CONHECIMENTOS ESPECÍFICOS

Acerca dos conceitos d3e estatística e dos parâmetros estatísticos, julgue os itens seguintes.

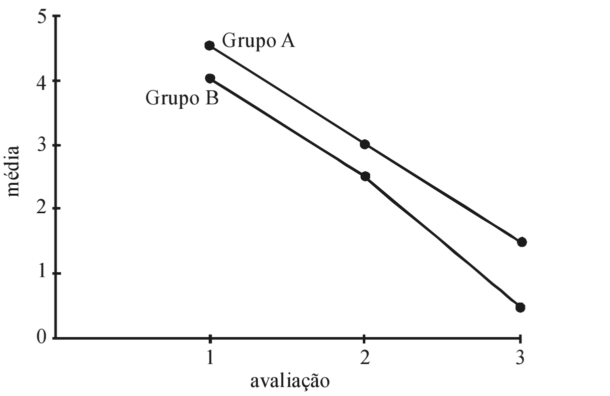
61 Em estatística, parâmetro pode ser uma quantidade desconhecida da população-alvo, à qual não se tem acesso diretamente, mas que se deseja estimar ou a respeito da qual se deseja avaliar hipóteses.

62 A estatística descritiva permite testar hipóteses a respeito da população de interesse.

Com relação a representações gráficas, julgue os itens de **63** a **67**.

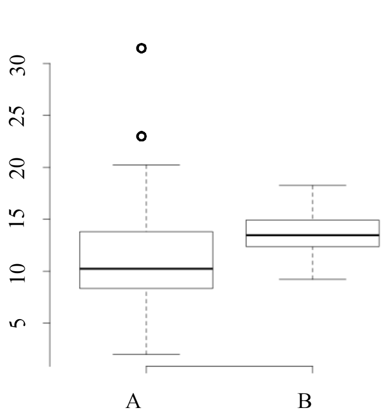
63 Um histograma é um gráfico que representa a dispersão tanto de variáveis discretas quanto de contínuas.

64 O gráfico de perfil de médias mostrado na figura abaixo sugere uma interação entre os grupos A e B na amostra em questão.



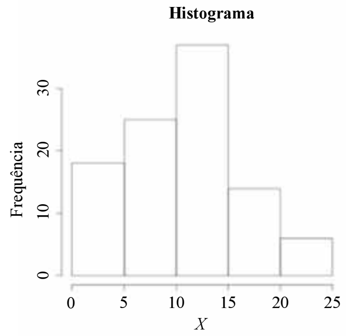
65 Considerando-se que o *boxplot* mostrado na figura a seguir compara a distribuição dos salários de dois grupos da população, é correto afirmar que existem, necessariamente, diferenças entre as distribuições populacionais desses salários.

Salários por grupo



66 O gráfico de setores, quando descreve a distribuição de uma variável quantitativa, pode ser usado para se obter uma estimativa da média amostral.

67 A partir do histograma mostrado na figura abaixo, é correto inferir que a distribuição da variável X é simétrica.



Considere o seguinte conjunto de dados composto por cinco elementos: {82,93; 94,54; 98,40; 115,41; 123,07}.

Com base nesses dados, julgue os itens subsequentes acerca das medidas de tendência central.

68 Em uma distribuição de dados unimodal, se a média e a mediana forem iguais, não é possível determinar o valor da moda se todos os dados não estiverem disponíveis.

69 Considerando-se um conjunto de n dados, $\{x_1, x_2, ..., x_n\}$, se não se dispõe de todos esses dados, mas apenas dos valores da média, do mínimo e do máximo dos dados, a expressão a seguir constitui um bom estimador para a mediana desse conjunto. $\tilde{x} = \dfrac{\min(x_1, x_2, ..., x_n) + \max(x_1, x_2, ..., x_n)}{2}$

70 A média do conjunto de dados em questão é 102,87 e a mediana é 98,40. Se o valor 123,07 for alterado para 200, a média irá aumentar, mas a mediana continuará sendo 98,40.

