

CONHECIMENTOS ESPECÍFICOS

clientes em atraso (N)	45	20	10	3	2
meses em atraso (X)	0	1	2	3	4

A tabela acima mostra a distribuição de frequências do número de meses em atraso nos pagamentos das prestações dos financiamentos de crédito em um grupo de 80 clientes de certa empresa. Considerando que esses clientes formam uma amostra aleatória simples e que atraso é considerado quando $X > 0$, julgue os itens que se seguem, com base nessas informações.

- 51 Considere que a referida amostra de 80 clientes tenha sido retirada da população de clientes da empresa, formada por 1.000 pessoas. Nesse caso, escolhendo-se ao acaso uma dessas pessoas, a probabilidade de ela ser parte da amostra é inferior a 0,05.
- 52 A mediana amostral da variável X é igual a 1.
- 53 O desvio padrão amostral de X é maior que a média amostral de X.
- 54 A variável X representa uma variável qualitativa em escala ordinal.
- 55 O percentual de clientes com atrasos iguais ou superiores a 3 meses nas prestações é inferior a 10%.
- 56 O diagrama de dispersão permite representar corretamente a distribuição de frequências da variável X.
- 57 A distribuição exponencial é um modelo de probabilidade que permite descrever adequadamente a variável X.
- 58 A média de X é superior a 1 mês.
- 59 A variável X apresenta assimetria à direita.
- 60 Suponha que seja selecionada aleatoriamente nova amostra aleatória simples de clientes da referida empresa. Nesse caso, a média amostral da variável X obtida com base nessa nova amostra será igual àquela produzida pela primeira amostra.

Em um jogo de apostas, o apostador escolhe e marca três dezenas distintas em uma cartela que contém 30 dezenas, de 01 a 30. Em seguida, faz-se a retirada aleatória e sem reposição de três bolas de uma urna que contém 30 bolas numeradas de 01 a 30 e verifica-se a coincidência entre as dezenas marcadas na cartela com aquelas nas bolas numeradas. O apostador será vitorioso se houver coincidência integral entre as três dezenas marcadas na cartela e as das bolas numeradas, independentemente da ordem de retirada das bolas. Com base nessas informações, julgue os itens de 61 a 66 acerca de conceitos de probabilidade.

- 61 A probabilidade de o apostador ter marcado duas das três dezenas nas bolas retiradas da urna é, pelo menos, 40 vezes a de ter marcado as três dezenas nas bolas retiradas.
- 62 A probabilidade de as dezenas nas bolas retiradas da urna serem 01 ou 30 é igual a $\frac{1}{15}$.
- 63 Com base em uma única cartela, a probabilidade de o apostador ser vitorioso é superior a 0,001.
- 64 Suponha que a dezena na primeira bola retirada seja 05. Nesse caso, a probabilidade de 06 ser a dezena na segunda bola retirada é igual a $\frac{1}{30}$.

- 65 Suponha que uma aposta — escolher e marcar três dezenas na cartela — custe R\$ 0,50 e que uma pessoa deseje apostar cada uma das combinações possíveis com base em cinco dezenas previamente escolhidas por ela. Nesse caso, o valor total das apostas será igual a R\$ 5,00.
- 66 A probabilidade de se retirarem da urna, em sequência, as bolas com as dezenas 01, 02, 03 é inferior à de se retirarem 03, 10, 18, também em sequência.

RASCUNHO

RASCUNHO

O setor de recursos humanos de uma instituição deseja avaliar a efetividade de um programa de treinamento que visa ao aumento da produtividade de seus empregados. Para essa avaliação, 30 empregados foram selecionados ao acaso para um estudo-piloto. As produtividades de cada empregado foram registradas, antes (X) e depois (Y) do treinamento. Com base nessas informações, julgue os itens a seguir.

