Probabilidades e Estatística

LEAmb, LEAN, LEEC21, LEGM, LEMat; LEE, LERC

2º Semestre – 2022/2023
28/06/2023
10:30–12:30

Duração: **120** minutos **Exame Época Normal – A**

- As questões de escolha múltipla têm uma e apenas uma resposta correta. Caso assinale uma resposta errada, serão descontados 0.5 valores.
- Se assinalar mais do que uma resposta numa questão de escolha múltipla, o resultado será classificado como errado e serão descontados 0.5 valores.
- As respostas às questões de escolha múltiplas têm de ser assinaladas na folha de instruções previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as suas respostas.

Pergunta 1 2 valores

Um sistema electrónico com três componentes (1, 2 e 3) funciona sempre que a componente 1 está operacional ou ambas as componentes 2 e 3 estão operacionais.

Sabendo que a componente 1 (respetivamente 2 e 3) está operacional com probabilidade 0.9 (respetivamente 0.85 e 0.8) e que as componentes estão (ou não) operacionais independentemente umas das outras, calcule a probabilidade de o sistema estar em funcionamento.

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
A_1 = "a componente 1 está operacional"	$P(A_1) = 0.9$
A_2 = "a componente 2 está operacional"	$P(A_2) = 0.85$
A_3 = "a componente 3 está operacional"	$P(A_3) = 0.8$
F = "o sistema está em funcionamento"	$P(F) = ?$

- **Cálculo da probabilidade pedida**

  Ao admitir-se que os acontecimentos A_1, A_2 e A_3 são [mutuamente] independentes, temos

$$
\begin{aligned}
P(F) &= P[A_1 \cup (A_2 \cap A_3)] \\
&= P(A_1) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3) \\
&= P(A_1) + P(A_2) \times P(A_3) - P(A_1) \times P(A_2) \times P(A_3) \\
&= 0.9 + 0.85 \times 0.8 - 0.9 \times 0.85 \times 0.8 \\
&= 0.968.
\end{aligned}
$$

Pergunta 2 2 valores

O número mensal de avarias de uma impressora utilizada num escritório é uma variável aleatória com distribuição de Poisson com valor esperado igual a duas avarias por mês (30 dias).

Calcule a probabilidade de que nos primeiros 15 dias do mês ocorram avarias de tal impressora.

- **V.a.**

  X_{30} = número mensal de avarias de uma impressora

$X_{30} \sim \text{Poisson}(\lambda)$, onde λ : $E(X_{30}) = 2 \Leftrightarrow \lambda = 2$

- **V.a. de interesse**

  X_{15} = número de avarias de uma impressora nos primeiros 15 dias do mês

- **Distribuição**

  $X_{15} \sim \text{Poisson}(\lambda \times 15/30 = 1)$, pela propriedade reprodutiva da distribuição de Poisson.

- **F.p. de** X

  $P(X_{15} = x) = \frac{e^{-1} \times 1^x}{x!} = \frac{e^{-1}}{x!}, \quad x \in \mathbb{N}_0$

- **Probabilidade pedida**

$$\begin{aligned} P(X_{15} > 0) &= 1 - P(X_{15} \leq 0) \\ &= 1 - F_{Poisson(1)}(0) \\ &\overset{tabelas, calc.}{\simeq} 1 - 0.3679 \\ &= 0.6321. \end{aligned}$$

Alternativamente, $P(X_{15} > 0) = 1 - P(X_{15} = 0) = 1 - e^{-1} \simeq 1 - 0.3679 = 0.6321$.

