

-- CONHECIMENTOS ESPECÍFICOS --**Questão 31**

A tabela a seguir apresenta valores de concentração de sólidos dissolvidos totais (em mg/L) em amostras de água coletadas de cinco estações de monitoramento de uma companhia de saneamento.

estação	concentração (mg/L)
I	120
II	150
III	135
IV	145
V	125

A partir dessa situação hipotética, é correto afirmar que o coeficiente de assimetria de Bowley (γ), definido como

$$\gamma = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1},$$

em que Q_1 , Q_2 e Q_3 representam, respectivamente, o primeiro, o segundo e o terceiro quartil desse conjunto de dados, é igual a

- ☐ A -0,8.
☐ B -0,2.
☐ C 0,0.
☐ D 0,2.
☐ E 0,25.

||JUSTIFICATIVAS||

||A|| - Incorreta. Se os dados fossem organizados em ordem crescente, 120, 125, 135, 145, 150, as diferenças entre as observações consecutivas seriam 5, 10, 10, 5. Portanto, esses dados não possuem assimetria de Bowley negativa já que esses dados ordenados são igualmente espaçados.*

||B|| - Incorreta. Se os dados fossem organizados em ordem crescente, 120, 125, 135, 145, 150, as diferenças entre as observações consecutivas seriam 5, 10, 10, 5. Portanto, esses dados não possuem assimetria de Bowley negativa já que esses dados ordenados são igualmente espaçados.*

||C|| - Correta. Como $Q_1 = 125$, $Q_2 = 135$ e $Q_3 = 145$, obtém-se $\gamma = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{10 - 10}{20} = 0$.*

||D|| - Incorreta. Se os dados fossem organizados em ordem crescente, 120, 125, 135, 145, 150, as diferenças entre as observações consecutivas seriam 5, 10, 10, 5. Portanto, esses dados não possuem assimetria de Bowley positiva já que esses dados ordenados são igualmente espaçados.*

||E|| - Incorreta. Se os dados fossem organizados em ordem crescente, 120, 125, 135, 145, 150, as diferenças entre as observações consecutivas seriam 5, 10, 10, 5. Portanto, esses dados não possuem assimetria de Bowley positiva já que esses dados ordenados são igualmente espaçados.*

Questão 32

Uma companhia de saneamento possui um sistema de monitoramento de qualidade da água. Durante uma inspeção, foi observado que a probabilidade de uma estação apresentar altos níveis de sólidos dissolvidos totais (A) é $P(A) = 0,3$, e que a probabilidade de essa mesma estação apresentar alto teor de nitratos (B) é $P(B) = 0,4$. Caso essa estação apresente altos níveis de sólidos dissolvidos totais, a probabilidade de ela apresentar altos níveis de nitratos aumentará para 0,5, ou seja, $P(B|A) = 0,5$.

Com base nessas informações, é correto concluir que a probabilidade condicional $P(A|B)$ é igual a

- ☐ A 0,375.
☐ B 0,300.
☐ C 0,500.
☐ D 0,667.
☐ E 0,750.

||JUSTIFICATIVAS||

||A|| - Correta. $P(A|B) = P(B|A)P(A)/P(B) = 0,5 \times 0,3/0,4 = 0,375$.*

||B|| - Incorreta. $P(A|B) = P(B|A)P(A)/P(B) = 0,5 \times 0,3/0,4 = 0,375$.*

||C|| - Incorreta. $P(A|B) = P(B|A)P(A)/P(B) = 0,5 \times 0,3/0,4 = 0,375$.*

||D|| - Incorreta. $P(A|B) = P(B|A)P(A)/P(B) = 0,5 \times 0,3/0,4 = 0,375$.*

||E|| - Incorreta. $P(A|B) = P(B|A)P(A)/P(B) = 0,5 \times 0,3/0,4 = 0,375$.*

Questão 33

O número diário de falhas ocorridas (Y) em determinado sistema hidráulico é uma variável aleatória distribuída da seguinte forma:

y	$P(Y=y)$
0	0,40
1	0,20
2	0,20
3	0,10
4	0,10

Quando esse sistema hidráulico falha, uma equipe técnica tenta, de imediato, corrigir tal falha. A probabilidade de sucesso nessa tentativa de correção é igual a 0,9, ou seja, $P(X=1) = 0,9$; e a probabilidade de fracasso é igual a 0,1, isto é, $P(X=0) = 0,1$.

Considerando-se a situação hipotética precedente e sabendo-se que X e Y são variáveis aleatórias independentes, é correto afirmar que o valor esperado do número diário de falhas corrigidas com sucesso é igual a

- ☐ A 0,9.
☐ B 1,17.
☐ C 1,3.
☐ D 1,9.
☐ E 2,25.

||JUSTIFICATIVAS||

||A|| - Incorreta. A média de Y é $E[Y] = 0,2 + 2 \times 0,2 + 3 \times 0,1 + 4 \times 0,1 = 1,3$. Esse é o total esperado de falhas observadas. Como X e Y são independentes, o total de falhas corrigidas representa 90% de 1,3, ou seja, 1,17. O número 0,9 denota uma probabilidade de sucesso, mas não representa o número diário de falhas corrigidas com sucesso.*

||B|| - Correta. A média de Y é $E[Y] = 0,2 + 2 \times 0,2 + 3 \times 0,1 + 4 \times 0,1 = 1,3$. Esse é o total esperado de falhas observadas. Como X e Y são independentes, o total de falhas corrigidas representa 90% de 1,3, ou seja, 1,17.*

||C|| - Incorreta. A média de Y é $E[Y] = 0,2 + 2 \times 0,2 + 3 \times 0,1 + 4 \times 0,1 = 1,3$. Esse é o total esperado de falhas observadas. Como X e Y são independentes, o total de falhas corrigidas representa 90% de 1,3, ou seja, 1,17. O número 1,3 representa o número médio diário de falhas ocorridas.*

||D|| - Incorreta. A média de Y é $E[Y] = 0,2 + 2 \times 0,2 + 3 \times 0,1 + 4 \times 0,1 = 1,3$. Esse é o total esperado de falhas observadas. Como X e Y são independentes, o total de falhas corrigidas representa 90% de 1,3, ou seja, 1,17. O número esperado de falhas corrigidas com sucesso não poderia ser superior ao número esperado de falhas ocorridas ($E[Y] = 1,3$).*

||E|| - Incorreta. A média de Y é $E[Y] = 0,2 + 2 \times 0,2 + 3 \times 0,1 + 4 \times 0,1 = 1,3$. Esse é o total esperado de falhas observadas. Como X e Y são independentes, o total de falhas corrigidas representa 90% de 1,3, ou seja, 1,17. O valor 0,9 não representa corretamente o valor esperado do número diário de falhas corrigidas com sucesso, pois 2,5, embora seja o ponto médio entre 0 e 4, não coincide com a média de Y .*

Questão 34

Considere que V e W sejam variáveis aleatórias tais que:

- I $P(V = -1 | W = w) = 1 - w^2$;
 II $P(V = +1 | W = w) = w^2$.

Sabendo-se que W segue uma distribuição uniforme contínua no intervalo $[0, 1]$, é correto afirmar que a variância de V será igual a

- A $\frac{1}{12}$.
 B $\frac{1}{2}$.
 C $\frac{4}{5}$.
 D $\frac{8}{9}$.
 E 1.

JUSTIFICATIVAS

||A|| - Incorreta. $\frac{1}{12}$ representa a variância de W .*/

||B|| - Incorreta. $\frac{1}{2}$ denota a média de W .*/

||C|| - Incorreta. $\frac{4}{5}$ não representa a variância de V .*/

||D|| - Correta. $P(V = 1) = \int_0^1 P(V = 1 | W = w) dw = \int_0^1 w^2 dw = \frac{1}{3}$, e, consequentemente, $P(V = -1) = 1 - \frac{1}{3} = \frac{2}{3}$.

Logo, $E(V) = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3}$. Além disso, $E(V^2) = 1 \times \frac{2}{3} + 1 \times \frac{1}{3} = 1$. Portanto, $Var(V) = E(V^2) - E^2(V) = 1 - \frac{1}{9} = \frac{8}{9}$.*/

||E|| - Incorreta. A variância de V deve ser inferior a 1, pois $E(V^2) = 1$.*/

Questão 35

A fração do dia de trabalho necessária para executar determinada tarefa é descrita como uma variável aleatória U , cuja função de densidade de probabilidade é dada por

$$f_U(u) = \begin{cases} A \cdot u^2, & \text{se } u \in [0, 1], \\ 0, & \text{se } u \notin [0, 1], \end{cases}$$

em que A representa a constante de normalização. Nesse caso, A será igual a

- A $1/3$.
 B $1/2$.
 C 1.
 D 2.
 E 3.

JUSTIFICATIVAS

||A|| - Incorreta. Se $A = 1/3$, então $\int_0^1 Au^2 du < 1$. Nesse caso, $f_U(u)$ não seria uma função de densidade de probabilidade.*/

||B|| - Incorreta. Se $A = 1/2$ então $\int_0^1 Au^2 du < 1$. Nesse caso, $f_U(u)$ não seria uma função de densidade de probabilidade.*/

||C|| - Incorreta. Se $A = 1$, então $\int_0^1 Au^2 du < 1$. Nesse caso, $f_U(u)$ não seria uma função de densidade de probabilidade.*/

||D|| - Incorreta. Se $A = 2$, então $\int_0^1 Au^2 du < 1$. Nesse caso, $f_U(u)$ não seria uma função de densidade de probabilidade.*/

||E|| - Correta. Como $\int_0^1 f_U(u) du = \int_0^1 Au^2 du = \frac{A}{3} = 1$, conclui-se que $A = 3$.*/

Questão 36

Se X for descrita por uma distribuição normal com média e desvio padrão iguais a 1, e se Y for uma variável aleatória tal que $E[Y|X = x] = 2x^2$, então o valor esperado de Y será igual a

- A 1.
 B 2.
 C 3.
 D 4.
 E 5.

