

-- CONHECIMENTOS ESPECÍFICOS --

Considerando os conceitos de estatística descritiva e análise exploratória de dados, especialmente no que concerne ao uso de gráficos, diagramas e tabelas, julgue os itens a seguir.

- 51 Gráficos em formato *pizza* devem ser utilizados apenas quando houver poucas categorias, pois um número elevado de setores dificulta a comparação visual das proporções.
- 52 O uso de gráficos, diagramas e tabelas, na análise exploratória de dados, substitui eficientemente métodos estatísticos inferenciais em processos de inferência populacional.
- 53 Um *boxplot* apresenta a mediana, os quartis e possíveis valores atípicos.

Em relação a medidas descritivas de posição, julgue os itens que se seguem.

- 54 Um conjunto de dados pode apresentar mais de uma moda.
- 55 Dados categóricos não admitem média aritmética, salvo quando transformados por codificações específicas que permitem interpretações limitadas, como proporções.
- 56 Para variáveis aleatórias em que o *skewness* seja diferente de zero, a mediana é sempre igual à média.
- 57 A mediana sofre grande influência de valores extremos.

Julgue os próximos itens, relativos a medidas descritivas de dispersão.

- 58 O terceiro quartil (Q3) representa o ponto abaixo do qual estão 75% dos dados.
- 59 A amplitude envolve todos os elementos da amostra no cálculo.
- 60 Considere que uma amostra com 50 elementos tenha sido extraída de um sistema de medição de tensão (V) do quadro elétrico de certa infraestrutura de telecomunicações. Nesse caso, a unidade da média e do desvio padrão será V.
- 61 A multiplicação de todos os valores de um conjunto de dados X por uma constante $k \neq 0$ implica que a média e a variância também sejam multiplicadas por k , ou seja, $E[kX] = kE[X]$ e $Var[kX] = kVar[X]$.

Julgue os itens seguintes, relacionados a medidas descritivas de assimetria.

- 62 Assimetria elevada frequentemente indica a presença de valores extremos, que impactam a forma da distribuição.
- 63 Assimetria igual a zero não implica, necessariamente, uma distribuição totalmente simétrica.
- 64 Distribuições com assimetria à esquerda geralmente apresentam média menor que a mediana.
- 65 Assimetria negativa indica que a distribuição possui maior concentração de valores à esquerda e uma cauda mais longa que se estende para a direita.

Com base no referencial teórico das medidas descritivas de curtose, julgue os itens subsequentes.

- 66 Em distribuições leptocúrticas, os valores se concentram mais próximos da média.
- 67 Curtose alta indica apenas um pico mais elevado, não estando relacionada às caudas.

Em estudo estatístico conduzido para apoiar decisões técnicas em telecomunicações, analisou-se o tempo (em minutos) até a ocorrência de uma falha em determinado equipamento de rede, tendo sido definidos os seguintes eventos aleatórios:

- A = "o equipamento falha antes de 30 minutos".
 B = "o equipamento falha antes de 60 minutos".
 C = "o equipamento falha entre 30 e 60 minutos".
 D = "o equipamento falha após 60 minutos".

Sabendo que, nessa situação hipotética, o tempo de falha segue uma variável aleatória contínua e que as probabilidades referentes a esses eventos são positivas, julgue os itens seguintes.

- 68 A e C são eventos independentes.
- 69 O evento A é o complementar do evento C .
- 70 $P(A) < P(B)$.
- 71 $P(C|B) = 1$.

Supondo que X , Y e Z sejam variáveis aleatórias independentes que seguem distribuições de Poisson, respectivamente, com médias 1, 2 e 3, julgue os próximos itens, em relação à soma $S = X + Y + Z$.

- 72 $S - X$ segue uma distribuição de Poisson.
- 73 O desvio padrão de S é igual a 6.
- 74 $\operatorname{argmax}_{x \geq 0} P(S = x) = \{5, 6\}$.
- 75 $P(S = 0) = P(X = 0) \cdot P(Y = 0) \cdot P(Z = 0)$.

Espaço livre

Considerando que a função de densidade de probabilidade de uma variável aleatória contínua X seja dada por

$$f(x) = \begin{cases} C \cdot x^6, & \text{se } |x| \leq 1, \\ 0, & \text{se } |x| > 1, \end{cases}$$

em que C é a constante normalizadora, julgue os itens a seguir.

