

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Modelo Estatístico de Avaliação dos Cuidados de
Saúde - Excelência Clínica**

Relatório de estágio - versão pública

Bruno da Silva Marques

Mestrado em Bioestatística

2013

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Modelo Estatístico de Avaliação dos Cuidados de
Saúde - Excelência Clínica**

Relatório de estágio - versão pública

Bruno da Silva Marques

Estágio orientado pela Professora Doutora Maria Antónia Amaral Turkman (FCUL) e pela Engenheira Inês Sá Dantas (Siemens, S.A.) para a obtenção do grau de Mestre em Bioestatística

2013

O presente relatório de estágio contém informação estritamente confidencial, pelo que, não pode ser copiado, transmitido ou divulgado, na sua parte ou na totalidade, sem o expreso consentimento por escrito do autor e da Siemens Sector Healthcare.

Para os meus pais, José e Marília.

Agradecimentos

Este trabalho não poderia ser realizado sem o apoio de algumas pessoas, às quais quero expressar o meu sincero agradecimento:

Aos Engenheiros Jorge Sequeira, Filipe Janela e Inês Sá Dantas pela oportunidade de realizar o estágio no sector de Healthcare da Siemens, S.A., permitindo assim concluir a componente não curricular do Mestrado em Bioestatística.

À Doutora Filipa Matos Baptista e à Engenheira Filipa Costa por todo o apoio, disponibilidade e profissionalismo que tornaram possíveis a realização e conclusão deste projecto.

A todos os elementos da equipa pela disponibilidade, pela forma como me acolheram na equipa e pelos bons momentos passados.

À professora Doutora Antónia Turkman pelo apoio, disponibilidade e orientação concedida durante todo o projecto, presente desde a licenciatura.

Em especial, aos meus pais, José e Marília, que sempre acreditaram em mim e que tornaram possível todo o caminho percorrido, a quem espero poder retribuir um dia todo o esforço que fizeram.

À minha irmã, Bruna, pela amizade, alegria e boa disposição contagiante, e a quem espero ver em igual situação brevemente.

À minha namorada, Sara, pelo incentivo diário e por todo o carinho, amor, paciência e bons momentos partilhados.

Ao resto da família, em especial aos meus avós, e a todos os amigos que me ajudaram e estiveram presentes nesta verdadeira aventura.

Resumo

A avaliação de cuidados de saúde é actualmente uma exigência nacional, pelo que são necessárias metodologias objectivas e consensuais que simultaneamente disponibilizem informação inteligível e promovam a melhoria dos cuidados prestados. Neste âmbito, o projecto “Modelo Estatístico de Avaliação dos Cuidados de Saúde - Excelência Clínica”, desenvolvido durante o estágio de 9 meses no sector de Healthcare da Siemens, S.A., visa redefinir o modelo estatístico de avaliação da qualidade dos cuidados de saúde prestados nos prestadores portugueses.

O modelo de avaliação actual contempla dois tipos de indicadores:

- Indicadores de resultado, que reflectem o efeito dos cuidados de saúde prestados no estado de saúde do utente, com o recurso a modelos que permitem ajustar o risco da ocorrência de um determinado acontecimento consoante as características do utente, permitindo comparar o desempenho observado com o desempenho esperado;
- Indicadores de processo, que reflectem o grau de concordância de determinadas práticas com as respectivas regras de conduta (*guidelines*), comparando-se os desempenhos individuais dos prestadores com o desempenho global.

O primeiro objectivo do projecto está ligado aos indicadores de resultado, uma vez que os modelos de ajuste de risco actualmente implementados em Portugal são os modelos da *Joint Commission International*, desenvolvidos com base nos dados de prestadores Norte-Americanos, sendo modelos mais adequados a essa realidade. Assim, o primeiro objectivo passa pela redefinição dos modelos de ajuste de risco actuais, utilizando modelos de regressão logística desenvolvidos com os dados recolhidos nos prestadores nacionais e todos os possíveis factores de risco/protecção suportados pela literatura. A construção de modelos mais adequados à realidade nacional é fundamental para uma classificação mais precisa dos prestadores nacionais, tendo em conta as diferenças existentes entre os dois países em diversos níveis.

O segundo objectivo do projecto está ligado aos indicadores de processo, passando pelo desenvolvimento de uma metodologia alternativa à actual. A metodologia actual permite comparar os desempenhos individuais dos prestadores com o desempenho médio global, o que trás limitações à perspectiva de evolução que se pretende na prestação de cuidados de saúde. A metodologia alternativa proposta neste projecto permite a introdução de valores de referência, construídos através da recolha e combinação de opiniões de especialistas para diversas áreas clínicas, classificando os prestadores através da comparação dos desempenhos individuais com os valores de referência.

Os resultados obtidos permitem perceber que é possível otimizar a metodologia de avaliação dos cuidados de saúde a nível nacional, através da implementação de modelos de ajuste de risco desenvolvidos com base em dados nacionais, algo inexistente em Portugal. A criação de uma metodologia de definição, validação e avaliação do desempenho dos modelos de ajuste de risco permite uniformizar o processo de desenvolvimento dos mesmos e a aplicação a futuros modelos.

A introdução de valores de referência para os indicadores de processo, mesmo apresentando bons resultados e minimizando algumas das limitações, significa uma rutura com a metodologia actual, sendo mais difícil a sua implementação no sistema de avaliação.

Palavras-Chave (Tema): cuidados de saúde, indicadores de resultado, indicadores de processo, valores de referência, opinião de especialistas

Palavras-Chave (Tecnologias): R, modelo de regressão logística, modelo *PERT*, risco

Abstract

The assessment of health care is currently a national requirement, therefore objective and consensual methodologies are necessary to simultaneously provide intelligible information and promote the improvement of the health care provided. In this context, the project “Statistical Model for the Evaluation of Health Care - Clinical Excellence”, developed during the 9-month internship in the Healthcare Department of Siemens, S.A., aims to redefine the statistical model for assessing the quality of health care in Portuguese providers.

The current assessment model contemplates two types of indicators:

- Outcome indicators, which reflect the effect of health care in the state of health of the user, by using models that allow to adjust the risk of occurrence of a certain event depending on the characteristics of the user, allowing to compare the observed performance with the expected performance;
- Process indicators, which reflect the degree of concordance of certain practices with their respective guidelines, comparing the performance of individual providers with the overall performance.

The first goal of the project is linked to the outcome indicators, since the risk adjustment models currently implemented in Portugal are the models of the Joint Commission International, developed based on the data from North American providers, which are models more appropriate to that reality. Therefore, the first goal is to redefine the current risk adjustment models, using logistic regression models developed with the data collected among the national providers and all the possible factors of risk/protection supported by literature. The creation of more appropriate models to the national reality is fundamental for a more precise classification of the national providers, taking into account the differences existing between the two countries at various levels.

The second goal of the project is linked to the process indicators, through the development of an alternative methodology to the current one. The current methodology allows to compare the performances of individual providers with the average overall performance, which results in limitations to the intended evolution perspective in the provision of health care. The alternative methodology proposed in this project allows the introduction of target values, created through the collection and combination of expert opinions for several clinical fields, classifying the providers through the comparison of individual performances with the target values.

The results obtained allow being aware that it's possible to optimize the assessment methodology of health care at a national level, through the implementation of risk adjustment models developed based on national data, something that doesn't exist in Portugal. The creation of a methodology for definition, validation and assessment of the performance of risk adjustment models allows standardizing their development process and the application to future models.

The introduction of target values for the process indicators, even with good results and minimizing some of the limitations, would mean a rupture with the current methodology, thus being more difficult to implement it in the assessment system.

Keywords (Theme): health care, outcome indicators, process indicators, target values, expert opinions

Keywords (Technology): R, logistic regression model, *PERT* model, risk

Conteúdo

Lista de Tabelas	xi
Lista de Figuras	xiii
1 Introdução	1
1.1 Enquadramento e Estado da arte	1
1.2 Apresentação do projecto	4
1.3 Contribuições do trabalho	5
1.4 Apresentação da empresa	5
1.5 Estrutura do relatório	7
2 Metodologia - modelos de ajuste de risco	9
2.1 Covariáveis	10
2.2 Modelos de regressão logística múltipla	12
2.2.1 Identificação de potenciais factores de risco	13
2.2.2 Análise preliminar dos dados	13
2.2.3 Selecção do modelo	14
2.2.4 Validação do modelo	18
2.2.5 Desempenho do modelo	22
2.2.6 Interpretação do modelo	25
3 Valores de referência	27
3.1 Metodologia actual	28
3.2 Limitações da metodologia actual	29
3.3 Objectivos da metodologia alternativa	30
3.4 Metodologias alternativas	31
3.5 Recolha e combinação de opiniões	33
3.5.1 Abordagem Comportamental	34
3.5.2 Abordagem Matemática	34
3.5.3 Recolha	36

3.5.4	Simulação	37
3.5.5	Combinação de opiniões	39
3.6	Determinação dos valores de referência	40
3.6.1	Classificação	41
3.7	Decisão estatística	44
4	Conclusões	51
4.1	Objectivos realizados	52
4.2	Outros trabalhos desenvolvidos	53
4.3	Limitações e trabalho futuro	53
4.4	Apreciação final	54
	Referências Bibliográficas	55
A	Código R	59
A.1	Modelos de ajuste de risco	60
A.2	Valores de referência	64

Lista de Tabelas

2.1	Exemplo do agrupamento para códigos ICD-9-CM de diagnóstico	11
2.2	Exemplo do agrupamento para códigos ICD-9-CM de procedimento	11
2.3	Poder de discriminação	24
3.1	Opiniões dos especialistas	38
3.2	Matriz perda	46
3.3	Matriz perda - caso de estudo	46

Lista de Figuras

2.1	Resíduos da desviância	19
2.2	<i>Leverage</i>	20
2.3	Distância de <i>Cook</i>	20
2.4	Verificação da relação linear da covariável idade	22
2.5	Curvas ROC	24
3.1	Opiniões dos especialistas - Funções Densidade de Probabilidade	38
3.2	Opiniões dos especialistas (individuais e combinada) - Funções Densidade de Probabilidade	39
3.3	Opinião combinada e valores de referência - Função de Distribuição Empírica	41
3.4	Classificação dos prestadores em 3 níveis	43
3.5	TI=0.655, TS=0.864	43
3.6	TI=0.864, TS=0.986	43
3.7	TI=0.345, TS=0.767	49
3.8	TI=0.735, TS=0.949	49

1

Introdução

1.1 Enquadramento e Estado da arte

A avaliação dos cuidados de saúde e a comparação do desempenho hospitalar tem, ao longo dos últimos anos, tomado grande importância a nível mundial, com o objectivo fundamental de promover uma cultura de melhoria contínua dos cuidados de saúde prestados.

Entidades como a *Joint Commission International (JCI)*, *Agency for Healthcare Research and Quality (AHRQ)*, entre outras, são especialistas na certificação de prestadores de cuidados de saúde com alto reconhecimento a nível internacional. Estas entidades utilizam modelos robustos e cientificamente validados que permitem fazer uma avaliação objectiva e transversal, em termos de excelência clínica e de segurança dos utentes, avaliando a qualidade dos cuidados de saúde prestados.

Esta avaliação é realizada utilizando indicadores de avaliação que estão divididos em duas categorias:

- Indicadores de processo, que avaliam o grau de concordância de determinadas práticas com as respectivas regras de conduta (*guidelines*);

- Indicadores de resultado, que reflectem o impacto dos prestadores e dos cuidados por estes prestados no estado de saúde do utente.

A avaliação dos prestadores de saúde permite disponibilizar aos utentes e profissionais de saúde informação objectiva sobre a qualidade dos cuidados prestados, permitindo a tomada de decisões mais informada, tornando-se ainda mais relevante para os países europeus após a introdução da directiva da União Europeia que indica que os Estados Membros terão de ajustar as leis actuais de forma a permitir a mobilidade e liberdade de escolha dos utentes até 25 de Outubro de 2013.[1]

Indicadores de processo

A metodologia aplicada aos indicadores de processo não sofre grandes alterações entre as diversas entidades, seguindo na sua maioria a metodologia aplicada pela JCI para estes indicadores. Nesta metodologia, o valor observado de cada prestador, que reflecte o desempenho do mesmo, é calculado como sendo a proporção de intervenções adequadas realizadas por cada prestador para um determinado indicador. Estas proporções são depois comparadas com um valor de referência, definido pela proporção global de intervenções adequadas, isto é, a proporção de intervenções adequadas de todos os prestadores para o determinado indicador.

Através da comparação dos valores observados com valores de referência, cujo cálculo pode variar entre as diferentes entidades, resulta um sistema de classificação habitualmente constituído por 3 níveis de classificação (abaixo da referência, dentro da referência, acima da referência).[2] Esta metodologia permite comparar os desempenhos individuais de uma forma prática e de fácil interpretação.

Indicadores de resultado

A metodologia aplicada aos indicadores de resultado pode ser dividida em duas fases: a primeira contempla a determinação de um modelo de ajuste de risco, permitindo ajustar o risco de um determinado acontecimento consoante as características ¹ de cada utente, existindo alguma homogeneidade no tipo de modelos utilizados por diferentes entidades e em diferentes países; a segunda utiliza o modelo de ajuste de risco para inferir sobre

¹Características sócio-demográficas como idade, sexo ou peso, mas também diagnósticos e procedimentos registados pelos prestadores de cuidados de saúde.

a classificação dos prestadores de cuidados de saúde, existindo neste caso uma maior variedade de alternativas.

O tipo de modelos de ajuste de risco mais utilizados neste tipo de indicadores são os modelos de regressão logística, uma vez que os indicadores de resultado são indicadores de resposta binária (realização ou não realização do acontecimento de interesse). Estes modelos são determinados com base nos dados recolhidos em diversos prestadores de cuidados de saúde para cada indicador de resultados, sendo este tipo de modelos bastante consensuais entre as várias entidades (JCI, AHRQ, entre outras).

As diferenças entre os processos de classificação das diferentes entidades situam-se sobretudo na forma como os modelos de regressão logística são utilizados para classificar os prestadores de cuidados de saúde. Utilizando estes modelos é possível comparar os valores estimados pelos mesmos para cada prestador, valores esperados, com a realidade observada em cada prestador, ou seja, com os valores observados de cada prestador. Esta comparação é habitualmente realizada de duas formas:

- utilizando um teste z-score para comparar os valores esperados e observados de cada prestador, classificando-os em 2 níveis consoante exista diferença significativa positiva ou negativa entre os valores esperados e os valores observados, ou num terceiro nível caso não exista diferença significativa entre os mesmos;[3]
- utilizando a razão de chances (*odds ratio* ou simplesmente OR), isto é, a razão entre a chance observada e a chance esperada para cada prestador, classificando os prestadores em 3 níveis.[4] Esta metodologia apresenta grandes vantagens a nível de interpretação e de computação face às restantes, calculando o intervalo de confiança para o OR.

Apesar de ser nesta segunda fase que as metodologias de classificação mais variam, estas podem ser aplicadas de forma transversal na avaliação de prestadores de diferentes países, permitindo assim uma “exportação” da metodologia de classificação.

Por outro lado, a construção dos modelos de ajuste de risco está dependente das observações dos utentes que foram utilizados para os determinar. Obviamente que um modelo desenvolvido com base nos dados recolhidos em prestadores do país X deverá, à partida, ser mais adequado e ter um maior poder de predição quando aplicado a outros dados do mesmo país do que quando aplicado aos dados do país Y , diminuindo o poder de predição e adequabilidade do modelo à medida que as diferenças relativas aos utentes dos dois países aumentam (características sócio-demográficas) ou às diferenças de técnicas de diagnóstico e de procedimentos utilizadas nos prestadores dos dois países.

Apesar destas condicionantes, esta “exportação” de modelos é frequentemente utilizada. No início da implementação deste tipo de avaliação no país *Y*, não existem dados suficientes para se determinarem modelos fiáveis que possam ser implementados, sendo utilizados os modelos importados de outros países até que o volume de dados disponível permita determinar modelos mais ajustados às realidades em que se inserem.

1.2 Apresentação do projecto

O projecto desenvolvido no âmbito do estágio compreende 2 objectivos principais:

1. Redefinição e implementação de modelos de ajuste de risco de suporte à avaliação dos cuidados de saúde adaptados à realidade nacional;
2. Implementação de valores de referência (*target*) para os indicadores de processo de suporte à avaliação dos cuidados de saúde nacionais.

Quanto ao primeiro objectivo, pretende-se definir modelos de ajuste de risco mais adaptados à realidade nacional que sejam alternativas válidas aos modelos actuais que, sendo modelos robustos e cientificamente validados, foram desenvolvidos pela JCI com base nos dados recolhidos de prestadores Norte-Americanos, sendo por essa razão modelos mais adequados à realidade dos Estados Unidos da América.

Os modelos de ajuste de risco são, por norma, modelos de regressão logística cujos factores de risco que os integram são características da população em estudo que estão relacionadas com a realização ou não realização da variável de interesse.² Estes factores de risco podem ir desde características demográficas, como a idade, peso ou sexo dos utentes, até comorbilidades diagnosticadas ou procedimentos efectuados no hospital. Admitindo que existem diferenças entre Portugal e os Estados Unidos, tanto nas características da população como nas técnicas de diagnóstico e procedimentos efectuados, é natural que existam diferenças entre os factores de risco que se associam significativamente com as variáveis resposta. Desta forma, torna-se fundamental a redefinição dos modelos actuais, de forma a garantir uma avaliação mais precisa dos prestadores nacionais.