71 Se o valor de um dos elementos do conjunto não for fornecido, esse valor pode ser determinado se a média do conjunto for conhecida, mas não será possível obter esse valor conhecendo-se apenas a mediana.

RASCUNHO

Com relação às medidas de dispersão, considerando que $\overline{X}$ representa a média aritmética de um conjunto de dados $\{X_1, X_2, ..., X_n\}$, julgue os itens que se seguem.

72 A expressão que se segue permite calcular corretamente a variância amostral do conjunto de dados.

$$V(X) = \left(\frac{1}{n-1}\right)\left(\sum_{i=1}^{n} X_i^2 - n \times \overline{X}^2\right)$$

73 A função a seguir não é uma medida de dispersão válida.

$$D(X) = \frac{1}{n} \cdot \sum_{i=1}^{n} e^{(X_i - \overline{X})}$$

74 Se R_i é o posto do valor X_i, $i = 1, 2, ..., n$, então a variância de R_i é uma medida de dispersão para X.

75 A expressão abaixo é uma medida correta de dispersão, para qualquer valor de $k > 0$.

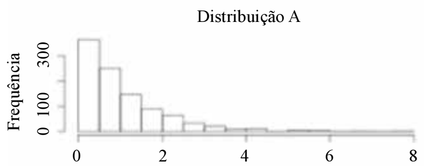
$$D(X) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} (X_i - \overline{X})^k,\ k > 0,\ k \in Z$$

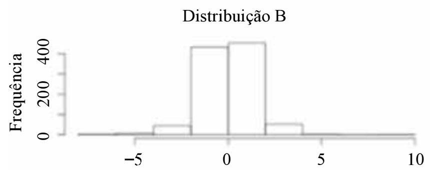
A respeito das medidas de assimetria e curtose, julgue os próximos itens.

76 A assimetria de uma distribuição correspondente a um conjunto de dados X com n elementos e média aritmetica $\overline{X}$, pode ser corretamente calculada a partir da equação abaixo, tal que o numerador dessa expressão somente será zero se a distribuição for simétrica.

$$AS(X) = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^3}{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2\right)^{\frac{3}{2}}}$$

77 Entre as distribuições A e B a seguir, aquela que melhor representa uma distribuição com assimetria negativa e curtose positiva é a distribuição A.





78 A curtose de uma distribuição correspondente a um conjunto de dados X com n elementos e média aritmetica $\overline{X}$, pode ser corretamente calculada a partir da expressão abaixo. No caso em que X tem distribuição normal, a curtose é zero.

$$K(X) = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^4}{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2\right)^2} - 3$$

RASCUNHO

Acerca da teoria de probabilidades, julgue os itens subsecutivos.

79 A função $f(x) = [1 - e^{-\lambda x}]^n$ representa a densidade da distribuição de probabilidades do valor máximo observado em uma amostra aleatória simples de tamanho n de uma distribuição exponencial com parâmetro $\lambda > 0$.

80 O número de possíveis retiradas de bolas de uma urna, sem reposição, é menor quando a ordem importa.

81 Se 80% de uma população pertence ao grupo A e 60%, ao grupo B, e sabendo que a interseção entre os grupos A e B não é vazia, então a probabilidade da interseção de A e B é maior ou igual a 0,4.

82 Se $X = I(A)$ é uma função indicadora da ocorrência do evento A, então $E(X) = P(A)$, em que $E(X)$ é o valor esperado de X e $P(A)$, a probabilidade de ocorrência do evento A.

83 A distribuição uniforme contínua em [0, 1] é um caso degenerado da distribuição beta.

84 Considere que $F(x)$ seja a função de distribuição acumulada de uma variável aleatória X. Nessa situação, $G(x) = \dfrac{1 - e^{F(x)}}{1 - e}$ não pode corresponder a uma função de distribuição acumulada.