76 $P(|X| \leq 0,5) = \frac{C}{448}$.

77 X possui assimetria nula.

78 A função de densidade de $Y = X^2$ é

$$f_Y(y) = \begin{cases} C \cdot y^3, & \text{se } y \in [0,1], \\ 0, & \text{se } y \notin [0,1]. \end{cases}$$

79 $C = \frac{2}{7}$.

80 A variância de X é igual a $\frac{2C}{9}$.

Considerando uma amostra aleatória simples $X_1, \dots, X_n$ retirada de uma distribuição exponencial com média igual a 2, julgue os seguintes itens, a respeito da soma $S_n = X_1 + \dots + X_n$.

81 Assintoticamente, S_n segue uma distribuição normal.

82 S_n segue uma distribuição quiquadrado com $2n$ graus de liberdade.

83 O teorema central do limite estabelece que $\frac{S_n - 2}{\sqrt{n}}$ converge em distribuição para a distribuição normal padrão, à medida que n tende para o infinito.

84 De acordo com a lei fraca dos grandes números, à medida que o tamanho da amostra aumenta, S_n converge em probabilidade para uma distribuição normal.

85 A variância de S_n é igual a $2n$.

Considerando que $X_1, X_2, \dots, X_n$ sejam cópias independentes e identicamente distribuídas de uma variável aleatória X com função de densidade de probabilidade $f(x, \phi)$, julgue os próximos itens.

86 O estimador de máxima verossimilhança para o parâmetro ϕ é a quantidade $\hat{\phi}$ que faz que os dados observados sejam mais prováveis (ou plausíveis) de ocorrerem.

87 Por definição, um estimador para o parâmetro ϕ é qualquer função da amostra observada que assume valores no espaço paramétrico e que não depende do parâmetro que está sendo estimado.

88 O estimador dos momentos para o parâmetro ϕ é a quantidade que minimiza a soma dos quadrados dos erros.

X_1, X_2 e X_3 denotam cópias independentes e identicamente distribuídas de uma variável aleatória X , com média $\mu \in (-\infty, +\infty)$ e variância $\sigma^2 > 0$. A seguir, apresentam-se alguns estimadores para o parâmetro μ .

$$\mu: \begin{cases} T_1 = \frac{2X_1 + X_2}{4} \\ T_2 = \frac{X_1 + X_2 + 3X_3}{5} \\ T_3 = \frac{X_1 + X_2}{2} \end{cases}$$

Com base nas informações precedentes, julgue os itens seguintes.

89 Apenas os estimadores T_2 e T_3 são consistentes para μ .

90 T_1 e T_2 são estimadores não viesados (ou centrados) para μ .

91 Entre os três estimadores apresentados, T_2 é o mais eficiente.

Julgue os itens a seguir, referentes a estimação pontual.

92 Os estimadores obtidos pelo método dos mínimos quadrados são aqueles que maximizam a soma dos quadrados dos erros.

93 Considerando-se que X_1, X_2 e X_3 denotem cópias independentes e identicamente distribuídas de uma variável aleatória X com distribuição de Poisson (θ), cuja função de probabilidade é dada por $P_\theta(X=x) = \left(\frac{\theta^x}{x!}\right) e^{-\theta}$, em que $x = 0, 1, 2, \dots$, é correto afirmar que $T = X_1 + X_2$ é uma estatística suficiente para θ .

94 Considere X uma variável aleatória proveniente de uma distribuição caracterizada pela função de probabilidade a seguir, em que $x = 0, 1, 2, \dots$ e $\beta > 1$.

$$f(x, \beta) = \frac{1}{\beta} \cdot \left(1 - \frac{1}{\beta}\right)^x$$

Nesse caso, se o conjunto 9, 8, 7, 5, 6, 4 denotar uma amostra observada de X , então, a estimativa de máxima verossimilhança para β será $\hat{\beta} = 7,50$.

95 Por definição, uma estatística é dita suficiente quando a distribuição condicional da amostra, dada a estatística, não depende do parâmetro de interesse.

