

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

1. Introdução

Os testes estudados até agora envolviam problemas nos quais a distribuição da população em estudo era conhecida, ou pelo menos nunca colocada em causa, e as hipóteses testadas apenas envolviam parâmetros populacionais. No entanto, outro tipo de problemas podem ser colocados: se a distribuição de uma população é desconhecida e se pretender testar a hipótese de uma distribuição particular para aquela população, que fazer? Por exemplo, ao efectuar uma ANOVA, as populações em estudo terão de facto distribuição normal?

Os testes que iremos abordar neste capítulo, denominados **testes não paramétricos** ou **testes de distribuição livre** constituem uma alternativa para este e outro tipos de problemas. O termo “distribuição livre” é vulgarmente usado para indicar que os métodos são aplicáveis independentemente da forma da distribuição, ou que são válidos para um ou mais largo espectro de distribuições.

Estes métodos são, em geral, fáceis de aplicar, pois podem ser usados quando as hipóteses exigidas por outras técnicas não são satisfeitas. Relembramos aqui que os testes paramétricos estudados até agora comportam uma diversidade de suposições fortes a que o seu emprego deve subordinar-se (as observações devem ser extraídas de populações com distribuição normal, as variáveis em estudo devem ser medidas em escala intervalar ou de rácios, de modo a que seja possível utilizar operações aritméticas sobre os valores obtidos das amostras (adição, multiplicação, obtenção de médias, ...), etc.).

Apesar de haver certas suposições básicas associadas à maioria das provas não paramétricas, essas suposições são em menor número e mais fracas do que as associadas às provas paramétricas. Servem para pequenas amostras e, além disso, a maior parte das provas não paramétricas aplicam-se a dados medidos em escala ordinal, e alguns mesmo a dados em escala nominal.

De entre uma vasta gama de testes não paramétricos disponíveis, foram seleccionados para análise neste capítulo apenas alguns testes de utilização frequente ou que complementam, de alguma forma, os testes paramétricos discutidos anteriormente.

2. Testes de Ajustamento

Os **testes de ajustamento**, também designados por **testes da bondade do ajustamento**, servem para testar a hipótese de que uma determinada amostra aleatória tenha sido extraída de uma população com distribuição especificada. Isto é, sendo X_1, X_2, \dots, X_n uma amostra aleatória de uma população X com função (densidade) de probabilidade f desconhecida, as hipóteses a testar são:

$$H_0: X \text{ tem função (densidade) de probabilidade } f_0$$

$$H_1: X \text{ não tem função (densidade) de probabilidade } f_0$$

ou, de modo mais simples,

$$H_0: f(x)=f_0(x)$$

$$H_1: f(x)\neq f_0(x),$$

com f_0 a função (densidade) de probabilidade proposta.

Exemplo 1: A procura diária de um certo produto foi, em 40 dias escolhidos ao acaso, a seguinte:

<u>Número de unidades</u>	<u>Número de dias</u>
0	6
1	14
2	10
3	7
4	2
5	1

Tabela I: Procura diária de um produto registada em 40 dias.

Será que tais observações foram extraídas de uma população com distribuição Poisson, isto é, será de admitir que tal procura segue uma distribuição de Poisson?

Exemplo 2: Pretende-se construir um modelo de simulação das operações de um determinado terminal de um porto situado na Europa. Uma das variáveis a considerar no

modelo é a diferença entre a data de chegada dos navios provenientes dos EU e a respectiva data planeada. Dado que tal diferença é influenciada por muitos factores, pode tomar-se como uma variável aleatória. Há razões para supor que tem distribuição Normal de média 0.1 e desvio padrão 7.2. Uma amostra de 30 navios revelou os resultados que se apresentam na tabela seguinte.

-6.6	-2	5	2.4	-1.8	-0.3	15	-7.6	-0.6	2.6
-7.4	12.4	-6	-5.8	15.2	-2.4	-8.9	-5.6	-3.7	2.2
8.2	-9	13.2	7.6	-2.8	-1.8	1.8	4.4	2.2	4

Tabela II: Diferença entre a data de chegada e a data planeada para 30 navios.

Será mesmo de admitir que tais dados foram extraídos de uma população $N(0.1, 7.2^2)$?

Tanto no primeiro como no segundo exemplo, estamos perante um problema de ajustamento de dados a uma determinada distribuição. Existem vários testes de ajustamento que nos permitem fazer uma análise de problemas deste tipo, entre os quais: o teste de ajustamento do Qui-quadrado sugerido por Karl Pearson, o teste de Kolmogorov ou Kolmogorov-Smirnov e o teste de normalidade de Lilliefors, que apresentamos a seguir.

2.1 Teste do Qui-quadrado

O teste de ajustamento do Qui-quadrado é um teste de fácil construção e baseia-se na comparação da distribuição dos dados da amostra (frequências observadas) com a distribuição teórica à qual se supõe pertencer a amostra.

Considere-se uma amostra aleatória de n elementos, extraída de uma população com distribuição desconhecida, sobre os quais se observa uma característica (qualitativa ou quantitativa). Os valores possíveis da característica em estudo são, num primeiro passo, repartidas por m classes mutuamente exclusivas, A_1, A_2, \dots, A_m (serão intervalos da recta real se a característica é quantitativa e contínua).

Denote-se por:

- O_i o nº de observações ou frequência absoluta observada da classe A_i ;

- p_i a probabilidade **desconhecida** de obter uma observação na classe A_i ;
- p_{0i} a probabilidade de obter uma observação na classe A_i assumindo que a observação foi extraída de uma população com a distribuição especificada em H_0 , i.e $p_{0i} = P(A_i | H_0)$.

