MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PAMPA – UNIPAMPA

CONCURSO PÚBLICO
NÍVEL SUPERIOR

MANHÃ

CADERNO DE PROVA
PARTE II
CONHECIMENTOS ESPECÍFICOS

CARGO 13:
Estatístico

ATENÇÃO!

Leia atentamente as instruções constantes na capa da Parte I do seu caderno de prova.

1 Nesta Parte II do seu caderno de prova, confira incialmente se os seus dados pessoais e se os dados identificadores do seu cargo transcritos acima coincidem com o que está registrado em sua **folha de respostas** e em cada página numerada desta Parte II do seu caderno. Caso o caderno esteja incompleto, tenha qualquer defeito, ou apresente divergência quanto aos seus dados pessoais ou aos dados identificadores do seu cargo, solicite ao fiscal de sala mais próximo que tome as providências cabíveis, pois não serão aceitas reclamações posteriores nesse sentido.

2 Quando autorizado pelo chefe de sala, no momento da identificação, escreva, no espaço apropriado da **folha de respostas,** com a sua caligrafia usual, a seguinte frase:

O pior acontece quando a ciência é considerada uma forma de arte.

OBSERVAÇÕES

- Não serão objeto de conhecimento recursos em desacordo com o estabelecido em edital.
- Informações adicionais: telefone 0(XX) 61 3448-0100; Internet — www.cespe.unb.br.
-

cespeUnB
Centro de Seleção e de Promoção de Eventos

CONHECIMENTOS ESPECÍFICOS

Uma universidade oferece um curso para capacitação profissional de jovens carentes. Ao final do curso, cada jovem participante será avaliado por meio de uma prova teórica e de uma prova prática, cujos resultados possíveis são apresentados na tabela seguinte.

prova	resultados possíveis
teórica	satisfatório (S) insatisfatório (I)
prática	bom (B) médio (M) fraco (F)

Considerando que a probabilidade de um jovem obter resultado bom na prova prática seja P(B) = 0,50; a probabilidade de um jovem obter resultado fraco na prova prática seja P(F) = 0,05; a probabilidade de um jovem obter resultado satisfatório na prova teórica seja P(S) = 0,80; e as probabilidades condicionais sejam $P(S|B) = 0{,}80$ e $P(S|F) = 0$, julgue os itens a seguir.

51 Os resultados da prova teórica são independentes dos resultados da prova prática.

52 A união $S \cup I \cup B \cup M$ é igual ao espaço amostral.

53 A probabilidade condicional $P(S|M \cup F)$ é igual a 0,20.

54 Se um jovem obteve resultado M na prova prática, então a probabilidade de ele obter resultado S na prova teórica é igual a 0,20.

55 $P(I \cap B) = P(I)\,P(B)$.

O tempo de espera (T, em minutos) em uma fila para atendimento em um hospital universitário segue uma distribuição exponencial cuja função de densidade de probabilidade é dada por $f(t) = 0{,}5e^{-0{,}5t}$, em que $t > 0$. Considerando essas informações e que $e^{-1} = 0{,}37$, julgue os itens subsequentes.

56 O tempo médio de espera na fila em questão é superior a 1 minuto.

57 O desvio padrão da distribuição dos tempos de espera na fila é inferior a 1 minuto ou é superior a 3 minutos.

58 A probabilidade $P(T > 4)$ é igual a $P(T \geq 4)$ e é inferior a 0,15.

59 A mediana da distribuição dos tempos de espera na fila é inferior a 2 minutos.

60 A taxa de atendimento no referido hospital é de 0,5 pessoa/minuto.

61 Em algum minuto, a probabilidade de a fila estar vazia é inferior a 0,6.

Um indicador de desempenho acadêmico X é uma variável aleatória cuja função de distribuição acumulada tem a forma apresentada a seguir.

$$F(x) = \begin{cases} 0 & \text{se, } x < 0; \\ 0{,}20 & \text{se, } 0 \leq x < 1; \\ 0{,}30 & \text{se, } 1 \leq x < 2; \\ 0{,}50 & \text{se, } 2 \leq x < 3; \\ 1 & \text{se, } x \geq 3. \end{cases}$$

Considerando esse indicador, julgue os próximos itens.

62 A variável aleatória X não é discreta.

63 A probabilidade $P(X = 0)$ é nula.

64 $P(X \geq 5) = 1$.