Espaço livre

Para avaliação da audiência do telejornal de certa emissora de televisão em determinado mês de 2025, mediram-se os pontos de audiência do programa em 25 dias desse mês, tendo sido obtido um valor médio de 34,7 pontos. Pesquisas anteriores apontaram um desvio padrão populacional de 3,50 pontos de audiência desse telejornal.

A partir dessa situação hipotética, e supondo que a audiência do telejornal siga uma distribuição normal, julgue os itens subsecutivos, assumindo um nível de confiança de 94% ($z_{1-\alpha/2} = 1,88$).

- 96 A amplitude do intervalo de confiança aumentaria caso o tamanho da amostra (número de dias de medição) fosse aumentado para 28, mantendo-se constantes as demais quantidades (nível de confiança e desvio-padrão).
- 97 Infere-se das informações apresentadas que o intervalo de confiança para a média dos pontos de audiência em todo o mês é dado por $IC_{0,94}(\mu) = [33,38; 36,02]$.

No que se refere a estimação intervalar, julgue os próximos itens.

- 98 Considere a situação em que uma especialista em gestão de telecomunicações esteja interessada em obter estimativas intervalares para o parâmetro θ , que representa o consumo médio de dados de Internet em determinada região do Brasil no primeiro semestre de 2025. Supondo que a especialista obteve duas estimativas intervalares, sendo: o intervalo de confiança (caso frequentista), satisfazendo $P(\theta \in I_{\text{conf}}(x)) = 0,90$, e o intervalo de credibilidade (caso bayesiano), tal que $P(\theta \in I_{\text{cred}} | X=x) = 0,90$.

É correto afirmar que, com base no intervalo de confiança, existe uma probabilidade de 90% do parâmetro θ pertencer ao intervalo obtido.

- 99 Considere que certo programa do governo federal forneça Internet via satélite para áreas rurais, comunidades isoladas, escolas, postos de saúde e centros comunitários onde o acesso à Internet é limitado ou inexistente. Considere, ainda, que, em determinado estado brasileiro, testes tenham mostrado que 385 dos 500 pontos avaliados apresentaram cobertura adequada de sinal. Com base nessa situação hipotética, considerando-se o nível de confiança de 95% com $z_{(1-\alpha/2)} = 1,96$, é correto afirmar que a amplitude do intervalo de confiança obtida por meio da abordagem otimista será sempre maior que a obtida por meio da abordagem conservadora.

A gerente de *marketing* de uma operadora de telefonia móvel afirmou, em uma entrevista, que o consumo médio de dados por usuário em Brasília era de 18,0 GB/mês. Após a implantação de uma nova tecnologia de Internet para as redes móveis, uma amostra de 36 usuários apresentou um consumo médio de 14,8 GB/mês e variância de 9,0 (GB/mês)².

Com base nas informações precedentes, considerando $H_0: \mu = 18,0$ GB/mês contra $H_1: \mu < 18,0$ GB/mês e $t_{(35; 0,97)} = 1,944$, julgue os próximos itens.

- 100 Para o caso em tela, o poder do teste estatístico de hipóteses é a probabilidade de não rejeitar uma hipótese nula, sendo ela falsa.
- 101 Ao nível de 3% de significância, há evidências suficientes para rejeitar H_0 .

A fim de testar se, nas cobranças de pênaltis no futebol, o pulo dos goleiros independe das direções da bola, foram coletados os dados mostrados na seguinte tabela, que resume os resultados de 290 cobranças de pênaltis registradas durante as últimas 5 temporadas de certo campeonato de futebol. Nessa tabela, as frequências esperadas estão entre parênteses.

direção da bola	pulo do goleiro		
	esquerda	centro	direita
esquerda	54 (46,7)	5 (7,3)	37 (42,0)
centro	41 (39,9)	10 (6,2)	31 (35,9)
direita	46 (54,5)	7 (8,5)	59 (49,0)

As hipóteses de interesse foram as seguintes.

- H_0 : A direção dos pulos do goleiro é independente da direção da bola.
 H_1 : A direção dos pulos do goleiro é dependente da direção da bola.

A partir da situação hipotética precedente, considerando o nível de significância de 3% e $\chi^2_{(4; 0,97)} = 10,7$, julgue o seguinte item.

- 102 Se a estatística observada for $\chi^2_{\text{obs}} = 9,10$, não haverá evidências suficientes para rejeitar H_0 ao nível de significância de 3%.

