TESTE DE MANN-WHITNEY

A importância deste teste é ser a alternativa não paramétrica ao teste t para a diferença de médias.

Sejam $(X_1,X_2,...,X_n)$ e $(Y_1,Y_2,...,Y_m)$ duas amostras independentes, de tamanhos n e m respectivamente, com $n \leq m$.

Suponhamos que $\mu_X = E(X)$ e $\mu_Y = E(Y)$

Pretende-se testar

$$\mathbf{H_0}: \mu_X = \mu_Y$$

$$\mathbf{H_1}: \mu_X \neq \mu_Y \quad \textbf{ou} \quad \mu_X > \mu_Y \quad \textbf{ou} \quad \mu_X < \mu_Y$$

Procedimentos:

1. Tome-se a amostra conjunta, isto é, sem fazer diferenciação entre os dois grupos, e ordenem-se os valores de 1 até n+m, mas sem perder o grupo de origem de cada observação.

2. Caso **não haja empates** a observação de valor mais baixo recebe o posto 1, a segunda mais baixa recebe o posto 2 e assim sucessivamente.

3. Caso **haja empates** às observações com o mesmo valor (empatadas) atribui-se o posto médio dos postos que lhe corresponderiam casos tais empates não existissem.

Estatística de teste

Supondo $(X_1, X_2, ..., X_i, ..., X_n)$ a **amostra de menor tamanho**.

Seja $R(X_i)$ o posto da observação X_i.

A **estatística de teste** é dada por:

$$T = \sum_{i=1}^{n} R(X_i)$$

Testes bilaterais (H_1: $\mu_X \neq \mu_Y$)

A região crítica será determinada com base na observação, na tabela de Mann-Whitney, de **dois pontos críticos**:

$T_{\alpha/2}$ – directo da tabela

$T_{1-\alpha/2} = nm - T_{\alpha/2}$

Regra de Decisão

Rejeitar H_0 se $T_{obs} < T_{\alpha/2}$ ou $T_{obs} > T_{1-\alpha/2}$

Não rejeitar H_0 se $T_{\alpha/2} \leq T_{obs} \leq T_{1-\alpha/2}$

Testes unilaterais (H_1: $\mu_X < \mu_Y$ ou $\mu_X > \mu_Y$)

A região crítica será determinada com base na observação, na tabela de Mann-Whitney, de **um ponto crítico**:

T_α – directo da tabela

Regra de decisão, para um teste unilateral à esquerda ($\mu_X < \mu_Y$)

Rejeitar H_0 se $T_{obs} < T_\alpha$

Regra de decisão, teste unilateral à direita ($\mu_X > \mu_Y$)

Rejeitar H_0 se $T_{obs} > T_{1-\alpha}$

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

Quando os valores de m e n **são elevados**, a variável aleatória T tem uma **distribuição aproximadamente Normal** com média μ_T e desvio padrão σ_T.

$$\mu_T = \frac{n(m+n+1)}{2} \qquad \text{e} \qquad \sigma_T = \frac{nm(m+n+1)}{12}$$

Isto é, a correspondente variável com valor médio nulo e variância unitária, é

$$\boxed{Z_T = \frac{T-\mu_T}{\sigma_T} \sim N(0,1)}$$

TESTE DE KRUSKALL-WALLIS

O teste de Kruskall-Wallis é uma generalização para **k > 2** amostras, do teste de Mann-Whitney.

A estatística de teste baseia-se nos postos das observações e como tal, a variável em estudo (nos diferentes grupos) é uma variável ordinal.

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

Suponha-se então a existência de **k** populações X_1, X_2, ..., X_k das quais foram retiradas **k** amostras aleatórias

$X_{11}, X_{12}, ..., X_{1n1}$ da população X_1

$X_{21}, X_{22}, ..., X_{1n2}$ da população X_2

... ...

$X_{k1}, X_{k2}, ..., X_{1nk}$ da população X_k

e que existe independência, não só entre os elementos de cada amostra mas também entre os elementos de amostras distintas.

$$\mathbf{H_0: \mu_1 = \mu_2 = ... = \mu_k}$$

$$\mathbf{H_1: \exists\ i, j: \mu_i \neq \mu_j, \quad sendo\ \mu_i = E(X_i),\ i = 1,2,...,k}$$

A **estatística de teste** baseia-se nos postos das observações:

1. Ordenem-se as k amostras conjuntamente. A observação de mais baixo valor tomará o posto 1, a segunda o posto 2 e assim sucessivamente.