65 O valor esperado da variável aleatória X é igual a 2.

66 A moda da distribuição de X é igual a 2.

67 $P(X = 0) + P(X = 1) + P(X = 2) = 1$.

68 A variância de X é superior a 1.

RASCUNHO

Considerando que $\{Z_1, Z_2,..., Z_n\}$ seja um conjunto de variáveis aleatórias contínuas independentes e que a função geratriz de momentos da variável aleatória Z_k seja $M_k(q) = \dfrac{e^q - 1}{q}$, em que $k = 1, 2,..., n$ e q é um número real, e que $S = \sum_{k=1}^{n} Z_k$, julgue os itens que se seguem.

69 A média de Z_k é nula.

70 A variância de S é igual a $\dfrac{n}{12}$.

71 As variáveis aleatórias $Z_1, Z_2, ..., Z_n$ são identicamente distribuídas.

72 A mediana de Z_k é superior a 0,3 e é inferior a 0,7.

73 A função geratriz de momentos da soma S é igual a $\left(\dfrac{e^q - 1}{q}\right)^n$.

74 Assintoticamente, a função geratriz de momentos da variável aleatória $\dfrac{S}{n}$ é igual a $\dfrac{e^q - 1}{q}$.

Um pesquisador deseja comparar dois instrumentos de avaliação que serão utilizados para avaliar as habilidades pessoais dos estudantes. Para isso, foram selecionados aleatoriamente 50 estudantes de certa escola, dispostos nos grupos A e B. Para os 32 estudantes do grupo A foi aplicado o instrumento de avaliação A; para os 18 estudantes do grupo B aplicou-se o instrumento B. Cada estudante obteve uma pontuação, e um resumo dos resultados (média e desvio padrão) encontra-se na tabela abaixo.

grupos de estudantes	números de estudantes	média aritmética	desvio padrão amostral
A	32	7,8	0,8
B	18	7,4	0,6

As pontuações produzidas pelos instrumentos A e B têm distribuições normais com variâncias populacionais diferentes, e o pesquisador deseja efetuar o seguinte teste de hipóteses: H_0: $\mu_A \leq \mu_B$ *versus* H_1: $\mu_A > \mu_B$, em que μ_A e μ_B são as médias populacionais das distribuições das pontuações nos grupos A e B, respectivamente.

Considerando as informações acima, e os valores aproximados $\Phi(2,0) = 0,98$ e $\Phi(3,0) = 0,99$, em que $\Phi(z)$ representa a função de distribuição acumulada da distribuição normal padrão, julgue os itens subsequentes.

75 O teste de hipótese pode ser escrito como H_0: as distribuições das pontuações nos grupos A e B são identicamente distribuídas *versus* H_1: as distribuições das pontuações nos grupos A e B não são identicamente distribuídas.

76 Como as pontuações produzidas pelos instrumentos A e B são normais, é correto concluir que a distribuição amostral da estatística do teste em questão é normal.

77 A estatística do teste é superior a 1,96.

78 O nível descritivo do teste é superior a 2%.

79 Se o nível de significância do teste for igual a 1%, então a hipótese nula H_0 será rejeitada.

80 Os valores dos desvios padrão apresentados na tabela são estimativas não tendenciosas dos respectivos desvios padrão populacionais.

RASCUNHO

Em um estudo realizado por uma instituição de ensino superior foram selecionados aleatoriamente 5 estudantes de graduação. Cada estudante foi submetido a duas provas de conhecimentos gerais. A primeira prova ocorreu no momento do ingresso do estudante na universidade e a segunda, no egresso. Os resultados dessas provas são apresentados na tabela abaixo.

estudante	notas na primeira prova (A)	notas na segunda prova (B)
1	4	6
2	4	8
3	2	5
4	6	8
5	5	9

Considerando que μ_A e μ_B são as notas médias populacionais e que as notas seguem distribuições normais, julgue os itens a seguir.

81 O delineamento do estudo é característico de uma amostragem aleatória estratificada, em que as notas da primeira prova formam o primeiro estrato e as notas da segunda prova formam o segundo estrato.

82 Se μ_A e μ_B são parâmetros desconhecidos, a variância do estimador da diferença $\mu_B - \mu_A$ é igual a 0,20.

83 O teste t de Student, com 4 graus de liberdade, pode ser aplicado para se testar a hipótese nula H_0: $\mu_A - \mu_B = 0$ contra a hipótese alternativa H_1: $\mu_A - \mu_B \neq 0$.

