DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidades e Estatística / Introd. às Probabilidades e Estatística** TODOS OS CURSOS	Exame Época Especial 2018/2019 26/07/2019 – **09:00**

Duração: 3 horas

Justifique convenientemente todas as respostas

Grupo I	5 valores

1. Uma aplicação prevê *sol* ou *chuva* para o dia seguinte numa dada região. Sabe-se que em 2% (resp. 24%) dos casos prevê-se *sol* (resp. *chuva*) e *chove* (resp. faz *sol*) no dia seguinte. Para além disso, em 10% (resp. 64%) dos casos prevê-se *chuva* (resp. *sol*) e *chove* (resp. faz *sol*) no dia seguinte.

(a) Qual é a probabilidade de a aplicação indicar uma previsão incorreta? (1.5)

- **Quadro de acontecimentos e probabilidades**

Acontecimento	**Probabilidade**
$C = \{\text{dia com chuva}\}$	$P(C) = ?$
$S = \{\text{dia com sol}\}$	$P(S) = ?$
$PC = \{\text{previsão de dia com chuva}\}$	$P(PC) = ?$
$PS = \{\text{previsão de dia com sol}\}$	$P(PS) = ?$
	$P(PS \cap C) = 0.02$
	$P(PC \cap S) = 0.24$
	$P(PS \cap S) = 0.64$
	$P(PC \cap C) = 0.10$

- **Probabilidade pedida**

$$\begin{aligned} P(\text{previsão incorreta}) &= P(PS \cap C) + P(PC \cap S) \\ &= 0.02 + 0.24 \\ &= 0.26. \end{aligned}$$

(b) Determine a probabilidade de fazer sol num dia em que a previsão da aplicação tenha sido de chuva? (1.0)

- **Probabilidade pedida**
Recorrendo à definição de probabilidade condicionada, temos:

$$\begin{aligned} P(S \mid PC) &= \frac{P(PC \cap S)}{P(PC)} \\ &= \frac{P(PC \cap S)}{P(PC \cap S) + P(PC \cap C)} \\ &= \frac{0.24}{0.24 + 0.10} \\ &= \frac{0.24}{0.34} \\ &\simeq 0.705882. \end{aligned}$$

2. Uma jornalista entrevista eleitores, escolhidos ao acaso, à saída de um local de voto até encontrar o primeiro que declara ter votado em determinado partido político. Considere que tal partido político obteve 3% dos votos nas eleições em causa e que nenhum dos eleitores mente à jornalista.

(a) Seja X a variável aleatória que representa o número de entrevistas efetuadas pela jornalista. Calcule $P(X > 7 \mid X > 3)$. (1.0)

- **V.a. de interesse**
  X = no. de entrevistas até se encontrar um primeiro eleitor com o perfil desejado

- **Distribuição de X**
  $X \sim \text{Geométrica}(p)$, onde $p = 0.03$.

- **F.p. de X**
  $P(X = x) = (1-0.03)^{x-1}\,0.03, \quad x = 1,2,\ldots$

- **Probabilidade pedida**
  Pela propriedade de falta de memória da distribuição geométrica segue-se

$$\begin{aligned}
P(X > 3+4 \mid X > 3) &= P(X > 4) \\
&= 1 - P(X \leq 4) \\
&= 1 - \sum_{x=1}^{4} (1-p)^{x-1}\, p \\
&= 1 - p\,\frac{1-(1-p)^4}{1-(1-p)} \\
&= (1-p)^4 \\
&= (1-0.03)^4 \\
&\simeq 0.885293.
\end{aligned}$$

[Alternativamente,

$$\begin{aligned}
P(X > 3+4 \mid X > 3) &= \frac{P(X > 3+4, X > 3)}{P(X > 3)} \\
&= \frac{P(X > 3+4)}{P(X > 3)} \\
&= \frac{0.97^{3+4}}{0.97^3} \\
&= 0.95^4 \quad \equiv P(X > 4) \\
&\simeq 0.885293.]
\end{aligned}$$

(b) Seja Y a variável aleatória que representa o número de entrevistas realizadas pela jornalista antes de encontrar o primeiro eleitor com o perfil desejado. Calcule o valor esperado e o desvio padrão de Y. (1.5)

- **Variável aleatória de interesse**
  Y = no. de entrevistas infrutíferas até se encontrar o primeiro eleitor com o perfil desejado
  $= X - 1$

