EXAME RECURSO 2022

15 July 2022 15:42

Instituto Superior de Engenharia de Coimbra
Departamento de Física e Matemática
Licenciatura em Engenharia Informática (LEI, LEI-PL, LEICE)
Exame de Métodos Estatísticos - Época de Recurso

19 de Julho de 2022 **Versão 101**/102 Duração: 2h30min

OBSERVAÇÕES:

- No **cabeçalho** da sua folha de resolução indique **o seu nome completo e número de estudante**;
- Na **primeira página** e antes de iniciar a resolução, **indique a versão do teste** que vai realizar e o **modelo de calculadora** que utilizará;
- As escolhas múltiplas erradas **NÃO** descontam;
- Ao longo da resolução **trabalhe com 4 casas decimais** (pelo menos);
- Na resolução de todas as questões que **não sejam** de escolha múltipla justifique todos os cálculos e deduções.

(1.0) **1.** Seja Ω o espaço de resultados associado a uma experiência aleatória. Sejam A e B dois acontecimentos ($A \subset \Omega$ e $B \subset \Omega$). Sabe-se que $P(A) = 0.75$, $P(B/\overline{A}) = \frac{3}{7}$ e $P(A \cap B) = \frac{3}{20}$. O valor de $P(B \cup \overline{A})$ é:

(A) $\frac{2}{5}$ (B) $\frac{1}{5}$ (C) $\frac{3}{5}$ (D) $\frac{9}{35}$

Ex 1:

$P(B \cup \overline{A}) = P(B) + P(\overline{A}) - P(B \cap \overline{A})$; $P(B \cap \overline{A}) = P(B) - P(A \cap B)$

$= \cancel{P(B)} + P(\overline{A}) - \cancel{P(B)} + P(B \cap A)$

$= \frac{1}{4} + \frac{3}{20} = \frac{8}{20} = \frac{2}{5}$ opção A / opção A

ou

$P(B) = P(B \cap A) + P(B \cap \overline{A})$

$= \frac{3}{20} + P(B/\overline{A}) \cdot P(\overline{A})$

$= \frac{3}{20} + \frac{3}{7} \times \frac{1}{4} = \frac{2}{5}$ opção A

(2.0) **2.** A cantina de uma determinada instituição de ensino, com horário diurno e pós-laboral, serve almoços e jantares. Relativamente aos estudantes desta instituição sabe-se que:

- 70% tomam pelo menos uma das refeições na cantina;
- 50% dos estudantes que jantam na cantina também almoçam;
- 10% dos estudantes almoçam e jantam na cantina.

Escolhe-se, ao acaso, um estudante desta instituição de ensino. Determine a probabilidade de:

(a) o estudante apenas almoçar na cantina;

a) A = "o estudante almoça na cantina"

J = " " " janta " " "

$P(A \cup J) = 0.7$; $P(A/J) = 0.5$; $P(A \cap J) = 0.1$

$$P(A \cap \overline{J}) = P(A) - P(A \cap J) = 0.6 - 0.1 = 0.5$$

$$P(A/J) = 0.5 \Leftrightarrow \frac{P(A \cap J)}{P(J)} = 0.5 \Leftrightarrow P(J) = \frac{0.1}{0.5} \Leftrightarrow P(J) = 0.2$$

$$P(A \cup J) = P(A) + P(J) - P(A \cap J) \Leftrightarrow 0.7 = P(A) + 0.2 - 0.1 \Leftrightarrow P(A) = 0.6$$

(b) o estudante jantar na cantina, sabendo que não almoçou lá.

$$P(J/\overline{A}) = \frac{P(J \cap \overline{A})}{P(\overline{A})} = \frac{P(J) - P(A \cap J)}{P(\overline{A})}$$

$$= \frac{0.2 - 0.1}{1 - 0.6} = \frac{0.1}{0.4} = 0.25$$

(3.0) **3.** Durante os Censos 2021, nos inquéritos à população feitos de forma independente uns dos outros, verificou-se que 25% das pessoas precisavam de ajuda para preencher o inquérito.

