CONHECIMENTOS ESPECÍFICOS

fonte de variação	graus de liberdade	soma de quadrados	quadrado médio	teste F
regressão	1	1,4	1,4	46,2
resíduo	28	0,8	0,03	
total		2,2		

A tabela de análise de variância (ANOVA) ilustrada acima resulta de uma regressão linear simples, cuja variável resposta — y — é a proporção de área devastada por uma queimada, e a variável explicativa — x — representa a umidade relativa do ar registrada momentos antes da ocorrência da queimada. Com base nessas informações e na tabela acima, e sabendo que as médias amostrais de y e x são 0,48 e 28,9, respectivamente, julgue os itens que se seguem.

71 Se o coeficiente de variação da variável x for inferior a 14%, o desvio padrão amostral dos registros de umidade relativa do ar será superior a 17,5.

72 O gráfico de dispersão permite verificar se as variáveis x e y são correlacionadas.

73 Considerando-se a análise de resíduos produzidos pelo modelo ajustado, é correto aplicar o teste de Kolmogorov-Smirnov para a verificação da normalidade dos dados e o teste de White para verificar se os resíduos são aleatórios.

74 Se os erros aleatórios seguirem distribuição normal com média zero e variância V, o terceiro momento central (assimetria) e o quarto momento central (curtose) da distribuição desses erros aleatórios deverão ser nulos.

75 A correlação linear entre as variáveis x e y é inferior a 0,7.

76 Na tabela ANOVA, o tamanho da amostra é igual a 29.

77 O valor absoluto da razão t referente ao teste de significância para o coeficiente angular é inferior a 7.

78 Considerando-se que os resultados mostrados na tabela ANOVA tenham sido produzidos com base em estimativas de mínimos quadrados ordinários (MQO) para os coeficientes do modelo de regressão, e assumindo-se que a variável resposta segue uma distribuição normal, então, se fosse empregado o método da máxima verossimilhança, a estimativa do coeficiente angular seria diferente da estimativa de MQO.

A respeito de probabilidade, julgue os itens de **79** a **86**.

79 Se A, B e C são eventos independentes com probabilidades não nulas, então $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

80 Se U segue uma distribuição uniforme (0, 1) e $\lambda > 0$, então a transformação $X = \dfrac{\ln(1-U)^{-1}}{\lambda}$ resulta em uma distribuição exponencial com densidade $f(x) = \lambda e^{-\lambda x}$, em que $x > 0$.

81 Considerando-se um espaço amostral W e a sequência de eventos $A_1, A_2, ..., A_n$, em que $A_1 \subset A_2 \subset ... \subset A_n \subset \Omega$ e $P(A_i) > 0$, então $0 < P\left(\bigcup_{i=1}^{n} A_i\right) < \sum_{i=1}^{n} P(A_i)$.

82 Se X e Y forem variáveis aleatórias com média condicional $g(y) = E(X | Y = y)$, então $E(X) = E(g(Y))$.

83 Se $F(x) = \dfrac{1}{1+e^{-x}}$ for a função de distribuição acumulada da variável aleatória X, então a função densidade de probabilidade será $f(x) = \dfrac{e^{-x}}{1+e^{-x}}$.

84 Considere uma distribuição binomial com parâmetros n e p. Se n for muito grande e se p for muito pequeno, então essa distribuição binomial poderá ser aproximada por uma distribuição de Poisson com média $\lambda = np$.

85 Se Z segue uma distribuição normal padrão, então a variável $X = 4Z + 2$ segue uma distribuição normal com média 2 e variância 4.

86 Considerando-se que uma amostra aleatória simples $X_1, ..., X_{49}$ tenha sido retirada de uma distribuição binomial negativa com média 2 e variância 4, é correto afirmar, com relação à média aritmética dessa amostra, que, de acordo com o teorema Limite Central, $P(\bar{X} \leq 3) > 0{,}55$.

RASCUNHO

Pesquisadores desenvolveram um novo dispositivo para medir a velocidade de uma aeronave e, em um oratório especial, submeteram uma amostra aleatória de 36 réplicas da aeronave (amostra A) a um teste de operação, medindo a temperatura mínima necessária para o bom funcionamento de cada réplica.

Considerando essa situação, julgue os itens que se seguem, acerca de inferência estatística.

87 Considere que os citados pesquisadores tenham selecionado uma amostra aleatória composta de n réplicas de outro protótipo (amostra B) para a medição de velocidade de uma aeronave, submetendo-a ao mesmo teste de operação que a amostra A. Nesse caso, se o p-valor do teste de Levene for igual a 0,005, é adequado o uso do teste de Wilcoxon para a comparação das médias amostrais das temperaturas mínimas de operação obtidas nas amostras A e B.