- **Valor esperado e desvio padrão pedidos**

$$\begin{aligned}
E(Y) &= E(X-1) \\
&= E(X) - 1 \\
&\stackrel{form.}{=} \frac{1}{p} - 1 \\
&= 32.3(3)
\end{aligned}$$

$$\begin{aligned}
\sqrt{V(Y)} &= \sqrt{V(X-1)} \\
&= \sqrt{V(X)} \\
&\stackrel{form.}{=} \sqrt{\frac{1-p}{p^2}} \\
&\simeq 32.829526.
\end{aligned}$$

Grupo II	5 valores

1. Admita que a variável aleatória X representa a proporção de impurezas em determinado produto de um processo químico e possui função de distribuição dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-(1-x)^{10}, & 0 \le x \le 1 \\ 1, & \text{caso contrário.} \end{cases}$$

(a) Qual é a probabilidade de a proporção de impurezas exceder 0.1 ? (0.5)

- **V.a. de interesse**
  X = proporção de impurezas em determinado produto
- **F.d. de X**
  $$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-(1-x)^{10}, & 0 \le x \le 1 \\ 1, & \text{caso contrário.} \end{cases}$$
- **Prob. pedida**
  $$\begin{aligned} P(X > 0.1) &= 1 - F_X(0.1) \\ &= 1 - [1-(1-0.1)^{10}] \\ &= 0.9^{10} \\ &\simeq 0.348678. \end{aligned}$$

(b) Obtenha o terceiro quartil de X. (1.0)

- **Terceiro quartil de X**
  Será representado por $\chi \equiv F_X^{-1}(3/4)$ e
  $$\begin{aligned} \chi \in (0,1) \quad : \quad & F_X(\chi) = \frac{3}{4} \\ & 1-(1-\chi)^{10} = 0.75 \\ & 1-\chi = 0.25^{\frac{1}{10}} \\ & \chi = 1 - 0.25^{\frac{1}{10}} \\ & \chi \simeq 0.129449. \end{aligned}$$

(c) Sabendo que $E(X) = \frac{1}{11}$ e $V(X) = \frac{5}{726}$, determine um valor aproximado para a probabilidade de a proporção média de impurezas em 50 desses produtos exceder 0.1. Suponha que as proporções de impurezas em tais produtos são variáveis aleatórias independentes e identicamente distribuídas a X. (1.5)

- **V.a.**
  X_i = proporção de impurezas no produto i, onde $i = 1, \ldots, n$
  $n = 50$
- **Distribuição, valor esperado e variância comuns**
  $X_i \overset{i.i.d.}{\sim} X, \quad i = 1, \ldots, n$
  $E(X_i) = E(X) = \mu = \frac{1}{11}, \quad i = 1, \ldots, n$
  $V(X_i) = V(X) = \sigma^2 = \frac{5}{726}, \quad i = 1, \ldots, n$
- **V.a. de interesse**
  $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ = média de impurezas em n produtos
- **Valor esperado e variância de $\bar{X}$**
  $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) \overset{X_i \sim X}{=} \frac{1}{n} \times nE(X) = E(X) = \mu$
  $V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \overset{X_i\, indep.}{=} \frac{1}{n^2} \times \sum_{i=1}^n V(X_i) \overset{X_i \sim X}{=} \frac{1}{n^2} \times nV(X) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$
- **Distribuição aproximada de $\bar{X}$**
  Pelo teorema do limite central (TLC) pode escrever-se

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned}
P(\bar{X} > 0.1) &= 1 - P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{0.1-\mu}{\frac{\sigma}{\sqrt{n}}}\right) \\
&\overset{TLC}{\simeq} 1 - \Phi\left(\frac{0.1 - \frac{1}{11}}{\sqrt{\frac{5}{726\times 50}}}\right) \\
&\simeq 1 - \Phi(0.77) \\
&\overset{tabela/calc}{\simeq} 1 - 0.7794 \\
&= 0.2206.
\end{aligned}$$

2. Na construção de um índice global da desigualdade de género em determinado país, considerou-se o par aleatório (X,Y), em que: X mede a desigualdade de género no que respeita à participação na atividade política desse país; Y traduz aspectos económicos e de oportunidade na atividade laboral entre homens e mulheres nesse país. Admita que a função de densidade de probabilidade conjunta de (X,Y) é dada por

$$f_{X,Y}(x,y) = \begin{cases} 60\,x^2\,y, & 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad x+y \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Calcule a função de densidade de probabilidade marginal de Y. (1.0)

