

- Nas questões a seguir, marque, para cada uma, a única opção correta, de acordo com o respectivo comando. Para as devidas marcações, use a **Folha de Respostas**, único documento válido para a correção das suas respostas.
- Eventuais espaços livres — identificados ou não pela expressão “Espaço livre” — que constarem deste caderno de provas poderão ser utilizados para rascunho.

PROVA OBJETIVA P₂
-- CONHECIMENTOS ESPECÍFICOS --

ESTATÍSTICA

Questão 1

X	frequência absoluta
0	10
1	7
2	7
3	10
total	34

Da análise da tabela de frequências precedente, em que é apresentada a distribuição do número diário X de ocorrências de determinado evento em uma amostra de 34 dias, conclui-se que o coeficiente de assimetria é igual a

- A 0,0.
- B 1,2.
- C 1,5.
- D 2,5.
- E 3,0.

Questão 2

Considerando-se, a partir de duas variáveis aleatórias discretas X e Y , a definição de dois eventos aleatórios, A e B , tais que $A = \{X = 1, Y = 0\}$ e $B = \{X = 0, Y = 1\}$, é correto afirmar que, se as variáveis aleatórias X e Y forem mutuamente independentes, e se $P(X = 0) = P(X = 1) = P(Y = 0) = P(Y = 1)$, então $P(A \cap B)$ é igual a

- A 0.
- B 0,0625.
- C 0,125.
- D 0,25.
- E 0,5.

Questão 3

Certo contribuinte entregou com atraso 30% das declarações eletrônicas que preencheu e 50% dessas declarações entregues continham erros de preenchimento. As declarações entregues por esse contribuinte dentro do prazo não continham erros de preenchimento.

Nessa situação hipotética, a probabilidade de se escolher ao acaso uma declaração entregue com erro de preenchimento é igual a

- A 0.
- B 0,15.
- C 0,30.
- D 0,50.
- E 0,80.

Questão 4

Caso uma variável aleatória contínua Y cuja função de densidade de probabilidade seja dada pela expressão

$$f_Y(y) = \begin{cases} Ay^5, & \text{se } y \in (0, 1), \\ 0, & \text{se } y \notin (0, 1), \end{cases}$$

em que A é uma constante positiva, então a probabilidade $P(Y < 0,5)$ é igual a

- A 0,000125.
- B 0,0078125.
- C 0,015625.
- D 0,03125.
- E 0,0625.

Questão 5

Se X for uma variável aleatória contínua tal que

$$P(X > c) = \exp(-3\sqrt{c}),$$

para $c > 0$, e $P(X > c) = 1$, para $c < 0$, então a função de densidade de probabilidade de X para $x > 0$ será dada pela expressão

- A $f(x) = \exp(-3\sqrt{x})$.
- B $f(x) = -3 \times \exp(-3\sqrt{x})$.
- C $f(x) = -3\sqrt{x} \times \exp(-3\sqrt{x})$.
- D $f(x) = \frac{3}{2\sqrt{x}} \times \exp(-3\sqrt{x})$.
- E $f(x) = 3\sqrt{x} \times \exp(-3\sqrt{x})$.

Questão 6

Considere uma variável aleatória W tal que $P(W > c) = P(W < -c)$, para qualquer valor real c . Nesse caso, se a variância de W^2 for igual a 10 e se a variância de W for igual a 3, então o valor esperado de W^4 será igual a

- A 0,5.
- B 1.
- C 7.
- D 13.
- E 19.

Questão 7

Em um posto de atendimento presencial a contribuintes, há três filas. As variáveis aleatórias X , Y e Z , que representam os números de atendimentos por minuto nessas filas, seguem distribuições de Poisson com médias, respectivamente, iguais a 1, 2 e 3.

Nessa situação hipotética, se as filas funcionam de forma independente e simultânea e se $T = X + Y + Z$ corresponde ao total de atendimentos por minuto no posto, então $P(T = 0)$ é igual a

- A $\exp(-1) + \exp(-2) + \exp(-3)$.
- B $\exp(-1) \times \exp(-2) \times \exp(-3)$.
- C $\exp(-1) + 2 \times \exp(-2) + 3 \times \exp(-3)$.
- D $\exp(-1) + \exp(-2) / 2 + \exp(-3) / 6$.
- E $\exp(-1) \times \exp(-2) \times \exp(-3) / 12$.

Questão 8

Se U e V forem variáveis aleatórias que sigam distribuições normais padrão, de tal forma que o desvio-padrão da soma $U + V$ seja igual a 1,5, então o valor esperado do produto $U \times V$ será igual a

- A -0,25.
- B -0,125.
- C 0,125.
- D 0,25.
- E 0,5.

Questão 9

Em uma entidade pública, em que o número X de processos instruídos para determinada finalidade segue uma distribuição condicional na forma

$$P(X = k|U = u) = \binom{5}{k} u^k (1 - u)^{5-k},$$

na qual $k \in \{0, 1, \dots, 5\}$ e U segue uma distribuição uniforme contínua no intervalo $[0, 1]$, 5 processos administrativos foram selecionados aleatoriamente de um sistema eletrônico.

Nessa situação hipotética, $P(X = 5)$ é igual a

- A 1/64.
- B 1/32.
- C 1/25.
- D 1/12.
- E 1/6.

Questão 10

Se $E[Y|X = x] = 2x^2$, em que $Y|X = x$ segue uma distribuição exponencial, então a probabilidade condicional $P(Y > 8|X = 2)$ será igual a

- A $0,125 \times \exp(1)$.
- B $1 - 0,125 \times \exp(-1)$.
- C $1 - \exp(-1)$.
- D $\exp(-1)$.
- E $0,125 \times \exp(-1)$.