Então, o problema que se põe é o de testar as hipóteses:

$$H_0: p_i = p_{0i}, \quad i=1, \dots, m$$

$$H_1: p_i \neq p_{0i} \quad \text{para algum } i.$$

Assim, a frequência esperada da classe A_i , quando H_0 é verdadeira, é dada por

$$e_i = n \times p_{0i}.$$

A estatística de teste, do teste de ajustamento do Qui-quadrado, é dada por

$$Q = \sum_{i=1}^m \frac{(O_i - e_i)^2}{e_i}$$

que, sendo verdadeira a hipótese nula, tem distribuição assintótica do Qui-quadrado com **m-k-1 graus de liberdade** (χ^2_{m-k-1}), onde k é o número de parâmetros desconhecidos da distribuição proposta em H_0 , estimados a partir da amostra.

Se a função (densidade) de probabilidade proposta em H_0 estiver completamente especificada, i.e., se não há parâmetros desconhecidos, então $Q \sim \chi^2_{m-1}$.

Notemos que, se a hipótese nula for de facto verdadeira, a diferença entre cada valor observado e o respectivo valor esperado, $O_i - e_i$, não deve ser muito grande, e consequentemente a estatística teste terá um valor observado, Q_{obs} , também não muito grande. De modo intuitivo, quanto maior for o valor observado de Q, menos plausível é a hipótese nula, isto é, mais nos encaminhamos de concluir que as frequências observadas não foram provenientes da população em que se baseou a hipótese nula, levando à rejeição desta. Trata-se portanto de um teste unilateral à direita.

Assim, para um nível de significância α , a hipótese nula é rejeitada se $Q_{\text{obs}} \geq \chi^2_{1-\alpha}$, com $\chi^2_{1-\alpha}$ o quantil de probabilidade $1-\alpha$ da distribuição χ^2_{m-k-1} .

Na aplicação deste teste deve-se ter particular atenção às frequências esperadas, e_i 's, pois se estas forem muito pequenas a aproximação ao Qui-quadrado não é a mais apropriada. São referidas na literatura várias regras práticas de aplicação do teste, das quais avançamos a seguinte. Se tivermos:

- mais de 20% das classes com e_i inferior a 5

ou,

- mais de uma classe com e_i inferior a 1,

devemos proceder à agregação de algumas classes contíguas, e iniciar novamente o teste, agora com menos classes.

Recorremos aos exemplos apresentados anteriormente para demonstrar a metodologia do teste descrito, neste caso com uma distribuição discreta em H_0 no Exemplo 1 e uma distribuição contínua no Exemplo 2.

No primeiro problema pretende-se verificar se as observações seleccionadas pertencem a uma população com distribuição de Poisson. Representando por X a procura diária do produto e por f a função de probabilidade de X , as hipóteses a testar são

$$H_0: f(x) = f_0(x) = e^{-\mu} \frac{\mu^x}{x!}, \quad x = 0, 1, 2, \dots \text{ e } \mu > 0$$

$$H_1: f(x) \neq f_0(x).$$

Notemos, no entanto, que nada é dito acerca do parâmetro/média da Poisson, μ , sendo assim necessária a sua estimação a partir dos dados da amostra. Uma estimativa para μ é dada pela média amostral $\bar{x} = (0 \times 6 + 1 \times 14 + 2 \times 10 + 3 \times 7 + 4 \times 2 + 5 \times 1) / 40 = 1.7$.

Os dados são inicialmente classificados em 7 classes, da forma apresentada na tabela III, e as probabilidades associadas a cada uma das classes, supondo H_0 verdadeira, são facilmente calculadas. Por exemplo:

$$p_{01} = P(A_1|H_0) = P(X \in \{0\}|H_0) = f_0(0) = e^{-1.7} \frac{1.7^0}{0!} = 0.1827;$$

$$p_{02} = P(A_2|H_0) = P(X \in \{1\}|H_0) = f_0(1) = e^{-1.7} \frac{1.7^1}{1!} = 0.3106;$$

$$p_{03} = e^{-1.7} \frac{1.7^2}{2!} = 0.2639;$$

⋮

Relembrando ainda que as frequências esperadas são calculadas a partir de $e_i = 40 \times p_{0i}$, podemos construir uma tabela como a seguinte :

Classes	Frequências observadas	$p_{0i} = P(A_i H_0)$	Frequências esperadas
$A_1 = \{0\}$	6	0.1827	7.308
$A_2 = \{1\}$	14	0.3106	12.424
$A_3 = \{2\}$	10	0.2639	10.556
$A_4 = \{3\}$	7	0.1496	5.984
$A_5 = \{4\}$	2	0.0636	2.544
$A_6 = \{5\}$	1	0.0216	0.864
$A_7 = \{6, 7, \dots\}$	0	0.008	0.32

$\left. \begin{array}{l} 2 \\ 1 \\ 0 \end{array} \right\} 3$
 $\left. \begin{array}{l} 0.0636 \\ 0.0216 \\ 0.008 \end{array} \right\} 0.0932$
 $\left. \begin{array}{l} 2.544 \\ 0.864 \\ 0.32 \end{array} \right\} 3.728$

Tabela III: Frequências observadas e esperadas.

De acordo com as regras atrás mencionadas as classes $A_5 = \{4\}$, $A_6 = \{5\}$ e $A_7 = \{6, 7, \dots\}$ foram agrupadas numa só, $\{4, 5, 6, 7, \dots\}$, ficando então os dados agrupados em $m = 5$ classes. Além disso, uma vez que estimámos um parâmetro (μ), a estatística teste Q, sob a hipótese H_0 , tem aproximadamente distribuição Qui-quadrado com $m - k - 1 = 5 - 1 - 1 = 3$ graus de liberdade.

Ao nível de significância de 0.05, o quantil de probabilidade 1-0.05 da distribuição χ_3^2 é 7.81, e logo a região crítica é $[7.81, +\infty[$.

Recorrendo à 2ª e 4ª coluna da tabela III, podemos finalmente calcular o valor observado da estatística de teste:

$$Q_{\text{obs}} = \frac{(6 - 7.308)^2}{7.308} + \frac{(14 - 12.424)^2}{12.424} + \frac{(10 - 10.556)^2}{10.556} + \frac{(7 - 5.984)^2}{5.984} + \frac{(3 - 3.728)^2}{3.728} = 0.778 .$$

Então, a hipótese H_0 não é rejeitada ao nível de significância de 0.05, isto é, não podemos rejeitar a hipótese de aquelas observações provirem de uma população com distribuição Poisson.