JUSTIFICATIVAS

||A|| - Incorreta. $E[Y] \neq E[X] = 1$.*/

||B|| - Incorreta. $E[Y] \neq 2Var[X] = 2$.*/

||C|| - Incorreta. $E[Y] \neq 3$.*/

||D|| - Correta. $E[Y] = E[E[Y|X]] = 2E[X^2] = 2Var(X) + 2E^2[X] = 2 + 2 = 4$.*/

||E|| - Incorreta. $E[Y] \neq 5$.*/

Questão 37

Considere a seguinte função de densidade de probabilidade:

$$f(x) = \begin{cases} 0, & \text{se } x < 0 \text{ ou } x \geq 3. \\ \frac{x}{4}, & \text{se } 0 \leq x < 2, \\ \frac{1}{2}, & \text{se } 2 \leq x < 3. \end{cases}$$

Assinale a opção que corresponde à função de distribuição acumulada relacionada à função de densidade de probabilidade apresentada.

A $F(x) = \begin{cases} 0, & \text{se } x < 0. \\ \frac{x^2}{8}, & \text{se } 0 \leq x < 2, \\ \frac{1}{2} + \frac{x-2}{2}, & \text{se } 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$

B $F(x) = \begin{cases} 0, & \text{se } x < 0. \\ \frac{x^2}{8}, & \text{se } 0 \leq x < 2, \\ \frac{x}{2}, & \text{se } 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$

C $F(x) = \begin{cases} 0, & \text{se } x < 0. \\ \frac{1}{4}, & \text{se } 0 \leq x < 2, \\ 0, & \text{se } 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$

D $F(x) = \begin{cases} 0, & \text{se } x < 0. \\ \frac{x}{4}, & \text{se } 0 \leq x < 2, \\ \frac{x}{4} + \frac{1}{2}, & \text{se } 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$

E $F(x) = \begin{cases} 0, & \text{se } x < 0 \text{ ou } x \geq 3. \\ \frac{x^2}{8}, & \text{se } 0 \leq x < 2, \\ \frac{x}{2}, & \text{se } 2 \leq x < 3. \end{cases}$

||JUSTIFICATIVAS||

||A|| - Correta.

 $F(x) = P(X \leq x) = 0$, se $x < 0$. $F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^0 0 dx + \int_0^x \frac{u}{4} du = \frac{x^2}{8}$, se $0 \leq x < 2$. $F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^0 0 dx + \int_0^2 \frac{u}{4} du + \int_2^x \frac{1}{2} du = \frac{1}{2} + \frac{(x-2)}{2}$.Finalmente, $F(x) = P(X \leq x) = 1$, se $x \geq 3$. Logo,

$$F(x) = \begin{cases} 0, & \text{se } x < 0. \\ \frac{x^2}{8}, & \text{se } 0 \leq x < 2, \\ \frac{1}{2} + \frac{x-2}{2}, & \text{se } 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases} *$$

||B|| - Incorreta. A função apresentada não é monotonicamente crescente de zero para 1.*

||C|| - Incorreta. A função apresentada não é monotonicamente crescente de zero para 1.*

||D|| - Incorreta. A função apresentada não é monotonicamente crescente de zero para 1.*

||E|| - Incorreta. A função apresentada produz probabilidades superiores a 1.*

Questão 38

Em determinada companhia, o número diário de atendimentos de emergência (Y), condicionado a determinada taxa $M = m$, segue uma distribuição de Poisson na forma apresentada a seguir, em que $y \in \{0, 1, 2, 3, \dots\}$, $m > 0$ e M segue uma distribuição exponencial com média igual a 1.

$$P(Y = y | M = m) = \frac{e^{-m} m^y}{y!}$$

Considerando essa situação hipotética, assinale a opção em que é corretamente apresentada a distribuição marginal Y .

A $P(Y = y) = \frac{e^{-1}}{y!}$.

B $P(Y = y) = 2^{-(y+1)}$.

C $P(Y = y) = \frac{e^{-2m} m^y}{y!}$.

D $P(Y = y) = \frac{e^{-y}}{1+e^{-y}}$.

E $P(Y = y) = e^{-y}$.

||JUSTIFICATIVAS||||A|| - Incorreta. Y não segue uma distribuição de Poisson com taxa unitária.*

||B|| - Correta.

$$P(Y = y) = \int_0^{+\infty} P(Y = y | M = m) e^{-m} dm = \frac{1}{y!} \int_0^{+\infty} e^{-2m} m^y dm$$

$$= \int_0^{+\infty} \frac{e^{-2m} m^{y+1-1}}{\Gamma(y+1)} dm = \frac{1}{2^{y+1}}, \text{ com base nas propriedades da distribuição especial gama.} *$$

||C|| - Incorreta. $P(Y = y)$ não pode depender de m .*

||D|| - Incorreta. Essa função não representa uma distribuição de probabilidade.*

||E|| - Incorreta. Essa função não representa uma distribuição de probabilidade.*

Questão 39

Caso X siga distribuição normal com média 3 e variância 1, e Y siga distribuição normal com média 2 e variância 4, então

A $P(X > 4) + P(Y < 4) > 1$.

B $P(X > 4) - P(Y < 4) = 0$.

C $P(X > 4) + P(Y < 4) < 1$.

D $P(X > 4) - P(Y < 4) = 0,5$.

E $P(X > 4) + P(Y < 4) = 1$.

||JUSTIFICATIVAS||

||A|| - Incorreta. A soma deve ser igual a 1.*

||B|| - Incorreta. A diferença não pode ser nula já que $P(Y < 4) > P(X > 4)$.*

||C|| - Incorreta. A soma deve ser igual a 1.*

||D|| - Incorreta. A diferença não pode ser positiva, já que $P(Y < 4) > P(X > 4)$.*

||E|| - Correta. $P(X > 4) = P\left(Z > \frac{4-3}{1}\right) = P(Z > 1)$ e $P(Y < 4) = P\left(Z < \frac{4-2}{2}\right) = P(Z < 1)$. Assim, como $P(Z = 1) = 0$, já que Z é variável aleatória contínua, conclui-se que $P(X > 4) + P(Y < 4) = 1$.*

Questão 40

Suponha que P seja uma probabilidade aleatória que se distribui conforme uma variável aleatória uniforme no intervalo $(0,1)$. Nesse caso, o valor esperado de $-\ln P$ será igual a

A 0,5.

B 0,75.

C 1,0.

D 1,5.

E 1,75.

||JUSTIFICATIVAS||

||A|| - Incorreta. A média deve ser igual a 1.*

||B|| - Incorreta. A média deve ser igual a 1.*

||C|| - Correta. Como P é uma probabilidade aleatória que se distribui conforme uma variável aleatória uniforme no intervalo $(0,1)$, temos $\text{Prob}[P < p] = p = \text{Prob}[-\ln P > -\ln p] = \text{Prob}[X > -\ln p = x] = e^{-x}$. Ou seja, $-\ln P$ segue uma distribuição exponencial com média igual a 1.*

||D|| - Incorreta. A média deve ser igual a 1.*

||E|| - Incorreta. A média deve ser igual a 1.*

Questão 41

Um analista está fazendo uma análise com base em dois indicadores de qualidade da água, A e B . Esses indicadores são variáveis aleatórias independentes, tais que A segue distribuição normal com média 10 e desvio padrão 2, enquanto B se distribui como uma distribuição normal padrão. O analista decide combinar os indicadores A e B na forma da razão a seguir.

$$R = \frac{A - 10}{2B}$$

A partir dessa situação hipotética, é correto afirmar que R

A segue uma distribuição normal com média zero.

B tem variância igual a 1.

C tem mediana igual a 10.

D tem assimetria positiva.

E segue uma distribuição t de Student com 1 grau de liberdade.

||JUSTIFICATIVAS||

||A|| - Incorreta. R não segue uma distribuição normal porque o denominador é uma variável aleatória.* /

||B|| - Incorreta. R não possui variância. Trata-se de uma distribuição com variância indefinida.* /

||C|| - Incorreta. A mediana de R é zero, porque ela é simétrica em torno de zero.* /

||D|| - Incorreta. A distribuição é simétrica em torno de erro.* /

||E|| - Correta. A razão entre duas distribuições normais padronizadas e independentes define uma distribuição t de Student com 1 grau de liberdade.* /

Questão 42

Supondo-se que T_n siga uma distribuição t de Student com n graus de liberdade, e que A_n convirja em probabilidade para 1 como resultado da aplicação da lei fraca dos grandes números, então é correto afirmar que, à medida que n aumentar, a distribuição da soma $T_n + A_n$ tenderá para uma distribuição

- Ⓐ qui-quadrado com 1 grau de liberdade.
- Ⓑ F de Snedecor com 2 graus de liberdade no numerador e 1 grau de liberdade no denominador.
- Ⓒ normal com média e desvio padrão iguais a 1.
- Ⓓ t de Student com 1 grau de liberdade.
- Ⓔ exponencial com média igual a 1.

||JUSTIFICATIVAS||

||A|| - Incorreta. A distribuição limite não possui restrição no seu suporte.* /

||B|| - Incorreta. A transformação não envolve uma razão de variáveis aleatórias.* /

||C|| - Correta. A distribuição converge para uma distribuição normal padrão à medida que o número de graus de liberdade aumenta. Como A_n convergem em probabilidade para 1 como resultado da aplicação da lei fraca dos grandes números. A transformação $T_n + A_n$ resulta no deslocamento da distribuição normal padrão em uma unidade, de modo que $E[T_n + A_n] = 0 + 1 = 1$. * /

||D|| - Incorreta. O número de graus de liberdade deve aumentar.* /

||E|| - Incorreta. A distribuição limite não possui restrição no seu suporte.* /

Questão 43

Considere uma amostra aleatória simples $X_1, X_2, X_3, \dots, X_n$, extraída de uma população com média 20 e desvio padrão 5, bem como a seguinte soma ponderada.

$$S_n = \sum_{k=1}^n (0,2)^k X_k$$

Com base nessas informações, assinale a opção correta.