2. Caso existam empates, será atribuído o mesmo posto às observações empatadas. Este é a média aritmética dos postos que lhe corresponderiam se tais empates não existissem.

Seja $R(X_{ij})$ o posto atribuído a X_{ij} e

$$R_i = \sum_{j=1}^{n_i} R(X_{ij})$$

a soma dos pontos das observações da i-ésima amostra (i=1,2,...,k).

Seja

$$N = \sum_{i=1}^{k} n_i$$

o número total de observações.

No caso de **não haver empates** a **estatística de teste** de Kruskall-Wallis é:

$$T = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

No caso de **haver empates** a **estatística de teste** é dada por:

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

onde

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

No caso de apenas **três grupos**, em que o **tamanho dos grupos não exceda 5** e **não existam empates**, os valores críticos da estatística do teste encontram-se tabelados.

Nas outras situações, utiliza-se como **distribuição aproximada** o χ^2 com **(k-1) graus de liberdade**, onde k é o número de amostras.

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

TESTES DE AJUSTAMENTO (TESTES DA BONDADE DO AJUSTAMENTO)

Os testes de ajustamento servem para testar a hipótese de que uma determinada amostra aleatória tenha sido extraída de uma população com distribuição especificada.

Seja $X_1, X_2, ..., X_n$ uma amostra aleatória de uma população X com função (densidade) de probabilidade f desconhecida e f_0 a função (densidade) de probabilidade proposta.

Hipóteses a testar:

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

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

Exemplo 1:

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

Número de unidades	Número de dias
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 de 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**
- o teste de normalidade de **Lilliefors**

TESTE DO QUI-QUADRADO

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, ... , 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., $p_{0i} = P(A_i / H_0)$.

Hipóteses a testar:

H_0: $p_i = p_{0i}$, $i=1,...,m$

H_1: $p_i \neq p_{0i}$ 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 assimptótica 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.

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

Se **a hipótese nula for 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 de 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**.

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.

Exemplo 1:

Número de unidades	Número de dias
0	6
1	14
2	10
3	7
4	2
5	1

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

Representando por X a procura diária do produto e por f a função de probabilidade de X, as hipóteses a testar são

$$\mathbf{H_0}: X \sim P(\mu) \qquad \left(f(x) = f_0(x) = \frac{e^{-\mu}\,\mu^{x}}{x!},\quad x = 0,1,2,\ldots \quad e \quad \mu>0\right)$$

$$\mathbf{H_1}: X \not\sim P(\mu)$$

É necessário estimar o parâmetro μ, média da Poisson:

$$\bar{x} = (0\times6 + 1\times14 + 2\times10 + 3\times7 + 4\times2 + 5\times1) / 40 = 1.7$$

Classes	Frequências observadas $\mathbf{O_i}$		$\mathbf{p_{0i}} = P(A_i / H_0)$		Frequências esperadas $\mathbf{e_i} = 40 \times p_{0i}$	
$A_1=\{0\}$	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	3	0.0216	0.0932	0.864	3.728
$A_7=\{6,7,\ldots\}$	0		0.008		0.32	

Probabilidades associadas a cada uma das classes supondo H_0 verdadeira:

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

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

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

$$\vdots$$

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^2_3 é 7.81, e logo a região crítica é $[7.81, +\infty[$.

Valor observado da estatística de teste:

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

$$= 0.778$$

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.

Exemplo 2:

Denotando por X a diferença entre a data de chegada dos navios e a data planeada, as hipóteses a testar são

$$\mathbf{H_0}: X \sim N(0.1, 7.2^2)$$
$$\mathbf{H_1}: X \not\sim N(0.1, 7.2^2)$$

Neste caso a distribuição proposta em H_0 é contínua e, deste modo, as classes A_i, i=1,...m, são intervalos da forma

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

Para a determinação das classes é sugerida a **regra de Mann e Wald**:

Número de classes = **m**, com m tal que n/m > 5. Os limites dos intervalos são tais que as probabilidades decorrentes da hipótese nula sejam iguais a **1/m** para todas as classes. Assim, as frequências esperadas são todas iguais a n/m > 5.

