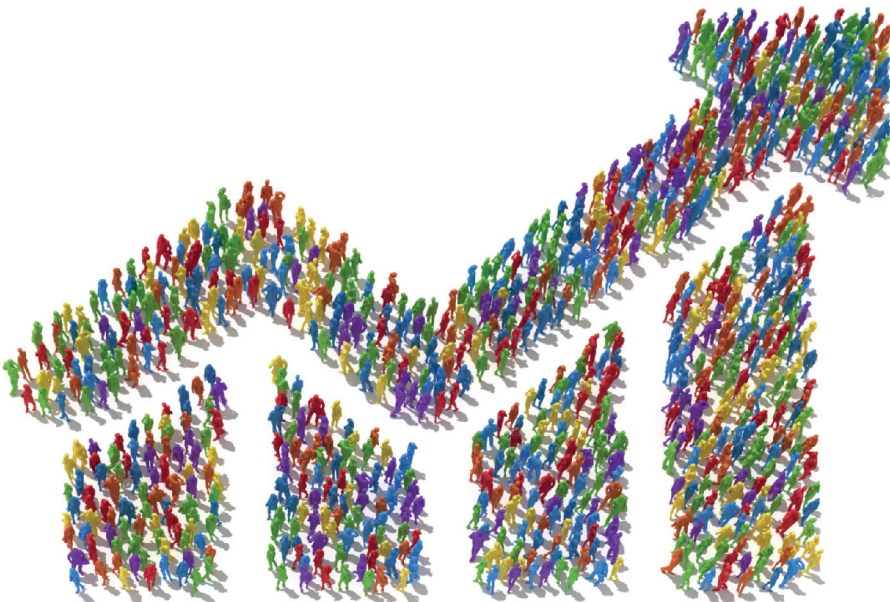


RAUL M. S. LAUREANO

Testes de Hipóteses e Regressão

O meu manual
de consulta rápida

Com **SPSS Statistics** e 



EDIÇÕES SÍLABO

Outros livros da coleção *O meu Manual de Consulta Rápida*

- IBM SPSS Statistics (3ª edição)
- Testes de Hipóteses com o IBM SPSS Statistics (3ª edição)
- Modelos de Regressão Linear (a publicar)

Apoios:



A PSE é uma empresa especializada em *Data Science* e Pesquisa Avançada que está no mercado desde 1994. A sua atividade materializa-se na implementação de soluções tecnológicas, na pesquisa de mercado avançada e na

prestação de serviços de consultoria e *data science*.

No âmbito da sua atividade promove e dinamiza a divulgação das técnicas estatísticas e das melhores práticas de execução de um projeto analítico.

Por esse motivo tivemos a honra em nos associar a esta publicação que permitirá aos seus leitores dispor de um manual de consulta rápida sobre procedimentos de testes de hipóteses e de métodos de regressão no desenvolvimento dos seus projetos de análise de dados.



TESTES DE HIPÓTESES E REGRESSÃO

O meu
MANUAL DE CONSULTA RÁPIDA

RAUL M. S. LAUREANO



EDIÇÕES SÍLABO

É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio gráfico, eletrónico ou mecânico, inclusive fotocópia, este livro.

As transgressões serão passíveis das penalizações previstas na legislação em vigor.

Não participe ou encoraje a pirataria eletrónica de materiais protegidos.

O seu apoio aos direitos dos autores será apreciado.

Visite a Sílabo na rede

www.silabo.pt

Esta obra teve revisão científica e pedagógica dos professores do ISCTE-IUL
Maria do Carmo Botelho e Luís Laureano.

FICHA TÉCNICA:

Título: Testes de Hipóteses e Regressão – O Meu Manual de Consulta Rápida

Autor: Raul M. S. Laureano

© Edições Sílabo, Lda.

Capa: Pedro Mota

Imagem da capa: Elcabron | Dreamstime.com

1ª Edição – Lisboa, outubro de 2020.

Impressão e acabamentos: Europress, Lda.

Depósito Legal: 474840/20

ISBN: 978-989-561-051-8



EDIÇÕES SÍLABO, Lda.