Pergunta 3 2 valores

A proporção de impurezas por 10g de minério de ferro é uma variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \begin{cases} 3(1-x)^2, & 0 < x < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Qual é o valor da mediana de X? **A**: 0.7937 **B**: 0.2063 **C**: 0.8750 **D**: 0.5000

- **V.a.**

  X = proporção de impurezas por 10g de minério de ferro

- **Mediana de** X

$$\begin{aligned} me = me(X) \in (0,1) \quad : \quad & F_X(me) = 0.5 \\ & \int_0^{me} 3(1-x)^2\, dx = 0.5 \\ & -(1-x)^3\Big|_0^{me} = 0.5 \\ & 1 - (1-me)^3 = 0.5 \\ & me = 1 - 0.5^{1/3} \\ & me \simeq 0.2063. \end{aligned}$$

[**Obs.**: Note que o valor da mediana é um número real no intervalo [0, 1].]

Pergunta 4 2 valores

Considere que as durações de 50 componentes eletrónicas são variáveis aleatórias independentes com distribuição comum exponencial com valor esperado de 10 dias.

Obtenha um valor aproximado para a probabilidade da duração total dessas 50 componentes eletrónicas não exceder um ano (365 dias).

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

  X_i = duração (em dias) da componente eletrónica i, $\quad i = 1,\ldots,n$

  $n = 50 >> 30$

  $X_i \overset{i.i.d.}{\sim} X \sim \text{exponencial}(\lambda), \quad i = 1,\ldots,n$

  $E(X_i) = E(X) = \mu = 1/\lambda = 10$

  $V(X_i) = V(X) = \sigma^2 = 1/\lambda^2 = 10^2$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i$ = duração de n componentes eletrónicas

- **Valor esperado e variância de S_n**

  $E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \overset{X_i \sim X}{=} n\,E(X) = n\,\mu$

  $V(S_n) = V\left(\sum_{i=1}^n X_i\right) \overset{X_i\ indep.}{=} \sum_{i=1}^n V(X_i) \overset{X_i \sim X}{=} n\,V(X) = n\,\sigma^2$

- **Distribuição aproximada de S_n**

  Segundo o teorema do limite central (TLC),

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\,E(X)}{\sqrt{n\,V(X)}} = \frac{S_n - n\,\mu}{\sqrt{n\,\sigma^2}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Probabilidade pedida (valor aproximado)**

$$\begin{aligned}
P(S_n \leq 365) &= P\left(\frac{S_n - n\,\mu}{\sqrt{n\,\sigma^2}} \leq \frac{365 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) \\
&\overset{TLC}{\simeq} \Phi\left(\frac{365 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) \\
&= \Phi\left(\frac{365 - 50 \times 10}{\sqrt{50 \times 10^2}}\right) \\
&\simeq \Phi(-1.91) \\
&= 1 - \Phi(1.91) \\
&\overset{tabelas/calc.}{=} 1 - 0.9719 \\
&= 0.0281.
\end{aligned}$$

Pergunta 5 — 2 valores

Sejam X e Y duas variáveis aleatórias com função de probabilidade conjunta

$$P(X = x, Y = y) = \begin{cases} \frac{1}{10}, & \text{se } x, y = 1,2,3,4 \text{ e } 1 \leq x + y \leq 5 \\ 0, & \text{caso contrário.} \end{cases}$$

Obtenha a função de probabilidade de Y condicional a $X = 2$.

- **Par aleatório**

  (X, Y) com a f.p. conjunta do enunciado

- **F.p. de Y condicional a $X = 2$**

$$P(Y = j \mid X = 2) = \frac{P(X = 2, Y = j)}{P(X = 2)} = \frac{P(X = 2, Y = j)}{\sum_y P(X = 2, Y = y)}$$

$$
\begin{aligned}
P(Y=j \mid X=2) &= \frac{P(X=2,Y=j)}{P(X=2,Y=1)+P(X=2,Y=2)+P(X=2,Y=3)} \\
&= \begin{cases} \frac{\frac{1}{10}}{\frac{1}{10}+\frac{1}{10}+\frac{1}{10}} = \frac{1}{3}, & j=1,2,3 \\ 0 & \text{caso contrário.} \end{cases}
\end{aligned}
$$

Pergunta 6	2 valores

Admita que a espessura (em *angstrom*) do óxido de semicondutores é representada pela variável aleatória X com função de distribuição

$$F_X(x) = \begin{cases} 1-\exp\left[-(x/\beta)^\alpha\right], & x>0 \\ 0, & \text{caso contrário,} \end{cases}$$

onde α e β são parâmetros positivos desconhecidos. A concretização de uma amostra aleatória de dimensão $n=24$ proveniente de X conduziu às seguintes estimativas de máxima verosimilhança de α e β, $\hat{\alpha}=54.129$ e $\hat{\beta}=427.642$.

