

Carla Patrícia

Santos Costa

Um estudo sobre a robustez do teste- F

**Carla Patrícia
Santos Costa**

Um estudo sobre a robustez do teste- F

Dissertação apresentada na Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, realizada sob a orientação científica da Doutora Maria Manuela Souto de Miranda, Professora Auxiliar da Universidade de Aveiro.

Aos pais e às irmãs, por todo o apoio.

o júri / the jury

presidente / president

Isabel Maria Simões Pereira

Professora Auxiliar da Universidade de Aveiro

vogais / examiners committee

Maria da Conceição Esperança Amado

Professora Auxiliar do Instituto Superior Técnico, Universidade Técnica de Lisboa

Maria Manuela Souto de Miranda

Professora Auxiliar da Universidade de Aveiro

**agradecimentos /
acknowledgements**

Antes de mais, quero deixar o meu profundo e sentido agradecimento à Professora Doutora Manuela Souto por toda a dedicação, compreensão e confiança que sempre manifestou. O seu apoio pedagógico, os desafios cada vez mais complexos que me foi colocando, o estímulo e exigência crescente foram determinantes para a conclusão deste trabalho.

Agradeço a todos os professores que me foram apoiando ao longo do meu percurso académico.

Institucionalmente estou grata à Universidade de Aveiro, por todas as boas condições que me concedeu, não esquecendo os Serviços de Acção Social pela atribuição de uma bolsa de estudo.

Um agradecimento muito especial aos meus pais e irmãs por terem acreditado nas minhas capacidades, me terem apoiado e ajudado nos piores momentos. A eles devo toda a minha força de vontade e persistência...

Não posso deixar de agradecer a todos os meus familiares e amigos pelas oportunas manifestações de companheirismo e encorajamento.

Por fim, peço desculpa por não nomear todas as pessoas que me ajudaram, de uma forma ou de outra, a todos o meu profundo Obrigado...

palavras-chave

ANOVA, *bootstrap*, efeitos fixos, estimador-M, estimador-S, Brown-Forsythe, teste- F , Levene, Kruskal-Wallis, Welch, robustez.

resumo

O presente trabalho consiste num estudo sobre a robustez do teste- F , tradicionalmente usado em Análise da Variância (ANOVA) para testar a igualdade de médias. Em particular, é considerado o modelo com um factor de efeitos fixos.

O teste- F é frequentemente usado ignorando que a validade das conclusões depende da confirmação das hipóteses assumidas no modelo. Mas a robustez do teste- F tem sido discutida por diversos autores quando ele é aplicado em populações que não seguem distribuição Normal, ou com observações com diferentes variâncias. Se não é possível validar tais hipóteses, são recomendadas técnicas alternativas.

Esta dissertação recolhe um conjunto de métodos a utilizar quando falham as hipóteses do modelo. Apresenta-se ainda uma nova proposta de modificação da estatística- F , com condições de aplicação menos restritivas. O método assenta na estimação robusta da variabilidade explicada por cada um dos termos do quociente da estatística- F , usando estimadores de Huber para as médias dos grupos e estimadores- S , para estimar a variabilidade entre grupos e para estimar a variabilidade total. A distribuição da estatística modificada é aproximada por reamostragem, para obter o percentil que define a regra de decisão do teste.

Os diversos métodos são aplicados a um conjunto de dados reais, seguindo-se uma análise comparativa dos resultados, relacionando-os com as hipóteses pressupostas na aplicação de cada técnica.

keywords

One-Way Anova, *bootstrap*, fixed effects, M -estimators, S -estimators, Brown-Forsythe, F -test, Levene, Kruskal-Wallis, Welch, robustness.

abstract

The present work consists of a study on the robustness of the Fisher-test for comparing different means, in the case of the one-way ANOVA with fixed effects. The F -test is often used without checking the assumptions of the model. If it is not possible to validate those assumptions, than one should use alternative procedures. Robustness of the F -test it is an open question, particularly when the Normal distribution does not hold or when the observations have different variances.

This thesis begins with a review of the one-way ANOVA model and it contains a brief description of alternative methods which are recommended when the validation of the assumptions it is not possible. It is also presented a proposal of a robust version of the F -statistic which can be used in much more wide conditions. The new statistic is developed using the Huber's function for estimating the means of the groups. Different S -estimators are used for estimating the variability between groups and the total variability. Its distribution is approached by resampling, so that it becomes possible to estimate the percentile that determines the critical region of the test.

The alternative techniques are used in a real data analyze and the results are compared taking in account the hypotheses imposed in each method.

Índice

Índice	i
Lista de Tabelas	iii
Lista de Figuras	v
Abreviaturas	vii
1 Introdução	1
2 Revisão do modelo ANOVA com um factor de efeitos fixos	5
2.1 Generalidades	5
2.2 Inferências usando métodos tradicionais	9
2.3 Validação das hipóteses do modelo	17
3 Métodos alternativos à ANOVA tradicional	25
3.1 Alternativas paramétricas em modelos heterocedásticos	26
3.1.1 Teste de Welch	27
3.1.2 Teste de Brown-Forsythe	28
3.1.3 Comparação entre os testes de <i>Welch</i> e de <i>Brown-Forsythe</i> . . .	29
3.2 Aproximação da distribuição de uma estatística por reamostragem . . .	30
3.3 Alternativa não paramétrica - Teste de Kruskal-Wallis	34
3.4 Modelação por Tabelas de Contingência	39

4	Robustez dos testes associados ao modelo ANOVA	43
4.1	Noções essenciais em robustez estatística	43
4.1.1	Generalidades	43
4.1.2	Estimadores- M	53
4.1.3	Estimadores- S	58
4.2	Robustez do teste- F	59
4.3	Versões robustas dos métodos paramétricos alternativos	62
4.4	Outros estudos	65
4.5	Proposta de uma estatística de teste robusta ($FHubS$)	66
4.6	Aspectos Computacionais	71
5	Modelação de um conjunto de dados reais	73
5.1	Descrição do Problema	74
5.2	Modelação por ANOVA com um factor de efeitos fixos	78
5.2.1	Aplicação da ANOVA supondo válidas as hipóteses do modelo	78
5.2.2	Validação das hipóteses do modelo	80
5.3	Aplicação dos métodos alternativos ao teste- F	83
5.4	Aplicação do método $FHubS$	85
5.5	Modelação por Tabelas de Contingência	86
5.6	Comentários comparativos	87
6	Comentários finais	91
A	Apêndice	101
B	Macros em R	103
B.1	Testes de Brown-Forsythe e de Welch (modificados)	103
B.2	Teste de $FHubS$	107
C	Anexos	111

Lista de Tabelas

2.1	Esquema da tabela ANOVA para o modelo com um factor fixo.	15
2.2	Tabela das classificações dos quatro grupos de estudantes, sujeitos a quatro diferentes técnicas de ensino.	16
2.3	Tabela ANOVA para o exemplo 2.2.1	17
3.1	Tabela de organização dos dados para o teste de Kruskal-Wallis	35
3.2	Forma geral de uma Tabela de Contingência bidimensional.	40
5.1	Tabela dos graus de risco e do tipo de obesidade em função do IMC.	75
5.2	Tabela ANOVA para o conjunto de dados reais.	79
5.3	Tabela de Contingência das opções da Prática de Exercício Físico relativa aos valores de IMC.	87
C.4	Conjunto dos dados reais.	114

Lista de Figuras

3.1	Ilustração da reamostragem <i>bootstrap</i>	33
5.1	Caixas de bigodes dos valores de IMC relativas aos vários tratamentos do modelo.	78
5.2	Q-Q plot da distribuição Normal dos resíduos.	80
5.3	Diagrama de dispersão dos resíduos estandardizados vs $\hat{y}_{i,j}$	81
5.4	Diagrama de dispersão dos resíduos estandardizados vs índice das observações.	82

Abreviaturas e símbolos

ANOVA	Análise da Variância
IMC	Índice de Massa Corporal
<i>vs</i>	<i>versus</i>
$F(x)$ ou <i>f.d.p.</i>	função distribuição de probabilidade
$f(x)$	função densidade de probabilidade
<i>i.i.d.</i>	independentes e identicamente distribuídos(as)
$N(\mu, \sigma^2)$	distribuição Normal de parâmetros μ e σ^2
$\chi^2_{(k)}$	distribuição de um Qui-quadrado com k graus de liberdade
$X \sim D(\dots)$	X segue uma distribuição D de parâmetros (...)
$X \approx D$	X segue aproximadamente uma distribuição D
$X_n \xrightarrow{D} X$	$\{X_n\}$ converge em distribuição para X
<i>SST</i>	soma total de quadrados
<i>SSG</i>	soma dos quadrados devidos aos tratamentos
<i>SSE</i>	soma dos quadrados devidos aos erros aleatórios
<i>MSG</i>	média quadrática da soma dos quadrados devidos aos tratamentos
<i>MSE</i>	média quadrática da soma dos quadrados devidos aos erros aleatórios
<i>d.f.</i>	número de graus de liberdade

Capítulo 1

Introdução

Este trabalho é dedicado ao estudo do teste de Fisher, no qual assenta o desenvolvimento da Análise de Variância (ANOVA). O modelo ANOVA constitui uma das principais contribuições de Sir Ronald Aylmer Fisher para o desenvolvimento da Estatística e é usado, sobretudo, para efectuar a comparação de médias de vários grupos de observações. É, por isso, um dos métodos estatísticos de maior repercussão na pesquisa científica das diversas áreas do saber.

A partir de registos de observações correspondentes a vários grupos de indivíduos, a ideia de comparar a variação existente entre os grupos, com a variação das observações dentro dos grupos, foi inicialmente estudada por Wilhelm Lexis (1837-1914), na Alemanha, como um teste de homogeneidade em relação à amostragem de atributos então designados por homógrados (note-se que o termo variância, não era conhecido na época). Contudo, foi graças ao trabalho desenvolvido em Fisher (1925) e em Fisher (1935), que essas ideias tiveram maior desenvolvimento. De acordo com Cochran (1980), o primeiro artigo sobre este tópico apareceu em 1923, (Fisher and Mackenzie (1923)). Embora esse artigo tenha sido um marco inicial na Análise de Variância, Fisher deu então preferência à questão algébrica da decomposição da soma dos quadrados. No entanto, o método foi rapidamente generalizado e utilizado em problemas de regressão

múltipla, sendo já bem conhecido no fim da 2^a Guerra Mundial. Pouco tempo depois, houve também o contributo importante de Eisenhart (1947), que introduziu e distinguiu o modelo de efeitos fixos, do modelo de efeitos aleatórios. Nas décadas seguintes surgiu uma grande diversidade de modelos estatísticos baseados na análise da variância, de crescente complexidade.

Os modelos de análise de variância admitem que existe uma característica estatística que é observada em vários grupos de indivíduos e procuram decompor a variabilidade existente nas observações, identificando as fontes dessa variabilidade. Para isso, a variabilidade é quantificada em termos do afastamento entre observações e respectivo valor médio. Em geral, o interesse na aplicação dos modelos reside em averiguar se, em média, existem diferenças significativas entre as observações dos diferentes grupos. A confirmação dessas hipóteses é decidida com base em testes estatísticos, nomeadamente através do teste de Fisher-Snedecor, vulgarmente designado por teste- F . A distribuição da estatística do teste- F é bem conhecida da Teoria da Amostragem e foi estabelecida para amostras constituídas por observações independentes, provenientes de populações com distribuição Normal, com variâncias iguais e independentes entre si. Nesse caso, o Teorema de Cochran foi um instrumento fundamental para o desenvolvimento da ANOVA, pois é com base nele que se garante a distribuição da estatística do teste- F . Como este teorema nem sempre aparece referido na literatura de maior divulgação, ele foi integrado num apêndice deste trabalho.

Mas, os resultados técnicos que sustentam a aplicação dos métodos estatísticos, apenas são válidos quando se verificam as hipóteses assumidas. No entanto, muitos utilizadores decidem aplicar técnicas de modelação estatísticas, sem se preocupar com a validação das hipóteses, e é nesse sentido que surge o presente trabalho. Quando não são verificadas as hipóteses que caracterizam um modelo estatístico, os métodos estatísticos que lhe estão associados podem perder a suas propriedades, surgindo assim a necessidade de encontrar procedimentos que não sejam muito sensíveis a pequenos

afastamentos das hipóteses. Estes métodos são genericamente designados por métodos estatísticos robustos.

Ao longo deste trabalho é feita uma breve análise sobre a robustez do teste- F , pois é um tema de grande interesse e que divide opiniões. De facto, existem autores que consideram que o teste- F é robusto quanto à falha das suposições do modelo, como é o caso de Shmider *et al.* (2010) e existem outras publicações onde se afirma que o teste- F não é robusto, como por exemplo em Rogan and Keselman (1977). É neste sentido que surgem na literatura várias alternativas ao teste- F , sobretudo paramétricas. De entre estas, destacam-se o teste de Welch e o de Brown-Forsythe. Como alternativa não paramétrica pode utilizar-se o teste de Kruskal-Wallis. No entanto, os modelos de Tabelas de Contingência também podem ser encarados como uma alternativa, uma vez que a não rejeição da hipótese de independência entre categorias ou entre factores (em Tabelas de Contingência), conduz a interpretações idênticas às da não rejeição da hipótese nula do teste- F (em ANOVA).

No trabalho que aqui se apresenta faz-se uma revisão do modelo ANOVA com um factor de efeitos fixos e resumem-se as principais opções de alternativas robustas. Depois de introduzir algumas noções básicas sobre robustez estatística, faz-se a apresentação de uma nova proposta de modificação do teste- F , a qual pode ser aplicada sem exigir as condições tradicionais. Por limitações de tempo, a metodologia é explicada no texto, mas não foi efectuado um estudo de simulação que permita fazer a sua avaliação em termos de diferentes cenários. De seguida, considera-se um conjunto de dados reais relativos a observações do Índice de Massa Corporal em estudantes do ensino secundário, que são analisados pelos diversos métodos anteriormente focados no trabalho, incluindo a nova proposta.

Todos os cálculos envolvidos foram efectuados com recurso ao programa **R**, na versão

2.11.1 (*The R Project for Statistical Computing*- ver em <http://cran.at.r-project.org/>), sem o qual teria sido muito difícil cumprir os objectivos propostos, não só pela facilidade de ser um programa de distribuição livre, mas também por incluir um vasto conjunto de técnicas pré-programadas e de métodos actuais em termos de investigação; em particular, no que diz respeito à robustez. Ao longo trabalho encontra-se a indicação dos comandos mais característicos para efectuar cada teste, pois considerou-se que essa indicação poderia vir a ser útil. Ainda em relação a aspectos computacionais, o trabalho inclui também um apêndice formado pelas *macros*, em R, para efectuar os testes de Brown-Forsythe e de Welch modificados, e também o teste de acordo com a nova proposta. Os três programas foram escritos pela autora, como um mero instrumento de cálculo, mas certamente que deverão ser ainda bastante otimizados futuramente, do ponto de vista da programação.

Assim, a dissertação encontra-se organizada em cinco capítulos, para além da presente introdução. O Capítulo 2 consiste na revisão do modelo ANOVA com um factor de efeitos fixos. O Capítulo 3 apresenta os métodos alternativos à ANOVA tradicional. O Capítulo 4 é dedicado à robustez dos testes associados ao modelo ANOVA e inclui uma secção com a explicação da nova proposta. No Capítulo 5, é feito o estudo de aplicação ao conjunto de dados reais. O Capítulo 6 consiste num resumo comentado das conclusões do trabalho. Finalmente, incluem-se o Apêndice A - com a demonstração do Teorema de Cochran -, o Apêndice B - com as macros em R -, e um Anexo com os dados do Índice de Massa Corporal.

Capítulo 2

Revisão do modelo ANOVA com um factor de efeitos fixos

No presente capítulo apresenta-se uma revisão dos principais métodos e resultados sobre a utilização do modelo Análise da Variância (vulgarmente referido por ANOVA), os quais se encontram descritos na maior parte dos livros de texto introdutórios à Estatística, como por exemplo, em Murteira *et al.* (2007).

2.1 Generalidades

Os modelos ANOVA admitem que existe uma relação de dependência de uma variável, cuja distribuição se pretende estudar em termos médios, relativamente a outras variáveis que a influenciam, não sendo explicitada a forma da relação de dependência. A variável que constitui o interesse principal do estudo é de natureza quantitativa e designa-se por *variável resposta*. As restantes variáveis, que supostamente influenciam a variável resposta, podem ser quantitativas ou qualitativas, e são geralmente designadas por *factors*. O principal objectivo na aplicação destes modelos consiste em explicar a variabilidade existente nas observações da variável resposta à custa dos factores. O

estudo da variabilidade é efectuado com base na análise da variância, o que justifica o nome dos modelos.

Ao longo do texto, admitir-se-á que o modelo contempla um só factor. Chamam-se *níveis do factor*, ou *tratamentos*, às diferentes categorias (ou estados) que o factor pode assumir. Quando o modelo inclui mais do que um factor, chama-se *tratamento* a cada combinação de níveis dos diferentes factores.

Um outro aspecto importante ao estabelecer o modelo ANOVA diz respeito à natureza determinística ou aleatória dos factores. Essa característica é estabelecida previamente pelo planeamento da experiência e conduz à classificação dos modelos ANOVA num de três tipos: modelo de *efeitos fixos*, quando os níveis de todos os factores são pré-estabelecidos (deterministicamente); modelo de *efeitos aleatórios*, quando todos os factores são aleatórios; e modelos *mistos*, que contemplam os dois tipos de factores.

Para ilustrar a utilização de terminologia a usar, apresenta-se de seguida um exemplo simples, introdutório.

Exemplo 2.1.1 (*Adaptado a partir de Fonseca (2001)*). Com o objectivo de averiguar se os resultados de um conjunto de estudantes dependiam da técnica de ensino utilizada numa certa disciplina, os alunos foram divididos em quatro grupos, e cada grupo, durante um certo período de tempo, foi sujeito a uma diferente técnica de ensino. Posteriormente foi atribuída uma classificação a cada um dos alunos, numa escala de 0 a 100, que constituiu cada resultado individual.

A variável resposta, que é o objectivo principal do estudo, é a variável “resultados obtidos pelos alunos”. Neste caso considera-se um só factor, que é a “técnica de ensino”, a qual assume quatro níveis diferentes. Quando, na recolha das observações, se selecciona um aluno ao acaso, de entre os que foram sujeitos a uma determinada técnica de ensino e, posteriormente, o seu resultado, então o nível de factor é pré-estabelecido. Procedendo do mesmo modo para todas as observações está-se perante um modelo de efeitos fixos, uma vez que previamente se fixaram os níveis do factor. Se, pelo contrá-

rio, se selecciona o aluno ao acaso, observando então qual a técnica de ensino a que foi sujeito e o posterior resultado, o nível do factor foi resultado de um processo aleatório; trata-se portanto de um factor aleatório e de um modelo de efeitos aleatórios.

□

Formalização do modelo de efeitos fixos

Considere-se que se dispõe de um conjunto de observações da variável resposta, denotada por $Y_{i,j}$, dependente de um factor com k níveis, ou seja, agrupadas em k grupos, cada um dos quais correspondente a um dos k níveis do factor. Cada um dos níveis será denotado pelo índice i e cada observação do nível i será denotada pelo índice j . Assim, o número total de observações do i -ésimo nível será n_i e o número total de indivíduos em estudo será denotado por $N = \sum_{i=1}^k n_i$. Quando os níveis do factor têm igual número de observações, isto é, quando $n_1 = n_2 = \dots = n_k = n$, o modelo diz-se *equilibrado* (ou *balanceado*); caso contrário, diz-se *não equilibrado*.

O modelo ANOVA de efeitos fixos pode ser caracterizado de duas formas: explicitando os modelos da variável resposta em termos de efeitos do factor, sendo designado por *modelo de efeitos*, ou em termos de médias dos níveis do factor, sendo então designado por *modelo de médias*. Apresentam-se separadamente as duas formulações, que como se verá estão relacionadas entre si.

Modelo de efeitos

De acordo com esta caracterização, o valor da variável resposta para a j -ésima observação do i -ésimo nível é representada por $Y_{i,j}$ e verifica uma equação da forma:

$$Y_{i,j} = \mu + \tau_i + \varepsilon_{i,j}, \quad j = 1, 2, \dots, n_i \quad e \quad i = 1, 2, \dots, k, \quad (2.1)$$

com $\sum_{i=1}^k n_i = N$ e $\sum_{i=1}^k n_i \tau_i = 0$, onde:

- μ é uma constante comum a todas as observações, designada por *efeito geral*;
- τ_i é o *efeito (específico)* do nível i , para $i = 1, 2, \dots, k$, uma vez que traduz a especificidade do i -ésimo grupo de observações em relação ao efeito geral;
- $\varepsilon_{i,j}$ é o termo de erro aleatório do modelo, associado à j -ésima observação do i -ésimo grupo. Admita-se que os erros do modelo têm esperança nula, que são não correlacionados e que têm variância constante. A maior parte dos autores supõe ainda que os erros seguem uma distribuição Normal, ou seja que $\varepsilon_{i,j} \sim N(0, \sigma^2)$, independentes e identicamente distribuídas (*i.i.d.*) $\forall i, j$.

Para explicar a restrição $\sum_{i=1}^k n_i \tau_i = 0$, denote-se por μ_i , a média das observações do nível i do factor e por μ , a média ponderada dos μ_i , ou seja,

$$\mu = \frac{\sum_{i=1}^k n_i \mu_i}{N}. \quad (2.2)$$

Então, por (2.1), $E[Y_{i,j}] = \mu_i = \mu + \tau_i$, pelo que $\tau_i = \mu_i - \mu$, para $i = 1, 2, \dots, k$. Daqui decorre que

$$\sum_{i=1}^k n_i \tau_i = \sum_{i=1}^k n_i (\mu_i - \mu) = \sum_{i=1}^k n_i \mu_i - \sum_{i=1}^k n_i \mu = \sum_{i=1}^k n_i \mu_i - N\mu = \sum_{i=1}^k n_i \mu_i - \sum_{i=1}^k n_i \mu_i = 0.$$

Modelo de médias

Esta formulação do modelo corresponde a uma simples reparametrização da equação do modelo de efeitos. Assim, tendo em conta que $\tau_i = \mu_i - \mu$, obtém-se que:

$$Y_{i,j} = \mu_i + \varepsilon_{i,j}, \quad j = 1, 2, \dots, n_i \quad e \quad i = 1, 2, \dots, k, \quad (2.3)$$

onde μ_i é a esperança das observações do nível i do factor.

Desde já se salientam as seguintes propriedades:

- Para um mesmo nível de factor, todas as observações da variável resposta têm o mesmo valor esperado, pois como $E[\varepsilon_{i,j}] = 0$, então $E[Y_{i,j}] = E[\mu_i + \varepsilon_{i,j}] = \mu_i = \mu + \tau_i$.

- Todas as observações possuem a mesma variância, uma vez que

$$Var[Y_{i,j}] = Var[\mu_i + \varepsilon_{i,j}] = Var[\varepsilon_{i,j}] = \sigma^2, \text{ pois } \mu_i \text{ é uma constante } \forall_i.$$

- Quando se assume a distribuição Normal dos erros do modelo, pode-se concluir que as variáveis $Y_{i,j}$ são independentes e seguem uma distribuição Normal, ou seja, que $Y_{i,j} \sim N(\mu + \tau_i, \sigma^2)$ são independentes para todo o $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n_i$.

2.2 Inferências usando métodos tradicionais

Quando se utiliza o modelo de efeitos, interessa averiguar se existem diferenças significativas entre os efeitos de factor, enquanto que quando se utiliza o modelo de médias interessa estudar se, em média, existem diferenças significativas entre os tratamentos.

A confirmação ou não de tais hipóteses é decidida com base em testes estatísticos e a formulação das hipóteses a testar é diferente. Para o caso do modelo de médias, temos as seguintes hipóteses:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i \neq j : \mu_i \neq \mu_j. \quad (2.4)$$

Para o caso do modelo de efeitos, temos:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0 \quad vs \quad H_1 : \exists i : \tau_i \neq 0. \quad (2.5)$$

Note-se que, se as médias dos tratamentos forem todas iguais, isto é, se $\mu_1 = \mu_2 = \dots = \mu_k$, isto significa que não existem efeitos específicos de nenhum tratamento, ou seja, que os efeitos são todos nulos. De facto, se

$$\mu_1 = \mu_2 = \dots = \mu_k \Leftrightarrow \mu + \tau_1 = \mu + \tau_2 = \dots = \mu + \tau_k \Leftrightarrow \tau_1 = \tau_2 = \dots = \tau_k = 0.$$

O procedimento geral para efectuar (estes e outros) testes em ANOVA recorre basicamente à decomposição da variabilidade através de partições de somas de quadrados, identificando as principais fontes de variação com as respectivas somas quadráticas.