- 67 O intervalo de 95 % de confiança para a variância populacional é da forma $s^2 \pm 1,96 \sqrt{\text{var}(s^2)}$.
- 68 A mediana amostral é uma estatística mais apropriada que a média amostral para representar a tendência central de uma distribuição amostral. Nesse caso, a distribuição amostral da mediana é igual à distribuição amostral da média aritmética.
- 69 Suponha que se deseje testar a normalidade da variável X, em que os parâmetros dessa distribuição sejam desconhecidos, mediante aplicação do teste qui-quadrado. Sabendo-se que o teste é válido, a estatística desse teste possui $k - 3$ graus de liberdade, em que k é o número de intervalos de classes.
- 70 Considere um teste qui-quadrado, em que se deseje testar se a variável X segue determinada distribuição hipotética W. Nessa situação, define-se a hipótese nula como aquela em que X, supostamente, não segue a distribuição W.
- 71 Para se avaliar a efetividade do treinamento, o teste estatístico recomendado é do tipo pareado.
- 72 De modo geral, o teste estatístico efetuado com base nos escores das produtividades dos empregados permite avaliar melhor a efetividade do treinamento em comparação com o teste dos sinais.
- 73 Suponha que as variáveis X e Y sejam independentes e que $X \sim N(14, 25)$ e $Y \sim N(16, 9)$. Nessa situação, a variância da diferença $X - Y$ será igual a 34.
- 74 Se $X \sim N(14, 25)$ e se \bar{X} é a produtividade média dos empregados antes do treinamento, então é correto afirmar que $\bar{X} \sim N(14, 25)$.
- 75 Suponha que X e Y sejam variáveis contínuas não gaussianas. Nesse caso, o teste não paramétrico de McNemar é apropriado para se testar a efetividade do treinamento para o aumento da produtividade dos empregados.

Um instituto de pesquisa deseja avaliar o efeito da redução do imposto sobre produtos industrializados (IPI) para veículos automotores na venda de veículos novos. Esse instituto obteve as seguintes estatísticas descritivas acerca do volume vendido, antes (X1) e depois (X2) da redução do IPI:

X1: $n_1 = 31$ $\bar{x}_1 = 90$ $s_1^2 = 12$; X2: $n_2 = 28$ $\bar{x}_2 = 115$ $s_2^2 = 9$,

em que n , \bar{x} , e s^2 correspondem, respectivamente, ao tamanho da amostra, à média aritmética e à variância amostral. Com base nessas informações, julgue os itens a seguir.

- 76** Sob a hipótese de igualdade das variâncias ($\sigma_1^2 = \sigma_2^2$), a estimativa da variância combinada será inferior a 10,5.
- 77** Considere a hipótese nula $\mu_1 = \mu_2$, que a estatística do teste t tenha sido igual a 29 e que as distribuições dos dados tenham sido normais. Nessa situação, há fortes evidências de que a redução do IPI produziu efeito significativo no volume de vendas de veículos novos.
- 78** Para se avaliar a tendência de aumento no volume das vendas com base no teste t , utiliza-se o teste bilateral, e, nesse caso, o P -valor do teste bilateral deverá ser calculado multiplicando-se por 2 o P -valor do teste unilateral correspondente.
- 79** A estatística do teste F para se testar a hipótese de igualdade das variâncias populacionais ($\sigma_1^2 = \sigma_2^2$) é superior a 1,2 ou inferior a 0,8.
- 80** Sob a hipótese de igualdade das variâncias populacionais ($\sigma_1^2 = \sigma_2^2$), o teste t para a hipótese (nula) de igualdade das médias populacionais ($\mu_1 = \mu_2$) possui 57 graus de liberdade.
- 81** O fato de a variância amostral da variável X2 ser menor que aquela observada para X1 deve-se exclusivamente ao fato de o tamanho da amostra de X2 ser inferior ao de X1.
- 82** Se os dados seguirem uma distribuição normal, o teste t é preferível ao teste não paramétrico de Wilcoxon.
- 83** Considere o teste de Wilcoxon e o teste t para a comparação das médias antes e depois da redução do IPI e suponha que os dados não sigam uma distribuição normal. Considere, ainda, que o P -valor do teste de Wilcoxon tenha sido igual a 0,071 e o do teste t , igual a 0,010. Nesse caso, considerando-se o nível de significância de 5%, pode-se concluir que há diferença entre os números médios de veículos vendidos antes e depois da redução do IPI.
- 84** Se o instituto utilizar os valores das variâncias populacionais, a estatística do teste de igualdade das médias terá 57 graus de liberdade.
- 85** Se a hipótese alternativa for $\mu_1 \leq \mu_2$ e se o numerador do teste for calculado com base na diferença $\bar{x}_1 - \bar{x}_2$, então os mesmos resultados serão obtidos alterando-se a hipótese alternativa para $\mu_1 \geq \mu_2$ ou $\mu_1 \neq \mu_2$.