Espaço livre

Na determinação do tamanho n de uma amostra aleatória simples para estimar uma proporção populacional p de modo a obter um erro de estimação e , pode-se utilizar a fórmula $n = z_y^2 p(1-p)/e^2$, em que z_y é o valor obtido da distribuição normal padrão, associado ao nível de confiança $0 < y < 1$.

Com base nas informações apresentadas, julgue os itens subsequentes.

- 103** Uma amostra por conglomerados utilizada para se estimar a proporção populacional terá tamanho maior que uma amostra aleatória simples com o mesmo objetivo, a mesma margem de erro e o mesmo nível de confiança.
- 104** Mesmo fixados y e e , caso nenhuma informação acerca da população seja conhecida, não é possível determinar o tamanho de uma amostra que possa ser utilizada para estimar a proporção populacional.
- 105** Se a população puder ser dividida em estratos, uma amostra estratificada utilizada para se estimar a proporção populacional terá tamanho maior que uma amostra aleatória simples com o mesmo objetivo, a mesma margem de erro e o mesmo nível de confiança.
- 106** Fixados y e e , o tamanho da amostra previsto pela fórmula $n = z_y^2 p(1-p)/e^2$ será tanto maior quanto maior for a estimativa inicial usada para a proporção populacional.
- 107** Se uma amostra-piloto fornece a estimativa de 10% para a proporção populacional, então é possível, com erro de 5% e nível de confiança de 95% ($z_y = 1,96$), estimar a proporção populacional com uma nova amostra aleatória simples de tamanho inferior a 140.

Um estatístico está planejando uma pesquisa para mensurar a proporção de motoristas insatisfeitos com o uso do estacionamento de um grande *shopping* da cidade, que abre todos os dias às 10 horas da manhã. O estacionamento tem somente uma entrada e uma saída. O estatístico sabe que cerca de dez mil motoristas entram no local diariamente e decide que, escolhido um dia de maior movimento, utilizará um dos seguintes planos amostrais.

Plano amostral I – Entrevista-se o primeiro motorista que chega ao estacionamento e, depois dele, entrevista-se sempre o décimo motorista após o último entrevistado, e assim por diante, até às 22 horas ou até completar 460 motoristas entrevistados.

Plano amostral II – Selecionam-se aleatoriamente 460 números de uma lista numerada de 1 a 10.000 e entrevistam-se os motoristas correspondentes aos números selecionados, descartando-se os números excedentes às 22 h, caso haja.

Plano amostral III – Categorizam-se previamente os automóveis nas seguintes classes: marca *premium* (lista previamente escolhida); veículos grandes (picapes e utilitários), desde que não sejam de marca *premium*; e veículos pequenos, desde que não sejam de marca *premium*. Entrevistam-se, na saída do estacionamento, 460 motoristas cujos veículos estão distribuídos entre essas classes, selecionados aleatoriamente com base na proporção entre as classes verificada no dia anterior.

Plano amostral IV – Setoriza-se o estacionamento em três setores: área descoberta, primeiro subsolo e segundo subsolo. Às 18h, selecionam-se aleatoriamente 200 automóveis no primeiro setor e 130 em cada um dos outros, entrevistando-se os respectivos motoristas no momento em que deixam o *shopping*.

Com base na situação hipotética precedente, julgue os itens seguintes.

- 108** O plano amostral III apresenta características de amostragem estratificada.
- 109** Uma amostra gerada pelo plano amostral II não pode ser classificada como uma amostra aleatória simples.
- 110** O plano amostral IV tem características de amostragem por conglomerados.
- 111** O plano amostral I apresenta características de amostragem sistemática.
- 112** Caso seja usado o plano amostral II, todos os motoristas que entrarem no estacionamento no dia escolhido terão igual probabilidade de serem entrevistados.

A tabela a seguir apresenta o tempo y , em anos, que cinco clientes levam até decidir trocar de operadora de telefone, bem como sua renda mensal x , em salários mínimos.

x	10	13	5	7	20
y	4	3	6	5	2

Nesse contexto, um modelo de regressão linear na forma $E(Y|x) = a + bx$ é construído para se ajustar aos dados por critério que minimize a soma dos quadrados dos erros.