Texto 5A1

Uma amostra aleatória simples X_1, X_2 e X_3 é retirada de uma população normal com média igual a 5 e desvio-padrão igual a 2, e definida nas seguintes transformações.

$$M = \frac{X_1 + X_2 + X_3}{3}$$

$$V^2 = \frac{(X_1 - 5)^2 + (X_2 - 5)^2 + (X_3 - 5)^2}{4}$$

$$U^2 = \frac{X_1^2 + X_2^2 + X_3^2}{3} - M^2$$

Questão 11

A partir das informações do texto 5A1, é correto afirmar que a variável aleatória V^2 segue uma distribuição

- A exponencial.
- B normal.
- C qui-quadrado.
- D Weibull.
- E t de Student.

Questão 12

Com base nas informações do texto 5A1, é correto afirmar que a razão $\frac{(M-5)\sqrt{2}}{U}$ tem variância

- A indefinida.
- B igual a 2.
- C zero.
- D negativa.
- E unitária.

Questão 13

No texto 5A1, o valor esperado de U^2 é igual a

- A 2.
- B 8/3.
- C 3.
- D 7/2.
- E 4.

Questão 14

Em um sistema de auditoria, cada processo administrativo é avaliado de forma independente para a detecção de indícios de fraude e a probabilidade de um processo apresentar fraude é de 0,19.

Nessa situação hipotética, definida a variável aleatória X como o número de processos analisados até que se encontre o primeiro caso de fraude, o desvio-padrão de X é igual a

- A 1/19.
- B 36/19.
- C 91/19.
- D 90/19.
- E 125/19.

Questão 15

Se $Z \sim N(0, 1)$, então $E(Z^4) + E(Z^3) + E(Z^2) + E(Z)$ é igual a

- A 4.
- B 3.
- C 2.
- D 1.
- E 0.

Questão 16

Considerando uma população com N elementos, assinale a opção em que é apresentado o estimador com o maior viés para a variância populacional, com base em uma amostra aleatória simples de tamanho $n > 4$.

- A $\hat{\sigma}^2 = \frac{n}{(n-1)(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$
- B $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2$
- C $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$
- D $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- E $\hat{\sigma}^2 = \frac{n-1}{n(n-3)} \sum_{i=1}^n (X_i - \bar{X})^2$

Texto 5A2-I

A fim de se estimar um parâmetro populacional Ω , foram construídas 5 fórmulas distintas, funções das observações amostrais. Para a escolha do melhor estimador, simulou-se uma situação em que $\Omega = 100$; da população retiraram-se 1.000 amostras, cada uma delas composta de dez unidades, e aplicaram-se todas as fórmulas às dez unidades de cada amostra, tendo sido obtidos 1.000 valores para cada fórmula, e, a partir deles, os dados a seguir.

fórmula	média: $E(w_i)$	variância: $\text{Var}(w_i)$
w_1	102	5
w_2	100	8
w_3	98	3
w_4	99	4
w_5	101	7

Questão 17

A partir dos dados apresentados na tabela que compõe o texto 5A2-I, assinale a opção que apresenta, empiricamente, em valor absoluto, o menor vício para a estimação de Ω .

- A w_5
- B w_1
- C w_3
- D w_4
- E w_2

Questão 18

A partir dos dados apresentados na tabela constante do texto 5A2-I, assinale a opção que corresponde, empiricamente, à estimação mais precisa de Ω .

- A w_5
- B w_3
- C w_4
- D w_1
- E w_2

Questão 19

A partir dos dados apresentados na tabela presente no texto 5A2-I, assinale a opção em que é apresentada, empiricamente, a estimação mais acurada de Ω .

- A w_3
- B w_4
- C w_5
- D w_1
- E w_2

Questão 20

Certo trompetista, com dificuldade em executar um trecho musical, decidiu anotar o número de tentativas realizadas até conseguir executar o solo sem erros. No primeiro dia, ele realizou 5 tentativas até acertar as notas, no segundo dia, 3 tentativas, no terceiro dia, 1 tentativa, no quarto dia, 4 tentativas e no quinto dia, 1 tentativa.

Com base nessas informações e considerando que cada tentativa corresponde a um ensaio de Bernoulli com parâmetro p desconhecido e que os experimentos realizados em dias diferentes são independentes entre si, assinale a opção na qual é apresentado o intervalo que contém a estimativa de máxima verossimilhança para a probabilidade p de o solo ser executado sem erros.

- A [32%; 40%[
- B [0; 8%[
- C [16%; 24%[
- D [24%; 32%[
- E [8%; 16%[

Questão 21

Considerando que os azeites de certa marca sejam vendidos em recipientes em cujo rótulo conste a indicação de 500 mL e que a avaliação de uma amostra de 100 recipientes indique a média $m = 485$ mL e a variância amostral $s^2 = 121$ mL², assinale a opção em que é apresentado o intervalo de confiança para a média populacional com coeficiente de confiança de 95% (considere que $z_{95\%} = 1,96$), com precisão de duas casas decimais, sendo s^2/n utilizado para calcular a variância populacional, em que n é o tamanho da amostra.

- A]483,90; 486,10[
- B]483,04; 486,96[
- C]483,79; 486,21[
- D]482,84; 487,16[
- E]482,63; 487,37[

Questão 22

Uma pesquisa realizada com 400 indivíduos condenados por sentença transitada em julgado, selecionados aleatoriamente, revelou que 70% deles são reincidentes, assim considerados, segundo o Código Penal brasileiro, os agentes que cometem novo crime depois de transitada em julgado a sentença que, no Brasil ou no estrangeiro, o tenha condenado por crime anterior.