Antes de prosseguir, é conveniente introduzir alguma notação adicional.

Notação:

- A soma das observações para o i -ésimo nível do factor é representada por $Y_{i,\cdot}$, isto é,

$$Y_{i,\cdot} = \sum_{j=1}^{n_i} Y_{i,j}. \quad (2.6)$$

- A média das observações do i -ésimo nível do factor, ou média do grupo, é representada por \bar{Y}_i , isto é,

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{i,j}}{n_i} = \frac{Y_{i,\cdot}}{n_i}, \quad i = 1, 2, \dots, k. \quad (2.7)$$

- A média global de todas as observações em estudo denota-se por $\bar{\bar{Y}}$, isto é,

$$\bar{\bar{Y}} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{i,j}}{N} = \frac{\sum_{i=1}^k Y_{i,\cdot}}{N} = \frac{\sum_{i=1}^k n_i \bar{Y}_i}{N}. \quad (2.8)$$

□

Rapidamente se pode confirmar que a média das observações \bar{Y}_i e a média global das observações $\bar{\bar{Y}}$ constituem os estimadores obtidos pelo Métodos dos Momentos, respectivamente, das médias μ_i e da média global μ . Coincidem também com os correspondentes estimadores de Máxima Verosimilhança, quando é suposta a hipótese de distribuição Normal dos erros do modelo.

Assim, a variabilidade total da variável $Y_{i,j}$ pode ser medida pelo seu afastamento em relação ao estimador da média global, isto é, por

$$Y_{i,j} - \bar{\bar{Y}}. \quad (2.9)$$

Contudo, a variabilidade de $Y_{i,j}$ relativamente ao estimador da média do nível i do factor, quando é utilizada a informação dada pelos níveis do mesmo, pode ser medida através da seguinte expressão

$$Y_{i,j} - \bar{Y}_i. \quad (2.10)$$

Tendo em conta (2.9) e (2.10), verifica-se que $(Y_{i,j} - \bar{Y})^2 = ((\bar{Y}_i - \bar{Y}) + (Y_{i,j} - \bar{Y}_i))^2$. Somando em i e em j , obtém-se que $\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y})^2$ é igual à expressão que se segue

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_i - \bar{Y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} 2(\bar{Y}_i - \bar{Y})(Y_{i,j} - \bar{Y}_i) + \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2$$

Mas, como facilmente se verifica que $\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i) = 0$, a segunda parcela da soma anterior anula-se e tem-se que

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 + \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2. \quad (2.11)$$

Designando a soma total dos quadrados dos afastamentos de $Y_{i,j}$ a \bar{Y} por *SST*,

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y})^2, \quad (2.12)$$

a soma dos quadrados devidos aos tratamentos por *SSG*,

$$SSG = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2, \quad (2.13)$$

e a soma dos quadrados devidos aos erros do modelo por *SSE*,

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2, \quad (2.14)$$

a expressão (2.11) representa-se por

$$SST = SSG + SSE.$$

Associada à anterior partição de SST , verifica-se a correspondente partição do número de graus de liberdade, que serão denotados por df :

$$df(SST) = df(SSG) + df(SSE) \Leftrightarrow (N - 1) = (k - 1) + (N - k).$$

Conhecidos os números de graus de liberdade associados a cada soma quadrática, definem-se as médias quadráticas

$$MSG = \frac{SSG}{k - 1}; \quad MSE = \frac{SSE}{N - k}.$$

Deve-se ter em conta que, se a modelação traduzir bem as causas da variabilidade, então a soma SSE deve ser pequena quando comparada com a soma SSG . Note-se que se SSE fosse nula, isso significaria que as observações de cada nível do factor eram todas iguais, logo não existiria variabilidade dentro de nenhum grupo. Por outro lado, se todas as estimativas das médias dos factores fossem iguais, elas coincidiriam com a média geral e então $SSG = 0$.

Logo, se o modelo estiver bem ajustado, a soma SSG deve ser muito superior à soma SSE . Em termos de médias quadráticas, o mesmo argumento justifica que a média quadrática MSG deva ser muito superior a MSE .

Como se referiu, quando se utiliza a ANOVA a um factor fixo, é usual que o objectivo seja o de testar a igualdade das médias dos níveis de factor (ou, equivalentemente, se são nulos os efeitos dos tratamentos).

Para responder a esta questão em geral, efectua-se um teste de *Fisher*, vulgarmente designado por teste- F , cujo desenvolvimento se passa a resumir.

Quando se assume a hipótese de que os erros do modelo seguem a distribuição Normal, os testes referenciados em (2.4) e (2.5) são desenvolvidos com base na distribuição da estatística

$$F = \frac{\frac{SSG}{k - 1}}{\frac{SSE}{N - k}} = \frac{MSG}{MSE}. \quad (2.15)$$

Supondo que as amostras são provenientes de populações com distribuição Normal com variâncias iguais e que, para além disso, são independentemente entre si e são constituídas por observações independentes, a distribuição da estatística F é conhecida a partir de resultados fundamentais da Teoria da Amostragem.

Com efeito, o Teorema de Cochran permite concluir que $\frac{SSG}{\sigma^2}$ e $\frac{SSE}{\sigma^2}$ são variáveis independentes e com distribuição de um Qui-quadrado com $k - 1$ e $N - k$ graus de liberdade, respectivamente. Como este resultado é fundamental no estudo dos modelos ANOVA e, em geral a sua demonstração não está disponível, considerou-se oportuno apresentar tal demonstração. A prova segue de perto a de Lindgren (1976) e encontra-se em Apêndice.

Teorema 2.2.1 (Cochran). *Sejam U_1, U_2, \dots, U_n variáveis aleatórias independentes, normais estandardizadas e tais que*

$$\sum_{i=1}^n U_i^2 = Q_1 + Q_2 + \dots + Q_k, \quad (2.16)$$

onde cada Q_i representa uma soma de quadrados de combinações linear dos U_i 's.

Denotem-se por r_i a característica de cada forma quadrática Q_i e por n a soma $n = r_1 + r_2 + \dots + r_k$.

Então, as variáveis Q_i são independentes e cada uma delas segue uma distribuição de um Qui-Quadrado cujo número de graus de liberdade é igual à característica da forma quadrática que a define.

□

Da aplicação do Teorema de Cochran à Teoria da Amostragem em populações com distribuição Normal, decorre o Corolário seguinte.

Corolário 2.2.2. *Se todas as n observações Y_i são provenientes da mesma distribuição Normal com média μ e variância σ^2 , e se $\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y})^2$ pode ser decomposto em k soma de quadrados, SS_i , cada uma com df_i graus de liberdade, então os termos*

$\frac{SS_i}{\sigma^2}$ são variáveis independentes com distribuição de um Qui-quadrado com df_i graus de liberdade, tais que $\sum_{i=1}^k df_i = n - 1$.

□

Atendendo ao teorema de Cochran e quando H_0 é verdadeira, pode-se concluir que a estatística de teste- F , em (2.4) e (2.5), segue uma distribuição de *Fisher* com $k - 1$ e $N - k$ graus de liberdade, pois:

- SSG e SSE são independentes;
- $\frac{SSG}{\sigma^2} \sim \chi_{(k-1)}^2$;
- $\frac{SSE}{\sigma^2} \sim \chi_{(N-k)}^2$.

Deste modo

$$F = \left(\frac{SSG}{k-1} \right) / \left(\frac{SSE}{N-k} \right) = \frac{MSG}{MSE} \sim F(k-1, N-k),$$

onde $F_{(u,v)}$ designa a distribuição de *Fisher* com (u, v) graus de liberdade.

Note-se que, se H_0 é falsa, então $E[MSG] > \sigma^2$, uma vez que

$$E[MSG] = E \left[\frac{SSG}{k-1} \right] = \sigma^2 + \frac{\sum_{i=1}^k n_i \tau_i^2}{k-1} > \sigma^2.$$

Por outro lado, $E[MSE] = \sigma^2$, em qualquer uma das hipóteses H_0 ou H_1 . Logo, verifica-se sempre que $E[MSG] \geq E[MSE]$.

Para além disso, de acordo com Krzanowski (1998), verifica-se também que:

$$E \left[\frac{MSG}{MSE} | H_0 \text{ é falsa} \right] > E \left[\frac{MSG}{MSE} | H_0 \text{ é verdadeira} \right],$$

ou de modo equivalente,

$$E \left[\frac{MSG}{MSE} | H_0 \text{ é falsa} \right] / E \left[\frac{MSG}{MSE} | H_0 \text{ é verdadeira} \right] > 1.$$

O resultado anterior conduz a uma região crítica unilateral à direita do teste- F . Denotando por $f_{(1-\alpha; k-1, N-k)}$ o quantil $(1-\alpha)$ da distribuição $F_{(k-1, N-k)}$ e por f_{obs} o valor observado da estatística F , a regra de decisão do teste, ao nível de significância α , passa a ser a seguinte:

- Se $f_{obs} \geq f_{(1-\alpha; k-1, N-k)}$, o que significa que o teste tem um valor- p igual a $P[F \geq f_{obs} | H_0] < \alpha$ e rejeita-se H_0 ao nível de significância α .
- Se $f_{obs} \leq f_{(1-\alpha; k-1, N-k)}$, então $p = P[F \geq f_{obs} | H_0] > \alpha$ e não há motivo para rejeitar H_0 .

Em suma, a informação anterior pode ser resumida numa tabela ANOVA de modo a facilitar os cálculos para a aplicação do teste- F :

Fonte de variação	df	SS	MS	Estatística de teste (F)	valor- p
Entre tratamentos	k-1	SSG	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSE}$	$P[F > F_{obs} H_0]$
Dentro dos tratamentos	N-k	SSE	$MSE = \frac{SSE}{N-k}$		
Total	N-1	SST	$MST = \frac{SST}{N-1}$		

Tabela 2.1: Esquema da tabela ANOVA para o modelo com um factor fixo.

O exemplo seguinte é continuação do anterior e ilustra a aplicação do teste F para comparação de médias. Foi adaptado a partir de Fonseca (2001).

Exemplo 2.2.1. Quatro grupos de estudantes estiveram sujeitos, durante certo período de tempo, a técnicas diferentes de ensino. No final desse período foram avaliados, tendo sido obtidos os resultados que constam da tabela seguinte, onde G_i designa o grupo de alunos sujeito à i -ésima técnica de ensino ($i = 1, 2, 3, 4$). Pretende-se testar se existem diferenças significativas entre as médias alcançadas pelos alunos sujeitos às diferentes técnicas de ensino.

Neste exemplo o factor em análise é a “técnica de ensino”, o qual assume quatro tratamentos correspondentes às quatro técnicas diferentes. Uma vez que os tratamentos

G1	G2	G3	G4
65	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88
81	72	83	
69	79	76	
	90		

Tabela 2.2: Tabela das classificações dos quatro grupos de estudantes, sujeitos a quatro diferentes técnicas de ensino.

são pré-determinados, o modelo ANOVA com um factor e efeitos fixos é o modelo adequado. Considerando a formalização em termos de modelo de médias. Admite-se que

$$Y_{i,j} = \mu_i + \varepsilon_{i,j}, \text{ com } j = 1, \dots, n_i \text{ e } i = 1, 2, 3, 4,$$

onde $Y_{i,j}$ indica a classificação do j -ésimo aluno sujeito à i -ésima técnica de ensino. Após uma análise preliminar dos dados, assume-se que os erros $\varepsilon_{i,j}$ são *i.i.d.* com distribuição $N(0, \sigma^2)$.

Pretende-se testar se as classificações médias dos alunos de cada grupo dependem da técnica de ensino a que foram sujeitos. Ou seja, interessa averiguar se as médias dos vários tratamentos do factor poderão ser consideradas iguais. Formalizando o teste, as hipóteses são

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu \quad vs \quad H_1 : \exists i : \mu_i \neq \mu.$$

Supondo verificadas as condições de normalidade, homocedasticidade e independência, a estatística de teste será $F = \frac{MSG}{MSE} \sim F_{(k-1, N-k)}$, com distribuição $F_{(3,19)}$. Efectuando os cálculos necessários obtém-se a seguinte Tabela ANOVA: o valor crítico do teste, correspondente ao nível de significância $\alpha = 0.05$, é o percentil da distribuição F , $f_{(0.05;3;19)} = 3.13$. Então a região crítica do teste é constituída pelo conjunto de

Fonte de variação	df	SS	MS	Estatística de teste (F)
Entre tratamentos	3	712.5864	$MSG = 237.5288$	$F_{obs} = 3.7715$
Dentro dos tratamentos	19	1196.631	$MSE = 62.9806$	

Tabela 2.3: Tabela ANOVA para o exemplo 2.2.1

amostras tais que $f_{obs} \geq 3.13$. Como $3.77 > 3.13$, com base nesta amostra e a um nível de significância de 5%, há motivos para rejeitar H_0 , concluindo-se que pelo menos um dos métodos de ensino conduz a resultados médios significativamente diferentes. A conclusão podia também ser observada com base no valor- p do teste, uma vez que $p = P[F > F_{obs} | H_0] = 0.02804 < 0.05$.

□

2.3 Validação das hipóteses do modelo

Antes de iniciar qualquer inferência estatística, deve-se verificar se os dados em estudo são compatíveis com as suposições exigidas pelo modelo. Os resultados técnicos que sustentam a aplicação dos métodos estatísticos apenas são válidos quando se verificam as hipóteses assumidas, ainda que aproximadamente. Assim, é da maior importância que se conheçam e executem procedimentos para a validação dos pressupostos de um modelo.

No modelo em estudo, para além da equação que o caracteriza, são assumidas as seguintes hipóteses sobre a distribuição dos erros do modelo:

- i. Distribuição Normal;
- ii. Homocedasticidade;
- iii. Independência.

De um modo geral, as hipóteses de qualquer modelo estatístico devem começar por ser

validadas por uma análise preliminar de dados e, em particular, por métodos gráficos, tendo em conta a distribuição empírica dos resíduos. No caso particular do modelo ANOVA, e dada a existência de diferentes níveis de *factor*, é importante completar a análise gráfica com gráficos de médias e/ou com caixas de bigodes.

No presente trabalho e tendo em conta os objectivos principais a atingir, isto é, uma reflexão sobre a robustez do teste- F , a questão da validação assume um papel de destaque. Efectivamente, para garantir que a estatística tradicionalmente usada no modelo ANOVA, segue uma distribuição de *Fisher*, é essencial que se validem as hipóteses i., ii. e iii. atrás referidas.

Tendo em vista o tratamento estatístico de dados reais, como os que serão explorados no Capítulo 5, interessa estudar até que ponto é que as conclusões sobre a comparação de médias vão diferir, dependendo da existência de motivos para rejeitar as hipóteses assumidas.

Deste modo, passa-se a fazer uma breve apresentação de técnicas úteis para validar as hipóteses de i. a iii..

i. Distribuição Normal

A concordância das observações com um modelo de distribuição Normal poderá ser analisada, graficamente, através da elaboração de um gráfico de quantis. Os pontos apresentados devem situar-se em torno de uma recta.

Os testes de ajustamento de *Shapiro-Wilks* ou o de *Lilliefors*, poderão ser utilizados para decidir sobre a distribuição Normal da população de que provêm os dados.

ii. Igualdade de Variâncias (Homocedasticidade do modelo)

Quando o modelo não é homocedástico, a análise da variância pode ficar comprometida, pois como foi visto pelo Corolário do teorema de Cochran, as n observações Y_i têm de ser provenientes da mesma distribuição Normal com médias μ e variâncias σ^2 , para que o quociente definido com as somas de quadrados e a variância siga uma distribuição de um Qui-quadrado. Desta forma, tem-se a certeza que a estatística de teste- F segue uma distribuição de Fisher.

Os testes mais conhecidos para testar a homogeneidade do modelo são o teste de *Levene*, o de *Bartlett* e o de *Hartley*. A referência a estes testes é motivada pelo facto dos mesmos serem os mais conhecidos na literatura e estarem disponíveis na maioria dos programas estatísticos como o R, SPSS, S-Plus, etc..

Tendo em conta considerações sobre a robustez estatística e de acordo com Almeida *et al.* (2008), desde já se adianta que o teste de *Bartlett* não é robusto quando a suposição de normalidade é violada. No entanto, este teste apresenta uma maior eficiência na validação das hipóteses quando comparado com o teste de *Hartley* (ver por exemplo em Neter *et al.* (1996)), o qual, apesar de ser um teste muito utilizado pela sua simplicidade de cálculo, deve ser evitado devido à sua baixa eficiência.

Os autores do artigo Almeida *et al.* (2008), propõem o uso do teste de *Levene* para a comparação de variâncias de grupos de observações provenientes de distribuições contínuas e não necessariamente Normais. Sem prejuízo pela antecipação, desde já se adianta que o teste de *Levene* é considerado robusto, já que na ausência da distribuição Normal das observações, a taxa de rejeição de H_0 é próxima do nível de significância fixado para uma grande variedade de distribuições de probabilidade.

Pelo referido anteriormente, o teste de *Levene* parece ser o teste mais usual e apropriado para o estudo da homocedasticidade. Sendo assim, dos testes referidos, será o único apresentado neste trabalho.

Teste de Levene (versão original)

O teste de *Levene* é aplicado quando se pretende testar a hipótese nula

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2, \quad (2.17)$$

contra a hipótese alternativa

$$H_1 : \exists i : \sigma_i^2 \neq \sigma^2, \text{ para algum } i = 1, \dots, k.$$

Para resumir o método, considere-se

$$Z_{i,j} = |Y_{i,j} - \bar{Y}_i|, \text{ com } j = 1, \dots, n_i \quad e \quad i = 1, \dots, k. \quad (2.18)$$

Deste modo, com

$$\bar{Z}_i = n_i^{-1} \sum_{j=1}^{n_i} Z_{i,j} \quad e \quad \bar{\bar{Z}} = N^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{i,j}.$$

Logo,

$$MSG_L = \frac{\sum_{i=1}^k n_i (\bar{Z}_i - \bar{\bar{Z}})^2}{k-1} \quad e \quad MSE_L = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{i,j} - \bar{Z}_i)^2}{n-k}.$$

A estatística de teste inicialmente proposta por *Levene* foi:

$$F_L = \frac{MSG_L}{MSE_L}.$$

Note-se que se H_0 é falsa, $E[MSG_L] \geq E[MSE_L]$ e a interpretação é análoga à do quociente da estatística de teste- F .

Quando H_0 é verdadeira, a estatística F_L possui distribuição aproximada de uma distribuição de Fisher, ou seja, $F_L \approx F_{(k-1; N-k)}$.

Então, denote-se por f_{Lobs} o valor observado da estatística de teste F_L e por $f_{(1-\alpha; k-1, N-k)}$, o quantil $(1 - \alpha)$ da distribuição $F_{(k-1, N-k)}$. A regra de decisão do teste, ao nível de significância α , passa a ser a seguinte:

- Se $f_{Lobs} \geq f_{(1-\alpha; k-1, N-k)}$ rejeita-se H_0 , ou seja, há evidência de que as variâncias das populações em questão são diferentes.
- Se $f_{Lobs} \leq f_{(1-\alpha; k-1, N-k)}$, então não há motivo para rejeitar H_0 .

De seguida é apresentada uma modificação do teste de Levene. A alteração consiste em substituir o estimador clássico do parâmetro de localização, \bar{Y}_i em (2.18), por estimadores mais robustos. Mais à frente neste trabalho, será aprofundado o significado do termo “robusto”.

Tal como foi apresentada, a versão original deste teste baseia-se em considerar os afastamentos $Z_{i,j} = |Y_{i,j} - \bar{Y}_i|$, isto é, tomando a média de cada grupo \bar{Y}_i , como estimador do parâmetro de localização μ_i . Uma vez que a média amostral não é um estimador robusto, foi sugerida uma versão robusta do teste (a qual se encontra programada no *software* R, na *package* `lawstat`). Apresenta-se de seguida a modificação do teste original de *Levene*.

Teste modificado de Levene

Substituindo \bar{Y}_i por \tilde{Y}_i , em (2.18), onde \tilde{Y}_i representa a mediana das observações da população i , tem-se que

$$Z_{i,j}^{(m)} = |Y_{i,j} - \tilde{Y}_i|, \text{ com } j = 1, \dots, n_i \text{ e } i = 1, \dots, k, \quad (2.19)$$

onde \tilde{Y}_i é a mediana das observações do grupo i .

Sendo assim, denotem-se por $\overline{Z_i^{(m)}}$ a média dos afastamentos $Z_{i,j}^{(m)}$ para as observações $Y_{i,j}$ pertencentes ao i -ésimo grupo e por $\overline{\overline{Z^{(m)}}}$ a correspondente média global, isto é,

$$\overline{Z_i^{(m)}} = n_i^{-1} \sum_{j=1}^{n_i} Z_{i,j}^{(m)}, \quad \text{e} \quad \overline{\overline{Z^{(m)}}} = N^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{i,j}^{(m)}.$$

Então, tem-se que

$$MSG_{Lm} = \frac{\sum_{i=1}^k n_i (\overline{Z_i^{(m)}} - \overline{\overline{Z^{(m)}}})^2}{k-1} \quad \text{e} \quad MSE_{Lm} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{i,j}^{(m)} - \overline{Z_i^{(m)}})^2}{N-k}.$$

A estatística de teste modificado de *Levene* é definida com as novas expressões de MSG_{Lm} e de MSE_{Lm} , através de

$$F_{Lm} = \frac{MSG_{Lm}}{MSE_{Lm}}.$$

Como foi referido em Almeida *et al.* (2008), quando H_0 é verdadeira, a estatística F_{Lm} segue aproximadamente uma distribuição de Fisher com $k - 1$ e $N - k$ graus de liberdade, ou seja, $F_{Lm} \sim F_{(k-1;N-k)}$.

Denote-se agora por f_{Lmobs} o valor observado de F_{Lm} e por $f_{(1-\alpha;k-1,N-k)}$, o quantil $(1 - \alpha)$ da distribuição $F_{(k-1,N-k)}$. A regra de decisão do teste, ao nível de significância α , passa a ser a seguinte:

- Se $f_{Lmobs} \geq f_{(1-\alpha;k-1,N-k)}$ rejeita-se H_0 , ou seja, há evidência de que as variâncias das populações em questão são diferentes.
- Se $f_{Lmobs} \leq f_{(1-\alpha;k-1,N-k)}$, então não há motivo para rejeitar H_0 .

No programa **R**, este teste pode ser efectuado usando o comando `levene.test`, com a opção `location="median"`.

iii. Independência

A avaliação da hipótese de independência das observações, em termos gráficos, pode ser efectuada através da análise de um gráfico sequencial de resíduos, onde se define o resíduo da observação $y_{i,j}$ por $r_{i,j} = y_{i,j} - \widehat{y}_i$. Deste modo, considere-se uma representação dos resíduos *vs* o índice da observação ($k = 1, 2, \dots, N$). Quando os erros são independentes espera-se que os resíduos flutuem num padrão mais ou menos aleatório em torno de zero.

Para além da análise gráfica, é possível efectuar um teste não paramétrico de aleatoriedade, vulgarmente designado por *teste de runs*, ou ainda o *teste de Durbin-Watson* (consultar, por exemplo Neter *et al.* (1996)).

A utilização destes testes numa análise preliminar de dados é, actualmente, muito rápida, uma vez que os testes se encontram disponíveis no *software* estatístico.

No programa R, o teste de Durbin-Watson, pode ser efectuado usando o comando `dwtest`.

Apenas após a execução destes procedimentos e a confirmação de que não há motivos estatisticamente significativos para rejeitar as hipóteses de que as amostras são independentes e provenientes de populações com a distribuição Normal e de variâncias iguais, e só nesse caso, é que se devem considerar reunidas as condições necessárias para aplicar o teste- F tradicional.

É importante realçar que, caso haja motivos para suspeitar que alguma das hipóteses não é verificada, nada garante que a estatística- F siga uma distribuição de *Fisher* $F_{(k-1, N-k)}$. De facto, embora se possa manter a utilização de tal estatística (do ponto de vista da interpretação do quociente), há a necessidade de conhecer a sua distribuição, ainda que aproximadamente, para que se possa estabelecer a regra de decisão do teste.