Com relação aos planos amostrais, julgue o próximo item.

85 A diferença principal entre amostragem estratificada e amostragem por conglomerados é que, no caso da estratificada, a população é dividida artificialmente em estratos, e, no caso da amostragem por conglomerados, a população já é naturalmente dividida em subpopulações.

A respeito de estimadores, julgue os itens a seguir.

86 Qualquer estimador obtido pelo método dos momentos é uma função de estatística suficiente.

87 Considere que $\overline{X}$ represente a média aritmética de um conjunto de dados $\{X_1, X_2, \ldots, X_n\}$, que $S^2 = \left(\dfrac{1}{n-1}\right)\sum_{i=1}^{n}(X_i - \overline{X})^2$ seja um estimador não viciado para a variância populacional e $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ seja um estimador viciado para essa mesma variância. Nessa situação, o estimador, $\hat{\sigma}^2$ será assintoticamente não viciado.

88 Para uma amostra aleatória simples $\{X_1, X_2, ..., X_n\}$, conhecendo-se a estatística $T(X) = \sum_{i=1}^{n} X_i$ e não sendo necessário que se conheçam os dados para se estimar a média populacional de X, é correto afirmar que $T(X)$ é uma estatística completa.

89 A média amostral obtida com base em uma amostra aleatória simples é um estimador inconsistente da média populacional.

90 Todo estimador viciado pode ser consistente.

RASCUNHO

Julgue os itens que se seguem, acerca de definições da teoria estatística.

91 O erro do tipo II de um teste de hipóteses ocorre quando se rejeita uma hipótese nula que é verdadeira.

92 Quanto menor for o p-valor associado a uma estatística de determinado teste, maior será a evidência para se rejeitar a hipótese nula desse teste.

93 No contexto da teoria da decisão estatística, ao se considerar uma função perda dada por erro médio absoluto, a mediana é obtida como estimador da localização da população.

Considerando que W seja um estimador pontual de um parâmetro $\theta > 0$ de uma distribuição D, julgue os itens a seguir.

94 Se W for obtido por máxima verossimilhança, então $\log W$ é o estimador de máxima verossimilhança de $\log\theta$.

95 Considerando-se que D seja uma distribuição qualquer de valores não negativos, se W for obtido por máxima verossimilhança e se $E(W) = \theta$, então $E(\log W) = \log\theta$, em que E corresponde ao valor esperado.

96 Independentemente da forma da distribuição D, o estimador W produzido pelo método dos momentos não é viciado para θ.

Os procedimentos estatísticos paramétricos incluem

97 a estimação de uma densidade por meio de suavização do histograma.

98 a estimação da densidade da distribuição Gama(a, b), estimando-se os parâmetros a e b pelo método dos momentos.

99 o teste dos postos de Wilcoxon.

100 os testes com base em permutação.

A respeito do problema de otimização, julgue os próximos itens.

101 Dualidade é um nome que se dá à característica de uma função objetivo que possui tanto ótimo global como ótimos locais.

102 O método *simplex* enumera todas as soluções básicas e procura a solução ótima por meio de derivadas primeiras.

No que concerne a diagnóstico em análises de regressão, julgue o item a seguir.

103 Uma observação pode ser discrepante e não influente.

Julgue os seguintes itens, acerca do coeficiente de determinação (R^2) de uma análise de regressão linear feita com base em estimação por mínimos quadrados ordinários.

104 O coeficiente de determinação R^2 da regressão linear simples $Y = b_0 + b_1X + \epsilon$, em que b_0 e b_1 são os coeficientes do modelo, corresponde ao quadrado da correlação estimada entre Y e ϵ.