Em relação ao Exemplo 2, denotando por X a diferença entre a data de chegada dos navios e a data planeada, as hipóteses a testar são

$$H_0: X \sim N(0.1, 7.2^2)$$

$$H_1: X \not\sim N(0.1, 7.2^2).$$

À semelhança do exemplo anterior, a variável em estudo é quantitativa sendo então necessário a agregação dos dados em classes. Notemos, no entanto, que neste caso a distribuição proposta em H_0 é contínua e, deste modo, as classes A_i , $i=1, \dots, m$, são intervalos da forma

$$A_1 =]-\infty, a_1[, \quad A_2 = [a_1, a_2[\quad A_3 = [a_2, a_3[\quad \dots \quad A_m = [a_{m-1}, +\infty[.$$

Para a determinação das classes, mais precisamente dos limites dos intervalos de classe é sugerida a *regra de Mann e Wald*:

Fixado o número de classes, m , com m tal que $n/m > 5$ para que não seja necessário a agregação de classes, definem-se os limites de modo a que as probabilidades decorrentes da hipótese nula sejam iguais a $1/m$ para todas as classes, isto é, de modo a que as frequências esperadas sejam todas iguais a n/m .

Para o exemplo escolheu-se $m=4$ classes, donde

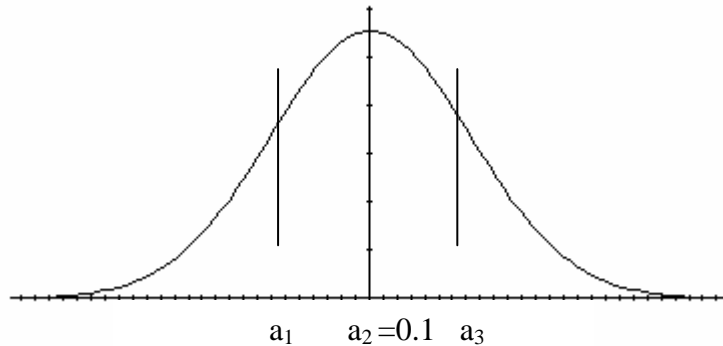
$$p_{0i} = P(A_i | H_0) = P(X \in A_i | X \sim N(0.1, 7.2^2)) = 1/4, \text{ para } i=1,2,3,4.$$

Note-se que deste modo as frequências esperadas $e_i = 30 \times 1/4 = 7.5 > 5$, para $i=1, \dots, 4$.

Cálculo dos limites dos intervalos de classe:

$$a_1: p_{01} = P(X \in A_1 | X \sim N(0.1, 7.2^2)) = 0.25 \Leftrightarrow P(X < a_1 | X \sim N(0.1, 7.2^2)) = 0.25 \Leftrightarrow \\ \Leftrightarrow P(Z < \frac{a_1 - 0.1}{7.2}) = 0.25 \Leftrightarrow a_1 = -4.724;$$

Da simetria da distribuição normal (ver gráfico), $a_2 = 0.1$ e $a_3 = 0.1 + 0.1 - (-4.724) = 4.924$.



Classes	Frequências observadas	p_{0i}	Frequências esperadas
$A_1 =]-\infty, -4.724[$	8	0.25	7.5
$A_2 = [-4.724, 0.1[$	8	0.25	7.5
$A_3 = [0.1, 4.924[$	7	0.25	7.5
$A_4 = [4.924, +\infty [$	7	0.25	7.5

Tabela IV: Frequências observadas e esperadas.

O valor observado da estatística de teste é

$$Q_{\text{obs}} = \frac{(8-7.5)^2}{7.5} + \frac{(8-7.5)^2}{7.5} + \frac{(7-7.5)^2}{7.5} + \frac{(7-7.5)^2}{7.5} = 0.1.$$

A estatística teste, sob o pressuposto de H_0 ser verdadeira, tem aproximadamente distribuição Qui-quadrado com $m-1=4-1=3$ graus de liberdade.

Ao nível de significância de 0.05, o quantil de probabilidade $1-0.05$ da distribuição χ_3^2 é igual a 7.81. Então, como Q_{obs} não pertence à região crítica, somos levados a não rejeitar a

hipótese de que a diferença entre os tempos de chegada e os tempos planeados tem distribuição $N(0.1, 7.2^2)$.

2.2 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov, abreviadamente K-S, ao contrário do teste do Qui-quadrado, não se aplica a dados qualitativos nem a variáveis discretas, pois a tabela disponível para este teste só é exacta caso a distribuição em teste seja contínua. No entanto, tem a vantagem de não estar dependente de classificações dos dados, que além de serem sempre algo arbitrárias envolvem perdas de informação. De facto, no ajustamento de uma distribuição contínua a uma amostra usando o teste do Qui-quadrado, temos de proceder à agregação dos dados em classes, sendo por isso mais adequado utilizar o teste K-S.

Por outro lado, o teste K-S só pode ser aplicado quando a distribuição indicada na hipótese nula está completamente especificada (o que não sucede com o teste do Qui-quadrado). No caso de pretendermos, por exemplo, efectuar um ajustamento de uma distribuição normal, sem especificar μ e σ , podemos recorrer a outro teste, neste caso o teste desenvolvido por Lilliefors (teste de normalidade de Lilliefors) que será abordado mais tarde.

Além disso, o teste do Qui-Quadrado está orientado essencialmente para grandes amostras, enquanto que o teste K-S é aplicável a pequenas amostras.

Definem-se de seguida função de distribuição da amostra e função de distribuição empírica, conceitos estes fundamentais para o desenvolvimento do teste de K-S.

Função de distribuição empírica e função de distribuição da amostra

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma certa população X e (x_1, x_2, \dots, x_n) uma sua realização.