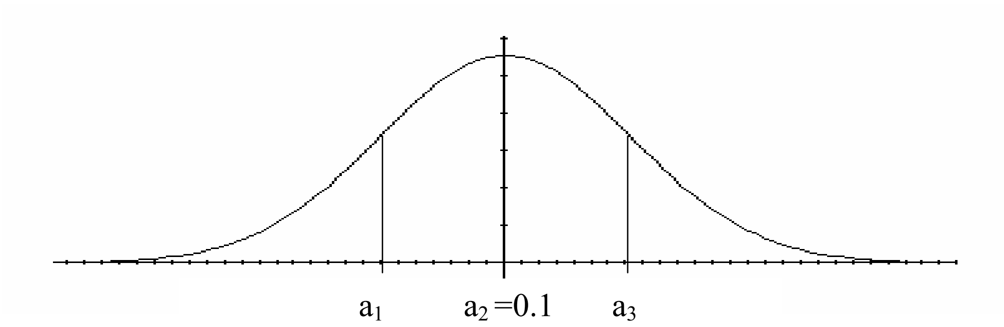
Para o exemplo escolheu-se m = 4 classes ($e_i = 30\times1/4 = 7.5 > \mathbf{5}$), donde

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

Cálculo dos limites dos intervalos de classe:

a_3: $p_{03} = P(X \in A_3 \,/\, X \sim N(0.1,\, 7.2^2)) = 0.25 \Leftrightarrow P(X < a_3 \,/\, X \sim N(0.1,\, 7.2^2)) = 0.75$

$\Leftrightarrow\ P(Z < \frac{a_3 - 0.1}{7.2}) = 0.75 \Leftrightarrow a_3 = 4.924;$



Da simetria da distribuição normal:

$a_2 = 0.1$ e $a_1 = 0.1-(4.924-0.1) = -4.724$

-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.

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

O valor observado da estatística de teste é

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

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.

Para $\alpha=0.05$: R.C. = $[7.81, +\infty[$

Como $Q_{obs} \notin$R.C., somos levados a não rejeitar a hipótese de que a diferença entre os tempos de chegada e os tempos planeados tem distribuição $N(0.1, 7.2^2)$.

TESTE DE KOLMOGOROV-SMIRNOV

O teste de Kolmogorov-Smirnov (**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 se pretender, por exemplo, efectuar um ajustamento de uma distribuição normal, sem especificar μ e σ, deve-se recorrer a outro teste, neste caso o teste desenvolvido por Lilliefors (teste de normalidade de Lilliefors).

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

FUNÇÃO DE DISTRIBUIÇÃO EMPÍRICA E FUNÇÃO DE DISTRIBUIÇÃO DA AMOSTRA

Seja $(X_1, X_2,...,X_n)$ uma amostra aleatória de uma certa população X e $(x_1, x_2,...,x_n)$ 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,...,X_n), por

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

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,...,X_n), ou seja, é uma estatística.

