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

LISBOA

UNIVERSIDADE DE LISBOA

Testes de hipóteses: uma abordagem não paramétrica

Maria José de Almeida Caetano de Sousa Firmino

Dissertação

Mestrado em Matemática para Professores

2015

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

Testes de hipóteses: uma abordagem não paramétrica

Maria José de Almeida Caetano de Sousa Firmino

Dissertação orientada pela Prof.ª Dr.ª Maria Fernanda Diamantino

2015

Dedicatória
Ao meu filho João

Agradecimentos

Este espaço é dedicado àqueles que deram a sua contribuição para que esta dissertação fosse realizada. A todos eles deixo aqui o meu agradecimento sincero. Este trabalho foi possível com o apoio de muitas pessoas.

- ◇ Ao meu marido João, antes de a quaisquer outros, devo o profundo agradecimento pelo modo como me aturou, pelo modo como sempre me apoiou e acompanhou ao longo da vida e em especial nesta árdua e custosa caminhada. Sempre que necessário soube aconselhar e soube criticar, como sempre e em tudo na vida. Pelas alegrias, momentos felizes, desânimos, angústias e essencialmente pela compreensão que durante já longos 38 anos me tem acompanhado incondicionalmente.

- ◇ Ao meu filho João que ao longe, no país distante que é o Chipre, ia vendo, criticando e corrigindo o que eu ia fazendo num programa que para mim era completamente novo o "LATEX".

- ◇ À minha orientadora Professora Fernanda Diamantino agradeço o seu apoio, o seu carinho e a sua inteira disponibilidade para me aconselhar e orientar ao longo de todo o desenvolvimento da tese.

- ◇ À minha mãe que embora não percebendo nada do assunto dizia "vai em frente".

- ◇ Aos amigos que ao longo do tempo se interessaram e foram dando incentivos para continuar.

- ◇ Aos meu colegas, Carla, Ilca e Tito que ao longo destes dois anos sempre colaboramos juntos e nos ajudámos nesta tarefa a que nos propusemos cumprir.

- ◇ Por fim dedico este meu trabalho ao meu pai que partiu em 2011. Se fosse vivo seria para ele um enorme orgulho.

Resumo

A Estatística é, hoje em dia, crucial para o desenvolvimento da sociedade em problemas tão diversos como o combate a doenças epidémicas variadas, a implementação de novos fármacos, o estudo de risco ambiental, o controlo de qualidade na indústria, estudos em ciências sociais, o desenvolvimento de modelos económicos apropriados, a disseminação da informação feita pela comunicação social. A intervenção da Estatística em cada uma destas áreas requer, hoje em dia, uma formação exigente, que permita aos profissionais terem um papel pró-activo junto dos diversos agentes.

Os testes estatísticos são fundamentalmente utilizados em pesquisas que têm por objectivo comparar condições experimentais. Os testes podem ser divididos em paramétricos e não paramétricos.

Uma justificação para o uso de métodos não paramétricos é a simplicidade. Em certos casos, até mesmo quando o uso de métodos paramétricos é justificado, os métodos não paramétricos são mais fáceis de usar. Devido tanto à simplicidade quanto à maior robustez, os métodos não paramétricos são vistos por algumas pessoas da área da estatística como o método que deixa menos espaço para usos indevidos e mal-entendidos.

A maior aplicabilidade e a maior robustez dos testes não paramétricos têm um custo: em alguns casos onde os testes paramétricos seriam apropriados, testes não paramétricos têm menos potência estatística. Por outras palavras, uma amostra maior pode ser necessária para retirar conclusões com o mesmo grau de confiança.

Os testes não paramétricos não têm exigências quanto ao conhecimento da distribuição da variável na população.

Estes testes são cada vez mais usados em análise estatística, sobretudo na área das Ciências Sociais, nas Ciências Admnistrativas (por exemplo em estudos de Marketing) e nas Ciências da Saúde, especialmente em Psiquiatria e Psicologia. A Estatística não paramétrica representa um conjunto de ferramentas de uso mais apropriado em pesquisas onde não se conhece bem a distribuição da população e os seus parâmetros.

Este trabalho teve como objectivo principal o estudo de testes não paramétricos e a sua aplicação em diversas situações.

Foram estudados alguns testes de hipóteses não paramétricos e, sempre que possível, foi dado um exemplo de aplicação desses mesmos testes.

Foi feita uma aplicação prática de um teste, neste caso do teste do Qui-Quadrado de independência para estudar a influência do grau de escolaridade dos pais no resultado académico dos alunos, tendo por base os dados recolhidos nas duas turmas leccionadas pela autora.

Palavras-Chave: Estatística não paramétrica, testes de hipóteses.

Abstract

Today Statistics is crucial to the development of the society, in issues as diverse as the fight against several epidemic diseases, the implementation of new drugs, the study of environmental risk, industry quality control, studies in social sciences, the development of appropriate economical models and the dissemination of information made by the media. Today the intervention of Statistics in each of these areas requires a demanding training, which allows professionals to have a proactive role among several agents.

Statistical tests are mainly used in research to compare experimental conditions. They can be divided into parametric and non-parametric tests.

A justification for the use of non-parametric methods is simplicity. In some cases, even where the use of parametric methods is justified, non-parametric methods are easier to use. Due both to simplicity and robustness, non-parametric methods are seen by some people in the statistical field as the method that allows less space for misunderstandings and inappropriate uses.

The wider applicability and robustness of nonparametric tests have a cost: in some cases where parametric tests would be appropriated, non-parametric tests have less statistical power. In other words, a larger sample may be required to draw conclusions with the same degree of confidence.

Non-parametric tests have no requirements concerning the knowledge of the variable distribution in the population.

These tests are increasingly used in statistical analysis, especially in the area of Social Sciences, in Administrative Sciences (e.g. in marketing studies) and in the Health Sciences, especially in Psychiatry and Psychology. The non-parametric statistics represents a set of more appropriate tools in research where the population distribution and its parameters are not very well defined.

This work had as main objective the study of non-parametric tests and their application in several situations.

Some statistical non-parametric tests were studied and whenever possible it has been given an example of an application of those tests.

A practical application of a test was done, in this case the Chi-Square independence was applied to study the influence of the educational level of parents on the academic results of the students, based on two classes taught by the author.

Key-words: Non-parametric statistics, hypothesis tests.

Conteúdo

Capítulo 1

Introdução

A palavra estatística, derivada do termo latino ≪status≫ (estado), parece ter sido introduzida na Alemanha, em 1748, por Achenwall. A Estatística é encarada, actualmente, como uma ciência capaz de obter, sintetizar, prever e tirar inferências sobre dados. Porém no século XVII em Inglaterra a estatística era a ≪Aritmética do Estado≫ (Political Arithmetic), consistindo basicamente na análise dos registos de nascimentos e mortes, originando mais tarde as primeiras tábuas de mortalidade. Ao longo da Idade Média e até ao século XVIII a estatística foi puramente descritiva, coexistindo duas escolas: a escola descritiva alemã, cujo representante mais conhecido é o economista G. Achenwall (1719-1772), professor na Universidade de Gottingen, considerado pelos alemães como o pai da estatística, e a escola dos matemáticos sociais que procuravam traduzir por leis a regularidade observada de certos fenómenos, de carácter económico e sociológico. Embora esta escola procurasse fundamentar a formulação de previsões com base em leis sugeridas pela experiência, a estatística confundia-se, praticamente, com a demografia à qual fornecia métodos sistemáticos de enumeração e organização. Na realidade, a necessidade sentida, em todas as épocas, de conhecer, numérica e quantitativamente, a realidade política e social tornou a análise demográfica uma preocupação constante.

No entanto, a estatística para adquirir o estatuto de disciplina científica, e não puramente ideográfica ou descritiva, teve que esperar pelo desenvolvimento do cálculo das probabilidades, que lhe viria a fornecer a linguagem e o aparelho conceptual permitindo a formulação de conclusões com base em regras indutivas. Data do século XVII o início do estudo sistemático dos problemas ligados aos fenómenos aleatórios, começando a ser manifesta a necessidade de instrumentos matemáticos, aptos a analisar este tipo de fenómenos, em todas as ciências que põem o problema do tratamento e interpretação de um grande número de dados. Pode datar-se dos fins do século XIX o desenvolvimento da estatística matemática e suas aplicações, com F. Galton (1822-1911), K. Pearson (1857-1936) e W. S. Gosset (1876-1936), conhecido sob o pseudónimo de Student, sendo lícito afirmar-se que a introdução sistemática dos métodos estatísticos na investigação experimental se fica a dever, fundamentalmente, aos trabalhos de K. Pearson e R. A. Fisher (1890-1962). A partir de Pearson e Fisher, o John Graunt (1620-1674), juntamente com William Petty (1623-1687), autor de Political Arithmetic, e o astrónomo Edmond Halley (1656-1742) são os principais representantes da escola inglesa, que dá um novo impulso à estatística, fazendo-a ultrapassar um estádio puramente descritivo: analisam-se os dados na procura de certas regularidades, permitindo enunciar leis e fazer previsões.

Estatística é uma ciência exacta que visa fornecer meios ao analista para organizar, resumir, analisar e apresentar dados. Está interessada na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises.

Em sentido mais estrito, o termo estatísticas é usado para designar os próprios dados ou algumas características que podemos calcular a partir deles tais como por exemplo média e variância.

O objetivo da Estatística consiste em extrair informação dos dados que nos são apresentados para obter uma melhor compreensão das situações que representam e sobre os problemas em estudo.

Antes de se recolher a amostra deve-se planear a experiência que nos vai permitir recolher os dados, de modo que, posteriormente, se possa extrair o máximo de informação relevante para o problema em estudo, ou seja para a população de onde os dados provêm.

Depois de recolher os dados, a análise inicial incide sobre a sua ordenação, resumo através do cálculo de características amostrais, agrupamento em classes (quando necessário) e representação gráfica.

Seguidamente o objectivo do estudo estatístico pode ser o de estimar parâmetros ou testar uma hipótese, utilizando-se técnicas estatísticas convenientes, as quais realçam toda a potencialidade da Estatística. Esta é a ciência que se ocupa da obtenção de informação (amostragem, planeamento de experiências), seu tratamento inicial (ordenação, cálculo de características amostrais, agrupamento em classes, representações gráficas - em suma, estatística descritiva e análise exploratória de dados), com a finalidade de, através de resultados probabilistas adequados, inferir de uma amostra para a população (decisão sobre hipóteses, estimação de parâmetros populacionais a partir de características amostrais relevantes, comparação de populações, relacionamento de uma variável com varáveis controladas), e eventualmente mesmo prever a evolução futura de um fenómeno (previsão).

A Estatística nos dias de hoje é uma ferramenta indispensável para qualquer profissional que necessita de analisar informações nas suas tomadas de decisões diárias, seja no seu trabalho ou na sua vida pessoal. Pode-se até pensar que as suas técnicas nasceram neste mundo contemporâneo em que se valoriza cada vez mais a rapidez e a agilidade das informações, de um mundo onde o avanço tecnológico (através da criação de computadores que processam uma imensa quantidade de dados num "*piscar de olhos*" é constante. Porém, a utilização da estatística como suporte para a tomada de decisões é verificada também no mundo antigo, e indícios da sua utilização são encontrados até na Era antes de Cristo.

Os Census são entendidos como processos normalizados de recolha, tratamento, avaliação, análise e difusão dos dados referenciados a um momento temporal específico e respeitantes, a todas as unidades estatísticas (indivíduos, famílias, alojamentos e edifícios) de uma zona geográfica bem delimitada, normalmente um país. Este não é um procedimento dos tempos passados. Na verdade, constitui uma importante área da Estatística..

No século XIX, surgiu outro campo da Estatística que se designa por Estatística Indutiva ou Inferência Estatística.

Esta área da Estatística preocupa-se em estimar o verdadeiro valor desconhecido do(s) parâmetro(s) de uma população e testar hipóteses com respeito ao valor dos parâmetros, ou à natureza da distribuição da população.

A análise paramétrica foi a primeira técnica de inferência estatística que apareceu em que se formulavam diversas hipóteses sobre a natureza dos parâmetros da população, da qual se retiraram os dados. Atendendo a que os valores relacionados com a população são vulgarmente designados de "parâmetros", estas técnicas chamar-se-iam de paramétricas. Os testes paramétricos visam analisar a variabilidade dos resultados da variável dependente, em função da manipulação das variáveis independentes, de forma a que se possa refutar ou aceitar a hipótese nula, a qual postula que os resultados da investigação são devidos, não aos efeitos previstos pela hipótese experimental, mas a diferenças aleatórias nos resultados, devidas a outras variáveis irrelevantes ou ao acaso.

Os testes paramétricos exigem que a(s) amostra(s) tenham uma distribuição normal, especialmente se tiverem uma dimensão inferior a 30. Em caso de dimensão superior, a distribuição tem de se aproximar da distribuição normal.

Os testes não paramétricos quando comparados com os testes paramétricos, requerem menos pressupostos para as distribuições. Baseiam-se em dados ordinais e nominais e são muito úteis para a análise de testes de hipóteses; são também úteis para a análise de amostras grandes, em que os pressupostos paramétricos não se verifiquem, assim como para as amostras muito pequenas e para as investigações que envolvam hipóteses cujos processos de medida sejam ordinais.

Centralizarei o meu estudo sobre a Estatística Não Paramétrica. Os primeiros métodos da estatística não paramétrica, embora com pouco uso até aos anos 40, foram referidos por John Arbuthnot em 1710. Estes começaram a ter maior impacto só a partir de 1942 com Wolfowitz. A partir daí o interesse aumentou de uma forma rápida.

Hoje a Estatística Não Paramétrica é considerada um dos campos mais importantes da Estatística. As técnicas que advêm desta categoria são usadas com grande frequência nas ciências físicas, biológicas e sociais ou até mesmo na comunicação. Outros autores, também dão importância a outros campos, tais como, na análise de dados da qualidade da água (Helsel), em aplicações na medicina (Brown and Hayden) ou mesmo na psicologia.

Um teste não paramétrico testa outras situações que não parâmetros populacionais. Estas situações podem ser modelos, dependência ou independência e factores aleatórios.

Estes testes são menos exigentes do que os paramétricos. Dispensam por exemplo, a normalidade dos dados, são independentes da forma da população da qual a amostra foi obtida.

Exemplo de alguns testes não-paramétricos: teste de Wilcoxon; teste de U Mann-Whitney; teste de Kruskal-wallis; teste de Qui-quadrado; teste de Friedman, entre outros.

Os testes não paramétricos não estão condicionados por qualquer distribuição de probabilidades dos dados em análise, sendo também designados por "*distribution-free test*".

Tal como não é estatisticamente rigorosa a utilização de testes paramétricos quando não se cumprem os pressupostos necessários, também deverá ser evitada a utilização dos testes não paramétricos em situações em que prevalecem as condições de utilização dos testes paramétricos, pois estes (paramétricos) são mais potentes que os testes não paramétricos.

Trate-se de um teste paramétrico ou não paramétrico, para lá dos pressupostos acima referidos, qualquer teste de hipóteses só tem validade estatística se as amostras sobre as que estão a ser aplicados forem aleatórias. Assim, dentro dos testes não paramétricos, veremos alguns que se aplicam para verificar a aleatoriedade das amostras.

De um modo geral, as variáveis qualitativas estão mais ligadas aos modelos não paramétricos, enquanto as variáveis quantitativas aos modelos paramétricos.