RASCUNHO

Julgue os seguintes itens, acerca de modelos lineares.

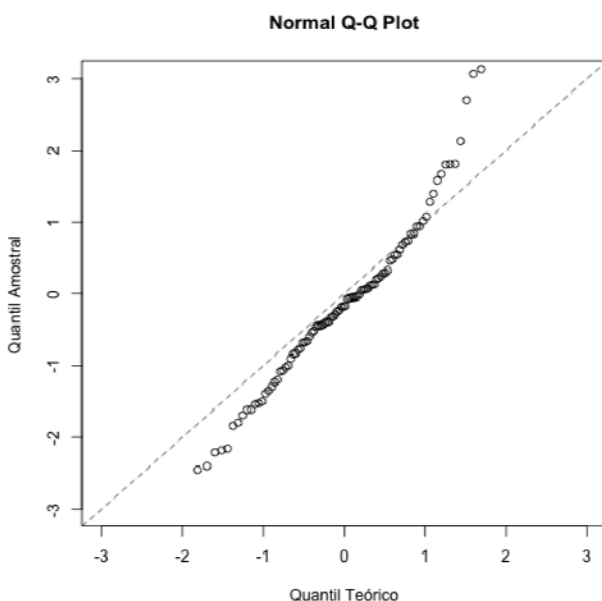
RASCUNHO

- 86 A presença de multicolinearidade na matriz de dados X pode tornar singular a matriz $X'X$.
- 87 Sob a hipótese de normalidade no modelo de regressão linear, o estimador de mínimos quadrados ordinários coincide com o estimador de máxima verossimilhança.
- 88 Do ponto de vista da análise de regressão, o modelo $\log y_i = \beta_0 + \beta_1 \log x_i + \log \epsilon_i$, $\log \epsilon_i \sim N(0, \sigma^2)$ não é considerado linear.
- 89 Suponha que, após o ajuste de um conjunto de dados por regressão linear simples, em que Y representa a variável resposta e X representa a variável regressora, foram obtidos os seguintes resultados: $R^2 = 0,81$; $\sigma_Y^2 = \text{var}(Y) = 1,00$; $\sigma_X^2 = \text{var}(X) = 4,00$ e $\rho_{XY} = \text{corr}(Y, X) < 0$. A partir desses dados, é correto concluir que a estimativa do coeficiente angular é um valor que se encontra entre $-0,250$ e $-0,120$.
- 90 Em um modelo de regressão linear simples, em que β_1 representa o intercepto do modelo, as hipóteses $H_0: \beta_1 = 0$; $H_1: \beta_1 \neq 0$ podem ser testadas por meio de uma tabela de análises de variâncias.
- 91 Considere um teste cujas hipóteses sejam $H_0: c\beta = 0$ e $H_1: c\beta \neq 0$, em que c é uma matriz de contrastes. Supondo-se que a matriz de covariância para as estimativas dos parâmetros (b) seja dada por $s^2(b) = \text{QMR}(X'X)^{-1}$, é correto afirmar que a matriz de covariância usada para testar os contrastes será igual a $s^2(c) = \text{QMR}(c'X'Xc)^{-1}$, em que X é a matriz de dados e QMR é o quadrado médio residual.
- 92 Relativamente à estimação dos coeficientes de um modelo de regressão linear múltipla pelo método dos mínimos quadrados ordinários, é correto afirmar que se X for a matriz de dados e Y o vetor de respostas, no produto matricial $H = (X'X)^{-1}X'$, denominado matriz *hat*, o número de colunas será igual ao número de linhas do vetor Y .
- 93 Considere que, em um modelo de regressão linear simples, a matriz de covariância para as estimativas dos parâmetros seja dada por $s^2(b) = \begin{bmatrix} 2,5 & 0,6 \\ 0,6 & 1,2 \end{bmatrix}$. Supondo-se que se deseje prever o valor da variável resposta concernente aos dados $x'_h = [1,0 \ 1,5]$, é correto afirmar que a variância do valor predito será igual a
- $$s^2_{\text{pred}} = [1,0 \ 1,5] \begin{bmatrix} 2,5 & 0,6 \\ 0,6 & 1,2 \end{bmatrix} \begin{bmatrix} 1,0 \\ 1,5 \end{bmatrix} = 3,4 + 3,6 = 7,0.$$