Com base na situação hipotética precedente, assinale a opção em que é apresentado um intervalo de confiança conservador para a proporção populacional de reincidentes com coeficiente de confiança de 95% (considere que $z_{95\%} = 1,96$).

- A [0,681; 0,719]
- B [0,641; 0,759]
- C [0,651; 0,749]
- D [0,661; 0,739]
- E [0,671; 0,729]

Espaço livre

Texto 5A2-II

O índice de reincidência de apenados (proporção de reincidentes em face do total de agentes condenados por sentença transitada em julgado) processados pelo sistema tradicional de justiça é de 70%. Acredita-se que o índice de reincidência de apenados processados pela justiça restaurativa é menor que o índice dos processados pelo sistema tradicional de justiça. Para a verificação dessa conjectura, foram estabelecidas as seguintes hipóteses nula e alternativa.

H_0 : o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual ou inferior a 70%.

H_1 : o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é superior a 70%.

Para a testagem da hipótese H_0 , selecionou-se uma amostra aleatória simples de 210 apenados com sentença transitada em julgado processados pela justiça restaurativa, tendo-se observado que 71% são reincidentes.

Questão 23

Tendo como base as informações do texto 5A2-II, assinale a opção em que é descrita uma situação que corresponde a um erro indesejável na realização do teste de hipóteses proposto.

- A não rejeitar a afirmação de que o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual a 70% quando na verdade esse índice é superior
- B não rejeitar a afirmação de que o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual ou inferior a 70% quando é verdade que esse índice é igual ou inferior
- C não rejeitar a afirmação de que o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual a 70% quando na verdade esse índice é inferior
- D rejeitar a afirmação de que o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual ou inferior a 70% quando é verdade que esse índice é igual ou inferior
- E rejeitar a afirmação de que o índice de reincidência de apenados processados pela abordagem da justiça restaurativa é igual ou inferior a 70% quando na verdade esse índice é superior

Questão 24

Suponha que, para se testar a hipótese H_0 , mencionada no texto 5A2-II, seja escolhido como região crítica o conjunto $RC = \{y \in \mathbb{R} / y > p_0\}$ de tal modo que a hipótese H_0 será rejeitada se a probabilidade de o valor observado da estatística pertencer a RC for igual a 5%, dado que a hipótese H_0 é verdadeira. Nesse caso, considerando-se $z_{95\%} = 1,96$ e $0,001^{1/2} = 0,032$, o menor p_0 nessas condições será

- A superior ou igual a 0,77.
- B inferior a 0,72.
- C superior ou igual a 0,74 e inferior a 0,76.
- D superior ou igual a 0,76 e inferior a 0,77.
- E superior ou igual a 0,72 e inferior a 0,74.

Texto 5A2-III

O dono de um supermercado contratou uma equipe especializada para investigar se a altura em que um produto é exposto aos consumidores afeta seu volume de vendas. Para a investigação, foi escolhido um produto com ampla aceitação dos consumidores, oferecido a preços compatíveis com suas rendas e por ampla variedade de marcas, tanto nacionais quanto estrangeiras.

Questão 25

Na situação hipotética apresentada no texto 5A2-III, a população a ser estudada corresponde ao conjunto

- A dos números reais positivos não superiores ao preço do produto mais caro vendido no supermercado.
- B de todos os habitantes da cidade onde se situa o supermercado.
- C dos números reais positivos não superiores à maior altura possível para exposição de produtos nesse supermercado.
- D de todos os produtos disponibilizados nas prateleiras do supermercado.
- E de todos os clientes que frequentam o supermercado regularmente.

Questão 26

Assinale a opção em que é apresentada a variável com menor influência nos resultados da investigação mencionada no texto 5A2-III, considerada a necessidade de se evitarem interferências de outros fatores que não a característica de interesse.

- A variação do preço do produto cujo volume de vendas esteja sendo mensurado relativamente à variação, durante o período de realização da investigação, dos preços de bens substitutos
- B variação do preço do produto cujo volume de vendas esteja sendo mensurado relativamente à variação, durante o período da investigação, dos preços de bens complementares
- C investimento em *marketing* e propaganda realizado pelo dono da marca do produto imediatamente antes e durante a realização da investigação
- D facilidade de acesso dos clientes a produtos substitutos do produto cujo volume de vendas esteja sendo mensurado
- E variação no preço do barril de petróleo no mercado internacional

Questão 27

Em sua investigação, a equipe especializada mencionada no texto 5A2-III selecionou um produto considerado adequado, procedeu ao controle de outros fatores que poderiam influir nas medições ou em suas interpretações e realizou algumas poucas mensurações, conforme a tabela a seguir, apenas para obter uma primeira noção do melhor modelo a ser utilizado.

altura de exposição (cm)	vendas semana 1 (itens)	vendas semana 2 (itens)
30	530	520
80	780	790
140	1.040	1.050
190	660	680

A partir dessas informações, e adotando, para cada altura de exposição, a média do número de itens vendidos nas semanas 1 e 2, assinale a opção em que é apresentada, sob um aspecto qualitativo, a distribuição mais factível para modelar a situação por meio da representação do volume de vendas mensurado pelo número de itens vendidos do produto, no eixo vertical, e da altura de exposição, no eixo horizontal.

- A geométrica
- B uniforme
- C t de Student
- D de Bernoulli
- E exponencial

Questão 28

Os modelos de regressão linear tornaram-se uma maneira comprovada de prever o futuro de forma científica e confiável. Devido ao fato de a regressão linear ser um procedimento estatístico consolidado, as propriedades dos modelos de regressão linear são bem conhecidas.

Internet: <ibm.com>.