88 Suponha que a distribuição das temperaturas mínimas de operação seja simétrica em torno da média populacional e que tenham sido observadas ocorrências de valores nas caudas da distribuição em frequência superior ao esperado considerando-se a distribuição dessas temperaturas normal. Nesse caso, a mediana amostral — Md — será um estimador mais eficiente que a média amostral — $\bar{X}$ —, pois $\frac{Var(Md)}{Var(\bar{X})} = \frac{2}{\pi} < 1$.

89 A estimativa da variância da média amostral da temperatura mínima é igual a $\sum_{i=1}^{36} \frac{(x_i - \bar{x})^2}{36 \times 35}$, em que x_i representa a temperatura mínima relativa ao dispositivo i e $\bar{x} = \sum_{i=1}^{36} \frac{x_i}{36}$.

90 O teste qui-quadrado é eficiente para testar a hipótese de normalidade das temperaturas mínimas.

91 Se o erro do tipo I for dado por $P(\bar{X} \leq x_c)$ para H_0: $\mu = -29$, então o poder do teste para H_1: $\mu = -31$ será dado por $1 - P(\bar{X} \geq x_c)$.

92 Considere que o intervalo de confiança de 95% para a média da temperatura mínima de operação do novo dispositivo tenha sido [−47 ºC; −45 ºC]. Nessa situação, se o nível de confiança aumentar de 95% para 99%, a amplitude do intervalo de confiança de 99% será inferior a 2 ºC.

93 Se os pesquisadores quiserem comparar a média amostral das temperaturas mínimas de operação da aeronave com determinado valor médio hipotético e se as temperaturas seguirem uma distribuição normal, o teste t de Student com 35 graus de liberdade torna-se viável para o estudo.

94 Considere que, em um teste de hipóteses unilateral para a média populacional, a média aritmética das temperaturas mínimas de operação observada nessa amostra aleatória tenha sido igual a −30 ºC, e a temperatura especificada na hipótese nula, igual a −29 ºC. Nesse caso, o p-valor desse teste será igual a $P(\bar{X} \leq -29)$, em que $\bar{X}$ representa a média amostral.

95 Os estimadores de momentos para a média μ e desvio padrão σ de uma distribuição normal, com base em uma amostra aleatória simples X_1, X_2, ..., X_n, podem ser obtidos a partir da solução do sistema de equações de estimação apresentado abaixo.

$$\begin{cases} \sum_{i=1}^{n} X_i - \mu = 0 \\ \sum_{i=1}^{n} \sqrt{X_i^2 - \mu^2} - \sigma = 0 \end{cases}$$

RASCUNHO

Com relação à amostragem aleatória simples em população de tamanho finito, julgue os itens seguintes.

96 Considere os métodos clássicos para a determinação do tamanho mínimo da amostra para a estimação da média populacional com base em uma amostra aleatória simples. Nesse caso, se o tamanho mínimo da amostra para uma seleção com reposição — *nc* — for $\frac{3}{2}$ do tamanho mínimo correspondente para uma seleção sem reposição — *ns* —, então o tamanho da população será igual ao triplo de *nc*.

97 Na amostragem aleatória simples sobre uma população de tamanho finito, a média amostral $\bar{X}$ é o estimador não viciado para a média populacional μ. Nesse caso, se a amostragem for realizada sem reposição, a variância de $\bar{X}$ será menor que a realizada com reposição.

Julgue os itens subsequentes, relativos a amostragem aleatória simples com reposição, considerando que τ_Y seja o estimador do total populacional da variável *Y*, que $T_R = \frac{\bar{y}}{\bar{x}} \times \tau_X$ seja o estimador razão para o total populacional da variável *Y* e que τ_X seja o estimador do total populacional da variável *X*.

98 Supondo-se que o estimador de regressão para a média populacional da variável *Y* seja dado pela expressão $\bar{y}_{\text{Reg}} = \bar{y}_R + \beta \times (\mu_X - \bar{x})$, em que $\bar{y}_R = \frac{\bar{y}}{\bar{x}} \times \mu_X$ e $\beta = \frac{\sigma_{XY}}{\sigma_X^2}$, σ_{XY} representa a covariância entre *X* e *Y*, e σ_X^2 é a variância de *X*, é correto afirmar que $Var(\bar{y}_{\text{Reg}}) \leq Var(\bar{y}_R)$.