Nas circunstâncias em que a análise preliminar da amostra conduz à rejeição das hipóteses assumidas no modelo ANOVA, é necessário recorrer a métodos alternativos ao tradicional teste- F , como os que se apresentam no capítulo seguinte.

Capítulo 3

Métodos alternativos à ANOVA tradicional

No capítulo anterior abordou-se o modelo ANOVA para comparação de médias, tendo-se realçado que, para que este modelo tenha uma aplicação correcta, é necessário que se verifiquem as hipóteses inicialmente assumidas. Assim, é importante conhecer métodos alternativos que permitam fazer o tratamento estatístico em situações cujos objectivos coincidem com os de uma ANOVA, mas em que falham as hipóteses do modelo. No presente capítulo serão apresentados alguns métodos para comparação de médias de $k > 2$ populações independentes. Em particular, serão focados testes paramétricos que não exigem homocedasticidade, como é o caso dos testes de Brown-Forsythe e de Welch, e testes não paramétricos. De entre estes últimos, serão apresentados o teste de Kruskal-Wallis, um método baseado em reamostragem que foi desenvolvido ao longo da preparação da presente dissertação e ainda a modelação por Tabelas de Contingência. Embora esta última técnica não seja, geralmente, colocada como uma alternativa à ANOVA, considera-se que, do ponto de vista prático, a não rejeição da hipótese de efeitos nulos em ANOVA, conduz a conclusões idênticas às da não rejeição da hipótese de independência entre os factores em Tabelas de Contingência.

Como se explicou na Introdução, os dados a analisar no Capítulo 5 tinham já

sido objecto de um estudo estatístico anterior, através de uma análise por Tabelas de Contingência, o que só por si, justificou a inclusão deste tópico no presente capítulo.

Antes de prosseguir com o resumo das alternativas ao teste- F para comparação de diversas médias, é de realçar que a designação de “comparação de médias” tem vindo a ser usada abusivamente. De facto, a expressão conveniente seria a de comparação de medidas de tendência central, pois na realidade é esse o objectivo da aplicação do teste- F . Esta explicação é pertinente no seguimento do presente capítulo, pois serão focados métodos paramétricos e não paramétricos, e no caso do teste de Kruskal-Wallis o parâmetro de interesse é a mediana e não a média. Note-se que em distribuições simétricas a mediana coincide com a média, pelo que é indiferente testar a igualdade de médias ou de medianas, em tais condições.

No sentido de esclarecer a terminologia utilizada, refere-se que a designação de inferência não paramétrica advém da possibilidade de delinear conclusões sobre a população de onde provêm os dados, sem que seja necessária a especificação completa da *f.d.p.* em causa (quando esta pertence a uma família paramétrica), ou quando não é suposta nenhuma família paramétrica específica. Segundo Murteira *et al.* (2007), a análise estatística clássica depende, num grau exageradamente elevado, do pressuposto de que a população tem distribuição Normal. Os métodos não paramétricos visam substituir, quando necessário, esse tão enraizado pressuposto, por outro de características mais gerais.

3.1 Alternativas paramétricas em modelos heterocedásticos

Para comparação de médias de várias populações existem alguns testes que se constituem como alternativas paramétricas à ANOVA.

Como resposta ao problema da comparação de médias num cenário de heterocedas-

ticidade, várias soluções foram propostas. Entre elas, destacam-se o teste de Brown-Forsythe e o teste de Welch, que se passam a apresentar.

3.1.1 Teste de Welch

Uma vez reconhecida a necessidade de se terem disponíveis testes que possam ser aplicados quando as observações têm diferentes variâncias, em 1951, Welch propôs um teste para comparação de $k > 2$ médias de populações com distribuição Normal, mas com variâncias distintas, o qual se passa a apresentar.

Como anteriormente, considere-se $Y_{i,j}$ com $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n_i$, variáveis aleatórias independentes e com distribuição Normal, ou seja, $Y_{i,j} \sim N(\mu_i, \sigma_i^2)$, com $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n_i$.

Pretende-se, então, testar as hipóteses:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i \neq j : \mu_i \neq \mu_j \quad (3.1)$$

A estatística do teste que foi proposta por Welch é definida por:

$$W^* = \frac{\sum_{i=1}^k w_i (\bar{Y}_i - \tilde{Y})^2 / (k-1)}{\left(1 + \frac{2(k-2)}{k^2-1} \sum_{i=1}^k \frac{\left(1 - \frac{w_i}{\sum_{i=1}^k w_i}\right)^2}{n_i - 1} \right)}, \quad (3.2)$$

onde $w_i = \frac{n_i}{S_i^2}$, com $S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2}{n_i - 1}$, para $i = 1, 2, \dots, k$, e $\tilde{Y} = \frac{\sum_{i=1}^k w_i \bar{Y}_i}{\sum_{i=1}^k w_i}$.

Supondo H_0 verdadeira, em (3.1), tem-se que $W^* \xrightarrow{D} W$, onde $W \sim F_{(k-1; df)}$, e

$$df = \left[\frac{3}{k^2-1} \sum_{i=1}^k \frac{\left(1 - \frac{w_i}{\sum_{i=1}^k w_i}\right)^2}{n_i - 1} \right]^{-1}. \quad (3.3)$$

A regra de decisão apropriada ao nível de significância α é a seguinte:

- Se $W^* \leq f_{(1-\alpha; k-1; df)}$, não se rejeita H_0 , ao nível de significância α ;
- Se $W^* > f_{(1-\alpha; k-1; df)}$ rejeita-se H_0 , ao nível de significância α ,

onde $f_{(1-\alpha; k-1; df)}$ é o quantil $1 - \alpha$ da distribuição $F_{(k-1, df)}$.

No programa R, este teste pode ser efectuado usando o comando `oneway.test`, com a opção `var.equal = FALSE`.

3.1.2 Teste de Brown-Forsythe

Uma outra solução alternativa ao teste- F , para o caso em que as observações têm variâncias diferentes é o teste de *Brown-Forsythe*.

Considere-se $Y_{i,j}$, a j -ésima observação da i -ésima amostra correspondente ao i -ésimo grupo, $j = 1, 2, \dots, n_i$ e $i = 1, 2, \dots, k$ de forma que $Y_{i,j} \sim N(\mu_i, \sigma_i^2)$.

Pretende-se novamente testar a hipótese nula:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k, \quad (3.4)$$

contra alternativa:

$$H_1 : \exists i \neq j : \mu_i \neq \mu_j.$$

A estatística deste teste, denotada por BF , é definida por:

$$BF = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{\bar{Y}})^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{N}\right) S_i^2}, \quad (3.5)$$

$$\text{onde } S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2}{n_i - 1} \text{ e } N = \sum_{i=1}^k n_i. \quad (3.6)$$

Os valores críticos de BF , sob a veracidade de H_0 , são obtidos a partir da distribuição $F_{(k-1,df)}$, onde o número de graus de liberdade df é obtido por

$$df = \left[\sum_{i=1}^k \frac{c_i^2}{n_i - 1} \right]^{-1} \quad e \quad c_i = \frac{\left(1 - \frac{n_i}{N}\right) S_i^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{N}\right) S_i^2}. \quad (3.7)$$

Para mais detalhes sobre o desenvolvimento deste teste, recomenda-se a leitura de Brown and Forsythe (1974b).

Sob H_0 , os valores elevados de BF levam a concluir que H_0 não é verdadeira. Assim, a regra de decisão apropriada para o nível de significância α é a seguinte:

- se $BF \leq f_{(k-1;df;1-\alpha)}$, não se rejeita H_0 , ao nível de significância α ;
- se $BF > f_{(k-1;df;1-\alpha)}$, rejeita-se H_0 , ao nível de significância α ,

onde $f_{(k-1;df;1-\alpha)}$ é o quantil $1 - \alpha$ da distribuição F com $k - 1$ e df graus de liberdade.

3.1.3 Comparação entre os testes de *Welch* e de *Brown-Forsythe*

Os autores de Brown *et al.* (1974a) realizaram um estudo de simulação para comparar quatro estatísticas que podem ser utilizadas para testar a igualdade de médias de populações independentes, supondo heterocedasticidade. As estatísticas consideradas foram a estatística de Fisher, a de Brown-Forsythe, a de Welch e ainda uma outra, designada por estatística de James; foi avaliada a sensibilidade dessas estatísticas em diferentes distribuições. O critério usado para avaliação da robustez foi o da comparação da potência empírica dos testes. Como é referido em Almeida *et al.* (2008), no trabalho de Brown e Forsythe foi verificado que, tratando-se de modelos heterocedásticos e para dimensões de grupos superiores a 10 observações, a diferença entre a taxa de rejeição e o nível de significância do teste- F é superior àquela que se verifica para os restantes testes. Assim, é de concluir que não há motivos para usar o teste- F nessas condições. Para além disso, nas mesmas condições, foi verificado que a diferença entre a taxa de

rejeição e o nível de significância era pequena, tanto no teste de Welch como no teste de Brown- Forsythe, e que no teste de Welch a variabilidade dessa diferença era menor.

É importante referir ainda que, para situações de homocedasticidade (e de acordo com o mesmo artigo de Almeida *et al.* (2007)), Brown e Forsythe observaram que, nas situações estudadas, a potência do teste- F é semelhante à dos testes de Brown-Forsythe e de Welch.

Deste modo, é de concluir que quando existe heterocedasticidade e as dimensões dos grupos são superiores a 10, o teste de Welch é o mais recomendado.

3.2 Aproximação da distribuição de uma estatística por reamostragem

Nesta secção será feita uma apresentação sumária do método *bootstrap* não paramétrico, pois será usado no seguimento, como componente de uma alternativa não paramétrica ao teste- F , em situações em que se desconhece a distribuição da estatística de teste.

O *bootstrap* é uma técnica de reamostragem bastante utilizada em diferentes estudos estatísticos. A metodologia *bootstrap* foi originalmente proposta por Efron (1979), num contexto de observações *i.i.d.*. A base da técnica consiste na obtenção de “novos” conjuntos de dados, gerados por reamostragem computacional do conjunto dos dados originais.

Para resumir esta técnica, considere-se que se pretende estudar a distribuição de uma certa estatística, em particular, de um estimador de um parâmetro θ , denotado por $\hat{\theta} = T(\mathbf{X})$, onde $\mathbf{X} = (X_1, X_2, \dots, X_n)$, a partir de uma realização de uma amostra aleatória, $\mathbf{x} = (x_1, x_2, \dots, x_n)$. A distribuição de $\hat{\theta}$ depende da distribuição da população, mas em diversas situações não se conhece a distribuição de probabilidade G

da população; outras vezes, embora se conheça a *f.d.p.*, a distribuição da estatística é muito difícil de determinar, ou mesmo impossível de tratar matematicamente. Em consequência, Efron propôs que se estudasse a distribuição de $\hat{\theta}$ aproximando G por uma sua estimativa \hat{G} , a qual é obtida através da realização amostral \mathbf{x} .

O método *bootstrap* pode ser resumido pelo seguinte algoritmo:

- 1º Encarar \hat{G} como se fosse a verdadeira *f.d.p.* da população.
- 2º A partir da população com *f.d.p.* \hat{G} , são geradas novas amostras aleatórias, com reposição de elementos de \mathbf{x} , amostras essas que são genericamente denotadas por, $\mathbf{X}^{(b)*} = (X_1^{(b)*}, X_2^{(b)*}, \dots, X_n^{(b)*})$, sendo $\mathbf{x}^{(b)*} = (x_1^{(b)*}, x_2^{(b)*}, \dots, x_n^{(b)*})$ as suas realizações.
- 3º Para cada uma das amostras *bootstrap* geradas, calcular o valor da estatística $\hat{\theta}$, obtendo assim B realizações $\hat{\theta}^{(b)*} = T(\mathbf{x}^{(b)*})$, para $b = 1, \dots, B$, as quais são designadas por *réplicas bootstrap*.
- 4º Calcular o erro padrão de $\hat{\theta}$, $S_{\hat{\theta}}$

$$S_{\hat{\theta}} = \sqrt{\frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}^{(b)*} - \hat{E}[\hat{\theta}] \right)^2},$$

$$\text{onde } \hat{E}[\hat{\theta}] \approx \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)*}.$$

Deste modo, aproxima-se a distribuição da estatística $\hat{\theta}$, pela distribuição empírica das B réplicas, sendo B um número elevado.

Para o cálculo das estimativas *bootstrap*, em muitas circunstâncias é suficiente um valor de $B = 100$. Contudo, para conhecer a distribuição por amostragem com alguma precisão, deve considerar-se um valor de B substancialmente mais elevado. Segundo Efron (1982), geralmente, $B = 1000$ proporciona bons resultados.

Existem dois tipos fundamentais de *bootstrap*, o *bootstrap* paramétrico e o não paramétrico. Ambos se distinguem pela função de distribuição \widehat{G} que utilizam para aproximar a *f.d.p.* G da população de onde é recolhida a amostra.

Para concretizar o *bootstrap* paramétrico, represente-se por $G(x|\theta)$ a *f.d.p.* G , evidenciando a sua dependência em relação ao valor de θ . Quando θ é estimado pelo estimador de máxima verosimilhança, então o verdadeiro valor de θ é aproximado pela estimativa $\widehat{\theta}_{MV}$; o *bootstrap* paramétrico assume que $\widehat{G}(x) = G(x|\widehat{\theta}_{MV})$. Então, as amostras *bootstrap* são geradas aleatoriamente a partir da distribuição $G(x|\widehat{\theta}_{MV})$.

O *bootstrap* não paramétrico caracteriza-se por considerar que \widehat{G} é a própria função de distribuição empírica da amostra observada \mathbf{x} . Neste caso, as amostras *bootstrap* são retiradas aleatoriamente e com reposição, directamente da amostra original.

De um ponto de vista prático, a principal diferença entre os dois tipos de *bootstrap* reside no facto do *bootstrap* paramétrico exigir que se assumam uma *f.d.p.* G para a população e que se conheça o estimador de máxima verosimilhança de θ . Por isso, o método que reúne mais adeptos é o *bootstrap* não paramétrico.

Ao longo deste trabalho usar-se-á a metodologia *bootstrap* não paramétrica, como se passa a explicar.

Quando surge o problema da não validação dos pressupostos do modelo ANOVA, nomeadamente, quando as observações não são independentes ou não têm variância comum, ou quando existe afastamento da distribuição Normal das observações, o utilizador não conhece a distribuição da estatística de teste, $F = \frac{MSG}{MSE}$, pois não é possível concluir que siga uma distribuição de *Fisher*. Neste caso, o método *bootstrap* surge como uma solução para estimar a distribuição do quociente F .

O paralelismo entre a metodologia usual e a metodologia *bootstrap* no âmbito do estudo deste trabalho, encontra-se esquematizado na Figura 3.1.

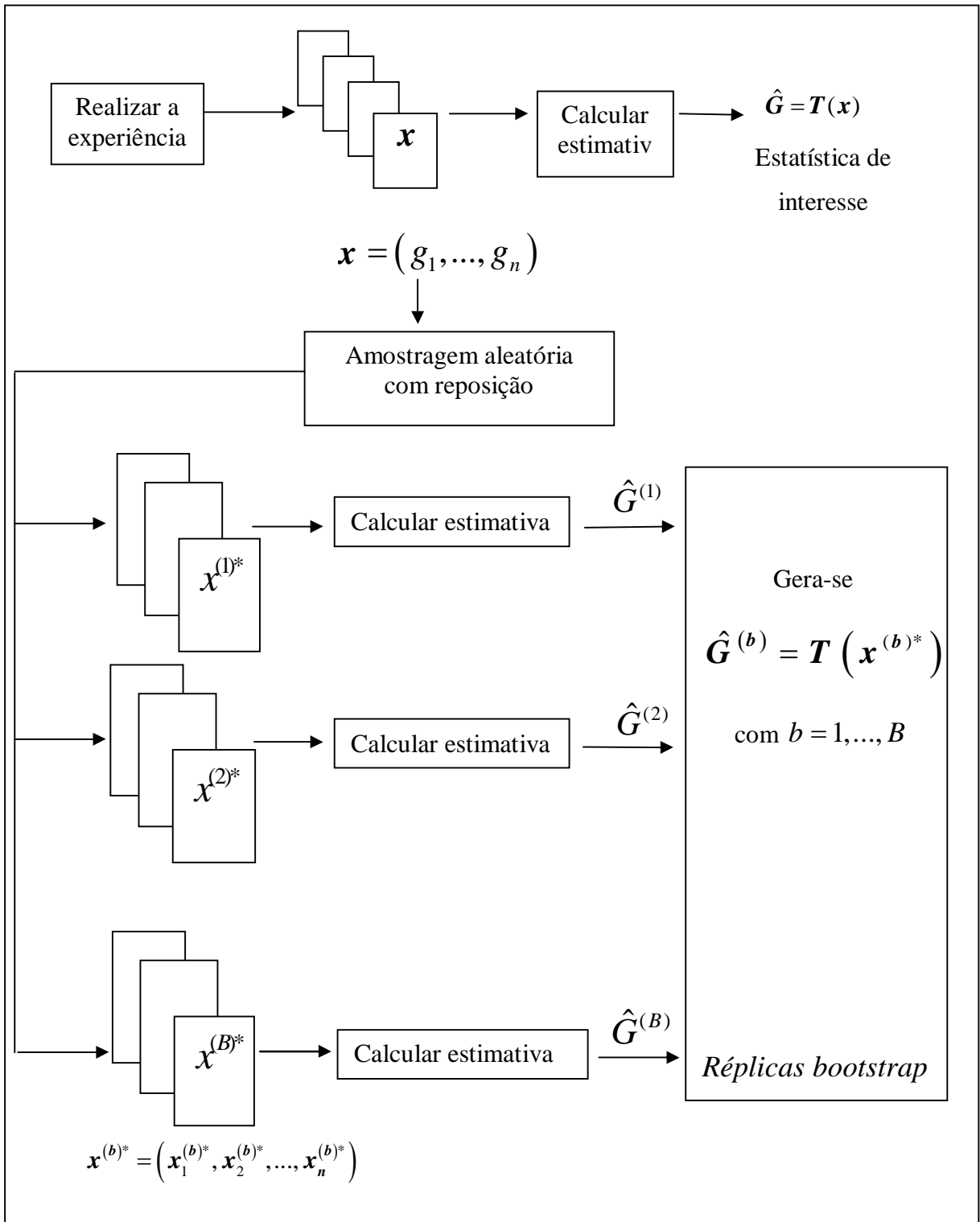


Figura 3.1: Ilustração da reamostragem *bootstrap*

3.3 Alternativa não paramétrica - Teste de Kruskal-Wallis

Esta secção é dedicada ao teste não paramétrico de Kruskal-Wallis. Este teste permite efectuar a comparação de medianas de k populações independentes, com $k \in \mathbb{N}$, sem assumir as hipóteses do modelo ANOVA, em particular, sem assumir a hipótese da distribuição Normal das populações de onde são retiradas as observações. O teste pode ser visto como uma generalização do teste de Mann-Whitney-Wilcoxon e em distribuições simétricas estabelece um teste para comparação de médias. O teste de Kruskal-Wallis constitui, por isso uma alternativa não paramétrica ao modelo ANOVA.

Teste de Kruskal-Wallis

Como se referiu, este teste serve para testar se k amostras provêm de uma mesma população ou de populações independentes com diferentes medianas. Tenha-se em conta que em distribuições simétricas a mediana coincide com a média.

Começar-se-á por apresentar a metodologia do teste, seguindo-se as condições de aplicabilidade e a distribuição da estatística do teste.

Considerem-se $k > 2$ amostras aleatórias, não necessariamente com a mesma dimensão, provenientes de k populações independentes, onde se denota por

$$\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,n_i}),$$

a i -ésima amostra de dimensão n_i , com $i = 1, 2, \dots, k$.

Para facilitar a organização dos dados, estes podem ser dispostos numa tabela como a Tabela 3.1.

Para elaboração deste teste, constrói-se a amostra combinada (formada pela reunião do conjunto de todas as amostras) a qual tem dimensão $N = \sum_{i=1}^k n_i$. De seguida, ordenam-se as observações da amostra combinada de forma crescente e atribuem-se

Amostra 1	Amostra 2	...	Amostra k
$X_{1,1}$	$X_{2,1}$...	$X_{k,1}$
$X_{1,2}$	$X_{2,2}$...	$X_{k,2}$
\vdots	\vdots	.	\vdots
X_{1,n_1}	X_{2,n_2}	...	X_{k,n_k}

Tabela 3.1: Tabela de organização dos dados para o teste de Kruskal-Wallis

números de ordem a cada uma das N observações. Se houver observações repetidas atribui-se a cada uma delas um número de ordem que é dado pela média dos números de ordem das observações repetidas.

Denote-se por $r(X_{i,j})$ o número de ordem atribuído a $X_{i,j}$ e por r_i e \bar{r}_i , respectivamente, a soma e a média das ordens atribuídas aos elementos da amostra i , i. e.,

$$r_i = \sum_{j=1}^{n_i} r(X_{i,j}), \quad i = 1, 2, \dots, k \quad (3.8)$$

$$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r(X_{i,j})}{n_i}, \quad i = 1, 2, \dots, k. \quad (3.9)$$

Pretende-se efectuar o teste de hipóteses

$$H_0 : m_1 = m_2 = \dots = m_k = m \quad vs \quad H_1 : \exists i : m_i \neq m, \quad (3.10)$$

onde m_i designa a mediana da i -ésima população. Em distribuições simétricas as hipóteses (3.10) são equivalentes às hipóteses (3.4).

Assim, se houver motivos para rejeitar a hipótese inicial aceita-se que pelo menos um par de amostras têm diferentes medianas.

Condições de aplicabilidade do teste:

- A escala de medida das observações é, pelo menos, ordinal.
- Todas as variáveis aleatórias $X_{i,j}$ são contínuas.
- Existe independência entre as várias amostras.

- As k amostras são amostras aleatórias das respectivas populações.

A ideia base do teste de Kruskal-Wallis é a de que, se H_0 for verdadeira, os *ranks* correspondentes aos vários grupos estarão misturados de forma aleatória; caso contrário, deverão existir grupos com predominância de *ranks* reduzidos e outros grupos com predominância de *ranks* elevados. Assim, a estatística de Kruskal-Wallis baseia-se nas posições centrais das observações de cada grupo, avaliando o grau de disparidade entre as várias médias dos *ranks*, r_i , tomados como aleatórios, e o seu valor esperado $\frac{N+1}{2}$. Quando a disparidade for muito grande, então H_0 deve ser rejeitado.

A estatística do teste de Kruskal-Wallis foi inicialmente apresentada em Kruskal and Wallis (1952b) e é definida pela seguinte expressão:

$$T = \frac{(N+1) \sum_{i=1}^k n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (r_{i,j} - \bar{r})^2}, \quad (3.11)$$

onde \bar{r} é a média das ordens de todos os elementos da amostra combinada, ou seja, $\bar{r} = \frac{N+1}{2}$. Entrando com as expressões de \bar{r}_i e de \bar{r} e simplificando os cálculos, obtém-se a forma seguinte, de maior divulgação:

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{\left(r_i - \frac{1}{2}n_i(N+1)\right)^2}{n_i} = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{r_i^2}{n_i} - 3(N+1). \quad (3.12)$$

No caso de existirem observações iguais, de acordo com Conover (1971) deve atribuir-se a ordem média a cada uma delas e corrigir a estatística de teste para:

$$T = \frac{1}{S^2} \left(\sum_{i=1}^k \frac{r_i^2}{n_i} - \frac{N(N+1)^2}{4} \right), \quad (3.13)$$

onde, $S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{n_i} R^2(X_{i,j}) - \frac{N(N+1)^2}{4}$

Para o cálculo da distribuição exacta de T são consideradas todas as observações provenientes de populações idênticas. Sendo assim, cada atribuição de ordens de 1 a N em grupos de dimensões n_1, n_2, \dots, n_k ocorre com igual probabilidade $\frac{n_1!, n_2!, \dots, n_k!}{k}$.

Passa a verificar-se de seguida que, para amostras de grandes dimensões, a variável T segue aproximadamente uma distribuição de um Qui-quadrado com $k - 1$ graus de liberdade. Para tal, será necessário usar o resultado auxiliar que se lembra na nota seguinte.

