

CONCURSO PÚBLICO

MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR
INSTITUTO NACIONAL DE METROLOGIA, NORMALIZAÇÃO
E QUALIDADE INDUSTRIAL (INMETRO)

CADERNO DE PROVAS PARTE II

PROVA DE CONHECIMENTOS ESPECÍFICOS
PROVA DISCURSIVA

CARGO

8

ANALISTA EXECUTIVO EM
METROLOGIA E QUALIDADE

ÁREA:

ESTATÍSTICA

ATENÇÃO!

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

- 1 Nesta parte II do seu caderno de provas, confira atentamente os seus dados pessoais e os dados identificadores de seu cargo transcritos acima com o que está registrado em sua **folha de respostas** e na capa de seu **caderno de texto definitivo da prova discursiva**. Confira também o seu nome e o nome do seu cargo no rodapé de cada página numerada desta parte II de seu caderno de provas. Caso o caderno esteja incompleto, tenha qualquer defeito, ou apresente divergência quanto aos seus dados pessoais ou aos dados identificadores de 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:

A alegria abre, a tristeza fecha o coração.

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.
- É permitida a reprodução deste material apenas para fins didáticos, desde que citada a fonte.

CONHECIMENTOS ESPECÍFICOS

A concentração X de determinada substância em um galão de água mineral é uma variável aleatória contínua cuja função de distribuição acumulada é dada por $F(x) = \begin{cases} 1 - \exp(-0,25x^2), & x \geq 0 \\ 0, & x < 0 \end{cases}$.

Nesse caso, considerando $\exp(-1) = 0,37$, julgue os itens subsequentes.

- 41 A probabilidade da concentração X dessa substância em um galão ser superior a $\sqrt{2}$ é inferior a 0,4.
- 42 A moda da distribuição X é igual a $\sqrt{2}$.
- 43 Se μ e v representam, respectivamente, a média e a mediana de X , então $\left(\frac{v}{\mu}\right)^2 = \frac{4 \times \ln 2}{\pi}$.
- 44 A variável aleatória $Y = \sqrt{X}$ segue uma distribuição exponencial.
- 45 O valor esperado e a variância da variável aleatória $W = \exp(-0,25X^2)$ são, respectivamente, iguais a $\frac{1}{2}$ e $\frac{1}{12}$.
- 46 Considere uma variável aleatória Y definida por $Y = X$. Nessa situação, $P(Y - X = 0) = 1$.
- 47 Considere que X_1 e X_2 sejam duas variáveis aleatórias independentes, identicamente distribuídas e que $F(x_1) = F(x_2) = F(x)$. Nessa situação, a função de densidade da soma $S = X_1 + X_2$ é igual a $f(s) = s \times \exp(-0,5s^2)$.

Considerando que $\{Z_t\}$, $t = 1, 2, 3, \dots$, seja uma sequência de ensaios de Bernoulli que forma uma cadeia de Markov com probabilidades de transição $P(Z_{t+1} = 0 \mid Z_t = 0) = 0,6$ e $P(Z_{t+1} = 1 \mid Z_t = 1) = 0,8$, julgue os itens a seguir.

- 48 A sequência $\{Z_t\}$ forma uma cadeia de Markov não irreduzível.
- 49 O processo estocástico $\{Z_t\}$ possui dois estados recorrentes.
- 50 $P(Z_{t+1} = 0 \mid Z_t = 0) = P(Z_{t+1} = 0 \mid Z_t = 1) + P(Z_{t+1} = 1 \mid Z_t = 0)$.
- 51 A matriz de transição de estados do processo $\{Z_t\}$ não é duplamente estocástica.
- 52 Considere que $Z_0 = 0$, isto é, inicialmente o processo encontra-se no estado 0, e que W é o número de ensaios de Bernoulli até que ocorra a primeira transição do estado 0 para o estado 1. Nessa situação, o valor esperado de W é maior ou igual a 1,5.
- 53 $P(Z_{t+6} = 1 \mid Z_t = 0, Z_{t+1} = 1, Z_{t+2} = 1, Z_{t+3} = 0, Z_{t+4} = 1, Z_{t+5} = 1) < 0,7$.
- 54 No limite estacionário do processo estocástico $\{Z_t\}$, tem-se que $\lim_{t \rightarrow \infty} P(Z_t = 0) = \lim_{t \rightarrow \infty} P(Z_t = 1) = 0,5$.
- 55 $0,42 < P(Z_{t+2} = 0 \mid Z_t = 0) < 0,46$.
- 56 Quando $t \rightarrow \infty$, a média e o desvio padrão de Z_t serão iguais.
- 57 A autorregressão de Z_{t+1} em Z_t pode ser escrita como $E(Z_{t+1} \mid Z_t = z) = 0,2 + 0,4z$, em que $z = 0$ ou $z = 1$.
- 58 No limite estacionário, a autocorrelação entre Z_{t+1} e Z_t é maior que 0,3 e menor que 0,6.