A **função de distribuição empírica** é definida por

$$\hat{F}_n(x) = \frac{1}{n} \#\{x_i: x_i \leq x\}, \quad -\infty < x < +\infty,$$

onde $\#\{x_i: x_i \leq x\}$ é o número de valores x_i que são inferiores ou iguais a x .

A **função de distribuição da amostra** é definida, para as variáveis aleatórias (X_1, X_2, \dots, X_n) , por

$$F_n(x) = \frac{1}{n} \#\{X_i: X_i \leq x\}, \quad -\infty < x < +\infty.$$

Convém fazer aqui uma clara distinção entre função de distribuição empírica \hat{F}_n , definida para uma particular realização (x_1, x_2, \dots, x_n) , e função de distribuição da amostra F_n , definida para as variáveis aleatórias (X_1, X_2, \dots, X_n) . Note-se que $\hat{F}_n(x)$ é uma função de distribuição do tipo discreto associado a uma particular amostra, enquanto que $F_n(x)$ é, para cada $-\infty < x < +\infty$ fixo, uma variável aleatória, função de (X_1, X_2, \dots, X_n) , ou seja, é uma estatística.

Veamos como construir uma função de distribuição empírica associada a uma determinada amostra. Consideremos, por exemplo, a amostra constituída pelas observações: 5, 7, 8, 8, 10 e 11. A função de distribuição empírica \hat{F}_6 , associada a esta amostra, é dada por

$$\hat{F}_6(x) = \begin{cases} 0 & \text{se } x < 5 \\ 1/6 & \text{se } 5 \leq x < 7 \\ 2/6 & \text{se } 7 \leq x < 8 \\ 4/6 & \text{se } 8 \leq x < 10 \\ 5/6 & \text{se } 10 \leq x < 11 \\ 1 & \text{se } x \geq 11 \end{cases}.$$

A representação gráfica de \hat{F}_6 , em forma de escada, é apresentada a seguir:

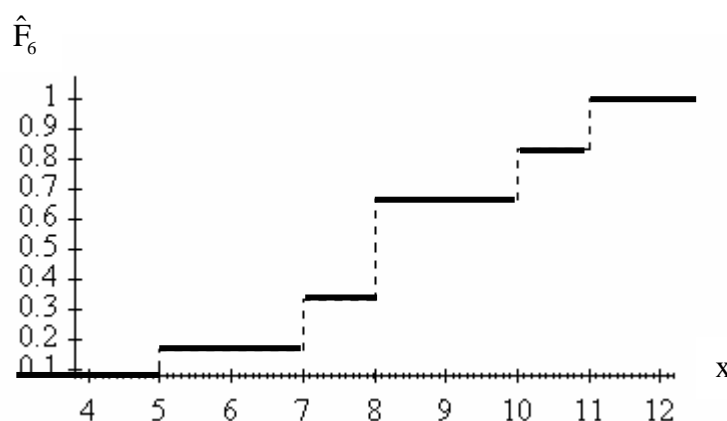


Gráfico I: Representação gráfica da f.d. empírica.

Note-se que a função \hat{F}_n é descontínua à esquerda em cada valor x_i .

Para uma variável aleatória X , o teste K-S baseia-se na análise do ajustamento entre a função de distribuição populacional admitida em H_0 , F_0 , e a função de distribuição empírica \hat{F}_n .

Assim, sendo F a função de distribuição (desconhecida) da população, as hipóteses a testar serão:

$$H_0: F(x)=F_0(x), -\infty < X < +\infty$$

$$H_1: F(x) \neq F_0(x),$$

onde F_0 é a função de distribuição proposta, contínua e completamente especificada.

No teste de Kolmogorov-Smirnov considera-se a estatística

$$D_n = \sup_{-\infty < X < +\infty} |F_n(x) - F_0(x)|,$$

como uma medida da discrepância entre a função de distribuição da amostra F_n e a função de distribuição proposta F_0 . Observe-se que D_n representa a distância vertical máxima entre as imagens da função de distribuição da amostra, $F_n(x)$, e da função de distribuição proposta $F_0(x)$, dando assim uma ideia do ajustamento, como aliás se pretendia.

Ao substituir em D_n a função de distribuição da amostra F_n pela função de distribuição empírica \hat{F}_n , obtém-se o valor observado da estatística teste:

$$d_n = \sup_{-\infty < X < +\infty} |\hat{F}_n(x) - F_0(x)|.$$

Uma vez que F_0 é uma função (contínua) crescente e \hat{F}_n é uma função em escada (ver gráfico II), o supremo d_n ocorre num ponto onde se verifica um salto de \hat{F}_n (numa observação x_i) ou imediatamente antes desse ponto. Isto é,

$$d_n = \max_{i=1, \dots, n} \left\{ |F_0(x_i) - \hat{F}_n(x_i)|, |F_0(x_i) - \hat{F}_n(x_i^-)| \right\}.$$

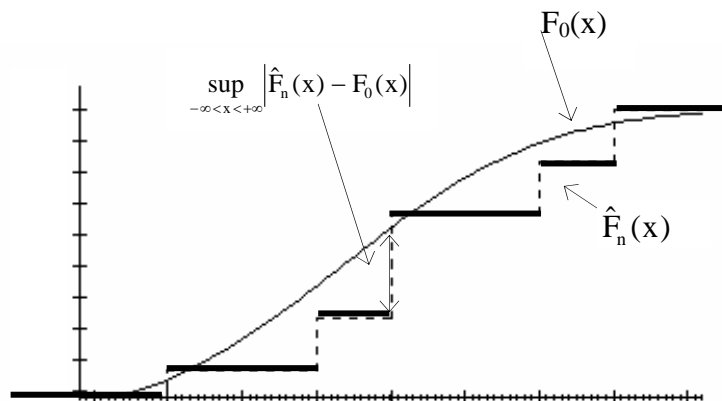


Gráfico II: Ajustamento de uma f.d. hipotética F_0 à f.d. empírica \hat{F}_n .

Assim, se H_0 for verdadeira, a distância vertical máxima entre as imagens das duas distribuições não deve de ser muito grande, e logo espera-se que D_n tome um valor pequeno.