84 Se μ_A for um parâmetro conhecido e igual a 4,2, então o estimador de regressão para μ_B será igual a 7,2.

85 Se μ_A for um parâmetro conhecido e igual a 5, então o estimador de razão para μ_B será superior a 8.

RASCUNHO

Um auditor foi encarregado de fazer uma avaliação dos erros cometidos em preços nas faturas emitidas por uma empresa. Utilizando métodos de amostragem que garantem a representatividade da amostra, ele registrou o número de erros em preços de 60 faturas amostradas, conforme apresentado a seguir.

0 0 2 1 4 1 0 1 3 2 2 0

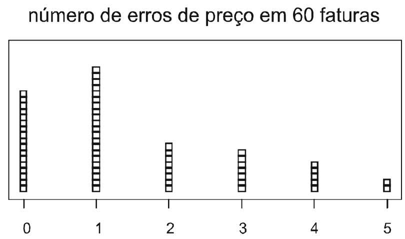
1 1 1 4 0 3 1 5 1 1 0 2

0 0 1 1 4 3 0 1 0 2 1 4

3 1 0 0 5 1 2 0 3 0 2 1

1 3 1 4 3 0 2 0 1 1 0 1

A partir desses dados, o auditor construiu o gráfico abaixo.



O auditor também realizou um teste de hipóteses e construiu um intervalo de confiança para a o número médio de erros, utilizando um *software* estatístico, que forneceu a saída a seguir, na qual os sinais ???? significam que a informação não estava disponível.

```
> t.test(erros)
 One Sample t-test
data: erros
t = 8.0698, df = 59, p-value = ??????
95 percent confidence interval:
 (1.102992; ??????)
sample estimates:
mean of x
 1.466667
```

Com base nos dados fornecidos e no gráfico acima e considerando que $\overline{X}$ seja a média dos dados, julgue os itens de **86** a **95**.

86 O gráfico apresentado é do tipo diagrama de pontos *dotplot* e permite visualizar a dispersão dos dados quando se tem poucas observações.

87 O número mediano de erros nos preços é superior a 1.

88 O gráfico apresenta assimetria à direita, na qual 85% das faturas apresentaram número de erros iguais ou inferiores a 3.

89 O coeficiente de variação é superior a 90%, e a média aparada a 10% é igual a 1.

90 Pelo tipo da variável que o auditor registrou e com base no gráfico apresentado, a melhor distribuição probabilística que descreve o comportamento aleatório do número de erros é a distribuição geométrica, cuja função de probabilidade é dada por $f(x) = p(1-p)^{x-1} I_{x \in \{1,2,\ldots\}}(x)$, em que p é a probabilidade de ocorrer um erro de preço em uma fatura inspecionada e x é o número de erros encontrados.

91 O estimador de máxima verossimilhança para a probabilidade de sucesso de uma distribuição geométrica é $\hat{p} = [\bar{x}]^{-1}$.

92 O teorema central do limite garante que, independentemente da distribuição dos dados, uma vez aplicada a padronização $z = \dfrac{x - \bar{x}}{s}$, em que s é o desvio padrão amostral, a distribuição dos dados amostrais passa a ter distribuição normal com média 0 e desvio padrão 1.

93 Considerando uma amostra aleatória $X_1, X_2, \ldots, X_n$ com média μ e variância igual a σ^2, então $Var(\bar{X}) = \dfrac{\sigma^2}{n-1}$.

94 A hipótese nula testada é H_0: $\mu = 0$, sendo μ a média populacional.

95 O limite superior do intervalo de 95% de confiança para o número médio de erros é inferior a 2, e a probabilidade de significância é inferior a 0,1%.

Considere uma amostra aleatória de tamanho n para a variável aleatória X, de esperança matemática $E(X) = \mu$ e variância finita $Var(X) = \sigma^2$. Considere γ um parâmetro desconhecido da distribuição de probabilidade de X e $\Gamma = h(X_1, X_2, \ldots, X_n)$ uma estatística que visa estimar γ. A partir dessas informações, julgue os seguintes itens.

96 O erro quadrático médio para um estimador não viciado é dado por $E[(\Gamma - \gamma)^2] = Var(\Gamma)$.

97 O erro quadrático médio de qualquer estimador é obtido por $Var(\Gamma) + \frac{1}{n}[B(\Gamma)]^2$, em que $B(\Gamma) = E(\Gamma) - \gamma$ é o vício do estimador.