- **Par aleatório** (X,Y)
  X = indicador da desigualdade de género no que respeita à participação na atividade política
  Y = indicador relativo a aspectos económicos etc. entre homens e mulheres

- **F.d.p. do par aleatório** (X,Y)

$$f_{X,Y}(x,y) = \begin{cases} 60\,x^2\,y, & 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad x+y \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

- **F.d.p. marginal de** Y

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx \\
&= \begin{cases} \int_0^{1-y} 60\,x^2\,y\,dx = 20\,y\,x^3\big|_0^{1-y} = 20\,y\,(1-y)^3, & 0 \leq y \leq 1 \\ 0, & \text{caso contrário} \end{cases}
\end{aligned}$$

(b) Obtenha $E(X \mid Y = 0.5)$. (1.0)

- **F.d.p. de** X **condicional a** $Y = 0.5$

$$\begin{aligned}
f_{X|Y=0.5}(x) &= \frac{f_{X,Y}(x,0.5)}{f_Y(0.5)} \\
&= \begin{cases} \frac{60\,x^2 \times 0.5}{20 \times 0.5 \times (1-0.5)^3} = \frac{3\,x^2}{(1-0.5)^3} = 24x^2, & 0 \leq x \leq 1 - 0.5 = 0.5 \\ 0, & \text{caso contrário} \end{cases}
\end{aligned}$$

- **Valor esperado de** X **condicional a** $Y = 0.5$

$$\begin{aligned}
E(X \mid Y = 0.5) &= \int_{-\infty}^{+\infty} x \times f_{X|Y=0.5}(x)\,dx \\
&= \int_0^{0.5} x \times 24x^2\,dx \\
&= 6x^4\big|_0^{0.5} = 6 \times 0.5^4 = 0.375.
\end{aligned}$$

Grupo III	5 valores

1. A variável aleatória X representa a proporção anual de área ardida em dada região e possui função de densidade de probabilidade

$$f_X(x) = \begin{cases} (\beta+1)\,x^{\beta}, & 0 < x < 1 \\ 0, & \text{caso contrário,} \end{cases}$$

onde β é um parâmetro desconhecido tal que $\beta > -1$.

(a) Mostre que o estimador de máxima verosimilhança de β, com base numa amostra aleatória $(X_1,\ldots,X_n)$, é dado por $-\frac{n}{\sum_{i=1}^n \ln(X_i)} - 1$. (1.5)

- **V.a. de interesse**
  X = proporção anual de área ardida em dada região
- **F.d.p. de** X
  $$f_X(x) = \begin{cases} (\beta+1)\,x^{\beta}, & 0 < x < 1 \\ 0, & \text{caso contrário,} \end{cases}$$
- **Parâmetro desconhecido**
  β
- **Amostra**
  $\underline{x} = (x_1,\ldots,x_n)$ é uma amostra de dimensão n proveniente da população X.
- **Obtenção do estimador de MV de** β

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\beta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\, indep}{=} \prod_{i=1}^n f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \left[(\beta+1)\,x_i^{\beta}\right] \\ &= (\beta+1)^n \left[\prod_{i=1}^n x_i\right]^{\beta}, \quad \beta > -1 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\beta \mid \underline{x}) = n\ln(\beta+1) + \beta \sum_{i=1}^n \ln(x_i)$$

  Passo 3 — Maximização
  A estimativa de MV de β é doravante representada por $\hat{\beta}$ e

$$\hat{\beta} \; : \; \begin{cases} \left.\frac{d\ln L(\beta|\underline{x})}{d\beta}\right|_{\beta=\hat{\beta}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\beta|\underline{x})}{d\beta^2}\right|_{\beta=\hat{\beta}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \frac{n}{\hat{\beta}+1} + \sum_{i=1}^n \ln(x_i) = 0 \quad \Leftrightarrow \quad \hat{\beta} = -\frac{n}{\sum_{i=1}^n \ln(x_i)} - 1 \\ -\frac{n}{(\hat{\beta}+1)^2} < 0 \quad \text{(proposição verdadeira já que } n \in \mathbb{N}) \end{cases}$$

