DM
DEPARTAMENTO
DE MATEMÁTICA
TÉCNICO LISBOA

Introdução à Probabilidade e Estatística

MA

2º Trimestre – 2024/2025
22/01/2025 – 15:30, V1.09
Recurso – 120 minutos

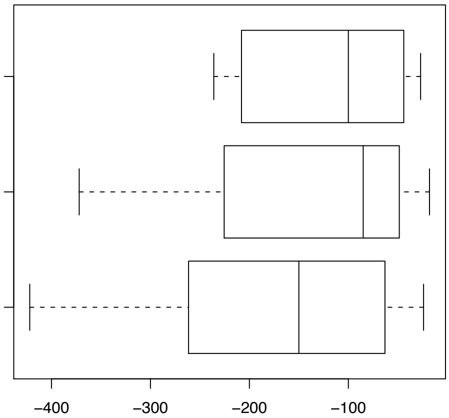
- Justifique convenientemente as suas respostas e escreva os resultados com casas decimais.
- Só pode sair da sala uma hora após o início do exame, devendo nesse caso entregar a sua prova ou desistir da mesma.

---

1. Numa experiência mediram-se fluxos de calor[1] de meia em meia hora, das 7h às 18h durante três dias consecutivos. Os $3 \times n = 69$ resultados (em $\mathrm{W\,m^{-2}}$) foram ordenados e organizados na tabela abaixo.

Dia 1	−236	−228	−216	−210	−208	−208	−208	−196	−175	−114	−100	−100	−67	−63	−53	−48	−47	−41	−41	−38	−32	−31	−27
Dia 2	−422	−405	−401	−359	−320	−295	−228	−184	−178	−175	−171	−150	−120	−98	−74	−67	−65	−61	−59	−54	−50	−38	−24
Dia 3	−372	−324	−294	−284	−259	−255	−196	−194	−138	−103	−92	−85	−74	−74	−59	−57	−49	−48	−34	−33	−32	−31	−18

(a) Foram elaborados diagramas de extremos-e-quartis dos resultados dos três dias, sem qualquer ordem aparente. **(1.0)**



Associe cada diagrama ao respectivo dia.

- **Identificação dos diagramas**
  Como não constam observações discordantes (*outliers*) nos diagramas acima, os limites à esquerda e à direita coincidem com o mínimo e o máximo amostrais. Ora,
  - o *Dia 1*, $x_{(1)} = \min_{i=1,\ldots,n} x_i = -236$ e $x_{(n)} = \max_{i=1,\ldots,n} x_i = -27$;
  - o *Dia 2*, $x_{(1)} = -422$ e $x_{(n)} = -24$;
  - o *Dia 3*, $x_{(1)} = -372$ e $x_{(n)} = -18$.

  Logo:
  - o *Dia 1* corresponde ao diagrama 1 (topo);
  - o *Dia 2* corresponde ao diagrama 3 (o que está em baixo);
  - o *Dia 3* corresponde ao diagrama do meio.

(b) Obtenha os cinco valores-chave usados para traçar o diagrama de extremos-e-quartis associado ao Dia 3. Por que razão não foram assinaladas quaisquer observações discordantes em tal diagrama? **(2.5)**

- **1o. quartil**
  Como $n/4 = 23/4$ não é um inteiro, temos $q_{1/4} = x_{([n/4]+1)} = x_{([5.75]+1)} = x_{(6)} = -255$.
- **2o. quartil ou mediana, $q_{1/2}$**
  Como $n = 23$ é ímpar, temos $me = x_{((n+1)/2)} = x_{(12)} = -85$.
- **3o. quartil**
  Como $3n/4 = 69/4$ não é um inteiro, temos $q_{3/4} = x_{([3n/4]+1)} = x_{([17.25]+1)} = x_{(18)} = -48$.

[1] Os fluxos de calor influenciam o conforto térmico das edificações e afectam a eficiência energética dos mesmos.