98 Se $\lim_{n \to \infty} P(|\Gamma - \gamma| \geq \varepsilon) = 0$ para qualquer $\varepsilon \geq 0$, então Γ é um estimador consistente para γ.

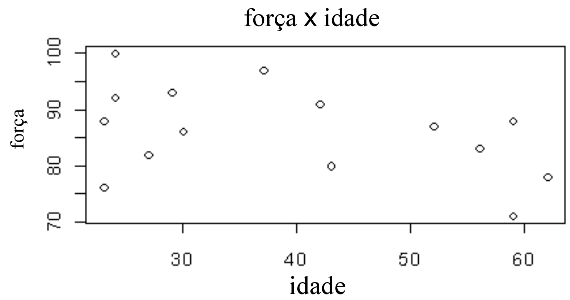
99 Um intervalo de $(1 - \alpha)100\%$ de confiança para Γ pode ser construído a partir da desigualdade $P(L_1 \leq \gamma \leq L_2) \geq 1 - \alpha$. A interpretação do intervalo $[L_1, L_2]$ pode ser expressa da forma: tem-se, pelo menos, $(1 - \alpha)100\%$ de confiança que o parâmetro γ esteja contido dentro do intervalo $[L_1, L_2]$.

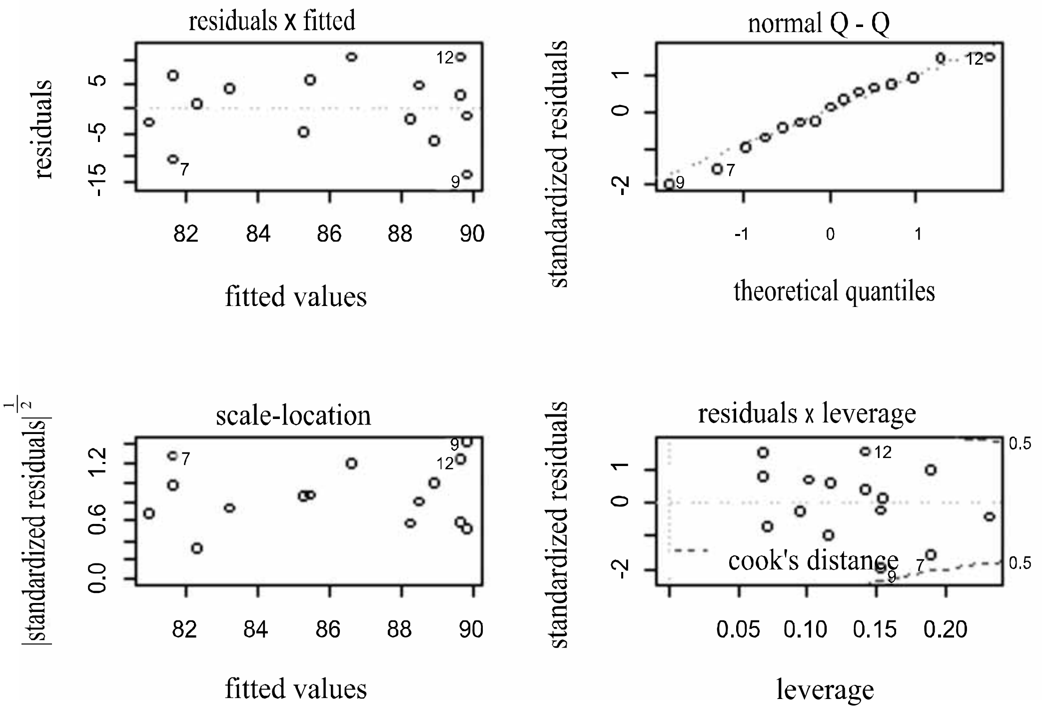
100 A distribuição probabilística de $X_1, X_2, \ldots, X_n$ é chamada distribuição amostral.