Capítulo 2

Introdução aos testes de hipóteses

Os testes de hipóteses são uma metodologia que nos permite fazer inferência sobre uma ou mais populações a partir do estudo de uma ou mais amostras.

O objectivo de um teste de hipóteses é determinar se uma hipótese ou conjectura que fazemos àcerca de um parâmetro de uma população é plausível, isto é, se tem razão de ser, com base na informação obtida a partir de uma amostra extraída dessa população.

Em Estatística, um teste de hipóteses é um método para verificar a validade ou não de uma hipótese. É um procedimento estatístico baseado na análise de amostras. O seu uso está condicionado à dimensão da amostra e à respectiva distribuição da variável em estudo. São constituídos por duas hipóteses, a hipótese a ser testada designamos por **Hipótese Nula** (H_0), que corresponde frequentemente ao estado actual, ao que é tradicionalmente aceite. Reflete a situação em que não há mudança. É a hipótese a refutar. A **Hipótese Alternativa** (H_1) corresponde a uma situação em que existe uma alteração face ao que é habitual; exprime, por exemplo, aquilo que um investigador está a tentar estabelecer com um novo estudo sobre o assunto.

Quando formulamos uma decisão sobre H_0 podem ocorrer dois erros distintos. O primeiro, designado por **erro tipo I**, consiste em rejeitar a hipótese nula quando ela é verdadeira. O segundo, designado por **erro tipo II**, consiste em não rejeitar H_0 quando ela é falsa.

A estes erros estão associadas probabilidades, isto é:

$$\text{P(rej } H_0|H_0 \text{ verd)}=\alpha$$

$$\text{P(não rej. } H_0|H_0 \text{ falsa)}=\beta$$

À probabilidade α damos o nome de **nível de significância do teste**.

	H_0 Verdadeira	H_0 Falsa
Não rejeitar H_0	Decisão correcta	Erro tipo II
Rejeitar H_0	Erro tipo I	Decisão correcta

Como o valor de α entra no processo de determinação de rejeição ou não rejeição de H_0 , a condição de objectividade da prova exige que o nível de significância seja fixado antes da recolha de dados. Os valores mais usuais para "alpha"são de 0,01, 0,05 e 0,01 de acordo com a importância prática dos resultados.

A probabilidade de não cometer um erro tipo II é o que se denomina potência do teste.

Sendo β a probabilidade de cometer um erro do tipo II, ou seja, a probabilidade de não rejeitar a hipótese nula quando esta é falsa, a potência do teste é dada por $1-\beta$.

A potência de um teste de hipóteses só pode ser determinada a partir de um valor concreto para o parâmetro que se pretende testar. Deste modo, não é geralmente possível determinar à *priori* a potência dum teste estatístico, pois o valor do parâmetro é desconhecido (por isso é que se realiza o teste).

Quanto mais pequena é a probabilidade β, mais potente é o teste, ou seja, o teste óptimo da hipótese H_0 *vs.* H_1 é aquele que para uma probabilidade de ocorrer o erro tipo I, torne mínima a probabilidade de ocorrer o erro tipo II.

À medida que se diminui o nível de significância dum teste, diminui também a sua potência.

A Estatística de teste é uma variável aleatória, função apenas da amostra, com base na qual será tomada a decisão de rejeitar ou não a hipótese nula. A sua distribuição é conhecida no caso de H_0 ser verdadeira.

2.1 Como realizar um teste de hipóteses?

• Formular a hipótese nula (H_0) e a hipótese alternativa H_1;

• Recolhida uma amostra, observamos uma função da amostra aleatória (valor da estatística de teste) cuja distribuição de probabilidade é conhecida pressupondo que H_0 é verdadeira.

• A decisão a tomar será rejeitar H_0 ou não rejeitar H_0.

• O nível de significância, α, e a distribuição de probabilidade da estatística de teste são utilizados para definir aquilo a que chamamos região crítica ou região de rejeição.

• Se o valor observado da estatística de teste pertencer à região de rejeição, a decisão é rejeitar H_0, caso contrário, a decisão é não rejeitar H_0.

Resumindo:

1. Formular a hipótese nula (H_0) e a hipótese alternativa H_1;

2. Estabelecer o nível de significância;

3. Escolher a estatística de teste a usar e encontrar qual a sua distribuição de probabilidade supondo que H_0 é verdadeira;

4. Determinar a região de rejeição;

5. Calcular o valor observado da estatística de teste;

6. Decidir rejeitar H_0 ou não rejeitar H_0,

7. Apresentar a conclusão de acordo com o problema.

2.2 Variáveis estatísticas. Escala de Stevens

O valor de um atributo de uma população pode variar de elemento para elemento. Chama-se ao atributo em estudo, variável. As variáveis podem ser do tipo:

• qualitativo, também designada factor se as suas diferentes modalidades não são mensuráveis;

- quantitativo, caso contrário.

As variáveis quantitativas podem ser:

- discretas, se tomam valores num conjunto finito ou infinito numerável;
- contínuas, se tomam valores num intervalo de números reais.

Recordemos a classificação de Stevens das "escalas" em que os dados são observados ou registados:

- nominais
- ordinais
- intervalares
- de razões, ou absoluta.

As duas primeiras são apropriadas para dados qualitativos, as duas últimas para dados quantitativos, e condicionam fortemente a escolha de métodos estatísticos que é legítimo usar. Claro que é sempre possível passar de dados mais sofisticados para menos sofisticados - por exemplo, considerando ordens e desprezando magnitudes, passar de dados em escala absoluta a dados ordinais; ou, por agrupamento em classes, e dados meramente nominais.

No que refere dados puramente nominais, apenas podemos *contar* quantos indivíduos pertencem a cada uma das classes, ou usar as correspondentes frequências relativas.

Já no que refere dados ordinais, ficam acessíveis todos os métodos "não paramétricos" baseados em *ranks.*

As escalas intervalares e as escalas de razões permitem operações aritméticas (plenamente na escala de razões), e o recurso a métodos estatísticos mais sofisticados.

"Se o modelo *Gaussiano* for aceitável, dispomos de métodos estatísticos simples, apoiados numa teoria sólida, pois nessa situação média e variância empíricas são estimadores *independentes* do valor médio e da variância populacionais, e o estudo das distribuições amostrais de estatísticas studentizadas ou de quocientes de quadrados médios (análise de variância) tem uma "elegância inexcedível".

Porém, prescindindo da hipótese de população parente Gaussiana, as dificuldades parecem inultrapassáveis: o teorema de Darmois-Skitovich estabelece que a independência entre $\overline{X}$ e S^2 é uma caraterização do modelo Gaussiano e a estrutura de dependência nos outros casos é regra geral complicada." In Dinis Pestana *Introdução à Probabilidade e à Estatística* (2002).

2.3 Testes não paramétricos

Existem fundamentalmente dois tipos de testes estatísticos, designados por testes paramétricos e não paramétricos. A principal diferença entre eles é a sofisticação das medidas utilizadas para calcular a variabilidade dos resultados. Uma das vantagens dos testes não paramétricos é que podem ser utilizados quando os dados experimentias apenas podem ser medidos numa escala ordinal, admitindo-se ainda a sua utilização em algumas situações, em que os dados são medidos numa escala nominal.

Muitos dos testes estatísticos não paramétricos respondem à mesma série de questões tal como os testes paramétricos. Com testes não paramétricos as hipóteses podem ser flexibilizadas consideravelmente. Por conseguinte, são utilizados métodos não paramétricos para situações que violem os pressupostos de procediemtos paramétricos.

Os testes não paramétricos requerem menos pressupostos em relação à população;

• Não exigem normalidade;

• Não se baseiam em parâmetros da distribuição (logo, não necessitam variâncias homogéneas);

• Ligeiramente menos eficientes que os testes paramétricos;

• Baseiam-se nas estatísticas ordinais (e não nos valores das observações);

• Mais fáceis de aplicar.

Vejamos ainda quais as vantagens e as desvantagens dos testes não paramétricos:

Vantagens

• Poucos pressupostos relativos à população

• Facilidade de implementação

• Maior perceptibilidade

• Aplicável em situações não abrangidas pela Normal

• Mais eficientes quando as populações não têm Distribuição Normal

• Os resultados podem ser tão exactos como nos procedimentos paramétricos.

Desvantagens

• As hipóteses testadas por testes não-paramétricos tendem a ser menos específicas;

- Não têm parâmetros. Dificultam as comparações quantitativas entre populações;
- Escasso aproveitamento de informação da amostra
- Pode ser de difícil cálculo à mão para grandes amostras
- As tabelas não são amplamente disponíveis

Região crítica ou de rejeição

É constituída por um conjunto de valores tomados pela estatística de teste, que conduzem à rejeição da hipótese nula.

Regra de Decisão Estatística

É uma regra que nos indica a decisão a tomar (rejeitar ou não H_0), a partir da comparação do valor da estatística de teste com um ou mais valores críticos (será um valor crítico para os testes unilaterais e dois para os testes bilaterais).

Região de aceitação

É constituída por um conjunto de valores tomados pela estatística de teste, que conduzem à não rejeição da hipótese nula.

Valor-p ou *p-value*

O valor- p, define-se como o menor nível de significância,α , a partir do qual se rejeita a hipótese nula. Calcular o valor- p, é calcular a probabilidade do erro de 1ª espécie, correspondente a rejeitar a hipótese nula para a amostra observada, ou seja, para o valor da estatística de teste que foi observado. Fixado o nível de significância ,α , a decisão de rejeitar a hipótese nula verifica-se se e só se $valor - p \leq \alpha$.

Potência do teste

Chama-se potência do teste à probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira. Ou seja, rejeitar a hipótese nula quando esta é de facto falsa. A potência de um teste é = 1-β.

Capítulo 3

Testes para o caso de uma amostra

No caso de uma amostra verifica-se, se há diferenças significativas entre frequências observadas e as frequências que poderíamos esperar com base em determinado princípio, se há diferenças significativas entre a proporção observada a a proporção esperada e se é razoável admitir que a amostra seja uma amostra aleatória proveniente de alguma população com distribuição conhecida.

3.1 Teste do Qui-Quadrado

O teste de ajustamento do Qui-Quadrado é o teste mais conhecido, porventura por ter sido um dos primeiros grandes êxitos da Estatística como esteio de descobertas científicas em outras ciências, e por a sua justificação intuitiva ser simples.

O teste do Qui-Quadrado é um teste de hipóteses que é adequado aplicar quando temos os elementos da amostra divididos em duas ou mais categorias. O propósito deste método é ver se existem diferenças significativas entre o número de indivíduos, de objectos ou de respostas, em determinada categoria, e o respectivo número esperado na hipótese nula. Isto é, o teste do Qui-Quadrado destina-se a averiguar se uma amostra pode ser considerada como proveniente de uma população com uma determinada distribuição sem restrições sobre esta. Este teste também pode ser usado para verificar se as categorias de uma variável estão equitativamente distribuidas.

É um teste não paramétrico e, como tal, não depende de parâmetros populacionais como o valor médio e a variância.

O objectivo básico deste método é comparar proporções, isto é, indagar sobre as possíveis divergências entre as frequências observadas e esperadas para um certo acontecimento.

Evidentemente, pode dizer-se que dois grupos se comportam de forma semelhante se as diferenças entre as frequências observadas e as esperadas em cada categoria forem muito pequenas, próximas de zero.

O teste do Qui-Quadrado de ajustamento, consiste em comparar os dados obtidos experimentalmente com os dados esperados para um determinado acontecimento.

As hipóteses a testar são as seguintes:

H_0: a população segue uma determinada distribuição D;

vs.

H_1: a população não segue distribuição D.

Das comparações surgem diferenças que podem ser grandes ou pequenas: se forem grandes, a hipótese H_0 que pressupõe um bom ajustamento deverá ser rejeitada em favor da hipótese alternativa H_1 ; se forem pequenas, a hipótese H_0 não será rejeitada e as diferenças são atribuíveis ao acaso. O objectivo é comparar frequências observadas com frequências teóricas ou esperadas.

Uma medida da discrepância existente entre as frequências observadas e esperadas é proporcionada pela expressão:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}, \tag{3.1}$$

em que

k - número de classes;

O_i - frequência observada, é a frequência absoluta em cada classe;

e_i - frequência esperada, dada por e_i=Np_i com p_i a probabilidade da classe i, se a hipótese H_0 verdadeira;

N - é o número total de observações independentes.

X^2 segue aproximadamente uma distribuição de χ^2 com k-1 graus de liberdade.

Quando $X^2 = 0$, as frequências teóricas e observadas coincidem exactamente, enquanto quando X^2>0, isso não se verifica. Quanto maior for o valor de X^2, maior será a discrepância entre as frequências observadas e esperadas.

A distribuição amostral de X^2, sob H_0, calculada pela expressão dada anteriormente, segue uma distribuição Qui-Quadrado com k-1 graus de liberdade.

Genericamente, o número de graus de liberdade é o número de variáveis independentes, que contribuem efectivamente para a variabilidade do resultado.

3.1.1 Exemplo

A descendência originada pelo cruzamento de dois dados tipos de plantas pode ser qualquer um dos três genótipos que representaremos por A, B e C. Um modelo teórico de sucessão genética indica que os tipos A, B e C devem aparecer na razão de 1 : 2 : 1. Efectuou-se o cruzamento daqueles dois tipos tendo-se classificado 90 plantas. A sua classificação genética foi registada na tabela:

Genótipos	A	B	C
	18	44	28

Estão estes dados de acordo com o modelo genético?

H_0 : p_1=0,25, p_2=0,5, p_3=0,25

vs.

H_1 : pelo menos uma das probabilidades é diferente do formulado.

A estatística de teste, $X^2 = \sum_{i=1}^{3} \frac{(O_i - e_i)^2}{e_i}$ segue uma distribuição de χ^2_2 se H_0 é verdadeira.

A tomada de decisão, para $\alpha = 0,05$ é feita comparando-se o valor observado da estatística de teste de X^2 e o valor de χ^2_{Calc} da estatística com o quantil.

Assim, neste caso a região crítica, para um nível de significância α, é definida por:

$X^2_{Calc} \geq \chi^2_{0,95:2}$, em que

$\chi^2_{1-\alpha,k-1}$ representa o quantil de probabilidade $(1-\alpha) \times 100\%$

- Se X^2 calculado for maior ou igual que $\chi^2_{0,95;2}$ tabelado, rejeita-se H_0;
- Se X^2 calculado for menor que $\chi^2_{0,95;2}$ tabelado, não se rejeita H_0.

	A	B	C
o_i	18	44	28
p_i	0,25	0,5	0,25
e_i=Np_i	22,5	45	22,5

Calculando o valor observado da estatística do teste,

$$X^2_{Calc} = \frac{(18-22{,}5)^2}{22{,}5} + \frac{(44-45)^2}{45} + \frac{(28-22{,}5)^2}{22{,}5} = 2,27$$

Consultando a tabela do χ^2, $\chi^2_{0{,}95;2} = 5,99$

Então como $X^2_{Calc} < \chi^2_{0{,}95;2}$ não se rejeita H_0 ao nível de significância de 5%. Portanto, podemos assumir que os dados estão de acordo com o modelo genético.

3.2 Teste da Binomial

Este teste é aplicado em amostras provenientes de populações que estão divididas em duas categorias, por exemplo, masculino e feminino, membro ou não membro de uma qualquer associação, doente ou não doente. Nestes casos, qualquer observação possível sobre a população recairá numa ou noutra dessas duas categorias.