Publicamos conhecimento

Editor: Manuel Robalo

R. Cidade de Manchester, 2

1170-100 Lisboa

Tel.: 218130345

e-mail: silabo@silabo.pt

www.silabo.pt

Índice

Introdução	11
Conceito de teste de hipóteses	13
Seleção de um teste estatístico	17

Capítulo 1

Testes de hipóteses paramétricos

Teste t para uma amostra	23
Problema 1 (média)	23
Procedimento no SPSS <i>Analyse</i> <i>Compare Means</i>	
<i>One-Sample T Test...</i>	23
Procedimento no R	26
Procedimento num relatório	27
Problema 2 (proporção – variável binária)	31
Procedimento num relatório	31
Teste t para duas amostras independentes	34
Problema	34
Procedimento no SPSS <i>Analyse</i> <i>Compare Means</i>	
<i>Independent-Samples T Test...</i>	35
Procedimento no R	36
Procedimento num relatório	38
Teste t para duas amostras emparelhadas	42
Problema	42
Procedimento no SPSS <i>Analyse</i> <i>Compare Means</i>	
<i>Paired-Samples T Test...</i>	43
Procedimento no R	45
Procedimento num relatório	46

Análise de Variância (ANOVA) a um fator	50
Problema 1	50
Procedimento no SPSS <i>Analyse Compare Means One-Way ANOVA...</i>	51
Procedimento no R	55
Procedimento num relatório	62
Problema 2	68
Procedimento num relatório	68
Análise de Variância (ANOVA) a dois fatores (efeitos fixos)	70
Problema 1 (sem efeito de interação)	70
Procedimento no SPSS <i>Analyse General Linear Model Univariate...</i>	71
Procedimento no R	79
Procedimento num relatório	83
Problema 2 (com efeito de interação)	93
Procedimento num relatório	94
Teste de <i>Levene</i> à homogeneidade de variâncias	104
Problema 1	104
Procedimento no SPSS <i>Analyse Descriptive Statistics Explore...</i>	105
Procedimento no R	107
Procedimento num relatório	107
Problema 2	111
Procedimento num relatório	111

Capítulo 2

Testes de hipóteses não-paramétricos

Testes de aderência de <i>Shapiro-Wilk</i> e <i>Kolmogorov-Smirnov</i> à distribuição normal	117
Problema	117
Procedimento no SPSS <i>Analyse Descriptive Statistics Explore...</i>	117
Procedimento no R	121
Procedimento num relatório	121
Teste de aderência de <i>Kolmogorov-Smirnov</i>	124
Problema	124
Procedimento no SPSS <i>Analyse Nonparametric Tests Legacy Dialogs 1-Sample K-S...</i>	125

Procedimento no R	128
Procedimento num relatório	128
Teste de aderência do <i>Qui-quadrado</i>	130
Problema 1	130
Procedimento no SPSS <i>Analyse Nonparametric Tests </i> <i> Legacy Dialogs Chi-Square...</i>	131
Procedimento no R	133
Procedimento num relatório	134
Problema 2	137
Procedimento num relatório	137
Teste de independência do <i>Qui-quadrado</i>	139
Problema	139
Procedimento no SPSS <i>Analyse Descriptive Statistics Crosstabs...</i>	140
Procedimento no R	143
Procedimento num relatório	145
Teste à medida de associação <i>V de Cramer</i>	149
Problema	149
Procedimento no R	150
Procedimento num relatório	151
Testes aos coeficientes de correlação de <i>Pearson</i> e de <i>Spearman</i>	153
Problema 1	154
Procedimento no SPSS <i>Analyse Correlate Bivariate...</i>	154
Procedimento no R	156
Procedimento num relatório	158
Problema 2	161
Procedimento num relatório	161
Problema 3	163
Procedimento num relatório	163
Teste de <i>Mann-Whitney U</i> para duas amostras independentes	165
Problema	165
Procedimento no SPSS <i>Analyse Nonparametric Tests </i> <i> Legacy Dialogs 2 Independent Samples...</i>	166
Procedimento no R	169
Procedimento num relatório	169