105 Se $R^2 = 1$, todos os dados estarão alinhados sobre uma reta de inclinação positiva ou negativa.

106 Se um segundo preditor (X_2) for adicionado ao modelo $Y = b_0 + b_1X_1 + \epsilon$, em que b_0 e b_1 são os coeficientes do modelo, o valor do erro padrão dos resíduos pode aumentar ou diminuir, mas o valor de R^2 não pode diminuir.

Considerando um modelo de regressão linear com intercepto b_0 e três variáveis regressoras cujos respectivos coeficientes são b_1, b_2 e b_3, julgue os itens subsequentes.

107 A hipótese nula H_0: $b_2 = b_3 = 1$ pode ser testada comparando-se o valor do coeficiente de determinação R^2 do modelo saturado com o R^2 correspondente ao modelo restrito à condição $b_2 = b_3 = 1$.

108 É correta a utilização de um teste t com base na estimativa da soma $b_1 + b_3$ para se testar H_0: $b_1 + b_3 = 5$.

109 Para se testar a hipótese nula H_0: $b_1 = b_2$, o teste F é feito comparando-se a soma de quadrados dos resíduos do modelo completo com a soma de quadrados dos resíduos do modelo restrito à hipótese $b_1 = b_2$.

Se, em determinada fábrica, 10% das peças produzidas são defeituosas, então, para fins de controle de qualidade, uma distribuição binomial negativa deve ser usada na situação em que

110 se deseje, em uma amostra aleatória simples com reposição, obter a probabilidade de a terceira peça defeituosa ocorrer na décima retirada.

111 é retirada uma amostra aleatória simples com reposição de 10 peças para se determinar a probabilidade de ocorrer exatamente 3 peças defeituosas nessa amostra.

112 se deseja calcular a probabilidade de a primeira peça defeituosa ocorrer na décima retirada, no caso de as peças serem retiradas por amostragem aleatória simples com reposição.

$$x_t = 0{,}7x_{t-1} + a_t - 2{,}5a_{t-1} + \theta a_{t-2}$$

$$y_t = 0{,}2y_{t-1} + a_t$$

$$z_t = z_{t-1} + a_t - 0{,}1a_{t-1}$$

Considere as séries temporais acima, em que a_t é um ruído branco com média 0 e variância não nula σ^2. Com base nessas informações, julgue os itens que se seguem.

113 A série $\{y_t\}$ pode ser escrita como um processo de médias móveis de ordem infinita:
$y_t = a_t + 0{,}2a_{t-1} + 0{,}2^2a_{t-2} + 0{,}2^3a_{t-3} + 0{,}2^4a_{t-4} + ...$

114 Se $\sigma < 1$, então a série z é estacionária.

115 A série $\{x_t\}$ é estacionária para qualquer valor θ.

116 A série temporal $\{x_t\}$ possui representação na forma AR(∞) se $\theta = 0$.

Considerando a série temporal $x_t = T_t + S_t + e_t$, em que T é o componente de tendência, S é o componente de sazonalidade e e é um componente aleatório de média 0 e variância constante, julgue os itens a seguir.

117 Um modelo de regressão com tendência polinomial e variáveis *dummies* para descrever a componente S pode ser usado corretamente para a estimação dos componentes da série temporal x_t.

118 O método de suavização por médias móveis não é aplicável para essa situação, pois a série x_t possui sazonalidade.

Acerca dos métodos de Paasche e Laspeyres para a construção de números índices e da utilização de índices de preços no deflacionamento de valores monetários, julgue os itens subsecutivos.

119 O índice de preços de Paasche requer, além dos preços, o conhecimento das quantidades vendidas em todos os períodos para os quais se pretende calcular o índice, ao passo que o índice de Laspeyres requer, além dos preços, as quantidades referentes ao ano base.

120 Se $D(1)$ é o valor monetário de uma dívida contraída no período 1 e ocorre uma inflação de 25% entre os períodos 1 e 2, então, o valor deflacionado (ou real) da dívida no período 2, medido em moeda do período 1, é $D(2) = 1{,}25 \times D(1)$.