- Ⓐ Pela lei fraca dos grandes números, S_n converge em probabilidade para 20 à medida que n aumenta.
- Ⓑ Pela lei forte dos grandes números, S_n converge para uma distribuição normal à medida que n aumenta.
- Ⓒ De acordo com o teorema central do limite, mesmo que a população não seja normal, S_n converge para uma distribuição normal padrão à medida que n aumenta.
- Ⓓ Se a população for normal, à medida que n aumenta, S_n converge para uma distribuição normal com média igual a 5.
- Ⓔ A média de S_n tende para 20 à medida que n aumenta.

||JUSTIFICATIVAS||

||A|| - Incorreta. Não há convergência em probabilidade porque os a soma dos pesos converge para 0,25.* /

||B|| - Incorreta. A lei forte dos grandes números não trata da convergência em distribuição.* /

||C|| - Incorreta. Como $Var(S_n)$ não decresce para zero à medida que n aumenta, não há convergência em distribuição para uma distribuição normal com média 20. * /

||D|| - Correta. Soma de variáveis aleatórias normais independentes resulta em uma distribuição normal, e $\lim_{n \rightarrow \infty} E[S_n] = \sum_{k=1}^{\infty} (0,2)^k E[X_k] = 20 \times \frac{0,2}{0,8} = 5$. * /

||E|| - Incorreta. $\lim_{n \rightarrow \infty} E[S_n] = \sum_{k=1}^{\infty} (0,2)^k E[X_k] = 20 \times \frac{0,2}{0,8} = 5$. * /

Questão 44

Considerando-se duas variáveis independentes X e Y , tais que $E[X] = E[Y] = 1$ e $Var[X] = Var[Y] = 2$, é correto afirmar que $E[(X - Y)^2]$ é igual a

- Ⓐ 4.
- Ⓑ 3.
- Ⓒ 2.
- Ⓓ 1.
- Ⓔ 0.

||JUSTIFICATIVAS||

||A|| - Correta. $E[(X - Y)^2] = EX^2 - 2EXEY + EY^2 = (2 + 1) - 2 + (2 + 1) = 4$. * /

||B|| - Incorreta. $E[(X - Y)^2] = 4$. * /

||C|| - Incorreta. $E[(X - Y)^2] = 4$. * /

||D|| - Incorreta. $E[(X - Y)^2] = 4$. * /

||E|| - Incorreta. $E[(X - Y)^2] = 4$. * /

Questão 45

Considere certa amostra aleatória simples $X_1, X_2, \dots, X_9$ retirada de uma população normal com média 0. Nesse caso, se $9\bar{X} = \sum_{i=1}^9 X_i$ e $8S^2 = \sum_{i=1}^9 (X_i - \bar{X})^2$, então a variância da razão $R = \frac{3\bar{X}}{S^2}$ é igual a

- Ⓐ 1.
- Ⓑ 4/3.
- Ⓒ 3/2.
- Ⓓ 8.
- Ⓔ 9.

||JUSTIFICATIVAS||

||A|| - Incorreta. A variância de R é 4/3.* /

||B|| - Correta. Como R segue distribuição t de Student com 8 graus de liberdade, sua variância é $8/6=4/3$. * /

||C|| - Incorreta. A variância de R é 4/3.* /

||D|| - Incorreta. A variância de R é 4/3.* /

||E|| - Incorreta. A variância de R é 4/3.* /

Questão 46

Considerando o contexto da inferência estatística sob a abordagem clássica frequentista, assinale a opção em que são apresentadas as propriedades desejáveis de um estimador.

- Ⓐ viés, consistência, eficiência, suficiência e robustez
- Ⓑ viés nulo, eficiência máxima, consistência, suficiência e simplicidade
- Ⓒ viés, eficiência, consistência, suficiência e simplicidade
- Ⓓ consistência, suficiência, eficiência e viés nulo
- Ⓔ consistência, suficiência, eficiência, robustez e variância nula

||JUSTIFICATIVAS||

||A|| - Incorreta. Viés não é uma propriedade de estimador sob a perspectiva da inferência clássica.*/
 ||B|| - Incorreta. Simplicidade não é uma propriedade de estimador sob a perspectiva da inferência clássica.*/
 ||C|| - Incorreta. Simplicidade não é uma propriedade de estimador sob a perspectiva da inferência clássica.*/
 ||D|| - Correta. As propriedades mais desejáveis em um estimador são as descritas a seguir.

Consistência: o estimador converge em probabilidade para o valor verdadeiro do parâmetro quando o tamanho da amostra aumenta.
 Suficiência: o estimador utiliza todas as informações disponíveis na amostra sobre o parâmetro.
 Eficiência: o estimador possui a menor variância possível entre os estimadores não viesados.
 Viés nulo: um estimador é não viesado se, em média, o valor estimado é igual ao valor verdadeiro do parâmetro.
 Variância mínima: quanto menor a variância, mais preciso é o estimador.*/
 ||E|| - Incorreta. Variância nula não é uma propriedade de estimador sob a perspectiva da inferência clássica.*/
Questão 47

De certa população normal, foi retirada uma amostra aleatória simples com reposição de tamanho $n = 100$, cuja média amostral é igual a 53 e cujo desvio padrão amostral é igual a 25.

Com base nas informações precedentes, assinale a opção que corresponde ao intervalo de 95% de confiança para a média μ dessa população, considerando $Z = 1,96$ para esse intervalo de confiança.

- Ⓐ $50,5 < \mu < 55,5$
- Ⓑ $48,875 < \mu < 57,125$
- Ⓒ $48,1 < \mu < 57,9$
- Ⓓ $47,375 < \mu < 58,625$
- Ⓔ $46,55 < \mu < 59,45$

||JUSTIFICATIVAS||

||A|| - Incorreta. Intervalo obtido para um nível de confiança de 68%.*/
 ||B|| - Incorreta. Intervalo obtido para um nível de confiança de 90%.*/
 ||C|| - Correta. Intervalo que corresponde ao nível de confiança de 95% para a média μ da população em questão é calculada de acordo com o que se segue, em que $\bar{x}$ corresponde à média amostral e s , ao desvio padrão amostral.

$P(|\frac{\bar{x}-\mu}{d/\sqrt{n}}| < Z) = 95\%$, em que $\bar{x} = 53$; $n = 100$; $d = 25$ e $Z = 1,96$.

Portanto, $(\frac{|53-\mu|}{2,5}) < 1,96$. Ou seja:

$$53 - \mu = 4,9$$

$$53 - \mu = -4,9$$

Por fim, $48,1 < \mu < 57,9$.*/
 ||D|| - Incorreta. Intervalo obtido para um nível de confiança de 97,5%.*/
 ||E|| - Incorreta. Intervalo obtido para um nível de confiança de 99%.*/
Questão 48

Questão 48

Considerando que se queira estimar, por meio de uma amostra, a mediana de uma população X que possua distribuição normal, assinale a opção correta.

- Ⓐ A mediana amostral será o melhor estimador nesse caso, qualquer que seja o critério escolhido.
- Ⓑ A média amostral terá viés maior que a mediana amostral.
- Ⓒ A mediana amostral terá viés maior que a média amostral.
- Ⓓ A mediana amostral será mais precisa que a média amostral.
- Ⓔ A média amostral será mais eficiente que a mediana amostral.

||JUSTIFICATIVAS||

||A|| - Incorreta. A média amostral é mais eficiente que a mediana amostral (tem menor variância).*/
 ||B|| - Incorreta. Ambos os estimadores possuem o mesmo viés.*/
 ||C|| - Incorreta. Ambos os estimadores possuem o mesmo viés.*/
 ||D|| - Incorreta. A média amostral tem menor variância que a mediana amostral.*/
 ||E|| - Correta. Ambos os estimadores são não viesados e a média amostral tem menor variância que a mediana amostral.*/
Questão 49

Ⓐ Ao estimar certo parâmetro populacional β da distribuição de uma variável aleatória X a partir de um estimador T baseado em uma amostra aleatória simples com reposição, um estatístico determinou que a variância e o viés de T são, respectivamente, iguais 25 e 9.

Considerando a situação hipotética precedente, assinale a opção que corresponde ao erro quadrático médio do estimador T .

- Ⓐ 14
- Ⓑ 16
- Ⓒ 34
- Ⓓ 106
- Ⓔ 225

||JUSTIFICATIVAS||

||A|| - Incorreta. O candidato pode chegar a essa resposta adicionando o desvio-padrão ao viés: $5 + 9 = 14$.*/
 ||B|| - Incorreta. O candidato pode chegar a essa resposta subtraindo o viés da variância: $25 - 9 = 16$.*/
 ||C|| - Incorreta. O candidato pode chegar a essa resposta adicionando o viés à variância: $25 + 9 = 34$.*/
 ||D|| - Correta. $EQM(T) = var(T) + [viés(T)]^2 = 25 + 81 = 106$.*/
 ||E|| - Incorreta. O candidato pode chegar a essa resposta multiplicando a variância pelo viés: $25 \times 9 = 225$.*/
Questão 50

Ⓐ Ao estudar o tempo de chegada, em minutos, de clientes ao caixa de uma farmácia, um pesquisador modelou essa variável por uma distribuição exponencial. A fim de estimar o parâmetro da distribuição, tomou a seguinte amostra: 2, 3, 1, 5, 1, 6.

A partir da situação hipotética precedente, assinale a opção em que é apresentada, para essa amostra, a estimativa de máxima verossimilhança do desvio padrão populacional.

