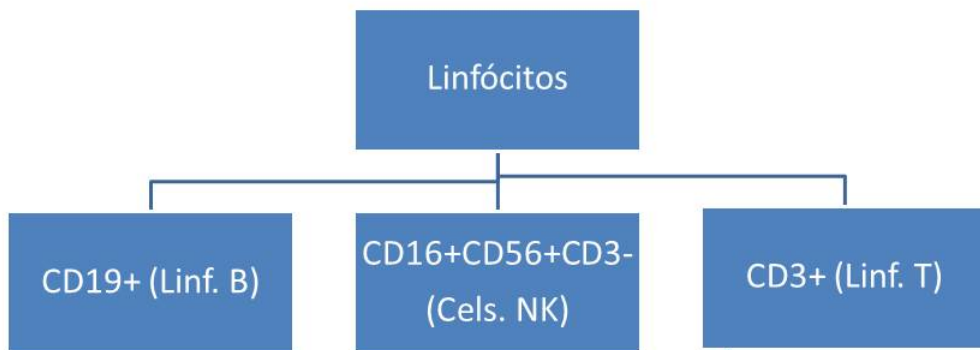
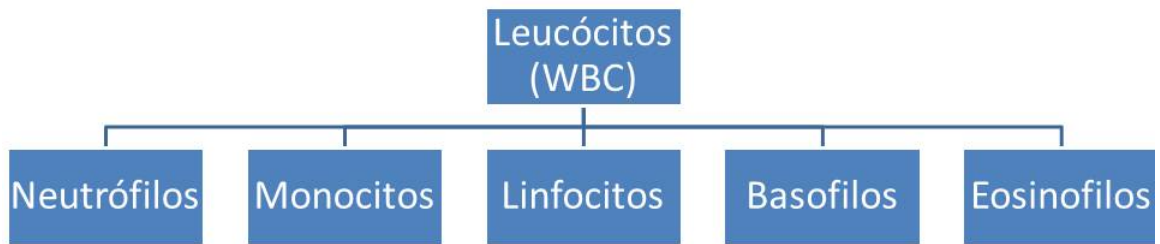
<b>Parâmetros</b>	
3+8-IL17+Total abs	Bm3+4 CD27+
3+8+IFNG+abs	Bm3+4 CD27-
3+8+IFN Total abs	3+IL10+ %
IFN1	3+IL10+ abs
IFN2	3+8-IL10+ %
3+8-IL10+ abs	MFI geo CD4+
3+8+IL10+ %	3+8+IL10+ abs
19+IL10+ %	19+IL10+ abs
FOX GeoMean	

## Apêndice C

### Classificação celular

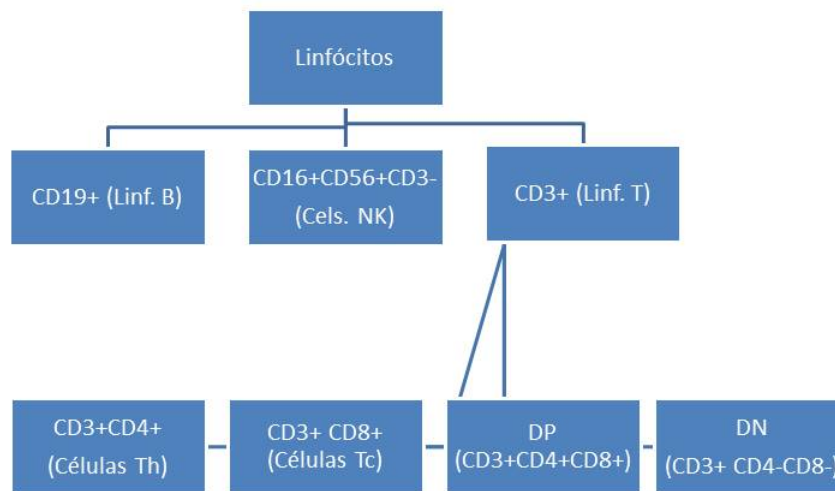
Para que o leitor possa compreender melhor as populações celulares estudadas ao longo deste trabalho, apresenta-se de seguida uma classificação/divisão das células estudadas.





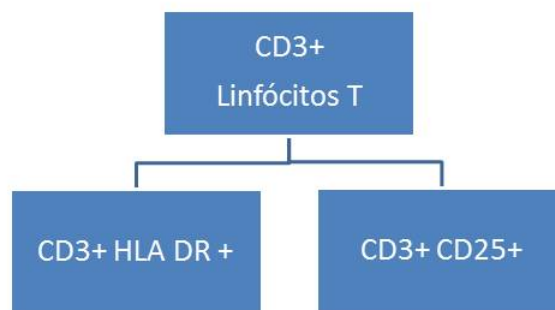
Apesar de serem linfócitos T (CD3+) – os valores % desta ultima população são relativas ao total de linfócitos