- **Restantes dois pontos**
  Uma vez que a amplitude inter-quartil é igual a $IQR = q_{3/4} - q_{1/4} = (-48) - (-255) = 207$, temos:
  $$\max\{x_{(1)}, q_{1/4} - 1.5 \times IQR\} = \max\{-372, -255 - 1.5 \times 207\} = \max\{-372, -565.5\} = -372;$$
  $$\min\{x_{(n)}, q_{3/4} + 1.5 \times IQR\} = \min\{-18, -48 + 1.5 \times 207\} = \min\{-18, 262.5\} = -18.$$
- **Comentário**
  Não foram assinaladas quaisquer observações discordantes neste diagrama, pois $x_i \in [-372, -18]$, $\forall i = 1, \ldots, n$.

(c) Qual é a percentagem de observações referentes ao Dia 3 que excedem -255 mas não excedem -85? **(0.7)**

- **Percentagem pedida**
  Atendendo a que a dimensão da amostra é igual a $n = 23$ e há 6 observações > -255 e ≤ -85 na amostra (ordenada) referente ao Dia 3, a percentagem pedida é dada por
  $$\frac{\text{no. observações } > -255 \text{ e } \leq -85}{n} \times 100\% \quad = \quad \frac{6}{23} \times 100\% \simeq 26.09\%.$$

2. A evaporação dos solventes que se usam nas tintas depende da humidade ambiente. O conhecimento desta relação poderá ser útil para melhorar a qualidade da operação de pintura.

Foi realizado um estudo para examinar a relação entre a humidade relativa ambiente (x, em %) e a quantidade de um determinado solvente evaporado durante a pintura (y, em % do peso), tendo-se registado: $n = 20$; $\bar{x} = 52.5$; $s_x^2 = \frac{1}{n-1}\left[\left(\sum_{i=1}^n x_i^2\right) - n\bar{x}^2\right] = 256.5789$; $\bar{y} = 9.5$; $s_y^2 = \frac{1}{n-1}\left[\left(\sum_{i=1}^n y_i^2\right) - n\bar{y}^2\right] = 10.2632$; $s_{xy} = \frac{1}{n-1}\left[\left(\sum_{i=1}^n x_i y_i\right) - n\bar{x}\bar{y}\right] = -46.0526$.

(a) Determine e compare os coeficientes de variação de x e de y. **(0.7)**

- **Coeficientes de variação pedidos**
  $cv(x) = \frac{s_x}{\bar{x}} = \frac{\sqrt{256.5789}}{52.5} \simeq 0.305106$
  $cv(y) = \frac{s_y}{\bar{y}} = \frac{\sqrt{10.2632}}{9.5} \simeq 0.337223$
  A fracção da dispersão por que a localização é responsável é ligeiramente inferior no caso da variável humidade relativa ambiente (x).

(b) Calcule o coeficiente de correlação amostral entre estas duas variáveis e comente brevemente o valor obtido. **(2.1)**

- **Coeficiente de correlação amostral (de Pearson)**
  $$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right)}} = \frac{(n-1)s_{xy}}{\sqrt{(n-1)s_x^2 \times (n-1)s_y^2}} = \frac{s_{xy}}{\sqrt{s_x^2 \times s_y^2}} = \frac{(-46.0526)}{\sqrt{256.5789 \times 10.2632}} \simeq -0.897434.$$
- **Comentário**
  As variáveis humidade relativa ambiente (x) e quantidade de determinado solvente evaporado durante a pintura (y) tendem a variar em sentidos opostos relativamente às respectivas médias (pois $r < 0$) e estão linear e fortemente correlacionadas (uma vez que r está próximo de -1).

3. Certo teste de ultra-som para detectar fissuras em edifícios resulta positivo: quando elas realmente existem, em 85% dos casos; quando não existem fissuras, em 5% das situações. Uma arquitecta assume que a prevalência de fissuras em edifícios de determinado centro urbano é de 10%.