Teste de <i>Wilcoxon</i> para duas amostras emparelhadas	173
Problema	173
Procedimento no SPSS <i>Analyse Nonparametric Tests </i> <i> Legacy Dialogs 2 Related Samples...</i>	174
Procedimento no R	177
Procedimento num relatório	177
Teste de <i>Kruskal-Wallis H</i> para k amostras independentes	181
Problema 1	182
Procedimento no SPSS <i>Analyse Nonparametric Tests </i> <i> Legacy Dialogs K Independent Samples...</i>	183
Procedimento no R	185
Procedimento num relatório	186
Problema 2	195
Procedimento num relatório	195

Capítulo 3

Regressão linear

Regressão linear simples	201
Problema	201
Procedimento no SPSS <i>Analyse Regression Linear...</i>	202
Procedimento no R	211
Procedimento num relatório	218
Regressão linear múltipla	232
Problema	232
Procedimento num relatório	233

Capítulo 4

Regressão logística simples

Problema	253
Procedimento no SPSS <i>Analyse Regression Binary Logistic...</i>	255
Procedimento no R	262
Procedimento num relatório	268

Apêndice 1

Cálculo da dimensão da amostra

Estimação de uma proporção	281
Relação entre dimensão da amostra e erro máximo admissível	283
Cálculo no Excel	284
Exemplos de cálculo	284
Estimação de uma média	286

Apêndice 2

Estatística descritiva

Seleção das técnicas adequadas	291
Estatística descritiva univariada	291
Estatística descritiva bivariada	292
Interpretação de indicadores de estatística descritiva: um exemplo	293
Análise univariada: tabela de frequências	293
Análise univariada: tabela de medidas de estatística descritiva	294
Análise bivariada	296

Apêndice 3

R com o R Studio: leitura de dados, tabelas e gráficos

Instalação do R (com o R Studio)	301
Leitura dos dados do Excel	303
Criação de tabelas	304
Tabela de frequências	304
Tabela de contingência	305
Medidas descritivas	306
Tabela de comparação de medidas descritivas	307
Criação de gráficos	309
Gráfico de circular (<i>pie</i>)	309
Gráfico de barras	310
Histograma	310

Diagrama de extremos e quartis (<i>boxplot</i>)	311
Gráfico de barras empilhadas a 100%	312
Gráfico de linhas com médias (perfil de médias)	312
Gráfico de barras com médias	313
Gráfico com médias	314
Diagrama de extremos e quartis (<i>boxplot</i>)	315
Diagrama de dispersão	316
Gráfico de correlações	316

Anexo

Questionário – Avaliação de ações de formação profissional	323
---	-----

Bibliografia	331
---------------------	-----

Introdução

O livro *Testes de Hipóteses e Regressão: o meu manual de consulta rápida* surge da necessidade de satisfazer a diversas solicitações de estudantes de diversas universidades. De facto, este livro vem complementar o meu outro livro *Testes de Hipóteses com o SPSS Statistics: o meu manual de consulta rápida*, ao acrescentar outras técnicas estatísticas muito utilizadas e integrantes de diversos programas de unidades curriculares de diversos cursos universitários e até de cursos de cariz profissional.

De facto, neste novo livro, para além dos principais testes de hipóteses, paramétricos e não paramétricos, abordam-se dois modelos de regressão. Por um lado, o modelo de regressão linear, quer a simples, com apenas uma variável independente, quer a múltipla, incluindo mais do que uma variável independente. Por outro, o modelo de regressão logística, adequado quando a variável dependente é qualitativa binária (e não quantitativa, como acontece na regressão linear).

Mas, as solicitações dos estudantes não remetem apenas para a inclusão de mais técnicas estatísticas, mas também para a inclusão de ferramentas de análise de dados que sejam gratuitas. É assim que surge neste livro, para além da utilização do *IBM SPSS Statistics*, a realização das diversas técnicas estatísticas com o *R*. São duas ferramentas muito poderosas de análise de dados e amplamente divulgadas e utilizadas, que, espero, possam satisfazer a maioria das necessidades dos leitores, no que se refere a técnicas de inferência estatística e de regressão.

Contudo, este livro mantém o foco na realização das análises com recurso a ferramentas de análise de dados e, principalmente, na análise dos resultados. Ou seja, não apresenta as, por vezes, complexas expressões matemáticas para o cálculo de diversos indicadores estatísticos, mas ensina a obter esses indicadores com o *SPSS Statistics* e com o *R* e a interpretar os diferentes indicadores obtidos.