- Ⓐ 1
- Ⓑ 2,5
- Ⓒ 3
- Ⓓ 5
- Ⓔ 18

||JUSTIFICATIVAS||

||A|| - Incorreta. Trata-se da moda da série.*/
 ||B|| - Incorreta. Trata-se da mediana da série.*/
 ||C|| - Correta. No caso da distribuição exponencial, o estimador de máxima verossimilhança do desvio padrão populacional é a média amostral, aplicando-se o princípio da invariância dessa classe de estimadores.*/
 ||D|| - Incorreta. Trata-se da amplitude da série.*/
 ||E|| - Incorreta. Trata-se do total da amostra.*/
Questão 51

Questão 51

Uma máquina é utilizada para encher garrafas PET com certo tipo de refrigerante. Quando regulada a máquina, a quantidade de líquido que ela despeja na garrafa segue uma distribuição normal com média $\mu = 2$ L (hipótese nula) e desvio padrão $\sigma = 50$ mL. Certo dia, para realizar um controle de qualidade, foi selecionada uma amostra aleatória simples com reposição de 100 garrafas, e a média amostral alcançou o valor $M = 1,93$ L.

Com relação a essa situação hipotética, assinale a opção correta.

- Ⓐ A distribuição da média amostral terá média μ e desvio padrão σ .
- Ⓑ A evidência amostral é suficiente para assegurar que a máquina está perfeitamente regulada.
- Ⓒ A amostra selecionada demonstra que a média amostral é um estimador viesado.
- Ⓓ Se a máquina estiver perfeitamente regulada, a probabilidade de uma garrafa, selecionada ao acaso na produção, ter menos de 1.950 mL do refrigerante é nula.
- Ⓔ A potência do teste de hipóteses (bilateral) para a média, na situação em tela, coincide com a probabilidade de se rejeitar a hipótese de que a máquina esteja regulada.

||JUSTIFICATIVAS||

||A|| - Incorreta. A variância da média amostral será $V = 2500/100 = 25$. Assim, o desvio padrão amostral será $5 \text{ mL} \neq \sigma$.*/

||B|| - Incorreta. Na verdade, a evidência amostral não pode dar segurança absoluta; seria necessário estabelecer, ainda, a precisão da afirmação: margem de erro e nível de confiança.*/

||C|| - Incorreta. A média amostral é sempre um estimador não viesado.*/

||D|| - Incorreta. Embora seja baixa a probabilidade de se obter alguma garrafa com conteúdo abaixo de $\mu - \sigma$, essa probabilidade ainda é positiva.*/

||E|| - Correta. Trata-se da aplicação imediata do conceito de potência de um teste.*/

Texto 17A2-I

No caminho entre a casa de João e seu trabalho, há 16 semáforos. Um dia, João saiu atrasado de casa para o trabalho e, dos 16 semáforos pelos quais passou, 13 estavam fechados. “Hoje estou muito azarado mesmo”, pensou. Todos os semáforos da cidade onde João mora e trabalha funcionam independentemente uns dos outros, e cada um somente pode estar ou aberto ou fechado.

Questão 52

Considerando a situação hipotética apresentada no texto 17A2-I e utilizando aproximação normal com base na **Tabela - Normal Padrão de 0 a z**, fornecida ao final do **Caderno de Provas**, assinale a opção que apresenta o intervalo de 80% de confiança da proporção populacional de semáforos fechados em uma observação feita por outro motorista no mesmo caminho adotado por João.

- Ⓐ]0,65; 0,97[
- Ⓑ]0,71; 0,91[
- Ⓒ]0,75; 0,87[
- Ⓓ]0,40; 0,60[
- Ⓔ]0,34; 0,66[

||JUSTIFICATIVAS||

||A|| - Incorreta. A plausibilidade da alternativa decorre de a margem de erro ser próxima do valor real, e o candidato escolher um intervalo centrado em $0,81 = 13/16$.*/

||B|| - Incorreta. A plausibilidade da alternativa decorre de a margem de erro ser próxima do valor real, e o candidato escolher um intervalo centrado em $0,81 = 13/16$.*/

||C|| - Incorreta. A plausibilidade da alternativa decorre de a margem de erro ser próxima do valor real, e o candidato escolher um intervalo centrado em $0,81 = 13/16$.*/

||D|| - Incorreta. A plausibilidade da alternativa decorre de a margem de erro ser próxima do valor real, e o candidato escolher um intervalo centrado em 0,5, que é a proporção esperada dentro das hipóteses da questão.*/

||E|| - Correta. Na hipótese em questão, na população de semáforos da cidade, a proporção daqueles que estão fechados em certo momento é $p = 0,5$. Assim, a variável amostral indicadora da situação de cada semáforo em dado momento tem distribuição de Bernoulli com valor esperado 0,5 e variância $V = 0,5 \times 0,5 = 0,25$. Com base na aproximação normal, em uma amostra de tamanho $n = 16$, a variável que quantifica a proporção de semáforos fechados tem distribuição normal com média 0,5 e variância $0,25/16$. Portanto, a probabilidade de que a proporção m observada pelo outro transeunte difira de 0,5 por um erro e é calculada de acordo com a relação a seguir.

$$0,8 = P(|m - 0,5| < \text{erro}) = P(|Z| < \text{erro}/(0,25/16)^{0,5}).$$

Na tabela normal, $\text{erro}/(0,25/16)^{0,5} = 1,29$.

Assim, $\text{erro} = 1,29 \times (0,25/16)^{0,5} = 1,29 \times 0,5 / 4 = 0,16125$. Portanto, o intervalo desejado é $]0,5 - 0,16125 ; 0,5 + 0,16125[=]0,33875 ; 0,66125[$.*/



Questão 53

Ainda no que concerne à situação do texto 17A2-I, aplicando-se uma aproximação normal com base na **Tabela - Normal Padrão de 0 a z**, fornecida ao final do **Caderno de Provas**, é correto afirmar que a evidência amostral será suficiente para um motorista ser considerado azarado se o critério de decisão for a proporção superior a 0,7 de semáforos fechados em uma amostra correspondente ao mesmo trajeto adotado por João, desde que o nível de significância seja igual a

- Ⓐ 2,75%.
- Ⓑ 5,48%.
- Ⓒ 8,37%.
- Ⓓ 12,36%.
- Ⓔ 18,75%.

||JUSTIFICATIVAS||

||A|| - Incorreta. A hipótese a ser testada é H_0 : João está azarado.

O critério de decisão é que a proporção observada de semáforos fechados seja superior a 0,7. A probabilidade α de se cometer o erro de tipo I é o número procurado na questão. Indicando-se por RC a região correspondente aos valores maiores que 0,7, ou seja, $RC = \{x \in \mathbb{R} | x > 0,7\}$ e por p a proporção amostral, a probabilidade de se cometer o erro tipo I pode ser escrita da seguinte forma.

$$P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando João for azarado, com base na aproximação normal, com média 0,5 e desvio padrão igual a $0,5/4 = 0,125$, isto é, $p \sim N(0,5; 0,125)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, então

$$\alpha = P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}) =$$

$$P(p > 0,7 | p \sim N(0,5; 0,125)) =$$

$$P(Z > (0,7 - 0,5)/0,125) = P(Z > 1,6) =$$

$$0,5 - 0,4452 = 5,48\%.*/$$

||B|| - Correta. A hipótese a ser testada é H_0 : João está azarado.

O critério de decisão é que a proporção observada de semáforos fechados seja superior a 0,7. A probabilidade α de se cometer o

erro de tipo I é o número procurado na questão. Indicando-se por RC a região correspondente aos valores maiores que 0,7, ou seja, $RC = \{x \in \mathbb{R} \mid x > 0,7\}$ e por p a proporção amostral, a probabilidade de se cometer o erro tipo I pode ser escrita da seguinte forma.

$$P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando João for azarado, com base na aproximação normal, com média 0,5 e desvio padrão igual a $0,5/4 = 0,125$, isto é, $p \sim N(0,5; 0,125)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, então

$$\alpha = P(\text{erro I}) = P(p \in RC | H_0 \text{ é verdadeira}) =$$

$$P(p > 0,7 | p \sim N(0,5; 0,125)) =$$

$$P(Z > (0,7 - 0,5)/0,125) = P(Z > 1,6) =$$

$$0,5 - 0,4452 = 5,48\%.*$$

||C|| - Incorreta. A hipótese a ser testada é H_0 : João está azarado.

O critério de decisão é que a proporção observada de semáforos fechados seja superior a 0,7. A probabilidade α de se cometer o erro de tipo I é o número procurado na questão. Indicando-se por RC a região correspondente aos valores maiores que 0,7, ou seja, $RC = \{x \in \mathbb{R} \mid x > 0,7\}$ e por p a proporção amostral, a probabilidade de se cometer o erro tipo I pode ser escrita da seguinte forma.

$$P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando João for azarado, com base na aproximação normal, com média 0,5 e desvio padrão igual a $0,5/4 = 0,125$, isto é, $p \sim N(0,5; 0,125)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, então

$$\alpha = P(\text{erro I}) = P(p \in RC | H_0 \text{ é verdadeira}) =$$

$$P(p > 0,7 | p \sim N(0,5; 0,125)) =$$

$$P(Z > (0,7 - 0,5)/0,125) = P(Z > 1,6) =$$

$$0,5 - 0,4452 = 5,48\%.*$$

||D|| - Incorreta. A hipótese a ser testada é H_0 : João está azarado.

O critério de decisão é que a proporção observada de semáforos fechados seja superior a 0,7. A probabilidade α de se cometer o erro de tipo I é o número procurado na questão. Indicando-se por RC a região correspondente aos valores maiores que 0,7, ou seja, $RC = \{x \in \mathbb{R} \mid x > 0,7\}$ e por p a proporção amostral, a probabilidade de se cometer o erro tipo I pode ser escrita da seguinte forma.

$$P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando João for azarado, com base na aproximação normal, com média 0,5 e desvio padrão igual a $0,5/4 = 0,125$, isto é, $p \sim N(0,5; 0,125)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, então

$$\alpha = P(\text{erro I}) = P(p \in RC | H_0 \text{ é verdadeira}) =$$

$$P(p > 0,7 | p \sim N(0,5; 0,125)) =$$

$$P(Z > (0,7 - 0,5)/0,125) = P(Z > 1,6) =$$

$$0,5 - 0,4452 = 5,48\%.*$$

||E|| - Incorreta. A hipótese a ser testada é H_0 : João está azarado.