(a) Escolhido casualmente um edifício de tal centro urbano, calcule a probabilidade de o teste de ultra-som resultar positivo. **(2.1)**

- **Quadro de eventos e probabilidades**

Evento	Probabilidade
F = existem fissuras no edifício	$P(F) = 0.10$
T = teste de ultra-som resultar positivo	$P(T) = ?$
	$P(T \mid F) = 0.85$
	$P(T \mid \overline{F}) = 0.05$

- **Prob. pedida**
A lei da probabilidade total (LPT) permite-nos afirmar que

$$\begin{aligned} P(T) &\overset{LPT}{=} P(T \mid F) \times P(F) + P(T \mid \overline{F}) \times P(\overline{F}) \\ &= 0.85 \times 0.10 + 0.05 \times (1 - 0.10) \\ &= 0.13. \end{aligned}$$

(b) Sabendo que o teste de ultra-som resultou positivo, qual é a probabilidade de o edifício ter realmente fissuras? **(1.6)**

- **Prob. pedida**
Ao invocarmos o teorema de Bayes (TB), temos

$$\begin{aligned} P(F \mid T) &\overset{TB}{=} \frac{P(T \mid F) \times P(F)}{P(T)} \\ &\overset{(a)}{=} \frac{0.85 \times 0.1}{0.13} \\ &\simeq 0.653846. \end{aligned}$$

4. Num gabinete de arquitectura existem duas licenças de *SketchUp*. Seja X a variável aleatória que representa o número diário de licenças activas nos computadores em utilização de tal gabinete. Suponha que a função de probabilidade de X é dada por **(1.0)**

$$P(X = x) = \begin{cases} 0.35, & x = 0 \\ 0.25, & x = 1 \\ 0.40, & x = 2 \\ 0, & \text{outros valores de } x. \end{cases}$$

Calcule $E(10 + 3X + 5X^2)$, o custo esperado diário (em 10^{-1} euros) da utilização das licenças activas de *SketchUp*.

- **V.a.**
X = número de licenças activas em dado momento...

- **Valor esperado pedido**

$$\begin{aligned} E(10 + 3X + 5X^2) &= 10 + 3E(X) + 5E(X^2) = 10 + 3 \times \sum_{x=0}^{2} x \times P(X = x) + 5 \times \sum_{x=0}^{2} x^2 \times P(X = x) \\ &= 10 + 3 \times 1.05 + 5 \times 1.85 = 22.4. \end{aligned}$$

5. Considere um edifício de escritórios com 5 pisos e uma escada e um elevador. Dos dados recolhidos junto das empresas que utilizavam o edifício nos últimos anos, pode concluir-se que a probabilidade de uma pessoa ter dificuldades no acesso aos pisos é igual a 0.1.

Seja X a variável aleatória que representa o número de frequentadores do edifício que referem ter tido dificuldades no acesso aos pisos, em 20 seleccionados ao acaso e com reposição entre os frequentadores do edifício.

(a) Qual é a probabilidade de X não exceder 3? **(1.0)**

- **V.a., distribuição e f.p.**
  X = no. de frequentadores com dificuldades..., em 20 selecc. ao acaso e com reposição...
  $X \sim \text{binomial}(n = 20, p = 0.1)$
  $P(X = x) \overset{form.}{=} \binom{20}{x} 0.1^x (1-0.1)^{10-x}, \quad x = 0, 1, \ldots, 20$
- **Prob. pedida**
  $$P(X \leq 3) \quad = \quad F_X(3) \overset{tabelas, calc}{\simeq} 0.8670.$$

(b) Obtenha a mediana de X. **(0.6)**

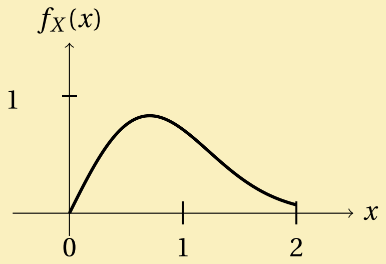
- **Mediana de X**
  É sabido que $me : P(X \leq me) \geq 0.5$ e $P(X \geq me) \geq 0.5$. Equivalentemente,
  $$F_X[(me)^-] \leq 0.5 \leq F_X(me).$$
  De acordo com as tabelas disponíveis,
  $$0.3917 = F_X(1) = P(X \leq 1) = F_X(2^-) \leq 0.5 \leq F_X(2) = 0.7852,$$
  pelo que a mediana da v.a. X é igual a $me = 2$.
  [Alternativamente, $me : 0.5 \leq F_X(me) \leq 0.5 + P(X = me)$ e temos $0.5 \leq F_X(2) = 0.6769 \leq 0.5 + P(X = 2) = 0.5 + (0.6769 - 0.3917) = 0.7852$, logo $me = 2$.]