De forma simples, este livro apresenta diferentes problemas que, para a sua resolução, obrigam a recorrer a testes de hipóteses ou a modelos de regressão. Todos os problemas apresentados têm por base os dados recolhidos através de um

inquérito por questionário,¹ e as análises foram realizados na versão 26 do *IBM SPSS Statistics* (mas todas elas compatíveis com versões anteriores) e, também, no *R* (versão 3.5.2). Para cada problema, apresenta-se uma justificação para a escolha do procedimento estatístico, ilustra-se como o procedimento é realizado no *SPSS Statistics* e no *R* e como os resultados devem ser apresentados e interpretados num relatório de análise de dados.

Assim, e em linha com os meus outros livros² também neste o mais importante é o *Como* fazer uma análise de dados. Deste modo, este livro destina-se, essencialmente, a estudantes do ensino superior, por exemplo, para os ajudar na realização das suas teses, e a profissionais que necessitem de elaborar relatórios de análise de dados.

Por fim, não posso deixar de agradecer a todos os que de alguma forma contribuíram para esta obra, estudantes de licenciatura, mestrado e, até, de doutoramento, colegas, amigos e familiares. E, ao Iscte – Instituto Universitário de Lisboa (ISCTE-IUL) onde exerço a minha atividade profissional desde o início dos anos 1990, como docente (da *Business School*) e como investigador (do grupo de *Data Analytics* da UNIDE e do grupo de Sistemas de Informação do ISTAR).

O Autor

Fevereiro de 2020

P.S.: não existe obra perfeita, nem tão pouco obra completa. Assim, agradecemos todos os contributos que possam melhorar este livro, quer introduzindo novos temas, quer melhorando a forma como são apresentados. O meu endereço de *e-mail* é raul.laureano@iscte-iul.pt.

(1) Ver Anexo: Questionário – Avaliação de ações de formação profissional. A tabela de dados encontra-se disponível na página do livro em www.silabo.pt.

(2) Não posso deixar de recomendar, para uma melhor utilização do *SPSS Statistics*, nomeadamente, para a transformação de variáveis e para a elaboração de tabelas e gráficos, isto é, para a realização de análises estatísticas descritivas e exploratórias, o livro *IBM SPSS Statistics: o meu manual de consulta rápida*. E para complementar a análise de regressão recomendo o livro *Modelos de Regressão Linear: o meu manual de consulta rápida*, que contempla modelos de regressão linear com moderação e com mediação, para além de apresentar os fundamentos da regressão linear e um conjunto alargado de exercícios resolvidos.

Conceito de teste de hipóteses

A estatística indutiva ou inferencial (ou inferência estatística) permite, com base nos elementos observados (amostra aleatória) e descritos, retirar conclusões para um domínio mais vasto de onde esses elementos provieram (população ou universo). Assim, este ramo da estatística permite obter generalizações aplicáveis a indivíduos que não tenham sido observados.

A estatística indutiva tem por base a teoria das probabilidades e a teoria da amostragem e divide-se em dois ramos: estimação e testes de hipóteses.

Com as técnicas de estimação visa-se obter estimativas pontuais, quando se usa um único valor para estimar o parâmetro populacional, e estimativas por intervalos, quando se indica um intervalo de valores dentro do qual o parâmetro populacional pode estar localizado com determinado nível de confiança. O nível de confiança (λ) corresponde à probabilidade de o intervalo de confiança conter o valor do parâmetro, sendo os mais usuais, expressos em percentagem, o 90%, 95% e 99%.

Os testes de hipóteses visam testar suposições que são efetuadas sobre a população. Pedrosa e Gama (2004, p. 442) definem um teste de hipóteses como sendo o «processo estatístico usado para se tirar uma conclusão do tipo sim ou não sobre uma ou mais populações, a partir de uma ou mais amostras dessas populações».

Desta forma, os testes de hipóteses visam testar se certas hipóteses (estatísticas) formuladas sobre, por exemplo, os parâmetros da população ou sobre as suas distribuições são ou não rejeitadas. A formulação das hipóteses deverá ser baseada na teoria e nunca tendo em conta o observado na amostra.