O critério de decisão é que a proporção observada de semáforos fechados seja superior a 0,7. A probabilidade α de se cometer o erro de tipo I é o número procurado na questão. Indicando-se por RC a região correspondente aos valores maiores que 0,7, ou seja, $RC = \{x \in \mathbb{R} \mid x > 0,7\}$ e por p a proporção amostral, a probabilidade de se cometer o erro tipo I pode ser escrita da seguinte forma.

$$P(\text{erro tipo I}) = P(p \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando João for azarado, com base na aproximação normal, com média 0,5 e desvio padrão igual a $0,5/4 = 0,125$, isto é, $p \sim N(0,5; 0,125)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, então

$$\alpha = P(\text{erro I}) = P(p \in RC | H_0 \text{ é verdadeira}) =$$

$$P(p > 0,7 | p \sim N(0,5; 0,125)) =$$

$$P(Z > (0,7 - 0,5)/0,125) = P(Z > 1,6) =$$

$$0,5 - 0,4452 = 5,48\%.*$$

Texto 17A2-II

Sabe-se que a indústria de computadores utiliza, como um dos componentes, GPUs (um tipo de processador especializado em cálculos paralelos, especialmente voltado para renderização de gráficos e vídeos). Em certo mercado, esses componentes são fornecidos por dois fornecedores, A e B. A velocidade de processamento, em teraflops, dos GPUs fornecidos por A, com distribuição normal, tem média de 145 e desvio padrão igual a 12. No caso dos GPUs fornecidos por B, a velocidade de processamento, em teraflops, com distribuição normal, tem média de 155 e desvio padrão igual a 20.

Um lote desses processadores, de origem não identificada, foi apreendido pela Receita Federal do Brasil e será leiloadado a um preço muito convidativo. A fabricante local de computadores (F) está avaliando a possibilidade de fazer uma oferta e, para tanto, deseja saber qual é o fornecedor dos equipamentos desse lote. O edital do leilão dispõe que, pouco antes do certame, será divulgada a velocidade média de uma amostra de 25 GPUs do lote.

Diante das informações disponíveis, a fabricante F estabeleceu que escolherá como produtor dos componentes o fornecedor A, se a média da amostra for inferior a certo número y — ainda a ser determinado com base na probabilidade de esse critério de decisão implicar erro; caso contrário, escolherá o fornecedor B.

Para modelar o problema e determinar o valor y , um pesquisador da fabricante F estabeleceu as seguintes notações.

hipóteses

- H_0 : Os GPUs são produzidos pelo fornecedor B.
- H_1 : Os GPUs são produzidos pelo fornecedor A.

erros

- tipo I: Inferir que os GPUs são de A, quando, na realidade, são de B.
- tipo II: Inferir que os GPUs são de B, quando, na realidade, são de A.



Questão 54

Considerando-se a situação hipotética apresentada no texto 17A2-II e aplicando-se aproximação normal de acordo com a Tabela - Normal Padrão de 0 a z , fornecida ao final do Caderno de Provas, é correto concluir que, caso y seja igual a 150, a probabilidade de ocorrer o erro do tipo I será de

- Ⓐ 8,28%.
- Ⓑ 10,56%.
- Ⓒ 12,91%.
- Ⓓ 16,13%.
- Ⓔ 17,24%.

JUSTIFICATIVAS

||A|| - Incorreta. O candidato pode chegar a essa resposta caso calcule $12/145$.*

||B|| - Correta. Indicando-se por RC a região correspondente aos valores menores que 150, ou seja, $RC = \{x \in \mathbb{R} \mid x < 150\}$, e por X a média amostral, a probabilidade de se cometer o erro tipo I pode ser escrita de acordo com a relação a seguir.

$$P(\text{erro do tipo I}) = P(X \in RC | H_0 \text{ é verdadeira}).$$

Quando H_0 for verdadeira, isto é, quando os GPUs forem de B, com base na aproximação normal, X terá média 155 e desvio padrão igual a $20/\sqrt{25} = 4$, isto é, $X \sim N(155, 16)$. Denotando-se por Z a variável aleatória com distribuição $N(0,1)$, a seguinte relação é obtida.

$$P(\text{erro tipo I}) = P(X \in RC | H_0 \text{ é verdadeira}) =$$

$$P(X < 150 | X \sim N(155, 16)) =$$

$$P(Z < (150 - 155)/4) =$$

$$P(Z < -1,25) = 0,10565 = 10,56\%.*$$

||C|| - Incorreta. O candidato pode chegar a essa resposta caso calcule $20/155$.*/

||D|| - Incorreta. O candidato pode chegar a essa resposta caso calcule $25/155$.*/

||E|| - Incorreta. O candidato pode chegar a essa resposta caso calcule $25/145$.*/

Questão 55

Considere que, na situação hipotética apresentada no texto 17A2-II, a fabricante F confie mais nos GPUs produzidos pelo fornecedor B, de modo que somente esteja disposta a incorrer no erro do tipo II com probabilidade de 1%. A partir dessas informações, e aplicando aproximação normal com base na **Tabela - Normal Padrão de 0 a z**, fornecida ao final do **Caderno de Provas**, assinale a opção em que é apresentado o valor de y que satisfaz essa disposição da fabricante F.

- A 149,59
- B 149,89
- C 150,59
- D 150,89
- E 151,59

||JUSTIFICATIVAS||

||A|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||B|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||C|| - Correta. Indicando-se por RC a região correspondente aos valores menores que y , ou seja, $RC = \{x \in \mathbb{R} \mid x < y\}$ e por X a média amostral, a probabilidade de se cometer um erro do tipo II pode ser escrita pela seguinte relação.

$$P(\text{erro tipo II}) = P(X \notin RC \mid H_1 \text{ é verdadeira}).$$

Quando H_1 for verdadeira, isto é, quando os GPUs forem de A, com base na aproximação normal, X terá média igual a 145 e desvio padrão igual a $12/\sqrt{25} = 2,4$, isto é, $X \sim N(145; 5,76)$. Denotando-se por Z a variável amostral com distribuição $N(0,1)$, a seguinte relação é obtida.

$$P(\text{erro II}) = P(X \notin RC \mid H_1 \text{ é verdadeira}) =$$

$$P(X > y \mid X \sim N(145; 5,76)) =$$

$$P(Z > (145 - y)/2,4).$$

Igualando a 1% é buscando na tabela normal a estatística correspondente temos $2,33 = (145 - y)/2,4$, e então $y = 150,59$.*/

||D|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||E|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

Questão 56

Ainda considerando a situação hipotética apresentada no texto 17A2-II e aplicando aproximação normal com base na **Tabela - Normal Padrão de 0 a z**, fornecida ao final do **Caderno de Provas**, assinale a opção em que é apresentado um valor de y para o qual a probabilidade de ocorrer o erro do tipo I seja igual à de ocorrer o erro do tipo II.

- A 148,75
- B 149,25
- C 149,75
- D 150,25
- E 150,75

||JUSTIFICATIVAS||

||A|| - Correta. Indicando por RC a região correspondente aos valores menores que y , ou seja, $RC = \{x \in \mathbb{R} \mid x < y\}$ e por X a média amostral, a probabilidade de se cometer o erro tipo I pode ser escrita: $P(\text{erro I}) = P(X \in RC \mid H_0 \text{ é verdadeira})$.

Quando H_0 for verdadeira, isto é, os GPUs forem de B, com base na aproximação, X terá distribuição normal, com média 155 e desvio padrão igual a $20/\sqrt{25} = 4$, isto é, $X \sim N(155, 16)$.

Denotando por Z a variável amostral com distribuição $N(0,1)$, tem-se

$$P(\text{erro tipo I}) = P(X \in RC \mid H_0 \text{ é verdadeira}) =$$

$$P(X < y \mid X \sim N(155, 16)) = P(Z < (y - 155)/4).$$

Procedendo de modo análogo,

$$P(\text{erro tipo II}) = P(X \notin RC \mid H_1 \text{ é verdadeira}) =$$

$$P(X > y \mid X \sim N(145; 5,76)) = P(Z > (145 - y)/2,4).$$

Para que as probabilidades sejam iguais, dada simetria da normal, o valor de y deve ser encolhido de modo que $(y - 155)/4 = (145 - y)/2,4$, e então $y = 148,75$.*/

||B|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||C|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||D|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

||E|| - Incorreta. A plausibilidade da alternativa decorre da proximidade com o valor correto.*/

Questão 57

Considere que, na situação hipotética apresentada no texto 17A2-II, seja escolhido um valor de y para o qual a probabilidade de ocorrer o erro do tipo I seja igual à de ocorrer o erro do tipo II. Nesse caso, aplicando-se a aproximação normal com base na **Tabela - Normal Padrão de 0 a z**, fornecida ao final do **Caderno de Provas**, a probabilidade de ocorrer o erro do tipo II será de

- A 5,94%.
- B 8,41%.
- C 12,7%.
- D 25%.
- E 50%.

||JUSTIFICATIVAS||

||A|| - Correta. Indicando por RC a região correspondente aos valores menores que y , ou seja, $RC = \{x \in \mathbb{R} \mid x < y\}$ e por X a média amostral, a probabilidade de se cometer o erro tipo I pode ser escrita: $P(\text{erro tipo I}) = P(X \in RC \mid H_0 \text{ é verdadeira})$.