RASCUNHO

Um órgão de controle ambiental efetuou um estudo que

considerou a distribuição de um vetor aleatório $\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, em

que a matriz de covariâncias é $\Phi = \begin{bmatrix} 3 & 2,1 & 2,1 \\ 2,1 & 3 & 2,1 \\ 2,1 & 2,1 & 3 \end{bmatrix}$ e o vetor de

médias é $\boldsymbol{\mu} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$. Nessa situação, julgue os itens que se seguem.

- 59** A matriz Φ possui três autovalores e 7,2 é o maior deles.
- 60** A primeira componente principal explica 80% da variação total.
- 61** A correlação entre X_2 e X_3 é igual a 2,1.
- 62** A primeira componente principal é o produto interno $\mathbf{u}^T \cdot \mathbf{x}$, em que $\mathbf{u}^T = \frac{1}{\sqrt{3}}[1 \ 1 \ 1]$.
- 63** Considerando que o vetor aleatório \mathbf{x} siga uma distribuição normal multivariada, a média condicional de X_2 , dado que $X_1 = x$, é igual a $0,2 + 0,7x$.
- 64** Se Φ fosse uma matriz de covariância amostral, então o intervalo de confiança assintótico de $100 \times (1 - \alpha)\%$ para o menor autovalor da matriz de covariância populacional tem a forma $0,9 \pm z_{\alpha/2} \sqrt{\frac{1}{n}}$, em que $z_{\alpha/2}$ é o quantil correspondente da distribuição normal padrão e n é o tamanho da amostra.
- 65** A função $\sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \Phi^{-1}(\mathbf{x} - \boldsymbol{\mu})}$ é conhecida como distância de Mahalanobis.
- 66** A correlação parcial entre X_1 e X_2 é inferior a 0,35.
- 67** Se o vetor aleatório \mathbf{x} segue uma distribuição normal multivariada, então X_1 , X_2 e X_3 são normais univariadas com variâncias iguais a 3.
- 68** A distribuição da soma $X_1 + X_2 + X_3$ possui média igual a 9 e variância menor que 16.

O total mensal de reclamações recebidas em uma central de atendimento ao consumidor segue um processo na forma $Y_t = 0,6Y_{t-1} - 0,09Y_{t-2} + \varepsilon_t - 0,3\varepsilon_{t-1}$, em que $Y_t = R_t - \mu$, e R_t representa o total mensal de reclamações recebidas no mês t ; μ é o total mensal médio de reclamações e ε_t representa um ruído branco no mês t com média 0 e variância σ^2 . Com base nessas informações, julgue os itens que se seguem.

- 69** A série temporal $\{Y_t\}$ segue um processo autorregressivo de primeira ordem.
- 70** A autocorrelação entre Y_t e Y_{t-1} é igual a 0,6.
- 71** A autocorrelação parcial entre Y_t e Y_{t-2} é igual a zero.
- 72** A função de autocorrelação inversa entre Y_t e Y_{t-h} é igual a $\frac{3(-0,3)^h}{1 + 0,6^h + 0,09^{2h}}$, em que $h = 0, 1, 2, 3, \dots$
- 73** A densidade espectral do processo Y_t é igual a $f(\omega) = \frac{\sigma^2}{2\pi(1 + 0,6\cos\omega)}$, em que $0 \leq \omega \leq \pi$.