As hipóteses (afirmações), competitivas, mutuamente exclusivas e coletivamente exaustivas (Doane e Seward, 2008, p. 350), formuladas sobre a população são sempre duas porque a afirmação verdadeira não é conhecida.

Assim, tem-se, por um lado, a hipótese nula (H_0), que corresponde à afirmação mais restritiva e que é considerada verdadeira até prova em contrário evidenciada pela amostra (isto é, até que se verifiquem evidências estatísticas que apontem em sentido contrário de H_0). A hipótese nula contém sempre uma igualdade (=) e é

aquela que se julga inverosímil.¹ Esta hipótese reflete, portanto, a não existência de diferenças entre o observado na amostra e o que se está a afirmar sobre a população.

Por outro, tem-se a hipótese alternativa (H_a ou H_1), que corresponde à afirmação contrária à hipótese nula e que se julga verosímil. Esta hipótese contém sempre uma desigualdade ($>$ ou $<$) ou a negação da igualdade (\neq). No caso desta hipótese conter uma desigualdade diz-se que o teste é unilateral (*one-tailed test*), à direita (*right tailed*), se o sinal for de maior ($>$), ou à esquerda (*left tailed*), se o sinal for de menor ($<$), e no caso de conter a negação da igualdade designa-se o teste por bilateral (*two-tailed test*).

Com o procedimento pretende-se verificar se existem evidências estatísticas que levem à rejeição da hipótese nula (H_0), aceitando-se, neste caso, que a hipótese alternativa (H_a) é verdadeira. Não se rejeitando a hipótese nula, não é possível provar que esta é verdadeira, mas deve assumir-se que sim porque não é possível rejeitá-la.

Os testes de hipóteses são usados para verificar se as diferenças entre os valores da amostra e os valores da população em teste são devidos ao acaso. Desta forma, quando o observado na amostra estiver próximo do afirmado sobre a população (em H_0), não se rejeita a hipótese nula, e quando estiver muito afastado, rejeita-se a hipótese nula.

A decisão de rejeitar ou não a hipótese nula (H_0) depende do valor do teste estatístico, que mede esse afastamento em termos do seu erro padrão:

- Se o afastamento é muito pequeno, então, é pouco provável que H_0 seja falsa, logo não se rejeita H_0 .
- Se o afastamento é grande, é pouco provável que H_0 seja verdadeira, logo rejeita-se H_0 e aceita-se H_a (se o afastamento vai no «sentido» de H_a).

Quer isto dizer que não rejeitar H_0 significa que a amostra não contém evidência suficiente para se poder rejeitar a hipótese nula, isto é, o teste dir-se-á, como sugerem Guimarães e Cabral (2007, p. 240), inconclusivo e, por isso, não se pode rejeitar H_0 . Por outro lado, rejeitar H_0 significa que a probabilidade de obter uma amostra que evidencie que H_0 é verdadeira é muito reduzida e, consequentemente, pode-se aceitar a hipótese alternativa (H_a).

(1) A hipótese nula tem sempre a igualdade pois a estatística de teste a utilizar na decisão tem uma determinada função de distribuição de probabilidades quando se verifica essa igualdade (pressupondo H_0 verdadeira). De facto, «o teste é efetuado considerando apenas a situação em que H_0 mais se aproxima de H_1 , ou seja, supondo que é verdadeira a afirmação de H_0 que corresponde à igualdade» (Guimarães e Cabral, 2007, p. 238).

Assim, torna-se necessário definir o valor a partir do qual se torna improvável a validade da hipótese nula, isto é, tem que se especificar a região de rejeição de H_0 (ou região crítica) e, conseqüentemente, a região de não rejeição de H_0 . A decisão de rejeitar ou não a hipótese nula é tomada a partir da comparação do valor da estatística de teste com os, designados, valores críticos que definem as regiões de rejeição e de não rejeição de H_0 .

Se o valor da estatística de teste pertencer à região de rejeição, então a hipótese nula é rejeitada. Caso contrário, a hipótese nula não é rejeitada.