  Passo 4 — Estimador de MV de β

$$EMV(\beta) = -\frac{n}{\sum_{i=1}^n \ln(X_i)} - 1$$

(b) Obtenha a estimativa de máxima verosimilhança de $P(X > 0.1) = 1 - 0.1^{\beta+1}$, tendo em conta uma amostra $(x_1, \ldots, x_6)$ para a qual $\sum_{i=1}^{6} \ln(x_i) \simeq -20.2941$. (0.5)

- **Estimativa de MV de β**

$$\begin{aligned} \hat{\beta} &= -1 - \frac{n}{\sum_{i=1}^{n} \ln(x_i)} \\ &= -1 - \frac{6}{-20.2941} \\ &\simeq -0.704347 \end{aligned}$$

- **Outro parâmetro desconhecido**

$$\begin{aligned} h(\beta) &= P(X > 0.1) \\ &= 1 - 0.1^{\beta+1} \end{aligned}$$

- **Estimativa de MV de $h(\beta)$**
Invocando a propriedade de invariância dos estimadores de máxima verosimilhança, pode concluir-se que a estimativa de MV de $h(\beta)$ é dada por

$$\begin{aligned} \widehat{h(\beta)} &= h(\hat{\beta}) \\ &= 1 - 0.1^{\hat{\beta}+1} \\ &= 1 - 0.1^{-0.704347+1} \\ &\simeq 0.493771. \end{aligned}$$

2. De modo a comparar a autonomia de baterias de iões de lítio de 2 marcas, escolheram-se ao acaso várias baterias de cada marca e registaram-se as durações (em hora) das suas cargas. A informação recolhida encontra-se sumariada na tabela seguinte:

Marca	Nº de baterias	Média amostral
1	40	3.5
2	50	3.2

Admita que as durações das cargas de baterias das marcas 1 e 2 são variáveis aleatórias independentes X_1 e X_2 (respetivamente), as quais possuem distribuição normal, com valores esperados desconhecidos μ_1 e μ_2 e desvios padrão conhecidos e iguais a 0.5.

(a) Obtenha um intervalo de confiança a 95% para $\mu_1 - \mu_2$. (1.5)

- **V.a. de interesse**
X_i = duração (em horas) da carga da bateria da marca i, $i = 1, 2$

- **Situação**
$X_i \stackrel{indep.}{\sim} \text{normal}(\mu_i, \sigma_i^2)$, $\quad i = 1, 2$
$(\mu_1 - \mu_2)$ DESCONHECIDO
$\sigma_1 = \sigma_2 = 0.5$ conhecido
$n_1 = 40$, $\quad n_2 = 50$

- **Obtenção do IC para $\mu_1 - \mu_2$**
Passo 1 — Selecção da v.a. fulcral para $\mu_1 - \mu_2$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{normal}(0, 1)$$

[dado que se pretende determinar um IC para a diferença de valores esperados de duas populações independentes com distribuições normais e com variâncias conhecidas.]

Passo 2 — Obtenção dos quantis de probabilidade
Ao ter-se em consideração que $(1 - \alpha) \times 100\% = 95\%$, far-se-á uso dos quantis

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = \Phi^{-1}(0.025) = -\Phi^{-1}(1-0.025) \stackrel{tabela/calc.}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) \stackrel{tabela/calc.}{=} 1.9600. \end{cases}$$

Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$P\left[a_\alpha \leq \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[(\bar{X}_1-\bar{X}_2) - b_\alpha \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq \mu_1-\mu_2 \leq (\bar{X}_1-\bar{X}_2) - a_\alpha \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] = 1-\alpha$$

Passo 4 — Concretização

Tendo em conta que

$$IC_{(1-\alpha)\times 100\%}(\mu_1-\mu_2) = \left[(\bar{x}_1-\bar{x}_2) \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

e atendendo aos valores dos quantis acima e de n_i, $\bar{x}_i$, σ_i^2 $(i=1,2)$, segue-se

$$\begin{aligned} IC_{95\%}(\mu_1-\mu_2) &= \left[(3.5-3.2) \pm 1.9600 \times \sqrt{\frac{0.5^2}{40}+\frac{0.5^2}{50}}\right] \\ &\simeq [0.3 \pm 1.9600 \times 0.106066] \\ &\simeq [0.092111,\ 0.507889]. \end{aligned}$$

(b) Confronte as hipóteses $H_0 : \mu_1 = \mu_2$ e $H_1 : \mu_1 > \mu_2$. Decida com base no valor-p. (1.5)

- **V.a. de interesse e situação**
  Ver alínea (a).