Para qualquer população dividida em duas categorias (isto é dicotomizada), se conhecermos a proporção, P, numa das categorias, a proporção na outra será $1 - P$.

O valor de P é fixo e desconhecido para uma determinada população. No entanto, mesmo que se saiba (ou se admita) o valor de P para determinada população, não podemos esperar que uma amostra aleatória extraída da referida população contenha exactamente a proporção P de casos numa categoria e a proporção $1 - P$ na outra.

A distribuição Binomial é o modelo probabilístico adequado para casos em que se consideram provas repetidas de Bernoulli, isto é, sucessões de experiências aleatórias independentes, em cada uma das quais se observa a realização ou não realização de um determinado acontecimento A, com probabilidade $P(A) = p$, constante de experiência para experiência. Por exemplo, lança-se uma moeda ao ar um certo número de vezes e pretende-se estudar a variável aleatória X, que representa o número de "caras"saídas nesses lançamentos. Suponhamos então que se lançou ao ar 20 vezes, uma moeda "equilibrada". Pretende-se estudar a variável aleatória X que representa o número de caras saídas nos 20 lançamentos.

A realização de A diz-se constituir um "*sucesso*" e a realização do seu complementar, $\overline{A}$, que tem probabilidade $P(\overline{A}) = 1 - p = q$, um "*insucesso*".

Se a variável aleatória X designa o número de sucessos em N provas independentes, a sua função massa de probabilidade é dada por:

$$P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}, x = 0, 1, ..., N \tag{3.2}$$

e dizemos que X segue uma distribuição Binomial com parâmetros N e p.

A distribuição Binomial é a distribuição amostral de uma proporção que podemos observar numa amostra aleatória extraída de uma população dicotomizada. Isto é, tal distribuição dá os diversos valores que podem ocorrer sob a hipótese H_0 em que $H_0 : P = p_0$. Portanto, quando os dados de uma pesquisa se apresentam dicotomizados, pode-se usar a distribuição Binomial para comprovar H_0.

Em resumo, os passos na aplicação do teste Binomial são os seguintes:

1. $H_0 : P = p_0$.

vs.

$H_1 : P \neq p_0$

2. Determinar o número total de casos observados N;

3. Determinar as frequências das ocorrências em cada uma das suas categorias;

4. O método para a determinação da probabilidade, sob H_0, da ocorrência dos valores observados ou valores extremos, varia:

4.1. Se $N \leq 25$ e se p=q=$\frac{1}{2}$, a tabela da binomial dá-nos "as probabilidades associadas a valores tão pequenos quanto os valores de x no teste Binomial". Ou seja, dá-nos as probabilidades unilaterais sob H_0. Emprega-se uma prova unilateral quando se pode especificar de antemão qual das categorias terá menos frequência. Para uma teste bilateral, é necessário duplicar os valores que se apresentam na referidada tabela;

4.2. Se $p \neq q$, determina-se a probabilidade, sob H_0, de ocorrência do valor observado x de acordo com

$$P(X \leq x) = \sum_{i=0}^{x} \binom{N}{i} p^i q^{N-i} \tag{3.3}$$

4.3. Para grandes amostras (N>25), quando N cresce, a distribuição Binomial tende para a distribuição Normal. Se p estiver próximo de $\frac{1}{2}$ utilizamos a aproximação pela Normal. Os parâmetros a usar serão o valor médio $\mu_X = Np$ e o desvio padrão $\sigma_X = \sqrt{Npq}$. Deste modo, Z tem distribuição aproximadamente Normal com valor médio 0 e variância 1, sendo:

$$Z = \frac{X-\mu_X}{\sigma_X} = \frac{X-Np}{\sqrt{Npq}}.$$

Devido à natureza da variável X ser discreta e a distribuição Normal ser contínua, deve-se introduzir um factor de correcção. Assim,

$$Z = \frac{(X \pm 0.5) - Np}{\sqrt{Npq}} \tag{3.4}$$

onde X+0.5 é utilizado quando X<Np e X-0.5 quando X>Np.

Então para grandes amostras e P próximo de $\frac{1}{2}$, testamos a hipótese aplicando a expressão 3.3. A tabela (de probabilidades associadas a valores tão extremos quanto os valores observados de Z na distribuição Normal) dá a probabilidade, sob H_0, associada à ocorrência de valores tão grandes quanto um valor de Z observado, dado por aquela expressão. A tabela dá os valores unilaterais de p, sendo necessário duplicá-los para teste bilateral.

Fixado um nível de significância α rejeita-se H_0 se o valor de p associado ao valor observado x, não superar α.

3.2.1 Exemplo

Num ensaio de degustação de café, cada mesa era constituída por 5 amostras, sendo duas delas de café "mole"e as 3 restantes de café "comum". Dos 8 degustadores que foram utilizados, 3 classificaram correctamente os tipos de café.

Teste a hipótese de que os degustadores conseguem distinguir o café "mole"dos demais.

Primeiro precisamos de saber qual a probabilidade de um degustador distinguir por acaso os dois cafés “moles”dentre as 5 amostras.

$$p_0 = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10} = 0,10$$

H_0: $p = 0,10$

vs.

H_1: $p > 0,10$

$X \cap Bi(8; 0,10)$ se H_0 é verdadeira.

$$P(X \geq 3) = \sum_{i=3}^{8} \binom{8}{i} 0,10^i 0,90^{8-i} = 0,0381$$

A probabilidade de 3 ou mais degustadores distinguirem correctamente os tipos de café, aleatoriamente, é de 0,0381 que significa o valor-p associado a este teste.

Ao nível de significância de 5% rejeitamos a hipótese nula. Há evidência para afirmar que os degustadores não conseguem distinguir o café “mole”dos demais.

3.3 Teste de Kolmogorov-Smirnov

Este teste foi proposto em 1933 por Kolmogorov e avalia o grau de concordância entre a distribuição de um conjunto de valores amostrais (observados) e uma determinada distribuição teórica. Determina se os valores da amostra podem ser considerados como provenientes de uma população com aquela distribuição teórica. Para isso utilizamos a função de distribuição empírica, compara-se com a distribuição teórica, determina-se o ponto em que estas distribuições mais divergem, e testamos se essa divergência é aleatória ou não.

Os dados devem seguir pelo menos uma escala ordinal.

Dada uma amostra de dimensão n $(x_1, x_2, ..., x_n)$, consideremos $S_n(X)$ (distribuição empírica) uma distribuição observada numa amostra de n observações e $F_0(X)$ uma distribuição teórica acumulada, sob H_0.

De seguida, determina-se o maior valor das diferenças entre $F_0(X)$ e $S_n(X)$, ou seja,

$$D = max|F_0(X) - S_n(X)|$$

O teste de Kolmogorov pode ser preferido em relação ao teste do Qui-Quadrado devido à qualidade do ajuste à amostra, se o tamanho desta for pequeno; o teste de Kolmogorov é exacto mesmo para pequenas amostras, enquanto que o teste do Qui-Quadrado assume que o número de observações é grande o suficiente para que a distribuição represente uma boa aproximação à estatística de teste. Há controvérsias sobre qual dos testes é o mais poderoso, mas actualmente é considerado que o teste de Kolmogorov é mais poderoso do que o teste do Qui-Quadrado na maioria das situações.

3.3.1 Exemplo

Efectuou-se uma experiência para calibrar a luminosidade adequada de uma nova máquina fotográfica. Foram tiradas 5 fotografias de cada uma das 10 pessoas que participaram na experiência. A cada pessoa perguntou-se qual das fotos apresentava uma maior qualidade, de 1 a 5, onde 1 representa um grau baixo e 5 um grau alto de luminosidade.

$H_0 : f_1 = f_2 = ... = f_5 = \frac{1}{5}$

vs.

$H_1 : f_1 \neq f_2 \neq ... \neq f_5$

	1	2	3	4	5
$F_0(X)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$
$S_{10}(X)$	$\frac{0}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{10}{10}$
$\lvert F_0(X) - S_{10}(X\rvert)$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	$\frac{2}{10}$	$\frac{0}{10}$

Observe-se que $F_0(X)$ é a distribuição acumulada teórica, sob H_0, onde H_0 é a hipótese de que cada uma das cinco cópias tenha precisamente $\frac{1}{5}$ das preferências. S_{10} é a distribuição acumulada das frequências observadas das escolhas dos 10 indivíduos.

Para $n = 10$ a $P(D \geq 0,5) < 0,01$, portanto rejeita-se H_0.

Conclui-se assim que os indivíduos apresentam uma preferência significativa em relação ao grau de luminosidade.

3.4 Teste dos sinais

O teste de hipóteses sobre a mediana é importante nas decisões sobre a localização da distribuição da população, até por não necessitar de qualquer pressuposto sobre a distribuição desta. Este é um teste para a mediana de uma população (m). Para este teste pressupõe-se que a distribuição da população é contínua.

As hipóteses a considerar são as seguintes:

$H_0 : m = m_0$;

vs.

$H_1 : m \neq m_0$.

O teste baseia-se no facto de que, se H_0 for verdadeira, então aproximadamente metade dos valores observados são inferiores a m_0. Assim, consideram-se as diferenças $x_i - m_0$ (ou $m_0 - x_i$), $i = 1, 2, ..., N$, não se rejeitando H_0 se o número de diferenças com sinal negativo for aproximadamente igual ao número de diferenças com sinal positivo.

A estatística de teste é S= número de observações abaixo (ou acima) de m.

Se a hipótese nula for verdadeira e a amostra for aleatória, o número de observações com valor inferior (ou superior) a m_0 é uma variável aleatória binomial com parâmetro $p = 0, 5$.

Retenha-se o sinal, positivo (+) ou negativo (-), das diferenças $x_i - m_0$.

A hipótese é posta em causa quando S é excessivamente "pequeno"ou excessivamente "grande"; se a hipótese é verdadeira, S tem distribuição Binomial $B(N, \frac{1}{2})$, e um teste de nível de significância α é o que leva a rejeitar a hipótese H_0 quando

$S \in \{0, 1, ..., s_0\}$ ou $S \in \{s_1, ..., N-1, N\}$

onde s_o é o maior inteiro tal que

$$P(S \leq s_o|H_0) = \sum_{m=0}^{s_o} \binom{N}{m} (\frac{1}{2})^N \leq \frac{\alpha}{2} \tag{3.5}$$

e s_1 é o menor inteiro tal que

$$P(S \geq s_1|H_0) = \sum_{m=s_1}^{N} \binom{N}{m} (\frac{1}{2})^N \leq \frac{\alpha}{2}. \tag{3.6}$$

Quando $N \geq 20$ pode utilizar-se a aproximação decorrente do Teorema de Moivre-Laplace (inicialmente formulado por De Moivre em 1733 e posteriormente tratado por Laplace em 1812, o teorema enuncia-se da forma seguinte:

Se a variável X segue uma distribuição Binomial $B(n, p)$ com $p \in]0, 1[$, então a variável

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1) \tag{3.7}$$

preferivelmente com correcção de continuidade. Este factor usa-se quando se pretende aproximar uma distribuição Binomial por uma distribuição Normal, aplica-se somando ou subtraindo 0,5 ao valor da variável).

Assim,

$$S^* = \frac{S - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \sim N(0,1) \tag{3.8}$$

rejeitando a hipótese H_0, ainda para o nível de significância α, se

$|S^*_{Calc}| \geq z_{1-\frac{\alpha}{2}}$

em que $z_{1-\frac{\alpha}{2}}$ é o quantil de probabilidade da N(0,1), isto é,

$$P(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} = \Phi(z_{1-\frac{\alpha}{2}})$$

Nota: Se a distribuição da população for simétrica devemos usar o teste de Wilcoxon.

3.4.1 Exemplo

Sabe-se que o rendimento familiar mediano numa determinada região é de 600 euros/mês. Uma amostra aleatória constituída por 24 famílias de uma vila daquela região revelou os seguintes rendimentos:

440, 466 482, 518 603, 617, 636, 727, 774, 824, 961, 1056,

650, 555, 1500, 750,820, 950, 828, 543, 1200, 1000, 790, 890

Denotando por m o rendimento mensal mediano naquela vila pretendemos testar

H_O: m=600

vs.

H_1: m$\neq$ 600

A hipótese nula estabelece que o rendimento mensal mediano é de 600 euros/mês; se esta hipótese é verdadeira, 50% das famílias terão um rendimento mensal inferior àquele valor (e 50% terá um rendimento mensal superior ao mesmo); isto é, o anterior teste pode escrever-se como:

H_0: p= 0,5

vs.

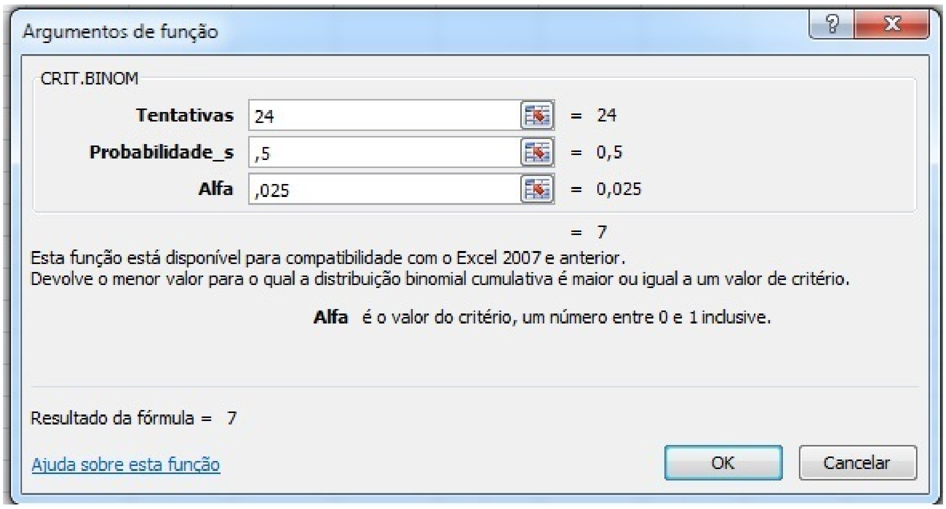
H_1: p $\neq$ 0,5.

Nestas condições, o número de famílias com rendimento inferior a 600 euros/mês numa amostra de 24 famílias segue uma distribuição Binomial $B(24; 0,5)$.

No nosso exemplo, S= 4 (número de famílias com rendimento inferior a 600 euros/mês).

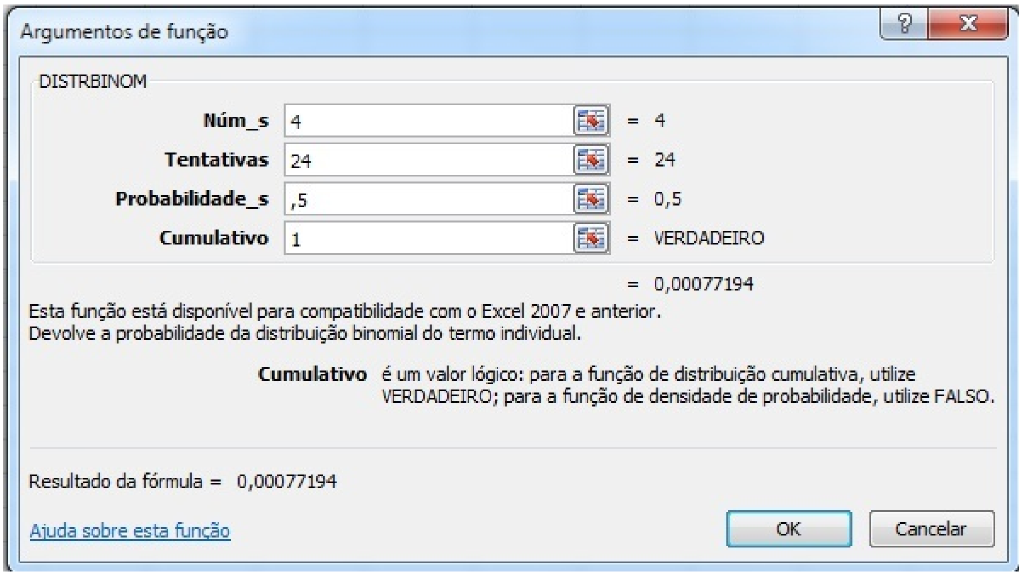
Para um nível de significância $\alpha = 5\%$, e sendo o teste bilateral, a hipótese nula seria rejeitada se na amostra ocorrerem menos de 7 famílias ou mais de 17 famílias com um rendimento mensal inferior a 600 euros/mês.