Relativamente ao segundo objectivo, pretende-se construir uma metodologia alternativa às metodologias actuais que permita incluir valores de referência (*target*). As metodologias actuais assentam fundamentalmente sobre uma ideia: a comparação dos desempenhos individuais com o desempenho médio global, funcionando o desempenho médio

²Também denominada de variável resposta, variável dependente ou desfecho.

global como o objectivo a atingir por parte dos prestadores, condicionando assim a perspectiva de evolução desejada.

A metodologia alternativa em estudo prevê a utilização de valores de referência que possam servir de referência para os prestadores, que não sejam condicionados pelos resultados dos próprios e que minimizem as limitações da metodologia actual. Pretende-se determinar os valores de referência através da recolha e combinação de opiniões de especialistas utilizando métodos estatísticos, seguida da comparação entre os desempenhos individuais dos prestadores e os valores de referência para dar origem a uma classificação em 3 níveis.

1.3 Contribuições do trabalho

O presente estudo apresenta aspectos inovadores nos dois objectivos:

- No primeiro, e apesar da metodologia utilizada ser bastante comum, a construção de modelos de ajuste de risco adaptados à realidade nacional representa uma clara optimização cuja implementação originará uma classificação dos prestadores nacionais mais precisa do que a actual. A implementação dos modelos desenvolvidos no sistema de avaliação será uma inovação, uma vez que são os primeiros modelos de ajuste de risco desenvolvidos com base nos dados de prestadores nacionais;
- No segundo, a introdução de valores de referência que são obtidos através da recolha e combinação da opinião de especialistas é algo inédito na classificação de cuidados de saúde, permitindo comparar os desempenhos individuais dos prestadores com valores de referência, ao contrário das metodologias actuais que comparam com o desempenho médio global; esta metodologia tem ainda a vantagem de minimizar as limitações de outras metodologias.

1.4 Apresentação da empresa

A Siemens está em Portugal há mais de 105 anos, sendo líder no fornecimento de soluções de engenharia nos sectores de Indústria, Energia, Saúde e Infra-estruturas & Cidades. Com cerca de 2000 colaboradores, duas unidades de produção e numerosas parcerias com o meio académico, a empresa desempenha um papel activo no desenvolvimento económico do país.

O Sector Energy é líder mundial no fornecimento de toda a gama de produtos, serviços e soluções para a produção de energia em centrais termoeléctricas, aproveitamento de energias renováveis, transmissão de energia e ainda tecnologias nas áreas da extracção, conversão e transporte de petróleo e gás.

O Sector Industry é um dos maiores fornecedores mundiais de produtos e soluções inovadoras e ecológicas para clientes industriais. Com uma gama completa de tecnologias de automação, soluções de software inteligente, vasta experiência industrial e serviços integrados complexos, o Sector promove a produtividade, eficiência e flexibilidade dos nossos clientes - e, ao mesmo tempo, reforça a sua competitividade.

Com um portefólio que inclui soluções de mobilidade integrada, sistemas de automação e de segurança para edifícios, equipamento de distribuição de energia, aplicações de rede eléctrica inteligente e produtos de baixa e média tensão, o novo Sector Infrastructure & Cities oferece tecnologias sustentáveis para centros metropolitanos e infra-estruturas urbanas em todo o mundo.

O Sector Healthcare é um dos maiores fornecedores mundiais na indústria dos cuidados de saúde e líder em sistemas de imagiologia, diagnósticos laboratoriais, tecnologia de informação médica e aparelhos auditivos. É a primeira empresa a nível mundial a oferecer um portefólio integrado de tecnologia que permite responder a todas as fases do ciclo de cuidados de saúde, disponibilizando produtos e soluções para todo o tipo de cuidados ao utente - desde a prevenção e diagnóstico precoce a situações de pós-tratamento. Ao otimizar os procedimentos clínicos associados às mais importantes condições clínicas, permite tornar os cuidados de saúde mais rápidos, melhores e mais rentáveis.

Em Portugal, o Sector Healthcare da Siemens é um dos líderes de mercado na área da saúde, reconhecido pelas suas competências, *know-how* e força de inovação em diagnóstico e tecnologias terapêuticas, assim como engenharia de conhecimento, incluindo tecnologias de informação, integração de sistemas e serviços de consultoria.

Nos últimos anos, o Sector Healthcare da Siemens tem promovido uma estratégia de contacto e parceria com a Comunidade Académica e Científica em Portugal, no sentido da criação de uma rede de conhecimento e parcerias estratégicas que potenciem a inovação, a investigação e o desenvolvimento (IDI) na área da Saúde.[5]

1.5 Estrutura do relatório

O presente relatório é constituído por 4 capítulos e 2 anexos. No capítulo actual, é realizada uma breve introdução ao tema e apresentação do projecto e respectivos objectivos.

No segundo capítulo, é apresentada a metodologia de construção e validação dos modelos de ajuste de risco desenvolvidos.

No capítulo 3 é apresentada a metodologia actualmente implementada em Portugal para os indicadores de processo, assim como a metodologia alternativa desenvolvida neste projecto.

No quarto capítulo são apresentadas as conclusões obtidas para os dois objectivos definidos, assim como o restante trabalho desenvolvido durante o estágio.

O anexo A contém o código **R** utilizado para a realização dos objectivos propostos.

Por fim, o anexo ??, capítulo que está disponível apenas na versão confidencial, contém os modelos de ajuste de risco desenvolvidos.

O *software* utilizado em todo o projecto é o **R** versão 2.15.2, uma linguagem e um ambiente de desenvolvimento integrado de utilização livre, amplamente utilizado para cálculos estatísticos e gráficos.[6] As funções utilizadas no desenvolvimento do projecto encontram-se definidas no Anexo A.

2

Metodologia - modelos de ajuste de risco

Como descrito anteriormente, o projecto está dividido em 2 objectivos sendo que neste capítulo será descrita a metodologia aplicada no primeiro objectivo, designadamente a redefinição e implementação de modelos de ajuste de risco de suporte à avaliação dos cuidados de saúde adaptados à realidade nacional.

Os modelos de ajuste de risco desenvolvidos neste projecto são apresentados no Anexo ?? e são estritamente confidenciais. Por esta razão, o Anexo ?? está disponível apenas para os membros do júri, não estando disponível na versão pública deste relatório.

Indicadores de resultado

Como referido no capítulo 1, os modelos de ajuste de risco utilizados na avaliação de cuidados de saúde são, por norma, modelos de regressão logística múltipla. Para cada indicador existe uma variável que indica o resultado ou desfecho do acontecimento de interesse. Essa variável, denominada de variável resposta e usualmente representada por Y , é uma variável dicotómica para todos os indicadores de resultado, isto é, é uma variável que assume apenas 2 valores (usualmente 0 e 1, em que 0 indica a não ocorrência do acontecimento de interesse e 1 a ocorrência do mesmo). Face à maior flexibilidade apresentada pelos modelos de regressão logística em termos de pressupostos[7], em relação por exemplo aos modelos de análise discriminante, e aos bons resultados apresentados nesta área da saúde, foram utilizados modelos de regressão logística.

A metodologia de construção dos modelos de regressão logística é bastante conhecida, sendo que a única inovação no desenvolvimento destes modelos é a utilização de dados nacionais. Neste capítulo serão introduzidas algumas noções que se consideram necessárias para o desenvolvimento dos modelos e os passos a seguir na construção dos mesmos para cada indicador de resultados, não sendo aprofundados detalhes mais técnicos. Pretende-se assim construir um documento que sirva de suporte à construção de novos modelos ou, futuramente, que sirva de suporte à actualização dos modelos desenvolvidos neste projecto, uniformizando assim o processo de construção e validação dos mesmos.

2.1 Covariáveis

Os dados disponíveis para a construção dos modelos de ajuste de risco foram recolhidos entre Janeiro de 2010 e Dezembro de 2011, sendo que o número de hospitais nacionais envolvidos varia entre as várias áreas clínicas. Estes dados apresentam dois tipos de variáveis:

- Variáveis demográficas → Idade (em anos), Sexo (Feminino, Masculino) e Peso (em gramas);
- Códigos ICD-9-CM¹, em vigor em Portugal desde 1980, que estão divididos em duas categorias:

¹“Conjunto de códigos de diagnósticos e de procedimentos utilizados para classificação e codificação da informação de morbilidade e mortalidade para fins estatísticos e para indexação dos registos hospitalares por doença e intervenções cirúrgicas, para armazenamento e pesquisa.” [8]

- Códigos de Diagnóstico - cada código corresponde a uma doença diagnosticada ao utente;
- Códigos de Procedimento - cada código corresponde a um procedimento efectuado no hospital.

Existem milhares de códigos ICD-9-CM de diagnóstico e de procedimento, pelo que seria impraticável considerar cada um destes códigos como sendo um possível factor de risco ou de protecção face à variável resposta. Assim, a solução utilizada foi a de agrupar estes códigos recorrendo a referências bibliográficas para os diversos indicadores. As tabelas 2.1 e 2.2 exemplificam a codificação que foi efectuada:

Código	Descrição	Variável
41400	Aterosclerose de vaso coronário	
41401	Aterosclerose de artéria coronária nativa	Aterosclerose
41402	Aterosclerose em bypass coronário venoso	
4280	Insuficiência cardíaca congestiva	
4281	Insuficiência cardíaca esquerda	Insuf. Cardíaca
42820	Insuficiência cardíaca sistólica	
584	Insuf. renal aguda	
5845	Insuf. renal aguda - lesão necr. tabular	Insuf. Renal
5846	Insuf. renal aguda - lesão necr. cortical	

Tabela 2.1: Exemplo do agrupamento para códigos ICD-9-CM de diagnóstico

Assim, se um utente apresentar um dos códigos de diagnóstico referente a uma determinada doença, a covariável correspondente toma o valor 1 (presente) para esse utente; caso contrário, toma o valor 0 (ausente).

Código	Descrição	Variável
720	Aplicação de forceps baixo	
721	Aplicação de forceps baixo com episiotomia	Forceps
724	Rotação da cabeça fetal com forceps	
7271	Extracção por ventosa com episiotomia	Ventosa
7279	Extracções por ventosa NCOP	
731	Induções cirúrgicas do trabalho de parto NCOP	Indução
734	Indução clínica do trabalho de parto	

Tabela 2.2: Exemplo do agrupamento para códigos ICD-9-CM de procedimento

De forma análoga, se um utente apresentar um dos códigos de procedimento referente a um determinado procedimento, a covariável correspondente toma o valor 1 (presente) para esse utente; caso contrário, toma o valor 0 (ausente).

2.2 Modelos de regressão logística múltipla

O objectivo, para cada indicador de resultado, passa por determinar o modelo de ajuste de risco mais adequado, parcimonioso e biologicamente válido para descrever a relação entre a variável resposta (Y) e um conjunto de variáveis independentes (x_k), covariáveis ou factores de risco ², introduzidas na secção 2.1. Assim, é possível ajustar o risco de um determinado acontecimento ($Y_i = 1$) consoante as características apresentadas pelos utentes (x_{ik}), utilizando os modelos de regressão logística.

Para cada indicador de resultado existem n variáveis resposta independentes $Y_i \cap B(1, \pi_i)$, *i.e.*,

$$f(y_i|\pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, y_i = 0, 1 \quad (2.2.1)$$

sendo que a cada indivíduo (leia-se utente) i está associado um vector de especificação $\mathbf{z}_i = (1, x_{i1}, \dots, x_{ik})^T, i = 1, \dots, n, k = 1, \dots, p-1$ (*covariáveis*). Porém, quando existem covariáveis qualitativas com q categorias ($q > 2$), estas covariáveis têm de ser codificadas através de variáveis mudas, sendo necessário criar $q - 1$ variáveis mudas que representem essas covariáveis que serão incluídas no vector \mathbf{z} . [9]

Sendo $E[Y_i] = \pi_i$ com função de ligação canónica $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \text{logit}(\pi_i)$, denominada de função *logit*, o modelo de regressão logística é dado por

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}, \quad (2.2.2)$$

onde a probabilidade de sucesso, $P(Y_i = 1) = \pi_i$, é dada por:

$$\pi_i = \frac{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{z}_i^T \boldsymbol{\beta})} \quad (2.2.3)$$

em que $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$ representa os coeficientes ou efeitos das covariáveis do modelo.

²Em sentido lato, referem-se também a factores de protecção em relação à variável resposta.

Foram delineadas 5 etapas para a construção de cada um dos modelos de regressão logística múltipla:

1. Identificação de potenciais factores de risco;
2. Análise preliminar dos dados;
3. Selecção do modelo;
4. Validação dos pressupostos do modelo;
5. Análise do desempenho do modelo.

2.2.1 Identificação de potenciais factores de risco

Antes da análise preliminar dos dados é necessário efectuar uma revisão bibliográfica que permita identificar um conjunto de possíveis factores que possam explicar os acontecimentos de interesse. É fundamental ter em conta que, por muito bom que seja o poder explicativo de uma covariável face à variável resposta, se, do ponto de vista biológico não fizer sentido estas variáveis terem qualquer relação, é imprudente utilizá-la no modelo de regressão.

Um dos maiores suportes para a determinação destes possíveis factores de risco são os modelos actualmente utilizados pela JCI, considerados modelos robustos e fidedignos, devendo os factores de risco que são considerados importantes para estes modelos ser considerados na análise. Artigos científicos, jornais científicos e outros estudos são também uma mais valia para esta fase de pré-selecção de factores de risco.

Assim, a primeira etapa para a construção do modelo de cada indicador de resultados passa por um levantamento exaustivo de possíveis factores de risco que possam explicar da melhor forma os acontecimentos de interesse de cada indicador.

2.2.2 Análise preliminar dos dados

A segunda etapa da construção dos modelos de regressão logística compreende uma análise exploratória das covariáveis que poderão ser importantes para cada indicador de resultado. Esta análise permite compreender, por exemplo, a proporção de ocorrências do acontecimento de interesse ou identificar covariáveis categóricas com poucas observações

em uma das categorias, o que conduz a uma revisão da forma como estas foram codificadas e que pode levar à exclusão das mesmas quando essa codificação não pode ser alterada (optou-se por excluir as covariáveis categóricas com menos de 10 observações numa categoria de forma a evitar problemas de convergência no ajustamento dos modelos, tentando assim evitar futuros problemas de separação - quando um preditor está completamente alinhado com a variável resposta, não permitindo estimar os parâmetros do modelo).

Nesta etapa foram construídos modelos de regressão logística simples, para cada covariável, obtendo-se um estudo preliminar da relação existente (ou inexistente) entre cada covariável e a variável resposta. Esta análise, sugerida por Hosmer e Lemeshow (1989), não é conclusiva, mas revela-se importante na detecção de possíveis contradições e de covariáveis que poderão ser importantes para o modelo final. São consideradas covariáveis importantes para o modelo final todas as covariáveis que apresentarem um valor p (*p-value*) associado ao teste de razão de verosimilhanças inferior a 0.25, em cada modelo de regressão logística simples.

2.2.3 Selecção do modelo

Após a identificação de possíveis factores de risco e de uma primeira análise das variáveis presentes em cada indicador, a etapa seguinte passa pela determinação do modelo, isto é, determinar as covariáveis e respectivos coeficientes a introduzir no modelo final. Quando existe um elevado número de covariáveis, é importante que o modelo final seja o mais parcimonioso possível, incluindo o menor número possível de covariáveis sem por em causa a capacidade de ajustamento e predição do modelo, e de forma a permitir uma boa interpretação do mesmo. Podemos então definir o “melhor modelo” como sendo aquele que melhor combina os princípios de ajustamento, parcimónia e interpretação.

O primeiro passo na selecção do modelo começa com a divisão dos dados em duas amostras de forma aleatória, amostra 0 e amostra 1, mutuamente exclusivas. A amostra 0 será utilizada para a determinação e validação dos pressupostos do modelo, enquanto que a amostra 1 será utilizada para avaliar o desempenho do modelo determinado. Esta técnica, denominada de **validação cruzada** [10], permite avaliar a capacidade de generalização do modelo, isto é, permite avaliar o desempenho do modelo para um novo conjunto de dados.

A avaliação da capacidade de generalização dos modelos é extremamente importante

para o projecto, uma vez que se pretende que os modelos desenvolvidos sejam implementados futuramente num processo de avaliação dos prestadores de cuidados de saúde nacionais. Nesse processo de avaliação, os modelos implementados serão utilizados em novos conjuntos de dados em cada avaliação efectuada, reforçando assim a importância da utilização da técnica de validação cruzada para avaliar o desempenho dos modelos em novos conjuntos de dados.

Utilizando a amostra 0, é iniciado o processo de selecção de covariáveis a incluir no modelo. Existem diversos métodos de selecção automática de covariáveis que podem dar origem a diferentes modelos. Entre os mais utilizados encontram-se os métodos:

- **Forward (Progressivo)** - Partindo do modelo nulo (modelo sem covariáveis), testa-se a adição de cada covariável ao modelo. As covariáveis vão sendo adicionadas sucessivamente até que não existam mais covariáveis cuja adição ao modelo produza alterações significativas ao mesmo, isto é, até não existirem covariáveis que “passem” um pré-determinado critério de inclusão. Neste método, qualquer covariável que seja adicionada ao modelo manter-se-á até ao modelo final.
- **Backward (Regressivo)** - Partindo do modelo com todas as covariáveis, testa-se a remoção de cada covariável. As covariáveis vão sendo retiradas sucessivamente do modelo até que não existam mais covariáveis cuja remoção do modelo produza alterações significativas ao mesmo, isto é, até não existirem covariáveis que “passem” um pré-determinado critério de exclusão. Neste método, qualquer covariável que seja excluída do modelo não pode voltar a ser adicionada.
- **Stepwise (Passo-a-passo)**[7, 9] - É uma combinação dos dois métodos anteriores, que testa em cada passo as covariáveis que devem ser incluídas ou excluídas, partindo do modelo nulo. O processo é repetido até não existir nenhuma covariável cuja adição ou remoção do modelo produza alterações significativas ao mesmo, isto é, até não existirem covariáveis que “passem” os critérios de inclusão e exclusão.