Uma amostra aleatória simples de tamanho $n = 4$ foi retirada para se estimar a quantidade de sódio em certo alimento, registrando-se os respectivos valores Y de sódio, em mg, encontrados em uma quantidade X , em mg, do alimento, conforme a tabela abaixo.

amostra i	y_i	x_i
1	2	40
2	3	50
3	3	50
4	4	60

As estatísticas de interesse são $r = \frac{\sum_{i=1}^4 y_i}{\sum_{i=1}^4 x_i}$ e $b = \frac{\sum_{i=1}^4 x_i y_i}{\sum_{i=1}^4 x_i^2}$, em que

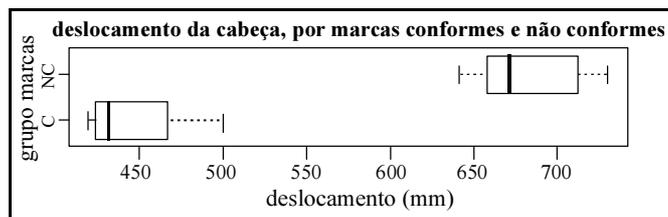
b é o estimador do coeficiente angular β da reta de regressão linear passando pela origem, $y_i = \beta x_i + \varepsilon_i$, em que os erros aleatórios $\varepsilon_1, \varepsilon_2, \varepsilon_3$ e ε_4 são independentes e identicamente distribuídos, com média zero e variância σ^2 . Com base nessas informações, julgue os itens subsequentes.

- 74 De acordo com a estatística r , a quantidade de sódio encontrada no alimento representa, em média, 6% da massa total desse alimento.
- 75 O uso da razão r tem sentido se houver uma relação linear da forma $y_i = \beta x_i + \varepsilon_i$, mesmo que a variância de y_i não seja constante, com aumento proporcional a x_i .
- 76 A estatística r não é viciada, pois seu valor esperado é igual a $\frac{\mu_y}{\mu_x}$, em que μ_x e μ_y são as médias populacionais de X e Y , respectivamente.
- 77 A estimativa da variância de r é um valor superior a 0,001.
- 78 Embora a reta ajustada $y = bx$ não passe pelos pontos médios amostrais $\frac{\sum_{i=1}^4 y_i}{n}$ e $\frac{\sum_{i=1}^4 x_i}{n}$, b é um estimador alternativo ao estimador de razão r . Além disso, b não é estimador tendencioso, pois seu valor esperado é igual a $\frac{\mu_y}{\mu_x}$, em que μ_x e μ_y são as médias populacionais de X e Y , respectivamente.
- 79 A estimativa da variância de b é superior a 0,003.
- 80 A estimativa da variância dos erros aleatórios σ^2 é inferior a 0,05.

Em um estudo sobre cadeiras infantis para automóveis, foram realizados diversos ensaios segundo uma adaptação da norma NBR 14.400, com o objetivo de avaliar a segurança na utilização do produto pelo usuário. Entre outros, foi realizado o ensaio de impacto, considerado o mais crítico, pois simula o deslocamento sofrido pela criança em frenagens bruscas e colisões e, portanto, avalia se o produto consegue reter a criança no assento e reduzir danos que podem ser fatais. A presença da criança na cadeira é simulada com a utilização de um boneco com todas as características antropométricas: peso, altura, articulações e pontos frágeis. A norma estabelece os seguintes limites para o deslocamento da cabeça e para a desaceleração máxima no tórax.

deslocamento máximo da cabeça (em mm)		desaceleração máxima no tórax (número vezes a força da gravidade)	
horizontal	vertical	horizontal	vertical
550	800	55	30

O analista responsável tomou as medidas de deslocamento horizontal da cabeça das 12 marcas que participaram do estudo, separadas nos grupos conformes (C) e não conformes (NC), construiu o seguinte gráfico e calculou as seguintes estatísticas.



marcas	média	desvio padrão	percentis					tam. amostra
			0%	25%	50%	75%	100%	
conformes	416,33	62,13	352	386,5	421	448,5	476	6
não conformes	675,33	24,21	658	661,5	665	684,0	703	6

A partir dessas informações, julgue os itens de **81** a **86**.