Em um modelo de regressão linear simples, o método de estimação por máxima verossimilhança (EMV)

- A** atribui a mesma verossimilhança para todo possível valor dos parâmetros.
- B** torna os dados observados mais prováveis sob o modelo estatístico assumido.
- C** sempre fornece os mesmos resultados do método dos mínimos quadrados, independentemente da distribuição dos dados.
- D** torna os valores preditos segundo o modelo os menores possíveis.
- E** minimiza o número de erros cometidos pelo modelo.

Questão 29

O método dos mínimos quadrados ordinários será utilizado em um modelo de regressão linear simples para se encontrar a reta que melhor se ajusta aos dados observados em determinada situação.

Nesse caso, o método dos mínimos quadrados ordinários permitirá

- A** encontrar uma reta na qual a soma algébrica simples de todos os desvios (positivos e negativos) seja igual a zero, independentemente da magnitude dos erros.
- B** maximizar a soma dos desvios entre os valores observados e os valores preditos pelo modelo, garantindo que a reta passe pelo maior número possível de pontos.
- C** minimizar a soma das diferenças absolutas entre os valores observados e os preditos, sem elevá-las ao quadrado, para evitar distorções causadas por valores extremos.
- D** maximizar a correlação entre as variáveis, ajustando uma reta que force todos os pontos a ficarem equidistantes da linha de regressão.
- E** minimizar a soma dos quadrados das diferenças verticais entre os valores observados e os valores preditos pela reta de regressão.

Questão 30

Um modelo de regressão linear múltipla com n observações e p variáveis explicativas (além do intercepto) é estimado por meio do método dos mínimos quadrados ordinários (com $p < n$). O coeficiente de determinação ajustado é substancialmente menor que o coeficiente de determinação não ajustado e a soma dos quadrados dos resíduos é numericamente próxima à soma de quadrados total. Adicionalmente, nenhum dos testes t individuais para os coeficientes apresentam significância estatística ao nível de 5%, embora o teste F global indique que o modelo como um todo é significativo ao mesmo nível.

Nessa situação hipotética,

- A** a aparente contradição resulta de heterocedasticidade severa nos resíduos, que afeta os testes t individuais, mas não o teste F global; a análise de resíduos studentizados revelaria padrões de variância crescente, e o uso de erros-padrão robustos de White corrigiria os testes t , tornando-os consistentes com o teste F .
- B** os resultados indicam violação grave do pressuposto de normalidade dos erros, pois, sob normalidade, os testes t e F devem sempre concordar em suas conclusões; a solução apropriada envolve a aplicação de transformações Box-Cox na variável resposta para restaurar a normalidade e reconciliar os testes.
- C** os resultados demonstram que o estimador de máxima verossimilhança seria mais apropriado que o estimador de mínimos quadrados para o conjunto de dados, pois, sob normalidade dos erros, ambos os métodos produzem estimativas idênticas, mas com propriedades de teste diferentes, sendo a verossimilhança mais robusta a estruturas de correlação complexas.
- D** a contradição entre os testes indica que os graus de liberdade foram calculados incorretamente, provavelmente devido à presença de observações duplicadas no conjunto de dados; a correção requer que se identifiquem e removam as duplicatas antes de se reestimar o modelo e recalcular todas as estatísticas de teste.
- E** os resultados sugerem presença de multicolinearidade substancial entre as variáveis explicativas; embora o conjunto de regressores tenha poder explicativo (teste F significativo), a alta correlação entre eles impede a identificação precisa das contribuições individuais (testes t não significativos), o que resulta em variâncias infladas dos estimadores individuais, mas não afeta a capacidade preditiva conjunta do modelo.

Questão 31

Em um modelo de regressão múltipla estimado por mínimos quadrados ordinários, observa-se que duas variáveis explicativas apresentam correlação muito alta entre si. Adicionalmente, um dos coeficientes estimados apresenta sinal contrário ao esperado pela teoria, e os erros-padrão dos coeficientes são substancialmente elevados.

A partir dessa situação hipotética, é correto concluir que o modelo em questão apresenta

- Ⓐ endogeneidade, o que viola o pressuposto de exogeneidade das variáveis explicativas e gera estimativas inconsistentes que não convergem para os valores verdadeiros mesmo com amostras grandes.
- Ⓑ autocorrelação dos resíduos, o que provoca perda da propriedade de eficiência dos estimadores e invalida completamente a capacidade preditiva do modelo ajustado.
- Ⓒ erro de especificação por omissão de variável relevante, sendo necessário identificar e incluir a variável faltante para corrigir os sinais inesperados dos coeficientes.
- Ⓓ multicolinearidade, o que resulta em estimativas com variância inflada e interpretações individuais dos coeficientes pouco confiáveis.
- Ⓔ heterocedasticidade, o que implica estimativas viesadas dos coeficientes e torna os testes de hipótese não confiáveis para inferências sobre os parâmetros populacionais.

Questão 32

Em um modelo de regressão linear estimado por mínimos quadrados ordinários, a análise gráfica dos resíduos revela que a dispersão dos erros aumenta sistematicamente conforme os valores ajustados crescem. Adicionalmente, observa-se que os testes de hipótese individuais indicam significância estatística para diversos coeficientes, mas o pesquisador suspeita que os intervalos de confiança e os p -valores podem estar incorretos.