Nos métodos referidos, é necessário definir à partida os critérios de inclusão e/ou exclusão de covariáveis. Estes critérios são definidos através dos valores p associados a um determinado teste à nulidade do coeficiente de cada covariável (ou vector de covariáveis).

Também neste caso existem diversos testes que podem ser utilizados, sendo usual recorrer à **função desviância**, D . Esta função é uma medida de distância entre o modelo saturado S (modelo com tantas covariáveis quanto observações que conduz a um ajustamento

perfeito aos dados e conseqüentemente ao maior valor que a função de verosimilhança atinge) e o modelo corrente (modelo que está a ser avaliado - \mathbf{M}), que é definida com base na verosimilhança dos dois modelos, $D = -2 \log \left(\frac{L_M}{L_S} \right)$, em que L_M e L_S representam as verosimilhanças dos modelos \mathbf{M} e \mathbf{S} , respectivamente.[11] Assim, quanto mais próximo o modelo estimado estiver dos dados observados, menor será o valor da função desviância, tendendo para zero à medida que o modelo em avaliação se aproximar do modelo saturado.

Esta função é ainda determinante para a comparação de modelos encaixados, devido à sua propriedade de aditividade. Sejam M_k e M_l dois modelos intermédios, com $M_k \subset M_l$, isto é, todas as k covariáveis de M_k estão presentes nas l covariáveis do modelo M_l . Para testar a hipótese de que os dois modelos não diferem significativamente, utiliza-se o **Teste de Razão de Verosimilhanças** (ou estatística de Wilks [12]) definido por

$$\lambda = D_k - D_l = -2 \log \left(\frac{L_{M_k}}{L_{M_l}} \right), \quad (2.2.4)$$

em que D_k e D_l representam as funções desviância dos modelos M_k e M_l , respectivamente. Sob a hipótese nula de que o modelo M_k é tão bom como o modelo M_l , λ tem distribuição Qui-Quadrado com $l - k$ graus de liberdade.

Em suma, a adição ou remoção de uma covariável na construção do “melhor modelo” pode ser determinada comparando dois modelos: o modelo com a covariável em estudo e o modelo sem a covariável em estudo. Quanto menor for o valor p associado ao teste da razão de verosimilhanças dos modelos que incluem e excluem essa covariável, maior é a importância da mesma para o modelo.³

Como referido anteriormente, existem outros testes e estatísticas de teste que podem ser utilizados como critérios de inclusão ou exclusão de covariáveis [9] como a estatística de Wald, a estatística de Rao ou o critério de informação de Akaike [7, 9, 13], mais utilizado na comparação de modelos não encaixados.

De forma geral, é preferível utilizar a estatística de Wilks quando o número de covariáveis é pequeno ou a amostra tem uma dimensão razoável. Quando a amostra tem uma grande dimensão, os resultados tendem a ser semelhantes para as várias estatísticas.[14] Neste caso, existindo indicadores com amostras grandes e com amostras mais reduzidas, a estatística utilizada foi a estatística de Wilks ou teste de razão de verosimilhanças.

Para os modelos em estudo foi utilizado o método de selecção *Stepwise*. Para assegurar

³No \mathbf{R} , utilizam-se as funções `add1()` e `drop1()`.

que entram apenas covariáveis com importância significativa para o modelo, é necessário definir o limite para a entrada, $P_e = 0.05$, isto é, são candidatas a entrar no modelo apenas as covariáveis cujo valor p associado ao teste da razão de verossimilhanças seja inferior a P_e . De forma análoga, define-se o limite para a exclusão de covariáveis, $P_s = 0.10$, isto é, são candidatas a sair do modelo todas as covariáveis cujo valor p associado ao teste da razão de verossimilhanças seja superior a P_s .

Os efeitos das covariáveis presentes no modelo são determinados através das estimativas de máxima verossimilhança para β . Estas estimativas são obtidas através da resolução do sistema de equações de verossimilhança $\frac{\partial l(\beta)}{\partial \beta_k} = 0, k = 1, \dots, p$, em que $l(\beta)$ representa o logaritmo da verossimilhança de β .

Geralmente, estas equações não têm solução analítica, sendo necessário recorrer a métodos iterativos para a sua resolução. Neste trabalho, as estimativas de máxima verossimilhança para β foram determinadas de forma automática através da utilização da função *glm* do pacote base do **R**, que utiliza o método de Newton-Raphson para resolver o sistema de equações.

Utilizando o método de selecção e o critério de inclusão/exclusão referidos anteriormente, é construído uma primeira versão do modelo final. De seguida, é testada a inclusão de possíveis interacções das covariáveis. A interacção representa a associação entre uma covariável e a variável resposta que varia de acordo com o valor de uma outra covariável, usualmente denominada de covariável modificadora de efeito, que provoca uma alteração no valor do coeficiente de uma das duas covariáveis envolvidas na interacção.

A inclusão de interacções no modelo deve estar sujeita a três princípios:

1. Se a interacção entre as variáveis tiver um significado clínico;
2. Se for estatisticamente significativa;
3. Mantendo os efeitos principais no modelo (princípio de marginalidade).

Incluídas as interacções no modelo, obtém-se uma segunda versão do modelo final. O passo seguinte é a validação dos efeitos de todas as covariáveis presentes no modelo, isto é, verificar se os efeitos de cada covariável têm sentido do ponto de vista clínico (por exemplo, se a presença de uma determinada doença aumenta a probabilidade de morte do utente). Caso sejam encontradas covariáveis com efeitos contrários ao esperado, é necessário perceber porque razão se verificam. Pode acontecer que a codificação, a escassez de observações de uma determinada covariável ou a presença de outras covariáveis no modelo, leve a que o efeito de uma covariável seja contrário ao que é esperado ou plausível.

Nesses casos, é conduzida uma análise à codificação da covariável com o objectivo de verificar se foi efectuada da forma mais correcta. É também realizada uma nova revisão da literatura que possa explicar o que foi encontrado, sendo que, nos casos em que se confirma que a covariável é de facto contraditória, esta é retirada do modelo e todo o processo de selecção reiniciado.

Sumariando os passos definidos até este ponto:

1. Dividir os dados aleatoriamente em duas amostras (validação cruzada), utilizando apenas a amostra 0 para os restantes passos de selecção do modelo;
2. Seleccionar as covariáveis através do método *Stepwise*, utilizando o valor p associado ao teste da razão de verosimilhanças dos modelos que incluem e excluem determinada covariável como critério de inclusão/exclusão das mesmas;
3. Refazer o passo anterior com o objectivo de seleccionar as interacções que envolvem as covariáveis presentes no modelo, quando o método *Stepwise* terminar;
4. Rever o modelo obtido, isto é, verificar se os efeitos associados às covariáveis que entram no modelo são válidos do ponto de vista clínico.

2.2.4 Validação do modelo

Na quarta etapa são validados os pressupostos do modelo. Os seguintes passos são efectuados utilizando a amostra 0.

Análise dos resíduos

O primeiro passo prende-se com a análise dos resíduos, que representam a distância entre o valor estimado pelo modelo e o valor observado, sendo esta análise essencialmente gráfica. Os resíduos mais utilizados quando se trabalha com modelos de regressão logística são os resíduos da desviância (*deviance residual*), $r_i^d = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}$, em que d_i representa a contribuição da i -ésima observação para a função desviância e $\hat{\mu}_i$ os valores ajustados pelo modelo.

Como os resíduos apresentam variabilidades diferentes, devem ser padronizados pelo erro padrão de y_i , considerando também os efeitos das covariáveis nos resíduos através da matriz de projecção (*hat*).

Assim, os resíduos da desviância padronizados são definidos por

$$r_i^{d'} = \frac{r_i^d}{\sqrt{(1 - h_{ii})}} \quad (2.2.5)$$

em que h_{ii} representa os valores da diagonal da matriz de projecção, que no caso do modelo logístico é definida por

$$\mathbf{H} = \mathbf{V}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{1/2}, \quad (2.2.6)$$

onde \mathbf{X} representa a matriz $n \times (k + 1)$ que contém os valores das k covariáveis e \mathbf{V} a matriz diagonal ($n \times n$) com elemento geral dado por $v_i = \hat{\pi}_i(1 - \hat{\pi}_i)$, onde $\hat{\pi}_i$ a probabilidade estimada pelo modelo.[7]

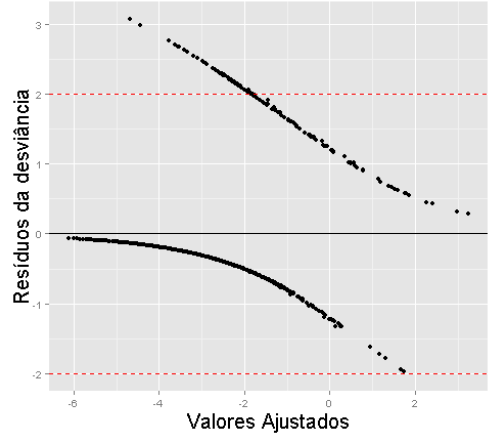


Figura 2.1: Resíduos da desviância

Representando graficamente os valores ajustados *vs* resíduos da desviância padronizados, espera-se que 95% dos resíduos estejam compreendidos entre -2 e 2 [15], não existindo anomalias nesta análise se esta condição se verificar, como se exemplifica na figura 2.1. Esta análise, em conjunto com a detecção de observações influentes, é de extrema importância para a detecção de observações discordantes que podem enviesar o modelo.

Detecção de observações influentes

A detecção de observações influentes é realizada essencialmente através do cálculo de duas medidas: *Leverage* (ponto de alavanca) e *Distância de Cook*.

Leverage

A *Leverage* de uma observação indica o efeito que essa observação tem nos valores ajustados, sendo que quanto maior for o seu valor de *Leverage*, maior é a sua influência. Os valores de *Leverage* são dados pelos elementos da diagonal da matriz de projecção, h_{ii} , sendo que a soma destes valores corresponde ao número de parâmetros do modelo, isto é, $\sum_{i=1}^n h_{ii} = k + 1$ com $h_{ii} \in [0, 1]$.

Assim, uma observação é considerada influente se $h_{ii} > 2k/n$ [9], mas só se considera preocupante se a mesma observação apresentar simultaneamente valores de resíduos elevados.

Distância de *Cook*

A Distância de *Cook* de uma observação indica a mudança dos coeficientes estimados da regressão quando se retira do modelo essa observação, quantificando assim o peso dessa observação no modelo. A distância de *Cook* é definida por

$$\Delta\hat{\beta}_i = (\hat{\beta} - \hat{\beta}_{(-i)})^T (\mathbf{X}^T \mathbf{V} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(-i)}), \quad (2.2.7)$$

onde $\hat{\beta}$ representa as estimativas de máxima verossimilhança dos parâmetros do modelo e $\hat{\beta}_{(-i)}$ as estimativas de máxima verossimilhança dos parâmetros do modelo retirando a observação i . Assim, consideram-se observações influentes as observações com $\Delta\hat{\beta}_i > 4/n$. [16]

Graficamente, é usual representar os valores de *Leverage* e de Distância de *Cook* de cada observação i contra i , como representado nas figuras 2.2 e 2.3.

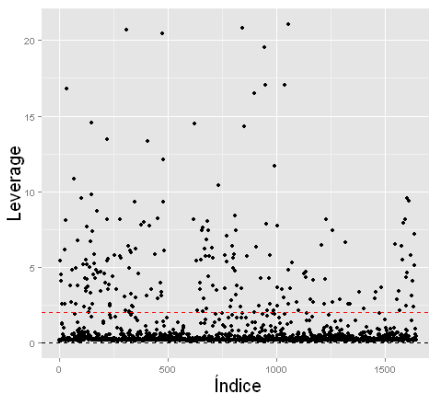


Figura 2.2: *Leverage*

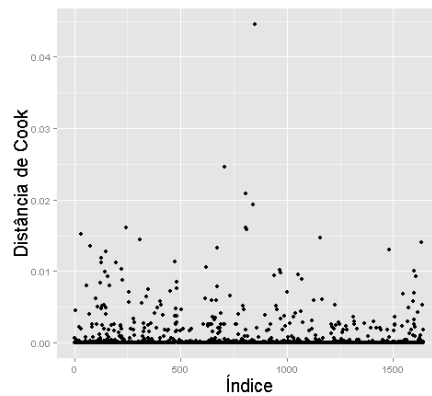


Figura 2.3: Distância de *Cook*

Todas as observações que sejam consideradas influentes e que apresentem também valores de resíduos elevados, podem ser consideradas observações discordantes, que influenciam negativamente a capacidade de predição do modelo. Estas observações devem ser identificadas e estudadas de forma a verificar se existe algum padrão entre as mesmas que provoque o afastamento da realidade. Um modelo que apresente uma grande proporção de observações discordantes deverá ser revisto, na tentativa de se encontrarem outras co-

variáveis que possam explicar mais correctamente o acontecimento de interesse. Porém, deve ser tido em consideração que o modelo obtido é uma aproximação da realidade e que, conseqüentemente, existirão naturalmente observações que o modelo não consegue explicar convenientemente.

Estudo da multicolinearidade

A multicolinearidade pode ser definida como uma forte relação entre duas ou mais covariáveis, que não permite estimar individualmente os parâmetros do modelo. A detecção de covariáveis colineares pode ser efectuada através do cálculo do *VIF* - *Variance Inflation Factor*, permitindo identificar covariáveis que sejam fortemente correlacionadas. O *VIF* é definido por:

$$VIF_k = \frac{1}{1 - R_k^2} \quad (2.2.8)$$

onde R_k^2 é o coeficiente de determinação não ajustado da regressão de x_k contra as restantes covariáveis do modelo. A solução mais utilizada para as covariáveis que apresentem *VIF* elevado, $VIF > 10$ [17], passa por retirar uma dessas covariáveis e refazer toda a análise.

Verificação da relação linear das covariáveis

Um dos pressupostos dos modelos de regressão logística é o de existir uma relação linear parcial entre a variável resposta e as covariáveis contínuas, controlando pelas restantes covariáveis presentes no modelo.

A avaliação do cumprimento deste pressuposto é feita através do gráfico *Component-plus-residuals*, representando o resíduo parcial da covariável x_k dado por $r_i^{(k)} = r_i^{d'} + \beta_k x_{ik}$ contra a própria covariável x_k , incluindo uma função de alisamento para uma detecção mais fácil da linearidade. Para o pressuposto ser considerado válido, é necessário existir uma recta como a representada na figura 2.4. No caso deste pressuposto ser violado, a solução passa pela utilização de uma transformação adequada ou pela categorização da covariável em causa utilizando referências bibliográficas. Esta categorização leva a uma perda de informação, pelo que só deve ser realizada em último caso.

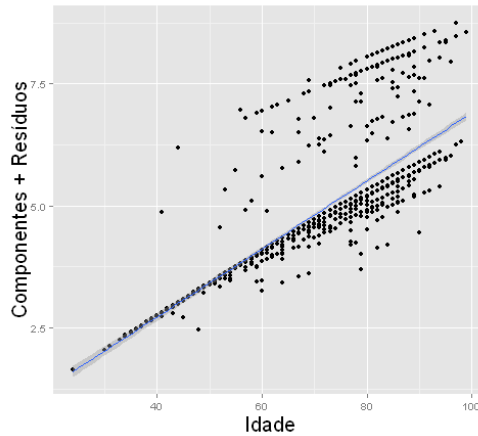


Figura 2.4: Verificação da relação linear da covariável idade

Análise da adequação da função de ligação

A última verificação em termos de diagnóstico passa por verificar se a função de ligação utilizada (*logit*) é adequada. Esta verificação é efectuada adicionando um preditor linear, η^2 , como covariável ao modelo e analisar a alteração da função desviância. Por outras palavras, é efectuada o teste da razão de verosimilhanças para o modelo com e sem a covariável η^2 . Se o teste da razão de verosimilhanças for significativo, concluí-se que a função de ligação não é adequada. Observações discordantes e a utilização de covariáveis com escalas erradas são as principais causas que podem levar à não adequação da função de ligação, pelo que o seu estudo deve ser aprofundado.

2.2.5 Desempenho do modelo

Terminada a validação dos pressupostos do modelo, é necessário avaliar o desempenho do mesmo. O desempenho do modelo pode ser avaliado utilizando diferentes metodologias, sendo necessário avaliar a qualidade de ajustamento aos dados e a capacidade de predição do modelo. Esta avaliação é realizada utilizando os dados da amostra 1, sendo assim possível avaliar o desempenho do modelo para um novo conjunto de dados.

Calibração

A calibração ou adequação do modelo é uma medida da qualidade de ajuste do modelo aos dados. Para o modelo de regressão logística, um dos testes mais utilizados para avaliar a qualidade de ajustamento é o teste de Hosmer e Lemeshow [7] baseado no agrupamento das probabilidades estimadas pelo modelo, cujas hipóteses são:

$$H_0 : \text{O modelo é adequado} \text{ vs } H_A : \text{O modelo não é adequado}$$

O primeiro passo é ordenar as n probabilidades estimadas pelo modelo, seguido do agrupamento das mesmas que pode ser efectuado de duas formas:

1. Baseado nos decis das probabilidades estimadas;
2. Com pontos de corte pré-definidos nos valores $m/10$, $m = 1, 2, \dots, 9$.