6. Assuma que o tempo (X, em horas) entre duas chegadas consecutivas de grupos de visitantes a um parque natural possui função de distribuição dada por

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x^2}, & x \geq 0. \end{cases}$$

(a) Determine a função de densidade de probabilidade de X, $f_X(x)$, $x \in \mathbb{R}$, e trace o gráfico de $f_X(x)$, para $0 \leq x \leq 2$. **(1.2)**

- **V.a.**
  X = tempo entre duas chegadas consecutivas de grupos de visitantes a um parque natural
- **F.d.p. de X e seu gráfico**
  [Suponhamos que $f_X(x)$ é contínua em x.] De acordo com a Prop. 4.12 do livro de PE da IST Press,
  $$f_X(x) \quad = \quad \frac{d\,F_X(x)}{dx} = \begin{cases} 0, & x < 0 \\ \frac{d\left(1-e^{-x^2}\right)}{dx} = 2xe^{-x^2}, & x \geq 0 \end{cases}$$



(b) Calcule a probabilidade condicionada $P[X \leq 1 \mid X > E(X)]$, onde $E(X) = \frac{\sqrt{\pi}}{2}$. **(0.7)**

- **Prob. pedida**
  $$\begin{aligned} P[X \leq 1 \mid X > E(X)] \quad &= \quad \frac{P[X \leq 1, X > E(X)]}{P[X > E(X)]} = \frac{P[E(X) < X \leq 1]}{1 - P[X \leq E(X)]} = \frac{F_X(1) - F_X[E(X)]}{1 - F_X[E(X)]} \\ &= \quad \frac{\left(1 - e^{-1^2}\right) - \left(1 - e^{-(\sqrt{\pi}/2)^2}\right)}{1 - \left(1 - e^{-(\sqrt{\pi}/2)^2}\right)} = \frac{e^{-\pi/4} - e^{-1}}{e^{-\pi/4}} \simeq 0.1931. \end{aligned}$$

7. Considere que X_i representa o tempo de execução (em anos) da obra de reabilitação i ($i = 1,\ldots,100$). Suponha que as variáveis aleatórias X_i são independentes e identicamente distribuídas com distribuição uniforme contínua no intervalo $[0.5, 2.5]$.

(a) Obtenha o valor esperado e a variância de X. **(0.7)**

- **V.a. e distribuição**

  X = tempo (em anos) de execução de obra de reabilitação

  $X \sim \text{uniforme}(a = 0.5, b = 2.5)$

- **Valor esperado e variância de** X

$$E(X) \stackrel{form.}{=} \frac{b+a}{2} = \frac{2.5+0.5}{2} = 1.5.$$

$$V(X) \stackrel{form.}{=} \frac{(b-a)^2}{12} = \frac{(2.5-0.5)^2}{12} = \frac{4}{12} = \frac{1}{3}.$$

(b) Calcule um valor aproximado para a probabilidade de o tempo médio de execução das 100 obras de reabilitação exceder um ano e meio, recorrendo para o efeito ao teorema do limite central. **(1.5)**

- **V.a. e distribuição**

  X = tempo (em anos) de execução de obra de reabilitação

  $X \sim \text{uniforme}(a = 0.5, b = 2.5)$

- **V.a.; valor esperado e variância comuns**

  X_i = tempo execução (em anos) da obra de reabilitação i, $\quad i = 1,\ldots,n$

  $X_i \stackrel{i.i.d.}{\sim} X, \quad i = 1,\ldots,n$

  $n = 100$

  $E(X_i) = E(X) = \mu \stackrel{(a)}{=} 1.5$

  $V(X_i) = V(X) = \sigma^2 \stackrel{(a)}{=} \frac{1}{3}$

- **V.a. de interesse**

  $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ = média dos tempos de execução de n obras de realibilitação