- **Hipóteses**
  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$
  $H_1 : \mu_1 - \mu_2 > \mu_0$

- **Estatística de teste**

$$T = \frac{(\bar{X}_1-\bar{X}_2)-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim_{H_0} \text{normal}(0,1)$$

  [uma vez que se pretende efectuar um teste sobre a diferença de valores esperados de duas populações independentes com distribuições normais e com variâncias conhecidas.]

- **Região de rejeição de** H_0 (para valores de T)
  Estamos a lidar com um teste unilateral superior ($H_1 : \mu_1 - \mu_2 > \mu_0$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**
  O valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \\ &= \frac{(3.5-3.2)-0}{\sqrt{\frac{0.5^2}{40}+\frac{0.5^2}{50}}} \\ &\simeq 2.83. \end{aligned}$$

e a região de rejeição deste teste é um intervalo à direita. Donde

$$\begin{aligned} valor-p &= P(T > t \mid H_0) \\ &= 1-\Phi(t) \\ &\simeq 1-\Phi(2.83) \\ &\overset{tabela/calc.}{=} 1-0.9977 \\ &= 0.0023. \end{aligned}$$

Logo é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.23\%$;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 0.23\%$ (e.g., aos n.u.s., 1%, 5% e 10%).

Grupo IV — 5 valores

1. Conjetura-se que a variável aleatória X, que representa o coeficiente de inteligência (QI), segue uma distribuição normal com valor esperado e desvio padrão iguais a 105 e 15, respetivamente. Depois de se testarem 50 indivíduos selecionados ao acaso, obteve-se a seguinte tabela de frequências:

Classe de QI	≤ 90	$]90,100]$	$]100,110]$	$]110,120]$	> 120
Frequência absoluta observada	7	10	14	11	8
Frequência absoluta esperada sob H_0	E_1	10.600	12.930	10.600	E_5

(a) Obtenha as frequências absolutas esperadas sob $H_0 : X \sim \text{normal}(105, 15^2)$, representadas na tabela acima por E_1 e E_5. Aproxime-as às milésimas. (0.5)

- **V.a. de interesse**
  $X =$ coeficiente de inteligência (QI)
- **Dist. conjecturada**
  $\text{normal}(105, 15^2)$
- **Frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n = 50$ e à dist. conjecturada, segue-se:

$$\begin{aligned} E_1 &= 50 \times \Phi\left(\frac{90-105}{15}\right) \\ &= 50 \times \Phi(-1) \\ &= 50 \times [1-\Phi(1)] \\ &\overset{tabela/calc.}{=} 50 \times (1-0.8413) \\ &= 7.935 \\ E_5 &= n - \sum_{i=1}^{4} E_i \\ &= 50-(7.935+10.600+12.930+10.600) \\ &= 7.935. \end{aligned}$$

(b) Teste H_0, ao nível de significância de 5%. (1.5)

- **Hipóteses**
  $H_0 : X \sim \text{normal}(105, 15^2)$
  $H_1 : X \not\sim \text{normal}(105, 15^2)$
- **Nível de significância**
  $\alpha_0 = 5\%$
- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)},$$

onde:

k = No. de classes = 5

O_i = Frequência absoluta observável da classe i

E_i =Frequência absoluta esperada, sob H_0, da classe i

β = No. de parâmetros a estimar = 0 [dado que em H_0 se conjectura uma distribuição específica.]

- **Frequências absolutas esperadas sob H_0**
De acordo com a tabela facultada e (a), os valores das freq. absolutas esperadas sob H_0 aproximados às milésimas são: $E_1 \simeq 7.935$; $E_2 \simeq 10.600$; $E_3 \simeq 12.930$; $E_4 \simeq 10.600$; $E_5 \simeq 7.935$.
[Constata-se que não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e que $E_i \geq 1$ para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

- **Região de rejeição de H_0** (para valores de T)
Tratando-se de um teste de ajustamento, a região de rejeição de H_0 escrita para valores de T é o intervalo à direita $W = (c, +\infty)$, onde

$$
\begin{aligned}
c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
&= F^{-1}_{\chi^2_{(5-0-1)}}(1-0.05) \\
&= F^{-1}_{\chi^2_{(4)}}(0.95) \\
&\overset{tabela/calc.}{=} 9.488.
\end{aligned}
$$