Enquanto que para o primeiro método serão sempre formados $g = 10$ grupos, o mesmo pode não acontecer para segundo método, sendo que no caso de o número de grupos ser inferior a 6, o teste perde a capacidade para rejeitar a hipótese nula, não se devendo utilizar este método nesta situação. Para os dois métodos de agrupamento, a estatística de teste C_v é definida por

$$C_v = \sum_{m=1}^g \frac{(o_m - e_m)^2}{n_m \bar{\pi}_m (1 - \bar{\pi}_m)} \quad (2.2.9)$$

em que m representa cada um dos g grupos, n_m o número de observações dentro do grupo m , o_m o número de respostas positivas dentro do grupo m ($o_m = \sum_{i=1}^n y_i$), e_m o valor esperado de respostas dentro do grupo m assumindo que o modelo é adequado ($e_m = \sum_{i=1}^n \hat{\pi}_i$) e $\bar{\pi}_m$ a probabilidade média estimada de cada grupo m ($\bar{\pi}_m = \sum_{i=1}^n \hat{\pi}_i / n_m$). Sob a hipótese nula, C_v tem distribuição Qui-Quadrado com g graus de liberdade ⁴. Factores como observações discordantes, função de ligação não adequada ou não inclusão de outros factores de risco relevantes podem condicionar a adequação do modelo aos dados, levando à rejeição da hipótese nula.

Discriminação

A discriminação ou capacidade do modelo classificar correctamente os utentes, é uma metodologia utilizada para medir a capacidade de predição do modelo através da curva ROC (Receiver Operating Characteristic). Esta curva representa a sensibilidade (probabilidade de se detectarem os verdadeiros positivos) contra a especificidade (probabilidade

⁴ g graus de liberdade em vez de $g - 2$, uma vez que estamos a utilizar um conjunto externo de dados para validar o modelo [7], a amostra 1.

de se detectarem os verdadeiros negativos), permitindo estudar a sua variação para diferentes pontos de corte ⁵. A área sob a curva ROC, denominada de AUC, indica a capacidade do modelo discriminar correctamente os utentes, variando entre 0 e 1.

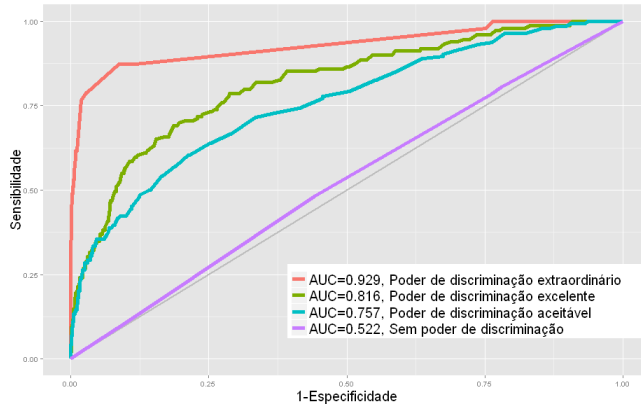


Figura 2.5: Curvas ROC

Na figura 2.5 estão representadas 4 curvas ROC correspondentes a 4 modelos distintos que, com o auxílio da tabela 2.3 [7], permitem interpretar o poder de discriminação dos mesmos com base nos valores da AUC.

<i>AUC</i>	Diagnóstico
$AUC \approx 0.5$	Sem poder de discriminação
$0.7 \leq AUC < 0.8$	Discriminação aceitável
$0.8 \leq AUC < 0.9$	Discriminação excelente
$AUC \geq 0.9$	Discriminação extraordinária

Tabela 2.3: Poder de discriminação

Graficamente, quanto mais a curva ROC se aproximar do ponto de coordenadas (0,1), maior é a área entre a curva e o eixo das abcissas, aumentando o poder de discriminação do modelo. Quando a curva ROC se aproxima da bissetriz dos quadrantes ímpares, o modelo não tem poder de discriminação, correspondendo a uma classificação aleatória. É ainda importante referir que o facto de um modelo ter um bom ajustamento aos dados não implica que o seu desempenho na classificação dos indivíduos seja igualmente bom, aumentando assim a importância da realização desta análise.

⁵Valores a partir dos quais se classificam os utentes como positivos ou negativos, com base na probabilidade estimada pelo modelo.

2.2.6 Interpretação do modelo

A interpretação da associação entre a variável resposta e as covariáveis presentes no modelo final é feita através da razão de chances ou *odds ratio* (OR). O OR permite quantificar a relação existente entre as covariáveis e a chance do acontecimento de interesse ocorrer.

Considere-se o modelo para a variável resposta Y , que representa o desfecho dos utentes diagnosticados com uma determinada doença ($Y = 0 \rightarrow$ sobrevivência, $Y = 1 \rightarrow$ morte), com $Y_i \in B(1, \pi_i)$, associando-se o valor esperado da variável resposta e as covariáveis através da função de ligação *logit*:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \cdot \text{Idade} + \beta_2 \cdot \text{Sexo} + \beta_3 \cdot (\text{Idade} : \text{Sexo}) \quad (2.2.10)$$

em que a Idade é uma variável contínua, Sexo uma variável dicotómica (0 - Feminino, 1 - Masculino) e Idade:Sexo representa a interacção entre as duas variáveis. Como referido anteriormente,

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (2.2.11)$$

que é o equivalente a $\log(\text{chance}_i)$ ou $\log(\text{odds}_i)$. Portanto, para calcular a chance ou *odds* do utente i dadas as suas características, basta aplicar a função exponencial aos coeficientes do modelo:

$$\text{chance}_i = \exp(\beta_0 + \beta_1 \cdot \text{Idade} + \beta_2 \cdot \text{Sexo} + \beta_3 \cdot (\text{Idade} : \text{Sexo})) \quad (2.2.12)$$

Porém, em termos de interpretação, a chance por si só não apresentam grandes vantagens. Por outro lado, com o cálculo da razão de chances é possível estabelecer uma comparação entre os utentes que possuem ou não diferentes características.

Por exemplo, o OR para utentes de 50 anos de diferentes sexos é dado por:

$$OR = \frac{\text{odds}|\text{sexo masculino}, 50 \text{ anos}}{\text{odds}|\text{sexo feminino}, 50 \text{ anos}} = \frac{\exp(\beta_0 + \beta_1 \cdot 50 + \beta_2 + \beta_3 \cdot 50)}{\exp(\beta_0 + \beta_1 \cdot 50)} = \exp(\beta_2 + \beta_3 \cdot 50) \quad (2.2.13)$$

Para $\widehat{\beta}_2 = 0.45$ e $\widehat{\beta}_3 = 0.03$, $\exp(0.45 + 0.03 \cdot 50) = 7.03$, ou seja, a chance de morte de um utente de 50 anos do sexo masculino é 7.03 vezes superior à chance de morte de um utente de igual idade do sexo feminino. Os coeficientes do modelo permitem antecipar o sentido da associação das covariáveis com a variável resposta. Coeficientes negativos são factores de protecção em relação ao acontecimento de interesse e coeficientes positivos são factores de risco em relação ao mesmo.

No exemplo anterior, fixou-se a idade dos utentes. Porém, é possível calcular o OR fazendo variar a idade dos utentes, controlando os restantes factores. Para os utentes do sexo masculino, para cada variação de t anos na idade dos utentes o OR é dado por $\exp(\beta_1 \cdot t + \beta_3 \cdot t)$, sendo que para os utentes do sexo feminino o valor da interacção é nulo e o OR é dado por $\exp(\beta_1 \cdot t)$. Para $\widehat{\beta}_1 = 0.014$ e $t = 10$, $\exp(0.014 \cdot 10) = 1.15$, o que significa que para cada aumento de 10 anos na idade dos utentes do sexo feminino, a chance de morte aumenta 1.15 vezes (aumenta 15%).

A interpretação do OR é realizada em comparação com o valor 1, sendo que um OR superior a 1 indica um aumento da chance do acontecimento de interesse ocorrer, e valores de OR no intervalo (0,1) representam uma diminuição da chance do acontecimento de interesse ocorrer. Assim, é também importante calcular o intervalo de confiança para o OR de cada factor. Se o intervalo contiver o valor 1, significa que não existe diferença significativa entre as chances de realização do acontecimento de interesse para esse factor. O intervalo de confiança para o OR é dado por:

$$\left[\exp\left(\widehat{\beta}_k - EP_k \cdot z_{1-\frac{\alpha}{2}}\right), \exp\left(\widehat{\beta}_k + EP_k \cdot z_{1-\frac{\alpha}{2}}\right) \right] \quad (2.2.14)$$

onde EP_k representa o erro padrão de $\widehat{\beta}_k$, $z_{1-\frac{\alpha}{2}}$ o quantil de probabilidade $1 - \frac{\alpha}{2}$ da distribuição $Normal(0, 1)$ e α o nível de significância.

Como referido no início deste capítulo, a metodologia utilizada é uma metodologia bastante usual, com uma aplicação morosa mas relativamente acessível utilizando o software **R** como ferramenta para a construção de modelos de regressão logística.

Durante a construção dos modelos de risco, as principais dificuldades encontradas prendiam-se com a selecção de possíveis factores de risco para os diversos indicadores, a escassez de observações de alguns desses factores ou a baixa proporção dos acontecimentos de interesse que dificultam a associação com os factores de risco.

Futuramente, a construção e/ou actualização dos modelos actuais para os indicadores de resultado no âmbito da avaliação dos cuidados de saúde a nível nacional, deverá utilizar os passos e definições enumerados neste capítulo, assim como as funções enunciadas no Anexo A, uniformizando assim o processo de construção deste tipo de modelos.

3

Valores de referência

Como descrito anteriormente, o projecto compreende 2 objectivos sendo que neste capítulo será descrita a metodologia aplicada para dar resposta ao segundo objectivo, designadamente a implementação de valores de referência (*target*) para os indicadores de processo de suporte à avaliação dos cuidados de saúde nacionais. Contrariamente ao primeiro objectivo, em que a metodologia utilizada se manteve, neste objectivo pretende-se determinar uma metodologia alternativa válida, que minimize as limitações da metodologia actual.

3.1 Metodologia actual

Os indicadores de processo referem-se à prestação de cuidados de saúde aos utentes. São medidas directas e de fácil interpretação, que avaliam a concordância de determinadas práticas realizadas pelos prestadores de cuidados de saúde com as respectivas regras de conduta (*guidelines*).

Os indicadores de processo estão presentes nas várias áreas de excelência clínica, como Obstetrícia, Enfarte Agudo do Miocárdio, Pediatria, Unidade de Cuidados Intensivos, entre outras. Exemplos de indicadores de processo:

- Administração de ácido acetilsalicílico à chegada ao hospital para utentes com diagnóstico de Enfarte Agudo do Miocárdio;
- Profilaxia de tromboembolismo venoso para utentes com diagnóstico de Acidente Vascular Cerebral;
- Profilaxia da úlcera de stress para utentes na Unidade de Cuidados Intensivos;
- Administração de antibiótico profiláctico na hora anterior à cirurgia a utentes submetidos a cirurgias cardíacas;
- Administração pré-natal de esteróides.

Para avaliar o desempenho dos prestadores é calculado um valor observado para cada prestador, que corresponde à proporção de intervenções adequadas realizadas pelo mesmo, isto é,

$$\text{valor observado}_k = \frac{\# \text{ intervenções adequadas do prestador } k}{\# \text{ oportunidades de intervenções adequadas do prestador } k} \quad (3.1.1)$$

A metodologia amplamente utilizada, definida pela JCI, compara os valores observados dos diferentes prestadores com um valor de referência. Este valor é dado pela proporção global de intervenções adequadas, isto é,

$$\text{valor de referência} = \frac{\# \text{ intervenções adequadas de todos os prestadores}}{\# \text{ oportunidades de intervenções adequadas de todos os prestadores}} \quad (3.1.2)$$

Por outras palavras, é comparado o desempenho individual dos prestadores com o desempenho médio global. Esta comparação é feita através de um teste de comparação de proporções, dando origem a uma classificação em 3 níveis:

- Se não existir diferença significativa entre o valor observado_k e o valor de referência, o prestador *k* é classificado como “dentro da média”;
- Se existir diferença significativa entre o valor observado_k e o valor de referência, o prestador *k* é classificado como “abaixo da média” ou “acima da média”, dependendo do sentido de melhoria.

3.2 Limitações da metodologia actual

Antes de trabalhar na metodologia alternativa, é importante estudar as limitações da metodologia actual, de forma a tentar que as mesmas não se repitam na nova metodologia. Algumas das principais limitações identificadas prendem-se com o próprio teste de comparação de proporções.

Seja X_k o número de intervenções adequadas realizadas pelo prestador *k* em n_k oportunidades, então X_k segue uma distribuição *Binomial*(n_k, p_k), em que p_k corresponde à proporção desconhecida de intervenções adequadas realizadas pelo prestador *k* e $\hat{p}_k = \frac{X_k}{n_k}$ o seu estimador pontual.

Seja X_g o número de intervenções adequadas realizadas por todos os prestadores em n_g oportunidades, então X_g segue uma distribuição *Binomial*(n_g, p_g), em que p_g corresponde à proporção desconhecida de intervenções adequadas realizadas por todos os prestadores e $\hat{p}_g = \frac{X_g}{n_g}$ o seu estimador pontual.

O teste de hipóteses para a diferença entre proporções, realiza-se sob as seguintes hipóteses:¹

$$H_0 : p_k = p_g \text{ vs } H_A : p_k \neq p_g$$

Estatística de teste:

$$Z = \frac{\hat{p}_k - \hat{p}_g}{\sqrt{\frac{\hat{p}_k(1-\hat{p}_k)}{n_k} + \frac{\hat{p}_g(1-\hat{p}_g)}{n_g}}} \stackrel{a}{\sim} Normal(0, 1) \text{ se } H_0 \text{ verdadeira.} \quad (3.2.1)$$

¹O teste de comparação de proporções assume que as duas populações em estudo são independentes, o que não acontece neste caso. Porém, dada a grande quantidade de prestadores, assume-se que existe independência entre as duas populações uma vez que retirando um dos prestadores do estudo, a proporção de referência sofre alterações mínimas. Estas alterações aumentam à medida que o número de prestadores em estudo diminui, sendo também esta uma das limitações da metodologia actual.

Para um nível de significância α , a hipótese nula é rejeitada se $|z| \geq z_{1-\frac{\alpha}{2}}$.

Se todos os prestadores cumprirem a 100% as regras de conduta definidas, então $p_k = 1$ para qualquer k , o que implica $p_g = 1$. Nestas condições, é impossível determinar o valor da estatística de teste, uma vez que o denominador da mesma é igual a zero. O mesmo acontece se todos os prestadores não cumprirem as regras de conduta, isto é, se $p_k = 0$ para qualquer k , o que implica $p_g = 0$.

Porém, as limitações não surgem apenas quando as proporções tomam os valores extremos, mas também quando se aproximam demasiado dos mesmos [18], sendo normalmente definido o limite para a utilização deste teste quando $np > 5$ e $n(1-p) > 5$. Assim, pode-se concluir que, quando a maioria dos prestadores apresentam desempenhos muito bons ou muito maus, o teste de comparação de proporções pode não ter a capacidade para distinguir estes prestadores dos que apresentam desempenhos razoáveis.

Outra limitação desta metodologia encontra-se no facto do valor de referência não ser fixo, variando em cada avaliação consoante os desempenhos dos prestadores. Um prestador que numa avaliação seja classificado como “acima da média”, na avaliação seguinte pode ser classificado como “abaixo da média”, mesmo que o seu desempenho tenha sido melhor.

Pode ainda identificar-se uma limitação na perspectiva de evolução.[19] Como referido anteriormente, o objectivo fundamental deste tipo de avaliações é promover uma cultura de melhoria contínua dos cuidados de saúde prestados. Se se considerar que esta melhoria ocorre, utilizar a média global como o valor de referência a atingir conduz a uma estagnação da evolução pretendida. Por outro lado, se para um determinado indicador os prestadores apresentarem desempenhos fracos e semelhantes, o valor de referência também será fraco, não podendo ser visto como um objectivo viável na perspectiva de evolução.

3.3 Objectivos da metodologia alternativa

Tendo em conta as limitações identificadas na metodologia actual, a metodologia alternativa deve reunir um conjunto de características:

- Ser aplicável a todos os indicadores de processo;

- Definir valores de referência fixos, de agora em diante denominados de valores de referência, que permitam classificar os prestadores em 3 níveis e que possam ser ajustados de forma a permitir a evolução dos prestadores de cuidados de saúde;
- Ter em conta que os prestadores têm dimensões diferentes, isto é, que o número de utentes recebidos pelos mesmos varia, sendo importante introduzir intervalos de confiança para a proporção de intervenções adequadas de cada prestador;
- Minimizar as limitações encontradas na metodologia actual.

3.4 Metodologias alternativas

Como referido anteriormente, são raras as metodologias que diferem da metodologia introduzida pela JCI. Na tentativa de corrigir algumas das limitações identificadas, a própria JCI iniciou uma reformulação à metodologia original em 2009, com a introdução de valores de referência.[2]

Assim, em vez do teste de comparações habitual, a JCI introduz os valores de referência como sendo os limites dos intervalos de confiança para a proporção global de intervenções adequadas. Esta alteração continua a não resolver a limitação que surge quando a proporção global é muito baixa ou muito elevada, não existindo capacidade para distinguir os prestadores com melhores ou piores desempenhos. Para precaver esta situação, foram definidas algumas regras ²:

- Se a proporção global for superior a 0.95, os valores de referência (inferior e superior) em vez de tomarem os valores dos limites do intervalo de confiança para a proporção global, são ambos iguais a 0.95. Assim, o valor de referência superior é reduzido de forma a poder existir uma distinção entre os prestadores “acima da média” e os prestadores “dentro da média”, assumindo-se que prestadores com proporção de intervenções adequadas significativamente maiores que 0.95 são “acima da média”;
- Se a proporção global for inferior a 0.95 e superior a 0.90, então o valor de referência superior é igual a 0.95 e o valor de referência inferior igual à proporção global;
- Se a proporção global for inferior a 0.90, então o valor de referência inferior é igual à proporção global e o valor de referência superior igual à média entre a proporção global e 1.