Com base nas informações apresentadas e supondo que o modelo apresentado possa ser usado para prever o comportamento das demais pessoas, julgue os próximos itens.

- 113** No modelo apresentado, o parâmetro b representa o quanto a média da variável que mensura o tempo médio que um cliente leva até decidir trocar de operadora varia para um aumento unitário na variável renda.
- 114** Infere-se do modelo apresentado que quanto maior a renda do cliente, menor é o tempo que ele leva até decidir trocar de operadora.
- 115** Segundo o modelo apresentado, um cliente com renda de 8 salários mínimos levará mais de 4 anos e 9 meses para decidir trocar de operadora.
- 116** Para um cliente com renda de 7 salários mínimos, o valor absoluto do erro ao se usar o modelo em questão frente ao dado amostral é superior a 2 meses.

Espaço livre

Espaço livre

O modelo de regressão para as duas v.a. quantitativas, X e Y , se escreve como $E(Y|x) = \alpha + \beta x$. Para se determinar os parâmetros α e β , são feitas n observações x_i e y_i e, nesse caso, o modelo pode ser escrito como $y_i = E(Y|x_i) + e_i = \alpha + \beta x_i + e_i$, em que $i = 1, 2, \dots, n$, e em que e_i é o erro do modelo frente à i -ésima observação, devendo-se encontrar os valores mais prováveis para α e β , segundo algum critério, a partir das n observações de pares de valores de (X, Y) .

No caso do critério dos mínimos quadrados, segundo o qual os valores das incógnitas α e β são determinados de modo a minimizar a soma dos quadrados dos erros, é necessário, a fim de encontrar os estimadores para os parâmetros do modelo, considerar as seguintes hipóteses para as v.a. envolvidas.

- 1 A variável X é controlada e não está sujeita a variações aleatórias, ou seja, X é uma variável fixa.
- 2 Para dado valor x de X , os erros distribuem-se ao redor da média $\alpha + \beta x$ com média zero, isto é, $E(e_i|x) = 0$.
- 3 Os erros têm a mesma variabilidade em torno dos níveis de X , ou seja, $Var(e_i|x) = \sigma_e^2$, para todo $i = 1, \dots, n$.
- 4 Os erros são não correlacionados.

Para verificar se o modelo é adequado aos dados, deve-se investigar se as suposições feitas para o desenvolvimento do modelo estão satisfeitas. Para tanto, deve-se fazer a análise dos resíduos. Uma técnica aplicável é a análise gráfica, que consiste em plotar os pares $(x_i, \hat{e}_i)$, em que $i = 1, \dots, n$; e $\hat{e}_i$ é a diferença entre o valor observado y_i e o valor previsto pelo modelo. Os gráficos a seguir ilustram situações típicas.

Gráfico de resíduos (a)

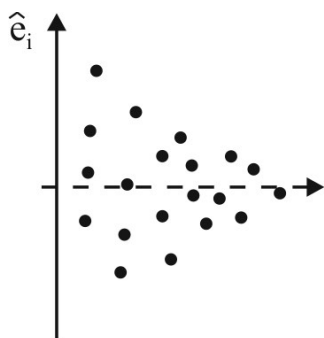
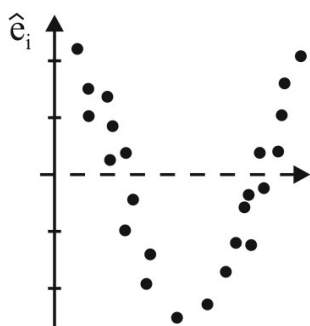


Gráfico de resíduos (b)



A partir das informações precedentes, julgue os seguintes itens.

- 117 No modelo de regressão linear simples estimado pelo método dos mínimos quadrados ordinários (MQO), a soma dos resíduos é sempre igual a zero.
- 118 Para que a hipótese 2 seja satisfeita, espera-se que o gráfico de resíduos apresente pontos $(x_i, \hat{e}_i)$ distribuídos aproximadamente em torno da reta $y = \alpha + \beta x$.
- 119 O gráfico de resíduos (a) apresenta uma situação em que a hipótese 3 é satisfeita.
- 120 O gráfico de resíduos (b) apresenta uma situação em que a hipótese 1 é satisfeita.