- 81** O gráfico, conhecido como *boxplot* ou diagrama de caixas, é construído utilizando-se as seguintes estatísticas: mínimo, 1.º quartil, média, 3.º quartil e máximo.
- 82** O gráfico e as estatísticas apresentadas não permitem concluir que as medidas de deslocamento horizontal provêm de uma população com distribuição moderadamente assimétrica à direita.
- 83** Supondo que as medidas de deslocamento horizontal para o grupo conformes seguem uma distribuição normal, a probabilidade de se obter valores inferiores a 352 mm é inferior a 1%.

- 84** No estudo da correlação entre deslocamento horizontal da cabeça e deslocamento vertical da cabeça, considerando o pequeno tamanho amostral e a falta de informação acerca da distribuição bivariada dessas medidas, o ideal é calcular o coeficiente de correlação de *Spearman*, com base nos postos das medidas das duas variáveis.
- 85** Considerando que, no teste da hipótese de que o coeficiente de correlação é igual a zero, o valor-P seja igual a 0,1719, é correto concluir que a probabilidade de que a correlação seja igual a zero é de 17,19%.
- 86** Um intervalo de 95% de confiança para o deslocamento médio horizontal das marcas não-conformes deve ser calculado utilizando-se a fórmula $\bar{x} \pm z_{\alpha/2} \hat{\sigma}$, em que \bar{x} é a média, $z_{\alpha/2}$ é o percentil da distribuição normal padronizada e $\hat{\sigma}$ é uma estimativa do desvio padrão das medidas.

RASCUNHO

Texto para os itens de 87 a 98

Uma indústria alimentícia possui uma fábrica que produz pacotes de farinha de trigo para utilização na culinária doméstica. Duas características físico-químicas associadas à qualidade da farinha são a umidade e a acidez graxa. A Portaria n.º 354/1996, do Ministério da Saúde, estabelece como padrão os seguintes limites:

- ▶ umidade: máximo de 15%;
- ▶ acidez graxa (base seca): máximo de 50%.

Para avaliar se a linha de produção de pacotes de farinha de trigo está atendendo a esses requisitos, um engenheiro de alimentos tomou amostras de alguns pacotes ao longo de um mês. Para a análise da umidade, ele tomou 3 pacotes por turno, em dois turnos diários, durante 6 dias por semana, totalizando 144 amostras. Para avaliar a acidez graxa, tomou um pacote por turno, considerando como amostra as duas medidas diárias. Assim, formaram-se 24 amostras para avaliar a acidez graxa. O engenheiro construiu os seguintes gráficos (ou cartas) de controle e histograma em que LIC, LSC e CL representam, respectivamente, os limites inferior e superior de controle e a linha central, e USL representa o limite máximo tolerado pela norma.

gráfico para médias de umidade

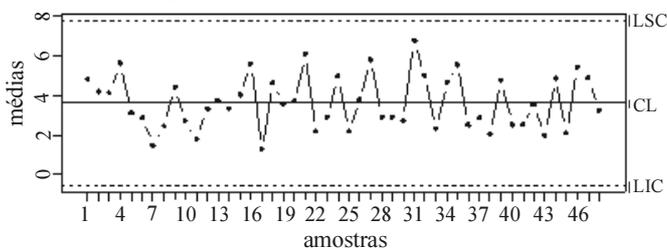
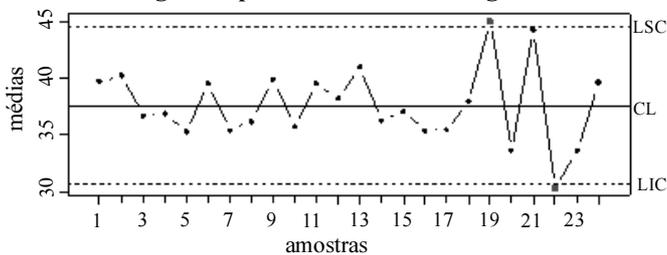
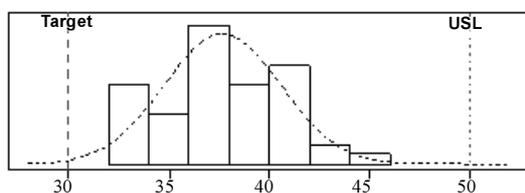