RASCUNHO

Em um estudo da relação entre a idade (X) e a condição física de adultos, mediu-se a força (Y) da mão direita em um grupo de homens adultos. De uma amostra de 15 voluntários, foram obtidos as estatísticas e os gráficos apresentados a seguir.

$$\sum_{i=1}^{15} x_i = \mathbf{590}; \quad \sum_{i=1}^{15} y_i = 1.292; \quad \sum_{i=1}^{15} x_i^2 = 26.308; \quad \sum_{i=1}^{15} y_i^2 = 112.170; \quad \sum_{i=1}^{15} x_i y_i = 50.111$$





Um *software* de análise estatística, utilizado para o processamento dos dados coletados, apresentou a seguinte saída.

```
Coefficients:
             Estimate Std. Error t value Pr(>t)
(Intercept) ???????       5.6130 16.944 3.04e-10 ***
Idade       -0.2282       0.1340
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: ????? on 13 degrees of freedom
Multiple R-squared: 0.1823, Adjusted R-squared: 0.1194
F-statistic: 2.898 on 1 and 13 DF, p-value: 0.1124
```

Com base nessas informações e nos gráficos apresentados, julgue os itens de **101** a **110**.

101 As equações normais para um modelo de regressão linear simples são mostradas abaixo, em que *SQE* é a soma dos quadrados dos erros.

$$\frac{\partial(SQE)}{\partial\alpha} = -2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i$$

$$\frac{\partial(SQE)}{\partial\beta} = -2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i^2,$$

102 Se o percentil superior 5% da distribuição *t* for igual a 1,771, então não há evidências amostrais de que a idade da pessoa influencia na força exercida pela mão direita, ao nível de significância de 10%.

103 O valor do intercepto do modelo é um valor negativo.

104 O quadrado médio do erro é superior a 150.

105 A análise do coeficiente de determinação permite afirmar que 18,23% da variabilidade total é determinado pela variabilidade residual.

106 Os pontos 9 e 12 nos gráficos de análise de resíduos são denominados pontos de influência.

107 A estatística do teste para a hipótese H_0: $\beta = 0$ é $t = \dfrac{\hat{\beta} - 0}{\sqrt{\dfrac{s^2}{S_{XX}}}}$, sendo $\hat{\beta}$ a estimativa de β, s^2 a variância amostral e $S_{XX} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$.

108 Somente quando um coeficiente de correlação é estatisticamente significativo é possível estabelecer uma relação de causa e efeito entre duas variáveis estatisticamente correlacionadas.

109 O coeficiente de correlação entre os parâmetros idade e força está entre 0,5 e 0,6.

110 As conclusões com base no teste *t* para o coeficiente angular e o teste *F* da análise da variância do modelo de regressão são sempre concordantes.

Seja uma variável resposta Y e uma variável regressora X. Uma amostra de tamanho 10 foi coletada e as observações foram registradas a seguir.

X: 1 1 2 3 4 4 4 5 6 6
Y: 4 4 3 2 1 2 2 2 3 3

Considerando uma abordagem matricial para um modelo de regressão, $\mathbf{y} = \mathbf{x}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ sendo $\boldsymbol{\varepsilon}$ um vetor de erros aleatórios com distribuição normal com vetor de médias 0 e uma matriz de variâncias-covariâncias $I\sigma^2$, julgue os itens de **111** a **119**.

111 A matriz de informação observada é $\frac{1}{10}\begin{bmatrix} 1 & 3{,}6 \\ 3{,}6 & 16 \end{bmatrix}^{-1}$.

112 O estimador do vetor β é dado por $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{Xy}$.

113 A soma de quadrados total corrigida para a média, na tabela de análise da variância, pode ser escrita como $\mathbf{y'y} - 10^{-1}\mathbf{y'\,II'y}$.

114 Considerando um valor adicional $\mathbf{X}_0$, a predição do valor de $\mathbf{y}$ é dado por $\hat{\mathbf{y}}_0 = \mathbf{b'X}_0$ com variância $Var(\hat{\mathbf{y}}_0) = \mathbf{X}_0'(\mathbf{X'X'})^{-1}\mathbf{X}_0\sigma^2$.