Este valor (ou quantil da distribuição binomial) pode ser calculado com a função CRIT.BINOM(N; p; α), do programa Excel e obtêm-se o seguinte:



(como se trata de um teste bilateral, o quantil que define o limite superior da região de não rejeição calcula-se colocando-o à mesma distância que separa o quantil inferior e a média).

A decisão do teste também se pode tomar, calculando a probabilidade limite (que geralmente todos os programas estatísticos apresentam nos testes de hipóteses). No programa Excel, a função DISTRIBINOM (k; N; p; cumulativo) calcula a função de distribuição cumulativa de probabilidades binomial, até k sucessos:



Tratando-se de um teste bilateral, valor-p=$2 \times 0,00077194 = 0,00154388$.

A decisão é rejeitar H_0, então a mediana do rendimento mensal é significativamente diferente de 600 euros.

Se o tamanho da amostra é muito grande, o cálculo das probabilidades da função binomial pode ser aproximado pela função de distribuição normal estandardizada, sendo:

$S \sim Bi(N, p)$

$S \sim N(\mu, \sigma)$ pelo Teorema do Limite Central

$\mu = N \cdot p$

$\sigma = \sqrt{N \cdot p \cdot (1-p)}$

e a estatística do teste é:

$Z = \frac{(k+0,5)-0,5 \cdot N}{\sqrt{N \cdot p \cdot (1-p)}} \sim N(0,1)$

No nosso exemplo apresentado, esta aproximação é:

$Z_{Calc} = \frac{(4+0,5)-0,5 \times 24}{\sqrt{24 \times 0,5 \times 0,5}} = -3,06186$

Rejeitar H_0 se $|Z_{Calc}| \geq z_{1-\frac{\alpha}{2}}$

Para $\alpha = 0,05$,

$\frac{\alpha}{2} = 0,025$

$1 - \frac{\alpha}{2} = 0,975$

Consultando a tabela da Normal vemos que $z_{0,975} = 1,96$

Calculando o valor-p temos que.

$Valor - p = 2(1 - \Phi(3,06)) = 2(1 - 0,99889) = 0,00222$

Donde se conclui que não se deve rejeitar H_0. O valor-p calculado pela aproximação à Normal é um valor muito aproximado ao estimado com a distribuição Binomial.

3.5 Teste de Wilcoxon

O teste de Wilcoxon tem a vantagem de ser mais potente do que o teste dos sinais, isto é, é menor a probabilidade de se cometer o erro de não rejeitar H_0 sendo H_0 falsa.

Quando se pretende estudar uma hipótese sobre a mediana e se considera como pressuposto a simetria da distribuição dos valores, o teste de Wilcoxon representa uma melhoria em relação ao teste dos sinais pois não despreza a informação dada pela ordem das diferenças.

Para testar H_0: $m = m_0$ contra a alternativa H_1: $m \neq m_0$, dada uma amostra de uma população com função de distribuição $F(x)$ desconhecida, mas simétrica, obtêem-se as diferenças

$d_i = x_i - m_0$, $i = 1, 2, ..., N$

Estas deverão distribuir-se de forma simétrica em torno de 0. Ou seja, observar-se-ão diferenças positivas e negativas com valores absolutos da mesma ordem de grandeza, e em número aproximadamente igual.

A avaliação relativa da magnitude das diferenças d_i pode ser efectuada ordenando de forma crescente, de 1 a N, os seus valores absolutos $|d_i|$ e atribuindo a cada um destes o respectivo número de ordem (em inglês esta ordenação designa-se por "rank", de onde vem o nome do teste), com o sinal negativo ou positivo, consoante d_i sejam negativo ou positivo.

Se a população for simétrica em torno me m_0 e H_0 for verdadeira, a soma dos números de ordem referentes às diferenças d_i negativas deverá ser aprosimadamente igual à soma dos números de ordem referentes às diferenças d_i positivas. Uma situação contrária a esta beneficia uma das hipóteses alternativas. Por exemplo, se a soma dos números de ordem relativos às diferenças positivas for muito maior do que a soma dos números de ordem das diferenças negativas, então a hipótese alternativa $H_1 : m \neq m_0$ tornar-se-á plausível. A estatística de teste de Wilcoxon é baseada justamente na propriedade que acaba de ser enunciada.

Os passos para o cálculo da estatística de teste de Wilcoxon são:

• Calculam-se as diferenças $d_i = x_i - m_0$;

• Ordenam-se as diferenças d_i por ordem crescente dos respectivos valores absolutos $|d_i|$;

• Atribui-se um número de ordem sequencialmente a cada $|d_i|$; os números de ordem referentes a d_i são precedidos do sinal "+"; os números de ordem referentes a d_i negativos são precedidos do sinal "-";

• Quando o valor absoluto de duas ou mais diferenças é o mesmo (isto é, quando existem "empates"ou "ties"), o número de ordem atribuído a cada uma dessas diferenças com o mesmo valor absoluto $|d_i|$ é a média aritmética dos números de ordem que tais observações receberiam se não estivessem empatadas. Sejam por exemplo as diferenças ordenadas a sequência 1, 3, -3, 5, 7, -7, -7, 8; os respectivos números de ordem seriam 1, 2.5, 2.5, 4, 6, 6, 6, 8:

• Quando existem zeros, isto é, quando $d_i = 0$, estes valores devem ignorar-se, e consequentemente, reduzir o tamnho da amostra em tantas unidades, tanto os zeros que existam;

• Calcula-se a estatística de teste, geralmente designada por T, e que resulta da soma dos números de ordem "positivos"(caso em que a estatística de teste se representa por T_+) ou dos números de ordem "negativos"(a estatística de teste é representada por T_-).

Note-se que a estatística de teste toma sempre um valor não negativo, e para uma amostra de tamanho N a soma de todos os números de ordem é:

$$T_{+} + T_{-} = \frac{N \times (N+1)}{2} \tag{3.9}$$

Se a hipótese nula é verdadeira, as distribuições de T_{+} e T_{-} são simétricas em torno do valor esperado:

$$E(T) = \frac{N \times (N+1)}{4} \tag{3.10}$$

de modo que seria indiferente usar T_{+} ou T_{-} como estatística de teste. Contudo, por comodidade, em cada uma das seguintes situações de hipótese alternativa, é usual considerar:

Hipótese nula	Hipótese alternativa	Estatística de teste usual
$m = m_0$	$m < m_0$	T_{+}
	$m \neq m_0$	Mínimo de T_{+} ou T_{-}
	$m > m_0$	T_{-}

Existem tabelas com os valores críticos de T_{+} ou T_{-} para decidir acerca da significância do teste. Para amostras com $N \geq 15$ demonstra-se que a distribuição amostral de T_{+} (ou T_{-}) se aproxima da distribuição normal de parâmetros:

- Valor médio:

$$\mu_{T_{+}} = \frac{N \times (N+1)}{4} \tag{3.11}$$

- Variância:

$$\sigma^2_{T_+} = \frac{N \times (N+1) \times (2N+1)}{24} \tag{3.12}$$

Se existem "empates"a variância deve ser corrigida, sendo neste caso a expressão para cálculo da variância:

- Variância: $\sigma^2_{T_+} = \frac{N\times(N+1)\times(2N+1)}{24} - \frac{\sum u_i^3 - \sum u_i}{48}$

em que u_i representa o número de "empates"no i-ésimo grupo de observações iguais.

Quando se faz a aproximação à função de distribuição normal, a estatística de teste é:

$$Z = \frac{T_+ - \mu_{T_+}}{\sigma_{T_+}} = \frac{T_+ - \frac{N\times(N+1)}{4}}{\sqrt{\frac{N\times(N+1)\times(2N+1)}{24}}} \sim N(0,1) \tag{3.13}$$

3.5.1 Exemplo 1

As pontuações totais (de 0 a 200) obtidas por 16 alunos, escolhidos ao acaso, num teste de Matemática foram as seguintes:

97, 140, 58, 60, 100, 31, 80, 27, 108, 73, 95, 58, 76, 69, 121, 117

Vamos testar a hipótese da mediana da população ser igual a 80, resorrendo ao teste de Wilcoxon.

Observando o seguinte quadro:

x_i	$z_i = x_i - 80$	$\|z_i\|$	$\|z_i\|ord$	r_i	S_i	iS_i	r_i^2
97	+17	17	0				
140	+60	60	4(-)	1	0	0	1
58	-22	22	7(-)	2	0	0	4
60	-20	20	11(-)	3	0	0	9
100	+20	20	15(+)	4	1	4	16
31	-49	49	17(+)	5	1	5	25
80	0	0	20(-)	6.5	0	0	42.25
27	-59	59	20(+)	6.5	1	6.5	42.25
108	+28	28	22(-)	8.5	0	0	72.25
73	-7	7	22(+)	8.5	0	0	72.25
95	+15	15	28(+)	10	1	10	100
58	-22	22	37(+)	11	1	11	121
76	-4	4	41(+)	12	1	12	144
69	-11	11	49(-)	13	0	0	169
121	+41	41	53(-)	14	0	0	196
117	+37	37	60(+)	15	1	15	225
					S=7	T=63.5	1239

verifica-se:

• Uma situação de valor nulo para Z_i. A respectiva observação é eliminada, passando a ter-se n=15.

• Duas situações de empate, atribuindo-se a cada uma delas a média das ordens em causa (6 e 7 substituídas por 6.5; 8 e 9 por 8.5).

Aplicando 3.7 e 3.8, obtêm-se o valor médio e a variância da estatística T de

Wilcoxon:

$E(T)=\frac{15\times16}{4}$; $Var(T)=\frac{1239}{4}=309.75$

O valor observado de T é então,

$T_{Calc}=\frac{63.5-60}{\sqrt{309.75}}=0.199$,

a que corresponde um valor-p igual a 0.4212. Desta forma, não se rejeita a hipótese formulada.

3.6 Teste de aleatorização das iterações

Para comprovar a propriedade de aleatoriedade de uma amostra utilizamos o teste de aleatorização, que faz uso da análise das sequências de símbolos idênticos.

Este teste, basicamente, verifica o número de iterações existentes na amostra; se o número de iterações é muito grande ou muito pequeno sugere-se falta de aleatoriedade.

Exige-se ao menos que os dados sigam uma escala nominal e que eles possam ser divididos em duas categorias.

Vejamos como se utiliza o teste:

• Seja n_1 o número de elementos da categoria 1, n_2 o número de elementos da categoria 2 e $N=n_1+n_2$

• Se n_1 e $n_2<20$ verificamos o número R de iterações, obtemos os limites inferior e superior, que definem o número aceitável de iterações em caso de aleatoriedade.

Se n_1 ou n_2 >20 usamos a seguinte aproximação:

$$\mu_R = \frac{2n_1n_2}{n_1+n_2} + 1 \tag{3.14}$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} \tag{3.15}$$

$$Z = \frac{R - \mu_R}{\sigma_R} \tag{3.16}$$

R segue uma distribuição Normal, $N(\mu_R, \sigma_R)$, e usando o Teorema do Limite Central, $Z \sim N(0,1)$ pelo que, fazemos posteriormente uso da tabela da distribuição normal padrão.

3.6.1 Exemplo 1

24 crianças foram avaliadas em relação a um índice de agressividade e em seguida converteram-se os dados em sinais positivos (+) e negativos (-) dependendo se o índice estava acima ou abaixo da mediana do grupo. Deseja-se verificar a aleatoriedade das pontuações de agressividade com relação à ordem em que foram obtidos.

H_0: As pontuações de agressividade ocorrem de forma aleatória

H_1: As pontuações de agressividade não ocorrem de forma aleatória

Sendo N=24, n_1=12 e n_2=12 temos a seguinte sequência de sinais:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
+	-	+	+	+	+	-	+	+	+	-	-	-	-	+	-	-	+	+	+	-	-	-	-
1	2	3	3	3	3	4	5	5	5	6	6	6	6	7	8	8	9	9	9	10	10	10	10

Conclusão: Consultando uma tabela, esta indica para $n_1 = n_2 = 12$ os limites 7 e 19, portanto $r = 10$ iterações não se encontra na região de rejeição. Não rejeitaremos H_0. Os dados parecem ter sido gerados de forma aleatória.

3.6.2 Exemplo 2

Deseja-se verificar se a disposição de homens e mulheres numa fila de cinema se dá de forma aleatória.

H_0: A ordem dos sexos na fila é aleatória

H_1: A ordem não é aleatória

Foram observados 30 homens e 20 mulheres, que forneceram os seguintes resultados:

$N = 30 \qquad n_1 = 30 \qquad n_2 = 20 \qquad r = 35$

$$\mu_R = \frac{2 \times 30 \times 20}{50} + 1 = 25 \tag{3.17}$$

$$\sigma_R = \sqrt{\frac{2 \times 30 \times 20(2 \times 30 \times 20 - 30 - 20)}{(30 + 20)^2(30 + 20 - 1)}} = 3,356 \tag{3.18}$$

$$Z = \frac{35 - 25}{3,356} = 2,98 \tag{3.19}$$

Conclusão: A Tabela da Normal, pois $Z \sim N(0, 1)$, mostra que a probabilidade de ocorrência, sob H_0, de Z≥2,98 é $p = 2 \times 0,0014 = 0,0028$ (a probabilidade é duas vezes a indicada na tabela porque se trata de uma prova bilateral). Como a probabilidade p=0,0028, associada à ocorrência observada, é inferior ao nível de significância 0,05, a decisão será rejeitar a hipótese H_0. Isto é, concluimos que, naquela fila, a ordem dos homens e das mulheres não foi aleatória.

Capítulo 4

Tabelas de contingência

Um processo de organizar a informação correspondente a dados bivariados é utilizando uma tabela de contingência.

De uma maneira geral, uma tabela de contingência é uma forma de organizar dados, quer de tipo qualitativo, quer de tipo quantitativo, especialmente quando são de tipo bivariado, isto é, podem ser classificados segundo dois critérios.

"No século XIII frei Roger Bacon, um cientista admirável (dos primeiros a libertar-se das limitações da escolástica, a questionar a autoridade dos clássicos e a considerar que a experiência é, em última análise, o que confirma ou infirma as nossas hipóteses científicas, defendendo simultaneamente a vantagem de usar modelos matemáticos no estudo da natureza), advogava o interesse das tabelas de presença e de ausência - aquilo a que hoje chamamos tabelas de contingência. (...)

