DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidades e Estatística** LEAer21, LEBiom21, LEMec21	2º Semestre – 2022/2023 28/06/2023 **18:00–20:00**

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

- 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

Sabe-se que determinado nutriente está presente no sangue de 10% dos indivíduos de uma dada cidade. Um teste para deteção deste nutriente em amostras de sangue conduz a resultado correto com probabilidade: 0.99 em amostras com a presença do nutriente; 0.97 em amostras sem a presença de tal nutriente.

Calcule a probabilidade de o teste realizado numa amostra de sangue de um indivíduo, escolhido ao acaso nessa cidade, detetar a presença do nutriente.

- **Quadro de acontecimentos e probabilidades**

  Para uma amostra de sangue de indivíduo escolhido ao acaso, temos

Acontecimento	Probabilidade
C = "a amostra de sangue contém o nutriente"	$P(C) = 0.1$
I = "o teste indica a presença do nutriente na amostra"	$P(I) = \ ?$
	$P(I \mid C) = 0.99$
	$P(\overline{I} \mid \overline{C}) = 0.97$

- **Cálculo da probabilidade pedida**

  Recorrendo à lei da probabilidade total, obtemos

$$\begin{aligned} P(I) &= P(C) \times P(I \mid C) + P(\overline{C}) \times P(I \mid \overline{C}) \\ &= P(C) \times P(I \mid C) + [1 - P(C)] \times [1 - P(\overline{I} \mid \overline{C})] \\ &= 0.1 \times 0.99 + (1 - 0.1) \times (1 - 0.97) \\ &= 0.126. \end{aligned}$$

Pergunta 2 2 valores

Os registos das consultas de um médico mostram que 20% dos pacientes que marcam consulta para 6a. feira não comparecem.

Calcule a probabilidade de, numa 6a. feira em que foram marcadas 20 consultas de modo independente, não compareçam 3 dos pacientes, sabendo que nesse dia o médico teve pacientes que não compareceram.

- **V.a.**

  X = número de pacientes que não comparecem à consulta, em 20 consultas marcadas à 6a.

- **Distribuição**

  $X \sim \text{binomial}(n, p)$

  $n = 20$

  $p = P(\text{paciente não comparecer à consulta marcada à 6a.}) = 0.2$

- **F.p. de X**

  $P(X = x) = \binom{20}{x} 0.2^x (1-0.2)^{20-x}, \quad x = 0, 1, \ldots, 20$

- **Probabilidade pedida**

$$
\begin{aligned}
P(X = 3 \mid X \geq 1) &= \frac{P(X = 3, X \geq 1)}{P(X \geq 1)} \\
&= \frac{P(X = 3)}{1 - P(X = 0)} \\
&= \frac{F_X(3) - F_X(2)}{1 - F_X(0)} \\
&\overset{tabelas/calc.}{=} \frac{0.4114 - 0.2061}{1 - 0.0115} \\
&= 0.207688.
\end{aligned}
$$

Pergunta 3	2 valores

Suponha que a pontuação de um aluno num teste é uma variável aleatória X com função de densidade de probabilidade

$$
f_X(x) = \begin{cases} 4x, & 0 \leq x \leq 1/2 \\ 4 - 4x, & 1/2 \leq x \leq 1 \\ 0, & \text{caso contrário.} \end{cases}
$$

Qual é valor do quantil de probabilidade 0.875 de X? **A**: 0.0897 **B**: 0.6250 **C**: 0.7500 **D**: 0.8750

- **V.a.**

  X = pontuação do aluno no teste

  Quantil de probabilidade 0.875 **de** X

  [Atendendo à f.d.p. de X e ao facto de a ordem do quantil ser próxima de 1]

  $\xi = F_X^{-1}(0.875) \in (1/2, 1) \quad : \quad F_X(\xi) = 0.875$

$$
\begin{aligned}
&\int_0^{0.5} 4x\,dx + \int_{0.5}^{\xi} (4 - 4x)\,dx = 0.875 \quad \Leftrightarrow \quad 2x^2\Big|_0^{0.5} + (4x - 2x^2)\Big|_{0.5}^{\xi} = 0.875 \\
&0.5 + 4\xi - 2 - 2\xi^2 + 0.5 = 0.875 \quad \Leftrightarrow \quad 2\xi^2 - 4\xi + 1.875 = 0 \\
&\xi = \frac{4 \pm \sqrt{16 - 8 \times 1.875}}{4} \quad \Leftrightarrow \quad \xi = 0.75 \quad [\text{ou } \xi = 1.25 \not\in (0, 1)].
\end{aligned}
$$