gráfico para médias de acidez graxa



análise de capacidade de processo para acidez graxa



As estatísticas calculadas sobre as medidas de umidade foram:

$$\sum \bar{X} = 174,2593; \sum R = 194,8813.$$

Algumas constantes utilizadas para o cálculo dos limites de controle dos gráficos são apresentadas a seguir.

n	A_2	A_3	c_4	c_4^{-1}	B_3	B_4	d_2	d_2^{-1}	D_4
2	1,880	2,659	0,7979	1,2533	0	3,267	1,128	0,8865	3,267
3	1,023	1,954	0,8862	1,1284	0	2,568	1,693	0,5907	2,575

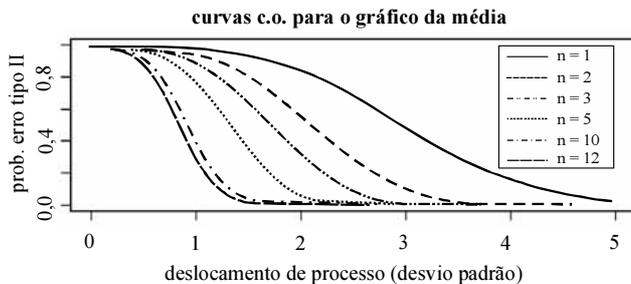
Considerando as informações apresentadas, julgue os itens a seguir.

- 87** Apesar de nenhum ponto se mostrar fora dos limites de controle, nota-se a alternância dos pontos entre as amostras 21 a 40, que sinalizam um padrão de processo fora de controle, sugerindo alguma mudança recorrente no processo, que resulta em instabilidade da medida de umidade em um mesmo dia.
- 88** A estimativa do desvio padrão da umidade deve ser calculada dividindo-se a média das amplitudes pela constante c_4 , que equivale a calcular o desvio padrão para todas as observações.
- 89** No gráfico de controle para a média de acidez graxa, verifica-se que o processo está fora de controle. A ação a ser tomada é verificar se ocorreu alguma mudança no processo que possa ter gerado os pontos fora dos limites. Uma vez bloqueada a causa atribuível, deve-se recalcular os limites de controle sem as observações associadas à causa atribuível.
- 90** Os índices de capacidade de processo são utilizados para avaliação do processo, uma vez que a distribuição das medidas de umidade não demonstra assimetria. Nesse caso, o índice de capacidade mais adequado é o C_{pm} , que mede a faixa da especificação utilizada.
- 91** Redefinir a formação dos subgrupos racionais por turno, considerando as medidas individuais, é mais adequado para se detectar variação entre amostras de acidez graxa.

RASCUNHO

Considere que as causas atribuíveis do processo fora de controle para a carta da acidez graxa foram bloqueadas, os dados associados a essas causas atribuíveis foram desconsiderados dos cálculos dos limites de controle e novas amostras foram coletadas, gerando as seguintes estatísticas de acidez graxa e o seguinte gráfico.

n.º de observações = 48	valor alvo = 30
média = 37,66461	USL = 50
desvio padrão = 2,912275	

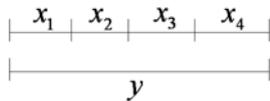


O gráfico, conhecido como curva característica de operação, possui, no eixo horizontal, o deslocamento da média do processo (na unidade de desvios padrão) e, no eixo vertical, a probabilidade do erro tipo-II; as curvas correspondem a um tamanho de amostra, conforme a legenda presente no gráfico.