As tabelas de contingência são uma apresentação tabular de contagens de efectivos de classes."In [2]".

Vejamos agora o que são dados categorizados: os indivíduos de uma dada população podem ser classificados em categorias ou classes, segundo determinado critério. Tal classificação consiste em detectar a categoria a que cada indivíduo pertence, devendo as categorias serem exaustivas e mutuamente exclusivas, isto é, qualquer indivíduo pertencer a uma e uma só categoria.

Para estudar dados categorizados procedemos ao estudo das frequências absolutas de cada categoria. Assim, perante uma amostra, efectuamos a contagem do número de observações em cada categoria, ou seja, calculamos as suas frequências observadas, organizadas, usualmente, em tabelas de contingência. Considerando A e B duas características (variáveis nominais) de uma determinada população, subdivididas em r e c categorias designadas por A_1,..., A_r e B_1, ..., B_c, respectivamente, a tabela de contingência que resulta da classificação de n observações ou indivíduos nas r×c categorias cruzadas tem a forma da seguinte tabela, onde n, a dimensão da amostra, se supõe fixa.

	B_1 B_2 ...B_j... B_c	Total Marginal
A_1	n_{11} n_{12} ... n_{1c}	$n_{1.}$
A_2	n_{21} n_{22} ... n_{2c}	$n_{2.}$
.	. . . .	.
A_i	n_{ij}	.
.	. . . .	.
.	. . . .	.
A_r	n_{r1} n_{r2} ... n_{rc}	$n_{r.}$
Total Marginal	$n_{.1}$ $n_{.2}$ ... $n_{.c}$	n

Nota: Esta é a forma geral de uma tabela de contingência r×c. A expressão n_{ij}, $i = 1, ..., r; j = 1, ..., c$ representa o número de observações pertencentes à categoria A_i de A e à categoria B_j de B, $n_{i.}$ representa o total de observações na categoria A_i da variável A e $n_{.j}$ o total de observações na categoria B_j da variável B, estes últimos designados por totais marginais.

Uma tabela de contingência é uma tabela de frequências que apresenta um conjunto de dados que foram classificados simultaneamente segundo duas (bidimensional) ou mais variáveis (multidimensional). As tabelas de contingência têm pelo menos, duas linhas e duas colunas.

As tabelas de contingência também se utilizam no caso que se pretende verificar se determinada característica categorizada se distribui de forma semelhante pelas diferentes categorias de duas ou mais populações, ou seja, quando se pretende averiguar se duas ou mais populações são homogéneas no que diz respeito à distribuição de determinada característica.

A palavra "contingência"pode estar associada a algo que não prevemos sobre uma pessoa ou entidade. Assim, para resolvermos este problema, quantificamos este contingente de pessoas ou entidades numa tabela, "Tabela de Contingência". O principal objectivo na construção deste tipo de tabela é que uma variável não seja influenciada pela outra, entretanto, em muitos casos esta influência ocorre. Este tipo de influência pode ser vista de dois modos.

A primeira é quando variáveis classificadoras causam uma dependência nos grupos ou populações. Para este tipo de influência, podemos citar um grupo de pessoas com doenças psiquiátricas, em que são classificadas como "actividade retardada"e "actividade não retardada"e que cada grupo pode ser classificado em três categorias, "desordem afectiva", "esquizofrenia"e "neurose". Para este tipo de aplicação queremos testar se o tipo de actividade sofre alguma influência das categorias de doenças psiquiátricas, ou seja, queremos testar se os grupos têm independência em relação as actividades retardadas ou não.

A segunda é usada quando pretendemos saber se os dados associados às categorias de uma das variáveis se comporta de modo homogéneo ou similar nas diversas classes ou populações definidas pelas categorias da outra variável classificadora. Para este tipo de influência podemos citar a eficácia de um medicamento, para isto seleccionamos 100 doentes, dentre eles 50 são medicados e os outros 50 recebem um placebo, neste estudo foram verificados os efeitos secundários presentes ou ausentes. Para estes efeitos podemos fazer uma classificação em diversos modos, como por exemplo, se o indivíduo teve ou não uma melhora na doença, ou ainda se obteve uma reação ao tipo de medicamento. Nesta aplicação, queremos testar se o grupo de indivíduos medicados e o grupo de indivíduos que usaram placebo têm comportamentos similares em relação a esses efeitos secundários, isto é, se as populações são homogêneas.

Através das tabelas de contingência é possível classificar os membros de uma população ou grupos dos mais diversos modos, tanto para o teste de homogeneidade, quanto para o teste de independência. Por exemplo, as pessoas podem ser classificadas quanto ao seu sexo, podem ser classificadas em solteiras ou casadas (classificações dicotómicas), classificadas em canhotas, destras ou ambidestras etc. A classificação pode ser feita sobre informações de dados contínuos, basta considerarmos classes de valores desses dados e depois classificarmos relativamente à classe a que pertencem.

De um modo geral, uma tabela de contingência é uma representação dos dados, sejam eles qualitativos ou quantitativos. Quando classificamos de modo bivariado, eles podem ser classificados segundo dois critérios. Caso classificarmos segundo mais de dois critérios estamos no caso multivariado.

4.1 Testes do Qui-Quadrado em tabelas de contingência

4.1.1 Teste de independência

O teste de independência do Qui-Quadrado permite verificar a independência entre duas variáveis de qualquer tipo que se apresentem agrupadas numa tabela de contingência.

Este teste não deve ser utilizado se mais do que 20% das frequências esperadas sob a hipótese da independência forem inferiores a 5 ou se algumas delas fou igual a 0.

As hipóteses em teste são as seguintes:

H_0: As variáveis são independentes;

vs.

H_1: As variáveis não são independentes.

Notemos que a hipótese H_1 não tem nenhuma indicação sobre o tipo de associação entre as variáveis.

A estatística de teste, é a variável:

$$X^2 = \sum_{i=1}^{r} \sum_{i=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{4.1}$$

Esta variável tem distribuição aproximadamente Qui-Quadrado com $gl = (r - 1)(c - 1)$, onde r é o número de linhas, c o número de colunas da tabela de contingência, no caso de H_0 ser verdadeira.

Para obter a frequência esperada (E_{ij}) em cada célula, multiplicamos os dois totais marginais comuns a uma determinada célula e dividimos o produto por N, total de casos.

Depois de encontrar o valor crítico na tabela do Qui-Quadrado, se $\chi^2_{calculado}$ for menor que o $\chi^2_{tabelado}$, rejeita-se H_0. Onde:

O_{ij} representa o número de casos observados na linha i da coluna j

E_{ij} representa o número de casos esperados, sob H_0 na linha i da coluna j

$\sum_{i=1}^{r}\sum_{i=1}^{k}$ indica o somatório sobre todas as r linhas e todas as k colunas, ou seja, sobre todas as células da tabela.

4.1.2 Exemplo

Suponha-se que desejamos comprovar se há diferenças de qualidades de liderança entre pessoas altas e pessoas baixas. A tabela seguinte exibe as frequências em que se classificaram 43 "baixos"e 52 "altos"quanto ao nível de liderança.

A hipótese nula é que a altura é independente da classificação ao nível da liderança.

No quadro seguinte apresenta-se os resultados:

	Baixo	Alto	Total
Líder	12	32	44
Liderado	22	14	36
Não classificável	9	6	15
Total	43	52	95

Determinemos então a frequência esperada em cada célula, utilizando o que atrás foi descrito. Obtemos então a seguinte tabela:

	Baixo	Alto	Total
Líder	19.9	24.1	44
Liderado	16.3	19.7	36
Não classificável	6.8	8.2	15
Total	43	52	95

Calculemos então χ^2 para os dados apresentados:

$$\chi^2_{Calc} = \frac{(12-19.9)^2}{19.9} + \frac{(32-24.1)^2}{24.1} + \frac{(22-16.3)^2}{16.3} + \frac{(14-19.7)^2}{19.7} + \frac{(9-6.8)^2}{6.8} + \frac{(6-8.2)^2}{8.2} = 10.67$$

Como $\chi^2_{Calc} > \chi^2_{(0,99;2)}$ rejeitamos H_0. Conclui-se que a altura não é independente da classificação ao nível da liderança.

4.1.3 Teste de homogeneidade

O teste de Qui-Quadrado de homogeneidade pode ser utilizado para comparar as populações em termos das proporções de elementos de determinada característica em estudo.

Este é usado quando pretendemos saber se os dados associados aos atributos de uma das variáveis se comporta de modo homogéneo ou similar nas diversas classes ou subpopulações definidas pelos atributos da outra variável estatística.

As hipóteses a testar são as seguintes:

H_0: Existe homogeneidade entre as subpopulações

vs.

H_1: Não existe homegeneidade entre as subpopulações.

Este teste constrói-se de maneira idêntica ao teste de Qui-Quadrado, sendo as hipóteses a testar referidas anteriormente.

4.2 Teste exacto de Fisher

O teste de exacto de Fisher constitui uma técnica não paramétrica muito útil para analisar dados discretos, quando a dimensão das amostras independentes é pequena e consiste em determinar a probabilidade exacta de ocorrência de uma frequência observada, ou de valores mais extremos.

Este teste exige que:

- Tenha duas populações;

- Cada população seja dividida em duas categorias exclusivas, categorias estas que têm de ser as mesmas para as duas populações;

- Duas classes mutuamente exclusivas, ou seja, cada elemento de uma população irá pertencer a exactamente uma das categorias.

Observa-se, em cada amostra, a quantidade de elementos pertencentes a cada categoria. O teste exacto de Fisher visa comprovar se as quantidades de ocorrências nestas categorias são ou não equivalentes nas duas populações.

Consideremos a definição de duas amostras I e II, agrupadas em classes - e +.

	-	+	
I	A	B	A+B
II	C	D	C+D
	A+C	B+D	N

A probabilidade p de ocorrência das frequências observadas nas células acima, faz-se com o uso da distribuição hipergeométrica, ou seja:

$$p = \frac{\binom{A+C}{A} * \binom{B+D}{B}}{\binom{N}{A+B}} \tag{4.2}$$

ou da mesma forma

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!} \tag{4.3}$$

Como a hipótese deseja testar a probabilidade de ocorrência de uma situação mais extrema, devemos calcular as probabilidades referentes às frequências observadas e das demais situações concretas.

Quando o valor esperado nalguma célula da tabela é menor que 5, não se usa o teste do Qui-Quadrado. A alternativa é usar o teste exacto de Fisher.

De um modo geral, usa-se o teste exacto de Fisher quando:

- o valor de $N < 20$;

- $20 < N < 40$ e a menor frequência esperada for menor que 5.

As hipóteses a testar neste teste são idênticas às hipóteses no teste do Qui-Quadrado.

A estatística de teste neste caso tem uma distribuição hipergeométrica.

No teste exacto de Fisher o valor-p tem também uma distribuição hipergeométrica.

Se a soma das probabilidades calculadas como descrito em cima for inferior ao nível de significância que escolhermos para o teste, devemos rejeitar H_0.

4.2.1 Exemplo

De uma maneira geral, os doentes psiquiátricos podem ser classificados em psicóticos e neuróticos. Um psiquiatra realiza um estudo sobre os sintomas suicidas em duas amostras de 20 doentes de cada grupo. A nossa hipótese é que a proporção de psicóticos com simtomas suicidas é igual à proporção de neuróticos com estes sintomas (num teste de independência, a hipótese nula seria, a presença ou ausência de sintomas suicidas é independente do tipo de doente envolvido). Assim, temos os dados resumidos na tabela seguinte:

	Psicótico	Neurótico	Total
Presente	2	6	8
Ausente	18	14	32
Total	20	20	40

Utilizando a expressão 3.2 O resultado obtido é o seguinte:

$P = P_2 + P_1 + P_0 = 0,095760 + 0,020160 + 0,001638 = 0,117558$

Este valor dá-nos a probabilidade de observar que, entre os 8 doentes com sintomas suicidas, 2 ou menos são psicóticos, quando a hipótese de igualdade da proporção de psicóticos e neuróticos com sintomas suicidas é verdadeira. Verificamos que a probabilidade da discrepância maior ou igual do que a observada ter ocorrido, é de 0,117558, que é consideravelmente elevada. Logo, a proporção de psicóticos e neuróticos são homogéneos no que diz respeito aos sintomas suicidas.

Capítulo 5

Testes para o caso de duas amostras independentes

Pode ser impossível delinear um projecto que utilize pares de dados, talvez por desconhecimento, de variáveis úteis que possam formar pares, ou pela impossibilidade de obter resultados adequados de alguma variável de reconhecida importância, ou, porque simplesmente não se dispõe de "pares" adequados.

Quando a utilização de duas amostras não independentes não é a melhor para o estudo que se quer fazer, podemos utilizar duas amostras independentes. Em tais estudos, as duas amostras podem ser obtidas por um dos dois métodos:

- podem ser extraídas aleatóriamente de duas populações

- podem decorrer da atribuição aleatória de dois tratamentos aos membros de uma amostra.

Em nenhum desses casos se exige que as amostras tenham a mesma dimensão.

Os testes de seguida apresentados, servem, de um modo geral, para determinar se as diferenças nas amostras constituem evidência convincente de uma diferença nos processos, ou tratamentos, aplicados a elas. A principal diferença é de que as amostras são independentes.

5.1 Teste U de Mann-Whitney

O teste U de Mann-Whitney (1947) pode-se aplicar para comprovar se dois grupos independentes foram ou não extraidos da mesma população. Trata-se de um teste não-paramétrico poderoso, e constitui uma alternativa extremamente útil quando se deseja evitar suposições exigidas pelo teste paramétrico t.

O objectivo deste teste é comprovar se dois grupos independentes foram ou não extraidos duma população com a mesma mediana. Para isso as amostras devem ser independentes e aleatórias: uma extraida duma população com mediana não conhecida M_1 e outra extraida de outra população com mediana desconhecida M_2.

A hipótese a comprovar é ver se as populações têm a mesma mediana, sendo a alternativa, as medianas serem diferentes ou uma maior do que a outra.

Vamos então ver como se aplica o teste U de Mann-Whitney:

. Determinar os valores n_1 (número de casos no menor dos dois grupos independentes) e n_2 (número de casos no maior grupo);

. Dispor em conjunto os valores dos dois grupos, ordenando-os de forma ascendente;

. Atribuir postos aos valores, em caso de empate, faz-se a média dos postos correndentes;

. Para determinar U basta recorrer ao seguinte:

$$U = min(U_1; U_2) \tag{5.1}$$

Sendo:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{5.2}$$

e

$$U_2 = n_1 n_2 - U_1 \tag{5.3}$$

com R_1 = soma das posições atribuidos à amostra 1;

. O método para determinar a significância do valor depende de n_2:

i) Se $n_2 \leq 8$ utiliza-se uma tabela que dá a probabilidade exacta associada a um valor tão pequeno quanto o valor de U. Para uma prova bilateral basta duplicar o valor obtido na tabela. Caso o valor de U não conste na tabela, deve ser interpretado como $U' = n_1 n_2 - U$

ii) Se $9 \leq n_2 \leq 20$, é utilizada uma outra tabela que dá os valores críticos de U para níveis de significância de 0,001, 0,01, 0,025 e 0,05 para um teste unilateral, duplicando estes valores para uma prova bilateral. Caso o valor observado de U seja maior que $\frac{n_1 n_2}{2}$ deve ser interpretado como U' descrito na alínea anterior;

iii) Se $n_2 > 20$, a probabilidade deve ser calculada através de uma aproximação à distribuição normal, através do valor z que nos é dado da seguinte maneira:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \tag{5.4}$$

Caso ocorram empates, em grandes amostras, a expressão utilizada será:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{N(N-1)}(\frac{(N^3-N)}{12} - \sum T)}} \tag{5.5}$$

onde $N = n_1 + n_2$ e $T = \frac{t^3-t}{12}$ sendo t o número de observações empatadas para uma dada posição.