Pergunta 4	2 valores

Um atleta admite que os tamanhos dos seus passos de corrida são variáveis aleatórias independentes e identicamente distribuídas com valor esperado de 0.97 metros e desvio padrão de 0.1 metros.
Determine um valor aproximado para a probabilidade de o tamanho total de 100 dos seus passos de corrida

não se encontrar no intervalo [95, 105].

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

  X_i = tamanho (em metro) do passo de corrida i, $\quad i = 1, \ldots, n$

  $n = 100 > 30$

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

  $E(X_i) = E(X) = \mu = 0.97$

  $V(X_i) = V(X) = 0.1^2$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i$ = tamanho de n passos de corrida

- **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**

  De acordo com 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)**

$$
\begin{aligned}
P(S_n \not\in [95, 105]) &= 1 - P(95 \le S_n \le 105) \\
&= 1 - P\left(\frac{95 - n\,\mu}{\sqrt{n\,\sigma^2}} \le \frac{S_n - n\,\mu}{\sqrt{n\,\sigma^2}} \le \frac{105 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) \\
&\overset{TLC}{\simeq} 1 - \left[\Phi\left(\frac{105 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) - \Phi\left(\frac{95 - n\,\mu}{\sqrt{n\,\sigma^2}}\right)\right] \\
&= 1 - \left[\Phi\left(\frac{105 - 100 \times 0.97}{\sqrt{100 \times 0.1^2}}\right) - \Phi\left(\frac{95 - 100 \times 0.97}{\sqrt{100 \times 0.1^2}}\right)\right] \\
&\simeq 1 - [\Phi(8) - \Phi(-2)] \\
&\simeq \Phi(-2) \\
&\simeq 1 - \Phi(2) \\
&\overset{tabelas/calc.}{=} 1 - 0.9719 \\
&= 0.0228.
\end{aligned}
$$

Pergunta 5	2 valores

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

$$P(X = x, Y = y) = \begin{cases} \frac{x+y-1}{27}, & \text{se } x, y = 1, 2, 3 \\ 0 & \text{caso contrário.} \end{cases}$$

Obtenha $P(X > Y)$.

- **Par aleatório**

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

- **Probabilidade pedida**

$$
\begin{aligned}
P(X > Y) &= \sum_{x=2}^{3}\sum_{y=1}^{x-1} P(X = x, Y = y) \\
&= P(X = 2, Y = 1) + P(X = 3, Y = 1) + P(X = 3, Y = 2) \\
&= \frac{1}{27}(2+1-1) + \frac{1}{27}(3+1-1) + \frac{1}{27}(3+2-1) \\
&= \frac{1}{27}(2+3+4) \\
&= \frac{1}{3}.
\end{aligned}
$$

Pergunta 6	2 valores

Admita que a força de tracção (em *libras*) para conectores utilizados em motores de automóvel é representada pela variável aleatória X com distribuição normal com valor esperado μ e desvio padrão σ desconhecidos. A concretização de uma amostra aleatória de dimensão $n = 26$ proveniente de X conduziu às seguintes estimativas de máxima verosimilhança de μ e σ, $\hat{\mu} = 75.615$ e $\hat{\sigma} = 1.623$.

Qual é o valor da estimativa de máxima verosimilhança de $P(X \leq 73)$?

A: 0.9463 **B**: 0.9799 **C**: 0.0537 **D**: 0.0201

- **V.a. de interesse**

  X = força de tracção...

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

- **Parâmetros desconhecidos**

  $\mu, \sigma \quad (\mu, \sigma > 0)$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 26$ proveniente da população X e tal que $\hat{\mu} = 75.615$ e $\hat{\sigma} = 1.623$.