Conhecendo a distribuição de D_n é possível avaliar o seu valor observado d_n , e decidir se este é suficientemente pequeno para que não se rejeite H_0 a um dado nível de significância α .

Mais concretamente, rejeita-se H_0 , para um nível de significância α , se o valor observado d_n da estatística teste D_n for superior ou igual ao ponto crítico $D_{n,\alpha}$ onde $D_{n,\alpha}$ é tal que,

$$P(D_n \geq D_{n,\alpha} \mid H_0 \text{ é verdadeira}) = \alpha.$$

Os valores críticos $D_{n,\alpha}$ podem ser consultados na Tabela 7 (em anexo).

Exemplo 3: Um certo Politécnico do país efectuou um contrato com uma determinada empresa que ficou responsável pelo abastecimento da carne que compunha as refeições na cantina dessa Escola. O contrato refere uma média de 290 gramas de carne por refeição, por estudante. No entanto, alguns alunos queixaram-se acerca da comida, em particular acerca da quantidade de carne servida por refeição. Os alunos falaram com o cozinheiro chefe, que lhes disse que a quantidade de carne servida por refeição a cada estudante tinha aproximadamente distribuição normal de média 290 gr com um desvio padrão de 56 gr.

Após esta conversa com o cozinheiro, alguns alunos concordaram em recolher as suas refeições ao longo de vários dias, resultando assim uma amostra de 10 refeições, que foram levadas para um laboratório afim de serem pesados os pedaços de carne nelas contidos. Os dados obtidos são os seguintes:

198 254 262 272 275 278 285 287 287 292

Ao nível de significância de 5%, há evidência para rejeitar a hipótese de que o cozinheiro seguia as regras que afirmou em relação à quantidade de carne servida?

Denote-se por X a quantidade, em gramas, de carne servida por refeição a cada estudante.

As hipóteses a testar são, neste caso,

$$H_0: X \sim N(290, 56^2) \quad H_1: X \not\sim N(290, 56^2).$$

Note-se que a função de distribuição proposta em H_0 , F_0 , é a função de distribuição normal com média 290 gr e desvio padrão 56. Assim, sendo uma distribuição contínua completamente especificada, podemos usar o teste de K-S.

A estatística de teste é $D_{10} = \sup_x |F_{10}(x) - F_0(x)|$.

O ponto crítico da estatística de teste D_{10} é, para $\alpha=0.05$, $D_{10,0.05} = 0.409$ (Tabela 7).

A função de distribuição empírica definida para a amostra dada foi calculada do modo descrito anteriormente, e é indicada na Tabela V.

Falta calcular as imagens das observações pela função distribuição F_0 . Indicam-se apenas os cálculos para as duas primeiras observações; os restantes encontram-se na respectiva coluna da tabela V.

$$F_0(198) = P(X \leq 198) = P\left(Z \leq \frac{198 - 290}{56}\right) = P(Z \leq -1.64) = 1 - 0.9495 = 0.0505;$$

$$F_0(254) = P(X \leq 254) = P\left(Z \leq \frac{254 - 290}{56}\right) = P(Z \leq -0.64) = 1 - 0.7389 = 0.2611$$

x_i	$\hat{F}_{10}(x_i)$	$\hat{F}_{10}(x_i^-)$	$z_i = \frac{x_i - 290}{56}$	$F_0(z_i)$	$ F_0(x_i) - \hat{F}_{10}(x_i) $	$ F_0(x_i) - \hat{F}_{10}(x_i^-) $
198	0,1	0	-1,64	0,0505	0,0495	0,0505
254	0,2	0,1	-0,64	0,2611	0,0611	0,1611
262	0,3	0,2	-0,5	0,3085	0,0085	0,1085
272	0,4	0,3	-0,32	0,3745	0,0255	0,0745
275	0,5	0,4	-0,27	0,3936	0,1064	0,0064
278	0,6	0,5	-0,21	0,4168	0,1832	0,0832
285	0,7	0,6	-0,09	0,4641	0,2359	0,1359
287	0,9	0,7	-0,05	0,4801	0,4199	0,2199
292	1	0,9	0,04	0,516	0,484	0,384

Tabela V: Cálculo do valor observado da estatística D_{10} .

Observamos, através das duas últimas colunas da tabela, que o valor observado da estatística teste d_{10} é igual a 0.484. Assim, como $d_{10} > 0.409$, ao nível de significância de 5%, rejeitamos a hipótese de a quantidade de carne servida por refeição a cada estudante seguir distribuição $N(290, 56^2)$.

2.3 Teste de Normalidade Lilliefors

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma população X com distribuição F desconhecida. Pretende-se testar se X tem distribuição $N(\mu, \sigma^2)$ **sem especificar** μ e σ , isto é, para algum μ e algum σ . As hipóteses a testar são

$$H_0: X \sim N(\mu, \sigma^2)$$

$$H_1: X \not\sim N(\mu, \sigma^2).$$

Foi dito anteriormente que a aplicação do teste de Kolmogorov-Smirnov carece da especificação completa da função de distribuição proposta em H_0 , o que não sucede no problema. É proposto então o **teste de normalidade de Lilliefors**.

Este teste processa-se como o teste de Kolmogorov-Smirnov, mas os dados originais são estandardizados, usando estimativas de μ e σ :

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (i=1, \dots, n).$$

Assim, as hipóteses a testar são:

$$H_0: Z \sim N(0, 1)$$

$$H_1: Z \not\sim N(0, 1).$$

Seja F_0 a função de distribuição $N(0,1)$, a estatística teste de Lilliefors é

$$D_n^* = \sup_{-\infty < z < +\infty} |F_n(z) - F_0(z)|,$$

onde F_n é a função de distribuição da amostra depois de estandardizada, i.e., definida para as variáveis aleatórias Z_1, Z_2, \dots, Z_n .

Mais uma vez, o valor observado da estatística teste D_n^* , d_n^* , é obtido substituindo na estatística a função distribuição da amostra estandardizada pela função distribuição empírica.