Nota 3.1. Seja X a soma de n inteiros seleccionados de forma aleatória, sem reposição, a partir dos primeiros N inteiros. Então, a média e a variância de X , são, respectivamente,

$$E[X] = n \frac{N+1}{2} \quad e \quad Var(X) = \frac{n(N+1)(N-n)}{12}.$$

□

A distribuição assintótica de T decorre do facto de R_i , definido por (3.8), ser a soma de n_i variáveis aleatórias independentes. Como, para grandes valores de n_i , o Teorema de Limite Central é válido, quando H_0 é verdadeira, a variável aleatória

$$\frac{R_i - E[R_i]}{\sqrt{Var(R_i)}},$$

possui distribuição assintoticamente Normal estandardizada. Assumindo que n_i é grande, pela Nota 3.2, tem-se que

$$\left[\frac{R_i - E[R_i]}{\sqrt{Var(R_i)}} \right]^2 = \frac{\left[R_i - \left(n_i \frac{N+1}{2} \right) \right]^2}{\frac{n_i(N+1)(N-n_i)}{12}} \quad (3.14)$$

possui aproximadamente uma distribuição de um Qui-quadrado com um grau de liberdade.

Se os R_i fossem independentes, então a distribuição de

$$T_1 = \sum_{i=1}^k \frac{\left(R_i - \left(n_i \frac{N+1}{2} \right) \right)^2}{\frac{n_i(N+1)(N-n_i)}{12}}$$

poderia ser aproximada pela distribuição de um *Qui-Quadrado* com k graus de liberdade.

No entanto, como $\sum_{i=1}^k R_i = \frac{N(N+1)}{2}$, existe uma dependência linear entre os R_i . Para além disso, Kruskal (1952a) mostra que se a i -ésima parcela de T_1 for multiplicada por $\frac{N-n_i}{N}$, $i = 1, 2, \dots, k$, então:

$$T = \sum_{i=1}^k \frac{\left(R_i - \left(n_i \frac{N+1}{2}\right)\right)^2}{\frac{n_i(N+1)N}{12}} \sim \chi_{k-1}^2, \quad (3.15)$$

Note-se que a expressão (3.15) é outra forma de escrever a expressão (3.11), que define a variável T . Pelo que se conclui o resultado anteriormente referido.

Para concretizar o teste em termos de distribuição exacta, os valores críticos para a rejeição de H_0 encontram-se tabelados, por exemplo em Conover (1971), mas apenas para $k=3$ e $n = i \leq 5$, quando não existem observações de igual valor na amostra combinada. Para $k > 3$ ou $n_i \geq 5$, a distribuição exacta não se encontra tabelada e deve utilizar-se a aproximação anteriormente referida, isto é $T \sim \chi_{k-1}^2$.

Regra de decisão:

Rejeita-se H_0 se $T \geq \chi_{(1-\alpha, k-1)}^2$, ou seja, se ao nível de significância α , T exceder o quantil $(1 - \alpha)$ obtido.

No programa R, este teste pode ser efectuado usando o comando `kruskal.test`.

3.4 Modelação por Tabelas de Contingência

Na presente subsecção apresenta-se uma revisão dos principais conceitos sobre Tabelas de Contingência.

Introdução

As Tabelas de Contingência são utilizadas para estudar a independência ou a associação entre duas variáveis categóricas (ou factores), descrevendo a frequência das categorias de uma das variáveis relativamente às categorias de outra.

A elaboração das Tabelas de Contingência deve ser *mutuamente exclusiva*, ou seja, deve ser tal que cada elemento da população pertença a uma e uma só categoria e deve ser *exaustiva*, i.e., deve contemplar todas as categorias possíveis das variáveis em estudo.

Deste modo, as tabelas de contingência aparecem num contexto de análise de variáveis qualitativas, mas também é possível utilizá-las para dados contínuos, considerando-os organizados em classes ou intervalos, de modo a que as classes representem as categorias.

Notação

Considerem-se duas variáveis R e C , respectivamente com r e com c categorias e seja $n_{i,j}$ o total de observações classificadas simultaneamente na categoria i da variável R e na categoria j da variável C , onde $i = 1, \dots, r$ e $j = 1, \dots, c$.

As frequências das observações na amostra são organizadas numa tabela como a Tabela 3.2, onde

		Colunas (Variável C)				Total
Linhas (Variável R)	1	$n_{1,1}$	$n_{1,2}$	\cdots	$n_{1,c}$	$n_{1,\cdot}$
	2	$n_{2,1}$	$n_{2,2}$	\cdots	$n_{2,c}$	$n_{2,\cdot}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	r	$n_{r,1}$	$n_{r,2}$	\cdots	$n_{r,c}$	$n_{r,\cdot}$
Total		$n_{\cdot,1}$	$n_{\cdot,2}$	\cdots	$n_{\cdot,c}$	$n_{\cdot,\cdot} = n$

Tabela 3.2: Forma geral de uma Tabela de Contingência bidimensional.

$$n_{i,\cdot} = \sum_{j=1}^c n_{i,j}, \text{ com } i = 1, 2, \dots, r \quad n_{\cdot,j} = \sum_{i=1}^r n_{i,j}, \text{ com } j = 1, 2, \dots, c$$

$$n_{\cdot,\cdot} = \sum_{i=1}^r \sum_{j=1}^c n_{i,j} = \sum_{i=1}^r n_{i,\cdot} = \sum_{j=1}^c n_{\cdot,j} = n.$$

Note-se que numa tabela de contingência ($r \times c$) existem $r + c - 1$ ligações independentes.

As tabelas de contingência podem ser estudadas de diferentes modos consoante o tipo de margens da mesma, pois estas podem ser fixadas anteriormente à experiência. Neste caso, as frequências absolutas que aparecem em cada entrada do interior da tabela (e as correspondentes probabilidades) são condicionadas, e não conjuntas. Sendo assim, as Tabelas de Contingência podem ser classificadas como de margens livres, com uma das margens fixas e com as duas margens fixas, ou pré-definidas.

Neste resumo apenas serão abordadas as tabelas de contingência de margens livres, pois serão estas as utilizadas no estudo prático, do Capítulo 5.

Tabelas de Contingência de margens livres

Seja $p_{i,j}$ a probabilidade de uma observação pertencer às classes i da variável R e j da variável C . Então o vector aleatório das frequências observadas, $(n_{1,1}, n_{1,2}, \dots, n_{r,c})$, tem distribuição multinomial, o que implica que cada variável $N_{i,j}$ tenha distribuição marginal binomial de parâmetros $(n, p_{i,j})$ e que a frequência esperada de n observações

classificadas nas categoria i e j seja

$$E[N_{i,j}] = np_{i,j}, \text{ com } i = 1, \dots, r \quad j = 1, \dots, c$$

Os estimadores de máxima verosimilhança de $p_{i,j}$ e das frequências esperadas, $E[N_{i,j}]$ são, respectivamente:

$$\begin{aligned} \widehat{p}_{i,j} &= \frac{n_{i,j}}{n} \quad i = 1, \dots, r \quad j = 1, \dots, c, \\ \widehat{E}[N_{i,j}] &= n\widehat{p}_{i,j} = n\frac{n_{i,j}}{n} = n_{i,j} \quad i = 1, \dots, r \quad j = 1, \dots, c. \end{aligned}$$

Para testar a independência das duas variáveis em estudo, consideram-se as distribuições marginais.

Se as variáveis forem independentes, a função de probabilidades conjunta factoriza no produto das funções de probabilidades marginais. Deste modo, a independência é testada com as hipóteses

$$H_0 : p_{i,j} = p_{i.} \times p_{.j} \quad i = 1, \dots, r \quad j = 1, \dots, c \quad vs \quad H_1 : \exists(i, j) : p_{i,j} \neq p_{i.} \times p_{.j}. \quad (3.16)$$

Face à distribuição multinomial do vector de frequência $(n_{1,1}, n_{1,2}, \dots, n_{r,c})$ e admitindo que é válida a hipótese H_0 , a estatística

$$X^* = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}, \quad (3.17)$$

onde $e_{i,j} = E[N_{i,j}]$, segue assintoticamente uma distribuição de Qui-quadrado com $(r-1)(c-1)$ graus de liberdade (ver demonstração, por exemplo, em Crámer (1946)).

A regra de decisão do teste ao nível de significância α é a seguinte:

- Se $x_{obs} \geq f_{\chi^2_{(r-1)(c-1)}(1-\alpha)}^{-1}$, rejeita-se H_0 , ao nível de significância α ;
- Se $x_{obs} \leq f_{\chi^2_{(r-1)(c-1)}(1-\alpha)}^{-1}$, não se rejeita H_0 , ao nível de significância α ,

onde $f_{\chi^2_{(r-1)(c-1)}(1-\alpha)}^{-1}$ representa o quantil de probabilidade $(1-\alpha)$ de uma distribuição $\chi^2_{(r-1)(c-1)}$ graus de liberdade e x_{obs} é o valor observado da estatística de teste X^* .

É de realçar que o resultado (3.17) é assintótico, o que exige cuidados especiais ao ser utilizado em amostras de pequenas dimensão ou de dimensões maiores, mas em que haja celas com frequência baixa. É usualmente recomendado que o teste só seja aplicado quando os valores de $e_{i,j}$ são superiores a 5 se o número de graus de liberdade for 2 ou mais. Quando $e_{i,j}$ desce abaixo desse valor, é habitual fundir as classes em que tal acontece com outras, a fim de obter novas categorias com frequência esperada superior a 10 - se trabalharmos com 1 grau de liberdade -, ou superior a 5 no caso geral. Contudo, este procedimento não está totalmente correcto, uma vez que afecta a aleatoriedade da amostra e provoca perda de informação na fusão de categorias.

Quando não são verificadas as recomendações referidas anteriormente, a aproximação da distribuição do Qui-quadrado para a estatística X^* , pode não ser suficientemente boa. Neste caso é preferível usar o teste exacto de *Fisher*.

No programa R, a análise por Tabelas de Contingência pode ser efectuada usando os comandos `CrossTable` e `fisher.test`.

Capítulo 4

Robustez dos testes associados ao modelo ANOVA

Este capítulo é dedicado à apresentação de noções fundamentais em estatística robusta, de modo a conter o material necessário para reflectir sobre a robustez do teste- F , bem como dos métodos alternativos, apresentados na capítulo anterior.

4.1 Noções essenciais em robustez estatística

4.1.1 Generalidades

Muitos utilizadores da Estatística decidem aplicar técnicas de modelação estatística sem se preocupar com a validação das hipóteses assumidas nos respectivos modelos. E, em muitas dessas situações, a realidade não se comporta de acordo com as hipóteses assumidas, ou estas apenas se verificam aproximadamente. Quando tal facto acontece, os métodos estatísticos podem perder as suas boas propriedades, deixando de produzir resultados com a qualidade esperada.

Por exemplo, um estimador de um parâmetro pode ter grande divulgação por gozar

de propriedades de optimalidade quando a população segue uma certa distribuição e não gozar das mesmas propriedades perante outras distribuições da população. Ao utilizar esse estimador ignorando a distribuição subjacente aos dados, as estimativas que se obtêm podem passar de excelentes a desastrosas. Será preferível usar métodos que sejam bons de acordo com um modelo, mesmo não sendo ótimos, mas que se mantenham com boas propriedades em condições próprias do modelo.

Surge assim a necessidade de encontrar procedimentos, que não sejam muito sensíveis a pequenos afastamentos das hipóteses do modelo e que possuam boas propriedades, quer no modelo que se assume verdadeiro, quer nas suas vizinhanças. Os métodos com essas características são genericamente designados por métodos estatísticos robustos.

A palavra robustez é usada na linguagem corrente para designar a qualidade daquilo que é robusto (termo derivado da palavra latina *robustus*). Por sua vez o significado do termo robusto é, entre outros: forte, vigoroso, resistente, saudável, bem constituído (ver Pires e Branco (2007)).

O conceito de robustez está intimamente relacionado com o de resistência, mas não é coincidente e é mais amplo. De facto, a noção de resistência de uma técnica estatística traduz a ideia de que essa técnica conduz a resultados numéricos que resistem bem a variações nos dados - trata-se de pouca sensibilidade em relação a alterações numéricas. A robustez estatística, ao colocar-se em relação a afastamentos dos próprios modelos, considera a questão da sensibilidade dos métodos relativamente a alterações numéricas nas observações, que podem até decorrer de falha de modelação, mas também contempla outros tipos de afastamento das condições assumidas nos modelos.

Para concretizar a noção de robustez existem três abordagens fundamentais (ver Huber and Ronchetti (1981)): a *abordagem qualitativa*, que se baseia na noção de continuidade, a *abordagem quantitativa* que se baseia na noção de ponto de rotura (ou *breakdown point*) e a *abordagem infinitesimal*, baseada na noção de função de influên-

cia. A cada uma das abordagens corresponde um critério de avaliação da robustez expresso através dessas noções fundamentais.

Antes de formalizar as principais definições relativas a cada uma das abordagens, introduz-se a noção de funcional estatístico, que é indispensável neste contexto.

Considere-se uma população X , da qual se retira uma amostra aleatória (X_1, \dots, X_n) . Seja Ω um espaço métrico geral e $\mathfrak{F}(\Omega)$ uma família de distribuições F_θ sobre Ω , onde θ é um parâmetro desconhecido pertencente a um espaço paramétrico Θ , ou seja tal que $\mathfrak{F}(\Omega) = \{F_\theta, \theta \in \Theta\}$. Para simplificar a exposição considere-se ainda que F_θ possui uma função de densidade de probabilidade que se representa por $f_\theta(x)$ ou por $f(x, \theta)$.

Definição 4.1. Um funcional estatístico T é uma aplicação que a cada elemento de $\mathfrak{F}(\Omega)$ faz corresponder um número real, ou um vector de números reais.

□

Deste modo, um mesmo funcional estatístico aplicado a diferentes *f.d.p.* de \mathfrak{F} , poderá fazer-lhes corresponder diferentes valores reais.

Quando as distribuições dependem de parâmetros, a representação desses parâmetros pode ser considerada através de funcionais estatísticos.

Exemplo 4.1.1. O exemplo mais simples e intuitivo de funcional estatístico é a média, enquanto parâmetro ou indicador de tendência central.

Considere-se que a variável aleatória X segue uma *f.d.p.* F e que existe a esperança de X , sendo esta a característica de interesse na distribuição.

Assim, defina-se o funcional T com $T(X) = E(X)$.

Calculando o funcional na distribuição da população X , tem-se que

$$T(F) = \int x f(x) dx = \mu.$$

Se F_n denotar a função de distribuição empírica da amostra aleatória (X_1, \dots, X_n) retirada da população X , a média, enquanto funcional, pode ser calculada pelo valor do funcional T na distribuição $F_n \in \mathfrak{F}(\Omega)$. Assim efectuando o cálculo do mesmo funcional na distribuição empírica, vem que

$$T(F_n) = \int x dF_n(x) = \sum_{i=1}^n x_i \times \frac{1}{n} = \bar{x},$$

ou seja, obtém-se a média amostral, pois a função de distribuição empírica atribui o peso $\frac{1}{n}$ a cada observação x_i .

□

A representação funcional de estimadores apresenta algumas limitações, como por exemplo, o facto de nem todos os estimadores poderem ser representados por um funcional da distribuição empírica, como é o caso da variância amostral (não corrigida). No entanto, em geral existe uma representação funcional que é assintoticamente equivalente ao estimador $T_n(X_1, \dots, X_n)$, no seguinte sentido: existe um funcional $T(F_n)$ tal que $T(F_n) \neq T_n(X_1, X_2, \dots, X_n)$, mas a diferença $T_n(X_1, \dots, X_n) - T(F_n) \rightarrow 0$, quando n tende para infinito.

Felizmente a maior parte dos estimadores habituais, incluindo os desenvolvidos no âmbito da teoria da robustez, pertence a um destes dois grupos, ou seja, ou admite uma representação funcional ou é assintoticamente equivalente a um funcional.

Normalmente, existe uma extensão natural de T a um subconjunto convexo de $\mathfrak{F}(\Omega)$ que vai constituir o domínio do funcional. Se G representar a distribuição exacta de X , então $T(G)$ define o vector dos parâmetros de facto estimados, enquanto que $T(G_n)$ define o vector das estimativas. É evidente que para esta afirmação faça sentido deve verificar-se que

$$\lim_{n \rightarrow +\infty} T_n = \lim_{n \rightarrow +\infty} T(G_n) = T(G) \quad (\text{em probabilidade});$$

o que significa que o estimador T_n é um estimador consistente da “quantidade” $T(G)$. Assume-se que a distribuição G pertence a uma “vizinhança” do modelo paramétrico adoptado mas pode não coincidir com nenhum elemento do modelo paramétrico.

Definição 4.2. Seja T_n um estimador do parâmetro θ de um modelo paramétrico F_θ e T a sua representação funcional ou uma representação funcional que lhe seja equivalente. T diz-se consistente segundo Fisher (ou consistente à Fisher), nesse modelo, se e só se

$$T(F_\theta) = \theta, \quad \forall \theta \in \Theta \quad (4.1)$$

□

Encontram-se agora reunidas as condições para apresentar as principais da Teoria da Robustez, de acordo com as diferentes abordagens.

Abordagem qualitativa

A ideia intuitiva de robustez qualitativa prende-se com o facto de que se uma distribuição G se encontra próxima de uma outra distribuição F , então, para um estimador T , as estimativas $T(G_n)$ e $T(F_n)$ também devem estar próximas.

Definição 4.3. Seja F uma *f.d.p.* e $d(\cdot)$ uma distância adequada entre funções de distribuição. O funcional T diz-se contínuo em F se, para qualquer sucessão de funções de distribuição $\{G_n\}_{n \in \mathbb{N}}$ tais que $d(G_n, F) \xrightarrow{n} 0$, se tem que $|T(G_n) - T(F)| \xrightarrow{n} 0$.

□

Definição 4.4. Se T for um funcional contínuo em F , então $T(F_n)$ diz-se robusto no sentido qualitativo.

□

Se T for um funcional contínuo, então $T(F_n)$ é consistente.

Para avaliar a noção de proximidade entre funções de distribuição é necessário usar distâncias adequadas. A distância mais utilizada é a distância de Kolmogorov, definida por:

$$d_K(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|.$$

Do ponto de vista teórico, o conceito de estimador qualitativamente robusto é muito importante. No entanto, o tratamento matemático da continuidade destes funcionais é muito delicado e por isso não tem sido muito desenvolvido nem muito aplicado.

Abordagem infinitesimal

A avaliação da robustez infinitesimal é feita através de um dos conceitos mais importantes da Teoria da Robustez, designadamente o de função de influência.

Para uma questão de simplicidade de exposição e de notação, considere-se a definição seguinte para o caso univariado e uniparamétrico.

Definição 4.5. Seja F uma *f.d.p.* pertencente ao domínio do funcional T e seja $F_{x,\varepsilon}$ a *f.d.p.* contaminada em x , da forma

$$F_{x,\varepsilon} = (1 - \varepsilon)F + \varepsilon\Delta_x, \quad (4.2)$$

onde Δ_x representa a função de distribuição com massa unitária em x e $0 < \varepsilon < 1$. Chama-se função de influência de T em F a

$$IF(x; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T(F_{x,\varepsilon}) - T(F)}{\varepsilon}, \quad (4.3)$$

definida pontualmente nos pontos $x \in \Omega \subset \mathbb{R}$ para os quais o limite existe.

□

Uma outra forma de interpretar a função de influência é a de encará-la como sendo uma derivada que define o estimador, ou seja

$$IF(x; T, F) = \left. \frac{\delta}{\delta\varepsilon} T(F_{x,\varepsilon}) \right|_{\varepsilon=0},$$

uma vez que $[T(F_{x,\varepsilon}) - T(F)]/\varepsilon$, representa a taxa de variação média do funcional T quando a distribuição F é sujeita a uma proporção ε de contaminação.

Supondo válidas as condições de regularidade necessárias à existência da quantidade de informação de Fisher, a função de influência pode ser útil para o conhecimento da distribuição assintótica de um estimador.

Definição 4.6. Seja T_n um estimador e T a sua representação funcional, consistente à Fisher em relação a determinado modelo paramétrico F_θ . A eficiência assintótica absoluta de um estimador T_n é

$$e(T, F_\theta) = [V(T, F_\theta)J(\theta)]^{-1}, \quad (4.4)$$

onde $V(T, F_\theta)$ é a variância assintótica do estimador T_n em F_θ e $J(\theta)$ é a quantidade de informação de Fisher em relação a F_θ .

□

Nota 4.1. Note-se que $V(T, F) = \lim_{n \rightarrow \infty} nVar(T_n)$ e que se $0 < J(\theta) < \infty$ e T for consistente à Fisher se tem que

$$\int IF(x, T, F_\theta)^2 dF_\theta(x) \geq \frac{1}{J(\theta)}. \quad (4.5)$$

Em (4.5) a igualdade é verificada se e só se

$$IF(x, T, F_\theta) = J(\theta)^{-1} \frac{\delta \log f(x, \theta)}{\delta \theta}$$

□

Um dos motivos que faz com que a abordagem infinitesimal seja muito popular, deve-se ao facto desta possibilitar a construção de novos estimadores, a partir de estimadores iniciais, melhorando-os quer em termos de robustez, quer em termos da eficiência.

Para além do conhecimento do interesse que a função de influência possa suscitar, pelas suas propriedades, é com base nessa função que se definem medidas importantes de robustez, que serão de seguida apresentadas:

Definição 4.7. Chama-se “*sensibilidade a grandes erros*” (*gross error sensitivity*) de T em F a

$$\gamma^* = \gamma^*(T, F) = \sup_x |IF(x; T, F)|. \quad (4.6)$$

□

Definição 4.8. T diz-se *B-robusto* em F se $\gamma^*(T, F) < \infty$.

□

A sensibilidade a grandes erros representa a perturbação máxima que uma contaminação infinitesimal ε , próxima de zero, pode ter no valor de uma estimativa.

Quanto maior for $\gamma^*(T, F)$, mais sensível é o estimador a erros grosseiros e, conseqüentemente, menos robusto; se a sensibilidade a grandes erros for infinita, então o estimador definido pelo funcional T não é robusto, de acordo com esta abordagem.

Abordagem quantitativa

A noção essencial a ter em conta nesta abordagem é a de ponto de rotura, o qual avalia o comportamento dos estimadores nas vizinhanças da distribuição $F \in \mathfrak{F}(\Omega)$.

Para introduzir esta noção, é necessário recorrer à definição de “vizinhanças de contaminação”.

Definição 4.9. Chama-se “vizinhança de contaminação” de F , de raio ε , ao conjunto das distribuições

$$\mathcal{V}_\varepsilon(F) = \{G : G = (1 - \varepsilon)F + \varepsilon H, H \in \mathfrak{F}\}, \quad (4.7)$$

onde $\varepsilon \in [0, 1]$.

□

As distribuições de $\mathcal{V}_\varepsilon(F)$ representam todas as possibilidades de contaminação do modelo F por uma qualquer distribuição H , com probabilidade ε . Se ε for igual a 0, $\mathcal{V}_0(F)$ apenas contém a distribuição F e não existe a possibilidade de contaminação. Se ε for igual a 1, $\mathcal{V}_1(F)$ inclui as piores contaminações possíveis que o estimador pode sofrer. Deste modo, é necessário saber qual é o pior efeito que um elemento de $\mathcal{V}_\varepsilon(F)$ pode causar no estimador T_n , designado por enviesamento assintótico máximo.

Definição 4.10. Chama-se *enviesamento assintótico máximo* de T na vizinhança de $\mathcal{V}_\varepsilon(F)$ a

$$b_{T,F}(\varepsilon) = \sup_{G \in \mathcal{V}_\varepsilon(F)} |T(G) - T(F)|. \quad (4.8)$$

□

É de notar que $b_{T,F}(1)$ é o pior valor que o enviesamento assintótico pode ter. O ponto de rotura é o maior valor de ε que faz com que $b_{T,F}(\varepsilon)$ não seja o pior possível.

Definição 4.11. Chama-se *ponto de rotura assintótico* do estimador T_n , representado pelo funcional T , ao valor dado por

$$\varepsilon^*(T, F) = \sup\{\varepsilon : b_{T,F}(\varepsilon) < b_{T,F}(1)\}. \quad (4.9)$$

□

A Definição 4.9 é muito útil teoricamente, mas, em geral, ε^* é muito difícil de calcular. Assim, tem bastante mais aplicação a versão amostral do ponto de rotura que se passa a apresentar.