Qual é o valor da estimativa de máxima verosimilhança de $P(X>430)$?

A: 0.260101 **B**: 0.739899 **C**: 0.369902 **D**: 0.475892

- **V.a. de interesse; f.d.**

  $X=$ espessura do óxido de semicondutores

  $$F_X(x) = \begin{cases} 1-\exp\left[-(x/\beta)^\alpha\right], & x>0 \\ 0, & \text{outros valores de } x \end{cases}$$

- **Parâmetros desconhecidos**

  $\alpha, \beta \quad (\alpha,\beta>0)$

- **Amostra**

  $\underline{x}=(x_1,\dots,x_n)$, amostra de dimensão $n=24$ da população X e tal que $\hat{\alpha}=54.129$ e $\hat{\beta}=427.642$.

- **Outro parâmetro desconhecido**

  $$h(\alpha,\beta) = P(X>430) = 1-P(X\leq 430) = 1-F_X(430) = \exp\left[-(430/\beta)^\alpha\right]$$

- **Estimativa de MV de $h(\alpha,\beta)$**

  Ao invocar a propriedade de invariância dos EMV, concluímos que a estimativa de MV de $h(\alpha,\beta)$ é

  $$
  \begin{aligned}
  \widehat{h(\alpha,\beta)} &= h(\hat{\alpha},\hat{\beta}) \\
  &= \exp\left[-(430/\hat{\beta})^{\hat{\alpha}}\right] \\
  &= \exp\left[-(430/427.642)^{54.129}\right] \\
  &= 0.260101.
  \end{aligned}
  $$

Pergunta 7	2 valores

A massa (em kg) dos aparelhos de ondas de choque radiais portáteis é uma variável aleatória X com distribuição normal com valor esperado desconhecido e variância igual a 1.2. Foram selecionados aleatoriamente 54 desses aparelhos, tendo-se observado uma média amostral de $\bar{x}=12.1$.

Determine um intervalo de confiança a 92% para o valor esperado de X.

- **V.a. de interesse**

  X = massa (em kg) de aparelho de ondas de choque radiais portátil

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  $\mu = E(X)$ DESCONHECIDO

  $\sigma^2 = V(X) = 1.2$

- **Obtenção do IC para μ**

  Passo 1 — Variável aleatória fulcral para μ

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{normal}(0, 1)$$

  Passo 2 — Quantis de probabilidade

  Dado que $(1-\alpha) \times 100\% = 92\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = \Phi^{-1}(0.04) = -1.7507 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.96) = 1.7507. \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[\bar{X} - b_\alpha \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

  Passo 4 — Concretização

  Atendendo à expressão geral do IC para μ,

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} \pm \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right],$$

  ao par de quantis acima e ao facto de $\bar{x} = 12.1$ e $\sigma^2 = 1.2$, temos:

$$\begin{aligned} IC_{92\%}(\mu) &= \left[12.1 - 1.7507 \times \frac{\sqrt{1.2}}{\sqrt{54}}, 12.1 + 1.7507 \times \frac{\sqrt{1.2}}{\sqrt{54}}\right] \\ &\simeq [11.8390, 12.3610]. \end{aligned}$$

Pergunta 8	2 valores

Um equipamento industrial está regulado para cortar peças metálicas com comprimento esperado de 100 mm e desvio padrão de 1 mm. Em uso, o equipamento vai ficando desregulado o que provoca um aumento da variância dos comprimentos das peças. Admita que o comprimento de uma peça é uma variável aleatória com distribuição normal com valor esperado e desvio padrão desconhecidos.