A hipótese H_0 é rejeitada, para um nível de significância α , se o valor observado d_n^* for superior ou igual ao ponto crítico $D_{n,\alpha}^*$, com $D_{n,\alpha}$ tal que,

$$P(D_n^* \geq D_{n,\alpha}^* \mid H_0 \text{ é verdadeira}) = \alpha.$$

Os pontos críticos são consultados na tabela elaborada por Lilliefors (Tabela 8).

Exemplo 4: Um distribuidor pretende estimar o tempo médio de entrega dos seus produtos a um cliente bastante importante. Foi recolhida uma amostra aleatória de cinco tempos: 29, 33, 35, 36 e 36.

O senhor quer estimar o tempo médio pretendido através de um intervalo de confiança, mas não sabe nada acerca da distribuição do tempo de entrega X , e além disso, a dimensão da amostra é muito pequena ($n=5$). Poderá fazê-lo?

Sabemos que caso X tenha distribuição normal o intervalo pode ser calculado usando a estatística $T = \frac{\bar{X} - \mu}{S} \sim t_{n-1}$.

Assim, interessa testar, em primeiro lugar, as hipóteses

$$H_0: X \sim N(\mu, \sigma^2)$$

$$H_1: X \not\sim N(\mu, \sigma^2).$$

Uma vez que nada sabemos acerca de μ e σ , podemos utilizar o teste de Lilliefors, recorrendo às estimativas $\bar{x}=33.8$ $s=2.95$.

O valor crítico da estatística teste, ao nível de significância de 0.05 é $D_{5,0.05}^* = 0.337$, e assim a distribuição normal é rejeitada se $d_5^* \geq 0.337$.

O cálculo do valor observado d_5^* é apresentado, de forma sucinta, na tabela seguinte.

x_i	$z_i = \frac{x_i - 33.8}{2.95}$	$\hat{F}_5(z_i)$	$\hat{F}_5(z_i^-)$	$F_0(z_i)$	$ F_0(z_i) - \hat{F}_5(z_i) $	$ F_0(z_i) - \hat{F}_5(z_i^-) $
29	-1,63	0,2	0	0,0516	0,1484	0,0516
33	-0,27	0,4	0,2	0,3936	0,0064	0,1936
35	0,41	0,6	0,4	0,6591	0,0591	0,2591
36	0,75	1	0,6	0,7734	0,2266	0,1734

Tabela VI: Cálculo do valor observado da estatística D_5^* .

Observamos que $d_5^* = 0.2591 < 0.337$, e logo, ao nível de significância de 5%, não rejeitamos a hipótese de a população em estudo ter distribuição normal.

O cálculo do intervalo de confiança pretendido fica como exercício.

3. Tabelas de Contingência

3.1 Teste do Qui-quadrado de Independência

Suponha que numa amostra aleatória de tamanho n de uma dada população são observados **dois atributos** ou **características** A e B (qualitativas ou quantitativas), uma com r e outra com s modalidades ou categorias, respectivamente A_1, A_2, \dots, A_r e B_1, B_2, \dots, B_c .

Cada indivíduo da amostra é classificado numa e numa só categoria (ou classe) de A e numa e numa só categoria (ou classe) de B . A classificação dos elementos da amostra dá origem a uma tabela de dupla entrada, designada por **tabela de contingência $r \times c$** , com o seguinte aspecto:

	B_1	B_2	...	B_c
A_1	O_{11}	O_{12}	...	O_{1c}
A_2	O_{21}	O_{22}	...	O_{2c}
\vdots	\vdots	\vdots	\ddots	\vdots
A_r	O_{r1}	O_{r2}	...	O_{rc}

Tabela I: Tabela de contingência $r \times c$.

Nesta tabela cada O_{ij} ($i=1, \dots, r$ e $j=1, \dots, c$) é uma variável aleatória que representa na amostra o número de elementos classificados simultaneamente nas categorias A_i de A e B_j de B .

Além disso, temos as variáveis aleatórias:

- $O_{.i} = \sum_{j=1}^c O_{ij}$ ($i=1, \dots, r$) que representa o número de elementos na amostra com modalidade A_i ;
- $O_{.j} = \sum_{i=1}^r O_{ij}$ ($j=1, \dots, c$) que representa o número de elementos na amostra com modalidade B_j .

Tem-se,

$$n = \sum_{i=1}^r \sum_{j=1}^c O_{ij} = \sum_{i=1}^r O_{i.} = \sum_{j=1}^c O_{.j},$$

onde n é a dimensão da amostra que se supõe fixa.

O objectivo a que nos propomos é o de tentar inferir sobre a existência ou não de qualquer relação ou associação entre os atributos (variáveis) A e B , mais concretamente, inferir se A e B são ou não independentes.

Pretende-se assim testar as hipóteses

H_0 : A e B são independentes

H_1 : A e B não são independentes.

Denote-se por:

- $p_{ij} = P(A_i \cap B_j)$ ($i=1, \dots, r$ e $j=1, \dots, c$) a probabilidade (desconhecida) de um indivíduo da população ser classificado simultaneamente nas categorias A_i de A e B_j de B ;
- $p_{i.} = P(A_i)$ ($i=1, \dots, r$) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria A_i de A ;
- $p_{.j} = P(B_j)$ ($j=1, \dots, c$) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria B_j de B .

Tem-se,

$$1 = \sum_{i=1}^r \sum_{j=1}^c p_{ij} = \sum_{i=1}^r p_{i.} = \sum_{j=1}^c p_{.j}.$$

Ora, se os atributos são independentes, verifica-se a conhecida relação,

$$P(A_i \cap B_j) = P(A_i)P(B_j),$$

isto é,

$$p_{ij} = p_{i.} p_{.j}.$$

Assim, as hipóteses anteriores podem ser formuladas do seguinte modo:

H_0 : $p_{ij} = p_{i.} p_{.j}$ (para todo i e j)

H_1 : $p_{ij} \neq p_{i.} p_{.j}$ (para algum $i \neq j$).