Exemplo

Consideremos 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

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

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

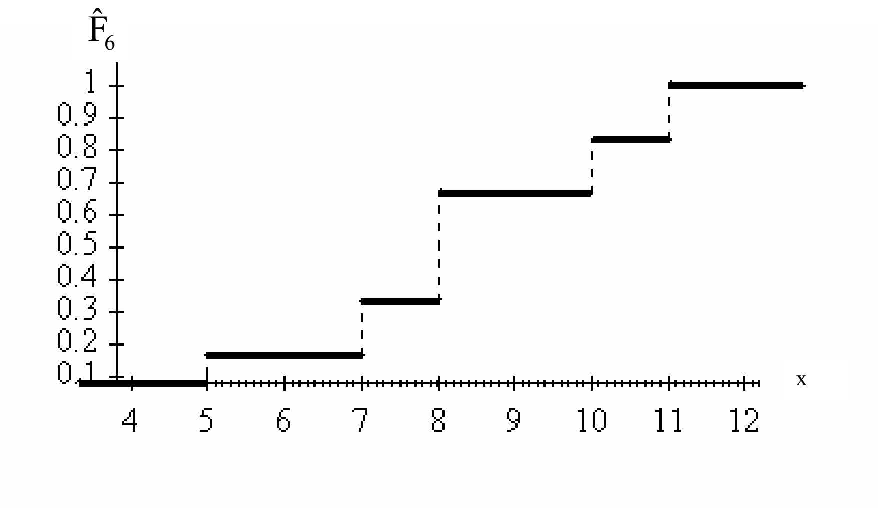


Gráfico I: Representação gráfica da f.d. empírica.

Seja F a função de distribuição da população e F_0 a função de distribuição proposta, contínua e completamente especificada.

Hipóteses a testar

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

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

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.

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:

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

Uma vez que F_0 é uma função (contínua) crescente e $\hat{F}_n$ é uma função em escada, 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,

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

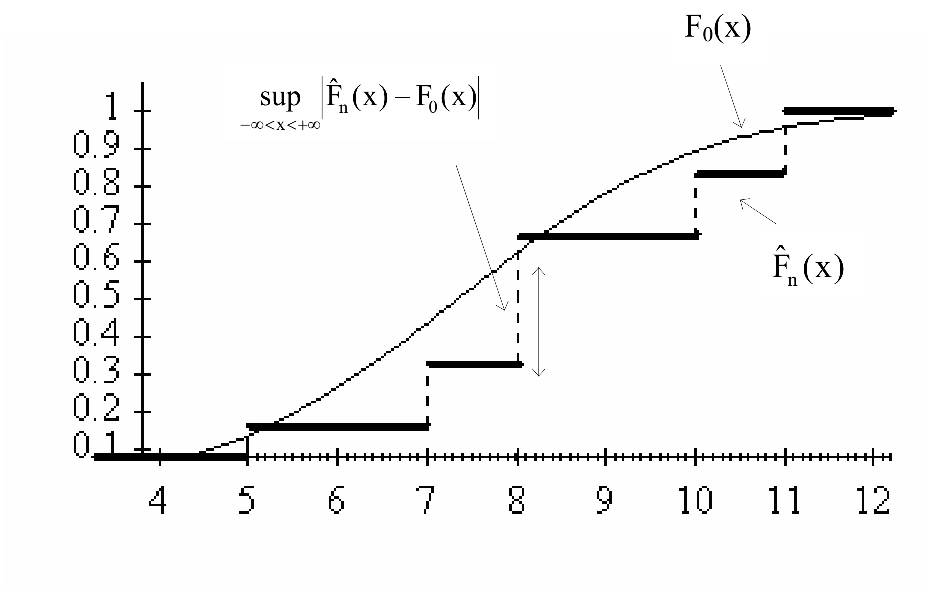


Gráfico II: Ajustamento de uma f.d hipotética F_0 à f.d. empírica $\hat{F}_n$

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.

Então, 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,

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

Os valores críticos $D_{n,\alpha}$ podem ser consultados numa tabela.

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,

$$H_0: X \sim N(290, 56^2)$$

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

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 $\alpha = 0.05$, $D_{10,0.05} = 0.409$

Cálculo do valor observado da estatística D_{10}

x_i	$\hat{F}_{10}(x_i)$	$\hat{F}_{10}(x_i^-)$	$z_i=\frac{x_i-290}{56}$	$F_0(x_i)=P(Z\leq z_i)$ $(Z\sim N(0,1))$	$\lvert F_0(x_i)-\hat{F}_{10}(x_i)\rvert$	$\lvert F_0(x_i)-\hat{F}_{10}(x_i^-)\rvert$
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

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

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

Como $d_{10} = 0.484 > 0.409$, ao nível de significância de 5%, rejeitamos a hipótese de a quantidade de carne servida por refeição a cada estudante seguir distribuição $N(290,56^2)$.

TABELAS DE CONTINGÊNCIA

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 $A_1, A_2, \ldots, A_r$ e $B_1, B_2, \ldots, B_s$.