99 Supondo-se que a correlação entre *X* e *Y* seja superior a $\frac{1}{3}$, que o coeficiente de variação de *X* seja igual a 2 e que o coeficiente de variação de *Y* seja igual a 3, é correto afirmar que, para uma amostra suficientemente grande, $Var(\tau_Y) = Var(T_R)$.

A respeito de amostragem estratificada, considerando que o estimador regressão para a média populacional de uma variável *Y* seja dado pela expressão $\bar{y}_{\text{Reg}} = \sum_i \omega_i \times \bar{y}_{\text{Reg}_i}$, em que $\bar{y}_{\text{Reg}_i}$ é o estimador de regressão para a média populacional de *Y* relativo ao estrato *i* e ω_i é o fator de expansão, julgue o próximo item.

100 A variância do estimador de regressão para a média populacional de *Y* pode ser representado pela expressão $Var(\bar{y}_{\text{Reg}}) = \sum_i \frac{\omega_i^2}{n_i} [\sigma_{Y_i}^2 - 2 \times \beta_i \times \sigma_{XY_i} + \beta_i^2 \times \sigma_{X_i}^2]$, em que n_i representa o tamanho do estrato *i*, $\sigma_{Y_i}^2$ e $\sigma_{X_i}^2$ são, respectivamente, as variâncias de *Y* e *X* no estrato *i*, $\sigma_{XY_i}^2$ representa a covariância entre *Y* e *X* no estrato *i*; e β_i é um coeficiente. Quando $\beta_i = \frac{\sigma_{XY_i}}{\sigma_{X_i}^2}$, essa variância atinge seu valor mínimo, cujo resultado é $Var(\bar{y}_{\text{Reg}}) = \sum_i \frac{\omega_i^2}{n_i} \times \sigma_{Y_i}^2 \, [1 - \rho_{XY_i}^2]$, em que $\rho_{XY_i}^2 = \frac{\sigma_{XY_i}^2}{\sigma_{X_i}^2 \times \sigma_{Y_i}^2}$.

RASCUNHO

Com relação ao gerador de números pseudoaleatórios, em que se utiliza a relação $X_{n+1} = [a \times X_n + b] \bmod m$, julgue os itens que se seguem.

101 Se $u_{n+1} = \dfrac{X_{n+1}}{m}$ segue uma distribuição uniforme em [0, 1], então $y_{n+1} = \dfrac{u_{n+1} - \frac{1}{2}}{\tau}$ segue uma distribuição uniforme em $\left[\dfrac{-1}{2\times\tau}, \dfrac{1}{2\times\tau}\right]$.

102 Se $\dfrac{X_{n+1}}{m} \sim U[0,1]$ e $Z_{n+1} = -\dfrac{1}{\lambda} \times \log\left(1 - \dfrac{X_{n+1}}{m}\right)$, então $E(Z_{n+1}^2) = 2 \times E(Z_{n+1})^2$.

103 Considere que, de acordo com um método de geração de variáveis aleatórias, se $u_{n+1} \sim U[0, 1]$, então $z_{n+1} = F^{-1}(u_{n+1})$ segue uma distribuição com função de probabilidade acumulada dada por $F(z_{n+1})$. Nessa situação, esse método é adequado para simular uma distribuição binomial de parâmetros n e p.

104 O período da sequência de números pseudoaleatórios é igual a $\dfrac{a}{m}$.

O comandante de um batalhão do Corpo de Bombeiros dispõe de uma amostra de tamanho n do tempo X, necessário para atender a uma chamada de emergência. Com o objetivo de conhecer a distribuição de tal variável, o comandante aplicou o seguinte esquema de reamostragem: dessa amostra original seleciona-se uma nova amostra aleatória simples com reposição de tamanho n e calcula-se o tempo mediano da chamada. Esse procedimento é replicado M vezes, em que M é um número grande o suficiente, resultando em uma distribuição amostral empírica de tempos medianos.

Com base nessa situação hipotética, julgue os próximos itens, relativos ao método computacional descrito.

105 Se o tamanho n da amostra original for grande o suficiente, independentemente da forma da distribuição dos tempos X, a distribuição amostral das estimativas das medianas terá distribuição aproximadamente normal.

106 Tal procedimento, denominado *jackknife*, produz uma estimativa não paramétrica da distribuição amostral da mediana dos tempos de atendimento de chamadas.

107 O número de replicações M não depende do tamanho n da amostra original.

RASCUNHO

Considerando um processo estocástico $\{X_n: n = 1, 2, ...\}$, tal que $E(|X_n|) < \infty$ e $E(X_{n+1}| X_1, X_2, ..., X_n)$, isto é, um processo martingale, julgue os itens seguintes.