- **Decisão**
No cálculo do valor obs. da estat. de teste convém recorrer à seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$]-\infty, 90]$	7	7.935	$\frac{(7-7.935)^2}{7.935} \simeq 0.110$
2	$]90, 100]$	10	10.600	0.034
3	$]100, 110]$	14	12.930	0.089
4	$]110, 120]$	11	10.600	0.015
5	$]120, +\infty[$	8	7.935	0.001
		$\sum_{i=1}^k o_i = n$ $= 50$	$\sum_{i=1}^k E_i = n$ $= 50$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i}$ $\simeq 0.249$

Como $t \simeq 0.249 \not\in W = (9.488, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [nem a qualquer outro n.s. inferior a α_0].

2. Um engenheiro químico deseja verificar se um instrumento para medir a concentração de determinada substância no sangue está bem calibrado. Com esse objectivo seleccionou ao acaso 15 amostras com concentrações conhecidas (x), determinou as respectivas concentrações recorrendo ao instrumento (Y) e obteve os seguintes resultados:

$$\bar{x} = 6, \quad \sum_{i=1}^{15} x_i^2 - 15\,\bar{x}^2 = 120, \quad \bar{y} = 6.04, \quad \sum_{i=1}^{15} y_i^2 - 15\,\bar{y}^2 = 116.156, \quad \sum_{i=1}^{15} x_i y_i - 15\,\bar{x}\bar{y} = 117.6.$$

Admitindo que os erros aleatórios associados ao modelo de regressão linear simples de Y em x satisfazem $\epsilon_i \overset{i.i.d.}{\sim} \text{normal}(0, \sigma^2)$, $i = 1, \ldots, 15$:

(a) Determine as estimativas de máxima verosimilhança dos coeficientes β_0 e β_1. (1.0)

- **Estimativas de MV de β_1 e β_1**
  Dado que

  $$n = 15$$
  $$\bar{x} = 6$$
  $$\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 120$$
  $$\bar{y} = 6.04$$
  $$\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 = 116.156$$
  $$\sum_{i=1}^{n} x_i\, y_i - n\,\bar{x}\,\bar{y} = 117.6,$$

  as estimativas de MV de β_1 e β_0 são, para este modelo de RLS, iguais a:

  $$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i\,y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\ &= \frac{117.6}{120} \\ &\simeq 0.98 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\ &\simeq 6.04 - 0.98 \times 6 \\ &\simeq 0.16.\end{aligned}$$

(b) Confronte $H_0 : \beta_1 = 1$ e $H_1 : \beta_1 \neq 1$, ao nível de significância de 1%. (1.5)

- **Hipóteses**
  $H_0 : \beta_1 = \beta_{1,0} = 1$
  $H_1 : \beta_1 \neq 1$

- **Estatística de teste**

  $$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)
  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq 1$), pelo que a região de rejeição de H_0 é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 \mid H_0) = 0.01$, i.e.,

  $$c = F^{-1}_{t_{(15-2)}}(1 - 0.01/2) = F^{-1}_{t_{(13)}}(0.995) \overset{tabela/calc.}{=} 3.012.$$

- **Decisão**
  Tendo em conta os valores obtidos em (a), bem como o valor de

  $$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2 \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\ &= \frac{1}{15-2}\left(116.156 - 0.98^2 \times 120\right) \\ &\simeq 0.069846.\end{aligned}$$

  concluímos que o valor observado da estatística de teste é igual a

  $$\begin{aligned}t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \\ &\simeq \frac{0.98 - 1}{\sqrt{\frac{0.069846}{120}}} \\ &= -0.828991,\end{aligned}$$

Como $t_0 \simeq -0.828991 \notin W = (-\infty, -3.012) \cup (3.012, +\infty)$ não devemos rejeitar H_0 ao n.s. de 1% [nem a qualquer outro n.s. inferior a 1%].

(c) Calcule e comente o valor do coeficiente de determinação do modelo ajustado. (0.5)

- **Cálculo do coeficiente de determinação**

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{117.6^2}{120 \times 116.156} \\ &\simeq 0.9922. \end{aligned}$$

- **Interpretação do coeficiente de determinação**

  Cerca de 99.22% da variação total de Y é explicada pela variável x, através do modelo de regressão linear simples considerado. A recta estimada ajusta-se muito bem ao conjunto de dados.

  [Mais, este resultado e o do teste em (b) leva a crer que o instrumento está bem calibrado.]