Uma vez que não se conhecem os verdadeiros valores das probabilidades envolvidas, estas terão que ser estimadas, a partir dos dados amostrais, por

$$\hat{p}_{i.} = \frac{O_{i.}}{n} \quad \text{e} \quad \hat{p}_{.j} = \frac{O_{.j}}{n},$$

onde $o_{i.}$ e $o_{.j}$ são os valores observados das variáveis aleatórias $O_{i.}$ e $O_{.j}$, respectivamente, para uma amostra concreta.

Seja ainda $e_{ij} = n p_{ij}$ o número esperado de indivíduos na classe A_i de A e B_j de B . Então, quando H_0 é verdadeira, i.e, $p_{ij} = p_{i.} p_{.j}$, teremos

$$e_{ij} = n p_{ij} = n p_{i.} p_{.j},$$

valor este que pode ser estimado por

$$\hat{e}_{ij} = n \hat{p}_{i.} \hat{p}_{.j}.$$

À semelhança do teste de ajustamento do Qui-quadrado, a estatística do teste de independência é

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}},$$

que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assintótica do Qui-quadrado com $(r-1)(c-1)$ graus de liberdade.

Vimos que quando H_0 é verdadeira e_{ij} pode ser estimado por $\hat{e}_{ij} = n \hat{p}_{i.} \hat{p}_{.j}$, e logo a diferença entre o_{ij} (frequência observada) e \hat{e}_{ij} (estimativa da frequência esperada supondo a independência) não deve ser grande. Assim, a estatística teste, tal como está definida, mede o afastamento dos dados em relação à hipótese de independência: valores pequenos da estatística teste são compatíveis com a hipótese nula, enquanto que valores muito grandes traduzem um maior afastamento dos dados em relação à hipótese nula, conduzindo à rejeição desta. Trata-se então de um teste unilateral à direita

Exemplo 1: Um supermercado quer testar ao nível de significância de 5% a hipótese de que o modo de pagamento dos clientes nesse estabelecimento é independente do período do dia em que fazem as compras. Existem três modos de efectuar os pagamentos: por cheque, dinheiro e cartão de crédito.

A seguinte tabela de contingência 3×3 apresenta os resultados obtidos numa amostra de 4000 clientes:

MODO DE PAGAMENTO	PERÍODO DO DIA		
	Manhã	Tarde	Noite
Cheque	750	1500	750
Dinheiro	125	300	75
Cartão de Crédito	125	200	175

Tabela II

Denotando por A o atributo *Modo de pagamento* e por B o atributo *Período do dia em que faz as compras*, as hipóteses a testar são

H_0 : A e B são independentes

H_0 : A e B não são independentes.

Uma vez que A e B assumem cada uma 3 modalidades, respectivamente (cheque, dinheiro, cartão de crédito) e (manhã, tarde, noite), sob H_0 , a estatística teste tem distribuição assintótica do Qui-quadrado com $(r-1)(c-1)=(3-1)(3-1)=4$ graus de liberdade.

Ao nível de significância de 0.05, a região crítica é então $[9.49, +\infty[$.

Como vimos, para obtermos o valor observado da estatística teste, temos de calcular as frequências esperadas. Estas calculam-se de modo muito simples:

$$\hat{e}_{ij} = n\hat{p}_i\hat{p}_j = n \frac{O_{i.}}{n} \frac{O_{.j}}{n} = \frac{O_{i.}O_{.j}}{n} : \text{produto dos totais da } i\text{-ésima linha pelos totais da } j\text{-ésima}$$

coluna, em n.

Assim, por exemplo,

$$\hat{e}_{11}=(3000 \times 1000) / 4000=750, \hat{e}_{12}=(3000 \times 2000) / 4000=1500 \text{ e } \hat{e}_{13}=(3000 \times 1000) / 4000=750.$$

MODO DE PAGAMENTO	PERÍODO DO DIA			Totais
	Manhã	Tarde	Noite	
Cheque	750	1500	750	3000
Dinheiro	125	250	125	500
Cartão de Crédito	125	250	125	500
Totais	1000	2000	1000	4000

Tabela III: Frequências esperadas.

Finalmente, o valor observado da estatística teste pode ser calculado, combinando os dados das tabelas anteriores. Tem-se,

$$\chi^2_{\text{obs}} = \frac{(750 - 750)^2}{750} + \frac{(1500 - 1500)^2}{1500} + \dots + \frac{(125 - 125)^2}{125} + \frac{(200 - 250)^2}{250} + \frac{(175 - 125)^2}{125} = 60.$$

Uma vez que 60 excede o valor crítico 9.49, ao nível de significância de 0.05, rejeitamos a hipótese de que o modo de pagamento é independente do período do dia em que as compras são feitas.

3.1.1 Medidas de Associação

No teste do Qui-Quadrado apresentado, se for rejeitada a hipótese de independência entre os atributos, pode interessar medir a intensidade da associação entre os mesmos, através de uma medida adequada.

Uma vez que a estatística do teste mede o afastamento em relação à hipótese de independência, o seu valor observado também poderá servir para avaliar a força da relação entre os atributos. No entanto, houve necessidade de introduzir algumas modificações, devido a diversas razões, por exemplo o facto do χ^2 não tomar valores apenas no intervalo [0,1], o que é salutar numa medida de associação.

O facto de ter havido já vários autores a dedicarem-se ao estudo de tais medidas faz com que o número de coeficientes propostos seja considerável. Vamos apenas referir alguns desses coeficientes.

- **Coefficiente de Contingência de Pearson:**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Este coeficiente varia entre 0 e $\sqrt{(q-1)/q}$ onde $q = \min\{r,s\}$ e portanto nunca assume o valor 1. Valores pequenos de C indicam fraca associação entre os atributos, enquanto que valores grandes de C indicam forte associação.

O facto deste coeficiente não assumir o valor 1 no caso de associação completa é uma sua limitação. Para obviar este problema, Tshuprow propôs o seguinte coeficiente.