(a) Condidere que num dia esse inquérito foi realizado, pelo mesmo profissional, a 16 pessoas.

i. A probabilidade de mais de 4 terem precisado de ajuda é:

(A) 0.4050 **(B)** 0.6302 **(C)** 0.3698 **(D)** 0.5950

$\mathcal{E}$ = "Observar e registar se uma pessoa precisou de ajuda"

Repete-se $\mathcal{E}$ sempre nas mesmas condições 16 vezes

X = "nº de pessoas, em 16, que necessitam de ajuda"

$X \sim B(16, 0.25)$

$$P(X > 4) = 1 - P(X \leq 4) = 1 - \text{binomcdf}(16, 0.25, 0, 4)$$

$$= 1 - 0.6302$$

$$= 0.3698$$ opção C / opção B

ii. As pessoas que precisam de ajuda demoram 60 minutos a preencher o inquérito e as que não precisam de ajuda demoram 30 minutos. Quantas horas é de esperar que, no total, o profissional tenha necessitado para recolher as respostas das 16 pessoas?

60 minutos = 1 hora ; 30 minutos = 0.5 hora

Y = "Tempo que 16 pessoas inquiridas demoram a responder ao inquérito"

$= 1 * X + 0.5(16 - x) = X + 8 - 0.5x = 0.5x + 8$

$E(Y) = 0.5\, E(X) + 8 = 0.5 * 4 + 8 = 10$ horas

$E(X) = 16 * 0.25 = 4$

$X \sim \mathcal{B}(16, 0.25)$

(b) Suponha agora que o inquérito foi realizado a 144 pessoas, qual a probabilidade *aproximada* de no máximo 40 pessoas terem precisado de ajuda?

A construção da variável é igual à feita na alínea a) i) mas o nº de repetições é 144.

$X \sim \mathcal{B}(144, 0.25)$

Como $n > 20$ e $0.25 \in]0.1, 0.9[$

$X \dot\sim \mathcal{N}(144 * 0.25, \sqrt{144 * 0.25 * 0.75})$

$= \mathcal{N}(36, 5.1962)$

Aproximações
• Se $X \sim \mathcal{H}(n, N, M)$ com $\frac{n}{N} \leq 0.1$ então $X \dot\sim \mathcal{B}\left(n, \frac{M}{N}\right)$
• Se $X \sim \mathcal{B}(n, p)$
• com $p \leq 0.1$ então $X \dot\sim \mathcal{P}(np)$
• com $p \in]0.1, 0.9[$ e $n > 20$ então $X \dot\sim \mathcal{N}(np, \sqrt{np(1-p)})$
• com $p \geq 0.9$ então $Y = n - X \dot\sim \mathcal{P}(n(1-p))$
• Se $X \sim \mathcal{P}(\lambda)$ com $\lambda > 20$ então $X \dot\sim \mathcal{N}(\lambda, \sqrt{\lambda})$

$P(X \leq 40) \simeq$ normalcdf$(-\infty, 40, 36, 5.1962) = 0.7793$

(4.0) **4.** Considere o vetor aleatório (X, Y) que representa o número de discos externos das marcas A e B, respectivamente, vendidos diariamente numa loja e que é caracterizado pela seguinte função de probabilidade conjunta.

X \ Y	0	1	2
0	a	0.06	0.04
1	0.09	b	0.01
2	0.05	0.02	0.01

a) Sabendo que em 80% dos dias não se vendem leitores da marca A, determine o valor das constantes a e b.

$\Leftrightarrow P(X=0) = 0.8$

$\Leftrightarrow a + 0.06 + 0.04 = 0.8$

$\Leftrightarrow \boxed{a = 0.7}$

Como $\sum_{(x,y) \in S_{(X,Y)}} P(X=x, Y=y) = 1 \Leftrightarrow 0.8 + b + 0.1 + 0.08 = 1$

$\Leftrightarrow b = 1 - 0.98$

$\Leftrightarrow \boxed{b = 0.02}$

b) Determine as leis de probabilidade marginal do número de discos externos das marcas A e B vendidas diariamente na loja.