Se o valor observado de U tem probabilidade associada não superior a α, rejeita-se a hipótese nula.

5.1.1 Exemplo

Numa disciplina de um curso universitário, onde se encontram inscritos alunos de dois cursos, registaram-se as seguintes classificações num dos exames:

Curso A	10.5	16.5	11	9.8	17.1	1.5	14.8	9.9	9.8	10.3	8.7
Curso B	11.4	12.9	10.1	7.9	8.8	12.8					

O que se pode concluir acerca das médias das ordens das classificações?

Formulemos as hipóteses:

H_O : Não há diferenças entre as médias das ordens das notas dos alunos do curso A e do curso B

vs.

H_1 : Há diferenças entre as médias das ordens (teste bilateral)

Após a contagem do número de casos em ambas as amostras temos:

$n_1 = 6$ e $n_2 = 11$

Calculemos U:

1.5	7.9	8.7	8.8	9.8	9.8	9.9	10.1	10.3	10.5	11	11.4	12.8	12.9	14.8	16.5	17.1
A	B	A	B	A	A	A	B	A	A	A	B	B	B	A	A	A
1	2	3	4	5.5	5.5	7	8	9	10	11	12	13	14	15	16	17

$U_1 = 6 \times 11 + \frac{6\times(6+1)}{2} - (2 + 4 + 8 + 12 + 13 + 14) = 34$

$U_2 = 6 \times 11 - 34 = 32$

$U = min(34; 32) = 32$

Como $9 \leq n_2 \leq 20$ temos que:

Para $n_1 = 6$, $n_2 = 11$ e $\alpha = 0.05$ (bilateral), temos:

$U_{tabelado} = 13$ (valor de tabela)

Como $U_{tabelado} < U_{calculado}$, podemos concluir que as duas amostras provêem de populações com a mesma média.

5.2 Teste de Moses para reacções extremas

Este teste aplica-se quando existe uma suspeita de que uma determinada condição experimental afectou um grupo de indivíduos e, de forma oposta, outro grupo.

É indicado quando é previsto que um dos grupos tenha valores altos, e o outro valores baixos.

O teste considera dois grupos (amostras) independentes, um grupo de controlo (C) e um grupo experimental (E).

Este teste exige que a escala seja, pelo menos, ordinal.

Embora o teste de Moses se destine especificamente ao tipo de dados mencionado anteriormente, é também aplicável quando se prevê que um grupo tenha resultado alto, e o outro grupo resultado baixo. Todavia, Moses salienta que, em tais casos, um teste baseado em medianas é mais eficiente do que este, por exemplo o teste U de Mann-Witney.

As hipóteses a considerar são então as seguintes:

H_0: Não há diferença entre o grupo C e o grupo E;

vs.

H_1: Há diferença entre os dois grupos.

Vejamos então o que fazer para se aplicar o teste de Moses:

. Sejam n_C e n_E os números de casos de controlo e experimentais, respectivamente;

. Colocar os dados dos dois grupos numa lista, dispondo-os em ordens e conservando a identidade do grupo em cada posição;

. Especificar o valor de h, número pequeno arbitrário;

. Dispôr os dados por ordem, numa única série conservando a identidade do grupo em cada posição;

. Determinar o valor de S_h, âmbito ou abrangência dos postos de controlo após eliminar os h postos mais extremos em cada extremidade da lista. Ou seja,

$$S_h = C_2 - C_1 + 1 \tag{5.6}$$

onde C_2 é o posto correspondente ao último grupo de controlo, depois de retirado h valores de controlo e C_1 corresponde ao primeiro posto do grupo de controlo, retirando h valores de controlo;

. Determinar o valor de g, excesso do valor observado de S_h sobre $n_C - 2h$, ou seja,

$$g = S_h - (n_C - 2h) \tag{5.7}$$

. Determinar valor de p pela fórmula:

$$p(S_h \leq n_C - 2h + g) = \frac{\sum_{i=0}^{g} \binom{i + n_C - 2h - 2}{i} \binom{n_E + 2h + 1 - i}{n_E - i}}{\binom{n_C + n_E}{n_C}} \tag{5.8}$$

Em caso de ocorrência de empates entre grupos, considerar esses empates de todos os modos possíveis e determinar o valor-p para cada um deles. A média desses valores é então utilizada para a decisão:

. Se o valor-p não for superior a α, rejeita-se H_0.

5.2.1 Exemplo

Num estudo efectuado para avaliar o grau de medo, perante ratos, escolheu-se dois grupos de indivíduos. O grupo controlo (C), constituído por 7 indivíduos, que trabalham diariamente com ratos e o grupo experimental (E), formado por 6 indivíduos, que têm dificuldade em controlar o medo, quando estão próximos de ratos. Os indivíduos dos grupos C e E estiveram em contacto com ratos durante 10 minutos e o grau de medo, numa escala de 0 a 20, onde o grau 20 significa que a pessoa tem pavor de ratos, foram anotados e são mostrados, para cada grupo, na tabela seguinte:

Grupo C	6	5	10	7	12	3	8
Grupo E	0	4	11	18	9	19	

Será que as duas amostras provêm da mesma população?

Formulemos as hipóteses:

H_0: Não há diferenças entre o grupo C e o grupo E;

vs.

H_1: Há diferença entre os dois grupos.

Vamos dividir em dois casos: o primeiro com $h = 0$ e o segundo com $h = 1$.

Disponhamos, então, os valores em postos, conservando o grupo:

1º caso:

Posto	1	2	3	4	5	6	7	8	9	10	11	12	13
Grupo	E	C	E	C	C	C	C	E	C	E	C	E	E

$S_h = 11 - 2 + 1 = 10$

Vamos determinar o valor de g, com $S_h = 10$ e $n_c = 7$:

$g = 10 - (7 - 2 \times 0) = 3$

Então utilizando a fórmula 5.9:

$$p(s_h \leq 10) = \frac{\sum_{i=0}^{3} \binom{i+5}{i}\binom{7-i}{6-i}}{\binom{13}{7}} = 0,2168$$

2º caso:

Posto	1	2	3	4	5	6	7	8	9	10	11	12	13
Grupo	E	C	E	C	C	C	C	E	C	E	C	E	E

$S_h = 9 - 4 + 1 = 6$

Vamos determinar o valor de g, com $S_h = 6$ e $n_c = 7$:

$g = 6 - (7 - 2 \times 1) = 1$

Então utilizando a fórmula 3.13:

$$p(s_h \leq 6) = \frac{\sum_{i=0}^{1} \binom{i+3}{i}\binom{9-i}{6-i}}{\binom{13}{7}} = 0,1795$$

Sendo $\alpha = 0,05$, concluimos que, para qualquer um dos casos, não existe diferenças entre os grupos C e E, sendo assim, as amostras provêm da mesma população.

Capítulo 6

Testes para o caso de duas amostras emparelhadas

Empregam-se testes para duas amostras emparelhadas quando queremos determinar, para uma mesma situação, se duas abordagens, tratamentos ou métodos são diferentes ou se um é melhor que o outro.

O método pode consistir numa diversidade de situações ou condições: treino de um atleta, modificação nas condições de habitação, alterações climáticas, etc.

6.1 Teste de McNemar

O teste desenvolvido por McNemar é usado para analisar a eficiência de determinada técnica, isto é, tem como objectivo avaliar a eficência de situações "*antes*" e "*depois*", em que cada indivíduo é utilizado como o seu próprio controlo. Utiliza-se a medição em escala nominal para avaliar alterações da situação "após" em relação à situação "antes".

Para comprovar a significância de qualquer mudança observada, por este método, constói-se uma tabela de frequências de quatro células para representar o primeiro e o segundo conjunto de reacções dos mesmos indivíduos. Os sinas de "+" e "-" utilizam-se para indicar diferentes reacções.

As hipóteses a considerar são as seguintes:

H_0: Não existe diferença antes e depois do tratamento

vs.

H_1: Existe diferença antes e depois do tratamento

Antes \ Depois	-	+
+	A	B
-	C	D

Note-se que os casos que acusam modificações entre a primeira e a segunda reacção aparecem nas células A e D. Um indivíduo é localizado na célula A passou de "+"para "-"; e na célula D passou de "-"para "+". Na ausência de modificação, o indivíduo é classificado na célula B ou na célula C, ou seja, as células A e D são consideradas células de mudança, enquanto as células B e C são células que não mudam de estado. O total de indivíduos que acusam mudança é $A + D$. Sendo assim, a perspectiva, sob a hipótese H_0, seria que $\frac{1}{2}(A + D)$ acusassem modificações num sentido, e $\frac{1}{2}(A + D)$ acusassem modificações noutro sentido. Por outras palavras, $\frac{1}{2}(A + D)$ é a frequência esperada, sob H_0, tanto na célula A como na célula D. Se as frequências esperadas são inferiores a 5, empregamos a prova binomial em substituição à de McNemar. Neste caso, $N = A + D$ e $x = min\{A, D\}$. Caso não se verifique que as frequências são inferiores a 5, McNemar propôs como estatística de teste o valor de χ^2 com a seguinte fórmula:

$$\chi^2 = \frac{(A - D)^2}{A + D} \sim \chi_1^2 \tag{6.1}$$

Nalguns casos, podemos usar uma modificação da estatística 6.1, a correcção torna-se necessária porque uma distribuição contínua, no caso, o qui-quadrado está a ser usada para aproximar uma distribuição discreta. Quando todas as frequências esperadas são pequenas, esta aproximação pode não ser boa.

A correcção de continuidade de Yates é uma tentativa de remover esta fonte de erro. Sendo assim, a expressão 6.1. fica então:

$$\chi_{Calc}^2 = \frac{(|A - D| - 1)^2}{A + D} comgl = 1 \tag{6.2}$$

O grau de significância de qualquer valor observado de χ^2, tal como calculado pela fórmula 6.2 é determinado mediante referência a uma tabela, que dá vários valores críticos de qui-quadrado para graus de liberdade de 1 a 30. Ou seja, se o valor observado de χ^2 é igual a, ou maior do que, o valor exibido na tabela para determinado nível de significância com $gl = 1$, a implicação é que existe efeito significativo nas reacções "antes"e "depois".

Mediante referência a uma tabela, determinamos a probabilidade, sob H_0, associada a um valor tão grande quanto o valor observado de χ^2. Se se tratar de um teste unilateral, basta dividir por dois o valor tabelado. Caso o valor de p exibido pela tabela, não superar α, rejeitamos H_0 em favor da hipótese alternativa.

6.1.1 Exemplo

Uma pesquisa realizada entre donos de automóveis sobre a necessidade do uso do cinto de segurança foi realizada antes e depois de um filme sobre acidentes, onde era focado os benefícios do uso do cinto de segurança.

Dos 70 motoristas entrevistados 20 eram a favor do uso do cinto de segurança antes e continuaram após, 30 eram contra antes e ficaram a favor após, 15 eram contra antes e continuaram cantra após e 5 eram a favor e ficaram contra após. Teste ao nível de 1%, a significância das mudanças.

H_0: A proporção de mudanças de A para B é igual à de B para A, isto é, $P_A = P_B = \frac{1}{2}$

vs.

$H_1 : P_A > P_B$

Antes \ Depois	-	+
+	5	20
-	15	30

Ora utilizando a expressão 6.2 tem-se:

$$\chi^2 = \frac{(|5-30|-1)^2}{5+30} = 16.457$$

Como pode ser visto o resultado encontrado é significativo a 1% ou menos, portanto as mudanças são significativas.

6.2 Teste de Wilcoxon

O teste de Wilcoxon atribui maior ponderação a um par que acusa grande diferença entre as condições, do que a um par em que essa diferença seja pequena.

O teste de Wilcoxon é extremamente útil para os cientistas do comportamento. Com dados sobre o comportamento não são raros os casos em que o observador pode dizer a qual membro de um par é "maior do que"o outro, e dispor as diferenças por ordem do seu valor absoluto. Isto é, o observador pode fazer o julgamento do tipo "maior do que"entre os resultados de qualquer par, bem como fazer esse julgamento em relação às difererenças relativas a dois pares quaisquer. Dispondo destas informações pode-se aplicar o teste de Wilcoxon.

A prova de Wilcoxon de duas amostras é a equivalente não paramétrica ao teste t para duas amostras dependentes. As hipóteses são as mesmas, embora às vezes elas possam ser colocadas em termos de mediana e não da média.

H_0: A diferença entre as médias (ou medianas) populacionais é zero;

vs.

H_1: A diferença entre as médias (ou medianas) não é zero.

A suposição básica por trás deste teste é que as distribuições populacionais são simétricas (médias e medianas idênticas).

Inicialmente calcula-se d_i = diferença do par "i". A seguir atribuir posições a cada d_i, independentemente do sinal. Ao menor d_i atribuir o valor 1; ao próximo 2, etc. A cada ordem atribuir o sinal da diferença, isto é, identificar quais as ordens que decorrem de diferenças negativas e quais de diferenças positivas.

Se as duas classificações são equivalentes, isto é, se H_0 é verdadeira, é de se esperar que algumas das maiores diferenças sejam positivas e outras negativas.

Desta forma, se forem somados as ordens com sinal mais e as ordens com sinal menos, deve-se esperar somas aproximadas iguais.

Se houver diferença entre estas duas somas é sinal de que as duas classificações (ou tratamentos) não se equivalem e deve-se então rejeitar a hipótese nula.

Se as duas amostras foram extraidas da mesma população, então espera-se que as distribuições acumuladas das amostras estejam próximas. Se as distribuições estão "distantes"isto sugere que as amostras provenham de populações distintas e um desvio grande pode levar à rejeição da hipótese h_0.

Eventualmente as pontuações de dois pares serão iguais. Neste caso eles devem ser excluidos da análise e o valor de n deve ser reduzido na mesma quantidade de valores em que a diferença for nula.

Pode ocorrer, ainda, um outro tipo de empate. Duas ou mais diferenças podem ter o mesmo valor absoluto. Neste caso, atribui-se o mesmo posto aos empates. Este posto é a média dos postos que teriam sido atribuidos se as diferenças fossem diferentes.

Por exemplo, se três pares acusam as diferenças -1,-1 e +1, a cada par será atribuido a ordem 2, que é a média entre 1, 2 e 3. O próximo valor, pela ordem, receberia o valor 4, porque já tinham sido utilizadas as ordens 1, 2 e 3.

Pequenas amostras (n<25)

Se $\tau =$ a menor soma das ordens de mesmo sinal (negativos ou positivos) então τ será significativo se não superar o valor dado na tabela, sob determinado nível de significância.

Grandes amostras (n>25)

Neste caso τ (menor soma) é aproximadamente normal com os seguintes parâmetros:

$$\mu_r = \frac{n(n+1)}{4} \tag{6.3}$$

$$\sigma_r = \sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{6.4}$$

Quando se têm pares de observações (X_1, Y_1), ..., (X_2, Y_2), e as diferenças $d_i = X_i - Y_i$ têm distribuição normal, usa-se o teste paramétrico t-Student para comparar as médias de duas amostras emparelhadas. Porém, se as diferenças $d_i = X_i - Y_i$ não se distribuem normalmente, pode usar-se o teste de Wilcoxon sobre as diferenças, desde que estas tenham um comportamento contínuo e simétrico.