Definição 4.12. Numa dada amostra (x_1, \dots, x_n) substituem-se m observações (quaisquer) x_{i_1}, \dots, x_{i_m} por valores arbitrários y_1, \dots, y_m e designe-se a nova amostra por (z_1, \dots, z_n) . O *ponto de rotura empírico* de T_n é definido por

$$\varepsilon_n^*(T_n) = \varepsilon_n^*(T_n; x_1, \dots, x_n) = \frac{1}{n} \max \left\{ m : \max_{i_1, \dots, i_m} \sup_{y_1, \dots, y_m} |T_n(z_1, \dots, z_n)| < \infty \right\}. \quad (4.10)$$

□

Note-se que em geral, $\varepsilon_n^*(T_n)$ não depende de (x_1, \dots, x_n) , ou seja, depende apenas do número de observações consideradas, e não dos seus valores. Em geral, $\varepsilon^*(T, F)$ é independente da distribuição de F .

Quando n tende para infinito, geralmente ε_n^* tende para $\varepsilon^*(T, F)$. O ponto de rotura empírico tem vindo a ser cada vez mais utilizado, uma vez que envolve conceitos simples e que não depende da distribuição de probabilidades subjacente.

Um bom estimador deve satisfazer diversos requisitos, nomeadamente em relação ao ponto de rotura, à eficiência, à sensibilidade a grandes erros, ou ao enviesamento assintótico máximo. No entanto, a harmonização e conciliação de todas essas propriedades não é tarefa fácil e tem sido um ponto fulcral de estudo recente da estatística robusta. Existem estimadores que são muito robustos mas que são pobres no que respeita à eficiência, particularmente em modelos Normais, como por exemplo, a mediana amostral no modelo de localização Normal. No entanto, existem outros estimadores que não apresentam boas propriedades no que diz respeito à robustez, mas são equivariantes, consistentes e eficientes sob o modelo Normal (como por exemplo, a média amostral e o desvio padrão amostral).

Sendo assim, surgiu a necessidade de procurar estimadores robustos, mantendo a eficiência em níveis aceitáveis ao trabalhar em modelos Normais. De seguida, apresentam-se algumas propostas da literatura que conseguem conciliar boas propriedades de robustez com boa eficiência em modelos Normais.

Em particular, os estimadores que se seguem inspiraram a construção da proposta da estatística robusta que será apresentada na secção 4.5.

4.1.2 Estimadores- M

Os estimadores- M foram inicialmente propostos por Huber (1964), como uma generalização dos estimadores de máxima verosimilhança (EMV).

Considere-se que (X_1, \dots, X_n) é uma amostra aleatória da população X , cuja distribuição depende de um parâmetro de localização θ desconhecido, com função de densidade $f(x, \theta)$.

O EMV de θ é o estimador $T_n = T_n(X_1, \dots, X_n)$ que maximiza a função de verosimilhança expressa pelo produto $\prod_{i=1}^n f(X_i, T_n)$; ou, de forma equivalente,

$$T_n \text{ minimiza } \sum_{i=1}^n (-\log f(X_i, T_n)).$$

A ideia inicial de Huber foi definir o estimador- M substituindo $-\log f$, por uma outra função ρ , onde ρ é uma função conveniente, definida em $\Omega \times \Theta$, unidimensionais.

Assim, supondo as condições de regularidade usuais, todo o EMV é um estimador- M (mas a implicação contrária não é verdadeira).

Definição 4.13. Um estimador- M é um estimador T_n tal que

$$T_n \text{ minimiza } \sum_{i=1}^n \rho(X_i, T_n), \tag{4.11}$$

onde ρ é uma função arbitrária (chamada *função objectivo*) definida em $\Omega \times \Theta$.

Em geral, a função objectivo satisfaz as seguintes propriedades:

- é não negativa, ou seja, $\forall x \in \mathbb{R}$;
- $\rho(0) = 0$;
- é par, isto é, $\rho(x) = \rho(-x), \forall x \in \mathbb{R}$;
- é não decrescente para $x > 0$;
- é uma função contínua.

Se existir a derivada de ρ , denotada por $\psi(x; \theta) = \frac{\delta\rho(x; \theta)}{\delta\theta}$, então o estimador- M também pode ser definido como o estimador T_n que satisfaz a equação

$$\sum_{i=1}^n \psi(X_i; T_n) = 0. \quad (4.12)$$

□

Neste contexto, a função ψ que consta na equação (4.12) é conhecida por função ψ do estimador- M e é de grande importância na construção de estimadores robustos. Pode provar-se, sob condições de regularidade, que

$$IF(x; T, F) = \frac{\psi(x, T(F))}{-\int [(\delta/\delta\theta)\psi(u, \theta)]_{\theta=T(F)} dF(u)},$$

ou seja, que a função de influência de uma estimador- M é proporcional à sua função ψ .

Este resultado é da maior importância para o desenvolvimento de estimadores robustos. De facto, partindo de funções ψ limitadas, é possível encontrar estimadores robustos por construção (veja-se, por exemplo, Hampel *et al.* (1986)).

Por outro lado, o mesmo resultado evidencia que nem todos os estimadores- M são robustos, dependendo de ser ou não limitada a sua função $\psi(x; \cdot)$. Por exemplo, a média é um estimador- M não robusto.

Por outro lado, dado um estimador suficientemente regular T , existe um estimador- M que lhe é assintoticamente equivalente sob o modelo considerado. Basta que se tome $\psi(x; \cdot) = IF(x; \cdot; F)$. Os dois estimadores terão as mesmas propriedades locais (tais como função de influência, sensibilidade, variância assintótica), embora possam diferir em relação a propriedades globais (tais como robustez qualitativa e ponto de rotura).

Por estes motivos, e porque a função ψ facilita a generalização a estimadores multivariados, a equação (4.12) é geralmente a preferida para caracterizar os estimadores- M .

Estas ideias fundamentais são mais facilmente compreendidas no contexto de modelos simples de localização e de escala.

Modelos de localização

Considere-se a definição de modelo de localização:

Definição 4.14. Um *modelo de localização*, com parâmetro de localização θ , consiste numa família de distribuições $\{F_\theta, \theta \in \mathbb{R}\}$, tal que para todo o x , $F_\theta(x) = F(x - \theta)$, onde $F = F_0$ representa uma função de distribuição univariada genérica que define a família do modelo.

□

Considerando as condições (4.11) e (4.12) reescritas para o caso particular do modelo de localização, o estimador- M T_n fica definido como solução do problema

$$T_n \text{ minimiza } \sum_{i=1}^n \rho(X_i - T_n), \quad (4.13)$$

ou por

$$\sum_{i=1}^n \psi(X_i - T_n) = 0. \quad (4.14)$$

Deste modo a expressão (4.14) é a mais utilizada para definir estimador- M no contexto do modelo de localização. Essa expressão é também de grande utilidade para o estudo das principais propriedades do estimador- M .

De entre as propriedades desejáveis num estimador, é importante que ele seja invariante em relação a certas transformações de dados, de modo a assegurar que as estimativas do parâmetro são modificadas de forma coerente com transformações efectuadas sobre os dados originais. A verificação da invariância de um estimador é feita através das definições de equivariância.

Definição 4.15. Um estimador T_n do parâmetro de um modelo de localização diz-se *equivariante* em relação à localização, se dada a amostra aleatória (X_1, \dots, X_n) ,

$$T_n(X_1 + k, \dots, X_n + k) = T_n(X_1, \dots, X_n) + k, \forall k \in \mathbb{R}.$$

□

Através da equação (4.14), é imediato verificar que os estimadores- M são equivariantes em relação à localização.

No entanto, a equivariância em relação à escala pode não ser verificada, o que, a acontecer, constituiria um problema grave na utilização destes estimadores, uma vez que as estimativas do parâmetro seriam afectadas por mudanças de escala das observações.

Definição 4.16. Um estimador T_n do parâmetro de localização diz-se *equivariante em relação à escala*, se dada a amostra aleatória (X_1, \dots, X_n) ,

$$T_n(kX_1, \dots, kX_n) = kT_n(X_1, \dots, X_n), \forall k > 0.$$

□

Sendo assim, para tornar os estimadores- M equivariantes em relação à escala deve considerar-se simultaneamente um estimador auxiliar de escala, S_n , e definir o estimador- M de localização pela solução de

$$T_n \text{ minimiza } \sum_{i=1}^n \rho \left(\frac{X_i - T_n}{S_n} \right).$$

O principal exemplo deste tipo de estimadores é conhecido por *estimador de Huber* - *Proposta 2*. Expresso em termos da função ψ , este estimador é definido por:

$$\psi_b(x) = \begin{cases} -b, & \text{se } x < -b \\ x, & \text{se } |x| \leq b \\ b, & \text{se } x > b. \end{cases}$$

É fácil ver que, se S_n for um estimador equivariante em relação à escala, então kS_n , é solução da equação que se obtém substituindo os X_1, \dots, X_n por kX_1, \dots, kX_n na equação anterior.

Sendo o parâmetro de interesse um parâmetro de localização, de acordo com Hampel *et al.* (1986), o estimador de escala S_n deve ser o mais robusto possível, não sendo por isso muito importante a sua eficiência, uma vez que este é apenas um estimador auxiliar.

Note-se que os estimadores- M são definidos por equações implícitas. Logo, na prática e na maior parte dos casos é necessário, utilizar métodos iterativos para o cálculo das estimativas, nomeadamente o método de Newton-Raphson.

Modelos de escala

Da mesma forma que o modelo de localização é focado numa medida de localização e na translações de dados, o conceito de modelo de escala está associado a uma medida de dispersão e relacionado com mudança de escalas dos dados.

Definição 4.17. Um modelo de escala, com parâmetro de escala σ , consiste numa família de distribuições $\{F_\sigma, \sigma \in \mathbb{R}^+\}$, tal que $F_\sigma(x) = F\left(\frac{x}{\sigma}\right)$, onde $F = F_1$ representa uma função de distribuição univariada genérica que define a família do modelo.

□

Num modelo de escala, as propriedades de equivariância (em relação à localização e à escala) são análogas às mesmas propriedades num modelo de localização.

Sendo assim, encontram-se reunidas as condições para definir um estimador- M de escala.

Definição 4.18. Um estimador $\hat{\sigma}_n$ diz-se um estimador- M de escala se satisfaz uma

equação da forma:

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}_n} \right) = b, \quad (4.15)$$

onde $\hat{\mu}$ é uma estimativa obtida através de um estimador equivariante em relação à localização e à escala, e b é uma constante positiva, $b \in]0, 1[$, tal que $\int \rho(u) dF(u) = E[\rho(X)]$, onde F representa a distribuição referida na Definição 4.17.

□

4.1.3 Estimadores- S

Os estimadores- S foram originalmente propostos por Rousseeuw and Yohai (1984), no contexto do modelo de regressão, derivando o seu nome do facto do processo de construção integrar um estimador de escala (*scale*).

Genericamente designa-se por estimador- S um estimador que minimiza uma estimativa de escala conveniente.

Definição 4.19. Considerem-se, para qualquer $t \in \mathbb{R}$, os resíduos $X_1 - t, \dots, X_n - t$, onde t é uma estimativa de localização e $\hat{\sigma}_n(t)$ é a respectiva estimativa- M de escala, a qual satisfaz a equação

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{X_i - t}{\hat{\sigma}_n(t)} \right) = b, \quad b \in]0, 1[.$$

O *estimador- S de escala* é solução do problema

$$S_n \text{ minimiza } \hat{\sigma}_n(t)_{t \in \mathbb{R}}. \quad (4.16)$$

□

Este tipo de estimadores têm sido muito recomendados na literatura mais recente. A evolução dos meios computacionais contribuiu para que os estimadores- S fossem melhor estudados e actualmente sejam dos estimadores de escala mais utilizados. Até porque permitem controlar o seu grau de eficiência quando aplicados a populações

Normais. Por outro lado, também são utilizados na construção de estimadores- MM no modelo de regressão linear.

4.2 Robustez do teste- F

O trabalho desta secção foi apoiado pela consulta da tese de mestrado de Pontes (2009).

Como se realçou antes, para garantir o bom funcionamento de muitos procedimentos estatísticos é indispensável que os dados verifiquem um conjunto de suposições.

O teste- F , utilizado em ANOVA para comparar médias de populações normais, independentes e com variâncias iguais, é um dos testes mais usados pelos utilizadores estatísticos. A utilização deste método pressupõe a verificação das suposições subjacentes ao seu bom funcionamento, de forma a que fique probabilisticamente garantida a validade das conclusões. Por isso, é essencial que se desenvolvam métodos para comparação de médias que sejam robustos, no sentido em que, se as observações forem provenientes de populações com distribuição Normal, esses métodos funcionem bem; mas que também funcionem tão bem (ou quase) quando as observações seguem outra distribuição que se afaste um pouco da Normal. Para além disso, há ainda que ter em conta a ocorrência de valores atípicos que, no caso da estatística do teste- F , podem introduzir grandes alterações nas estimativas, conduzindo a estimativas de valores médios bem longe de traduzir a tendência central dos dados. Finalmente, é ainda de relembrar que, se as observações se afastarem da distribuição Normal, e mesmo que ambos os termos do quociente da estatística F se aproximem da distribuição de um Qui-quadrado, não é possível assegurar a independência necessária para se concluir a distribuição de Fisher do quociente.

Em seguida, apresenta-se um pequeno resumo de uma pesquisa realizada com base num conjunto de publicações sobre este tema, onde se dá conta das consequências das violações das hipóteses do modelo ANOVA, onde se indica um caminho ou uma melhor

alternativa a seguir.

Analisando livros de nível introdutório mais elementar como, por exemplo, Spiegel (1977), observa-se que na maior parte dos casos estes apenas apresentam o modelo ANOVA e as suas hipóteses de validação, sem qualquer salvaguarda perante o facto de estas suposições não serem verificadas.

À medida que se progride no nível de especialização das obras, consegue-se perceber uma preocupação acrescida com a aplicação correcta do modelo ANOVA, referindo não só as preocupações com o modelo quando não são verificados os pressupostos do mesmo, mas também as formas de os verificar.

Como exemplo, refere-se o livro de Fonseca (2001), que alerta para o facto da violação da hipótese de independência, pois “pode afectar seriamente as inferências sobre os valores médios, mais até do que a violação da normalidade”. Afirma também que a violação da normalidade “afectará mais as inferências sobre variâncias, e que a não homocedasticidade das variáveis resposta terá pouco efeito se as dimensões amostrais forem todas iguais”. Estas afirmações encontram-se na mesma linha das apresentadas por Ross (1987) e Montgomery and Ruger (1994), que acrescentam que a potência do teste é maximizada se as dimensões das amostras forem iguais.

É de realçar que tanto Scheffé (1959) como Lindman (1992) consideram que o teste- F é robusto relativamente à falha de suposição da distribuição Normal e concordam também que pequenos desvios em relação à falha da homocedasticidade, quando as amostras possuem iguais dimensões, afecta pouco as conclusões do teste- F , incluindo a sua função potência. No entanto, Scheffé (1959) não deixa de referir que acredita que existam testes que são tão robustos como o teste- F , em termos de nível de significância e menos potentes quando se trata de populações normais, mas mais potentes para populações de distribuições não normais. Existe um aspecto importante de referir em Lindman (1992), pois o autor preocupa-se em sugerir formas para testar se as variâncias são iguais e mostra como procurar métodos alternativos, chamando a atenção para a possibilidade de se efectuar transformações nos dados, de modo a que respeitem os

pressupostos.

Pesquisando a bibliografia para leitores de formação mais avançada na área da estatística, encontram-se livros como Neter *et al.* (1996) e Driscoll (1996), que afirmam que o teste- F é robusto quanto à falha da Normalidade. Driscoll (1996) argumenta a sua afirmação através de um pequeno estudo de simulação, onde analisa o comportamento do nível de significância do teste no caso de amostras provenientes de populações não normais. Esse autor conclui que o leitor não precisa de se preocupar com a falta de Normalidade dos dados, podendo dessa forma elaborar o teste sem grandes consequências.

Contudo Neter *et al.* (1996) é mais elucidativo e justifica a sua afirmação explicando que o nível de significância atingido é ligeiramente superior ao valor nominal e que, para além disso, a potência do teste atingida é ligeiramente inferior à que se teria se os dados fossem normais. Os mesmos autores fazem também referência à robustez: para amostras de iguais dimensões, quando há ausência de homocedasticidade, o nível de significância é apenas ligeiramente superior ao valor especificado. No entanto, os autores realçam a preocupação com a falha da suposição de independência, pois esta tem sérios efeitos sobre os resultados do teste- F . Sugerem para esta situação uma transformação dos dados; quando esta possível solução não for suficiente, aconselham o uso de procedimentos não paramétricos. Estes autores vão mais longe e sugerem que se atribuem às observações os respectivos números de ordem de forma crescente. De seguida elaboram o teste- F como usualmente, mas agora baseado nos números de ordem das observações, ou seja, substituindo os valores observados pelos respectivos números de ordem, numa linha semelhante à da definição do coeficiente de correlação de Spearman a partir do de Pearson.

Depois de elaborada esta pesquisa por diversos livros disponíveis que abordavam o modelo ANOVA, concluiu-se que existem autores que se preocupam com a validação das hipóteses do modelo, referindo que quando estas não são validadas, a melhor solução é o uso de testes não paramétricos, pois estes não dependem da distribuição dos dados.

No entanto, existem muitos autores que não dão a devida importância a este assunto, transmitindo ao utilizador total liberdade para o uso do modelo, sem qualquer tipo de precaução.

4.3 Versões robustas dos métodos paramétricos alternativos

Nesta secção serão apresentadas versões robustas dos métodos paramétricos alternativos resumidos no capítulo anterior, nomeadamente o Teste de Welch e o Teste de Brown-Forsythe.

Estas versões são apresentadas em Almeida *et al.* (2008), pelo que esse artigo constitui o suporte da presente secção.

Teste de Welch modificado

Os autores de Keselman and Wilcox (1999) propõem uma alternativa robusta do teste de Welch que consiste num teste de igualdade de parâmetros de localização. A estatística do teste utiliza estimadores robustos dos parâmetros de localização e de dispersão, nomeadamente a média aparada e a variâncias *winsorizadas*, ao invés dos estimadores usuais. Sugerem ainda que os valores críticos associados a um particular nível de significância sejam obtidos através do método de *bootstrap*.

Apresenta-se de seguida a versão robusta do teste de Welch:

Considere-se $Y_{i,j}$ com $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n_i$ variáveis aleatórias independentes e com distribuição Normal, ou seja, $Y_{i,j} \sim N(\mu_i, \sigma_i^2)$, com $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n_i$.

Formalizando o teste de igualdade de médias (3.1) em termos de igualdade de médias

aparadas, as hipóteses são:

$$H_0 : \mu_{a_1} = \mu_{a_2} = \dots = \mu_{a_k} \quad vs \quad H_1 : \exists i \neq j : \mu_{a_i} \neq \mu_{a_j}. \quad (4.17)$$

Sejam $Y_{(1)i} \leq Y_{(2)i} \leq \dots \leq Y_{(n_i)i}$ as observações ordenadas do i -ésimo grupo.

A i -ésima média amostral aparada a $(\alpha \times 100)\%$ de cada lado, com $0 < \alpha < \frac{1}{2}$, representada por $\bar{Y}_{\alpha_{i..}}$, é calculada do modo seguinte: ordenam-se as observações por ordem crescente e em seguida, calcula-se a média amostral das observações centrais, omitindo as $[\alpha n_i]$ inferiores e as $[\alpha n_i]$ superiores (onde $[t]$ significa o maior inteiro contido em t). Desta forma, a dimensão amostral do i -ésimo grupo de observações efectivamente considerada para o cálculo da média do grupo é $n_i - 2[\alpha n_i]$.

Sendo assim,

$$\bar{Y}_{\alpha_{i..}} = (n_i - 2[\alpha n_i])^{-1} \sum_{j=[\alpha n_i]+1}^{n_i - [\alpha n_i]} Y_{(j)i}. \quad (4.18)$$

Defina-se também a i -ésima média *winsorizada* por:

$$\bar{Y}_{w_{i..}} = n^{-1} \sum_{j=1}^{n_i} X_{i,j},$$

onde

$$X_{i,j} = \begin{cases} Y_{([\alpha n_i]+1)i}, & \text{se } Y_{i,j} \leq Y_{[\alpha n_i]i} \\ Y_{i,j}, & \text{se } Y_{[\alpha n_i]i} < Y_{i,j} < Y_{(n_i - [\alpha n_i])i}, \\ Y_{(n_i - [\alpha n_i])i}, & \text{se } Y_{i,j} \geq Y_{(n_i - [\alpha n_i])i} \end{cases} \quad (4.19)$$

e a variância amostral *winsorizada* do i -ésimo grupo, por:

$$S_{w_i}^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{i,j} - \bar{Y}_{w_{i..}})^2. \quad (4.20)$$

De acordo com Wilcox (1996), uma estimativa da variância da média amostral aparada é dada por:

$$\widetilde{S}_{w_i}^2 = \frac{(n_i - 1)S_{w_i}^2}{(n_i - 2[\alpha n_i])(n_i - 2[\alpha n_i] - 1)}. \quad (4.21)$$

Note-se que tanto a média α -aparada como a média *winsorizada* têm função de influência limitada, ao contrário da média amostral.

Fazendo uso dessas estatísticas robustas, poder-se-á apresentar uma versão robusta da estatística de teste de Welch, apresentada em (3.2), onde são utilizadas as médias aparadas dos grupos $\bar{Y}_{\alpha_i, \cdot}$, as variâncias amostrais *winsorizadas* $S_{w_i}^2$ e $\sum_{i=1}^k (n_i - 2[\alpha n_i])$, em vez de n , de forma a obter a estatística

$$W_{\alpha}^* = \frac{\sum_{i=1}^k w_{\alpha_i} \left(\bar{Y}_{\alpha_i, \cdot} - \tilde{Y}_{\alpha, \cdot} \right)^2 / (k-1)}{\left(1 + \frac{2(k-2)}{(k^2-1)} \sum_{i=1}^k \frac{\left(1 - \frac{w_{\alpha_i}}{\sum_{i=1}^k w_{\alpha_i}} \right)^2}{n_i - 2[\alpha n_i] - 1} \right)}, \quad (4.22)$$

onde $w_{\alpha_i} = \frac{n_i - 2[\alpha n_i]}{S_{w_i}^2}$ e $\tilde{Y}_{\alpha, \cdot} = \sum_{i=1}^k \left(\frac{w_{\alpha_i} \bar{Y}_{\alpha_i, \cdot}}{\sum_{i=1}^k w_{\alpha_i}} \right)$.

Sob H_0 , em (4.17), tem-se que $W_{\alpha}^* \xrightarrow{D} X$, onde $X \sim F(k-1; df_{\alpha})$, em que

$$df_{\alpha} = \left[\frac{3}{k^2-1} \sum_{i=1}^k \frac{\left(1 - \frac{w_{\alpha_i}}{\sum_{i=1}^k w_{\alpha_i}} \right)^2}{n_i - 2[\alpha n_i] - 1} \right]^{-1}. \quad (4.23)$$

e \xrightarrow{D} significa a convergência em distribuição.

O teste apresentado, não se encontra implementado, no programa R. No entanto, como vai ser utilizado no capítulo seguinte, o algoritmo deste teste encontra-se em apêndice.

Teste de Brown-Forsythe modificado

Mehrotra (1997) apontou uma inadequação na aproximação proposta por Brown and Forsythe (1974b) para a distribuição da estatística de teste BF sob H_0 em (3.4).

Sendo assim, a estatística do teste de Brown-Forsythe, apresentada em (3.5) como seguindo aproximadamente a distribuição $F_{(k-1, df)}$, onde df é dado por (3.7), sofreu

uma correcção. Em Mehrotra (1997) mostra-se que sob H_0 , $BF_M \approx F_{(df_1, df)}$, onde

$$df_1 = \frac{\left(\sum_{i=1}^k S_i^2 - \frac{\sum_{i=1}^k n_i S_i^2}{n} \right)^2}{\sum_{i=1}^k S_i^4 + \left(\frac{\sum_{i=1}^k n_i S_i^2}{n} \right)^2 - 2 \frac{\sum_{i=1}^k n_i S_i^4}{n}}. \quad (4.24)$$

O algoritmo de implementação deste teste, encontra-se em apêndice.

4.4 Outros estudos

Para além das versões robustas dos teste de Welch e de Brown-Forsythe descritas na secção anterior, têm surgido diversas publicações que confirmam a importância que continua a ser dada à investigação do presente tema. De facto, fazendo uma breve pesquisa bibliográfica constata-se a publicação de muitos artigos, mesmo quando se restringe a pesquisa a trabalhos posteriores aos anos 90. É provável que o desenvolvimento computacional e informático das últimas décadas seja o grande responsável pela motivação dos investigadores, uma vez que actualmente existem meios que facilitam o estudo computacional de metodologias já antigas e por outro lado, estimulam o desenvolvimento de novas propostas teóricas.