Lei marginal de X (marca A)

x	0	1	2	c.c
$P(X=x)$	0.8	0.12	0.08	0

Lei marginal de Y (marca B)

y	0	1	2	c.c
$P(Y=y)$	0.84	0.1	0.06	0

Lei marginal de X (marca A)

x	0	1	2	c.c
$P(X=x)$	0.8	0.12	0.08	0

Lei marginal de Y (marca B)

y	0	1	2	c.c
$P(Y=y)$	0.84	0.1	0.06	0

c) Calcule $P(Y=y/X=2), \forall y \in \mathbb{R}$ e comente o seu significado.

$$P(Y=y/X=2) = \begin{cases} \dfrac{P(Y=0, X=2)}{P(X=2)} & , y=0 \\[2ex] \dfrac{P(Y=1, X=2)}{P(X=2)} & , y=1 \\[2ex] \dfrac{P(Y=2, X=2)}{P(X=2)} & , y=2 \\[2ex] 0 & , y \notin \{0,1,2\} \end{cases}$$

proporção de vendas de y carros da marca B tendo sido vendidos 2 carros da marca A

$$= \begin{cases} \dfrac{0.05}{0.08} & , y=0 \\[2ex] \dfrac{0.02}{0.08} & , y=1 \\[2ex] \dfrac{0.01}{0.08} & , y=2 \\[2ex] 0 & , y \notin \{0,1,2\} \end{cases} = \begin{cases} \dfrac{5}{8} & , y=0 \\[2ex] \dfrac{1}{4} & , y=1 \\[2ex] 1/8 & , y=2 \\ 0 & , y \notin \{0,1,2\} \end{cases}$$

d) Determine a função distribuição do $\underbrace{\text{número total de discos externos vendidos diariamente na loja}}_{W = X+Y}$.

w	0	1	2	3	4
$P(W=w)$	0.7	0.15	0.11	0.03	0.01

$x+y$	0	1	2
0	0	1	2
1	1	2	3
2	2	3	4

$P(W=0) = P(X=0, Y=0) = 0.7$

$P(W=1) = P(X=0, Y=1) + P(X=1, Y=0) = 0.15$

$P(W=2) = P(X=1, Y=1) + P(X=0, Y=2) + P(X=2, Y=0) = 0.11$

$P(W=3) = P(X=1, Y=2) + P(X=2, Y=1) = 0.03$

$P(W=4) = P(X=2, Y=2) = 0.01$

$\forall w \in \mathbb{R}$

$$F(w) = P(W \leq w) = \begin{cases} 0 & , w < 0 \\ 0.7 & , 0 \leq w < 1 \\ 0.85 & , 1 \leq w < 2 \end{cases}$$

$$F(w) = P(W \leq w) = \begin{cases} 0.7 & , 0 \leq w < 1 \\ 0.85 & , 1 \leq w < 2 \\ 0.96 & , 2 \leq w < 3 \\ 0.99 & , 3 \leq w < 4 \\ 1 & , w \geq 4 \end{cases}$$

(2.5) **5.** Um monitor emite um sinal de alerta quando uma acção é necessária como parte de um processo de produção. O intervalo, X (em horas), entre sinais sucessivos segue uma distribuição exponencial com as seguintes funções densidade e distribuição, respectivamente.

$$f(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ 0.08\, e^{-0.08x} & \text{se} \quad x \geq 0 \end{cases}; \qquad F(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ 1 - e^{-0.08x} & \text{se} \quad x \geq 0 \end{cases}.$$

(a) A probabilidade do intervalo entre dois sinais consecutivos ser superior a 10 horas e inferior a 20 horas é:

(A) 0.2474 **(B)** 0.6512 **(C)** 0.6680 **(D)** 0.2306

$$P(10 < X < 20) = P(X \leq 20) - P(X \leq 10)$$

X e' contínua

$$= F(20) - F(10)$$

$$= (1 - e^{-0.08*20}) - (1 - e^{-0.08*10})$$

$= 0.2474$ opção A / opção B

(b) Qual é a média e o desvio-padrão do intervalo entre dois sinais consecutivos?