- **Outro parâmetro desconhecido**

$$
h(\mu,\sigma) = P(X \leq 73) = P\left(\frac{X-\mu}{\sigma} \leq \frac{73-\mu}{\sigma}\right) = \Phi\left(\frac{73-\mu}{\sigma}\right)
$$

- **Estimativa de MV de $h(\mu,\sigma)$**

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

$$
\begin{aligned}
\widehat{h(\mu,\sigma)} &= h(\hat{\mu},\hat{\sigma}) \\
&= \Phi\left(\frac{73-\hat{\mu}}{\hat{\sigma}}\right) = \Phi\left(\frac{73-75.615}{1.623}\right) \simeq \Phi(-1.61) = 1-\Phi(-1.61) \\
&\overset{tabela/calc.}{=} 1-0.9463 \\
&= 0.0537.
\end{aligned}
$$

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 e variância desconhecidos. Foram selecionados aleatoriamente 22 desses aparelhos, tendo-se observado uma variância amostral corrigida de $s^2 = 2.5$.

Determine um intervalo de confiança a 90% para o desvio padrão 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

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

- **Obtenção do IC para σ**

  Passo 1 — Variável aleatória fulcral para σ

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

  Passo 2 — Quantis de probabilidade

  Dado que $n = 22$ e $(1-\alpha) \times 100\% = 90\%$, usaremos os quantis

$$(a_\alpha, b_\alpha): \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(21)}}(0.05) \overset{tabela/calc.}{=} 11.59 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(21)}}(0.95) \overset{tabela/calc.}{=} 32.67. \end{cases}$$

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

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

$$P\left[a_\alpha \le \frac{(n-1)S^2}{\sigma^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{1}{b_\alpha} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\sqrt{\frac{(n-1)S^2}{b_\alpha}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{a_\alpha}}\right] = 1-\alpha$$

  Passo 4 — Concretização

  Atendendo à expressão geral do IC para σ,

$$IC_{(1-\alpha)\times 100\%}(\sigma) = \left[\sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}}, \sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}}\right],$$

  ao par de quantis acima e ao facto de $s^2 = 2.5$, temos:

$$\begin{aligned} IC_{90\%}(\sigma) &= \left[\sqrt{\frac{(22-1)\times 2.5}{32.67}}, \sqrt{\frac{(22-1)\times 2.5}{11.59}}\right] \\ &\simeq [1.2677, 2.1283]. \end{aligned}$$

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 do comprimento esperado das peças. Admita que o comprimento de uma peça metálica é uma variável aleatória com distribuição normal com valor esperado desconhecido e desvio padrão igual a 1 mm.

Teste a hipótese $H_0 : \mu = 100$ contra $H_1 : \mu > 100$, num dia em que a soma dos comprimentos das 16 peças recolhidas ao acaso foi de 1 610 mm. Decida com base no valor-p do teste.

- **V.a. de interesse**

  X = comprimento de uma peça metálica

- **Situação**

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

  μ DESCONHECIDO

  $\sigma = 1$

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 100$

  $H_1 : \mu > \mu_0$

- **Estatística de teste**

  $$T = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim_{H_0} \text{normal}(0, 1)$$

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

  O teste é unilateral superior $(H_1 : \mu > \mu_0)$, 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 aproximado são iguais a

  $$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \\ &= \frac{\frac{1610}{16} - 100}{1 / \sqrt{16}} \\ &\simeq 2.5 \\ valor - p &= P(T > t \mid H_0) \\ &= 1 - \Phi(t) \\ &\simeq 1 - \Phi(2.5) \\ &\overset{tabelas/calc.}{=} 1 - 0.9938 \\ &= 0.0062, \end{aligned}$$

  é suposto:

  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor - p = 0.62\%$, nomeadamente aos níveis de significância habituais $(\alpha_0 \in [0.01, 0.1]$, e.g., 1%, 5%, 10%)[;
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor - p = 0.62\%$]

Pergunta 9	2 valores

Numa experiência laboratorial, pretende-se averiguar se o número de contactos até contágio em certos micro-organismos segue uma distribuição geométrica de valor esperado 2.5 (hipótese H_0). Para tal, recolheu-se uma amostra casual de dimensão 40, tendo-se obtido a seguinte tabela de frequências:

Número de contactos até contágio	1	2	3	4 ou mais
Frequência absoluta observada	22	12	0	6
Frequência absoluta esperada sob H_0	16.00	9.60	5.76	8.64

Com base no teste de ajustamento do qui-quadrado, teste tal conjetura ao nível de significância de 1%.