Com base nessas informações, julgue os itens a seguir.

- 92 O índice C_{pu} é inferior a 1.
- 93 A probabilidade do erro tipo-II, nesse caso, está associada à ocorrência de alarmes falsos, ou seja, assume-se que o processo está fora de controle quando, na realidade, não está.
- 94 A média das amplitudes amostrais, quando multiplicada por uma constante adequada, permite obter uma estimativa não viciada da variância.
- 95 O comprimento médio de sequência (CMS) calculado, assumindo-se o tamanho amostral igual a 2 e um deslocamento de dois desvios padrão, é inferior a 4 amostras.
- 96 A faixa utilizada da especificação é o recíproco do índice C_p .
- 97 Considerando o tamanho amostral igual a 2, é aproximadamente igual a 40% a probabilidade de se detectar um deslocamento do processo de dois desvios padrão.
- 98 Se a probabilidade de não se detectar o processo fora de controle na primeira amostra é 0,8, então a probabilidade de se detectar o processo fora de controle na terceira amostra é inferior a 20%.

Considere que um estudo foi realizado utilizando-se peças mecânicas de precisão dentro de uma câmara climática sob determinada temperatura. Os técnicos sabem que essas peças foram produzidas em processos industriais de 4 linhas de produção diferentes, cujas medidas de comprimento x_1, \dots, x_4 apresentam distribuição normal com média de 2,0 cm e desvio padrão de 0,2 cm. A medida y representa o comprimento total da montagem, conforme a figura abaixo.



Com base nessa situação hipotética, julgue os itens que se seguem.

- 99** É superior a 5% a probabilidade de que o comprimento total da montagem seja inferior a 7,5 cm.
- 100** Considere que o pesquisador responsável por esse estudo, utilizando um computador, deseja simular medidas do comprimento total da montagem, a partir da simulação pseudoaleatória de uma amostra com distribuição normal padrão. Por meio dessas medidas, obtêm-se as medidas na escala do comprimento total da montagem. Supondo que ele obteve uma medida simulada a partir da distribuição normal padronizada igual a $-0,75$, então a medida simulada do comprimento de montagem é inferior a 7 cm.
- 101** O método de Box-Müller é computacionalmente eficiente para geração de números pseudoaleatórios com distribuição normal padronizada, a partir de dois números pseudoaleatórios com distribuição normal com média zero e variância igual a 1.
- 102** Se x_1, \dots, x_4 apresentassem, individualmente, uma distribuição exponencial com média 2,0 cm, então a distribuição do comprimento total da montagem seguiria uma distribuição gama $(4; 2,0)$.
- 103** O método de estimação por máxima verossimilhança garante a obtenção de estimadores não viciados, o que o tornou difundido em diversas áreas da ciência.
- 104** Caso não seja possível obter uma solução analítica na solução das equações envolvidas no método de estimação por máxima verossimilhança, deve-se utilizar métodos numéricos. Nessa situação, o método da quadratura de Gauss-Hermite é um dos mais eficientes.
- 105** O método de Newton para maximizar funções de log-verossimilhança consiste em calcular, iterativamente, a equação $\theta^{(t+1)} = \theta^{(t)} - \frac{l'(\theta^{(t)})}{l''(\theta^{(t)})}$, sendo $\theta^{(t)}$ e $\theta^{(t+1)}$ o parâmetro a ser estimado nos momentos t e $t + 1$, respectivamente; $l'(\cdot)$ e $l''(\cdot)$ são, respectivamente, a primeira e a segunda derivadas do logaritmo da função de verossimilhança.
- 106** O método *scoring* de Fisher para maximizar funções de log-verossimilhança consiste em calcular, iterativamente, a equação $\theta^{(t+1)} = \theta^{(t)} + \frac{l'(\theta^{(t)})}{I(\theta^{(t)})}$, sendo $\theta^{(t)}$ e $\theta^{(t+1)}$ o parâmetro a ser estimado nos momentos t e $t + 1$, respectivamente; $l'(\cdot)$ e $I(\cdot)$ são, respectivamente, utilizando um computador, a primeira derivada do logaritmo da função de verossimilhança e a informação esperada de Fisher.
- 107** O método *scoring* de Fisher diverge, assintoticamente, do método de Newton, sendo aquele mais confiável para a obtenção de estimativas não viciadas.