A dificuldade em se obterem os valores críticos que delimitam as duas regiões e atendendo a que ao se retirarem conclusões da amostra para a população não se podem ter certezas, isto é, que existe sempre uma probabilidade de errar, leva a que a decisão de rejeitar ou não a hipótese nula seja tomada de forma a limitar o erro de decisão.

Há dois possíveis tipos de erros quando se realiza um teste de hipóteses:

- O erro de tipo I (tem probabilidade máxima α) consiste em rejeitar a hipótese nula quando esta é, de facto, verdadeira.
- O erro de tipo II (tem probabilidade β) consiste em não rejeitar a hipótese nula quando esta é, de facto, falsa.

Em esquema tem-se:

Decisão baseada na amostra	Situação real na população	
	H_0 é verdadeira	H_0 é falsa
Não rejeitar H_0	Decisão correta (probabilidade $1 - \alpha$)	Erro tipo II (probabilidade β)
Rejeitar H_0	Erro tipo I (probabilidade máxima α)	Decisão correta (probabilidade $1 - \beta$, indica a potência do teste).

Desta forma pretende-se reduzir ao mínimo as probabilidades α e β dos dois tipos de erros. No entanto, diminuir a probabilidade de se cometer um erro traduz-se no aumento da probabilidade de se cometer o outro erro. Por este facto, a decisão a tomar é usualmente baseada no erro tipo I.¹

Assim, define-se à partida o nível de significância (α), isto é, a probabilidade máxima de incorrer no erro tipo I.

⁽¹⁾ Note-se que só é possível diminuir os dois tipos de erro em simultâneo com o aumento da dimensão da amostra.

Ao realizar-se um teste calcula-se a probabilidade de significância (*valor-p* ou valor de prova ou, em inglês, *p-value*¹⁾) que corresponde ao menor nível de significância (α) em que a hipótese nula, admitida como verdadeira, pode ser rejeitada, isto é, representa uma medida do grau com que os dados amostrais contradizem a hipótese nula. Assim, corresponde à probabilidade de a estatística de teste tomar um valor igual ou mais extremo (mais desfavorável para a hipótese nula, admitindo que esta é verdadeira) do que aquele que, de facto, é observado.

Obviamente, quanto menor for a probabilidade de significância associada ao valor do teste estatístico, maior será o grau com que a hipótese nula é contradita, ou seja, menor será o erro de tipo I (rejeitar H_0 quando H_0 é verdadeira).

Em suma, se $p\text{-value} \leq \alpha$ então rejeitar H_0
se $p\text{-value} > \alpha$ então não rejeitar H_0

Por exemplo, se *p-value* for 0,001 então significa que se rejeita H_0 admitindo uma probabilidade de erro (tipo I) muitíssimo pequena e, claro está, que se se admitisse uma probabilidade máxima de erro de 0,1, 0,05 ou 0,01 (valores de α mais usuais em testes), então, também se rejeitaria a hipótese nula.

Assim, a finalidade dos procedimentos estatísticos é reduzir o nível de incerteza associada à decisão, isto é, limitando a probabilidade da decisão tomada ser errada.

A realização de um teste de hipóteses segue uma metodologia de trabalho que visa a minimização dos erros de decisão. Por exemplo, Doane e Seward (2008, p. 347) referem uma metodologia com sete fases: 1) ter uma ideia; 2) formular hipóteses testáveis; 3) desenhar a experiência; 4) estabelecer a regra de decisão; 5) recolher os dados; 6) tomar a decisão; e 7) rever as ideias (voltar ao passo 2).

Já Guimarães e Cabral (2007, p. 237) definem uma metodologia com quatro fases: 1) definição das hipóteses; 2) identificação da estatística de teste e caracterização da sua distribuição; 3) definição da regra de decisão; e 4) cálculo da estatística de teste e tomada de decisão. Será esta metodologia que vai ser seguida nos diferentes exemplos, em que, no entanto, a decisão será tomada com base na probabilidade de significância associada ao valor do teste e não com base no próprio valor do teste.