Teste a hipótese $H_0 : \sigma^2 = 1$ contra $H_1 : \sigma^2 > 1$, num dia em que a variância corrigida dos comprimentos das 25 peças recolhidas casualmente foi de $1.95\,\text{mm}^2$. Decida com base no valor-p (ou num intervalo para o valor-p) do teste.

- **V.a. de interesse**

  X = comprimento de uma peça metálica

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ desconhecido

  σ DESCONHECIDO

- **Hipóteses**

  $H_0 : \sigma^2 = \sigma_0^2 = 1$ vs. $H_1 : \sigma^2 > \sigma_0^2$

- **Estatística de teste**

  $$\frac{(n-1)\,S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{(n-1)}$$

- **Região de rejeição de H_0 (para valores da estatística de teste)**

  O teste é unilateral superior ($H_1 : \sigma^2 > \sigma_0^2$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

  Atendendo a que o valor observado da estatística de teste e o valor-p são iguais a

  $$\begin{aligned} t &= \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{(25-1)\times 1.95}{1^2} = 46.8 \\ valor-p &= P(T > t \mid H_0) = 1 - F_{\chi^2_{(n-1)}}(t) = 1 - F_{\chi^2_{(24)}}(46.8) \overset{calc.}{=} 0.0035, \end{aligned}$$

  é suposto:

  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor-p = 0.35\%$, nomeadamente ao níveis de significância habituais ($\alpha_0 \in [0.01, 0.1]$, e.g., 1%, 5%, 10%)[;
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor-p = 0.35\%$]

  Alternativamente, poderíamos recorrer às tabelas de quantis da distribuição do qui-quadrado e adiantar um intervalo para o valor-p:

  $$\begin{aligned} F^{-1}_{\chi^2_{(24)}}(0.995) = 45.56 &< t = 46.8 < 51.18 = F^{-1}_{\chi^2_{(24)}}(0.999) \\ 0.995 &< F_{\chi^2_{(24)}}(46.8) < 0.999 \\ 1-0.999 &< 1 - F_{\chi^2_{(24)}}(46.8) < 1-0.995 \\ 0.001 &< valor-p < 0.005. \end{aligned}$$

  Assim, é suposto:

  - rejeitarmos H_0 a qualquer n.s. $\alpha_0 \geq 0.5\%$, designadamente aos n.u.s. $(1\%, 5\%, 10\%)$[;
  - não rejeitarmos H_0 a qualquer n.s. $\alpha_0 \leq 0.1\%$].

Pergunta 9 2 valores

Uma engenheira de materiais pretende averiguar se o tempo até fractura de um material segue uma distribuição exponencial de valor esperado 4 (hipótese H_0). Com base no teste de ajustamento do qui-quadrado e na tabela de frequências seguinte, teste H_0 ao nível de significância de 10%.

Tempo até fractura	$]0,2]$	$]2,4]$	$]4,6]$	$]6,+\infty[$
Frequência absoluta observada	23	16	9	12
Frequência absoluta esperada sob H_0	23.61	14.32	8.68	13.39

- **V.a. de interesse**

  X = tempo até fractura de um material

- **Hipóteses**

  $H_0 : X \sim$ exponencial$(1/4 = 0.25)$

  $H_1 : X \not\sim$ exponencial$(1/4 = 0.25)$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

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

  onde: k = no. de classes = 4; O_i = frequência absoluta observável da classe i; E_i = frequência absoluta esperada, sob H_0, da classe i.