²Estas regras são utilizadas quando o sentido de melhoria é crescente. Para a situação em que o sentido de melhoria é decrescente, as regras são obtidas de forma análoga.

Apesar desta metodologia conseguir corrigir a limitação da proporção global muito elevada/baixa, trabalha com valores de referência que variam em cada avaliação e que dependem do desempenho global dos prestadores. Define ainda 0.95 como o valor de referência superior sempre que a proporção global lhe seja superior, para todos os indicadores de processo, assumindo assim que este valor é a referência máxima para indicadores completamente diferentes.

Analisando de uma forma mais global o tipo de metodologias que podem ser aplicadas, podem-se definir dois grupos: metodologias que utilizam os dados e metodologias que utilizam fontes de informação externas.

Utilizando os dados

Dentro das metodologias que utilizam os dados, foi ainda estudada uma alternativa às duas metodologias já referidas da JCI: utilizar os dados de avaliações anteriores de forma a definir os valores de referência a utilizar nas avaliações seguintes.

O primeiro passo desta metodologia é ajustar uma distribuição aos dados, sendo que, tratando-se da modelação de uma variável aleatória que toma valores no intervalo $[0, 1]$, foram ponderados dois modelos:

- Um modelo *Uniforme*(0, 1), em que a proporção média de intervenções adequadas seria de 0.5 para todos os prestadores;
- Um modelo *Beta*(α, β), em que a proporção média de intervenções adequadas depende de 2 parâmetros, podendo tomar valores diferentes no intervalo $(0, 1)$, variando entre os prestadores.

A utilização do segundo modelo é a escolha mais lógica e que melhor pode modelar a realidade, sendo os parâmetros α e β estimados pelo método da máxima verosimilhança ³.

Porém, esta metodologia não pode ser aplicada sempre que a informação disponível de anos anteriores não seja suficiente para determinar a distribuição (por exemplo, para novos indicadores ou indicadores com poucos prestadores em avaliação), ou que simplesmente não seja possível ajustar a distribuição *Beta* aos dados. Existe ainda a limitação

³A estimação dos parâmetros da distribuição *Beta*(α, β) pelo método de máxima verosimilhança, com α e β desconhecidos, não é possível analiticamente. Os parâmetros têm de ser estimados iterativamente, recorrendo-se normalmente ao método Newton-Raphson.

identificada anteriormente de dependência dos dados na determinação dos valores de referência, que leva à limitação da perspectiva de evolução.

Posto isto, tornou-se evidente que as metodologias que utilizam os dados para determinar valores de referência, apresentam limitações semelhantes. Assim, procurou-se uma metodologia diferente que não dependesse dos dados para definir os valores de referência e que cumprisse com os objectivos propostos.

Utilizando fontes de informação externa

Na primeira metodologia equacionada, o objectivo consistia na definição de valores de referência com base na literatura, através de estudos, artigos ou comunidades científicas que fornecessem esses valores para cada indicador. Considerando que o número de indicadores de processo é elevado, que aumenta sempre que uma nova área clínica é avaliada e que não existem estudos individuais sobre alguns destes indicadores, seria extremamente difícil definir valores de referência adequados à realidade nacional para todos eles.

Assim, pensou-se numa metodologia que pudesse ser directamente aplicada a todos os indicadores e que, uma vez que os indicadores de processo são definidos por um conjunto de especialistas, a opinião desses mesmos especialistas pudesse contribuir para determinar os valores de referência. Através da recolha de opinião de especialistas das diferentes áreas clínicas sobre os indicadores de processo de cada área, pretende-se definir valores de referência que possam ser utilizados como um objectivo real a atingir, que possibilitem uma evolução contínua dos prestadores de cuidados de saúde e que, ao mesmo tempo, permitam criar um sistema de classificação dos mesmos.

3.5 Recolha e combinação de opiniões

A recolha de opinião de especialistas é utilizada em diversas áreas, com especial ênfase para a análise de risco. O processo habitual passa por uma entrevista ou questionário efectuado aos especialistas que fornecem a sua opinião sobre o acontecimento em estudo. A recolha de opinião de especialistas pode ser realizada utilizando uma abordagem comportamental, em que se tenta atingir um consenso entre os especialistas através de algum tipo de interacção entre os mesmos, ou utilizando uma abordagem matemática, com a construção de modelos probabilísticos para representar as diferentes opiniões.[20]

3.5.1 Abordagem Comportamental

A abordagem comportamental requer que haja algum tipo de interacção entre os especialistas, através de reuniões de grupo ou qualquer outro tipo de troca de informação e conhecimento. O grupo de especialistas pode avaliar probabilidades, fazer previsões ou apenas discutir assuntos relevantes sem uma avaliação formal. O objectivo é que o grupo de especialistas chegue a um consenso ou acordo para a solução.

Uma das abordagens mais antigas, introduzida em 1969, é o método Delphi.[21] Neste método, os especialistas começam por determinar individualmente a distribuição de probabilidade que melhor se ajusta à sua opinião. Num segundo momento, as diferentes opiniões são divulgadas entre os diversos especialistas, mantendo o anonimato, existindo a possibilidade de reformularem as suas opiniões iniciais mas não existindo um interacção directa entre os diferentes especialistas. Este processo repete-se até que seja atingido o consenso entre as opiniões dos especialistas, isto é, que não exista um grande distanciamento entre as diferentes opiniões. Numa perspectiva ideal, o consenso deverá ser atingido após algumas rondas, mas na prática isto raramente acontece. Depois de atingido o consenso, o último passo desta metodologia passa por combinar as diferentes opiniões através de métodos matemáticos.

Outra abordagem possível é a Técnica do Grupo Nominal[22], introduzida em 1975. De forma muito resumida, nesta técnica os especialistas começam por determinar individualmente a distribuição de probabilidade que melhor se ajusta à sua opinião e depois apresentam-na ao restante grupo. Segue-se uma discussão entre os especialistas sobre as diferentes opiniões, existindo também a possibilidade dos especialistas reformularem as suas opiniões. Atingido o consenso, as diferentes opiniões são combinadas através de métodos matemáticos, como no método Delphi.

3.5.2 Abordagem Matemática

Existe uma variedade de abordagens matemáticas bastante superior face às abordagens comportamentais. Estas podem variar desde simples técnicas de agregação de estatísticas descritivas, a abordagens axiomáticas ou bayesianas. As técnicas de agregação de estatísticas descritivas são práticas de aplicar e bastante intuitivas, como o método Cumulativo [23] cuja distribuição de opiniões combinada corresponde à média dos percentis das distribuições determinadas pelos especialistas, ou ainda outros métodos mais simples que não utilizam toda a informação fornecida pelos especialistas, mas que pela sua sim-

plicidade de aplicação e de interpretação são utilizados.

Um método bastante apelativo dentro da abordagem axiomática é o *linear opinion pool*. [24] Este método, também ele prático de aplicar e de fácil percepção, é uma combinação linear ponderada das densidades de probabilidade atribuídas pelos especialistas, isto é,

$$f(\theta) = \sum_{i=1}^n w_i f_i(\theta) \quad (3.5.1)$$

em que n representa o número de especialistas, w_i o peso atribuído a cada opinião (com $w_i \in [0, 1]$ e $\sum_{i=1}^n w_i = 1$), θ o parâmetro em estudo (desconhecido), $f_i(\theta)$ a função densidade de probabilidade que corresponde à opinião do especialista i e $f(\theta)$ a função densidade de probabilidade que corresponde à opinião combinada, determinada pela média aritmética ponderada das densidades de probabilidade dos especialistas, tendo a vantagem de satisfazer importantes axiomas.

Outro método bastante utilizado é o *logarithmic opinion pool*, uma combinação geométrica ponderada das densidades de probabilidade atribuídas pelos especialistas, isto é,

$$f(\theta) = k \prod_{i=1}^n f_i(\theta)^{w_i} \quad (3.5.2)$$

em que k é uma constante normalizadora e w_i está sujeita às restrições supra-mencionadas. De uma maneira geral, o segundo método produz distribuições menos dispersas que o primeiro, mas com a desvantagem de não traduzir de uma forma tão fiel as opiniões dos especialistas.

Existem outros métodos matemáticos para combinar distribuições de probabilidade baseadas nas opiniões de especialistas, uns mais simples que utilizam a mediana das estimativas fornecidas para construir a opinião combinada, não utilizando toda a informação disponível, e outros mais complexos como os métodos bayesianos. Uma revisão mais aprofundada da literatura leva a uma conclusão: de uma forma geral, no que diz respeito à combinação de opiniões de especialistas, métodos de agregação simples têm um desempenho melhor do que métodos mais complexos. [20, 25, 26, 27]

Por outro lado, a comparação entre as diferentes abordagens comportamentais leva a diferentes resultados, com desempenhos variáveis consoante as situações em estudo. Por fim, a comparação entre as abordagens comportamentais e matemáticas leva a uma conclusão de desempenhos equilibrados, com as regras matemáticas a terem uma ligeira

vantagem.[20] Em qualquer dos casos, existe um problema inerente à recolha de opinião de diferentes especialistas: as opiniões podem divergir. Esta divergência, que pode condicionar o estudo, deve ser encarada como uma mais valia, em que cada opinião acrescenta algo de novo à opinião combinada. Na verdade, se não existisse divergência nas opiniões bastava recolher a opinião de um especialista. Assim, o princípio fundamental destas abordagens é que a opinião de diversos especialistas acrescenta mais informação relevante do que a opinião de um único especialista.

Posto isto, a abordagem utilizada neste trabalho é uma abordagem matemática, com menos limitações de aplicação do que a abordagem comportamental, utilizando um método de fácil interpretação, prático em termos computacionais e com excelente desempenho, o método *linear opinion pool*.

3.5.3 Recolha

Os diferentes métodos de combinação de opiniões têm um factor em comum: para determinar a opinião combinada é necessário obter as opiniões individuais dos especialistas em termos probabilísticos. Para os diferentes indicadores de processo, a variável em estudo é definida como a proporção de intervenções adequadas esperadas num prestador de referência ⁴, θ , com $\theta \in [0, 1]$. Compreendendo que não é fácil ao especialista determinar uma distribuição para a proporção em estudo, é importante definir uma estratégia que possa traduzir a opinião dos especialistas em modelos de probabilidade. Este objectivo pode ser concretizado recorrendo ao modelo *PERT*. [28]

A distribuição *PERT* é uma variação da distribuição *Beta* que requer 3 parâmetros para ser definida: mínimo (a), moda (b) e máximo (c) (tal que $a < b < c$). Esta distribuição é exclusivamente utilizada para modelar opiniões de especialistas, tendo sido inicialmente desenvolvida para estimar a duração de projectos com base nas estimativas fornecidas pelos especialistas do tempo mínimo, mais provável e máximo para a duração das tarefas que os compunham. A sua função densidade de probabilidade é dada por:

$$f(\theta) = \frac{(\theta - a)^{\alpha-1}(c - \theta)^{\beta-1}}{B(\alpha, \beta)(c - a)^{\alpha+\beta-1}}, \quad a < \theta < c \quad (3.5.3)$$

⁴Define-se prestador de referência como sendo um modelo teórico que contempla os valores aceitáveis para cada indicador, adaptado à realidade nacional.

com $\alpha = \frac{6(\mu - a)}{c - a}$, $\beta = \frac{6(c - \mu)}{c - a}$, $\mu = \frac{a + 4b + c}{6}$ e $B(\alpha, \beta)$ representa a função Beta.

Este modelo transforma a opinião de cada especialista em modelos probabilísticos, em que cada opinião i corresponde a uma distribuição *PERT*, sendo $f_i(\theta)$ a função densidade de probabilidade correspondente à opinião do especialista i .

Para cada indicador de processo, cada especialista seleccionado determina o valor máximo, mínimo e mais provável ⁵ para a proporção de intervenções adequadas num prestador de referência. O exemplo 3.5.1 demonstra como a recolha de opinião pode ser efectuada de forma a definir o modelo *PERT*.

Exemplo 3.5.1. *As regras de conduta relativas à área clínica de Enfarte Agudo do Miocárdio consideram que os utentes-padrão que apresentem este diagnóstico, devem ser medicados com a administração de ácido acetilsalicílico (aspirina) à chegada ao hospital. Porém, este procedimento não se verifica em todos os utentes-padrão. Na sua opinião, qual o valor:*

- *mais optimista (máximo) da proporção de utentes-padrão que recebem aspirina à chegada ao hospital?*
- *mais pessimista (mínimo) da proporção de utentes-padrão que recebem aspirina à chegada ao hospital?*
- *mais provável (moda) da proporção de utentes-padrão que receberam aspirina à chegada ao hospital?*

3.5.4 Simulação

Para cada indicador é recolhida a opinião dos n especialistas, dando origem a n distribuições *PERT*(a, b, c). Utilizando a função *rpert* do pacote *mc2d* [29], que permite a geração de números aleatórios com distribuição *PERT*, é possível simular amostras que correspondem às opiniões de cada especialista. Supondo que 5 especialistas responderam às questões introduzidas no exemplo anterior (tabela 3.1):

⁵Os especialistas têm tendência para estimar em primeiro lugar o valor que pensam ser mais provável, e depois ajustar o máximo e o mínimo a esse valor. O problema é que esses ajustamentos são, na maioria das vezes, insuficientes para incluir todos os possíveis valores que podem de facto ocorrer, dando a ideia de que existe uma ancoragem ao primeiro valor estimado. Para minimizar este efeito, os especialistas devem definir os valores pela ordem: máximo, mínimo, moda.[28]

Especialista	Máximo	Mínimo	Moda
A	1	0.8	0.95
B	0.95	0.7	0.9
C	0.92	0.6	0.85
D	0.9	0.6	0.75
E	1	0.7	0.9

Tabela 3.1: Opiniões dos especialistas

O passo seguinte utiliza a função *rpert* para gerar amostras aleatórias de dimensão 20000 com distribuição *PERT*, com os parâmetros especificados pelos especialistas. Para cada opinião tem-se:

```
Especialista i <- rpert (20000, mínimo, moda, máximo)
```

Calculando e representando graficamente as respectivas funções densidade, $f_i(\theta)$:

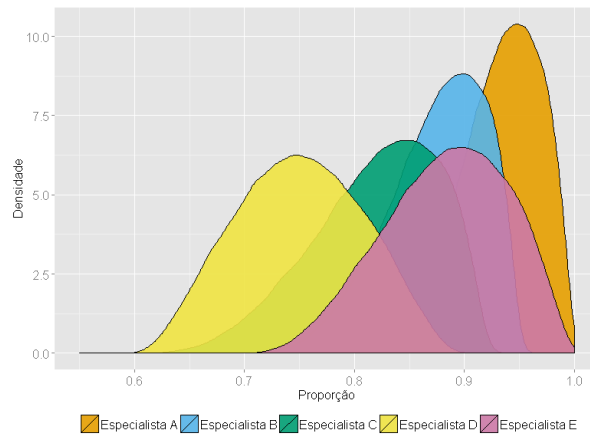


Figura 3.1: Opiniões dos especialistas - Funções Densidade de Probabilidade

As densidades representadas na figura 3.1 reflectem a opinião de cada especialista em termos probabilísticos e permitem perceber a forma como o modelo *PERT* distribui a probabilidade face aos parâmetros fornecidos pelos especialistas. A densidade de probabilidade atribuída a cada valor de θ varia entre os valores mínimos e máximos de cada densidade, sendo que estas tomam valores mais elevados à medida que se aproximam das respectivas modas e diminuem quando se aproximam dos extremos. A distribuição *PERT* pode ainda ser caracterizada em termos de simetria através da localização da moda, sendo simétrica quando a moda se encontra a igual distância do mínimo e do

máximo, simétrica positiva quando se aproxima mais do mínimo do que do máximo, e simétrica negativa no caso contrário.

3.5.5 Combinação de opiniões

Como referido anteriormente, o método *linear opinion pool* permite ponderar as opiniões dos especialistas atribuindo um peso w_i a cada opinião i . Assim, os pesos representam a qualidade relativa de cada especialista, sendo que quanto maior é o peso atribuído, mais precisa e fiável é a opinião do especialista. Porém, a determinação do peso a atribuir a cada especialista é bastante subjectiva, podendo ser determinada, por exemplo, pela experiência dos mesmos, pela “distância” da opinião de cada especialista em relação às restantes opiniões ou questionando o especialista sobre um determinado valor (conhecido) para avaliar a precisão e conhecimento do mesmo em relação à área de interesse, não existindo uma forma que se possa considerar mais correcta na determinação destes pesos.

Diferentes estudos concluem que a atribuição de pesos diferentes não conduzem a melhores resultados do que a atribuição de pesos iguais. [20, 30] Assim, na metodologia utilizada neste trabalho foram atribuídos pesos iguais a todos os especialistas, correspondendo a função densidade de probabilidade da opinião combinada a uma simples média aritmética das funções densidade de probabilidade de cada especialista.