Sem pretender apresentar uma lista exaustiva, exemplificam-se alguns estudos feitos neste sentido.

Existe um grande número de artigos com estudos de simulação onde é comparado o desempenho dos métodos já existentes. Referem-se, por exemplo, Reed and Stark (1995), Buning (1997), Driscoll (1996) e a tese de mestrado de Pontes (2009). Muitos outros trabalhos têm em vista a divulgação de novos métodos baseados na ponderação de observações e/ou na substituição de médias amostrais por outras estatísticas mais robustas, como por exemplo em Babu (1999) e em Amado *et al.* (2010). Neste último,

os autores sugerem que se use a função de influência, para ponderar as observações nos estimadores de mínimos quadrados das médias que contribuem para o cálculo da estatística do teste. Kulinskaya and Dollinger (2007) propõem que a contribuição das observações para a estatística do teste seja determinada pela utilização de diferentes estatísticas robustas. A ideia generaliza o trabalho de Wilcox (1995), em que era considerada a média α -aparada em alternativa à média amostral. Também alguns autores têm estudado o assunto do ponto de vista *Bayesiano*, como por exemplo, Bian (1995) ou Gelman (2005).

Não se pode deixar de referir o imenso trabalho realizado em campos de aplicações específicas, como é o caso a título de exemplo recente, em De Haan *et al.* (2009).

4.5 Proposta de uma estatística de teste robusta (*FHubS*)

Como se focou nas secções anteriores, continua a haver uma grande preocupação sobre a robustez da estatística de Fisher, no sentido de procurar os melhores processos para efectuar o teste de igualdade de médias de modo a que produzam bons resultados, quando as condições do modelo ANOVA são verificadas, e também sob condições mais gerais do que as tradicionalmente assumidas.

Em termos de grande generalidade, pode dizer-se que a maioria das alternativas à estatística de Fisher procura comparar a variabilidade existente nos grupos com a variabilidade devida aos erros do modelo: quer seja pela modificação da distância que mede o afastamento existente nos dados; quer seja pela substituição de médias amostrais por outras medidas de tendência central; quer seja por ponderação das observações; quer seja ainda através de números de ordem.

Nesta secção propõe-se mais uma sugestão alternativa para a estatística do teste, cuja motivação se passa a resumir através das linhas gerais.

- (i) Em primeiro lugar, considera-se que é da maior importância preservar a interpretação do quociente que define a estatística- F , no sentido em que compara a proporção de variabilidade existente entre os grupos com a variabilidade dos erros - se a variabilidade entre os grupos é grande em relação à variabilidade dos erros, há motivos para rejeitar a igualdade de médias dos grupos.
- (ii) Em segundo lugar, procura-se um processo que permita concretizar a ideia básica contida em (i), mas que seja aplicável a qualquer distribuição conjunta das observações - Normal ou não, homocedástica ou não; quer haja, ou não, independência entre o numerador e o denominador da estatística. Neste sentido, a proposta é de natureza não paramétrica pois, como se verá, não depende da distribuição subjacente aos dados. Embora tenha sido delineada para distribuições contínuas e simétricas ou aproximadamente simétricas das observações, a ideia básica poderá ser posteriormente estendida a outras distribuições contínuas, com observações em escala numérica e até eventualmente, fora do contexto de independência, com as devidas adaptações.

Assim, a segunda linha mestra do desenvolvimento assenta no facto da regra de decisão do teste ser estabelecida a partir de um quantil elevado da verdadeira distribuição do quociente, a qual precisa de ser aproximada, pois é desconhecida. Sugere-se que essa aproximação seja efectuada por reamostragem, uma vez que são bem conhecidas as vantagens do processo, desde que seja usado com as necessárias cautela o que, hoje em dia, não exige meios computacionais sofisticados.

- (iii) A terceira ideia fundamental contida na proposta é a de concretizar os pontos anteriores através de estatísticas robustas. Note-se que, globalmente, a metodologia da proposta já é robusta, no sentido em que é aplicável a distribuições nas vizinhanças de algum modelo probabilístico eventualmente suposto. Refere-se agora o termo “estatística robusta” num sentido mais próximo do de resistência, ou seja, pensando em estatísticas que desempenhem o papel dos tradicionais indicadores de tendência central e de dispersão, mas de uma forma mais resistente

relativamente a observações atípicas - sem descurar preocupações de eficiência no caso da distribuição Normal das observações.

Um outro aspecto a ter em conta associado à robustez do processo, é o facto de ser necessário tomar as devidas cautelas, como foi referido em (ii), com a utilização de reamostragem. Efectivamente, uma das principais desvantagens da reamostragem é o risco de reamostrar, com reposição, observações atípicas, tendo como consequência que a proporção em que as observações atípicas existem na amostra original, seja artificialmente muito ampliada numa amostra *bootstrap*, o que é forçoso evitar.

Tendo em conta as considerações anteriores, tomaram-se as opções que se passam a descrever.

Seja $Y_{i,j}$, o valor da variável resposta para a j -ésima observação do i -ésimo grupo.

Pretende-se testar

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i \neq j : \mu_i \neq \mu_j \quad (4.25)$$

com base numa “modificação” da estatística

$$F = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{\bar{Y}})^2 / (k - 1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 / (N - k)},$$

explicitada em (2.15). No seguimento da presente secção denotou-se por n a dimensão da amostra.

Considere-se o numerador dessa estatística. No essencial, esse numerador mede a dispersão que existe entre as médias das observações dos diferentes grupos e a média global, tomando todas essas médias como indicadores de tendência central, respectivamente, dos grupos e de todas as observações. Por outro lado, o denominador é um indicador da dispersão que existe nas observações dentro de cada grupo.

No seguimento desta interpretação, i.e., tendo como objectivo medir duas formas distintas de variabilidade, seleccionou-se o quociente entre dois estimadores- S (pois

conforme se apresentou na secção 4.1.3, os estimadores- S são estimadores robustos de dispersão e que, combinados com estimadores- M , têm sido bastante recomendados na literatura sobre robustez).

Precisando melhor a proposta, a ideia é usar um estimador robusto de dispersão ($\hat{\sigma}_g^2$) para avaliar a dispersão existente entre estimativas robustas das médias dos grupos ($\hat{\mu}_i$), ou seja, no numerador utiliza-se um estimador- S ; no denominador, optou-se por um estimador- S da dispersão entre todas as observações ($\hat{\sigma}_\varepsilon^2$). Assim, considerou-se o quociente

$$FHubS = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_\varepsilon^2}.$$

Para o cálculo das estimativas robustas das médias dos grupos $\hat{\mu}_i$, usou-se um estimador de Huber.

No que diz respeito ao tópico (ii), sobre a aproximação da distribuição da estatística $FHubS$, decidiu-se usar o método *bootstrap* não paramétrico.

Pelo que foi referido, e mantendo presente que interessa aproximar a distribuição da estatística, prevenindo contaminação artificialmente ampliada nos dados, considerou-se conveniente fazer a reamostragem apenas sobre uma proporção elevada e central das observações. Deste modo, a reamostragem incidiu sobre a distribuição amostral aparada, sem ter em conta valores extremos de observações.

A partir dessa subamostra, o *bootstrap* decorre nos moldes usuais.

Para uma amostra de dimensão n , determina-se a amostra *bootstrap* seleccionando aleatoriamente e com reposição, um número n_a de observações de entre as $Y_{i,j}$ originais, com $n_a < n$, de forma a obter uma amostra *bootstrap* ($Y_1^*, Y_2^*, \dots, Y_{n_a}^*$). Para cada amostra *bootstrap*, calcula-se o valor da estatística $FHubS$. Repete-se o procedimento um número B de vezes, com B elevado, obtendo $\hat{F}HubS^{(1)*}, \hat{F}HubS^{(2)*}, \dots, \hat{F}HubS^{(B)*}$.

A distribuição empírica das estimativas $FHubS^{(j)*}$, para $j = 1, \dots, B$, constitui uma aproximação da verdadeira distribuição da estatística $FHubS$. Em particular, pode

estimar-se um quantil elevado da distribuição de $FHubS$, usando como estimador o correspondente quantil empírico $q_{(n_a, 1-\alpha)}^*$, determinado na amostragem *bootstrap*.

Deste modo, a regra de rejeição do teste (4.25) fica determinada a partir de um percentil elevado da distribuição obtida por *bootstrap*; ou seja, para efectuar o teste a um nível de significância $\alpha = P(F \geq q_{(n_a, 1-\alpha)})$, o nível é aproximado por $\alpha^* = \hat{\alpha} = \hat{P}(F \geq q_{(n_a, 1-\alpha)})$.

Assim, rejeita-se a hipótese nula quando, para a amostra original, o valor observado do quociente, for maior do que o percentil estimado, ou seja $FHubS_{n, obs} \geq \hat{q}_{n_a, 1-\alpha} = q_{(n_a, 1-\alpha)}^*$.

A metodologia que se acaba de explicar pode ser resumida no seguinte algoritmo:

- 1^o Determinar percentis $q_{n, \alpha}$ e $q_{n, 1-\alpha}$ da amostra observada, com α pequeno.
- 2^o Formar a subamostra de trabalho com base na subamostra das n_a observações contidas entre $q_{n, \alpha}$ e $q_{n, 1-\alpha}$.
- 3^o Com base na subamostra, calcular estimativas robustas das médias dos grupos $\hat{\mu}_i$ ($i = 1, 2, \dots, k$), usando o estimador de Huber.
- 4^o Calcular a estimativa- S da dispersão entre $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k$.
- 5^o Calcular a estimativa- S da dispersão entre as n_a observações consideradas;
- 6^o Calcular o quociente entre as duas estimativas- S , que será designado por $FHubS$.
- 7^o Repetir os passos de 1. a 4. para cada amostra *bootstrap* $\mathbf{Y}^{(b)*} = (Y_1^{(b)*}, Y_2^{(b)*}, \dots, Y_{n_a}^{(b)*})$, para $b = 1, \dots, B$, obtendo $FHubS^{(1)*}, FHubS^{(2)*}, \dots, FHubS^{(B)*}$.
- 8^o Calcular um quantil elevado $q_{(n_a, 1-\alpha)}^*$ da distribuição empírica de $FHubS^*$, obtida por *bootstrap* e tomar $\alpha^* = \hat{P}(F \geq q_{(n_a, 1-\alpha)})$, onde $q_{(n_a, 1-\alpha)}$ é o verdadeiro quantil da distribuição desconhecida da estatística $FHubS$.

9^o Tomar como estimativa do ponto crítico do teste $\hat{q}_{n_a, 1-\alpha} = q_{(n_a, 1-\alpha)}^*$.

10^o Comparar o valor observado da estatística calculada com todo o conjunto das observações originais, $FHubS_{n, obs}$ com o ponto crítico estimado. Se $FHubS_{n, obs} \geq \hat{q}_{n_a, 1-\alpha}$, rejeitar H_0 com um nível de significância α^* , que é aproximadamente α .

O algoritmo desta proposta encontra-se implementado no Apêndice B.

4.6 Aspectos Computacionais

Em termos computacionais também existem propostas implementadas que permitem aplicação robusta do modelo ANOVA, nomeadamente nos programas R, S-Plus e o SAS.

No S-Plus pode-se encontrar a *package* `robust`, com a função `anova.lmRob` que tem por base de implementação uma versão robusta do teste de Wald, utilizando estimadores-*MM*.

No *software* SAS, existe a função `ROBUSTREG`. No entanto, esta não facilita ao utilizador conhecer a teoria subjacente na sua construção. No mesmo sentido, existe a função `anova.lmrob`, apresentada na *package* `robustbase` do programa R, que permite que o utilizador opte por duas alternativas, uma com aplicação directa do teste de Wald e a outra com uma versão robusta do teste-*F*.

No capítulo seguinte será analisado um conjunto de dados reais, utilizando algumas funções já implementadas no programa R. Sendo assim, para além da base do programa, usaram-se as seguintes *packages*: `asbio`, `gregmisc`, `lawstat`, `lntest`, `MASS` e `robustbase`.

A *package* `asbio` foi utilizada para *winsorizar* os dados, no contexto do cálculo da

estatística modificada do teste de Welch. Recorreu-se às *packages* `lawstat` e `lntest` para a utilização dos testes implementados de Levene e de Durbin-Watson, respectivamente. A `gregmisc` foi usada para a análise das tabelas de contingência, e por fim, as *packages* `MASS` e `robustbase`, foram utilizadas para o cálculo dos estimadores- M e - S , respectivamente.

Capítulo 5

Modelação de um conjunto de dados reais

O presente capítulo ilustra a aplicação dos métodos focados nos capítulos anteriores, a um conjunto de dados reais.

Depois de descrever o problema de aplicação, apresenta-se a modelação por ANOVA com um factor de efeitos fixos. A modelação será efectuada supondo válidas as hipóteses do modelo; ou melhor, ignorando a necessidade de validar essas hipóteses. O objectivo desse procedimento é o de alertar os leitores para o facto de uma aplicação correcta do método exigir que se cumpram os pressupostos em que a sua construção assenta. A validação das hipóteses será apresentada na subsecção seguinte.

De seguida, considera-se a possibilidade de efectuar a comparação de médias, usando os métodos alternativos apresentados no Capítulo 3 e no Capítulo 4.

Este capítulo termina com alguns comentários sobre as conclusões encontradas com a utilização dos diferentes métodos.

5.1 Descrição do Problema

Há algum tempo atrás, a autora realizou um trabalho de análise de dados relativos à obesidade em alunos do Ensino Secundário da cidade de Aveiro, no contexto da disciplina de Seminário da Licenciatura em Ensino de Matemática.

Com esse estudo pretendeu-se perceber, no contexto escolar, qual a importância da obesidade e da hipertensão arterial e relacionar essas características com factores como a alimentação e a actividade física, numa população de jovens entre os 15 e os 19 anos de idade. Para isso procedeu-se à elaboração de uma sondagem, sendo a população constituída por todos os alunos matriculados no 10^o, 11^o e 12^o anos, nas Escolas Secundárias Homem Cristo, Dr. Jaime Magalhães Lima, José Estêvão e Mário Sacramento, no ano lectivo de 2008/2009.

Tendo em conta o objectivo geral já referido, foram definidos alguns objectivos específicos, nomeadamente, conhecer o padrão alimentar dos jovens e os hábitos relativos à prática de actividade desportiva.

Actualmente, a forma que é mais recomendada para avaliar o peso corporal é baseado no cálculo do Índice de Massa Corporal (IMC) - esta técnica é recomendada, inclusive, pela Organização Mundial de Saúde (OMS). O IMC de uma determinada pessoa é definido pelo quociente entre o seu peso (em quilogramas) e o quadrado da sua altura (em metros), ou seja, por

$$IMC = \frac{Peso(Kg)}{(Altura(m))^2} \quad (5.1)$$

O valor assim obtido estabelece o diagnóstico da obesidade e caracteriza também os riscos associados.

A tabela que se encontra a seguir, foi adaptada de acordo com as normas da OMS, e foi a utilizada na avaliação do diagnóstico do IMC dos inquiridos. No estudo que se segue, interessava apenas registar se os indivíduos eram ou não, classificados como obesos. Por isso consideram-se como casos de obesidade todos os alunos inquiridos que possuíam $IMC \geq 25kg/m^2$.

IMC (Kg/m^2)	Grau de Risco	Tipo de Obesidade
≤ 18	Magreza	Ausente
]18, 25]	Peso Saudável	Ausente
]25, 30]	Moderado	Sobrecarga ponderal (Pré-Obesidade)
]30, 35]	Alto	Obesidade de grau I
]35, 40[Muito Alto	Obesidade de grau II
≥ 40	Extremo	Obesidade de grau III ("Mórbida")

Tabela 5.1: Tabela dos graus de risco e do tipo de obesidade em função do IMC.

Para atingir os propósitos atrás apontados, foi elaborado um inquérito, a qual foi avaliado e validado pela Direcção-Geral de Inovação e de Desenvolvimento Curricular. O inquérito foi administrado a 500 alunos das escolas Secundárias de Aveiro. Este valor foi determinado a partir de um financiamento pré-definido de um total 30 euros.

Para efectuar esse estudo, foi então delineado um plano de amostragem considerando amostragem estratificada. Os estratos considerados foram as quatro escolas secundárias da cidade. Dentro de cada estrato, considerou-se o plano de amostragem aleatória simples sem reposição, por ser aquele que permitiria posteriormente realizar testes estatísticos baseados em amostras aleatórias, pois confere, a cada conjunto possível de n elementos, a mesma probabilidade de serem seleccionados para fazerem a amostra.

Considerando uma estratificação com repartição ou alocação proporcional, as dimensões dos estratos foram determinadas por

$$n_i = N_i/N \times n, \quad i = 1, \dots, 4,$$

onde N_i representa o número de elementos da população no estrato i , N representa o número total de elementos da população e n representa o número total de elementos da amostra, (neste caso $n = 500$).

Determinou-se o número de inquiridos para cada estrato, concretamente, $n_1 = 121$ alunos da E.S. Homem Cristo, $n_2 = 116$ alunos da E.S. José Estêvão, $n_3 = 127$ da

E.S. Dr. Jaime Magalhães Lima e $n_4 = 136$ da E.S. Mário Sacramento. No entanto, por questões de organização e logística das escolas, não foi possível recolher a totalidade dos inquéritos inicialmente prevista, tendo-se efectuado o estudo com base em 336 inquéritos.

Ao longo desse estudo inicial, efectuou-se uma análise estatística através de tabelas de contingência, testando (entre outros) a independência entre o IMC e as diferentes opções relativas à prática de exercício físico.

Aproveitando agora o conjunto de dados reais obtidos anteriormente em Costa (2009), no presente trabalho pretende-se aplicar o modelo ANOVA. Sem perda de generalidade o trabalho foi desenvolvido supondo amostragem aleatória simples. O objectivo é analisar a existência de diferenças significativas entre as médias de IMC dos inquiridos quanto às diferentes opções de prática de exercício físico, fora do contexto escolar.

Sendo assim, o factor em análise foi a “Prática de Exercício Físico”. Tendo em conta o género feminino e masculino dos alunos, definiram-se oito tratamentos correspondentes aos quatro hábitos de frequência de exercício sugeridos nos inquéritos, combinados com o género dos inquiridos. As opções referidas, para a Prática de Exercício Físico foram as seguintes:

- prática em anos anteriores e no ano do inquérito (SS);
- em anos anteriores, mas não no ano do inquérito (SN);
- no ano do inquérito, mas não em anos anteriores (NS);
- nunca efectuado (NN).

Uma vez que os tratamentos foram pré-determinados, o modelo ANOVA com um factor de efeitos fixos pareceu ser o modelo mais adequado.

Numa análise preliminar, averiguou-se a existência de interacção entre o género e as

várias opções de Exercício Físico dos inquiridos. Concluiu-se que deveria ser considerado o efeito de interacção. Deste modo, redefiniu-se um novo modelo ANOVA com apenas um factor de efeitos fixos e 8 tratamentos:

- prática em anos anteriores e no ano do inquérito do género feminino (FSS);
- em anos anteriores, mas não no ano do inquérito do género feminino (FSN);
- no ano do inquérito, mas não em anos anteriores do género feminino (FNS);
- nunca efectuado do género feminino (FNN);
- prática em anos anteriores e no ano do inquérito do género masculino (MSS);
- em anos anteriores, mas não no ano do inquérito do género masculino (SN);
- no ano do inquérito, mas não em anos anteriores do género masculino (MNS);
- nunca efectuado do género masculino (MNN).

Na Figura 5.1 apresenta-se um gráfico com as caixas de bigodes correspondentes aos valores de IMC nos 8 tratamentos considerados. Embora as representações dos diversos tratamentos não sejam muito idênticas, observa-se que as medianas estão relativamente próximas, verificando-se que o maior afastamento se situa entre as medianas dos grupos FSN e MNS, isto é, das raparigas que tinham praticado desporto em anos anteriores e já não praticavam no ano do inquérito e os rapazes que não praticavam desporto nem no ano do inquérito, nem em anos anteriores. Mesmo assim, esse afastamento máximo entre medianas é de 2.74 Kg/m^2 .

Não é por isso muito claro, nem evidente, se a prática de desporto e o género (feminino e masculino) têm influência significativa sobre os valores médios de IMC.

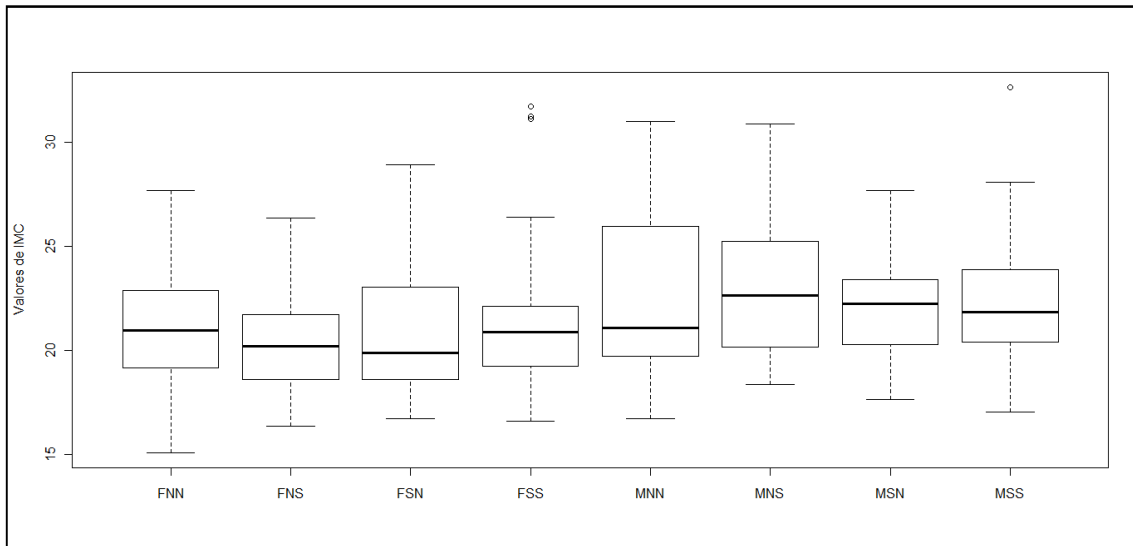


Figura 5.1: Caixas de bigodes dos valores de IMC relativas aos vários tratamentos do modelo.

5.2 Modelação por ANOVA com um factor de efeitos fixos

Nesta secção, em primeiro lugar, será feita a análise dos resultados obtidos aplicando aos dados um modelo ANOVA, sem efectuar a validação das hipóteses que o mesmo modelo confere.

Será considerada a formalização em termos de modelo de médias.

5.2.1 Aplicação da ANOVA supondo válidas as hipóteses do modelo

Considere-se que

$$Y_{i,j} = \mu_i + \varepsilon_{i,j}, \quad \text{com } j = 1, \dots, 336 \quad e \quad i = 1, \dots, 8,$$

onde $Y_{i,j}$ indica o valor de IMC do j -ésimo aluno que optou pela i -ésima opção de Prática de Exercício Físico conjugada com o género, e onde os erros $\varepsilon_{i,j}$ são *i.i.d.* com

distribuição $N(0, \sigma^2)$.

Pretende-se saber se o valor de IMC dos alunos depende da frequência de Prática de Exercício Físico conjugada com o género. Ou seja, interessa testar se as médias do IMC correspondentes aos vários tratamentos do factor poderão ser consideradas iguais.

Formalizando o teste, as hipóteses a testar são

$$H_0 = \mu_1 = \mu_2 = \dots = \mu_8 = \mu \quad vs \quad H_1 : \exists i : \mu_i \neq \mu.$$

Supondo verificadas as condições de normalidade, homocedasticidade e independência, a estatística de teste será:

$$F = \frac{MSG}{MSE} \sim F_{(k-1, N-k)}$$

e, neste caso, como há $n = 336$ alunos e $k = 8$ tratamentos, $n - k = 328$ e $F \sim F_{(7, 328)}$.

Efectuando os cálculos necessários com ajuda do programa **R**, obtém-se a seguinte tabela ANOVA:

Fonte de variação	df	SS	MS	Estatística	p_value
Entre Tratamentos	7	181.83	25.9763	2.9451	0.005265
Dentro dos Tratamentos	328	2893.02	8.8202		

Tabela 5.2: Tabela ANOVA para o conjunto de dados reais.