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando o espaço para rascunho indicado no presente caderno. Em seguida, transcreva o texto para a **FOLHA DE TEXTO DEFINITIVO DA PROVA DISCURSIVA**, no local apropriado, pois **não serão avaliados fragmentos de texto escritos em locais indevidos**.
- Qualquer fragmento de texto além da extensão máxima de linhas disponibilizadas será desconsiderado.
- Na **folha de texto definitivo**, identifique-se apenas no cabeçalho da primeira página, pois **não será avaliado** texto que tenha qualquer assinatura ou marca identificadora fora do local apropriado.

Inferências acerca da média de uma população normal constituem um dos procedimentos estatísticos mais comuns na análise de observações univariadas independentes e identicamente distribuídas. Com respeito a inferências sobre médias populacionais, a comparação entre duas populações requer atenção especial no que se refere à normalidade e aleatoriedade dos dados, homocedasticidade, independência entre as amostras, entre outros aspectos. Nesse contexto, considere as seguintes situações estatísticas que envolvem duas amostras.

Situação I. Duas amostras aleatórias independentes, denotadas por $X_{11},\ldots,X_{1n_1}$ e $X_{21},\ldots,X_{2n_2}$, são provenientes de populações normais, em que cada X_{ij} tem média μ_i e variância σ_i^2, estimadas, respectivamente, por $\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ e $S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2$.

Situação II. Vinte servidores de um órgão público federal realizarão uma tarefa padrão sob condições idealizadas. O tempo X_i que cada servidor i, i = 1, 2, ..., 20, gastará para a execução dessa tarefa será medido, formando-se uma amostra aleatória $X_1, \ldots, X_{20}$. Em seguida, esses mesmos vinte servidores serão submetidos a um treinamento que visa a melhoria da produtividade. Após o treinamento, eles efetuarão novamente a tarefa padrão, medindo-se os tempos de execução Y_i, i = 1, 2, ...,20, produzindo uma amostra aleatória $Y_1, \ldots, Y_{20}$.

Situação III. Determinada população possui dois estratos. No estrato 1, a variável de interesse X_1 segue uma distribuição normal com média μ e variância $\sigma_1^2 > 0$. No estrato 2, a variável de interesse X_2 possui distribuição normal com a mesma média μ, porém com variância $\sigma_2^2 \neq \sigma_1^2 > 0$. Dos estratos 1 e 2, são retiradas amostras aleatórias simples, independentes entre si, de tamanhos n_1 e n_2, respectivamente.

Considerando as três situações descritas no texto acima, redija um texto dissertativo, que tem como foco testes estatísticos para comparação de duas médias, abordando, necessariamente, os seguintes aspectos:

- na **Situação I**, utilizando a notação apresentada, identifique a estatística do teste *t* e sua distribuição sob a hipótese nula para testar se as médias populacionais são iguais, considerando o caso de populações com variâncias iguais;
- na **Situação II**, identifique a estatística teste e sua distribuição sob a hipótese nula para se testar o efeito do treinamento no tempo médio de execução da tarefa;
- para a **Situação III**, com o intuito de utilizar as duas amostras para se estimar a média μ, considere que: Pedro proponha utilizar a média aritmética simples de $\overline{X}_1$ e $\overline{X}_2$ como estimador de μ, em que $\overline{X}_i$ é a média amostral no estrato i = 1 ou 2; Marcos sugira juntar todas as $(n_1 + n_2)$ observações e utilizar a média global como estimador de μ; João afirme que uma média ponderada de $\overline{X}_1$ e $\overline{X}_2$ com pesos dependendo de σ_1^2 e σ_2^2 deveria ser utilizada, desde que essas duas variabilidades σ_1^2 e σ_2^2 sejam conhecidas. Com base no critério estatístico de erro médio quadrático mínimo, discuta quem tem a melhor proposta.

RASCUNHO