Ainda com relação aos modelos lineares, julgue os itens subsequentes.

RASCUNHO

- 94 Considere que um pesquisador tenha ajustado um modelo de regressão linear simples com base em amostra com 10 observações, tendo observado que $R^2 = 0,80$ e $\text{var}(Y) = 2,50$, em que Y é a variável resposta. Nesse caso, se modelos com estatística $F \geq 5,32$ forem considerados bem ajustados, é correto afirmar que o referido modelo não apresentou um bom ajuste.
- 95 O gráfico resíduos padronizados da regressão (no eixo das ordenadas) *versus* valores observados da variável resposta (no eixo das abscissas) permite verificar a suposição de normalidade residual.
- 96 Sabendo-se que a análise de variância (ANOVA) permite testar a igualdade entre médias de grupos que constituem determinado conjunto de dados, e sendo as médias entre os grupos idênticas, é correto afirmar que a maior parte da variabilidade seja devida à aleatoriedade dos dados.
- 97 Em um modelo de análise de variância com um fator em 3 níveis e com um fator em 4 níveis, é correto afirmar que é igual a 13 o menor tamanho amostral que permite o ajuste desse modelo.
- 98 Suponha que uma variável resposta Y esteja em distribuição simétrica de cauda pesada. Nesse caso, ao se ajustar um modelo de regressão simples, o gráfico normal Q-Q Plot referente a essa variável apresentará padrão semelhante ao ilustrado na figura abaixo.



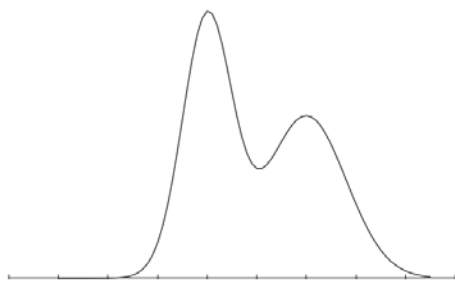
RASCUNHO

Um pesquisador buscou a opinião de clientes de uma agência bancária acerca do tempo de atendimento e efetuou uma amostragem por cotas. Nessa agência, circulam diariamente 5.000 clientes, e o tamanho amostral foi calculado fixando-se uma margem de erro (ξ) em um plano de amostragem aleatória simples sem reposição (AASs). A respeito dessa pesquisa, julgue o seguinte item.

- 99 O erro da pesquisa realizada por cotas tende a ser diferente do valor ξ fixado no plano amostral por AASs, e, na amostragem por cotas, a magnitude do erro pode não ser facilmente calculada.

Com relação a métodos computacionais e geração de números aleatórios, julgue os itens que se seguem.