Considere que, com base em levantamentos acerca da presença de bolores/leveduras em três marcas de pizza congelada, foram obtidas as seguintes frequências em 90 caixas de pizza analisadas, descritas na tabela de contingência abaixo.

		marcas de pizza congelada		
		A	B	C
concentração de bolores/leveduras	$< 10^4$ UFC/g	22	13	10
	$\geq 10^4$ UFC/g	8	17	20

Com base nessa situação hipotética, julgue os itens a seguir.

- 108** Dado que a concentração de leveduras foi inferior a 10.000 UFC/g, a probabilidade de que a pizza seja da marca A é inferior a 45%.
- 109** A frequência esperada para as células da tabela de contingência é igual a 15.
- 110** A contribuição de cada célula associada à marca C na estatística χ^2 para o teste de homogeneidade é superior a 2.
- 111** A estatística χ^2 tem distribuição qui-quadrado com $l \times c$ graus de liberdade, em que l é o número de linhas da tabela e c é o número de colunas da tabela.
- 112** Considere que a estatística χ^2 para o teste de homogeneidade tenha sido igual a 10,4. Supondo que o valor-P seja igual a 0,006, então é improvável a probabilidade de se obter uma estatística χ^2 superior à que foi obtida nesse estudo.

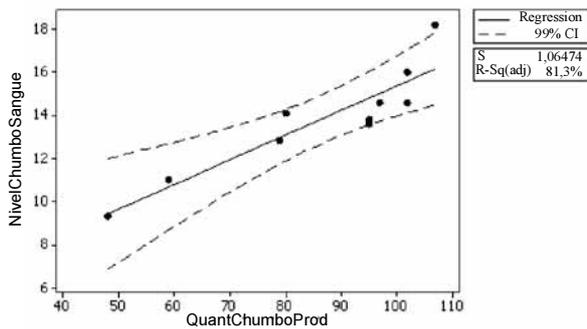
RASCUNHO

Um estudo foi feito para avaliar o impacto da utilização de chumbo na produção de gasolina sobre a saúde de crianças expostas à poluição gerada pela queima desse combustível em veículos automotores. Duas variáveis foram consideradas: y é igual ao nível médio de chumbo no sangue de crianças brancas entre 6 meses e 5 anos de idade e x é igual à quantidade de chumbo usada na produção de gasolina, em 1.000 toneladas, em dez períodos de seis meses. O analista responsável realizou uma análise estatística dos dados e obteve os seguintes resultados a partir de um *software* estatístico.

Descriptive Statistics: QuantChumboProd; NivelChumboSangue

Variable	Mean	TrMean	CoefVar	Median	Skewness	Kurtosis
QuantChumboProd	86,40	88,63	22,80	95,00	?????	0,10
NivelChumboSangue	13,800	13,812	17,86	13,950	?????	0,81

Quant. de chumbo na prod. de gasolina versus nível médio de chumbo no sangue
 NivelChumboSangue = 3,924 + 0,1143 QuantChumboProd



The regression equation is

$$\text{NivelChumboSangue} = 3,92 + 0,114 \text{ QuantChumboProd}$$

Predictor	Coef	SE Coef	T	P
Constant	3,924	1,593	????	?????
QuantChumboProd	0,11431	0,01802	????	?????

S = 1,06474 R-Sq = ????? R-Sq (adj) = 81,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	45,631	?????	?????	?????
Residual Error	8	9,069	?????		
Total	9	54,700			

Com base nos resultados apresentados, julgue os itens subsequentes.

113 O intervalo desenhado no gráfico representa o intervalo de 99% de predição para a média, calculado mediante a fórmula

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} \cdot s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

sendo x^* o valor futuro da variável x , $t_{\alpha/2, n-2}$ o quantil da distribuição t de Student e $S_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$.