Nessa situação hipotética, o modelo apresenta

- Ⓐ endogeneidade resultante da correlação entre variáveis explicativas e o termo de erro, o que torna os estimadores viesados e exige o uso de métodos de variáveis instrumentais para a obtenção de estimativas consistentes.
- Ⓑ multicolinearidade severa entre os regressores, o que provoca instabilidade nas estimativas e inflação artificial da dispersão dos resíduos conforme os valores ajustados aumentam.
- Ⓒ não linearidade na forma funcional, sendo necessário aplicar transformações logarítmicas ou polinomiais para a correção do viés sistemático introduzido pela especificação linear incorreta.
- Ⓓ autocorrelação dos resíduos, o que viola o pressuposto de independência temporal, resultando em estimadores viesados e inconsistentes, que não convergem para os valores verdadeiros populacionais.
- Ⓔ heterocedasticidade, o que viola o pressuposto de variância constante dos erros, resultando em erros-padrão incorretos e inferências estatísticas não confiáveis.

Questão 33

Ao estimar um modelo de regressão linear simples para explicar o salário em função dos anos de experiência profissional, um pesquisador obteve um coeficiente angular (inclinação) estimado de 2.500 reais por ano de experiência, com um p -valor de 0,001.

Com base nessa situação hipotética e considerando um nível de significância de 5%, assinale a opção correta acerca da relação entre experiência e salário.

- Ⓐ Existe evidência estatística significativa de que há uma relação entre anos de experiência e salário na população, pois o p -valor é menor que o nível de significância estabelecido.
- Ⓑ Não é possível concluir nada sobre a relação entre experiência e salário, pois o p -valor é muito pequeno e indica que os dados são insuficientes para qualquer inferência.
- Ⓒ O p -valor de 0,001 significa que o modelo está 99,9% incorreto, portanto deve-se rejeitar a hipótese de que experiência afeta salário.
- Ⓓ Como o p -valor é diferente de zero, não há relação entre experiência e salário, sendo o coeficiente estimado resultado apenas do acaso.
- Ⓔ O p -valor indica que existe 0,1% de chance de que a experiência realmente influencie o salário, o que é uma probabilidade muito baixa para se considerar a relação como verdadeira.

Questão 34

Um analista tributário analisou a relação entre a cotação do dólar e a arrecadação mensal de ICMS (em milhões de reais). Após estimar o modelo de regressão, ele construiu um intervalo de confiança de 95% para o coeficiente angular (inclinação da reta), tendo obtido o seguinte resultado: [0,8 ; 3,2].

Tendo por base a situação hipotética apresentada, assinale a opção correta acerca da relação entre a cotação do dólar e a arrecadação mensal do ICMS.

- Ⓐ O intervalo indica que, para cada real adicional na cotação do dólar, a arrecadação do ICMS aumenta exatamente em 2 milhões de reais, que é a média entre 0,8 e 3,2, com margem de erro de 95%.
- Ⓑ O intervalo mostra que há 95% de probabilidade de que o verdadeiro valor do coeficiente esteja entre 0,8 e 3,2 e, portanto, há 5% de chance de que o crescimento da cotação do dólar resulte na diminuição da arrecadação do ICMS.
- Ⓒ Como o intervalo não contém o valor 1, não há relação estatisticamente significativa entre a cotação do dólar e a arrecadação do ICMS, devendo o modelo estimado ser rejeitado.
- Ⓓ Pode-se afirmar com 95% de confiança que existe uma relação positiva entre a cotação do dólar e a arrecadação de ICMS, pois todos os valores do intervalo são positivos e não incluem o zero.
- Ⓔ O intervalo indica que não se pode ter certeza sobre a direção do efeito da cotação do dólar na arrecadação de ICMS, pois o intervalo é muito amplo e contém valores tanto positivos quanto negativos.

Questão 35

Utilizando um modelo de regressão linear simples para relacionar o peso de veículos com seu consumo de combustível, um analista realizou a análise de variância (ANOVA) e obteve um valor F bastante elevado, com p -valor muito próximo de zero.

Nessa situação hipotética,

- Ⓐ o resultado indica que exatamente 50% da variação no consumo pode ser explicada pelo peso do veículo, pois a estatística F mede a proporção de variância explicada em escala percentual.
- Ⓑ o valor F elevado indica a existência de muita variabilidade nos dados, que não pode ser explicada pelo modelo, o que sugere que o peso não é uma boa variável explicativa para o consumo.
- Ⓒ um valor F muito alto é sempre problemático e indica que o modelo está superajustado aos dados, devendo ser descartado e substituído por um modelo mais simples.
- Ⓓ o valor F elevado com p -valor próximo de zero indica que o modelo de regressão como um todo é estatisticamente significativo, ou seja, a variável peso ajuda a explicar a variação no consumo de combustível.
- Ⓔ o p -valor próximo de zero significa que praticamente não há relação entre peso e consumo, pois a probabilidade de associação entre as variáveis é quase nula.

Espaço livre

Questão 36

Na aplicação de um modelo de regressão linear múltipla com intercepto, um pesquisador observou que a soma dos quadrados da regressão representava 75% da soma dos quadrados total, o que indicava coeficiente de determinação de 0,75. Ao testar a significância global do modelo por meio do teste F , ele obteve um p -valor de 0,08 ao nível de 5%. O pesquisador, então, decidiu remover uma variável explicativa que apresentou o maior p -valor no teste t individual e reestimar o modelo. No novo modelo, a soma dos quadrados da regressão diminuiu, mas surpreendentemente o teste F global indicou significância ao nível de 5%.

A partir dessa situação hipotética, assinale a opção correta no que diz respeito à explicação para o fenômeno aparentemente paradoxal verificado.