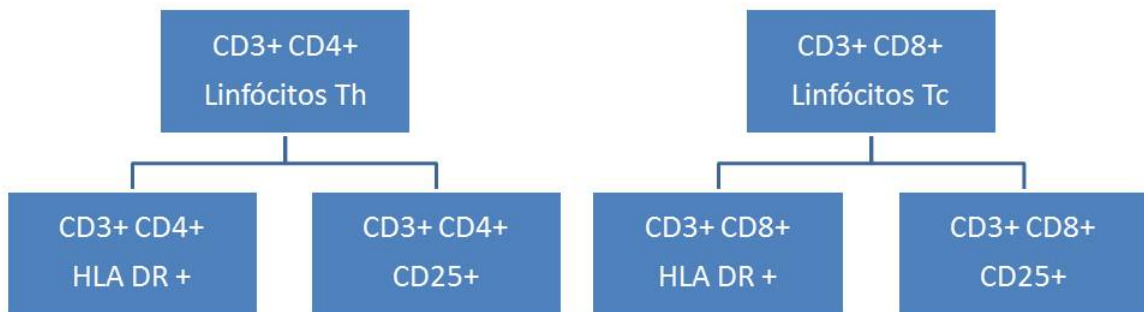


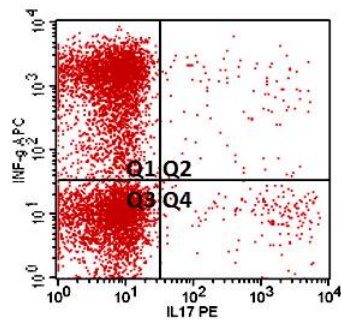
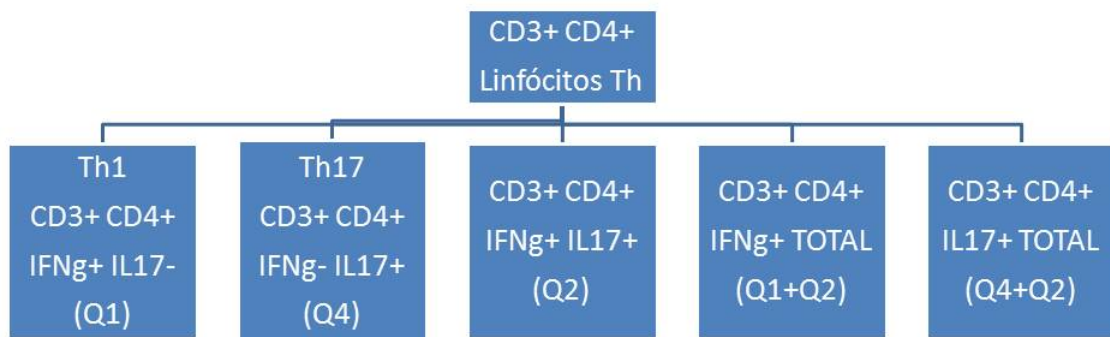
Apesar de serem linfócitos T (CD3+)

– os valores % destas ultimas 4 populações são relativas ao total de linfócitos



Subpopulações independentes dentro dos linfócitos T/Th/Tc,  
sem relação entre si, do ponto de vista da proporção  
**HLA DR+ ≠ CD25+**





CD3+  
Linfócitos T



CD3+ IL10+

CD4+  
Linfócitos Th



CD4+ IL10+

CD8+  
Linfócitos Tc



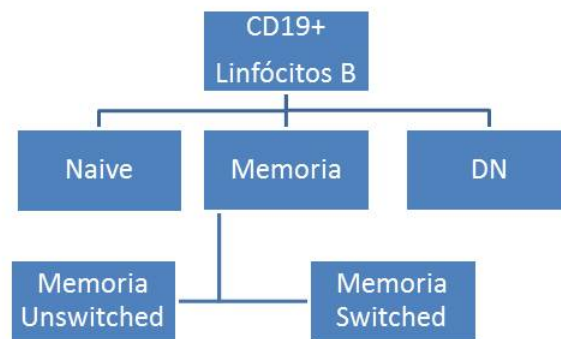
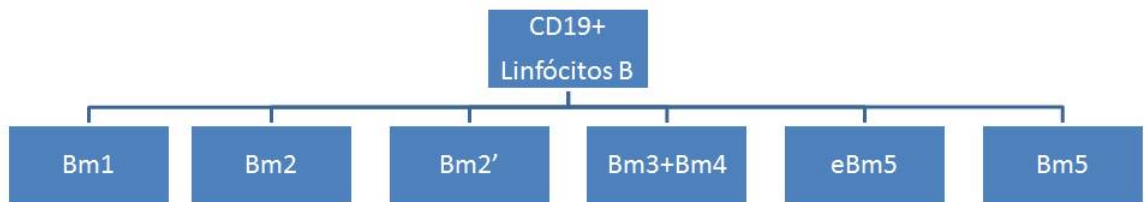
CD8+ IL10+

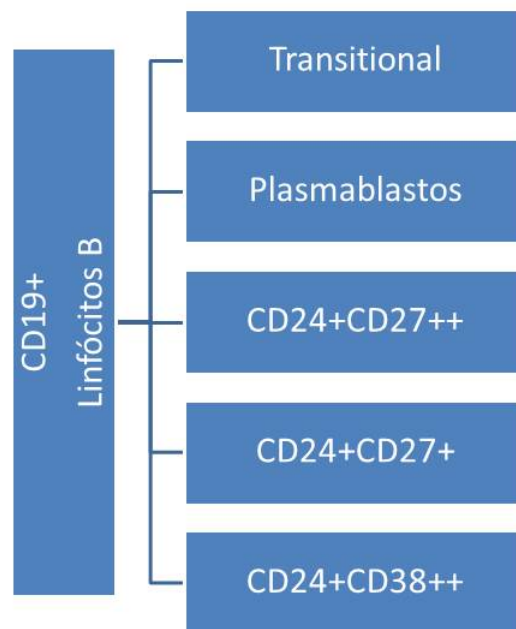
CD19+  
Linfócitos B



CD19+ IL10+







Subpopulações independentes dentro dos linfócitos B, sem relação entre si, do ponto de vista da proporção