114 Sabendo que o quantil da distribuição t para uma probabilidade superior de 5% e 1% (considerando os graus de liberdade apropriados) seja igual a 1,83 e 2,82, respectivamente, então o analista concluiu que a variável quantidade de chumbo na produção de gasolina mostrou-se significativa ao nível de confiança de 5%, mas não significativa ao nível descritivo de 1%.

115 É correta a seguinte afirmação: a cada mil toneladas adicionais de chumbo usado na produção de gasolina gera, em média, um aumento de 0,114 unidades no nível médio de chumbo no sangue das crianças.

- 116 O coeficiente de determinação foi superior a 85%.
- 117 Para se obter os estimadores de mínimos quadrados desse modelo de regressão, é necessário tomar as derivadas parciais da função $f(b_0, b_1) = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2$, igualá-las a zero e resolver o sistema de duas equações lineares.
- 118 Os coeficientes de assimetria de ambas as variáveis são negativos.
- 119 Se existissem 6 outras variáveis independentes, potencialmente influentes no nível médio de chumbo no sangue das crianças, um método para selecionar um subconjunto parcimonioso de variáveis é o método passo-a-passo (ou *stepwise*) que requer que o analista ajuste todas as equações de regressão que envolvem uma variável candidata, todas as equações que envolvem duas variáveis candidatas, e assim por diante, até o ajuste do modelo maximal. O desempenho de todos os modelos é comparado por meio do coeficiente de determinação, do quadrado médio dos erros ou da estatística c_p .
- 120 Se uma variável independente está fortemente correlacionada com outra variável independente, a matriz X do modelo de regressão $y = X\beta + \epsilon$ tende a ter colunas linearmente dependentes, fazendo que as variâncias e covariâncias tenham valores muito grandes, implicando coeficientes de regressão mal estimados, devido à instabilidade numérica da inversão da matriz $X^T X$. Esse fenômeno é conhecido como multicolinearidade.

RASCUNHO

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando os espaços para rascunho indicados no presente caderno. Em seguida, transcreva o texto para o **CADERNO DE TEXTO DEFINITIVO DA PROVA DISCURSIVA**, nos locais apropriados, pois **não serão avaliados fragmentos de texto escritos em locais indevidos**.
- Qualquer fragmento de texto além da extensão máxima de **sessenta** linhas será desconsiderado.
- No **caderno 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.

Um pesquisador deseja investigar se dois tipos de cerâmica, A e B, possuem os mesmos coeficientes médios de atrito. Para o estudo, ele selecionará n_A peças do tipo A e n_B peças do tipo B, perfazendo um total de $n = n_A + n_B$ peças na amostra. As variâncias dos coeficientes de atrito das cerâmicas A e B são, respectivamente, iguais a σ_A^2 e σ_B^2 , e \bar{X}_A e \bar{X}_B são as médias amostrais dos coeficientes de atrito a serem observadas no estudo.

Esse pesquisador, no entanto, tem algumas dúvidas sobre os seguintes pontos:

- a) como se calcula um valor para o tamanho da amostra n que tenha relação direta com o poder do teste de comparação de médias de duas populações?
- b) como se determinam os valores de n_A e n_B , de modo que se minimize o valor da variância da diferença $\bar{X}_A - \bar{X}_B$?
- c) como ficaria a alocação da amostra (que minimiza o valor da variância da diferença $\bar{X}_A - \bar{X}_B$) quando se considera um custo total de amostragem proporcional a $n_A + \gamma \times n_B$, em que $\gamma > 1$?
- d) o que são os erros não amostrais? Quais são as recomendações essenciais para que esse estudo tenha valor científico?

Considerando a situação hipotética acima, elabore um texto que responda detalhadamente as quatro dúvidas levantadas pelo pesquisador, apresentando (com as devidas justificativas ou demonstrações) as fórmulas matemáticas e os procedimentos necessários para a aplicação dessas fórmulas, de tal modo que o pesquisador possa efetuar, na prática, o estudo pretendido, seguindo as suas orientações.

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	