Usualmente os diferentes testes são classificados em dois grandes grupos:

- **Paramétricos** – quando os testes envolvem hipóteses relativas a um parâmetro da população ou à comparação de parâmetros de duas ou mais populações. Tem-se uma «hipótese paramétrica quando a forma da função de distri-

⁽¹⁾ No SPSS é identificado com *Sig.*

buição ou da função densidade (função probabilidade) é (suposta) conhecida e a conjectura diz respeito apenas ao parâmetro» (Murteira *et al.*, 2002, p. 386).

Desta forma, são efetuados testes paramétricos quando se conhece a distribuição amostral (sendo a normal a mais comum) e, regra geral, quando a característica em estudo é quantitativa (ou tratada como tal). No entanto, alguns testes paramétricos requerem a verificação de outros pressupostos para a sua realização (por exemplo, a homogeneidade das variâncias).

- **Não-paramétricos** – são procedimentos estatísticos que podem ser utilizados para testar hipóteses quando não são definidas proposições (feitas suposições) sobre os parâmetros ou distribuições populacionais (Webster, 1998).

Reis *et al.* (2019, p. 231) referem que «um método estatístico diz-se não-paramétrico se satisfaz pelo menos uma das seguintes condições: 1) é utilizado com variáveis medidas numa escala nominal ou ordinal, isto é, com características qualitativas; 2) é utilizado com dados quantitativos, mas a função distribuição da variável (aleatória) não está especificada ou estando, está a menos de um número infinito de parâmetros desconhecidos».

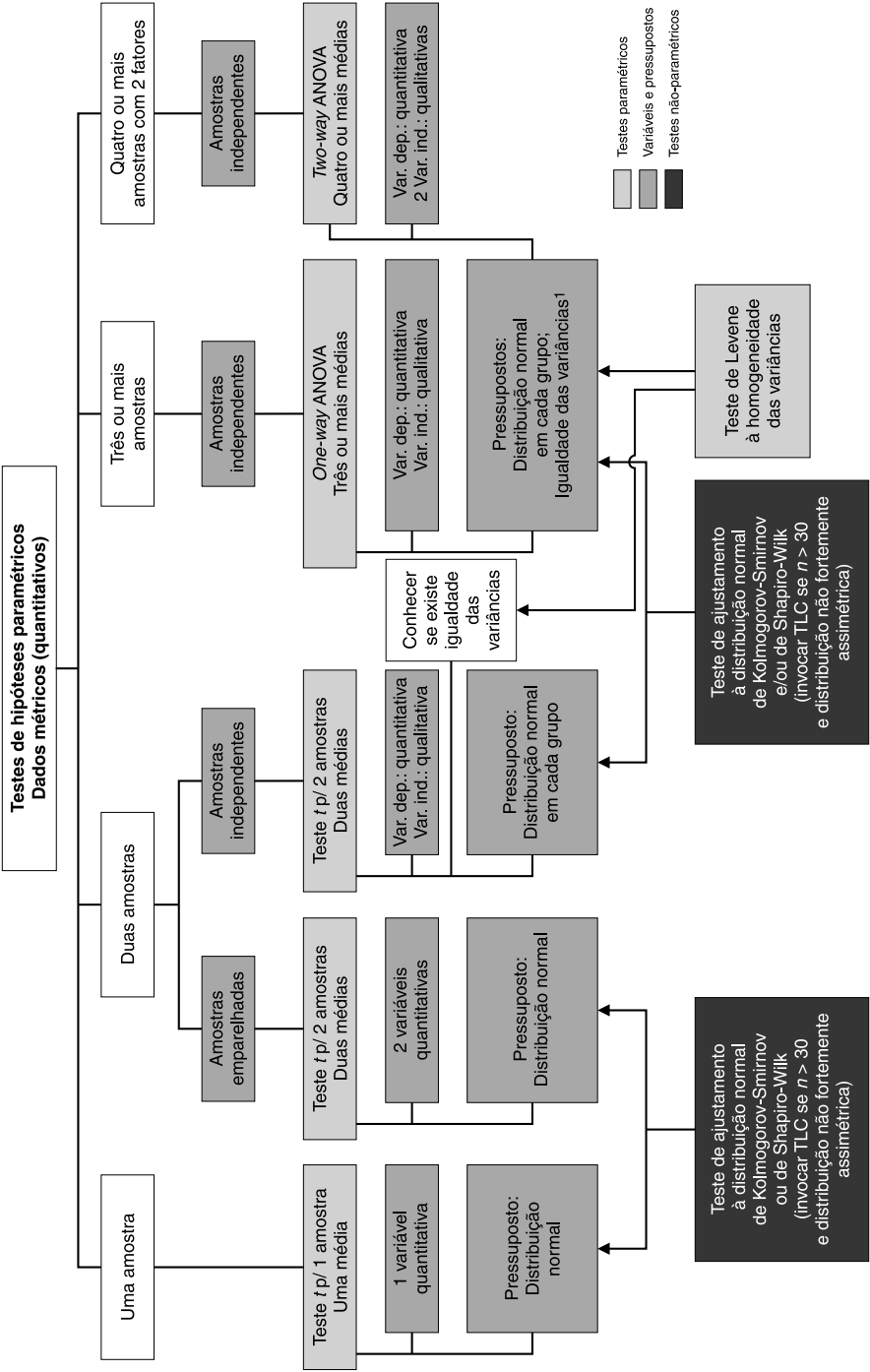
Desta forma, os testes não-paramétricos podem ser realizados como alternativa aos testes paramétricos para variáveis quantitativas, quando não se verificam os seus pressupostos. No entanto, Maroco (2007, p. 213) adverte que os testes paramétricos são mais potentes em certas situações, ou seja, a probabilidade de decidir corretamente (a probabilidade de rejeitar corretamente a hipótese nula) é inferior nos testes não-paramétricos.