Quando H_0 for verdadeira, isto é, os GPUs forem de B, com base na aproximação, X tem distribuição normal, com média 155 e desvio padrão igual a $20/\sqrt{25} = 4$, isto é, $X \sim N(155, 16)$. Denotando por Z a variável amostral com distribuição $N(0,1)$, tem-se

$$P(\text{erro tipo I}) = P(X \in RC \mid H_0 \text{ é verdadeira}) =$$

$$P(X < y \mid X \sim N(155, 16)) = P(Z < (y - 155)/4).$$

De modo análogo,

$$P(\text{erro II}) = P(X \notin RC \mid H_1 \text{ é verdadeira}) =$$

$$P(X > y \mid X \sim N(145; 5,76)) = P(Z > (145 - y)/2,4).$$

Para que as probabilidades sejam iguais, dada simetria da normal, o valor de y deve ser encolhido de modo que

$$(y - 155)/4 = (145 - y)/2,4, \text{ e então } y = 148,75.$$

Atribuindo-se tal valor em qualquer uma das probabilidades mencionadas, tem-se

$$P(Z < (148,75 - 155)/4) = P(Z < -1,5625) = 5,94\%$$

(sendo a última igualdade obtida a partir da busca da referida probabilidade na tabela de distribuição normal.*/

||B|| - Incorreta. A plausibilidade dessa alternativa está em sua similaridade com a correta.*/

||C|| - Incorreta. A plausibilidade dessa alternativa está em sua similaridade com a correta.*/

||D|| - Incorreta. O candidato pode acreditar que são igualmente prováveis, além dos dois erros, as situações “inferir que os GPUs são de A, sendo efetivamente de A” e “inferir que os GPUs são de B, sendo efetivamente de B”, e assim dividir 100% pelas 4 situações.*/

||E|| - Incorreta. O candidato pode acreditar que as probabilidades de ocorrer os dois erros devem somar 100%, e então, dado que são iguais, atribuir 50% para cada.*/

Texto 17A3

Conforme a literatura básica de estatística, o coeficiente de correlação linear r é adimensional e pode variar de -1 a 1 , ou seja $-1 \leq r \leq 1$.

Questão 58

Considerando as informações do texto 17A3, julgue os itens a seguir.

- I Se o valor de r estiver próximo de $+1$, a reta será crescente e representará a correlação entre os valores das variáveis, com uma mínima dispersão entre os pontos obtidos pelas variáveis e os pontos da reta.
- II Para duas variáveis, X e Y , se o coeficiente de correlação for, aproximadamente, $r = 0,9813$, então 96,29% das variações totais serão explicadas pela reta de regressão $Y = a + bX$.
- III Caso o coeficiente de correlação seja $r = -1$, a reta ajustada explicará toda a variação de Y e, por consequência, o ajuste linear será excelente.

Assinale a opção correta.

- A** Apenas o item I está certo.
- B** Apenas o item II está certo.
- C** Apenas os itens I e III estão certos.
- D** Apenas os itens II e III estão certos.
- E** Todos os itens estão certos.

||JUSTIFICATIVAS||

||A|| - Incorreta. Todos os itens estão certos. Quando $r = -1$, tem-se a correlação linear negativa perfeita, enquanto que para $r = +1$ a correlação linear é positiva perfeita. Portanto, o item I está certo. Além disso, sabemos que o coeficiente de determinação para $r = 0,9813$ será dado por $r^2 = 0,9629$. Ou seja, a relação entre as variáveis X e Y serão explicadas em 96,29% dos casos e, desta forma, o item II também está certo. Também é sabido que o coeficiente de determinação será $r^2 = 1$ quando $r = 1$ ou $r = -1$. Dessa forma, as variáveis X e Y serão explicadas em 100% dos casos, de modo que o item III também está certo. Portanto, todos os itens estão certos. */

||B|| - Incorreta. Todos os itens estão certos. Quando $r = -1$, tem-se a correlação linear negativa perfeita, enquanto que para $r = +1$ a correlação linear é positiva perfeita. Portanto, o item I está certo. Além disso, sabemos que o coeficiente de determinação para $r = 0,9813$ será dado por $r^2 = 0,9629$. Ou seja, a relação entre as variáveis X e Y serão explicadas em 96,29% dos casos e, desta forma, o item II também está certo. Também é sabido que o coeficiente de determinação será $r^2 = 1$ quando $r = 1$ ou $r = -1$. Dessa forma, as variáveis X e Y serão explicadas em 100% dos casos, de modo que o item III também está certo. Portanto, todos os itens estão certos. */

||C|| - Incorreta. Todos os itens estão certos. Quando $r = -1$, tem-se a correlação linear negativa perfeita, enquanto que para $r = +1$ a correlação linear é positiva perfeita. Portanto, o item I está certo. Além disso, sabemos que o coeficiente de determinação para $r = 0,9813$ será dado por $r^2 = 0,9629$. Ou seja, a relação entre as variáveis X e Y serão explicadas em 96,29% dos casos e, desta forma, o item II também está certo. Também é sabido que o coeficiente de determinação será $r^2 = 1$ quando $r = 1$ ou $r = -1$. Dessa forma, as variáveis X e Y serão explicadas em 100% dos casos, de modo que o item III também está certo. Portanto, todos os itens estão certos. */

||D|| - Incorreta. Todos os itens estão certos. Quando $r = -1$, tem-se a correlação linear negativa perfeita, enquanto que para $r = +1$ a correlação linear é positiva perfeita. Portanto, o item I está certo. Além disso, sabemos que o coeficiente de determinação para $r = 0,9813$ será dado por $r^2 = 0,9629$. Ou seja, a relação entre as variáveis X e Y serão explicadas em 96,29% dos casos e, desta forma, o item II também está certo. Também é sabido que o coeficiente de determinação será $r^2 = 1$ quando $r = 1$ ou $r = -1$. Dessa forma, as variáveis X e Y serão explicadas em 100% dos casos, de modo que o item III também está certo. Portanto, todos os itens estão certos. */

||E|| - Correta. Todos os itens estão certos. Quando $r = -1$, tem-se a correlação linear negativa perfeita, enquanto que para $r = +1$ a correlação linear é positiva perfeita. Portanto, o item I está certo. Além disso, sabemos que o coeficiente de determinação para $r = 0,9813$ será dado por $r^2 = 0,9629$. Ou seja, a relação entre as variáveis X e Y serão explicadas em 96,29% dos casos e, desta forma, o item II também está certo. Também é sabido que o coeficiente de determinação será $r^2 = 1$ quando $r = 1$ ou $r = -1$. Dessa forma, as variáveis X e Y serão explicadas em 100% dos casos, de modo que o item III também está certo. Portanto, todos os itens estão certos. */

Questão 59

Para as variáveis aleatórias X e Y , 6 amostras foram obtidas, de modo que $\Sigma X = 21$, $\Sigma Y = 47$, $\Sigma X^2 = 91$, $\Sigma Y^2 = 409$ e $\Sigma(X \cdot Y) = 186$. Assim, o coeficiente de correlação ($r_{X,Y}$) é dado pela seguinte relação.

$$r_{X,Y} = \frac{n \cdot \Sigma X \cdot Y - (\Sigma X) \cdot (\Sigma Y)}{\sqrt{[n \cdot \Sigma X^2 - (\Sigma X)^2] \cdot [n \Sigma Y^2 - (\Sigma Y)^2]}}$$

A partir das informações precedentes e daquelas apresentadas no texto 17A3, é correto afirmar, considerando-se $\sqrt{25.725} = 160,39$, que o nível de ajuste entre as variáveis é

- A** inferior a 82% e superior ou igual a 77%.
- B** inferior a 77% e superior ou igual a 67%.
- C** inferior a 67% e superior ou igual a 50%.
- D** inferior a 50% e superior ou igual a 45%.
- E** inferior a 45%.

||JUSTIFICATIVAS||

||A|| - Incorreta. A partir dos dados apresentados, tem-se que $r = (6.186 - 21.47)/([6.91-212].[6.409-472]1/2) = 129/(25725)^{1/2} = 129/160,39 = 0,8043$. Porém, o nível de ajuste é dado por $r^2 = 0,80432 \cong 0,6470$. Ou seja, 64,70%. Portanto, a opção está errada.*/

||B|| - Incorreta. A partir dos dados apresentados, tem-se que $r = (6.186 - 21.47)/([6.91-212].[6.409-472]1/2) = 129/(25725)^{1/2} = 129/160,39 = 0,8043$. Porém, o nível de ajuste é dado por $r^2 = 0,80432 \cong 0,6470$. Ou seja, 64,70%. Portanto, a opção está errada.*/

||C|| - Correta. A partir dos dados apresentados, tem-se que $r = (6.186 - 21.47)/([6.91-212].[6.409-472]1/2) = 129/(25725)^{1/2} = 129/160,39 = 0,8043$. Porém, o nível de ajuste é dado por $r^2 = 0,80432 \cong 0,6470$. Ou seja, 64,70%. Portanto, a opção está correta.*/

||D|| - Incorreta. A partir dos dados apresentados, tem-se que $r = (6.186 - 21.47)/([6.91-212].[6.409-472]1/2) = 129/(25725)^{1/2} = 129/160,39 = 0,8043$. Porém, o nível de ajuste é dado por $r^2 = 0,80432 \cong 0,6470$. Ou seja, 64,70%. Portanto, a opção está errada.*/

||E|| - Incorreta. A partir dos dados apresentados, tem-se que $r = (6.186 - 21.47)/([6.91-212].[6.409-472]1/2) = 129/(25725)^{1/2} = 129/160,39 = 0,8043$. Porém, o nível de ajuste é dado por $r^2 = 0,80432 \cong 0,6470$. Ou seja, 64,70%. Portanto, a opção está errada.*/

Questão 60

Uma pesquisa científica foi realizada para investigar a relação entre o tempo médio, por dia, que um indivíduo gasta navegando nas redes sociais (em minutos) e sua idade (em anos). Os dados foram obtidos em uma entrevista com os usuários, que responderam sua idade e seu tempo de navegação diário nas redes. A partir dos dados observados experimentalmente, um pesquisador elaborou o seguinte modelo de regressão linear simples ajustado, em que I corresponde à idade e T , ao tempo.

$$T = 163,12 - 0,9532 \times I$$

Considerando a situação hipotética precedente, julgue os itens a seguir.