- **Coefficiente de Tshuprow:**

$$T = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)/(c-1)}}$$

Este coeficiente varia entre 0 e 1, tomando o valor 0 no caso de existir independência e o valor 1 quando $r=c$ e houver associação completa.

Por último, referimos o coeficiente proposto por Cramer que atinge o valor 1 quando há associação completa.

- **Coefficiente V de Cramer:**

$$V = \sqrt{\frac{\chi^2}{n(q-1)}}, \text{ com } q = \min\{r,s\} \quad 0 \leq V \leq 1.$$

Para o exemplo anterior, rejeitamos a hipótese de independência entre o modo de pagamento e o período do dia em que as compras eram efectuadas. Para ter uma ideia da intensidade de associação entre estes dois atributos, calcula-se, por exemplo, o coeficiente V de Cramer. Assim, tem-se

$$V = \sqrt{\frac{60}{4000 \times 2}} = 0.087.$$

Verificamos, segundo o coeficiente V que, apesar de haver associação entre os atributos, esta pode considerar-se fraca.

3.2 Teste de Homogeneidade

Suponha que são recolhidas amostras aleatórias de c **populações** (subpopulações ou estratos) B_1, B_2, \dots, B_c , nas quais se observa **um atributo** A com r categorias A_1, A_2, \dots, A_r .

Neste contexto, surge também uma tabela de contingência $r \times c$ da forma apresentada na tabela I, mas com leitura diferente.

Assim, cada O_{ij} ($i=1, \dots, r$ e $j=1, \dots, c$) é uma variável aleatória que representa o número de elementos classificados na categoria A_i de A , na amostra da **população** B_j .

$O_{i.} = \sum_{j=1}^c O_{ij}$ ($i=1, \dots, r$) é uma variável aleatória que representa o número de elementos na categoria A_i de A em todas as amostras.

$O_{.j} = \sum_{i=1}^r O_{ij}$ ($j=1, \dots, c$) é uma constante prefixada (e não uma variável aleatória como acontece no teste de independência), pois é o tamanho da amostra recolhida na população B_j .

Neste caso, cada B_j ($j=1, \dots, c$) rotula uma subpopulação cujos elementos se distribuem pelas r modalidades do atributo A , e o que se pretende saber é se existe homogeneidade, isto é, se não há diferença entre as populações no modo como os seus elementos se distribuem pelas modalidades do atributo A .

Suponhamos, por exemplo, que dispomos dos resultados de vacinação contra a cólera num conjunto de 279 indivíduos escolhidos aleatoriamente entre os vacinados, e num conjunto de 539 indivíduos escolhidos aleatoriamente entre os não vacinados:

	Vacinados	Não Vacinados
Atacados	3	66
Não Atacados	276	473
<i>Totais</i>	279	539

Isto corresponde a ter duas amostras, uma em cada coluna da tabela, obtidas de modo independente e de dimensões, respectivamente $o_{.1}=279$ e $o_{.2}=539$.

A hipótese a testar é a de homogeneidade entre as amostras:

H_0 : os atacados e não atacados distribuem-se de forma idêntica (homogênea) nos vacinados e não vacinados

H_1 : os atacados e não atacados distribuem-se de modo diferente nos vacinados e não vacinados.

As proporções de atacados e não atacados são dadas, respectivamente, por

$$\frac{o_{1.}}{n} = \frac{69}{818} = 0.084 \quad \text{e} \quad \frac{o_{2.}}{n} = \frac{749}{818} = 0.916 .$$

Assim, sob o pressuposto de H_0 ser verdadeira, em cada um dos grupos dos vacinados e não vacinados, deviam ser atacados (não atacados) uma proporção de indivíduos igual a 0.084 (0.916), isto é:

- nos vacinados espera-se que sejam:

$$\text{atacados} \quad \hat{e}_{11} = o_{.1} \times \frac{o_{1.}}{n} = 279 \times 0.084 = 23.44 \quad \text{indivíduos}$$

$$\text{e não atacados} \quad \hat{e}_{21} = o_{.1} \times \frac{o_{2.}}{n} = 279 \times 0.916 = 255.56;$$

- nos não vacinados espera-se que sejam

$$\text{atacados} \quad \hat{e}_{12} = o_{.2} \times \frac{o_{1.}}{n} = 539 \times 0.084 = 45.276 \quad \text{indivíduos}$$

$$\text{e não atacados} \quad \hat{e}_{22} = o_{.2} \times \frac{o_{2.}}{n} = 539 \times 0.916 = 493.724.$$

O quadro seguinte apresenta as frequências esperadas sob o pressuposto de homogeneidade:

	Vacinados	Não Vacinados
Atacados	23.44	45.276
Não Atacados	255.56	493.724
Totais	279	539

À semelhança do teste de independência, a estatística do teste é

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}},$$

que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assintótica do Qui-Quadrado com $(r-1)(c-1)$ graus de liberdade.

As frequências observadas O_{ij} e as estimativas das frequências esperadas \hat{e}_{ij} calculadas sob o pressuposto de H_0 ser verdadeira, devem diferir pouco se H_0 for de facto verdadeira. Assim, valores muito grandes da estatística teste traduzem um grande afastamento dos dados em relação à hipótese nula, conduzindo à rejeição desta. Mais uma vez, a estatística teste mede o afastamento dos dados em relação à hipótese de homogeneidade.

Calculo do valor observado da estatística teste:

$$\chi^2 = \frac{(3 - 23.44)^2}{23.44} + \frac{(66 - 45.276)^2}{45.276} + \frac{(276 - 255.56)^2}{255.56} + \frac{(473 - 493.724)^2}{493.724} = 29.8$$

O quantil de probabilidade 0.995 da distribuição χ_1^2 é 7.88 .

Como o valor observado da estatística teste é $29.8 > 7.88$ então, para um nível de significância 0.005, rejeita-se a hipótese de homogeneidade entre as duas amostras, isto é, a população dos vacinados difere da dos não vacinados no que se refere ao facto de terem ou não sido atacados.