Como o valor- p é bastante pequeno ($p_value = 0.005265$) verifica-se que a hipótese nula deve ser rejeitada aos níveis de significância usuais, concluindo-se que pelo menos uma das opções de Prática de Exercício Físico conjugada com o género conduz a resultados médios de IMC significativamente diferentes.

5.2.2 Validação das hipóteses do modelo

Para garantir que a estatística do teste segue uma distribuição de Fisher, é essencial que se validem as hipóteses de Normalidade, homocedasticidade e independência dos erros do modelo. No entanto, na secção anterior, esses procedimentos de validação não foram efectuados. Em consequência, não existe nenhuma certeza de que o valor- p do teste anterior traduza a probabilidade pretendida.

Seguidamente, apresenta-se o trabalho efectuado para confirmar as hipóteses do modelo. Começar-se-á por averiguar a hipótese de distribuição Normal dos erros, seguidamente a homocedasticidade do modelo e por fim a independência dos erros.

Distribuição Normal

Esta análise poderá ser feita graficamente, através da elaboração de um gráfico de quantis, e de uma forma mais precisa, através da realização de um teste de ajustamento.

A partir das observações de IMC obteve-se o seguinte gráfico dos quantis:

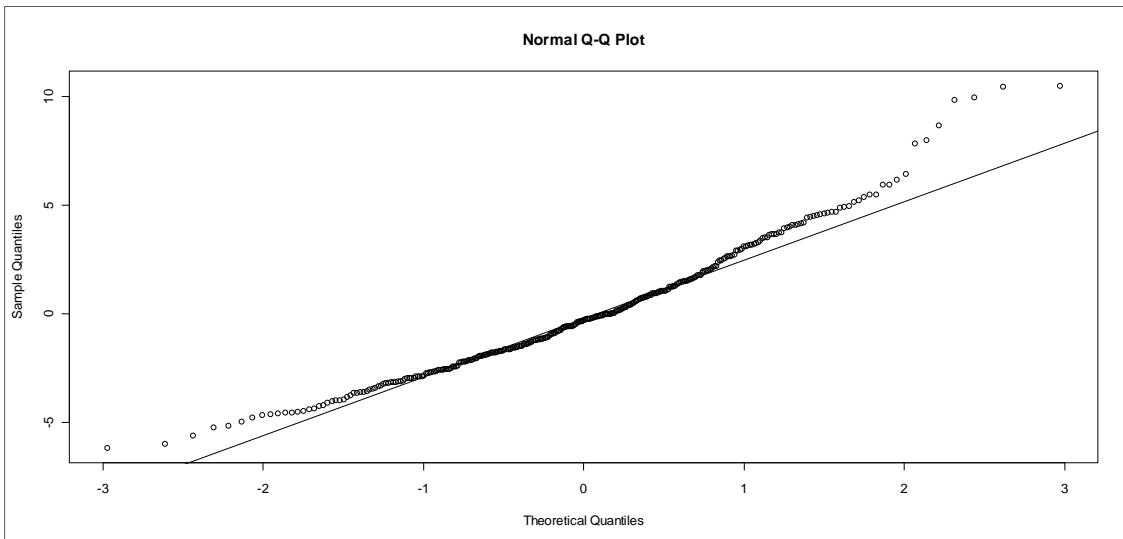


Figura 5.2: Q-Q plot da distribuição Normal dos resíduos.

Pela observação do gráfico, verifica-se que existe algum desvio do comportamento

linear, o que leva a suspeitar que a distribuição Normal não seja a distribuição mais indicada para modelar os valores de IMC dos inquiridos.

Contudo, como esta conclusão é subjectiva, efectuou-se também um teste de ajustamento, nomeadamente, o teste de Shapiro-Wilk.

O valor da estatística do teste de Shapiro-Wilk (S_w), obtido a partir do programa **R**, é de $S_w = 0.967$, com um $p_value = 6.422 \times 10^{-7}$. Como o p_value é reduzido, há motivos para rejeitar a hipótese nula, ou seja, pode concluir-se que os erros não segue uma distribuição Normal, como se poderia prever pela análise do Q-Q plot.

Homocedasticidade

Uma análise um pouco subjectiva poderá ser feita através da observação do seguinte gráfico, o qual não sugere que se rejeite a hipótese de variância constante dos erros do modelo.

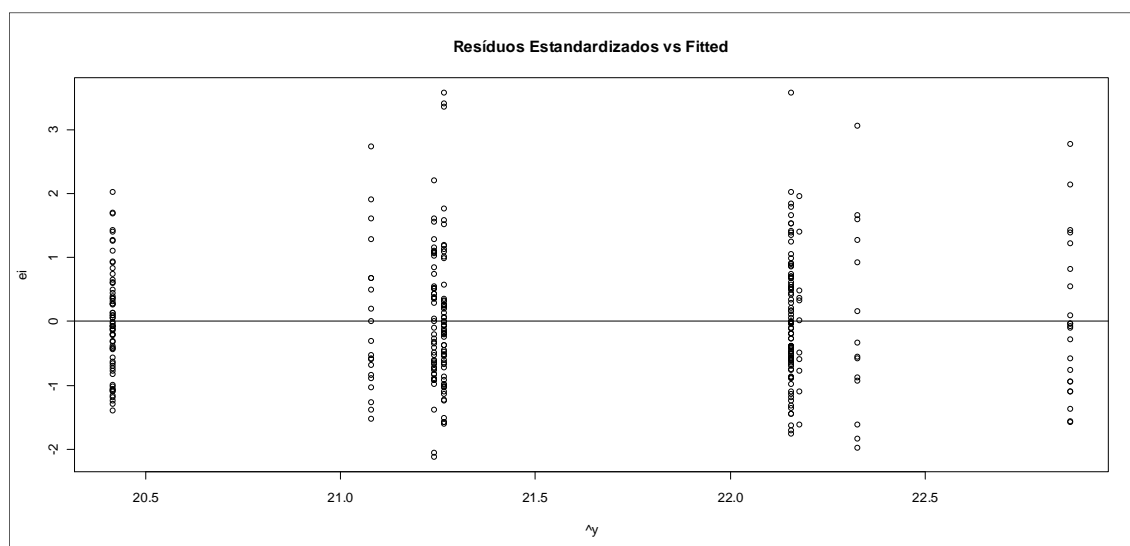


Figura 5.3: Diagrama de dispersão dos resíduos estandardizados vs $\hat{y}_{i,j}$.

A confirmação da homocedasticidade foi efectuada a partir do teste modificado de Levene, desenvolvido e aconselhado na secção 2.3.

Sendo assim, efectuou-se o teste de hipóteses

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_8^2 = \sigma^2 \quad vs \quad H_1 : \exists i : \sigma_i^2 \neq \sigma^2, \text{ para algum } i = 1, \dots, 8.$$

O valor da estatística do teste modificado de Levene (F_{ML}^*), obtido a partir do programa **R**, foi $F_{ML}^* = 1.3797$, com um $p\text{-value} = 0.213$. Concluiu-se, então, que não havia motivos para rejeitar a hipótese nula, ou seja, não havia evidências de que as variâncias do IMC para os vários níveis do factor em questão fossem diferentes.

Independência

Dada a natureza do problema, a confirmação desta hipótese é sobretudo académica, pois não é de crer que exista dependência entre os valores de IMC para os diferentes estudantes (salvo raras excepções).

Calculando os resíduos estandardizados e representando-os graficamente, a Figura 5.4 evidencia que os resíduos flutuam num padrão mais ou menos aleatório em torno de zero, sugerindo desta forma a eventual independência dos erros.

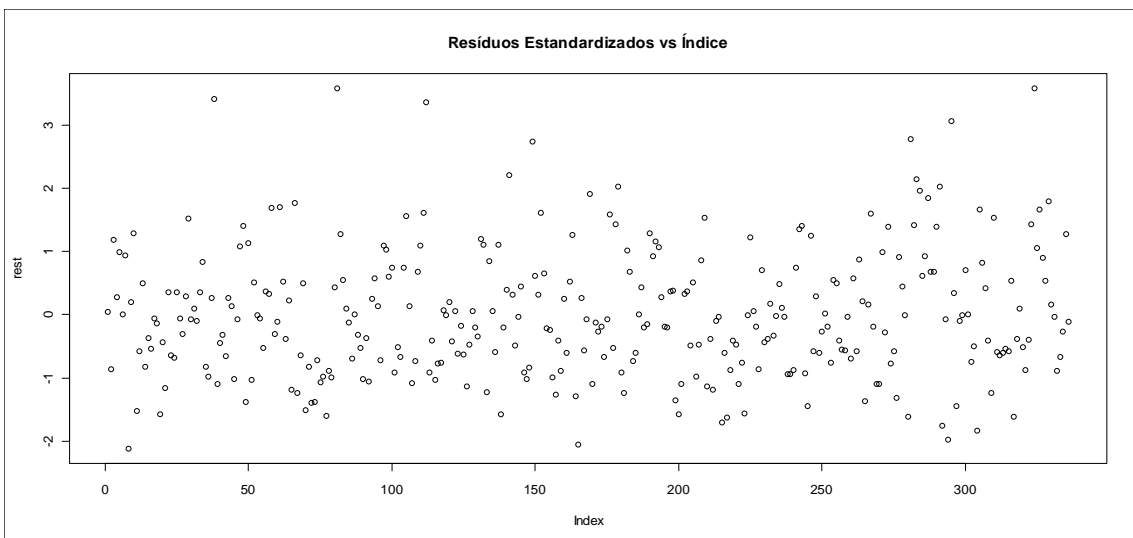


Figura 5.4: Diagrama de dispersão dos resíduos estandardizados vs índice das observações.

Ainda assim, a análise gráfica do diagrama de dispersão foi completada com um teste de aleatoriedade. O teste mais divulgado para detectar correlação entre termos sucessivos é o teste de Durbin-Watson (ver secção 2.3), que resultou no valor da estatística $D_w = 1.7114$, com um $p_value = 0.00668$. Portanto verifica-se que de acordo com o teste, a hipótese nula deveria ser rejeitada concluindo-se que os $Y_{i,j}$ não são independentes, para todo o $i = 1, 2, \dots, 8$ e $j = 1, 2, \dots, 336$.

No entanto, reafirma-se que o IMC diz respeito à individualidade física de cada inquirido. Por isso, e apesar da conclusão do teste, não parecer ser muito realista colocar em causa a validade desta hipótese.

Resumindo: em face das observações, há motivos para crer que os erros não seguem uma distribuição Normal. Uma vez que a hipótese da distribuição Normal foi claramente rejeitada, poder-se-ia considerar a solução de aplicar uma transformação aos valores de IMC, de modo que a variável transformada seguisse uma distribuição mais próxima da distribuição Normal. Não se optou por essa via, pelo facto deste trabalho ter como objectivo a análise de situações em que as observações se afastam das hipóteses supostas nos modelos.

5.3 Aplicação dos métodos alternativos ao teste- F

Nesta secção pretende-se ilustrar a aplicação de métodos alternativos ao teste- F , nomeadamente, o teste de Brown-Forsythe e o teste de Welch, com as respectivas modificações apresentadas no Capítulo 4 e o teste de Kruskal-Wallis.

Relembra-se que se pretende testar

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_8 = \mu \quad vs \quad H_1 : \exists i : \mu_i \neq \mu, \quad (5.2)$$

mas tendo agora em atenção que não estão reunidas as condições necessárias para que se possa usar o teste- F .

Teste Brown-Forsythe modificado e teste de Welch modificado

Ambos os testes de Brown-Forsythe modificado e teste de Welch modificado, pressupõem a distribuição Normal das observações e, como se viu, há motivos para rejeitar essa hipótese. Assim, nenhum dos dois testes deve ser utilizado no presente exemplo; Mas, realizaram-se esses testes para poder comparar conclusões.

Ao efectuar o teste de Brown-Forsythe modificado (secção 4.3), obteve-se o valor da estatística de teste $BF = 2.472603$, com $p_value = 0.02538148$. Isto significa que, para um nível de 1%, não há motivos para rejeitar H_0 ; no entanto, para $\alpha = 0.05$, já há motivos para rejeitar H_0 e concluir que existe pelo menos um dos tratamentos que conduz a resultados médios de IMC significativamente diferentes.

Ao utilizar o teste de Welch modificado, descrito na secção 4.3, com médias aparadas a 10% (5% para cada cauda), obteve-se o valor da estatística de teste $W_{0.1}^* = 3.014124$, com $p_value = 0.009183859$. Logo, a hipótese nula deve ser rejeitada, concluindo-se que pelo menos uma das opções de Prática de Exercício Físico conjugada com o género conduz a resultados médios de IMC significativamente diferentes.

Teste de Kruskal-Wallis

A região crítica de nível aproximadamente $\alpha = 0.05$ corresponde a valores da estatística T superiores ao quantil 0.95 de um $\chi_{(7)}^2$, que se encontra tabelado como 14.06714. Usando a expressão (3.15) obtém-se o valor observado $T = 19.6629$. Como $19.6629 > 14.06714$, há motivos para rejeitar H_0 , mesmo ao nível de significância $\alpha = 0.01$.

Deste modo, este teste também leva a concluir que existem diferenças significativas entre os valores médios de IMC para os diferentes tratamentos, embora neste caso, não seja assumida a distribuição Normal das observações.

5.4 Aplicação do método $FHubS$

Nesta secção é feita uma aplicação do método apresentado como proposta na secção 4.5.

No Apêndice B encontra-se o programa usado na obtenção dos resultados, que segue a metodologia apresentada anteriormente (ver secção 4.5).

Decidiu-se efectuar o teste com um nível de significância aproximadamente de 5%, o que levou à estimação do percentil 95% da distribuição da estatística $FHubS$. Para isso, formou-se uma subamostra aparando o conjunto dos dados iniciais em 2.5% em cada cauda. A subamostra ficou com 301 observações contidas entre os quantis $\hat{q}_{0.025}$ e $\hat{q}_{0.975}$ da amostra inicial. Com base nessa subamostra, calculou-se o quociente entre as duas estimativas- S , para cada uma das 1000 réplicas efectuadas.

Uma vez obtidas as réplicas do quociente $FHubS$ determinou-se a sua distribuição empírica, tendo-se estimado o seu percentil em $\hat{q}_{0.95} = 0.3586120$.

Tomando $\hat{q}_{0.95}$ como estimativa do ponto crítico do teste, verifica-se que este valor é superior ao valor observado da estatística de teste calculada a partir do conjunto das observações reais, pois $FHubS = 0.163458$.

Como $FHubS_{336,obs} < \hat{q}_{0.95}$ não temos motivos para rejeitar H_0 a um nível de significância de 5% (aproximadamente), o que permite não rejeitar a hipótese de que todas as opções de Prática de Exercício Físico conjugadas com o género, conduzem a resultados médios de IMC que não diferem significativamente.

Repetiu-se o procedimento para um subconjunto aparado num total de 10%, obtendo-se como estimativa do mesmo percentil, o valor $\hat{q}_{0.95} = 0.2515341$. Apesar das estimativas serem diferentes nos dois casos, as conclusões obtidas vão precisamente no mesmo sentido.

5.5 Modelação por Tabelas de Contingência

Como se referiu anteriormente, os dados em causa tinham já sido objecto de estudo através de Tabelas de Contingência.

A relação entre os dois tipos de tratamento estatístico baseia-se no seguinte: se uma ANOVA conduz à conclusão de que não há diferença significativa entre as médias dos tratamentos, ou de outro modo, que não há efeitos específicos dos tratamentos, isso significa que o IMC é interpretado como “independente” dos tratamentos, o que pode ser analisado com Tabelas de Contingência, efectuando um teste de independência entre o IMC e a Prática de Exercício Físico conjugada com o género.

Para isso, a variável IMC foi categorizada de acordo com a Tabela 5.1.

Sejam

$p_{i.}$ =P(uma observação ser classificada na categoria i da variável IMC), com $i = 1, \dots, 4$
e

$p_{.j}$ =P(uma observação ser classificada na categoria j da variável Prática de Exercício Físico conjugado com o género).

Formalizando o teste, as hipóteses que interessa testar são:

$$H_0 : p_{i,j} = p_{i.} \times p_{.j} \quad vs \quad H_1 : \exists(i, j) : p_{i,j} \neq p_{i.} \times p_{.j}, \quad (5.3)$$

com $i = 1, \dots, 4$ e $j = 1, \dots, 8$.

Através da análise dos inquéritos, construiu-se a seguinte Tabela de Contingência. Note-se que apenas se consideram quatro tratamentos, pois por um lado havia celas com uma frequência baixa e por outro lado, são os tratamentos previstos de acordo pela Organização Mundial da Saúde.

Relembra-se que no modelo ANOVA, foram considerados inicialmente os mesmos 4 tratamentos, mas que foi necessário passar a 8 tratamentos depois de ser efectuado o teste de interacção.

Analisou-se a Tabela de Contingência apresentada anteriormente a partir de um

		Opções de Prática de Exercício Físico			
		SS	SN	NS	NN
IMC (Kg/m ²)	≤ 18	12	4	12	6
]18, 25]	124	22	66	50
]25, 30]	15	5	8	6
]30, 35]	4	0	1	1
]35, 40[0	0	0	0
	≥ 40	0	0	0	0

Tabela 5.3: Tabela de Contingência das opções da Prática de Exercício Físico relativa aos valores de IMC.

Teste Exacto de Fisher. Não se usou aproximação pela distribuição de um Qui-quadrado, uma vez que existiam células da tabela que apresentavam frequências esperadas inferiores a 5.

Como o $p_value = 0.8316 > \alpha$, onde α representa os níveis de significância usuais, pode-se concluir que, de acordo com o Teste Exacto de Fisher, não existem motivos para rejeitar H_0 . Sendo assim, é de concluir que o IMC é independente dos tratamentos, ou seja das diferentes opções de prática de exercício físico.

Remetendo esta conclusão para os princípios em que assenta o modelo ANOVA, pode-se afirmar que não há diferenças significativas, entre as médias de IMC, devidas aos diferentes tratamentos.

5.6 Comentários comparativos

Numa análise preliminar, os valores de IMC foram modelados por um modelo ANOVA com dois factores de efeitos fixos. Verificou-se a existência de interacção significativa entre as opções de Prática de Exercício Físico e o género dos indivíduos. Sendo assim, considerou-se esse factor de interacção, definindo um novo modelo ANOVA com apenas um factor de efeitos fixos e 8 tratamentos. A partir da observação gráfica das caixas

de bigodes (Figura 5.1) relativas aos vários tratamentos do modelo, é plausível que não existam diferenças significativas entre as medianas.

Logo, a questão seguinte foi a de decidir, através de um teste de hipóteses de comparação de medidas de localização, se se confirmaria ou não, a conclusão retirada por observação gráfica.

Efectuada a análise dos dados através dos métodos apresentados neste trabalho, podem-se retirar conclusões a esse nível.

Considerando o modelo ANOVA e usando os testes de Welch e de Brown-Forsythe modificados concluiu-se que existiam diferenças significativas entre os valores médios de IMC nos vários tratamentos. A mesma conclusão foi apontada pelo teste de Kruskal-Wallis.

No entanto, as conclusões retiradas a partir da análise das Tabelas de Contingência e do teste *FHubS* foram em sentido oposto. Estes métodos não rejeitaram a hipótese inicial de igualdade de valores médios de IMC entre os 8 tratamentos.

Mas, deve-se ter em conta que ao utilizar o teste-*F*, o de Brown-Forsythe e o teste de Welch, a aplicação do modelo ANOVA estava incorrecta, pois não foi possível validar os pressupostos exigidos por esses testes.

Mediante a verificação dos pressupostos, concluiu-se que há motivos para rejeitar a distribuição Normal dos dados. Perante esta conclusão, poderia ter sido feita alguma transformação de dados para melhorar a proximidade à distribuição Normal, como por exemplo a utilização da transformação de Box-Cox. Contudo, esse procedimento não foi utilizado, tendo em conta que interessava investigar as limitações dos diferentes métodos, quando falham as hipóteses do modelo e, em particular a da distribuição Normal.

Sendo assim, os métodos propostos neste trabalho que se adequavam à análise dos dados, seriam apenas o teste de Kruskal-Wallis, as Tabelas de Contingência e o método proposto, *FHubS*. De entre estes, infelizmente não se obtiveram as mesmas conclusões.

Mas os resultados com a proposta $FHubS$ são interpretáveis com a mesma facilidade dos que se obteriam com o teste- F e coincidem com os apontados pela análise com Tabelas de Contingência.

Finalmente, aproveita-se para referir uma questão de natureza computacional que surgiu ao calcular o valor observado da estatística $FHubS$ para este exemplo, mas que pode ocorrer noutras exemplos de aplicação.

Durante a utilização do método, inicialmente registaram-se dificuldades de convergência na obtenção das estimativas- M . O problema ocorria com a reamostragem dos dados, por haver pouca variabilidade entre eles, fazendo com que o valor do MAD das observações fosse muito próximo do valor nulo. Esse facto causava dificuldades no cálculo do estimador de Huber, usando o comando pré-programado `hubers` do R. Para ultrapassar essa dificuldade alterou-se o algoritmo de cálculo, usando apenas as amostras *bootstrap* que não provocavam esse problema. O programa em apêndice salvaguardou essas situações.

Capítulo 6

Comentários finais

Com este trabalho pode-se concluir que a discussão sobre o tema em questão está longe de estar esgotada. São inúmeras as publicações neste âmbito, muitas delas recentes. Contudo, é de realçar que a maior parte dos estudos efectuados, tem sido feita relativamente às falhas da distribuição Normal e da homocedasticidade, não existindo grande análise perante a falha da independência dos dados.

Neste trabalho fez-se uma revisão dos principais métodos disponíveis para efectuar a comparação de diversas médias (ou medianas). Para além disso, investigou-se a possibilidade de modificar a estatística do teste- F , no sentido de usar o quociente entre duas estatísticas robustas destinadas a comparar a variabilidade entre grupos com a variabilidade dependente do acaso. Desse estudo resultou a apresentação de uma proposta original, que foi designada por estatística $FHubS$, pelo facto de utilizar o estimador de Huber, para estimar as médias dos grupos, e estimadores- S como estimadores da dispersão.

A distribuição da estatística $FHubS$ foi aproximada por reamostragem, considerando o método de *bootstrap*. Escolhendo um percentil empírico elevado e notando que esse percentil empírico estima o verdadeiro percentil da distribuição, é possível efectuar o teste de comparação de médias, comparando o valor observado da estatística com o

percentil estimado.

O teste que, deste modo, se propõe no trabalho é aplicável em situações em que possam falhar as hipóteses assumidas no modelo ANOVA, nomeadamente a da independência das observações.

O conjunto dos testes estudados foi posteriormente aplicado a um conjunto de dados reais, relativos ao Índice de Massa Corporal de estudantes do ensino secundário em Aveiro. Pretendia-se avaliar se a Prática de Exercício Físico conjugada com o género (feminino ou masculino) exerciam efeitos significativos sobre o valor médio do IMC.

Tendo em atenção a impossibilidade de validar as hipóteses assumidas pelos diferentes métodos, verificou-se que os dois testes disponíveis para usar com esses dados, eram o teste de Kruskal-Wallis e a modelação por Tabelas de Contingência. Mas, os dois testes apontaram resultados opostos. Ao usar o teste com a estatística $FHubS$ a conclusão do teste coincidiu com a que se obteve usando Tabelas de Contingência.

Toda a componente computacional foi executada com o programa **R** e as suas *packages*, de modo a que os procedimentos estudados e o proposto usem programas devidamente testados e, por outro lado, sejam facilmente usados por todos os interessados.

A presente dissertação deixa em aberto diversas questões que poderão vir a ser objecto de futuras pesquisas. A título de exemplo, será interessante estudar, as propriedades do método robusto proposto neste trabalho.

Por motivos temporais não houve oportunidade para fazer um estudo de simulação em diferentes distribuições e com várias dimensões amostrais, que permitisse analisar o desempenho da estatística $FHubS$. No entanto, considera-se que o comportamento da estatística $FHubS$ só pode ser avaliado na sequência de um tal estudo de simulação, o que deverá vir a ser feito no futuro.

Um outro melhoramento deste trabalho consistiria no aperfeiçoamento da imple-

mentação do algoritmo *FHubS*, pois este recorre a diversos ciclos, o que diminui o desempenho do mesmo.

Para finalizar, espera-se que esta tese possa dar um pequeno contributo para o estudo da robustez do teste-*F*, alertando os utilizadores para esta preocupação.