- I A correlação entre as variáveis T e I é linear positiva.
- II T é uma variável dependente de I .
- III Se uma pessoa de 20 anos de idade responder que gasta diariamente 152 minutos nas redes sociais, então o erro, ou seja, a diferença entre o valor efetivo e o valor previsto, será superior a 7 minutos.

Assinale a opção correta.

- ☐ A Apenas o item I está certo.
- ☐ B Apenas o item II está certo.
- ☐ C Apenas os itens I e III estão certos.
- ☐ D Apenas os itens II e III estão certos.
- ☐ E Todos os itens estão certos.

JUSTIFICATIVAS

||A|| - Incorreta. O item I está errado, porque o modelo linear apresentado possui coeficiente angular negativo. Portanto, a reta dos mínimos quadrados possui inclinação decrescente, mostrando que a correlação é linear negativa.*/

||B|| - Incorreta. O item II está certo, mas o item III também está. O item II está certo porque o modelo matemático apresenta a variável T como resultado da variação da variável I . Portanto, a variável I é independente, e a variável T é dependente de I .*/

||C|| - Incorreta. Embora o item III esteja certo, o item I está errado, porque o modelo linear apresentado possui coeficiente angular negativo. Portanto, a reta dos mínimos quadrados possui inclinação decrescente, mostrando que a correlação é linear negativa.*/

||D|| - Correta. O item II está certo, porque o modelo matemático apresenta a variável T como resultado da variação da variável I . Portanto, a variável I é independente, e a variável T é dependente de I . O item III também está certo, pois, para $I = 20$, tem-se que $T = 163,12 - 0,9532 \cdot 20 = 144,056$. Assim, $152 - 144,056 = 7,944$ (ou seja, mais que 7 minutos).*/

||E|| - Incorreta. Embora os itens II e III estejam certos, o item I está errado, porque o modelo linear apresentado possui coeficiente angular negativo. Portanto, a reta dos mínimos quadrados possui inclinação decrescente, mostrando que a correlação é linear negativa.*/

Questão 61

As técnicas de amostragem podem ser classificadas em probabilísticas (aleatórias) ou não probabilísticas (não aleatórias). Assinale a opção em que são apresentados apenas exemplos de técnicas de amostragem aleatória.

- ☐ A amostragem sistemática, amostragem estratificada e amostragem por cotas
- ☐ B amostragem aleatória simples, amostragem por julgamento e amostragem a esmo
- ☐ C amostragem aleatória simples, amostragem estratificada e amostragem por conglomerados
- ☐ D amostragem a esmo, amostragem intencional e amostragem por cotas
- ☐ E amostragem por julgamento, amostragem por conglomerado e amostragem sistemática

JUSTIFICATIVAS

||A|| - Incorreta. Conforme a literatura da área, são exemplos de técnica de amostragem aleatória a amostragem simples, a amostragem sistemática, a amostragem estratificada e a amostragem por conglomerado.*/

||B|| - Incorreta. Conforme a literatura da área, são exemplos de técnica de amostragem aleatória a amostragem simples, a amostragem sistemática, a amostragem estratificada e a amostragem por conglomerado.*/

||C|| - Correta. Conforme a literatura da área, são exemplos de técnica de amostragem aleatória a amostragem simples, a amostragem sistemática, a amostragem estratificada e a amostragem por conglomerado.*/

||D|| - Incorreta. Conforme a literatura da área, são exemplos de técnica de amostragem aleatória a amostragem simples, a amostragem sistemática, a amostragem estratificada e a amostragem por conglomerado.*/

||E|| - Incorreta. Conforme a literatura da área, são exemplos de técnica de amostragem aleatória a amostragem simples, a amostragem sistemática, a amostragem estratificada e a amostragem por conglomerado.*/

Questão 62

Um professor fará uma pesquisa amostral de tamanho 5 em uma sala de aula, que possui 50 alunos. Para a seleção da amostra probabilística, o professor utilizou a lista de presença, organizada em ordem alfabética, dividindo a relação de nomes em 5 grupos de 10 alunos. Do primeiro grupo de 10 alunos, ele selecionou um nome de forma aleatória e verificou a posição dele na lista de chamada. A seleção dos demais foi feita adicionando-se 10 unidades à posição do nome anteriormente selecionado.

É correto concluir das informações apresentadas na situação precedente que a pesquisa estatística em apreço utilizará a metodologia de amostragem

- ☐ A sistemática.
- ☐ B por cotas.
- ☐ C estratificada.
- ☐ D aleatória simples.
- ☐ E intencional.

JUSTIFICATIVAS

||A|| - Correta. Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, tem-se uma amostragem sistemática. A situação hipotética afirma ser um modelo probabilístico, e a amostra sistemática é a que adequadamente se enquadra à situação descrita.*/

||B|| - Incorreta. Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, tem-se uma amostragem sistemática. A situação hipotética afirma ser um modelo probabilístico, e a amostra sistemática é a que adequadamente se enquadra à situação descrita.*/

||C|| - Incorreta. Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, tem-se uma amostragem sistemática. A situação hipotética afirma ser um modelo probabilístico, e a amostra sistemática é a que adequadamente se enquadra à situação descrita.*/

||D|| - Incorreta. Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, tem-se uma amostragem sistemática. A situação hipotética afirma ser um modelo probabilístico, e a amostra sistemática é a que adequadamente se enquadra à situação descrita.*/

||E|| - Incorreta. Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, tem-se uma amostragem sistemática. A situação hipotética afirma ser um modelo probabilístico, e a amostra sistemática é a que adequadamente se enquadra à situação descrita.* /

Questão 63

Suponha que, no ajuste de uma reta de regressão linear simples de uma variável Y em uma variável X , o coeficiente de determinação observado tenha sido igual a 0,1296. Nesse caso, o módulo do coeficiente de correlação amostral entre X e Y é igual a

- A 0,75.
- B 0,64.
- C 0,51.
- D 0,42.
- E 0,36.

||JUSTIFICATIVAS||

||A|| - Incorreta. O coeficiente de determinação é calculado por r^2 . Assim, se $r^2 = 0,1296$, o valor positivo para r será de 0,36.* /

||B|| - Incorreta. O coeficiente de determinação é calculado por r^2 . Assim, se $r^2 = 0,1296$, o valor positivo para r será de 0,36.* /

||C|| - Incorreta. O coeficiente de determinação é calculado por r^2 . Assim, se $r^2 = 0,1296$, o valor positivo para r será de 0,36.* /

||D|| - Incorreta. O coeficiente de determinação é calculado por r^2 . Assim, se $r^2 = 0,1296$, o valor positivo para r será de 0,36.* /

||E|| - Correta. O coeficiente de determinação é calculado por r^2 . Assim, se $r^2 = 0,1296$, o valor positivo para r será de 0,36.* /

Questão 64

Considere que, em uma pesquisa realizada com 500 consumidores, se pretenda estimar a porcentagem daqueles satisfeitos com o abastecimento de água de uma região administrativa do Distrito Federal. Nessa situação, o tamanho mínimo que a amostra aleatória simples deve apresentar para que o erro amostral não seja superior a 5% é

- A inferior a 100 e superior ou igual a 50.
- B inferior a 150 e superior ou igual a 100.
- C inferior a 250 e superior ou igual a 200.
- D inferior a 350 e superior ou igual a 300.
- E inferior a 450 e superior ou igual a 400.

||JUSTIFICATIVAS||

||A|| - Incorreta. Para os dados apresentados, tem-se $n_0 = 1/0,052 = 400$. Assim, $n = 500.400/(500+400) = 222$. Portanto a opção está errada.* /

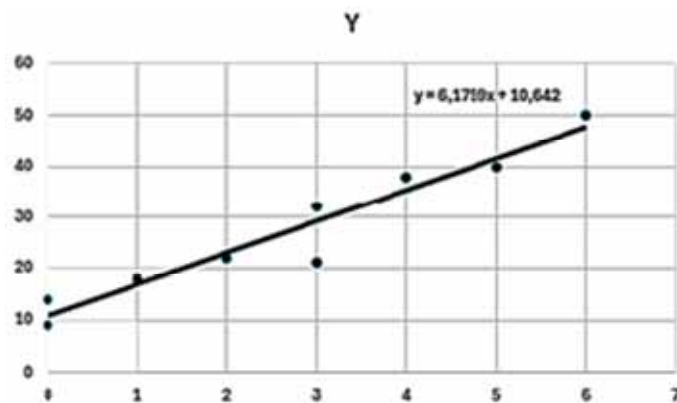
||B|| - Incorreta. Para os dados apresentados, tem-se $n_0 = 1/0,052 = 400$. Assim, $n = 500.400/(500+400) = 222$. Portanto a opção está errada.* /

||C|| - Correta. Para os dados apresentados, tem-se $n_0 = 1/0,052 = 400$. Assim, $n = 500.400/(500+400) = 222$. Portanto a opção está correta.* /

||D|| - Incorreta. Para os dados apresentados, tem-se $n_0 = 1/0,052 = 400$. Assim, $n = 500.400/(500+400) = 222$. Portanto a opção está errada.* /

||E|| - Incorreta. Para os dados apresentados, tem-se $n_0 = 1/0,052 = 400$. Assim, $n = 500.400/(500+400) = 222$. Portanto a opção está errada.* /

Questão 65



A figura precedente corresponde ao diagrama de dispersão para as variáveis X e Y , e a sua linha de tendência é dada pela reta dos mínimos quadrados da equação $Y = 6,1759X + 10,642$. Considerando-se que o valor observado da variável Y seja igual a 50, quando $X = 6$, o valor do erro de predição será

- A 1,2653.
- B 2,3026.
- C 3,6589.
- D 4,0251.
- E 5,2589.