- [**Frequências absolutas esperadas sob** H_0 — De acordo com a tabela facultada, as frequências absolutas esperadas sob H_0 aproximadas às centésimas são: $E_1 \simeq 23.61$; $E_2 \simeq 14.32$; $E_3 \simeq 8.68$; $E_4 \simeq 13.39$. Não é necessário fazer qualquer agrupamento de classes uma vez que se verifica $E_i \geq 5$, em pelo menos 80% das classes, e que $E_i \geq 1$, para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

- **Região de rejeição de** H_0 (para valores de T)

  Lidamos com um teste de ajustamento, donde a região de rejeição de H_0 é o intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(4-1)}}(1-0.1) \overset{tabela/calc.}{=} 6.251.$$

- **Decisão**

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$]0,2]$	23	23.61	$\frac{(23-23.61)^2}{23.61} \simeq 0.016$
2	$]2,4]$	16	14.32	0.197
3	$]4,6]$	9	8.68	0.012
4	$]6,+\infty[$	12	13.39	0.144
		$\sum_{i=1}^{k} o_i = n = 60$	$\sum_{i=1}^{k} E_i = n = 60$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 0.369$

  Uma vez que $t \simeq 0.369 \not\in W = (6.251, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 10\%$ [nem a qualquer outro n.s. inferior a α_0].

Pergunta 10 — 2 valores

Um fisiatra está a investigar a relação entre a carga de treino em cicloergómetro (x, em rotações por minuto) e a variação da glicémia (Y, em miligramas por decilitro). Numa amostra de 5 utentes de um centro de reabilitação, obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{5} x_i = 380, \quad \sum_{i=1}^{5} x_i^2 = 29\,774, \quad \sum_{i=1}^{5} y_i = 452, \quad \sum_{i=1}^{5} y_i^2 = 41\,454, \quad \sum_{i=1}^{5} x_i y_i = 33\,654,$$

onde $[\min_{i=1,\ldots,5} x_i, \max_{i=1,\ldots,5} x_i] = [55, 90]$. Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$.

Estime $E(Y \mid x = x_0 = 60)$. Calcule o coeficiente de determinação e interprete o seu valor.

- **[Modelo de RLS e hipóteses de trabalho**

  Y = variação da glicémia (v.a. resposta)

  x = carga de treino em cicloergómetro (variável explicativa)

  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$

  $E(\epsilon_i) = 0, \ V(\epsilon_i) = \sigma^2, \quad i = 1, \ldots, n$

  $cov(\epsilon_i, \epsilon_j) = 0, \ i \neq j, \ i, j = 1, \ldots, n]$

- **Estimativas de MQ dos parâmetros desconhecidos β_0 e β_1**

  Temos

  - $n = 5$
  - $\sum_{i=1}^{n} x_i = 380$
    $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{380}{5} = 76$
    $\sum_{i=1}^{n} x_i^2 = 29\,774$
    $\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 29\,774 - 5 \times 76^2 = 894$
  - $\sum_{i=1}^{n} y_i = 452$
    $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{452}{5} = 90.4$
    $\sum_{i=1}^{n} y_i^2 = 41\,454$
    $\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 = 41\,454 - 5 \times 90.4^2 = 593.2$
  - $\sum_{i=1}^{n} x_i y_i = 33\,654$
    $\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 33\,654 - 5 \times 76 \times 90.4 = -698.$

  Logo, as estimativas de MQ dos parâmetros desconhecidos β_1 e β_0 são dadas por:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} = \frac{-698}{894} \simeq -0.780761. \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 90.4 - (-0.780761) \times 76 = 149.737808. \end{aligned}$$

- **Estimativa pedida**

$$\hat{E}(Y \mid x = 60) = \hat{\beta}_0 + \hat{\beta}_1 \times 60 \simeq 149.7408 + (-0.780761) \times 60 = 102.892170.$$

- **Coeficiente de determinação**

$$r^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right)} = \frac{(-698)^2}{894 \times 593.2} \simeq 0.918697.$$

- **Interpretação coeficiente de determinação**

  Cerca de 92% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado. Podemos concluir que a reta estimada, $\hat{E}(Y \mid x) \simeq 149.7408 + (-0.780761) \times x$, parece ajustar muito bem ao conjunto de dados.