- 100 Sabe-se que o método da transformação inversa consiste em gerar uma realização u da distribuição uniforme no intervalo $[0, 1]$. Considere que a função de probabilidade acumulada da distribuição desejada X seja $F(x)$ e que uma realização de X possa ser obtida com base na transformação inversa $x = F^{-1}(u)$. Nesse caso, é correto afirmar que esse método é comumente utilizado para simular tanto variáveis aleatórias discretas quanto a distribuição normal.
- 101 Considere que um pesquisador se proponha a simular uma variável cuja função distribuição de probabilidade acumulada seja representada por $F(x)$, pela determinação do valor x , tal que $G(x) = F(x) - u = 0$, $u \sim U[0; 1]$. Nesse caso, é correto afirmar que esse método somente funcionará se $F(x)$ for estritamente crescente.
- 102 Considere que a função densidade de probabilidade de certa variável aleatória possua o gráfico ilustrado na figura abaixo. Nesse caso, se o algoritmo EM (*expectation-maximization*) for aplicado a essa distribuição, independentemente do lado da cauda de início, o algoritmo convergirá para o ponto de máximo global da distribuição.



- 103 Para a geração de realizações de duas variáveis X e Y , os amostrados de Gibbs consideram alternadamente as distribuições condicionais $X|Y=y$ e $Y|X=x$. Assim, é correto afirmar que, se X segue uma distribuição de Bernoulli com parâmetro Y e se Y segue uma distribuição Beta com parâmetros a e b , então a distribuição conjunta da amostra gerada pelo amostrador de Gibbs segue aproximadamente uma distribuição Beta com parâmetros $a + X$ e $b + 1 - X$.
- 104 Considere que uma sequência de números seja gerada de acordo com a seguinte fórmula: $X_{n+1} = (aX_n + b) \bmod w$. Nesse caso, o valor a deve ser escolhido de modo que se garanta um longo ciclo de números pseudoaleatórios, isto é, o valor a determina o tamanho do ciclo do algoritmo.

A respeito de séries temporais, julgue os itens a seguir.

RASCUNHO

- 105** O modelo ARIMA(3, 1, 2) é um filtro linear que permite descrever uma série temporal estacionária com período sazonal igual a 3.
- 106** A função de densidade espectral $f(\lambda)$ representa o espaço de estados de um processo estocástico no domínio de Fourier. Para um processo AR(1), é correto afirmar que essa função é expressa na forma $f(\lambda) = \sigma \times \{2\pi(1 - 2\phi\cos\lambda)\}^{-1}$, em que $|\lambda| \leq \pi$ e $|\phi| > 1$.
- 107** Considere o processo de médias móveis definido como $M_t(z) = (z+1)^{-1} \times [X_t + X_{t-1} + \dots + X_{t-z+1}]$, em que t é um número inteiro positivo e $\{X_t\}$ é um processo fracamente estacionário e não gaussiano. Nesse caso, é correto afirmar que, à medida que o denominador z aumenta, o processo $M_t(z)$ converge em distribuição para um processo gaussiano.
- 108** Supondo-se que $\{Y_t\}$ seja uma série temporal que segue um processo ARMA(p, q), em que $Y_t = X_t - X_{t-1}$, é correto afirmar que, para que Y_t seja estacionário, é necessário que X_t também o seja.

No que se refere a processos estocásticos, julgue os próximos itens.

- 109** Considere que, na fila do tipo $M/M/1/K$, o sistema seja finito e comporte até K elementos. Nesse caso, é correto afirmar que a probabilidade limite de haver N_t elementos no sistema no

$$\text{instante } t \text{ é dada por } p_n(t) = P(N_t = n) \rightarrow \left(\frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{K+1}} \right) \left(\frac{\lambda}{\mu} \right)^n,$$

em que λ é a taxa de chegadas por unidade de tempo de elementos na fila e μ é a taxa de atendimentos por unidade de tempo, e que tal probabilidade para a fila $M/M/1$ é obtida no limite $K \rightarrow \infty$.