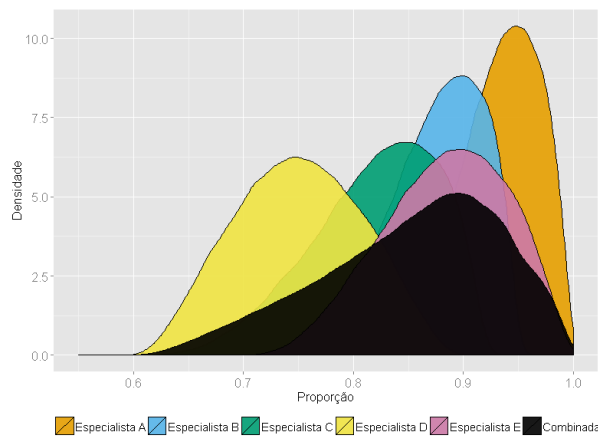


Figura 3.2: Opiniões dos especialistas (individuais e combinada) - Funções Densidade de Probabilidade

Na prática, a aplicação do método *linear opinion pool* com pesos iguais ao exemplo an-

terior, corresponde a agregar as 5 amostras geradas aleatoriamente numa única amostra, denominada de amostra combinada. Assim, é possível calcular a densidade da amostra combinada para cada valor da variável θ , representada graficamente na figura 3.2. A estimação das densidades que correspondem às opiniões individuais é feita no **R** pela função *density*, utilizando as amostras simuladas. Da mesma forma, é determinada a função de densidade correspondente à amostra combinada, sendo que esta não tem uma forma paramétrica conhecida.

3.6 Determinação dos valores de referência

Utilizando a amostra combinada, fruto da agregação das 5 amostras simuladas, é então possível determinar a sua função densidade de probabilidade, como referido anteriormente, e ainda a função de distribuição empírica para a proporção de intervenções adequadas da opinião combinada, através da função *ecdf*⁶. A partir da função de distribuição combinada, é possível determinar os valores de referência inferior e superior, *TI* e *TS* respectivamente, que servirão como valores de referência para distinguir os prestadores entre os 3 níveis de classificação.

De forma a permitir um ajustamento posterior destes valores em relação ao grau de exigência pretendido e da evolução esperada, os valores de referência são definidos com base em 2 quantis da amostra combinada. Assim, e partindo do pressuposto que existe uma melhoria contínua dos cuidados prestados e um conseqüente aumento das proporções de intervenções adequadas realizadas em cada prestador, é possível actualizar os valores de referência entre avaliações sempre que se considere necessário e não dependendo da variação dos dados como sucede na metodologia actual. Por exemplo, o *TS* numa determinada avaliação pode corresponder ao quantil 0.8 da opinião combinada. Existindo uma melhoria contínua dos cuidados prestados, poder-se-á actualizar este valor para um valor mais elevado, aumentando assim o nível de exigência para a classificação no melhor nível de qualidade, podendo *TS* passar a corresponder ao quantil 0.9 da opinião combinada.

⁶Ou da função *ewcdf* do pacote *spatstat* [31] que permite encontrar a função de distribuição empírica ponderada.

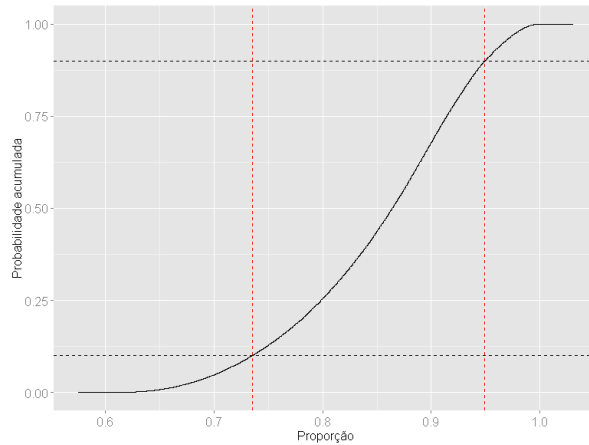


Figura 3.3: Opinião combinada e valores de referência - Função de Distribuição Empírica

Define-se o quantil de probabilidade p como sendo o número θ tal que $F(\theta) = p$, em que $F(\theta)$ representa a função de distribuição empírica combinada. Para o exemplo anterior, os valores de referência inferior e superior correspondem aos quantis 0.1 e 0.9 respectivamente. No **R**, os quantis são determinados recorrendo à função `quantile`, obtendo-se $TI = 0.735$ e $TS = 0.949$, representados pelas linhas a vermelho na figura 3.3.

3.6.1 Classificação

Utilizando esta metodologia, já não se podem considerar os níveis “abaixo da média”, “dentro da média” e “acima da média”, uma vez que já não existe uma comparação com a média global. Assim, os níveis desta metodologia são representados por 1, 2 e 3, em que 1 e 3 correspondem ao pior e ao melhor nível de classificação, respectivamente, e o nível 2 a um nível de qualidade intermédio.

A forma como os valores de referência foram construídos permite perceber que TI é a proporção mínima de intervenções adequadas necessárias para um prestador poder ser classificado como nível 2 e TS para poder ser classificado como nível 3. Porém, como referido anteriormente na secção 3.3, as diferentes dimensões dos prestadores devem ser tidas em conta, sendo necessário trabalhar com intervalos de confiança para a proporção de intervenções adequadas realizadas por cada prestador, p_k .

Assim, os limites do intervalo de $100(1 - \alpha)\%$ confiança para p_k são definidos por:

$$LI(X_k) = \frac{X_k}{n_k} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X_k}{n_k} \left(1 - \frac{X_k}{n_k}\right)}{n}}, \quad LS(X_k) = \frac{X_k}{n_k} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X_k}{n_k} \left(1 - \frac{X_k}{n_k}\right)}{n}} \quad (3.6.1)$$

em que $z_{1-\frac{\alpha}{2}}$ representa o quantil de probabilidade $1 - \frac{\alpha}{2}$ da distribuição $Normal(0, 1)$ e α o nível de significância.

Posto isto, e como prolongamento da definição anterior, é possível definir a regra de classificação dos prestadores em 3 níveis:⁷

- Se $LS(X_k) < TI \Rightarrow$ O prestador k é classificado como nível 1;
- Se $LI(X_k) > TS \Rightarrow$ O prestador k é classificado como nível 3;
- Caso contrário \Rightarrow O prestador k é classificado como nível 2.

Por outras palavras, se o limite inferior do intervalo de confiança para a proporção de intervenções adequadas realizadas pelo prestador k for superior ao valor de referência superior, então o prestador realiza um número de intervenções adequadas acima do que é esperado, pelo que é classificado como nível 3. Por outro lado, se o limite superior do intervalo de confiança for inferior ao valor de referência inferior, o prestador está aquém do esperado e é classificado como nível 1. Para os restantes casos, não existe uma diferença significativa entre a proporção de intervenções adequadas realizadas pelo prestador e o que é de esperar, sendo assim classificado como nível 2.

A aplicação da regra de classificação é facilmente perceptível com o auxílio da figura 3.4, onde estão representados os intervalos de confiança para p_k , $k = 1, 2, 3, 4$, em que as linhas horizontais dos intervalos de confiança correspondem aos limites $LI(X_k)$ e $LS(X_k)$, os pontos a negro à estimativa pontual \hat{p}_k e as linhas horizontais descontínuas a vermelho correspondem aos valores de referência $TI = 0.735$ e $TS = 0.949$.

⁷A regra definida é aplicável aos casos em que o sentido de melhoria é crescente. Para os indicadores em que o sentido de melhoria é decrescente, é necessário trocar os níveis de classificação 1 e 3 na regra de classificação.

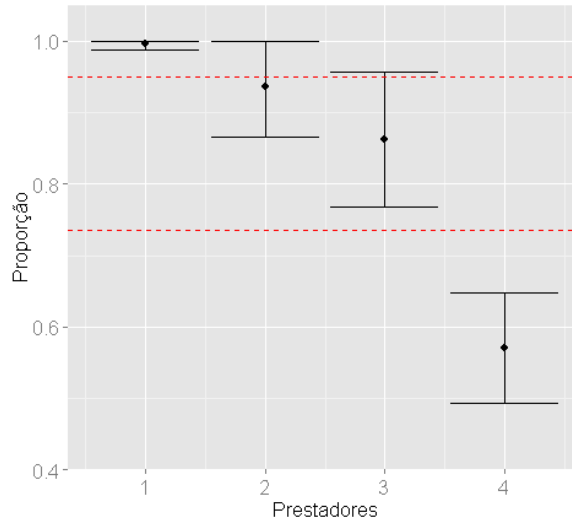


Figura 3.4: Classificação dos prestadores em 3 níveis

Sempre que existir uma intercepção entre o intervalo de confiança e o valor de referência inferior, superior ou ambos, o prestador é classificado como nível 2. Se o intervalo de confiança for completamente superior ao valor de referência superior, o prestador é classificado como nível 3. De forma análoga, se o intervalo de confiança for completamente inferior ao valor de referência inferior, o prestador é classificado como nível 1. Na figura 3.4, o prestador 1 é classificado como nível 3, os prestadores 2 e 3 são classificados como nível 2 e o prestador 4 é classificado como nível 1.

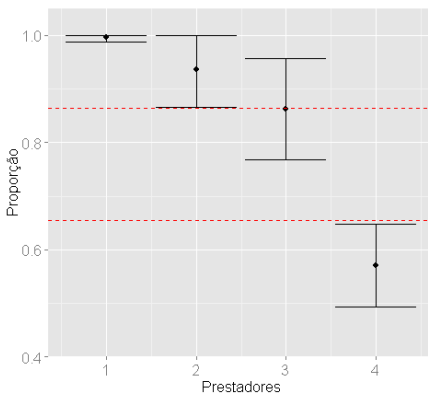


Figura 3.5: TI=0.655, TS=0.864

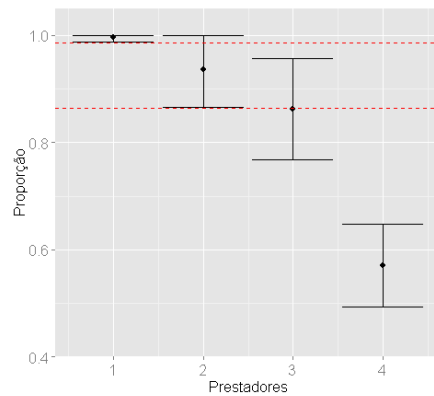


Figura 3.6: TI=0.864, TS=0.986

Como referido anteriormente, os valores de referência inferior e superior representados na

figura 3.4 correspondem aos quantis 0.1 e 0.9 da amostra combinada, respectivamente. As figuras 3.5 e 3.6 ilustram a variação na classificação dos mesmos prestadores representados anteriormente para diferentes valores de referência:

- Utilizando um critério menos exigente, baixando os limites para os prestadores serem classificados como nível 2 e 3, figura 3.5;
- Utilizando um critério mais exigente, aumentando os limites para os prestadores serem classificados como nível 2 e 3, figura 3.6.

Utilizando um critério menos exigente, em que é necessário uma proporção superior a 86% de administração de ácido acetilsalicílico para ser classificado como nível 3 e superior a 65% para ser classificado como nível 2, verifica-se que o prestador 2, inicialmente classificado como nível 2, é agora classificado como nível 3, mantendo-se a classificação inicial para os restantes prestadores. Utilizando um critério mais exigente, com proporções mais elevadas para os prestadores serem classificados como nível 2 e 3, verifica-se que a classificação inicial se mantém apesar do valor de referência superior ser muito elevado.

3.7 Decisão estatística

No desenvolvimento desta metodologia, surgiu a possibilidade de introduzir uma abordagem adicional que permitisse calcular o risco da regra de classificação utilizada face aos estados dos hospitais, aumentando assim as potencialidades da própria metodologia. Este objectivo pode ser atingido formulando o problema de classificação como um problema de decisão estatística, utilizando a regra de classificação definida anteriormente. Na estruturação de problemas de decisão, são importantes três elementos: acções, estados da natureza e consequências.

Acções

O primeiro passo a realizar consiste no levantamento de todas as acções possíveis, tipicamente representado por \mathcal{A} , em que cada $a \in \mathcal{A}$ representa uma acção diferente. As acções devem ser definidas atendendo a duas propriedades:

- exaustividade - não se devem ignorar ou não identificar todas as opções possíveis;
- exclusividade - evitar acções duplicadas ou que ocorram simultaneamente.

O conjunto de possíveis acções pode ser finito ou infinito sendo que, neste caso, é finito e composto por 3 acções: a_1 - classificar como nível 1, a_2 - classificar como nível 2 e a_3 - classificar como nível 3.

Estados da natureza

As diferentes acções podem ser tomadas considerando cenários diferentes, isto é, dependem do meio envolvente. As várias circunstâncias ou situações que caracterizam o meio envolvente em que a decisão é tomada, são designados por estados da natureza, tipicamente representado por Θ , em que cada $\theta \in \Theta$ representa um estado diferente. Os estados da natureza também são definidos tendo em conta as duas propriedades referidas para as acções.

O conjunto de estados da natureza pode ser finito ou infinito sendo que, neste caso, por simplificação admite-se ser finito e composto por 3 estados ⁸ que definem os verdadeiros estados dos prestadores: θ_1 - prestador de nível 1, θ_2 - prestador de nível 2 e θ_3 - prestador de nível 3. Porém, as decisões tomadas são efectuadas em condições de incerteza, não se conhecendo os verdadeiros estados em que os prestadores se encontram, embora se conheça por hipótese o conjunto Θ .

Função perda

O conjunto das consequências associa-se com o produto cartesiano $\mathcal{A} \times \Theta$, em que cada par ordenado (a, θ) designa uma consequência correspondente ao resultado da interacção da acção $a \in \mathcal{A}$ com o estado da natureza $\theta \in \Theta$. Estas consequências podem ser quantificadas definindo-se a função perda (ou perca), $L(a, \theta)$, que traduz o “custo” do resultado obtido ao escolher a acção $a \in \mathcal{A}$ quando o estado da natureza é $\theta \in \Theta$. [32]

Quando $\mathcal{A} = a_1, a_2, \dots, a_m$ e $\Theta = \theta_1, \theta_2, \dots, \theta_n$, a consequência ou perda de cada interacção acção-estado é designada por $L(a_i, \theta_j)$ ou L_{ij} , $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Tipicamente, é designado um “custo” positivo quando a interacção resulta numa perda, $L_{ij} > 0$, e um “custo” nulo quando se trata de um ganho, $L_{ij} = 0$, sendo a função perda representada em forma de matriz (tabela 3.2):

⁸Poderia-se admitir $\theta \in (0, 1)$.

\mathcal{A}/Θ	θ_1	θ_2	\dots	θ_n
a_1	L_{11}	L_{12}	\dots	L_{1n}
a_2	L_{21}	L_{22}	\dots	L_{2n}
\dots	\dots	\dots	\dots	\dots
a_m	L_{m1}	L_{m2}	\dots	L_{mn}

Tabela 3.2: Matriz perda

Desta forma, pode-se definir a função perda para o caso actual como (tabela 3.3:

\mathcal{A}/Θ	θ_1	θ_2	θ_3
a_1	0	1	2
a_2	1	0	1
a_3	2	1	0

Tabela 3.3: Matriz perda - caso de estudo

Esta função é também denominada de função perda absoluta, podendo ser definida analiticamente por $L(a_i, \theta_j) = |a_i - \theta_j|$. Assim, é atribuído um “custo” nulo quando os prestadores são classificados no nível em que se apresentam, incrementando o “custo” em uma unidade por cada nível de diferença entre o verdadeiro estado e a classificação atribuída. Por exemplo, um prestador cujo estado verdadeiro seja nível 1, ao ser classificado como nível 3 representa uma perda duas vezes superior ao de ser classificado como nível 2.

Risco

Para ser possível calcular o risco associado à regra de classificação escolhida, é necessário reformular alguns dos conceitos introduzidos. Assim, seja $(LI(X), LS(X))$ o intervalo de confiança para o parâmetro θ baseado em $X \sim Binomial(n, \theta)$ número de intervenções adequadas em n oportunidades,

$$LI(X) = \frac{X}{n} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X}{n} (1 - \frac{X}{n})}{n}}, \quad LS(X) = \frac{X}{n} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X}{n} (1 - \frac{X}{n})}{n}} \quad (3.7.1)$$

e sejam (TI, TS) os valores de referência inferior e superior previamente fixados.

A regra de classificação, agora designada por regra de decisão $d(X)$, é dada por:

- Se $LS(X) < TI$, classificar em 1;
- Se $LI(X) > TS$, classificar em 3;
- Caso contrário, classificar em 2.

A função risco (ou custo médio) da regra de decisão, é calculada para cada estado da natureza $\theta_j, j = 1, 2, 3$:

$$R(\theta_j) = \sum_{i=1}^m L(a_i, \theta_j)P[a_i|\theta_j] \quad (3.7.2)$$

isto é,

$$R(\theta_j) = L(a_1, \theta_j)P[a_1|\theta_j] + L(a_2, \theta_j)P[a_2|\theta_j] + L(a_3, \theta_j)P[a_3|\theta_j] \quad (3.7.3)$$

Pela regra de decisão vem

$$\begin{aligned} &= L(a_1, \theta_j)P[LS(X) < TI|\theta_j] + L(a_3, \theta_j)P[LI(X) > TS|\theta_j] \\ &+ L(a_2, \theta_j)(1 - P[LS(X) < TI|\theta_j] - P[LI(X) > TS|\theta_j]) \end{aligned} \quad (3.7.4)$$

Para calcular as probabilidades $P[a_i|\theta_j]$, é necessário determinar a distribuição de $LI(X)$ e de $LS(X)$. Uma vez que $LI(X)$ e $LS(X)$ são funções do estimador de máxima verosimilhança $\hat{\theta} = X/n$, sob condições gerais de regularidade $g(\hat{\theta}) = (LI(X), LS(X))$ tem distribuição aproximadamente $Normal(g(\theta), var(g(\theta)))$, onde $g(\cdot)$ é uma função contínua.