Se $X \sim Exp(0.08)$ então usando as tabelas temos

$$E(X) = \frac{1}{0.08} = 12{,}5 \text{ horas} \quad (= \sigma(X) \text{ ou })$$

$$V(X) = \frac{1}{(0.08)^2} = 156.25 \iff \sigma(X) = \sqrt{156.25} = 12.5 \text{ horas}$$

(c) Após um sinal de alerta, qual é o maior intervalo de tempo em horas que o processo de produção pode ser deixado sem supervisão, garantindo que a probabilidade de perder o próximo sinal seja menos de 0.01?

$$P(X < t) < 0.01$$

$$\iff 1 - e^{-0.08t} = 0.01$$

$$\iff e^{-0.08t} = 0.99$$

$$\iff -0.08t = \ln 0.99$$

$$\iff t = -\frac{1}{0.08} \ln 0.99$$

$\Leftrightarrow \quad t = 0.1256$ horas

(2.5) **6.** Admita que foi realizado um estudo sobre as despesas mensais de alojamento de um estudante em Coimbra. Nesse estudo apurou-se que essa despesa obedece a uma distribuição normal de média 250 euros e desvio-padrão 50 euros.

(a) Qual a probabilidade de um estudante, escolhido ao acaso, gastar mais de 324.5 euros por mês em alojamento?

$X =$ "despesa mensal de alojamento de um estudante em Coimbra"

$X \sim \mathcal{N}(250, 50)$

$P(X > 324.5) = \text{normalcdf}(324.5, +\infty, 250, 50) = 0.0681$

(b) Suponha que a Câmara de Coimbra concluiu que 20% dessas despesas de alojamento são demasiado elevadas. Qual o valor, arredondado às unidades, a partir do qual se considera que essas despesas são demasiado elevadas?

(A) 210€ **(B)** 207€ **(C)** 290€ **(D)** 292€

$P(X \geq x) = 0.20$

$\Leftrightarrow \quad 1 - P(X < x) = 0.20$

$\Leftrightarrow \quad P(X \leq x) = 0.80$

X é contínua

$\Leftrightarrow \quad x = \text{invNorm}(0.80, 250, 50)$

$\Leftrightarrow \quad x = 292.081\text{€} \simeq 292\text{€}$ opção D / opção A

(c) Considere agora que se reuniu uma amostra com despesas de alojamento de 64 estudantes. Qual é a probabilidade da despesa média com alojamento desses estudantes estar compreendida entre 250 euros e 262.5 euros?

$P(250 < \overline{X} < 262.5) = ?$

$\overline{X} = \frac{1}{64} \sum_{i=1}^{64} X_i$

Como $X_1, \dots, X_{64}$ são i.i.d com $X \sim \mathcal{N}(250, 50)$, pela estabilidade da Lei normal

$$\overline{X} \sim \mathcal{N}\left(250, \frac{50}{\sqrt{64}}\right) = \mathcal{N}(250, 6.25)$$

$E(\overline{X}) = E(X) \;;\quad V(\overline{X}) = \frac{V(X)}{64} = \frac{50^2}{64}$

então

$P(250 < \overline{X} < 262.5) = \text{normalcdf}(250, 262.5, 250, 6.25)$

$= 0.4773$

(1.0) **7.** Considere uma variável aleatória X cuja distribuição depende de um parâmetro θ, sendo $E(X) = \theta$ e $V(X) = \theta^2 + 1$. De uma amostra aleatória de dimensão 320 sabe-se que $\overline{x} = 0.0694$ e que $s^2 = 1.6748$. Indique, **justificando**, estimativas cêntricas para a média e para o desvio-padrão de X e para o parâmetro θ.

Estimador cêntrico para $E(x)$ é $\overline{X}$ pois $E(\overline{X}) = E(X)$

logo uma estimativa cêntrica para $E(x)$ é $\overline{x} = 0.0694$

Estimador cêntrico para $V(x)$ é S^2 pois $E(S^2) = \sigma^2$

logo uma estimativa cêntrica para $V(x)$ é $s^2 = 1.6748$

$\Rightarrow$ uma estimativa cêntrica para $\sigma(x)$ é $s = \sqrt{1.6748}$
$= 1.2941$

Como $E(x) = \theta$ uma estimativa cêntrica para θ é também $\overline{x} = 0.0694$

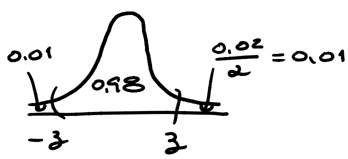
(4.0) **8.** Um conjunto de professores está preocupado com o impacto do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares. Em primeiro lugar será necessário perceber em média quantas horas por semana os alunos dedicam à realização desses trabalhos. Admite-se que o desvio-padrão é de aproximadamente 5 horas. Um estudo realizado com 200 estudantes forneceu como média do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares, o valor de 15 horas.