- Ⓐ O resultado demonstra que a variável removida estava causando multicolinearidade severa que inflava artificialmente a soma de quadrados da regressão, mas reduzia a estatística F , e sua remoção restaurou a ortogonalidade entre os regressores, o que permitiu que o teste F detectasse corretamente a significância do modelo.
- Ⓑ A aparente melhora na significância resulta de violação do pressuposto de normalidade dos erros no modelo original, que foi parcialmente corrigida pela remoção da variável; a variável excluída estava introduzindo assimetria nos resíduos, e sua ausência aproximou a distribuição dos erros à normalidade, o que validou o teste F .
- Ⓒ O fenômeno é possível e resulta da mudança na estrutura dos graus de liberdade: embora a soma de quadrados da regressão tenha diminuído com a remoção da variável, o quadrado médio da regressão pode ter aumentado e, simultaneamente, o quadrado médio dos resíduos pode ter diminuído suficientemente, o que alterou favoravelmente a razão F , considerando-se especialmente a redução de um grau de liberdade nos regressores.
- Ⓓ O resultado é impossível e indica erro computacional, pois a remoção de qualquer variável explicativa necessariamente reduz tanto a soma de quadrados da regressão quanto a estatística F , dado que o modelo mais complexo sempre se ajusta de forma melhor ou igual aos dados que o modelo mais simples.
- Ⓔ O fenômeno ocorre porque o R -quadrado ajustado, que penaliza a inclusão de variáveis irrelevantes, aumentou no modelo reduzido, e, como o teste F é baseado no R -quadrado ajustado em vez de no R -quadrado usual, a estatística F necessariamente aumenta quando o R -quadrado ajustado aumenta.

Questão 37

Um pesquisador estima dois modelos de regressão aninhados: o modelo completo contém p variáveis explicativas além do intercepto, e o modelo reduzido, q variáveis explicativas além do intercepto, em que $q < p$. Ambos os modelos são estimados com os mesmos n dados. Para testar se as $(p - q)$ variáveis adicionais do modelo completo contribuem significativamente para explicar a variação na variável resposta, o pesquisador calcula a diferença entre as somas dos quadrados dos resíduos dos dois modelos e constrói uma estatística F parcial. O teste resulta em um p -valor de 0,12 ao nível de 5%, o que implica a não rejeição da hipótese nula. Simultaneamente, ao examinar o modelo completo, o pesquisador observa que o teste F global é altamente significativo (p -valor $< 0,001$).

Nessa situação hipotética,

- Ⓐ a discrepância entre os testes indica violação do pressuposto de homogeneidade de variâncias, em que o modelo completo apresenta heterocedasticidade que não estava presente no modelo reduzido; o teste F parcial é mais sensível a essa violação que o teste F global, o que explica por que um é não significativo enquanto o outro permanece altamente significativo.
- Ⓑ os resultados são consistentes e indicam que, embora o modelo completo como um todo tenha poder explicativo significativo comparado ao modelo apenas com intercepto, as $(p - q)$ variáveis adicionais não contribuem significativamente além da explicação já fornecida pelas q variáveis do modelo reduzido, o que sugere que o modelo mais parcimonioso (reduzido) é adequado e as variáveis extras são redundantes, dado o que já está no modelo.
- Ⓒ os resultados são contraditórios e matematicamente impossíveis, pois, se o modelo completo é globalmente significativo, então qualquer subconjunto de variáveis também deve ser significativo; o erro provavelmente resulta de uso incorreto das somas dos quadrados, devendo o pesquisador utilizar a soma dos quadrados da regressão em vez da soma dos quadrados dos resíduos.
- Ⓓ a aparente inconsistência resulta de multicolinearidade entre as variáveis do modelo completo e as do modelo reduzido, que infla os erros-padrão no teste F parcial, mas não afeta o teste F global, o que resulta em conclusões divergentes que invalidam ambos os testes até que a multicolinearidade seja resolvida.
- Ⓔ os resultados demonstram que as $(p - q)$ variáveis adicionais estão causando sobreajuste no modelo completo, capturando ruído amostral em vez de padrões verdadeiros. O teste F parcial não significativo indica corretamente que essas variáveis deterioram a qualidade preditiva, ao passo que o teste F global permanece significativo devido às q variáveis genuinamente importantes.

Questão 38

Na análise diagnóstica dos resíduos de um modelo de regressão múltipla estimado por mínimos quadrados ordinários, identificam-se três observações com características distintas: a i -ésima observação apresenta resíduo studentizado externamente superior a 3 em valor absoluto, mas baixo valor de alavancagem; a j -ésima observação apresenta alta alavancagem, mas resíduo próximo de zero; e a k -ésima observação apresenta, simultaneamente, resíduo studentizado moderado (aproximadamente 2) e alta alavancagem, o que resulta em distância de Cook substancialmente elevada.

Considerando a situação hipotética apresentada e os conceitos de *leverage*, *outliers* e observações influentes, assinale a opção correta.

- Ⓐ A presença simultânea dessas três observações com características distintas invalida o uso de mínimos quadrados ordinários, sendo necessário utilizar métodos robustos de estimação, como regressão quantílica ou M -estimadores, para a obtenção de resultados confiáveis.
- Ⓑ As três observações são igualmente problemáticas e devem ser removidas do conjunto de dados, pois qualquer observação com resíduo elevado ou alta alavancagem viola os pressupostos do modelo de regressão e compromete a validade das inferências estatísticas.
- Ⓒ A i -ésima observação indica heterocedasticidade localizada e deve ser corrigida por meio de ponderação; a j -ésima observação sugere multicolinearidade que envolve essa observação específica; e a k -ésima observação demonstra não linearidade, que requer inclusão de termos polinomiais no modelo.
- Ⓓ Como a k -ésima observação apresenta a maior distância de Cook, ela é necessariamente a única observação que deve ser investigada e potencialmente removida, ao passo que as observações i e j podem ser ignoradas na análise, independentemente de seus valores de resíduo ou alavancagem.
- Ⓔ A i -ésima observação é um *outlier* sem influência substancial sobre as estimativas; a j -ésima observação tem potencial de influência devido ao *leverage*, mas está bem ajustada pelo modelo; e a k -ésima observação é genuinamente influente, pois combina posição extrema no espaço dos regressores com inadequação do ajuste, sendo a mais preocupante para a estabilidade das estimativas dos coeficientes.