O valor médio e variância assintóticos de $g(\theta)$ são determinados recorrendo ao método delta de 1ª ordem.[33] Seja $\hat{\theta}$ o estimador de máxima verosimilhança de θ e $g(\theta)$ uma função diferenciável de θ , então

$$E(g(\hat{\theta})) \approx g(\theta), \quad var(g(\hat{\theta})) \approx var(\hat{\theta}) \cdot \left(\frac{d}{d\theta}g(\theta)\right)^2 \quad (3.7.5)$$

Aplicando o método delta a $LI(X)$ e $LS(X)$, obtêm-se os respectivos valores médios e

variâncias assintóticas:

$$\begin{aligned}
 E(LI(X)) &\approx \theta - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta(1-\theta)}{n}} \\
 E(LS(X)) &\approx \theta + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta(1-\theta)}{n}} \\
 \text{var}(LI(X)) &\approx \frac{\theta(1-\theta)}{n} \cdot \left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \cdot \frac{1-2\theta}{2\sqrt{\theta(1-\theta)}}\right)^2 \\
 \text{var}(LS(X)) &\approx \frac{\theta(1-\theta)}{n} \cdot \left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \cdot \frac{1-2\theta}{2\sqrt{\theta(1-\theta)}}\right)^2
 \end{aligned} \tag{3.7.6}$$

Porém, as probabilidades pretendidas são condicionais aos estados da natureza. Essa ligação pode ser atingida atribuindo valores a θ determinados pela opinião combinada dos especialistas. Como esta opinião é relativa à proporção de referência de intervenções adequadas realizadas nos hospitais, θ_1 , θ_2 e θ_3 tomam os valores correspondentes ao mínimo, mediana⁹ e máximo da amostra da opinião combinada, respectivamente. Assim, as distribuições de $LI(X)$ e $LS(X)$ condicionais aos valores de θ_j , $j = 1, 2, 3$ são dadas por:

$$\begin{aligned}
 LI(X)|\theta_j &\sim N\left(\theta_j - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta_j(1-\theta_j)}{n}}, \frac{\theta_j(1-\theta_j)}{n} \cdot \left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \cdot \frac{1-2\theta_j}{2\sqrt{\theta_j(1-\theta_j)}}\right)^2\right) \\
 LS(X)|\theta_j &\sim N\left(\theta_j + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta_j(1-\theta_j)}{n}}, \frac{\theta_j(1-\theta_j)}{n} \cdot \left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \cdot \frac{1-2\theta_j}{2\sqrt{\theta_j(1-\theta_j)}}\right)^2\right)
 \end{aligned} \tag{3.7.7}$$

Exemplo 3.7.1. Calculando θ_j , $j = 1, 2, 3$ com base na amostra de opiniões combinada referida anteriormente, obtém-se $\theta_1 = 0.607$, $\theta_2 = 0.864$ e $\theta_3 = 0.999$.

Considerando ainda os valores de referência definidos anteriormente, $TI = 0.735$ e $TS = 0.949$, a função perda absoluta, e fixando a dimensão da amostra ($n = 200$), é possível calcular o risco da regra de decisão para cada valor de θ .

⁹É usada a mediana em vez da moda porque a opinião combinada pode ser multimodal.

$$\begin{aligned}
 R(\theta_1) &= L(a_1, \theta_1)P[a_1|\theta_1] + L(a_2, \theta_1)P[a_2|\theta_1] + L(a_3, \theta_1)P[a_3|\theta_1] = \\
 &= 2 \cdot P[LI(X) > TS|\theta_1] + (1 - P[LS(X) \leq TI|\theta_1] - P[LI(X) > TS|\theta_1]) = 0.036 \\
 R(\theta_2) &= L(a_1, \theta_2)P[a_1|\theta_2] + L(a_2, \theta_2)P[a_2|\theta_2] + L(a_3, \theta_2)P[a_3|\theta_2] = 0.035 \\
 R(\theta_3) &= L(a_1, \theta_3)P[a_1|\theta_3] + L(a_2, \theta_3)P[a_2|\theta_3] + L(a_3, \theta_3)P[a_3|\theta_3] = 0
 \end{aligned}
 \tag{3.7.8}$$

Os riscos obtidos no exemplo anterior são praticamente nulos, o que seria de esperar uma vez que os valores de θ utilizados são, à partida, os valores de referência para as proporções de intervenções adequadas para cada nível de classificação.

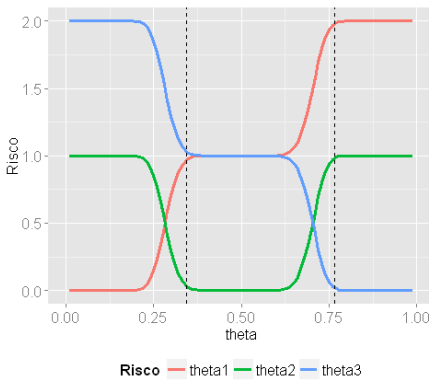


Figura 3.7: TI=0.345, TS=0.767

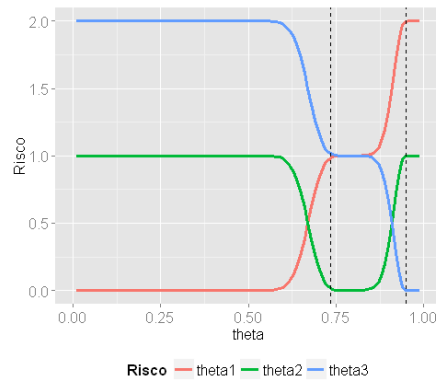


Figura 3.8: TI=0.735, TS=0.949

A variação do risco pode ser analisada fazendo variar os valores de θ_j no intervalo $(0, 1)$ e para diferentes valores de referência, obtendo-se resultados interessantes sobre a forma como o risco varia para os diferentes estados da natureza.

Nas figuras 3.7 e 3.8 é representado o risco da regra de decisão para θ_1 , θ_2 e θ_3 , para valores de $\theta \in (0, 1)$, e para diferentes valores de referência. Em ambos os casos, verifica-se que o risco da decisão para θ_1 aumenta à medida que os valores de θ aumentam. Em sentido inverso, o risco da decisão para θ_3 diminui à medida que os valores de θ aumentam. Para θ_2 , o risco mínimo é atingido quando os valores de θ estão entre os dois valores de referência, aumentando à medida que os valores de θ se aproximam de 0 ou de 1.

Tendo em conta as metodologias actuais para classificação de prestadores de cuidados de saúde em indicadores de processo, a metodologia apresentada representa um novo tipo

de classificação com um conceito base diferente: classificar os prestadores de acordo com níveis de referência construídos a partir da opinião de especialistas, em vez de comparar o desempenho individual com o desempenho global que apresenta as limitações anteriormente referidas.

Para além de conseguir minimizar as limitações encontradas em outras metodologias, a capacidade dos investigadores (ou responsáveis pelo processo de classificação) poderem controlar os valores de referência pelo grau de exigência pretendido é de grande utilidade para uma classificação correcta dos prestadores. Qualquer que seja o indicador de processo, é expectável que exista uma melhoria contínua por parte dos prestadores. Esta capacidade de ajuste permitirá alterar os valores de referência de acordo com a evolução registada, mas não dependendo da mesma como acontecia anteriormente.

4

Conclusões

À semelhança dos capítulos anteriores, as conclusões retiradas serão repartidas pelos dois objectivos endereçados no presente estudo.

1º Objectivo

Os modelos de ajuste de risco desenvolvidos mostraram-se, de um ponto de vista geral, mais ajustados à realidade nacional e com maior poder de predição do que os modelos da JCI, o que terá reflexo numa avaliação mais precisa dos cuidados prestados pelos prestadores nacionais. Apresentam ainda a vantagem de serem modelos mais parcimoniosos, necessitando de menos informação para obter melhores resultados, permitindo ao mesmo tempo aumentar a velocidade de execução dos algoritmos de classificação.

A criação de uma metodologia de definição, validação e avaliação do desempenho dos modelos de ajuste de risco permite uniformizar o processo de desenvolvimento dos mesmos e a aplicação a futuros modelos. Este ponto é bastante importante uma vez que apenas se conhece o “produto final” dos modelos disponibilizados pela JCI, não sendo possível detectar potenciais problemas na construção dos mesmos. O desenvolvimento de modelos de ajuste de risco com base em dados nacionais é um ponto de inovação nos

sistemas de avaliação de cuidados de saúde nacionais.

Por outro lado, a menor quantidade de observações e baixa proporção de acontecimentos de interesse verificados em alguns dos modelos desenvolvidos pode limitar o desempenho dos mesmos, sendo aconselhável rever, e se necessário actualizar, estes modelos quando estiverem disponíveis mais observações.

2º Objectivo

A metodologia desenvolvida permite comparar os desempenhos individuais dos prestadores com valores de referência. A determinação dos valores de referência através dos quantis da amostra de opinião combinada tem a vantagem de permitir ajustar os mesmos consoante o grau de exigência pretendido e ainda de actualizar-los entre avaliações, deixando de existir uma dependência dos dados. É também de destacar o carácter inovador desta metodologia, com a introdução da opinião de especialistas no processo de avaliação dos prestadores de cuidados de saúde, não tendo sido encontrada nenhuma metodologia semelhante na revisão bibliográfica efectuada.

Por outro lado, a necessidade de recolher a opinião de especialistas para cada indicador de processos torna o processo de implementação desta metodologia num processo moroso quando comparado com a metodologia actual. A implementação da metodologia desenvolvida pode também estar limitada pela rotura que faz com a metodologia actual, não sendo possível aplicar as duas metodologias simultaneamente no mesmo sistema de avaliação.

4.1 Objectivos realizados

O primeiro objectivo do projecto, a redefinição e implementação de modelos de ajuste de risco de suporte à avaliação dos cuidados de saúde adaptados à realidade nacional, foi cumprido. A metodologia utilizada na construção, validação e avaliação do desempenho dos modelos é uma metodologia robusta, prática de aplicar, bastante utilizada e documentada, tendo sido obtidos modelos mais adaptados à realidade nacional quando comparados com os modelos da JCI. Até ao momento de conclusão deste relatório, dois dos modelos de ajuste de risco da área de Ortopedia foram aplicados, perspectivando-se ainda uma aplicação futura do modelo da área de Enfarte Agudo do Miocárdio e de dois

modelos da área de Obstetrícia, cujos desempenhos foram bastante positivos.

O segundo objectivo, a implementação de valores de referência para os indicadores de processo de suporte à avaliação dos cuidados de saúde nacionais, foi cumprido, tendo-se construído uma metodologia alternativa aplicável a todos os indicadores de processo, que minimiza as limitações encontradas na metodologia actual. Porém, as duas metodologias não devem ser comparadas directamente. Apesar de terem sido construídas com a finalidade de classificar os prestadores de cuidados de saúde em 3 níveis de classificação, a metodologia actual efectua uma comparação do desempenho individual dos prestadores com o desempenho global, enquanto que a metodologia alternativa efectua uma comparação do desempenho individual dos prestadores com níveis de referência, não limitando a evolução contínua dos prestadores a nível individual pelo desempenho global.

4.2 Outros trabalhos desenvolvidos

Durante a realização do estágio foram estipulados alguns objectivos adicionais, permitindo participar em projectos a nível nacional, tais como:

- Revisão dos ficheiros de código **R** implementados no âmbito do modelo de avaliação de cuidados de saúde, de forma a assegurar a correcta implementação do mesmo;
- Optimização dos ficheiros de conversão de códigos ICD-9-CM através da inclusão de funções que permitiram aumentar a velocidade de execução dos mesmos;
- Apoio na validação dos resultados finais da última avaliação do modelo de avaliação de cuidados de saúde actual;
- Criação de casos de teste para as novas áreas do modelo de avaliação de cuidados de saúde.

4.3 Limitações e trabalho futuro

A definição dos modelos de ajuste de risco apresentou algumas limitações na construção de alguns dos modelos, especialmente nos indicadores em que o volume de dados é ainda reduzido e em especial nos indicadores de resultado cujos factores de risco não são muito estudados. No futuro, é expectável que alguns dos modelos determinados possam ser aplicados num sistema de avaliação, substituindo assim os modelos da JCI.

Em termos de metodologia, não existiram limitações relevantes porque é uma metodologia bastante comum e de fácil aplicação, sendo a análise do desempenho e a validação dos modelos para os dados nacionais um ponto forte em comparação com os modelos JCI.

Quanto ao segundo objectivo, apesar de a metodologia alternativa estar completamente definida e de a sua aplicação ser simples, tem também uma aplicação morosa, sendo necessário recolher a opinião de diferentes especialistas para cada indicador de resultados. Para além disso, a avaliação efectuada actualmente em Portugal utiliza a comparação dos desempenhos individuais com o desempenho global, sendo necessária uma mudança na abordagem desta classificação para a metodologia alternativa poder ser implementada.

4.4 Apreciação final

Os 9 meses de estágio no sector de Healthcare da Siemens, S.A. permitiram-me aplicar conhecimentos adquiridos durante a licenciatura e mestrado, mas também trabalhar com novas metodologias que permitiram aprofundar conhecimentos em novas áreas.

A quantidade e diversidade de indicadores de processo e resultado com que tive oportunidade de trabalhar, permitiram adquirir um maior conhecimento na área da saúde em diversas áreas clínicas, mais especificamente na vertente de avaliação de cuidados de saúde.

A nível de software, a utilização diária do **R** permitiu adquirir uma prática e proficiência bastante assinaláveis que certamente trará vantagens no futuro dada a importância desta ferramenta na análise de dados. De destacar ainda a aprendizagem de \LaTeX , linguagem utilizada para a produção de textos matemáticos e científicos, para a escrita do presente relatório e que será, certamente, uma mais valia em termos futuros.

Todas as experiências durante o estágio permitiram-me crescer, tanto a nível pessoal como profissional, adquirir experiência e compreender melhor o ambiente que existe numa empresa de grande dimensão, pelo que a apreciação final só pode ser positiva.

Referências Bibliográficas

- [1] J. Buzek and E. Gyori. Directive 2011/24/EU of the European parliament and of the council. *Official Journal of the European Union*, pages 45–65, 2011.
- [2] The Joint Commission. Target analysis methodology for assessing hospital performance on the aligned cms/joint commission national hospital quality measures, 2009.
- [3] Health Grades. The Twelfth Annual HealthGrades Hospital Quality in America Study, 2009.
- [4] G. Grunkemeier and Y. Wu. What Are the Odds? *The Annals of thoracic surgery*, 83:1240–1244, April 2007.
- [5] Sector Healthcare Siemens. Apresentação da Empresa, 2012.
- [6] R Core Team. R: A language and environment for statistical computing, 2012.
- [7] D. Hosmer and S. Lemeshow. *Applied Logistic Regression, Second Edition*. Wiley, Ohio, 2000.
- [8] ACSS. Portal da Codificação Clínica e dos GDH, 2011.
- [9] M. A. Turkman and G. L. Silva. *Modelos Lineares Generalizados - da teoria à prática*. Lisboa, 2000.
- [10] S. Haykin. *Neural networks - A comprehensive foundation*, volume 8. Prentice Hall International, Inc., August 1999.
- [11] J. Hardin and J. Hilbe. *Generalized Linear Models and Extensions*. Stata Press, Texas, 2nd edition, 2007.
- [12] S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 1938.

- [13] H. Akaike. A new look at the statistical model identification. *Transactions on Automatic Control*, 1974.
- [14] L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, 2nd edition, 2001.
- [15] W. Dupont. *Statistical modeling for biomedical researchers: A simple introduction to the analysis of complex data*. Cambridge University Press, Cambridge, 2002.
- [16] K. Boolen and R. Jackman. *Regression diagnostics: An expository treatment of outliers and influential cases*. Sage Publications, Newbury Park, 1990.
- [17] M. Kutner, C. Nachtsheim, and J. Neter. *Applied Linear Regression Models*. McGraw-Hill, 2004.
- [18] L. Brown, T. Cai, and A. DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–116, 2001.
- [19] Joint Commission International. Target Analysis. Technical report, 2011.
- [20] R. Clemen and R. Winkler. Combining Probability Distributions From Experts in Risk Analysis. *Risk Analysis*, 9, 1999.
- [21] H. Linstone and M. Turoff. *The Delphi Method: techniques and applications*. 2002.
- [22] A. Delbecq, A. Van de Ven, and D. Gustafson. *Group techniques for program planning: A guide*. Scott Foreman, Glenview, IL, 1975.
- [23] K. Stark, H. Horst, and L. Kelly. Combining expert opinions: a comparison of different approaches. *9th International Symposium on Veterinary Epidemiology and Economics*, pages 3–5, 2000.
- [24] M. Stone. The opinion pool. *Annals of Mathematical Statistics*, 1961.
- [25] R. Clemen and R. Winkler. Calibrating and combining precipitation probability forecasts. *Probability and Bayesian Statistics*, pages 97–110, 1987.
- [26] W. Ferrell. Combining individual judgments. *Behavioral Decision Making*, pages 111–145, 1985.
- [27] D. Seaver. Assessing probability with multiple individuals: Group interactions versus mathematical aggregation. Technical report, Social Science Research Institute, California, 1978.
- [28] D. Vose. *Risk analysis: a quantitative guide*. John Wiley & Sons, Chichester, 1996.