(a) Com base nesta amostra, determine um intervalo com 98% de confiança para a média do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares.

$\sigma = 5$ horas ; $n = 200$; $\overline{x} = 15$ horas

$(1-\alpha)\% = 98\%$; intervalo de confiança para μ com σ conhecido

$Z = \sqrt{n}\,\dfrac{\overline{X} - \mu}{\sigma} \sim N(0,1)$



$?z: \; P(-z < Z < z) = 0.98$

$\Leftrightarrow ?z: \; P(Z < z) = 0.99$

$\Rightarrow z \simeq \text{invnorm}(0.99, 0, 1) = 2.3264$

$\therefore \; P(-2.3264 < Z < 2.3264) = 0.98$

$-2.3264 < \sqrt{n}\,\dfrac{\overline{X} - \mu}{\sigma} < 2.3264 \Leftrightarrow \overline{X} - \dfrac{2.3264 \times \sigma}{\sqrt{n}} < \mu < \overline{X} + \dfrac{2.3264 \times \sigma}{\sqrt{n}}$

O intervalo aleatório para μ pretendido é

$$\left] \overline{X} - \frac{2.3264 \times \sigma}{\sqrt{n}} , \overline{X} + \frac{2.3264 \times \sigma}{\sqrt{n}} \right[$$

Da amostra recolhida, um intervalo de confiança para μ será
$\overline{x} = 15$h ; $\sigma = 5$; $n = 200$

Da amostra recolhida, um intervalo de confiança para μ será
$\bar{x} = 15h$; $\sigma = 5$; $n = 200$

$$\left]15 - \frac{2.3264 * 5}{\sqrt{200}},\ 15 + \frac{2.3264 * 5}{\sqrt{200}}\right[= \left]14.1775; 15.8225\right[$$

(b) A interpretação correcta do intervalo de confiança obtido na alínea anterior é (escolha a opção correcta)

(A) Existe uma probabilidade de 98% de que o verdadeiro valor da média do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares pertença a esse intervalo.

(B) Existe uma probabilidade de 98% de que um estudante, seleccionado aleatoriamente, tenha dedicado um número de horas semanais a executar os trabalhos das unidades curriculares que pertence a esse intervalo.

(C) Estamos 98% confiantes de que a média do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares pertence a esse intervalo.

(D) Estamos 98% confiantes de que a média da amostra recolhida do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares, pertence a esse intervalo.

opção C / opção C

(c) Se pretendermos um intervalo de confiança de menor amplitude do que o obtido na alínea a) o que podemos mudar no novo estudo (escolha a opção correcta)

(A) Aumentar o grau de confiança e aumentar a dimensão da amostra;

(B) Aumentar o grau de confiança e diminuir a dimensão da amostra;

(C) Diminuir o grau de confiança e aumentar a dimensão da amostra;

(D) Diminuir o grau de confiança e diminuir a dimensão da amostra.

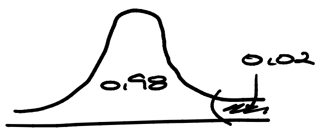
opção C / opção C

(d) Os alunos afirmam que a média do número de horas semanais necessárias para executar os trabalhos de outras unidades curriculares é superior a 15 horas. Efectue um teste com 2% de significância que permita averiguar se os alunos têm razão.

$H_0: \mu = 15$ vs $H_1: \mu > 15$

$\alpha = 0.02$

$T = \sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0,1)$



Região Crítica $=]z, +\infty[$ onde $z: P(Z < z) = 0.98$

$=]2.0538, +\infty[$

$\Leftrightarrow z = \text{invNorm}(0.98, 0, 1)$

$\Leftrightarrow z = 2.0538$

$T_{H_0} = \sqrt{200}\,\frac{15 - 15}{5} = 0 \notin$ Região crítica logo, ao nível 0.02 não rejeitamos Ho. Efetivamente o número médio de horas, ao nível 0.02, não é superior a 15 horas.