Bibliografia

- Almeida, A., Elian, S. and Nobre, J. (2008). Modificações e alternativas aos testes de Levene e de Brown e Forsythe para igualdade de variâncias e médias, *Revista Colombiana de Estadística* **31**, 241-260.
- Amado, C., Branco, J. A. and Pires, Ana M. (2010). *Robust methods for one and two-way ANOVA*. Comunicação oral em LinStat 's 2010, Tomar, Portugal.
- Babu, G. J. (1999). Robust One-Way ANOVA under Possibly Non-Regular Conditions, *Biometrical Journal* **41**, 321-339.
- Bian, G. (1995). Robust Bayesian Estimators in a one-way ANOVA models, *Test* **4**, 115-135.
- Brown, Morton B. and Forsythe, Alan B. (1974a). Robust Tests for the Equality of Variances, *Journal of the American Statistical Association* **69**, 364-367.
- Brown, Morton B. and Forsythe, Alan B. (1974b). The Small Sample Behavior of Some Statistics which Test the Equality of Several Means, *Technometrics* **16**, 129-132.
- Bünig, Herbert (1997). Robust Analysis of Variance, *Journal of Applied Statistics* **24**, 319-332.
- Cochran, W. G. (1980). Fisher and the analysis of variance. In: S. S. Fienberg, S. S.; Hinkley, D. V. (Ed.). R. A. *Füler*. Fisher: an appreciation. New York: Springer Verlag, p. 17-34.

- Conover, W. J. (1971). *Practical Nonparametric Statistics*. New York: John Wiley and Sons Inc..
- Costa, C. P. S. (2009). *O problema da Obesidade nos alunos do Ensino Secundário, da cidade de Aveiro*. Departamento de Matemática: Universidade de Aveiro.
- Crámer, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- De Haan, J. R., Bauerschmidt, S., Schaik, R. C., Piek, E., Buydens, L. M. C. and Wehrens, R. (2009). Robust ANOVA for microarray data, *Chemometrics and Intelligent Laboratory Systems* **98**, 38-44.
- Driscoll, W. C. (1996). Robustness of the ANOVA and Tukey-Kramer statistical tests, *Computers and Industrial Engineering* **31**, 265-268.
- Efron, B. (1979). Bootstrap methods: another look at the Jackknife, *The Annals of Statistics* **7**, 1-26.
- Efron, B. (1982). *The Jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.
- Eisenhart, C. (1947). The assumptions underlying the analysis of variance, *Biometrics* **3**, 1-21.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. and Mackenzie, W. A. (1923). Studies in crop variation II. The manurial response of different potato varieties, *Journal of Agricultural Science* **13**, 311-320.
- Fonseca, Jaime S. S. (2001). *Estatística Matemática*, Vol 2, 2^a ed., Lisboa: Edições Sílabo.

- Gelman, A. (2005). Analysis of Variance - Why it more important than ever (with discussion), *Annals of Statistics* **33**, 1-53.
- Hampel, Frank R. and Rousseeuw, Peter J. and Stahel, Werner A. (1986). *Robust Statistics- The Approach based on Influence Functions*. New York: John Wiley and Sons Inc..
- Huber, P.J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics* **35**, 73-101.
- Huber, Peter J. and Ronchetti, Elvezio (1981). *Robust Statistics*, 2^a ed., New York: John Wiley and Sons Inc..
- Keselman, H.J. and Wilcox, R.R. (1999). The Improved Brown and Forsythe Test for Mean Equality: Some things can't be fixed, *Communications in Statistics-Simulation* **28**, 687-698.
- Kruskal, William H. (1952a). A nonparametric test for the several sample problem, *The Annals of Mathematical Statistics* **23**, 525-540.
- Kruskal, William H. and Wallis, W. Allen (1952b). Use of Ranks in One-Criterion Variance Analysis, *Journal of the American Statistical Association* **47**, 583-621.
- Krzanowski, Wojtek J. (1998). *An introduction to statistical modelling*. London: Arnold.
- Kulinskaya, E. and Dollinger, M. B. (2007). Robust weighted one-way ANOVA: Improved approximation and efficiency, *Journal of Statistical Planning and Inference* **137**, 462-472.
- Lindgren, Bernard W. (1973). *Statistical theory*. New York: MacMillan.
- Lindman, H. (1992). *Analysis of variance in experimental design*. New York: Springer-Verlag.

- Mehrotra, O.V. (1997). Improving the Brown-Forsythe Solution to the Generalized Behrens-Fisher Problem, *Communications in Statistics-Simulation and Computation* **26**, 1139-1145.
- Miranda, Hilário A. S. R. (2009). *Métodos Robustos em Geoestatística*. Departamento de Matemática: Universidade de Aveiro.
- Montgomery, D. C. and Runger, G. C. (1994). *Applied statistics and probability for engineers*. New York: John Wiley and Sons Inc..
- Murteira, B. and Ribeiro, C.S. and Silva, J.A. and Pimenta, C. (2007). *Introdução à Estatística*, 2^a ed., Lisboa: McGraw-Hill.
- Neter, J. and Kutner, M. H. and Wasserman, W. and Nachtsheim, C. (1996). *Applied Linear Statistical models*. Chicago: Irwin.
- Pires, A. M. e Branco, J. A. (2007). *Introdução aos Métodos Estatísticos Robustos*. Lisboa: SPE.
- Pontes, D. N. (2009). *Testes t e F: Robustos ou não, eis a questão!*. Instituto Superior Técnico: Universidade Técnica de Lisboa.
- Reed, J. F. and Stark, D. B. (1995). Robust analysis of variance:simulation study, *Journal of Applied Statistics* **22**, 87-104.
- Rogan, J.C. and Keselman, H.J. (1977). Is the ANOVA F-Test Robust to Variance Heterogeneity When Sample Sizes are Equal?: An Investigation via a Coefficient of Variation, *American Educational Research Journal* **14**, 493-498.
- Ross, S. M. (1987). *Introduction to probability and statistics for engineers and scientists*. New York: John Wiley and Sons Inc..
- Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators, Robust and Nonlinear Time Series Analysis, *Lecture Notes in Statistics*, Springer **26**, 256-272.

Scheffé, H. (1959). *The analysis of variance*. New York: John Wiley and Sons Inc..

Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M. (2010). Is It Really Robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption, *European Journal of Research Methods for the Behavioral and Social Sciences* **6**, 147-151.

Spiegel, M. R. (1977). *Probabilidades e Estatística traduzido por Alfredo Alves de Faria, resumo de teoria, 760 problemas resolvidos*. Lisboa: McGraw-Hill.

Todorov, V. and Filzmoser, P. (2010). Robust Statistic for the one-way MANOVA, *Computational Statistics and Data Analysis* **54**, 37-48.

Wilcox, R.R. (1995). The Practical Importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies, *British J. Math. Statist. Psych* **48**, 99-114.

Wilcox, R.R. (1996). *Statistics for the Social Sciences*. New York: Academic Press.

Apêndice A

Apêndice

Demonstração do Teorema de Cochran

Demonstração A.0.1 (Cochran). Por hipótese, os U'_i s são variáveis aleatórias independentes, normais estandardizadas e tais que:

$$\sum_{i=1}^n U_i^2 = Q_1 + Q_2 + \dots + Q_k,$$

onde cada Q_i representa uma soma de quadrados de combinações lineares dos U' s.

Seja r_i a característica da respectiva forma quadrática de Q_i . Cada Q_i pode ser expresso numa soma de exactamente r_i quadrados das combinações lineares dos U' s, ou seja:

$$\left\{ \begin{array}{l} Q_1 = W_1^2 + W_2^2 + \dots + W_{r_1}^2 \\ Q_2 = W_{r_1+1}^2 + \dots + W_{r_1+r_2}^2 \\ \vdots \\ Q_k = W_{r_1+r_2+\dots+r_{k-1}+1}^2 + \dots + W_{r_1+r_2+\dots+r_k}^2 \end{array} \right.$$

onde $\mathbf{W} = \mathbf{C}\mathbf{U}$, sendo assim existem exactamente $r_1 + r_2 + \dots + r_k$ combinações lineares dos U' s.

Como $\mathbf{U} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$ e

$$\begin{aligned} \mathbf{U}^T \mathbf{U} &= \sum_{i=1}^n U_i^2 = Q_1 + Q_2 + \dots + Q_k \\ &= W_1^2 + W_2^2 + \dots + W_{r_1}^2 + W_{r_1+1}^2 + \dots + W_{r_1+r_2}^2 + \dots + W_{r_1+r_2+\dots+r_k}^2 \\ &= \mathbf{W}^T \mathbf{W}. \end{aligned}$$

Então $\mathbf{U}^T \mathbf{U} = \mathbf{W}^T \mathbf{W} = (\mathbf{C}\mathbf{U})^T \mathbf{C}\mathbf{U} = \mathbf{U}^T \mathbf{C}^T \mathbf{U}\mathbf{C}$, o que implica que $\mathbf{C}_T \mathbf{C} = \mathbf{I}_n$. Como \mathbf{C} é quadrada, $\mathbf{C}^T = \mathbf{C}^{-1}$, logo $\mathbf{C}\mathbf{C}^T = \mathbf{I}_n$.

Sendo assim, $\mathbf{W} \sim \mathcal{N}_n(\mathbf{C}\mathbf{0}, \mathbf{C}\mathbf{I}_n\mathbf{C}^T)$, ou seja \mathbf{W} segue uma distribuição Normal multivariada, com um vector de médias $(0, 0, \dots, 0)$ e matriz de covariâncias \mathbf{I}_n , $\mathbf{W} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$.

Pode-se concluir que as variáveis Q'_i s são independentes e cada uma delas segue uma distribuição de um qui-quadrado com r_i graus de liberdade, para $i = 1, 2, \dots, n$.

□

Apêndice B

Macros em R

Este apêndice contém algumas funções implementadas para o cálculo das estatísticas de teste referidas no Capítulo 5. Os programas apresentados não estão comentados, mas têm uma estrutura simples, pelo que se assume que são de fácil leitura para um leitor familiarizado com o *software* estatístico R.

B.1 Testes de Brown-Forsythe e de Welch (modificados)

```
testeBFModif<-function(d)
{
  library(robustbase)
  library(MASS)

  lista<-list(d[[2]])
  md<-tapply(d[[1]],lista,list)
  md<-as.vector(md)
  dimAmostras<-tapply(d[[1]],lista, length)
  dimAmostras<-as.vector(dimAmostras)
  vector.medias<-tapply(d[[1]],lista, mean)
```

```

vector.medias<-as.vector(vector.medias)

media<-mean(d[[1]]);
k<-1;
nTotal<-sum(dimAmostras);
numerador<-0;
denominador<-0;
numeradorCk=NULL;
s22=NULL;
aux3=NULL;
aux4=NULL;
aux5=NULL;
while(k<=length(dimAmostras))
{
numerador<-numerador+dimAmostras[k]*(vector.medias[k]-media)^2;
j<-1;
s2<-0;
while(j<=dimAmostras[k])
{
s2<-s2+(md[[k]][j]-vector.medias[k])^2/(dimAmostras[k]-1)
j<-j+1;
}
aux<-(1-dimAmostras[k]/nTotal)*s2;
numeradorCk<-append(numeradorCk,aux);
denominador<-denominador+aux;
s22<-sum(append(s22,s2));
aux3<-sum(append(aux3,dimAmostras[k]*s2));
aux4<-sum(append(aux4,dimAmostras[k]*s2^2));
aux5<-sum(append(aux5,s2^2));
k<-k+1;
}
valorEstadistica<-numerador/denominador;

f1<-((s22-(aux3/nTotal))^2)/(aux5+((aux3/nTotal)^2)-2*(aux4/nTotal));
r<-1;

```

```

f2<-0;
while(r<=length(dimAmostras))
{
ck<-numeradorCk[r]/denominador;
f2<-f2+ck^2/(dimAmostras[r]-1)
r<-r+1;
}
valorP<-1-pf(valorEstatistica,f1,f2^(-1));
valorP
}

#definir o data.frame dos dados "\emph{winsorizados}", com alpha=0.1, para a
# estatística de Teste de Welch
library(robustbase)
library(MASS)
library(asbio)
lista<-list(dados.pef[[2]])
md<-tapply(dados.pef[[1]],lista,list)
md<-as.vector(md)
c1<-win(md[[1]],lambda=0.1)
c2<-win(md[[2]],lambda=0.1)
c3<-win(md[[3]],lambda=0.1)
c4<-win(md[[4]],lambda=0.1)
c5<-win(md[[5]],lambda=0.1)
c6<-win(md[[6]],lambda=0.1)
c7<-win(md[[7]],lambda=0.1)
c8<-win(md[[8]],lambda=0.1)
winIMC<-c(c1,c2,c3,c4,c5,c6,c7,c8)
ef<-c("FSS","FSN","FNS","FNN","MSS","MSN","MNS","MNN")
winpef<-factor(c(rep(ef[1],49),rep(ef[2],65),rep(ef[3],20),rep(ef[4],64),
rep(ef[5],14),rep(ef[6],22),rep(ef[7],11),rep(ef[8],91)))
windados.pef<-data.frame(winIMC,winpef)

testeWelchModif<-function(d,wind,alpha)
{
lista<-list(d[[2]])

```

```

listaw<-list(wind[[2]])
winmd<-tapply(wind[[1]],listaw,list)
winmd<-as.vector(winmd)
dimAmostrasw<-tapply(wind[[1]],listaw, length)
dimAmostrasw<-as.vector(dimAmostrasw)
vector.medias.ap<-tapply(d[[1]],lista, mean,trim=alpha*2)
vector.medias.ap<-as.vector(vector.medias.ap)
vector.medias.win<-tapply(wind[[1]],listaw,mean)
vector.medias.win<-as.vector(vector.medias.win)

t<-1
j.i=NULL
h.i=NULL
while(t<=length(dimAmostrasw))
{
j<-as.integer(dimAmostrasw[t]*alpha)
h.i<-append(h.i,dimAmostrasw[t]-2*j)
j.i<-append(j.i,j)
t<-t+1
}

k<-1;
denon<-0;
num.mediaTotal=NULL;
den.mediaTotal=NULL;
w=NULL;
numerador<-0;

while(k<=length(dimAmostrasw))
{
j<-1;
s2w<-0;
while(j<=dimAmostrasw[k])
{
s2w<-s2w+(winmd[[k]][j]-vector.medias.win[k])^2/(dimAmostrasw[k]-1);

```



```

nume.mediaTotal<-(h.i[k]/s2w)*vector.medias.ap[k];
deno.mediaTotal<-(h.i[k]/s2w);
j<-j+1;
}
w<-append(w,deno.mediaTotal);
num.mediaTotal<-sum(append(num.mediaTotal,nume.mediaTotal));
den.mediaTotal<-sum(append(den.mediaTotal,deno.mediaTotal));
k<-k+1;
}
mediaTotal<-num.mediaTotal/den.mediaTotal;
u<-sum(w);
t<-1;
while(t<=length(dimAmostrasw))
{
aux<-((1-(w[t]/u))^2)/(h.i[t]-1);
denon<-append(denon,aux);
numerador<-numerador+w[t]*(vector.medias.ap[t]-mediaTotal)^2;
t<-t+1;
}
numerador<-sum(numerador)
denominador<-sum(denon)
f<-(numerador/(length(dimAmostrasw)-1))/((1+((2*(length(dimAmostrasw)-2)/
((length(dimAmostrasw))^2-1)))*denominador));

fa<-(3/((length(dimAmostrasw))^2-1))*denominador;
valorP<-1-pf(f,length(dimAmostrasw)-1,fa^(-1));
valorP
f
}

```

B.2 Teste de $FHubS$

```

library(robustbase)
library(MASS)

```

```

estatistica.teste.FHubS<-function(d)
{
  lista<-list(d[[2]])
  md<-tapply(d[[1]],lista,list)
  md<-as.vector(md)
  n<-tapply(d[[1]],lista, length)
  n<-as.vector(n)
  i<-1
  mu.g=NULL

  for(i in 1:length(n))
  {
    mu.g<-append(mu.g,hubers(md[[i]])[[1]])
    i=i+1
  }

  mu.g
  sigma2.g<-(s_Sn(mu.g))
  sigma2.g
  sigma2.erro<-s_Sn(d[[1]])
  Fpr<-(sigma2.g/sigma2.erro)^2
  Fpr
}

bootstrap.prop<-function(dados)
{
  d<-subset(dados,(quantile(dados[[1]],0.05)<dados[[1]])&
(dados[[1]]<quantile(dados[[1]],0.95)))
  t<-1
  mat.ind<-c(1:length(d[[1]]))
  for(t in 1:1100)
  {
    indices<-sample(1:nrow(d),length(d[[1]]), replace = TRUE)
    n.data<-d[indices,]
  }
}

```

```

lista<-list(n.data[[2]])
md<-tapply(n.data[[1]],lista,md)
md<-as.vector(md)

if((md[[1]]!=0)& (md[[2]]!=0)& (md[[3]]!=0)&(md[[4]]!=0)&
(md[[5]]!=0)&(md[[6]]!=0)&(md[[7]]!=0)&(md[[8]]!=0))
{
  mat.ind<-cbind(mat.ind,indices)
  t<-t+1
}

else
{
  t<-t+1
}
}

mat.ind
vector.est=NULL
i<-1
for(i in 1:1000)
{
  n.data<-d[mat.ind[,i],]
  vector.est<-append(vector.est,estadistica.teste.FHubS(n.data))
  i<-i+1
}

vector.est
}

```


Apêndice C

Anexos

A tabela que se segue contém as observações de Índice de Massa Corporal (Kg/m^2) nos tratamentos onde foram inseridos os indivíduos.

Prática de E.F.	IMC	Prática de E.F.	IMC	Prática de E.F.	IMC
FNN	21.36	FSS	18.73	FSS	24.74
FNS	21.23	FSS	24.17	FSS	21.3
FNS	23.15	FNN	15.06	FSS	21.85
FSN	24.77	FSN	16.71	FSN	19.43
FNS	21.88	FNN	18.82	FSS	20.2
FSS	19.69	FSS	21.09	FNS	20.03
FSS	16.65	FNS	19.14	FNS	17.01
FNS	21.45	FSS	19.38	FSN	19.14
FSS	22.31	FSS	21.1	FSN	20.2
FNN	22.1	FSS	25.71	FNS	20.2
FNS	20.7	FNN	20.96	FNN	22.27
FNS	22.86	FNS	18	FNN	18.37
FSS	22.03	FSS	31.25	FNS	17.22
FSS	19.95	FNS	19.49	FNS	18.5
FNS	21.19	FNS	20.82	FSS	18.29
FSS	21.04	FNN	24.38	FNS	24.52
FSN	17.1	FSS	24.56	FSS	18.26

Prática de E.F.	IMC	Prática de E.F.	IMC	Prática de E.F.	IMC
FNN	22.72	FNS	20.4	FSS	21.1
FNN	19.71	FNS	21.48	FSS	22.22
FNS	25.35	FNS	19.53	FNS	20.08
FNS	25.39	FNN	22.77	FNS	19.29
FSS	21.93	FNS	16.96	FSS	26.42
FSS	17.65	FNN	19.38	FSN	22.49
FSS	16.85	FNN	18.82	FNS	16.33
FNN	17.21	FNN	19.13	FNS	17.3
FSS	18.42	FSS	16.59	FSN	18.52
FNS	17.53	FNN	22.5	FSS	31.74
FNS	24.16	FNN	22.86	FNS	20.7
FNS	20.06	FNS	18.37	FSN	21.08
FNN	20.31	FSS	19.72	FSS	18.29
FSS	20.2	FNS	17.31	FSS	21.99
FSS	22.96	FNS	20.82	FSS	19.16
FSS	24.46	FNN	24.24	FNS	22.19
FNS	22.58	FNN	18.59	FSS	19.78
FNN	19.29	FNN	23.42	FNN	25.81
FSS	21.67	FNS	17.26	FNN	19.1
FSN	23.03	FNN	24.44	FSN	25.71
FSS	31.11	FSS	18.59	FNS	19.23
FSN	18.13	FNS	18.17	FNN	19.03
FSS	21.48	FSS	21.23	FSN	21.67
FNS	19.2	FNS	20.57	FNN	19.44
FSS	20.76	FNS	18.59	FSS	17.96
FSS	19.88	FNS	20.57	FNS	19.81
FNN	20.24	FSS	24.77	FNS	23.66
FSS	17.69	FNN	23.71	FSS	21.45
FSN	19.38	FNN	24.46	FSS	16.67
FNN	20.66	FNS	21.56	FNN	27.69
FNS	21.36	FNN	19.83	FNS	20.31
FNS	21.71	FNN	18.59	FNS	17.44

Prática de E.F.	IMC	Prática de E.F.	IMC	Prática de E.F.	IMC
FSN	18.67	FSN	28.93	FNS	22.21
FSS	22.21	FNN	25.95	FNS	22.32
FNS	19.78	FSS	20.57	FSS	18.37
FSN	17.45	FNN	20.06	FNN	18.66
FSS	21.99	FSS	19.49	FNN	22.77
FNS	24.09	FNS	16.65	FNN	15.24
FSS	22.04	FNS	18.78	FNS	20.2
FSN	26.56	FSS	18.07	FSS	20.9
FNN	20.48	FSS	20.7	FSS	19.33
FNS	20.2	FSS	25.91	FSN	19.57
FNS	24.61	FNS	26.35	FSN	23.03
FNN	18.56	FNS	16.8	FSS	24.22
FNS	18.29	FNN	19.49	FNN	21.26
FNN	22.49	FNS	19.81	FSS	20.81
FNN	24.98	FNS	23.12	FNN	24.61
FNN	24.34	FNS	21.22	FSS	20.7
FSS	20.69	FNN	22.31	FNN	22.35
MSS	18.17	MNS	18.34	MSS	18.93
MSN	23.11	MSN	23.23	MSN	20.8
MSS	23.66	MSS	19.28	MSS	20.76
MSS	24.69	MSS	26.64	MSS	18.83
MSS	21.05	MSS	18.67	MNS	22.59
MNS	22.76	MSS	17.17	MSS	20.38
MSS	17.37	MSS	19.6	MSS	20.94
MSS	20.76	MNS	19.72	MSS	19.92
MNS	18.38	MSS	22.14	MNS	26.4
MSS	22.34	MSS	21.6	MSS	19.61
MSS	24.22	MSS	20.88	MSS	21.02
MSS	22.68	MNN	21.39	MSS	22.09
MSN	23.53	MSS	22.49	MSS	22.04
MNS	20.15	MNS	20.15	MSS	19.59
MSS	21.36	MNN	25.95	MSS	21.83

Prática de E.F.	IMC	Prática de E.F.	IMC	Prática de E.F.	IMC
MSS	24.34	MSS	26.12	MSN	26.12
MNN	19.69	MSS	17.93	MSS	25.83
MSS	20.45	MSS	23.01	MSS	20.4
MSS	21.38	MSN	22.23	MSS	21.62
MNS	20.68	MNS	24.45	MSS	23.6
MSS	20.96	MNN	20.75	MSS	20.52
MSS	22.06	MSS	20.11	MSS	23.83
MSS	20.45	MSS	24.72	MSS	22.79
MNS	18.94	MNN	22.78	MNN	26.87
MSS	21.6	MSN	19.1	MNS	19.72
MSS	25.06	MNS	22.06	MSS	26.23
MSN	20	MNN	20.7	MSS	18.31
MSS	24.82	MSS	23.46	MSS	22.15
MNN	17.76	MNS	30.88	MSS	26.33
MNS	29.04	MSN	27.68	MSS	23.94
MNN	24.97	MSS	27.55	MSS	24.15
MSS	24.15	MNS	26.87	MSS	28.09
MSS	17.01	MNS	22.66	MNN	16.71
MNN	31.02	MSS	23.15	MSS	17.92
MSS	21.86	MSS	22.15	MSS	24.21
MSS	22.16	MSS	19.96	MSS	20.68
MNN	17.1	MSS	27.04	MNS	25.25
MSS	23.39	MSS	20.94	MSS	18.52
MSS	26.67	MSN	20.52	MSS	20.29
MSS	20.37	MSS	20.57	MNS	21.22
MSS	23.74	MSN	17.64	MSS	21.04
MNS	23.15	MSS	20.66	MNN	19.83
MSS	21.01	MNS	26.99	MSS	32.66
MSS	25.26	MNN	27.04	MSS	24.8
MSS	23.72	MSS	27.4	MSS	22.64
MNS	22.78	MSS	19.57	MSS	20.2

Tabela C.4: Conjunto dos dados reais.