Questão 39

Em um modelo de regressão linear simples estimado por mínimos quadrados ordinários, um analista observa que a análise de variância decompõe apropriadamente a soma de quadrados total em soma de quadrados da regressão e soma de quadrados dos resíduos. Ao examinar o gráfico de resíduos *versus* valores ajustados, o analista identifica um padrão claro em formato de parábola (U invertido). Adicionalmente, ao calcular a correlação entre os resíduos ordinários e os valores ajustados, ele obtém valor próximo de zero. O analista também nota que, ao estimar o modelo por meio de máxima verossimilhança sob pressuposto de normalidade, os estimadores dos coeficientes de regressão são numericamente idênticos aos obtidos por mínimos quadrados, mas o estimador de máxima verossimilhança da variância do erro é ligeiramente menor que o estimador usual não viesado.

Nessa situação hipotética,

- Ⓐ a identidade numérica entre estimadores MQO e MV dos coeficientes implica que a decomposição da variância pela análise de variância é ótima e, portanto, o padrão parabólico nos resíduos não representa problema real, mas flutuação amostral aleatória, que desapareceria com o aumento do tamanho amostral, o que valida o modelo linear simples estimado.
- Ⓑ o padrão parabólico indica inadequação da forma funcional linear, o que sugere não linearidade na relação verdadeira entre as variáveis, ao passo que a correlação zero entre resíduos ordinários e valores ajustados é uma propriedade algébrica que sempre vale em regressão por mínimos quadrados, não servindo como diagnóstico de adequação do modelo; a diferença entre estimadores MV e MQO da variância reflete que o estimador MV usa divisor n enquanto o não viesado usa $n - 2$.
- Ⓒ o padrão parabólico demonstra violação do pressuposto de homogeneidade de variâncias, e a solução adequada envolve a utilização de estimadores de mínimos quadrados ponderados, em que os pesos são inversamente proporcionais à variância estimada em cada ponto, o que corrige simultaneamente o padrão residual e aumenta a eficiência dos estimadores.
- Ⓓ o padrão parabólico nos resíduos combinado com correlação zero entre resíduos e valores ajustados indica que o modelo está corretamente especificado e todos os pressupostos são atendidos, pois a propriedade de ortogonalidade entre resíduos e predições é garantida algebricamente pelo método de mínimos quadrados, independentemente da forma funcional.
- Ⓔ a correlação zero entre resíduos e valores ajustados prova que não há heterocedasticidade no modelo, e o padrão parabólico observado é artefato visual causado por observações influentes que devem ser identificadas por meio da distância de Cook e removidas para revelar a verdadeira estrutura linear dos dados.

Questão 40

Certo pesquisador deseja avaliar o comportamento dos motoristas em um estacionamento. Ele decide entrevistar o motorista de cada décimo carro que entra no local, até obter uma amostra de 50 motoristas.

Na situação hipotética precedente, o pesquisador utilizou a técnica de amostragem

- Ⓐ por conveniência.
- Ⓑ por conglomerados.
- Ⓒ sistemática.
- Ⓓ estratificada.
- Ⓔ aleatória simples.

Questão 41

Uma empresa possui 800 funcionárias e 200 funcionários. Para garantir que a amostra mantenha o equilíbrio de gênero presente na empresa, a população é dividida em dois grupos conforme o sexo. Em seguida, realiza-se amostragem aleatória em cada grupo, de modo a selecionar 40 mulheres e 10 homens, o que resulta em uma amostra representativa de 50 pessoas.

Na situação hipotética apresentada, utilizou-se a técnica de amostragem

- Ⓐ estratificada.
- Ⓑ aleatória simples.
- Ⓒ por conveniência.
- Ⓓ por conglomerados.
- Ⓔ sistemática.

Questão 42

Amostragem é um processo estatístico que consiste na seleção criteriosa de um subconjunto representativo (a amostra) a partir de um conjunto total de dados ou indivíduos (a população). Para que uma amostragem seja classificada tecnicamente como aleatória simples (AAS),

- Ⓐ a amostra deve conter, obrigatoriamente, pelo menos 10% da população total para garantir precisão estatística.
- Ⓑ a população deve ser dividida em subgrupos homogêneos antes do sorteio.
- Ⓒ a seleção dos elementos deve ocorrer em intervalos fixos a partir de uma lista ordenada.
- Ⓓ o pesquisador deve selecionar os elementos que julgar mais representativos da média populacional.
- Ⓔ qualquer subconjunto de tamanho n da população deve ter a mesma probabilidade de ser selecionado que qualquer outro subconjunto de mesmo tamanho.

Questão 43

De uma população composta por $N = 500$ elementos deseja-se obter uma amostra aleatória simples, sem reposição, de tamanho $n = 50$.

A partir dessas informações, assinale a opção que corresponde à probabilidade de um elemento específico da população ser selecionado na extração da amostra.

- Ⓐ 0,50
- Ⓑ 0,002
- Ⓒ 0,01
- Ⓓ 0,10
- Ⓔ 0,25

Questão 44

Em uma pesquisa de clima organizacional, o setor de recursos humanos de certa empresa utilizou um *software* de sorteio aleatório para selecionar 200 funcionários entre os 5.000 existentes.

A partir da situação hipotética precedente, assinale a opção que corresponde à técnica de amostragem que foi usada no caso em tela.