||JUSTIFICATIVAS||

||A|| - Incorreta. Para $Y = 6,1759X + 10,642$ e $X = 6$, tem-se $Y = 47,6974$. Assim, o erro será dado por $e = 50 - 47,6974 = 2,3026$. Portanto, a opção está errada.* /

||B|| - Correta. Para $Y = 6,1759X + 10,642$ e $X = 6$, tem-se $Y = 47,6974$. Assim, o erro será dado por $e = 50 - 47,6974 = 2,3026$. Portanto, a opção está correta.* /

||C|| - Incorreta. Para $Y = 6,1759X + 10,642$ e $X = 6$, tem-se $Y = 47,6974$. Assim, o erro será dado por $e = 50 - 47,6974 = 2,3026$. Portanto, a opção está errada.* /

||D|| - Incorreta. Para $Y = 6,1759X + 10,642$ e $X = 6$, tem-se $Y = 47,6974$. Assim, o erro será dado por $e = 50 - 47,6974 = 2,3026$. Portanto, a opção está errada.* /

||E|| - Incorreta. Para $Y = 6,1759X + 10,642$ e $X = 6$, tem-se $Y = 47,6974$. Assim, o erro será dado por $e = 50 - 47,6974 = 2,3026$. Portanto, a opção está errada.* /

Questão 66

Um estatístico está desenvolvendo um modelo de regressão linear simples para prever o valor de uma variável dependente Y com base em uma variável independente X . Ele obteve a seguinte equação estimada.

$$Y = 3 + 2X$$

Com base nessas informações, assinale a opção correta.

- A O coeficiente angular da reta é 3, o que significa que, para cada unidade de aumento em X , Y aumenta em 3 unidades.
- B O coeficiente linear da reta é 2, que é o valor esperado de Y quando $X = 0$.
- C Quando X for igual a 0, o valor estimado de Y será 3.
- D O coeficiente angular da reta indica que Y diminui 2 unidades para cada unidade de aumento em X .
- E O modelo de regressão apresentado não pode ser usado para previsão, pois não apresenta um coeficiente de determinação (r^2).

||JUSTIFICATIVAS||

||A|| - Incorreta. O coeficiente angular da reta (2) indica a variação de Y para cada unidade de aumento em X . Portanto, para cada unidade de X , Y aumentará em 2 unidades. */

||B|| - Incorreta. O coeficiente linear (3) indica o valor estimado de Y quando $X = 0$. */

||C|| - Correta. Se $X = 0$, $Y = 3 + 2 \cdot 0$. Ou seja, $Y = 3$. Portanto, a opção está correta.*/

||D|| - Incorreta. O coeficiente angular é positivo. Portanto, a variação é crescente, e não decrescente. */

||E|| - Incorreta. O modelo pode ser usado para previsão, mesmo sem apresentar diretamente o coeficiente de determinação (r^2). */

Questão 67

A análise de variância (ANOVA) é uma técnica estatística amplamente utilizada para comparar as médias de diferentes grupos. Acerca da ANOVA, assinale a opção correta.

- Ⓐ A ANOVA só pode ser utilizada quando há exatamente dois grupos para comparação.
- Ⓑ O pressuposto de homocedasticidade (igualdade de variâncias entre os grupos) não é necessário para se realizar a ANOVA.
- Ⓒ O teste F na ANOVA avalia se as variâncias entre os grupos são iguais.
- Ⓓ A ANOVA de uma via permite testar se as médias de três ou mais grupos são significativamente diferentes entre si.
- Ⓔ A ANOVA só é válida quando os tamanhos dos grupos são exatamente iguais.

||JUSTIFICATIVAS||

||A|| - Incorreta. A análise de variância (ANOVA) pode ser utilizada para comparar três ou mais tratamentos. */

||B|| - Incorreta. Para realizar a análise de variância (ANOVA), é necessário que os dados satisfaçam alguns critérios, entre eles: variável de resposta contínua; pelo menos um fator categórico com dois ou mais níveis; dados de populações que sigam a distribuição normal; variâncias iguais entre fatores; independência das observações; e homogeneidade de variâncias.*/

||C|| - Incorreta. O teste F na ANOVA avalia a razão entre a variabilidade explicada pelos grupos (variabilidade entre grupos) e a variabilidade residual (variabilidade dentro dos grupos). Ele não testa diretamente as variâncias, mas sim as médias. */

||D|| - Correta. A ANOVA é usada para avaliar diferenças nas médias entre dois ou mais grupos, sendo particularmente útil para identificar se pelo menos um grupo difere significativamente dos demais. */

||E|| - Incorreta. A ANOVA pode ser aplicada mesmo quando os tamanhos dos grupos são diferentes, desde que os demais pressupostos sejam atendidos. */

Questão 68

Acerca dos critérios de mínimos quadrados ordinários (MQO) e de máxima verossimilhança (MV), amplamente utilizados na estimação de parâmetros em modelos estatísticos, assinale a opção correta.

- Ⓐ O critério de MQO é sempre preferível ao critério de MV, pois não faz pressupostos sobre a distribuição dos erros.
- Ⓑ O critério de MV requer que os erros sigam uma distribuição normal, para ser aplicado em qualquer situação.
- Ⓒ Os critérios de MQO e de MV fornecem os mesmos resultados em todos os casos, independentemente das suposições feitas sobre os erros ou da forma do modelo.
- Ⓓ O critério de MV é usado exclusivamente para modelos lineares, enquanto o critério de MQO pode ser usado em qualquer tipo de modelo.
- Ⓔ O critério de MQO minimiza a soma dos quadrados dos resíduos, enquanto o critério de MV maximiza a probabilidade (ou a densidade) de se observarem os dados amostrais, conforme os parâmetros do modelo.

||JUSTIFICATIVAS||

||A|| - Incorreta. O MQO não faz pressupostos sobre a distribuição dos erros, mas isso não implica que ele seja sempre preferível ao MV. O MV é mais flexível, especialmente em contextos em que a distribuição dos erros é conhecida. */

||B|| - Incorreta. O critério de MV não exige normalidade dos erros em todos os casos, podendo ser aplicado a qualquer distribuição especificada. */

||C|| - Incorreta. Os resultados fornecidos pelos critérios de MQO e MV podem diferir, de acordo com as suposições feitas sobre os erros e com o tipo de modelo adotado. */

||D|| - Incorreta. O MV não é exclusivo de modelos lineares; pode ser usado em modelos não lineares e com diferentes distribuições de erro. Já o MQO é mais restrito a modelos lineares ou que possam ser linearizados. */

||E|| - Correta. A literatura afirma que o MQO minimiza a soma dos quadrados dos resíduos, enquanto o MV maximiza a função de verossimilhança, que reflete a probabilidade de se observarem os dados amostrais. */

Questão 69

A análise de resíduos é uma etapa essencial na validação de modelos estatísticos, especialmente em regressão linear. Acerca dos resíduos e da sua análise, assinale a opção correta.

- Ⓐ Em estatística, resíduos são as diferenças entre os valores previstos pelo modelo e os valores ajustados.
- Ⓑ A análise de resíduos é usada para verificar se o modelo atende aos pressupostos de linearidade, homocedasticidade, normalidade e independência.
- Ⓒ Resíduos padronizados são calculados por meio da divisão de cada resíduo pela soma total dos resíduos.
- Ⓓ Em um gráfico de resíduos *versus* valores ajustados, a presença de resíduos não aleatórios indica que o modelo é apropriado.
- Ⓔ Se os resíduos seguem uma distribuição uniforme, o pressuposto de normalidade é considerado atendido.

||JUSTIFICATIVAS||

||A|| - Incorreta. Resíduos são as diferenças entre os valores previstos pelo modelo e os valores ajustados.*/

||B|| - Correta. A análise de resíduos é usada para verificar se o modelo atende aos pressupostos de linearidade, homocedasticidade, normalidade e independência.*/

||C|| - Incorreta. Resíduos padronizados são calculados dividindo cada resíduo pela soma total dos resíduos.*/

||D|| - Incorreta. A presença de resíduos não aleatórios em um gráfico de resíduos *versus* valores ajustados indica que o modelo é apropriado.*/

||E|| - Incorreta. Se os resíduos seguem uma distribuição uniforme, o pressuposto de normalidade é considerado atendido.*/

Questão 70

Em relação aos modelos de regressão linear simples sob o método de mínimos quadrados ordinários, assinale a opção correta.

- Ⓐ A regressão linear simples é usada para modelar a relação entre uma variável dependente e duas ou mais variáveis independentes.
- Ⓑ O coeficiente angular da equação de regressão representa o valor previsto da variável dependente quando a variável independente é igual a zero.
- Ⓒ A equação de regressão linear simples pode ser expressa como $y = \beta_0 + \beta_1 x + \epsilon$, onde β_0 é o intercepto, β_1 é o coeficiente angular, e ϵ é o termo de erro.
- Ⓓ Um dos pressupostos da regressão linear simples é que os resíduos apresentem variâncias diferentes ao longo dos valores previstos.
- Ⓔ A principal medida para avaliar o ajuste de um modelo de regressão linear simples é a mediana dos resíduos.

||JUSTIFICATIVAS||

||A|| - Incorreta. A regressão linear simples envolve apenas uma variável independente, enquanto a regressão linear múltipla lida com duas ou mais variáveis independentes. */

||B|| - Incorreta. O coeficiente angular representa a variação de y em relação a x , e não o valor previsto quando $x = 0$. */

||C|| - Correta. A equação apresentada é a forma correta da regressão linear simples, com uma única variável independente e incluindo o termo de erro. */

||D|| - Incorreta. Um pressuposto da regressão linear é a homocedasticidade, ou seja, os resíduos devem ter variâncias constantes ao longo dos valores previstos. */

||E|| - Incorreta. A principal medida para avaliar o ajuste de um modelo é o coeficiente de determinação (r^2), que indica a proporção da variância de y explicada por x . */