115 O coeficiente de determinação está limitado aos valores 0 e 1 (ou 0% a 100%).

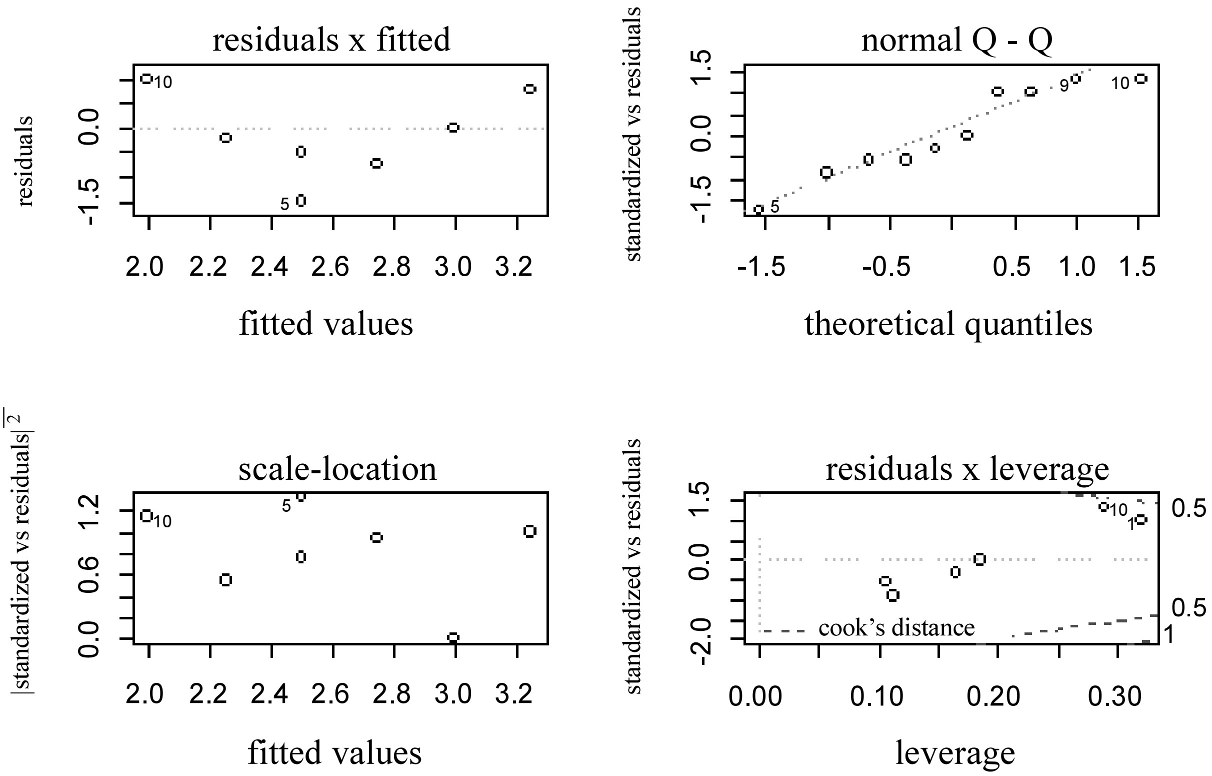
116 Em caso de heterogeneidade da variância, a matriz de variâncias-covariâncias pode ser modificada para $\mathbf{V} = \begin{bmatrix} \sigma_1^2 & & & \mathbf{0} \\ & \sigma_2^2 & & \\ & & \ddots & \\ \mathbf{0} & & & \sigma_{10}^2 \end{bmatrix}$, em um método de estimação sob dados heterocedásticos chamado de mínimos quadrados ponderados.

117 A cada aumento de uma unidade no regressor X há um aumento de 0,25 unidades em Y.

118 O teste de Mann-Whitney é usualmente utilizando para avaliar a presença de dependência temporal entre elementos do vetor ε, ou seja, avalia se a hipótese de independência dos erros está sendo atendida. Junto com esse teste, podem ser utilizadas técnicas gráficas, tais como: construção de gráficos sequenciais para os erros, gráficos da função de autocorrelação ou um diagrama de dispersão com os erros defasados em uma unidade, buscando correlação entre os pontos.

119 Na necessidade de se ajustarem modelos polinomiais de grau superior a 2 (modelos cúbicos, quárticos etc.), é aconselhável utilizar polinômios ortogonais que garantem a independência entre os termos dentro do preditor linear do modelo de regressão.

RASCUNHO



Considerando os gráficos acima, julgue o próximo item.

120 A partir dos gráficos de diagnósticos dos resíduos e influência, é correto afirmar que, apesar de poucas observações, não há evidências contra as suposições de normalidade e homogeneidade da variância; também não há evidências de forte influência de pontos extremos. Entretanto, há evidência de curvatura não acomodada pelo modelo de regressão linear simples, sugerindo, assim, um modelo com um termo quadrático no preditor linear.

RASCUNHO