108 Considere um jogo no qual se ganham k unidades com probabilidade p e se perdem w unidades com probabilidade $1 - p$. Nesse caso, o ganho a cada passo do jogo será um processo martingale, para $k \neq w$, somente se $p = \dfrac{k}{k+w}$.

109 Considere que $\{N_t: t = 1, 2, ...\}$ seja um processo de Poisson com parâmetro λ. Nesse caso, o processo de Poisson compensado $\{N_t^* = N_t - \lambda t: t = 1, 2, ...\}$ é um processo martingale.

110 Considere a seguinte situação hipotética.

A população de uma cidade tem sido ameaçada por duas espécies de animais: A e B. Na floresta próxima a essa cidade, existem a casais de animais do tipo A e b casais de animais do tipo B. A cada instante, um casal de uma das espécies é selecionado ao acaso (pela natureza) e gera um novo casal.

Nessa situação, o processo $\{P_n: n = 1, 2, ...\}$, que representa a probabilidade de, no instante n, um casal de animais do tipo A procriar, é um processo martingale.

111 O valor esperado $E(X_{n+1} \mid X_1, X_2, ..., X_n)$ para a próxima observação no movimento browniano é constante no tempo.

Considerando que uma população seja formada por dois tipos de indivíduos — A e B — e que $X_A \sim N(\mu_1, \Sigma_1)$ e $X_B \sim N(\mu_2, \Sigma_2)$, julgue os itens subsequentes, a respeito de análise discriminante.

112 Considere que um pesquisador tenha calculado o logaritmo entre as densidades dos grupos A e B, supondo que as matrizes de variância-covariância fossem diferentes, e tenha obtido, para um vetor x, $\ln\dfrac{f_A(x)}{f_B(x)} = 4{,}75$. Nesse caso, sabendo-se que o ponto foi alocado pelo método dos discriminantes no grupo A, que apenas 1% dos indivíduos da população pertence a tal grupo e que $\ln 99 = 4{,}60$ e $\exp 0{,}15 = 1{,}16$, é correto afirmar que o custo de classificar erradamente um indivíduo do grupo A como pertencente ao grupo B é superior ao custo correspondente à classificação de um indivíduo do grupo B no grupo A.

113 Considerando que as matrizes de variância-covariância sejam iguais nos dois grupos — nesse caso, $\Sigma_1 = \Sigma_2 = \Sigma$ —, o logaritmo entre as densidades dos grupos A e B — $f_A(x)$ e $f_B(x)$ — é corretamente representado pela expressão:

$$\ln\frac{f_A(x)}{f_B(x)} = -\frac{(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)}{2} + (\mu_1 - \mu_2)'\Sigma^{-1}x.$$

RASCUNHO

$$\sum = \begin{bmatrix} 10 & 1 & -1 \\ 1 & 7 & 3 \\ -1 & 3 & 2 \end{bmatrix}$$

Considerando a matriz de variância-covariância acima, julgue o item seguinte.

114 Considere que os dois maiores autovalores dessa matriz sejam λ_1 = 10,30 e λ_2 = 8,29. Nessa situação, os dois componentes principais de maior explicação responderão por 97,84% da variabilidade total dos dados.

$$\sum = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

Considerando a matriz de variância-covariância acima, julgue os itens a seguir.

115 Na análise fatorial, as matrizes de cargas fatoriais obtidas da matriz de variância-covariância apresentada e da matriz de correlação correspondente são idênticas.

116 Cada um dos componentes principais obtidos da matriz de variância-covariância apresentada ou da matriz de correlação correspondente são idênticos e explicam o mesmo percentual da variabilidade.

Julgue os itens que se seguem, relativos aos processos AR(p), assumindo que $\sqrt{17} = 4{,}12$ e $\sqrt{2} = 1{,}41$ e que $\{a_t\}$ é uma sequência de choques aleatórios independentes e identicamente distribuídos $N\,(0, \sigma^2)$.

117 O processo $Y_t = Y_{t-1} - b\,Y_{t-2} + a_t$ é estacionário somente se $b \in [-2,41; 0,41]$.

118 O processo autoregressivo na forma $Y_t = 2\,Y_{t-1} + 4\,Y_{t-2} + a_t$ é estacionário.

Julgue os próximos itens, a respeito do processo $Z_t = -2a_t + a_{t-1} + a_{t-2}$, em que a_t segue um processo de ruído branco.

119 O processo em questão é não estacionário.

120 O referido processo é um filtro linear invertível.

RASCUNHO