Neste caso, o teste de hipóteses é:

$H_0 : \mu_d = \delta_0$

vs.

$H_1 : \mu_d \neq \delta_0$

A estatística de teste é $min(T_+; T_-)$, isto é, o valor mínimo da soma dos números de ordem associados aos valores positivos ou negativos de $\delta_i - \delta_0$.

6.2.1 Exemplo

Existem diversos métodos de estimação do volume de madeira produzido pelas árvores, nomeadamente modelos de estimação baseados no diâmetro basal e modelos de estimação baseados no diâmetro à altura do peito (dap).

Pretende-se comparar um método de estimação baseado no diâmetro basal com outro método baseado no dap. Para tal, os volumes (m^3) de madeira dos mesmas

15 pinheiros foram estimados pelos dois métodos:

Basal	1.06	1.08	1.12	0.98	1.05	0.85	1.06	0.87	1.03	1.1	0.95	0.78	1.23	1.04	0.88
Dap	1.12	0.97	1.15	1.07	0.89	0.98	1.13	0.82	1.15	1.25	0.86	0.83	1.05	0.89	1.02

Como exposto, pretendendo testar se as estimativas pelos dois métodos são idênticas, então a média das diferenças entre as observações será nula, e o teste de hipóteses é:

$H_0 : \mu = 0$

vs.

$H_1 : \mu \neq 0$

em que μ_d é a média das diferenças $d_i = V_{basal} - V_{dap}$

Na tabela seguinte apresentam-se os cálculos do teste:

V_{basal}	V_{dap}	$d_i = V_{basal} - V_{dap}$	$\|d_i\|$	Ordem (+)	Ordem (-)
1.06	1.12	-0.06	0.6		4
1.08	0.97	0.11	0.11	8	
1.12	1.15	-0.03	0.03		1
0.98	1.07	-0.09	0.09		6.5
1.05	0.89	0.16	0.16	14	
0.85	0.98	-0.13	0.13		10
1.06	1.13	-0.07	0.07		5
0.87	0.82	0.05	0.05	2.5	
1.03	1.15	-0.12	0.12		9
1.1	1.25	-0.15	0.15		12.5
0.95	0.86	0.09	0.09	6.5	
0.78	0.83	-0.05	0.05		2.5
1.23	1.05	0.18	0.18	15	
1.04	0.89	0.15	0.15	12.5	
0.88	1.02	-0.14	0.14		11
				$T_+ = 58.5$	$T_- = 61.5$

A fim de calcular a estatística de teste para proceder à decisão do teste, temos em primeiro lugar de fazer a aproximação à função de distribuição normal. Os parâmetros desta aproximação são:

- Média: $\mu_{T_+} = \frac{N.(N+1)}{4} = \frac{15\times16}{4} = 60$
- Variância (note-se que existem três grupo de observações iguais, cada um com 2 observações):

$$\sigma^2_{T_+} = \frac{N.(N+1)(2N+1)}{24} - \frac{\sum u_i^3 - \sum u_i}{48} = \frac{15\times16\times31}{24} - \frac{(2^3+2^3+2^3)-(2+2+2)}{48}$$
$$= 309.625$$

A estatística do teste é então:

$$Z = \frac{T_+ - \mu_{T_+}}{\sigma_{T_+}} = \frac{58.5-60}{\sqrt{309.625}} = -0,0853$$

Para um nível de significância $\alpha = 5\%$, e tratando-se de um teste bilateral, o quantil crítico da distribuição normal N(0,1) é $Z_{0.05} = \pm 1.96$, pelo que se conclui que não há evidência estatística para rejeitar a hipótese nula.

6.3 Teste dos Sinais

O teste dos Sinais tem a sua denominação devida ao facto de utilizar como dados sinais "mais"e "menos", em vez de medidas quantitativas. É particularmente útil nos trabalhos de pesquisa em que é impossível ou inviável a obtenção de uma medida quantitativa, mas é possível estabelecer ordens em relação a cada um dos dois membros de cada par.

O teste dos Sinais é aplicável no caso de duas amostras emparelhadas, quando se deseja determinar se duas condições são diferentes. A única suposição que o teste dos Sinais exige é que a variável em estudo tenha distribuição contínua. O teste não faz qualquer suposição sobre a forma da distribuição das diferenças, nem supõe que todos os indivíduos tenham sido extraídos da mesma população.

Os diferentes pares podem provir de populações diferentes com respeito a várias características. A única exigência é que, dentro de cada par, se tenha conseguido um nivelamento quanto às variáveis extrínsecas importantes.

As hipóteses deste teste são as seguintes:

H_0: O número de sinais "+"é o mesmo de sinais "-",

vs.

H_1: H_0 é falsa.

Este teste é na verdade uma prova Binomail com $p_0 = \frac{1}{2}$.

O teste exige que os pares (X_i, Y_i) sejam mutuamente independentes e a escala de medida seja ordinal.

Caso $N \leq 25$, fazemos uso da Prova Binomial considerando $p_0 = \frac{1}{2}$, sendo N o número de pares, e x o número de sinais que corresponde à menos frequência. Deve-se depois calcular $P \leq x)$.

Quando $N > 25$ utilizamos a aproximação Normal fazendo:

$$z = \frac{x - \frac{N}{2}}{\sqrt{\frac{N}{4}}} = \frac{2x - N}{\sqrt{N}} \tag{6.5}$$

Pode ocorrer por vezes que $X_i = Y_i$, ou seja, não há diferenças entre as pontuações do par i. Neste caso, os empates são eliminados da análise.

Se o valor-p obtido no teste não for superior a α, rejeitamos H_0.

6.3.1 Exemplo

Extraiu-se uma amostra de 100 adultos de uma comunidade e perguntou-se a cada um sobre o tipo de punição a ser aplicado em casos de delinquência juvenil (se mais forte ou mais fraca). De seguida exibiu-se um filme sobre instituições de reabilitação, e posteriormente repetiu-se a pergunta. Os resultados obtidos foram os seguintes:

	Depois	
Antes	-	+
+	59	7
-	8	26

H_0: O filme não produz efeito

vs.

H_1: O filme produz efeito

Utilizamos o teste do sinal por ser uma escala ordinal e temos uma amostra consideravelmente grande. Como se verificaram 15 empates, estes são excluidos da análise. Assim, sob H_0, é de esperar que metade dos restantes 85 entrevistados mudem a sua opinião de + para - e a outra metade de - para +. Assim

$$z = \frac{59-42{,}5}{\sqrt{\frac{85}{4}}} = 3,85$$

Sendo assim, rejeita-se a hipótese H_0, ou seja, o filme teve um efeito muito significativo sobre a atitude dos indivíduos.

Capítulo 7

Testes para o caso de k (k $>$ 2) amostras emparelhadas

Estudemos agora a hipótese de que k (k >2) amostras tenham sido extraidas da mesma população ou de populações idênticas. Quando se trata de comparar três ou mais amostras ou condições numa experiência, é necessário aplicar testes estatísticos que indiquem se há uma diferença geral entre as K amostras, antes de podermos comprovar a significância da diferença entre duas amostras quaisquer.

Há dois processos básicos para comparar k (k > 2) grupos. No primeiro deles, as k (k > 2) amostras de igual tamanho são postas em correspondência de acordo com determinado(s) critério(s) que pode(m) afectar os valores das observações. O segundo plano envolve k (k > 2) amostras aleatórias independentes, uma de cada população.

7.1 Teste de Q de Cochran

O Teste Q de Cochran é uma extensão do teste de McNemar para amostras emparelhadas, que fornece um método para testar as diferenças entre três ou mais conjuntos combinados de frequências ou proporções.

Este teste proporciona um método para comprovar se três ou mais conjuntos correspondentes de frequências ou proporções diferem entre si significativamente. A correspondência pode basear-se em características relevantes dos diferentes indivíduos, ou no facto de os mesmos indivíduos serem observados em condições diferentes. O teste Q de Cochran adapta-se especialmente ao caso em que os dados se apresentam em escala nominal ou sob a forma de informação ordinal dicotomizada.

As hipóteses a considerar são as seguintes:

H_0: "sucessos" ou "insucessos"distribuem-se aleatoriamente pelas linhas e colunas de uma tabela,

vs.

H_1: "sucessos"ou "insucessos não se distribuem aleatoriamente pelas linhas e colunas de uma tabela.

Este teste exige um nível de medida em escala nominal ou ordinal dicotomizada.

Para utilizar este teste procedemos da seguinte forma:

Para dados dicotomizados, atribuimos o valor "1"a cada "sucesso"e o valor "0"a cada "insucesso";

Dispor os dados numa tabela k×N, com k colunas e N linhas, N = número de casos em cada k grupos;

Determinar o valor observado da estatística de teste Q, utilizando a fórmula:

$$Q = \frac{(k-1)[k\sum_{j=1}^{k} G_j^2 - (\sum_{j=1}^{k} G_j)^2}{k\sum_{i=1}^{N} L_i - \sum_{i=1}^{N} L_i^2} \tag{7.1}$$

Onde G_j é a soma dos valores das j colunas;

L_i é a soma dos valores das i linhas

A significância do valor observado de Q pode ser determinada mediante referência à tabela do Qui-Quadrado, pois Q tem distribuição aproximadamente Qui-Quadrado com $gl = k - 1$. Se a probabilidade associada à ocorrência sob H_0 de um valor tão grande quanto um valor observado de Q não supera α, rejeita-se a hipótese H_0.

7.1.1 Exemplo

Um fabricante de sapatos mostra quatro modelos dos seus últimos lançamentos (A, B, C e D) a sete comerciantes que têm lojas de calçado. Para cada modelo as encomendas de cada comprador estão resumidas na tabela seguinte.

Comprador	Modelo			
	A	B	C	D
1	X	X		
2		X	X	
3		X	X	X
4				X
5			X	
6			X	
7			X	

Existem diferenças significativas entre os quatro modelos de sapatos?

Vamos testar o seguinte:

H_0: Os modelos diferem no número de encomendas efectuadas.

vs.

H_1: Os modelos não diferem no número de encomendas efectuadas.

Utilizando a fórmula 7.1 obtemos o seguinte resultado:

$$Q = \frac{3[4(1^2+3^2+5^2+2^2)-11^2]}{4.11-21} = \frac{3{,}35}{23} = 4,565$$

Pela tabela do Qui-Quadrado vemos que:

$0,25 < P(\chi^2_3 \geq 4.565) < 0,1.$

Utilizando o programa Excel verificamos que temos um valor-p de 0,21.

Logo não existem diferenças significativas no número de encomendas efectuadas para cada modelo de sapatos.

7.2 Teste de Friedman

Este teste é útil quando se deseja comprovar a hipótese de que as k amostras emparelhadas provêm da mesma população. Neste tipo de estudo observa-se o mesmo grupo de indivíduos sob cada uma das k condições, ou então formam-se conjuntos de indivíduos homogéneos entre si, e estes são colocados aleatoriamente em cada uma das condições.

As hipóteses a testar são as seguintes:

H_0: As distribuições das k amostras são idênticas

vs.

H_1: As distribuições das k amostras diferem na localização

Os valores são dispostos numa tabela de dupla entrada com k colunas e N linhas.

A estatística de teste (designada por χ^2) é dada pela expressão:

$$\chi^2 = \frac{12}{Nk(k+10} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1) \tag{7.2}$$

onde: N é o número de linhas;

k é o número de colunas

R_j a soma das ordens na coluna.

Esta variável segue uma distribuição Qui-Quadrado com $k-1$ graus de liberdade.

Se a probabilidade obtida por este método não superar α rejeita-se H_0.

7.2.1 Exemplo

Um teste de consumo de combustível envolvendo carros produzidos por três fabricantes foi realizado e os resultados, em quilómetros por litro de combustível estão apresentados na tabela abaixo. Verificar se existem diferenças significativas entre os fabricantes.

Modelo	Fabricante		
	G	**F**	**C**
Pequeno	9.0	11.3	10.6
Médio - 6 cil	9.4	10.9	10.2
Médio - 8 cil	8.1	8.6	9.1
Grande - 9 cil	8.3	8.6	8.8
Desportivo	8.2	9.2	9.5

As hipóteses a testar são as seguintes:

H_0: Não existem diferenças no consumo dos diferentes automóveis.

vs.

H_1: Existem diferenças no consumo dos diferentes automóveis.

Utilizando a expressão 7.2 obtemos o seguinte:

$$\chi^2_{calc} = \frac{12}{5\times 3\times(3+1)}(5^2 + 12^2 + 13^2) - 3\times 5\times 4 = 7,6$$

Para um nível de significância $\alpha = 5\%$, o valor crítico da distribuição χ^2 é $\chi^2_{(0,05;2)} = 5,991$. Como $\chi^2_{Calc} > \chi^2_{(0,05;2)}$ deve rejeitar-se a hipótese H_0.

Donde se conclui que existem diferenças significativas no consumo nos três fabricantes.

Capítulo 8

Testes para o caso de k ($k > 2$) amostras independentes

Na análise de dados de pesquisa, é necessário decidir se diversas variáveis independentes devem ser consideradas como provenientes da mesma população. Os valores amostrais quase sempre são um tanto diferentes, e o problema é determinar se as diferenças amostrais observadas sugerem realmente diferenças entre as populações ou se são apenas variações casuais que podem ser esperadas entre amostras aleatórias da mesma população.

8.1 Teste de Kruskal-Wallis

O objetivo deste teste é ver se as diferentes k ($k > 2$) amostras provêm da mesma população ou de populações idênticas em relação à mediana.

Ele indica-nos se há diferenças entre pelo menos duas amostras.

É na verdade uma extensão do teste de Wilcoxon para duas amostras independentes.

São os seguintes os passos a percorrer:

1. Dispor, em postos, as observações de todos os k grupos numa única série, atribuindo-lhes postos de 1 a N:

2. Determinar o valor de R (soma das ordens) para cada um dos k grupos de postos;

3. Caso não ocorram empates, calcular o valor de H, estatística de teste, pela seguinte expressão:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1) \tag{8.1}$$

onde:

k= número de amostras;

n_j = número de casos na amostra j;

N = $\sum n_j$, número de elementos em todas as amostras combinadas;

R_j = soma das ordens na amostra j.

Se houver empates, atribui-se a cada uma delas a média das respectivas ordens. O valor de H é influenciado pelos empates, sendo assim é necessário introduzir um factor de correcção. Deste modo, para o cálculo de H deve-se utilizar a fórmula:

$$H = \frac{\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)}{1 - \frac{\sum T}{N^3 - N}} \tag{8.2}$$

onde:

$T = t^3 - t$ (sendo o número de observações empatadas num grupo de valores empatados);

Esta estatística de teste tem, aproximadamente, uma distribuição Qui-Quadrado com k-1 graus de liberdade.

4. O método para determinar a significância do valor observado de H depende da dimensão de k e da dimensão dos grupos;

5. Se a probabilidade associada ao valor observado de H, valor-p, não superar o nível de significância previamente fixado, rejeita-se H_0.

8.1.1 Exemplo

A tabela em baixo, mostra os dados de três grupos de indivíduos relativos ao número de vezes que os mesmos realizam algum tipo de compra num centro comercial durante um mês.

Os grupos apresentam a mesma distribuição?