- 110** Considere um processo de Poisson em que N_t representa a quantidade de ocorrências registradas até o instante t , de modo que $P(N_t = n) = (n!)^{-1} \times e^{-\lambda t} (\lambda t)^n$. Considere, ainda, que a probabilidade de transição do estado i para o estado j seja dada por $p_{i,j}(t) = [(j - i)!]^{-1} \times e^{-\lambda t} (\lambda t)^{j - i}$. Nesse caso, se $p_{1,2}(t) = p_{1,3}(s)$ e se $s \rightarrow t$, então $\lambda > 2$.
- 111** Um processo gaussiano de Wiener, definido como $W_0 = 0$; $W_t - W_s \sim N(0; t - s)$, é estacionário e homocedástico.
- 112** Considere que um processo estocástico seja gerado com base no modelo $Z_0 = 0$; $Z_1 = 0$; $Z_{n+1} = Z_n + Z_{n-1} + X_n$, em que X_1, X_2, \dots sejam variáveis aleatórias de Bernoulli, independentes, com parâmetro p . Nesse caso, o processo Z_n será de Markov se, e somente se, $p = 0,5$.
- 113** Suponha que, em um processo de Poisson $\{N_t; t \geq 0\}$, a probabilidade de não se registrar uma ocorrência até o instante t seja $P(N_t = 0) = e^{-\lambda t}$. Nesse caso, se T_k representa o tempo para o registro da k -ésima ocorrência, é correto afirmar que $P(T_k > t) > P(T_k > t + s | T_k \geq s)$.
- 114** Em uma fila do tipo $M/M/1$, são exponenciais as distribuições dos tempos entre chegadas e dos tempos de atendimento.

Julgue os seguintes itens, acerca de análise multivariada de dados.

RASCUNHO

115 Considere a aplicação de uma análise fatorial $\Sigma = LL' + \Psi$ sob

a matriz de covariâncias $\Sigma = \sigma^2 \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$. Nesse caso, para

que exista solução para a análise fatorial com um fator, é necessário que $ab < c$ com $c \neq 0$.

116 Na análise discriminante para populações normais, para se avaliar a igualdade entre as matrizes de covariâncias populacionais entre diferentes grupos, o escore quadrático de classificação corresponde ao escore linear de classificação.

117 Com relação ao emprego do método de Fisher para análise discriminante, considerando-se que \mathbf{B} seja a matriz de covariâncias entre as populações e que \mathbf{W} represente a matriz de covariâncias dentro da população, é correto afirmar que a função discriminante será obtida com base no produto $\mathbf{W} \times \mathbf{B}$.

118 Se um vetor aleatório segue uma distribuição normal multivariada de dimensão p , então é correto afirmar que o quadrado da distância de Mahalanobis segue uma distribuição t de Student com $p - 1$ graus de liberdade.

119 Considere a aplicação da técnica das componentes principais com o objetivo de reduzir a dimensão de um conjunto de dados constituído de p variáveis. Considere, ainda, supondo que os autovalores da matriz das correlações entre essas variáveis sejam tais que $\lambda_2 = \lambda_3 = \dots = \lambda_p = 1 - \rho$, em que ρ representa uma medida de correlação. Nessa situação, considerando-se $p = 10$, é correto afirmar que $\rho < 0,75$.

120 A tabela a seguir mostra os resultados de uma pesquisa de mercado, por meio da qual foram obtidas informações acerca de quatro produtos comercializados na praça, em que $P_k = 1$ se o entrevistado X possui o item k e $P_k = 0$ se X não possui o item k . Considerando-se que a métrica de similaridade seja $(A+B)/4$, em que A indica a contagem de resultados 1 – 1 e B representa a contagem de resultados 0 – 0, é correto afirmar que o primeiro par de entrevistados a ser formado é o (2;5) ou (1;3).

X	P ₁	P ₂	P ₃	P ₄
1	0	0	0	1
2	1	1	1	0
3	0	1	0	1
4	0	0	1	0
5	1	1	1	0