Seleção de um teste estatístico

Para a seleção de um teste de hipóteses devem-se ter em consideração alguns aspetos, tais como o tipo de dados, o número e a natureza das amostras e os objetivos pretendidos.

Para auxiliar a escolha do teste adequado apresentam-se de seguida dois esquemas.¹ O primeiro para testes paramétricos e o segundo para os não paramétricos.

(1) Estes esquemas estão disponíveis na página do livro em www.silabo.pt.



(1) No caso de não se verificar a igualdade das variâncias realiza-se o teste de Welch ou de Brown-Forsythe.



RAUL MANUEL SILVA LAUREANO (Iscte – Instituto Universitário de Lisboa). É doutor em Gestão com especialização em Métodos Quantitativos para Gestão, professor da Iscte-Business School e investigador do grupo de *Data Analytics* da BRU-Iscte e do grupo de Sistemas de Informação do ISTAR-Iscte. Tem lecionado unidades curriculares de Estatística e Análise de Dados. Atualmente é diretor do mestrado em *Business Analytics* da Iscte-Business School e foi diretor da pós-graduação em *Analytics for Business* da Iscte-Executive Education. É formador de cursos profissionais de utilização do IBM SPSS Statistics e do Microsoft Excel e é (co)autor de outros livros e de mais de 100 publicações científicas.

A análise de dados é necessária em todas as áreas do conhecimento e tem vindo a ganhar maior relevância no apoio à tomada de decisão, num contexto de crescente adoção de uma cultura analítica por parte das organizações.

Esta obra apresenta e explica, numa linguagem acessível, mas com o necessário rigor científico, as técnicas de análise de dados mais utilizadas, permitindo, quer a estudantes, quer a profissionais, realizarem análises, que envolvam técnicas descritivas, inferenciais e de regressão, com recurso ao IBM SPSS Statistics ou ao R.

Com este manual ganha competências para:

- Interpretar medidas descritivas.
- Calcular a dimensão adequada para uma amostra.
- Selecionar o teste de hipóteses adequado para uma, duas ou mais amostras.
- Realizar testes de hipóteses paramétricos e não-paramétricos.
- Estimar modelos de regressão linear.
- Estimar modelos de regressão logística.
- Elaborar uma ficha técnica para uma análise estatística.
- Reportar e interpretar os resultados numa tese ou relatório de análise de dados.
- Utilizar o IBM SPSS Statistics em testes de hipóteses e em regressão.
- Utilizar o R para análise descritiva (gráficos e tabelas), análise inferencial (testes de hipóteses) e análise preditiva (modelos de regressão).



ISBN 978-989-561-051-8



9 789895 610518

659