G1	G2	G3
20	12	8
4	21	22
7	9	10
2	0	5
17	14	6
3	1	20

As hipóteses a considerar são as seguintes:

H_0: Os grupos apresentam a mesma distribuição.

vs.

H_1: Os grupos não apresentam a mesma distribuição

Vamos fazer a ordenação das posições de 1 a 18 de todos os dados. Quando existir empates divide-se o número da posição pelo número de empates para continuar a ordenação.

G1	Pos	G2	Pos	G3	Pos
20	15,5	12	12	8	9
4	5	21	17	22	18
7	8	9	10	10	11
2	3	0	1	5	6
17	14	14	13	6	7
3	4	1	2	20	15,5

Calcula-se a soma das posições para cada grupo e calcula-se a estatística de teste:

$$H = [\frac{12}{18 \times (18+1)}] \times [\frac{49{,}5^2}{6} + \frac{55{,}0^2}{6} + \frac{66{,}5^2}{6}] - 3 \times (18+1) = 0,73$$

Como H tem uma distribuição de Qui-Quadrado com 2 graus de liberdade, segundo a tabela de distribuição do Qui-Quadrado este valor é inferior ao valor tabelado logo não existem diferenças significativas entre os grupos, ou seja rejeita-se a hipótese H_0.

Capítulo 9

Uma aplicação

Em duas turmas de 11^{O} ano, fizemos um inquérito aos alunos para saber as habilitações dos pais e a sua classificação obtida no ano anterior.

O objectivo foi o de saber se a habilitação literária dos pais condicionava o resultado académico obtido pelos alunos.

Elaborámos uma tabela constituida da seguinte forma: alunos em que pelo menos um dos progenitores tem como habilitação literária oensino superior e os alunos em que nenhum dos pais tem ensino superior como habilitação literária, e dividimos os alunos pela classificação obtida; os alunos de suficiente com classificação no intervalo [10, 14[e os alunos bons e muito bons com uma classificação no intervalo [14, 20].

Utilizámos o teste do Qui-QUadrado de independência para testar as seguintes hipóteses:

H_0: Há independência entre o nível de escolaridade dos pais e o resultado académico dos alunos.

vs.

H_1: Não há independência entre o nível de escolaridade dos pais e o resultado académico dos alunos.

Consideremos então a seguinte tabela de contingência em que:

A representa alunos em que pelo menos um dos pais tem habilitação literária ao nível do ensino superior

B representa alunos em que nenhum dos pais tem habilitação literária ao nível do ensino superior.

Classificação	A	B
[10, 14[	10	15
[14, 20]	14	9

Vamos construir agora a tabela dos valores esperados:

Classificação	A	B
[10, 14[	12,5	12,5
[14, 20]	11,5	11,5

Fazendo os cálculos para obter o valor da estatística de teste, obtemos o seguinte:

$$\chi^2_{calc} = \frac{(10-12{,}5)^2}{12{,}5} + \frac{(15-12{,}5)^2}{12{,}5} + \frac{(14-11{,}5)^2}{11{,}5} + \frac{(9-11{,}5)^2}{11{,}5} = 2,08696$$

Rejeita-se H_0 quando $\chi^2_{Calc}) > \chi^2_{(0{,}95;1)}$

Utilizando a tabela do Qui-Quadrado, vemos que $\chi^2_{(0{,}95;1)} = 3,841$

Como $\chi^2_{calc} < \chi^2_{(0{,}95;1)}$ não se rejeita H_0. Logo, com base nestes resultados, não podemos afirmar que o grau de ensino dos pais influencia os resultados escolares.

Calculando o valor-p podemos ver que $0,1 < P(\chi^2_1 \geq 2,086969) < 0,25$ e utilizando o programa Excel podemos afirmar que $P(\chi^2_1 \geq 2,086969) = 0,149$, o que nos permite chegar à mesma conclusão.

Capítulo 10

Conclusão

As estatísticas não paramétricas são técnicas de inferência estatística. Podem ser utilizadas com distribuição de resultados que não obedeçam aos pressupostos da distribuição normal.

De um modo geral são as variáveis qualitativas que estão mais ligadas aos modelos não paramétricos.

Dentro da Estatística não paramétrica, estudamos os testes de hipóteses. Trata-se de uma técnica para se fazer inferência estatística sobre uma população a partir de uma amostra. É uma regra de decisão para rejeitar ou não rejeitar uma hipótese estatística com base nos elementos amostrais.

Existem muitos testes estatísticos não paramétricos. Deve-se ter em atenção alguns pressupostos na sua escolha: a maneira como a amostra foi obtida, a natureza da população da qual se extraiu a amostra, o tipo de variável envolvida e o tamanho da amostra disponível.

Vejamos então quais as etapas a seguir para formular um teste de hipóteses:

- Formular as hipóteses;
- Definir ou fixar o nível de significância α;

- Identificar a estatística de teste e a respectiva distribuição;
- Definir a região crítica;
- Calcular o valor observado da estatística de teste;
- Tomar uma decisão;
- Formular a conclusão.

Testes para o caso de uma amostra:

Teste do Qui-Quadrado (teste de ajustamento): Este teste é adequado aplicar quando se tem todos os elementos da amostra divididos em duas ou mais categorias. Serve para averiguar se uma amostra pode ser considerada como proveniente de uma população com uma determinada distribuição sem restrições sobre esta. Pode também ser usado para verificar se as categorias de uma variável estão equitativamente distribuídas. A estatística de teste segue uma distribuição Qui-Quadrado.

Teste da Binomial: Teste aplicado em amostras provenientes de populações que estão divididas em duas categorias, por exemplo, masculino e feminino, membro ou não membro de uma qualquer associação, doente ou não doente. Para qualquer população dividida em duas categorias (isto é, dicotomizada), se conhecermos a proporção P, numa das categorias, a proporção na outra será $1 - P$. A estatística de teste segue uma distribuição Binomial.

Teste de Kolmogorov-Smirnov: Foi proposto em 1933 por Kolmogorov e avalia o grau de concordância entre a distribuição de um conjunto de valores amostrais (observados) e uma determinada distribuição teórica. Determina se os valores da amostra podem ser considerados como provenientes de uma população com aquela distribuição teórica.

O teste de Kolmogorov-Smirnov pode ser preferido em relação ao teste do Qui-Quadrado devido à forma como se ajusta à amostra, se o tamanho desta for pequeno; o teste de Kolmogorov-Smirnov é exacto mesmo para pequenas amostras, enquanto o teste do Qui-Quadrado assume que o número de observações é grande o suficiente para que a distribuição represente uma boa aproximação à distribuição da estatística de teste. Há controvérsias sobre qual dos testes é o mais potente, mas actualmente é considerado que o teste de Kolmogorov-Smirnov é mais potente do que o teste do Qui-Quadrado na maioria das situações.

Teste dos Sinais: O teste de hipóteses sobre a mediana (m) é importante nas decisões sobre a localização da distribuição da população, até por não necessitar de qualquer pressupos to sobre a distribuição desta. Para este teste pressupõe-se que a distribuição da população é contínua. A estatística de teste é o número de observações abaixo (ou acima) de m.

Teste de Wilcoxon: O teste de Wilcoxon tem a vantagem de ser mais potente do que o teste dos sinais, isto é, é menor a probabilidade de se cometer o erro de não rejeitar H_0 sendo H_0 falsa.

Quando se pretende estudar uma hipótese sobre a mediana e se considera como pressuposto a simetria da distribuição dos valores, o teste de Wilcoxon representa uma melhoria em relação ao teste dos sinais pois não despreza a informação dada pela ordem das diferenças.

Teste de Aleatorização das Iterações: Este teste faz uso da análise das sequências de símbolos idênticos. Verifica o número de iterações existentes na amostra; se o número de iterações é muito grande ou muito pequeno sugere-se falta de aleatoriedade da amostra.

Tabelas de contingência: Um processo de organizar a informação correspondente a dados bivariados é utilizando uma tabela de contingência.

De uma maneira geral, uma tabela de contingência é uma forma de organizar dados, quer de tipo qualitativo, quer de tipo quantitativo, especialmente quando são de tipo bivariado, isto é, podem ser classificados segundo dois critérios.

As tabelas de contingência são uma apresentação tabular de contagens de efectivos de classes.

Uma tabela de contingência é uma tabela de frequências que apresenta um conjunto de dados que foram classificados simultaneamente segundo duas (bidimensional) ou mais variáveis (multidimensional). As tabelas de contingência têm pelo menos, duas linhas e duas colunas.

As tabelas de contingência também se utilizam no caso em que se pretende verificar se determinada característica categorizada se distribui de forma semelhante pelas diferentes categorias de duas ou mais populações, ou seja, quando se pretende averiguar se duas ou mais populações são homogéneas no que diz respeito à distribuição de determinada característica.

Teste do Qui-Quadrado para duas amostras independentes: O teste de independência do Qui-Quadrado permite verificar a independência entre duas variáveis de qualquer tipo que se apresentem agrupadas numa tabela de contingência.

Este teste não deve ser utilizado se mais do que 20% das frequências esperadas sob a hipótese da independência forem inferiores a 5 ou se algumas delas for igual a zero.

Teste do Qui-Quadrado de homogeneidade: Este teste constrói-se de maneira idêntica ao teste de Qui-Quadrado, sendo apenas diferente nas hipóteses a testar.

O teste de Qui-Quadrado de homogeneidade pode ser utilizado para comparar as populações em termos das proporções de elementos de determinada característica em estudo.

Teste exacto de Fisher: O teste exacto de Fisher constitui uma técnica não paramétrica muito útil para analisar dados discretos quando a dimensão das amostras independentes é pequena e consiste em determinar a probabilidade exacta de ocorrência de uma frequência observada, ou de valores mais extremos.

Este teste exige que:

- Se tenha duas populações;

- Cada população seja dividida em duas categorias, categorias estas que têm de ser as mesmas para as duas populações;

- Se tenha duas classes mutuamente exclusivas, ou seja, cada elemento de uma população irá pertencer a exactamente uma das categorias.

Testes para o caso de duas amostras independentes:

Teste U de Mann-Whitney: O teste exige que os grupos tenham a mesma distribuição (que não precisa ser normal).

O teste U de Mann-Whitney (1947) pode-se aplicar para comprovar se dois grupos independentes foram ou não extraídos da mesma população. Trata-se de um teste não paramétrico potente, e constitui uma alternativa extremamente útil quando se deseja evitar suposições exigidas pelo teste paramétrico t.

As amostras devem ser independentes e aleatórias: uma extraída duma população com mediana não conhecida M_1 e outra extraída de outra população com mediana desconhecida M_2.

A hipótese a testar é ver se as populações têm a mesma mediana, sendo a alternativa, as medianas serem diferentes ou uma maior do que a outra.

Teste de Moses para reacções extremas: Este teste aplica-se quando existe uma suspeita de que uma determinada condição experimental afectou um grupo de indivíduos e, de forma oposta, outro grupo.

O teste considera dois grupos (amostras) independentes, um grupo de controlo (C) e um grupo experimental (E).

Embora o teste de Moses se destine especificamente ao tipo de dados do grupo de controlo e do grupo experimental, é também aplicável quando se prevê que um grupo tenha resultado alto, e o outro grupo resultado baixo. Todavia, Moses salienta que, em tais casos, um teste baseado em medianas é mais eficiente do que este, por exemplo o teste U de Mann-Witney.

Testes para o caso de duas amostras emparelhadas:

Teste de McNemar: É um teste aplicado a variáveis dicotómicas, ou seja, a variáveis que apenas tomam dois valores (por exemplo sim/não). Por exemplo, testar se numa determinada licenciatura, a hipótese nula de que há igualdade entre rapazes e raparigas de terem irmãos.

Teste de Wilcoxon: O teste de Wilcoxon atribui maior ponderação a um par que acusa grande diferença entre as condições, do que a um par em que essa diferença seja pequena.

O teste de Wilcoxon é extremamente útil para os cientistas do comportamento. Com dados sobre o comportamento, não são raros os casos em que o observador pode dizer qual membro de um par é “maior do que”o outro, e dispor as diferenças por ordem do seu valor absoluto. Isto é, o observador pode fazer o julgamento do tipo “maior do que”entre os resultados de qualquer par, bem como fazer esse julgamento em relação às difererenças relativas a dois pares quaisquer. Dispondo destas informações pode-se aplicar o teste de Wilcoxon.

Teste dos Sinais: O teste dos Sinais tem a sua denominação devida ao facto de utilizar como dados, sinais “mais”e “menos”, em vez de medidas quantitativas. É particularmente útil nos trabalhos de pesquisa em que é impossível ou inviável a obtenção de uma medida quantitativa, mas é possível estabelecer posições em relação a cada um dos dois membros de cada par de valores.

O teste dos Sinais é aplicável no caso de duas amostras emparelhadas, quando se deseja determinar se duas condições são diferentes. A única suposição que o teste dos Sinais exige é que a variável em estudo tenha distribuição contínua. O teste não faz qualquer suposição sobre a forma da distribuição das diferenças, nem supõe que todos os indivíduos tenham sido extraídos da mesma população. Os diferentes pares podem provir de populações diferentes com respeito a várias características. A única exigência é que, dentro de cada par, se tenha conseguido um nivelamento quanto às variáveis extrínsecas importantes.

Testes para o caso de k (k >2) amostras emparelhadas:

Teste de Q de Cochran: O Teste Q de Cochran é uma extensão do teste de McNemar para amostras emparelhadas, que fornece um método para testar as diferenças entre três ou mais conjuntos combinados de frequências ou proporções.

Este teste proporciona um método para comprovar se três ou mais conjuntos correspondentes de frequências ou proporções diferem entre si significativamente. A correspondência pode basear-se em características relevantes dos diferentes indivíduos, ou no facto de os mesmos indivíduos serem observados em condições diferentes. O teste Q de Cochran adapta-se especialmente ao caso em que os dados se apresentam em escala nominal ou sob a forma de informação ordinal dicotomizada.

Teste de Friedman: Este teste é útil quando se deseja comprovar a hipótese de que k amostras emparelhadas provêm da mesma população. Neste tipo de estudo observa-se o mesmo grupo de indivíduos sob cada uma das k condições, ou então formam-se conjuntos de indivíduos homogéneos entre si, e estes são colocados aleatoriamente em cada uma das condições.

Testes para o caso de k (k >2) amostras independentes:

Teste de Kruskal-Wallis: O objetivo deste teste é ver se as diferentes k ($k > 2$) amostras provêm da mesma população ou de populações idênticas em relação à mediana.

É na verdade uma extensão do teste de Wilcoxon para duas amostras independentes.

Capítulo 11

Bibliografia

[1] Siegel, Sidney (1975) Estatística Não Paramétrica para as ciências do comportamento, McGraw-Hill, Brasil;

[2] Pestana, Dinis Duarte; Velosa, Sílvio Filipe (2002) Introdução à Probabilidade e à Estatística. Fundação Calouste Gulbenkian, Lisboa;

[3] Conover, William Jay (1980) Practical Nonparametric Statistics, United Sates of America;

[4] Murteira, Bento, Ribeiro, Carlos Silva, Silva João Andrade, Pimenta Carlos, (2001) Introdução à Estatística, McGraw-Hill, Portugal;

[5] Martins, Maria Eugénia Graça (2005) Introdução à Probabilidade e à Estatística, Departamento de Estatística e Investigação Operacional da Faculdade de Ciências de Lisboa, Lisboa;

[6] Murteira, Bento; Antunes, Marília (2012) Probabilidades e Estatística, Escolar Editora.