-
- [29] R. Pouillot and M.L. Delignette-Muller. Evaluating variability and uncertainty in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology*, 2010.
- [30] R. Clemen. Comment on Cooke’s Classical Method, 2006.
- [31] A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 2005.
- [32] C. D. Paulino, M. A. Turkman, and B. Murteira. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa, 2003.
- [33] L. Sousa and M. A. Turkman. Complementos de Estatística 2012/2013, 2013.

A

Código R

A.1 Modelos de ajuste de risco

Pacotes utilizados na construção e validação dos modelos de ajuste de risco:

```
library(car) #Para calcular a multicolinearidade
library(ROCR) #Para a construção das curvas ROC
library(ggplot2) #Pacote gráfico
library(MKmisc) #Para realizar o teste de Hosmer-Lemeshow
library(reshape2) #Pacote com operações para data frames
```

Amostras para a validação cruzada:

```
#cria um vector de dimensão igual aos dados,
#preenchido com valores 0 e 1 de forma aleatória
amostra<-sample(c(0,1), length(dados[,1]), replace = TRUE)

dados<-cbind(dados,amostra) #junta os dados e o vector criado

#separa os dados em duas amostras (amostra 0 e amostra 1)
dados0<-dados[-which(amostra==1),]
dados1<-dados[-which(amostra==0),]

#retira as colunas do vector criado
dados0<-dados0[,-match("amostra",names(dados0))]
dados1<-dados1[,-match("amostra",names(dados1))]
```

Seleccção do modelo:

```
#modelo de regressão logística
modelo<-glm(y~x1+x2+...+xk,data=dados0,family=binomial)
summary(modelo) #resumo do modelo

#realiza o teste de razão de verossimilhanças para todas as covariáveis
#que ainda não foram incluídas no modelo
add1(modelo,dados0,test="Chisq")

#realiza o teste de razão de verossimilhanças para todas as covariáveis
#incluídas no modelo
drop1(modelo,test="Chisq")
```

Cálculo OR:

```
#Função que calcula o odds ratio e respectivos intervalos de confiança
or.glm<-function(x,ic=0.95){
  if(!all(class(x)==c("glm","lm"))){
    stop("Só pode ser usada em objectos da classe 'glm'")}
  x<-summary.glm(x)
  tabela<-cbind(Coeficientes=x$coefficients[,1],
               "Erro Padrão"=x$coefficients[,2],
               OR=exp(x$coefficients[,1]),
               "Inf"=rep(0,length(x$coefficients[,1])),
               Sup=rep(0, length(x$coefficients[,1])))
  tabela[,4]<-exp(tabela[,1]-(tabela[,2]*qnorm(1-((1-ic)/2))))
  tabela[,5]<-exp(tabela[,1]+(tabela[,2]*qnorm(1-((1-ic)/2))))
  dimnames(tabela)[[1]]<-dimnames(x$coeff)[[1]][]
  round(tabela,4)
}
```

Multicolinearidade:

```
vif(glm(y~x1+x2+...+xk,data=dados0,family=binomial)) #Calcula o VIF
```

Discriminação:

```
#Calcula a sensibilidade e especificidade
#para diferentes pontos de corte
pred<-predict(modelo,newdata=dados1)
pred1<-prediction(pred,dados1$y)
perf<-performance(pred1,"tpr","fpr")

#Calcula a AUC
auc<-performance(pred1,measure="auc")@y.values[[1]]

#data.frame para o gráfico
rd<-data.frame(x=perf@x.values[[1]],y=perf@y.values[[1]])

#Curva ROC
ggplot(rd,aes(x=x,y=y))+geom_path(size=1)+
  geom_segment(aes(x=0,y=0,xend=1,yend=1),colour="black",linetype=2)+
  geom_text(aes(x=1,y=0,hjust=1,vjust=0,
               label=paste(sep="","AUC modelo Revisto = ",round(auc,3))),
```

```

    colour="black",size=6)+
  scale_x_continuous(name="1-Especificidade")+
  scale_y_continuous(name="Sensibilidade")+
  guides(colour=guide_legend(title.hjust=0.5))+
  theme(legend.title=element_text(size=20))+
  theme(legend.text=element_text(size=20))+
  theme(axis.title.x=element_text(size=20))+
  theme(axis.title.y=element_text(size=20))+
  theme(legend.position="top")

```

Calibração:

```

#valores ajustados
ajustados<-predict(modelo,newdata=dados1,type="response")

#teste de Hosmer-Lemeshow
HLgof.test(fit=ajustados, obs=dados1$y,verbose=TRUE)

#utilizar a estatística de teste e calcular o p-value, para 10 g.l.
1-pchisq("valor da estatística de teste",10)

```

Adequação da função de ligação:

```

#preditor linear
eta2<-(modelo$linear.predictors)^2

#Função que avalia a adequação da função de ligação
goodlink<-function(modelo){
  if(any(class(modelo)=="glm")){
    eta2<-(modelo$linear.predictors)^2
    novo_modelo<-update(modelo, ~ . + eta2)
    teste<-ifelse(modelo$family$family!="gaussian","Chisq","F")
    lrt<-anova(modelo,novo_modelo,test=teste)
    pvalue<-lrt[nrow(lrt),ncol(lrt)]
    cat("Hipótese nula: A função de ligação é adequada.
        \n LRT P-value:",format.pval(pvalue),"\n")
  }
  else warning("Objecto deve ser da classe 'glm'")
}

```


Resíduos:

```
#valores ajustados vs resíduos deviance padronizados
res<-rstandard(modelo,type="deviance")

ggplot(modelo,aes(.fitted,res))+geom_hline(yintercept=0)+
  geom_point()+geom_hline(y=c(-2,2),lty=2,col="red")+
  scale_y_continuous(name="Resíduos Deviance")+
  scale_x_continuous(name="Valores Ajustados")+
  theme(axis.title.x=element_text(size=20))+
  theme(axis.title.y=element_text(size=20))
```

Leverage:

```
h<-hatvalues(modelo) #matriz hat
p<-dim(model.matrix(modelo))[[2]] #número de parâmetros
n<-dim(model.matrix(modelo))[[1]] #número de observações
leverage<-h/(p/n)

ggplot(modelo,aes(1:length(leverage),leverage))+
  geom_point()+geom_hline(linetype=2, size=.2)+
  scale_x_continuous("Índice")+
  scale_y_continuous("Leverage")+
  geom_hline(y=2,lty=2,col="red")+
  theme(axis.title.x=element_text(size=20))+
  theme(axis.title.y=element_text(size=20))
```

Distância de Cook:

```
ggplot(modelo,aes(1:length(.cooks),.cooks,ymin=0,ymax=.cooks))+
  geom_point()+scale_x_continuous("Índice")+
  scale_y_continuous("Distância de Cook")+
  theme(axis.title.x=element_text(size=20))+
  theme(axis.title.y=element_text(size=20))
```

Linearidade das variáveis contínuas:

```
#gráfico components+residuals
ndf<-cbind(modelo$model[,-1],residual=residuals(modelo))
ndf <- melt(ndf,id.vars="residual")
coefs<-coef(modelo)[-1]
names(coefs)<-c("x1","x2",..., "xk")
```

```

ndf$cr<-rep(NA,nrow(ndf))
ndf[,3]<-as.numeric(ndf[,3])

#coluna com os valores das componentes*coeficientes+resíduos
for(i in 1:length(coefs)){
  idx<-which(ndf$variable==names(coefs)[i])
  ndf$cr[idx]<-coefs[[i]]*ndf[idx,3]+ndf[idx,1]
}
ndf<-ndf[ndf$variable=="xi",] #xi = nome da variável contínua

qplot(value,cr,data=ndf)+
  stat_smooth(method="lm")+
  ylab("Componentes+Resíduos")+xlab("x1")+
  theme(axis.title.x=element_text(size=20))+
  theme(axis.title.y=element_text(size=20))

```

A.2 Valores de referência

Pacotes utilizados na recolha, simulação e combinação de opiniões de especialistas:

```

library(mc2d) #Permite a utilização de modelos PERT
library(ggplot2) #Pacote gráfico
library(binom) #Intervalo de confiança para a proporção

```

Recolha de opiniões:

```

n_experts<-5 #número de especialistas

#opiniões dos especialistas (máximo, mínimo, moda)
opinions<-matrix(scan(nlines=n_experts),nrow=n_experts,byrow=TRUE)
1 1 0.8 0.95
2 0.95 0.7 0.9
3 0.92 0.6 0.85
4 0.9 0.6 0.75
5 1 0.7 0.9

colnames(opinions)<-c("expert","maximo","minimo","moda")
opinions<-as.data.frame(opinions)

```

```
#nomes dos especialistas
opinions[,1]<-c("Especialista A","Especialista B","Especialista C",
              "Especialista D","Especialista E")
```

Simulação de amostras aleatórias com distribuição PERT:

```
n<-20000 #dimensão das amostras a gerar
#matrix com as amostras aleatórias
rand<-matrix(NA,ncol=n_experts,nrow=n)
for(i in 1:n_experts){
  set.seed(0)
  rand[,i]<-rpert(n,opinions[i,3],opinions[i,4],opinions[i,2])
}
```

Representação gráfica das funções de densidade de probabilidade correspondentes às opiniões dos especialistas:

```
#para fazer o gráfico utilizando o ggplot2,
#é necessário juntar as amostras geradas numa única coluna
mistura<-c(rand[,1],rand[,2],rand[,3],rand[,4],rand[,5])
simul<-data.frame(dens=mistura,lines=rep(opinions[,1],each=n))

#cores atribuídas às opiniões (colorblind friendly)
cbPalette <- c("#E69F00","#56B4E9","#009E73",
              "#F0E442","#CC79A7","#000000")

#gráfico das opiniões individuais
ggplot(simul,aes(x=dens,fill=lines))+
  geom_density(alpha=0.9)+
  scale_x_continuous("Proporção",limits=c(0.55,1))+
  scale_y_continuous("Densidade")+
  theme(legend.position="bottom")+
  guides(fill=guide_legend(title=NULL))+
  scale_fill_manual(values=cbPalette)+
  theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))+
  theme(legend.text=element_text(size=13))
```

Representação gráfica das funções de densidade de probabilidade correspondentes às opiniões dos especialistas e da opinião combinada:

```
#amostra da opinião combinada
mix<-c(rand[,1],rand[,2],rand[,3],rand[,4],rand[,5])
mix<-data.frame(dens=mix,lines=rep("Combinada",length(mix)))

todas<-rbind(simul,mix) #dataframe com todas as amostras

#gráfico das opiniões individuais e opinião combinada
ggplot(todas,aes(x=dens,fill=lines))+
  geom_density(alpha=0.9)+
  scale_fill_manual(values=cbPalette)+
  guides(fill=guide_legend(title=NULL))+
  scale_x_continuous("Proporção",limits =c(0.55,1))+
  scale_y_continuous("Densidade")+
  theme(legend.position="bottom")+
  theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))+
  theme(legend.text=element_text(size=13))
```

Cálculo da função de distribuição empírica, moda, mínimo, máximo e mediana:

```
f<-ecdf(mistura) #função de distribuição da amostra combinada

#função que calcula a moda
Mode<-function(x) {
  ux<-unique(x)
  ux[which.max(tabulate(match(x,ux)))]
}
moda<-Mode(mistura) #moda amostra combinada
min(mistura) #mínimo amostra combinada
max(mistura) #máximo amostra combinada
median(mistura) #mediana amostra combinada
```

Definição dos valores de referência através dos quantis da amostra combinada:

```
#valor de referência superior (quantil 0.9)
```

```
TS<-quantile(mistura,0.9)
#valor de referência inferior (quantil 0.1)
TI<-quantile(mistura,0.1)
```

Gráficos das funções densidade de probabilidade e de distribuição da amostra combinada com os valores de referência:

```
#gráfico da função densidade de probabilidade
ggplot(mix,aes(x=dens,fill=lines))+
  geom_density(alpha=0.9)+
  scale_fill_manual(values="black")+
  guides(fill=guide_legend(title=NULL))+
  scale_x_continuous("Proporção",limits=c(0.55,1))+
  scale_y_continuous("Densidade")+
  theme(legend.position="bottom")+
  theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))+
  theme(legend.text=element_text(size=13))+
  geom_vline(x=TS,lty=2,col="red")+
  geom_vline(x=TI,lty=2,col="red")

#gráfico da função de distribuição
ggplot(mix, aes(dens))+stat_ecdf()+
  theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))+
  scale_x_continuous("Proporção")+
  scale_y_continuous("Probabilidade acumulada")+
  geom_vline(x=TS,lty=2,col="red")+geom_vline(x=TI,lty=2,col="red")+
  geom_hline(y=0.1,lty=2)+geom_hline(y=0.9,lty=2)
```

Cálculo dos intervalos de confiança para a proporção de intervenções adequadas realizadas para 4 hospitais:

```
n_prestadores<-4 #número de hospitais
dados<-matrix(scan(nlines=n_prestadores),nrow=n_prestadores,byrow=TRUE)
247 246
47 44
```

```

51 44
158 90

#número de oportunidades, número de intervenções adequadas
colnames(dados)<-c("denominador","numerador")
dados<-data.frame(1:n_prestadores,dados)
colnames(dados)[1]<-c("id_prest")

#intervalos de confiança
IC<-binom.confint(dados[,3],dados[,2],
  conf.level = 0.95,method="asymptotic")
IC$upper<-ifelse(IC$upper>1,1,IC$upper) #correção

```

Classificação dos hospitais:

```

rank<-c(1:n_prestadores)
#classificação
classif<-rep(2,n_prestadores)
for(i in 1:n_prestadores){
  if(IC[i,5]>TS)
    classif[i]<-3
  if(IC[i,6]<TI)
    classif[i]<-1
}

#gráfico da classificação
ggplot(IC,aes(x=1:n_prestadores,y=IC[,4]))+
  geom_point(size=3)+coord_cartesian(ylim=c(0.4,1.05))+
  geom_errorbar(aes(ymin=IC[,5],ymax=IC[,6]))+
  geom_hline(y=TS,lty=2,col="red")+
  geom_hline(y=TI,lty=2,col="red")+
  geom_text(aes(1:n_prestadores,0,label=classif))+
  scale_y_continuous("Proporção")+
  scale_x_continuous("Prestadores")+
  theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))

```

Pacotes utilizados no cálculo do risco:

```
library(ggplot2) #Pacote gráfico
library(reshape) #Pacote de agregação de estruturas
```

Funções utilizadas no cálculo do risco:

```
#valor esperado do limite inferior
EspInf<-function(p){
  p-qnorm(0.975,0,1)*sqrt((p*(1-p))/n)
}

#valor esperado do limite superior
EspSup<-function(p){
  p+qnorm(0.975,0,1)*sqrt((p*(1-p))/n)
}

#variância do limite inferior
VarInf<-function(p){
  ((p*(1-p))/n)*(1-(qnorm(0.975,0,1)/sqrt(n))*
  ((1-2*p)/(2*sqrt(p*(1-p))))))^2
}

#variância do limite superior
VarSup<-function(p){
  ((p*(1-p))/n)*(1+(qnorm(0.975,0,1)/sqrt(n))*
  ((1-2*p)/(2*sqrt(p*(1-p))))))^2
}

#probabilidades
#P(LS<=TI)
prob1<-function(ti,mu,sigma){
  pnorm(ti,mu,sigma)
}

#P(LI>TS)
prob3<-function(ts,mu,sigma){
  1-pnorm(ts,mu,sigma)
}
```

Cálculo do risco para diferentes valores de θ :

```
theta<-seq(0.01,0.99,0.01) #theta varia no intervalo (0,1)
n<-200 #fixa a dimensão da amostra

#Risco - theta1
risco1<-0*prob1(TI,EspSup(theta),sqrt(VarSup(theta)))+
  2*prob3(TS,EspSup(theta),sqrt(VarSup(theta)))+
  1*(1-prob1(TI,EspSup(theta),sqrt(VarSup(theta)))-
    prob3(TS,EspSup(theta),sqrt(VarSup(theta))))

#Risco - theta2
risco2<-1*prob1(TI,EspSup(theta),sqrt(VarSup(theta)))+
  1*prob3(TS,EspSup(theta),sqrt(VarSup(theta)))+
  0*(1-prob1(TI,EspSup(theta),sqrt(VarSup(theta)))-
    prob3(TS,EspSup(theta),sqrt(VarSup(theta))))

#Risco - theta3
risco3<-2*prob1(TI,EspSup(theta),sqrt(VarSup(theta)))+
  0*prob3(TS,EspSup(theta),sqrt(VarSup(theta)))+
  1*(1-prob1(TI,EspSup(theta),sqrt(VarSup(theta)))-
    prob3(TS,EspSup(theta),sqrt(VarSup(theta))))

risco<-cbind(theta,risco1,risco2,risco3)
```

Representação gráfica da variação do risco para diferentes valores de θ :

```
#necessário para utilizar o pacote ggplot2
data<-data.frame(prop=theta,theta1=risco1,
  theta2=risco2,theta3=risco3)
risco_graf<- melt(data, id.vars = "prop")

ggplot(risco_graf,aes(x=prop,y=value,colour=variable))+
  geom_line(size=1.2)+theme(axis.title.x=element_text(size=16))+
  theme(axis.title.y=element_text(size=16))+
  theme(legend.text=element_text(size=16))+
  theme(legend.position="bottom")+
  theme(axis.text.x=element_text(size=16))+
  theme(axis.text.y=element_text(size=16))+
```

```
guides(colour=guide_legend(title="Risco"))+
theme(legend.title=element_text(size=16))+
scale_x_continuous("theta",limits=c(0,1))+
scale_y_continuous("Risco")+
geom_vline(x=TS,lty=2,col="black")+
geom_vline(x=TI,lty=2,col="black")
```