- Ⓐ amostragem por conveniência
- Ⓑ amostragem por conglomerados
- Ⓒ amostragem estratificada
- Ⓓ amostragem aleatória simples
- Ⓔ amostragem sistemática

Questão 45

O objetivo central do processo de amostragem é permitir que características, comportamentos ou opiniões observados em um grupo menor possam ser extrapolados e generalizados para todo o universo estudado, economizando-se tempo, recursos financeiros e esforço logístico.

Assinale a opção que apresenta a principal diferença operacional entre a amostragem estratificada e a amostragem por conglomerados.

- Ⓐ A amostragem estratificada exige menos informações prévias acerca da população que a amostragem por conglomerados.
- Ⓑ A amostragem estratificada é caracterizada pela mistura dos grupos antes da realização do sorteio, enquanto a amostragem por conglomerados é caracterizada por manter os grupos separados antes da realização do sorteio.
- Ⓒ A amostragem estratificada tem como objetivo reduzir o custo e a logística com o sorteio dos grupos, enquanto a amostragem por conglomerados tem como objetivo garantir a representação de todos os grupos.
- Ⓓ A amostragem estratificada é caracterizada pela realização do sorteio dentro de todos os grupos, enquanto a amostragem por conglomerados é caracterizada por sortear alguns grupos inteiros, sendo os outros ignorados.
- Ⓔ A amostragem estratificada é não probabilística, enquanto a amostragem por conglomerados é probabilística.

Questão 46

A secretaria de mobilidade urbana de determinado estado pretende realizar uma pesquisa de satisfação sobre o serviço de transporte público prestado no estado, sendo consultadas as populações urbana e rural que habitam seu território. No planejamento da pesquisa, foi sugerida a escolha do método de amostragem estratificada, em vez da amostragem aleatória simples.

A amostragem estratificada é particularmente recomendada, em lugar da amostragem aleatória simples, quando

- Ⓐ há um prazo muito curto para a pesquisa.
- Ⓑ se deseja comparar resultados entre subgrupos específicos (p. ex. zona rural *versus* zona urbana) com precisão controlada.
- Ⓒ há uma população muito homogênea, sem grupos distintos.
- Ⓓ não há nenhuma lista ou cadastro dos membros da população.
- Ⓔ o objetivo principal da realização da amostragem é reduzir o custo logístico de deslocamento do pesquisador.

Texto 5A4

Para a realização de uma fiscalização, a secretaria de fazenda de determinado estado executou um plano amostral da seguinte forma: primeiro, foram escolhidos, aleatoriamente, sem reposição, três municípios entre os 70 do estado; em seguida, para cada município escolhido, selecionou-se aleatoriamente um setor entre comércio, serviços e indústria; por fim, dentro do conjunto de empresas sediadas no município e atuantes no setor selecionado, escolheu-se uma ao acaso.

Questão 47

A partir da situação hipotética apresentada no texto 5A4, considerando a probabilidade de as empresas do município selecionado serem fiscalizadas, assinale a opção correta.

- Ⓐ Após a execução da segunda etapa, todas as empresas do município selecionado na primeira etapa têm igual probabilidade de serem fiscalizadas.
- Ⓑ Antes de se iniciar a execução do plano amostral, todas as empresas do estado têm igual probabilidade de serem fiscalizadas.
- Ⓒ Após a execução da primeira etapa e antes da execução da segunda etapa, há empresas sediadas nos municípios selecionados que têm probabilidade nula de virem a ser fiscalizadas.
- Ⓓ Após a execução da segunda etapa, há empresas do município selecionado na primeira etapa que têm probabilidade nula de serem fiscalizadas.
- Ⓔ Antes de se iniciar a execução do plano amostral, há empresas no estado que têm probabilidade nula de virem a ser fiscalizadas.

Questão 48

Considere que certa empresa seja sediada em um dos municípios do estado referido no texto 5A4, e que tal município sedie 250 empresas, sendo 110 do setor de serviços. Nesse caso, a probabilidade de essa empresa vir a ser fiscalizada pela secretaria de fazenda do estado será de

- Ⓐ 1/7.700.
- Ⓑ 1/17.500.
- Ⓒ 1/27.500.
- Ⓓ 1/70.
- Ⓔ 1/110.

Questão 49

Para conduzir uma pesquisa de satisfação em uma empresa com 2.000 colaboradores, certo pesquisador estabeleceu um erro amostral de 5%, usual em estudos sociais. O tamanho da amostra n foi calculado por meio da fórmula de Yamane a seguir, em que N representa o tamanho da população e E , o erro amostral tolerado.

$$n = \frac{N}{1 + NE^2}$$

A partir das informações presentes na situação hipotética precedente, assinale a opção que apresenta o intervalo que corresponde ao tamanho da amostra necessário para garantir que os resultados representem adequadamente a população que participou da pesquisa.

- Ⓐ $500 \leq n$
- Ⓑ $100 \leq n < 200$
- Ⓒ $300 \leq n < 400$
- Ⓓ $400 \leq n < 500$
- Ⓔ $200 \leq n < 300$

Questão 50

Em uma cidade com 20.000 habitantes, a população se divide em 6.000 jovens, 10.000 adultos e 4.000 idosos. Certo pesquisador deseja selecionar uma amostra de 400 pessoas por amostragem estratificada proporcional.

Com relação à situação hipotética precedente, assinale a opção que corresponde à quantidade de pessoas que devem ser selecionadas em cada estrato.

- A** 100 jovens, 200 adultos, 100 idosos
- B** 200 jovens, 120 adultos, 80 idosos
- C** 150 jovens, 150 adultos, 100 idosos
- D** 80 jovens, 240 adultos, 80 idosos
- E** 120 jovens, 200 adultos, 80 idosos

Espaço livre