- **Valor esperado, variância e distribuição aproximada de** $\bar{X}$

  $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \stackrel{X_i \sim X}{=} E(X) = \mu$

  $V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \stackrel{X_i \sim_{i.i.d.} X}{=} \frac{V(X)}{n} = \frac{\sigma^2}{n}$

  De acordo com o teorema do limite central (TLC),

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} \stackrel{a}{\sim} \text{normal}(0,1).$$

- **Probabilidade pedida**

$$\begin{aligned} P(\bar{X} > 1.5) &= 1 - P\left[Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} \leq \frac{1.5 - E(\bar{X})}{\sqrt{V(\bar{X})}}\right] = 1 - P\left(Z \leq \frac{1.5 - 1.5}{\sqrt{\frac{1/3}{100}}}\right) \\ &\stackrel{TLC}{\simeq} 1 - \Phi(0) \stackrel{tabelas/calc.}{=} 1 - 0.5 = 0.5. \end{aligned}$$

8. Um fabricante de janelas de PVC dá uma garantia de 5 anos quer para alterações nas características do perfil, quer para corrosão nas ferragens. Seja X a variável aleatória que indica se um comprador de uma janela de PVC apresentou uma reclamação a tal fabricante dentro do prazo de garantia.

Admita que $X \sim \text{Bernoulli}(p)$, onde p é uma probabilidade desconhecida, e que numa amostra casual de 100 registos do fabricante foram identificadas 10 reclamações.

(a) Obtenha a estimativa de máxima verosimilhança de $V(X) = p(1-p)$. **(1.2)**

- **V.a. de interesse**

$$X = \begin{cases} 1, & \text{reclamação ocorreu dentro do prazo de garantia} \\ 0, & \text{caso contrário} \end{cases}$$

- **Situação**
  $X \sim \text{Bernoulli}(p)$
  p DESCONHECIDO
- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 100$ proveniente de X e tal que $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 10$.
- **Parâmetro desconhecido**
  $p \quad (p \in (0,1))$
- **Estimativa de MV de p**
  De acordo com o Exemplo 6.55 das páginas 305–307 do livro de PE da IST Press,
  $$\hat{p} \quad = \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10}{100} = 0.1.$$
- **Outro parâmetro desconhecido**
  $h(p) = V(X) = p(1-p)$
- **Estimativa de MV de $h(\lambda)$**
  Ao invocar a propriedade de invariância dos estimadores de máxima verosimilhança (Prop. 6.58, página 311 do livro de PE da IST Press), a estimativa de MV de $h(p) = p(1-p)$ é
  $$\widehat{h(p)} \quad = \quad h(\hat{p}) = \hat{p}(1-\hat{p}) = 0.1 \times (1-0.1) = 0.09.$$

(b) Determine um intervalo de confiança aproximado a 95% para p. **(1.4)**

- **IC aproximado para p**
  A expressão geral do intervalo aproximado de confiança a $(1-\alpha)\times 100\%$ para p é
  $$IC_{(1-\alpha)\times 100\%}(p) \quad \simeq \quad \left[\bar{x} \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right].$$
  Ao termos em conta que $n = 100$, $(1-\alpha)\times 100\% = 0.95$, $\alpha = 0.05$, $\Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.9600$ e $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0.1$,
  $$\begin{aligned} IC_{95\%}(p) \quad &\simeq \quad \left[0.1 - 1.9600 \times \sqrt{\frac{0.1\times(1-0.1)}{100}}, 0.1 + 1.9600 \times \sqrt{\frac{0.1\times(1-0.1)}{100}}\right] \\ &\simeq \quad [0.0412\ ,\ 0.1588]. \end{aligned}$$