- **V.a. de interesse**

  X = número de contactos até contagio

- **Hipóteses**

  $H_0 : X \sim \text{geométrica}(1/2.5 = 0.4)$ vs. $H_1 : X \not\sim \text{geométrica}(1/2.5 = 0.4)$

- **Nível de significância**

  $\alpha_0 = 1\%$

- **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: $E_1 \simeq 16.00$; $E_2 \simeq 9.60$; $E_3 \simeq 5.76$; $E_4 \simeq 8.64$. 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

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1 - \alpha_0) = F^{-1}_{\chi^2_{(4-1)}}(1 - 0.01) \overset{tabela/calc.}{=} 11.34.$$

- **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	$\{1\}$	22	16.00	$\frac{(22-16.00)^2}{16.00} = 2.250$
2	$\{2\}$	12	9.60	0.6
3	$\{3\}$	0	5.76	5.76
4	$\{4,5,\ldots\}$	6	8.64	0.807
		$\sum_{i=1}^{k} o_i = n = 40$	$\sum_{i=1}^{k} E_i = n = 40$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 9.417$

  Uma vez que $t \simeq 9.417 \not\in W = (11.34, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 1\%$ [nem a qualquer outro n.s. inferior a α_0].

Pergunta 10	2 valores

Uma engenheira considera que a idade (x, em semanas) dos dois propulsores, quando o motor do foguete é fundido, e a resistência ao cisalhamento da ligação dos propulsores (Y, em 10^3 psi) estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1\, x + \epsilon$. Uma amostra de 20 observações casuais conduziu a:

$$\sum_{i=1}^{20} x_i = 266.9, \quad \sum_{i=1}^{20} x_i^2 = 4\,676.17, \quad \sum_{i=1}^{20} y_i = 43.0, \quad \sum_{i=1}^{20} y_i^2 = 94.12, \quad \sum_{i=1}^{20} x_i\, y_i = 533.62,$$

onde $\min_{i=1,\ldots,20} x_i = 2.0$ e $\max_{i=1,\ldots,20} x_i = 25.0$.

Estime $E(Y \mid x = x_0 = 20)$. Calcule o coeficiente de determinação e interprete o valor obtido.

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

  Y = resistência ao cisalhamento da ligação... (v.a. resposta)

  x = idade dos dois propulsores... (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

  - $n = 20$
  - $\sum_{i=1}^n x_i = 266.9$
    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{266.9}{20} = 13.345$
    $\sum_{i=1}^n x_i^2 = 4\,676.17$
    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 4\,676.17 - 20 \times 13.345^2 = 1114.3895$
  - $\sum_{i=1}^n y_i = 43.0$
    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{43.0}{20} = 2.15$
    $\sum_{i=1}^n y_i^2 = 94.12$
    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 94.12 - 20 \times 2.15^2 = 1.67$
  - $\sum_{i=1}^n x_i\, y_i = 533.62$
    $\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y} = 533.62 - 20 \times 13.345 \times 2.15 = -40.215.$

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

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{-40.215}{1114.3895} \simeq -0.036087; \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 2.15 - (-0.036087) \times 13.345 \simeq 2.631581; \end{aligned}$$

- **Estimativa pedida**

$$\hat{E}(Y \mid x = 20) \;=\; \hat{\beta}_0 + \hat{\beta}_1 \times 20 \simeq 2.631581 + (-0.036087) \times 20 = 1.909841 \quad (\text{em } 10^3 \text{ psi}).$$

- **Cálculo do coeficiente de determinação**

$$r^2 = \frac{\left(\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} = \frac{(-40.215)^2}{1114.3895 \times 1.67} \simeq 0.869006$$

- **Interpretação coeficiente de determinação**

  Cerca de 87% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, donde possamos afirmar que a reta estimada, $\hat{E}(Y \mid x) \simeq 2.631581 + (-0.036087) \times x$, parece ajustar-se bem ao conjunto de dados.