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×s**, com o seguinte aspecto:

	B_1	B_2	...	B_s
A_1	O_{11}	O_{12}	...	O_{1s}
A_2	O_{21}	O_{22}	...	O_{2s}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_r	O_{r1}	O_{r2}	...	O_{rs}

Nesta tabela cada $\mathbf{O_{ij}}$ (i=1,...,r e j=1,...,s) 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:

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

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

Tem-se,

$$n = \sum_{i=1}^{r}\sum_{j=1}^{s} O_{ij} = \sum_{i=1}^{r} O_{i\cdot} = \sum_{j=1}^{s} O_{\cdot j}$$

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.

Hipóteses a testar:

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,..,r e j=1,...,s) a probabilidade (desconhecida) de um indivíduo da população ser classificado simultaneamente nas categorias A_i de A e B_j de B;

- $p_{i\cdot} = P(A_i)$ (i=1,...,r) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria A_i de A;

- $p_{\cdot j} = P(B_j)$ (j=1,...,s) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria B_j de B.

Tem-se,

$$1 = \sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} = \sum_{i=1}^{r} p_{i\cdot} = \sum_{j=1}^{s} p_{\cdot j}$$

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,

$$p_{ij} = p_{i\cdot}\, p_{\cdot j}$$

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

H_0: $p_{ij} = p_{i\cdot}\, p_{\cdot j}$ (para todo i e j)

H_1: $p_{ij} \neq p_{i\cdot}\, p_{\cdot j}$ (para algum i≠j)

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

Os verdadeiros valores das probabilidades $p_{i\cdot}$ e $p_{\cdot j}$ são estimadas, a partir dos dados amostrais, por

$$\hat{p}_{i\cdot} = \frac{o_{i\cdot}}{n} \qquad \text{e} \qquad \hat{p}_{\cdot j} = \frac{o_{\cdot j}}{n}$$

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

$e_{ij} = n\, p_{ij}$ → número esperado de indivíduos na classe A_i de A e B_j de B.

Quando H_0 é verdadeira, i.e, $p_{ij} = p_{i\cdot}\, p_{\cdot j}$, temos

$$e_{ij} = n\, p_{ij} = n\, p_{i\cdot}\, p_{\cdot j} \xrightarrow{\text{estimado por}} \hat{e}_{ij} = n\, \hat{p}_{i\cdot} \hat{p}_{\cdot j}$$

A **estatística do teste** de independência é

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

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

Vimos que quando H_0 é verdadeira e_{ij} pode ser estimado por $\hat{e}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j}$, 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.

Trata-se então de um **teste unilateral à direita**.

Exemplo 4

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:

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

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

H_0: A e B são independentes

H_1: A e B não são independentes

Uma vez que A e B assumem cada uma 3 modalidades, sob H_0, a estatística teste tem distribuição assintótica do Qui-quadrado com (r-1)(s-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:

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

Assim, por exemplo,

$$\hat{e}_{11} = (3000\times1000)/4000 = 750$$

$$\hat{e}_{12} = (3000\times2000)/4000 = 1500$$

$$\hat{e}_{13} = (3000\times1000)/4000 = 750$$

Frequências esperadas

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

Valor observado da estatística teste:

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

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.

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.

- **COEFICIENTE 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.

- **COEFICIENTE DE TSHUPROW:**

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

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

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

- **COEFICIENTE V DE CRAMER:**

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

Exemplo 4

Neste exemplo, 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.

TESTE DE HOMOGENEIDADE

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

Neste contexto, surge também uma tabela de contingência r×s da forma apresentada na tabela I, mas com leitura diferente.

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

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

$O_{\cdot j} = \sum_{i=1}^{r} O_{ij}$ (j=1,...,s) 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 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**.

Exemplo 5

Suponhamos 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
Totais	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.

Hipóteses a testar:

H_0: os atacados e não atacados distribuem-se de forma idêntica (homogénea) 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

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

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:

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

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

- nos não vacinados espera-se que sejam

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

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

TESTES DE HIPÓTESES NÃO PARAMÉTRICOS

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

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

que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assintótica do **Qui-Quadrado com (r-1)(s-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**.

Cálculo do valor observado da estatística teste:

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

O quantil de probabilidade 0.995 da distribuição χ^2_1 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.