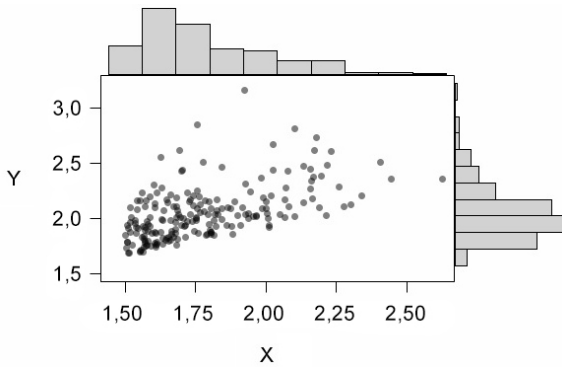


CONHECIMENTOS ESPECÍFICOS



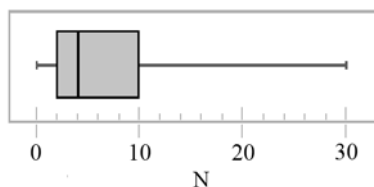
A figura acima mostra a dispersão dos valores previstos (X, em R\$ milhões) e dos valores efetivamente gastos (Y, em R\$ milhões) em 200 obras de pavimentação em determinado estado, e os respectivos histogramas das distribuições dos valores X e Y. Com base nessas informações, julgue os itens a seguir.

- 51 A distribuição de X apresentou assimetria positiva, enquanto a de Y exibiu assimetria negativa.
- 52 A amplitude total da distribuição de X foi igual ou superior a R\$ 2 milhões.

estatística	quantidade de acidentes (N)
mínimo	0
primeiro quartil	2
segundo quartil	4
terceiro quartil	10
máximo	30

A tabela acima mostra algumas estatísticas descritivas produzidas por um estudo acerca da quantidade de acidentes de trabalho (N), ocorridos em 2012, a partir de uma amostra aleatória simples de 200 indústrias de pequeno porte. Com base nessas informações, julgue os próximos itens.

- 53 O coeficiente de variação quartil foi superior a 50% e inferior a 60%.
- 54 O intervalo interquartil da variável N foi igual a 8.
- 55 O número médio de acidentes ocorridos em 2012 em cada uma dessas 200 indústrias foi inferior a 4.
- 56 A moda da distribuição foi superior a 2.
- 57 A figura abaixo mostra corretamente o diagrama de *box-plot* da distribuição do número de acidentes de trabalho ocorridos no ano de 2012 na referida amostra.



As variáveis aleatórias X e Y seguem uma distribuição de Bernoulli com probabilidade de sucesso igual a 0,4. Considerando $S = X + Y$ e que os eventos aleatórios $A = [X = 1]$ e $B = [Y = 1]$ sejam mutuamente exclusivos, julgue os itens subsequentes.

- 58 A variável aleatória S segue uma distribuição de Bernoulli.
- 59 $P([X = 0] \cap [Y = 0]) < 0,1$.
- 60 O desvio padrão da soma S é igual a 0,4.
- 61 Quanto à distribuição condicional, a regressão linear de Y em $X = x$ é expressa por $E(Y | X = x) = \frac{5+x}{15}$.
- 62 É correto afirmar que $P(A \cup B) = P(S = 1) = 0,8$.

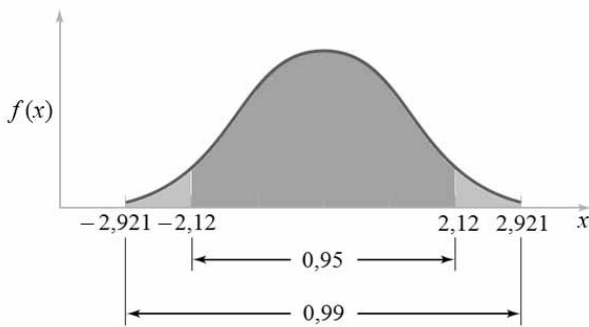
RASCUNHO

Um estudo mostrou que a distribuição do tempo de reação (em segundos) dos operários que trabalham em minas de carvão frente a situações de perigo segue uma distribuição W cuja função de densidade de probabilidade é dada por $f(w) = 2^{-w} \ln 2$, se $w \geq 0$, e $f(w) = 0$, se $w < 0$. Com base nessas informações, julgue os itens a seguir.

- 63 A média dos tempos W é igual a $\ln 2$.
- 64 É correto afirmar que $2 \times P(W > k + 1) = P(W > k)$, em que $k \geq 0$.
- 65 A mediana e o terceiro quartil da distribuição W são, respectivamente, iguais a 1 s e 2 s.
- 66 O tempo de reação W se distribui conforme uma distribuição exponencial.

A distribuição conjunta de dois indicadores de qualidade do ar, X e Y , é expressa por $f(x, y) = \alpha xy$, em que $0 \leq x \leq 1$, $0 \leq y \leq 1$ e $\alpha > 0$. Para outros valores de x e de y , $f(x, y) = 0$. Com base nessas informações, julgue os próximos itens.

- 67 A distribuição condicional $Y|X=0,5$ é uniforme no intervalo $(0, 1)$.
- 68 É correto afirmar que α é um parâmetro da distribuição que pode assumir qualquer valor real positivo, e, a partir de uma amostra aleatória simples, esse parâmetro pode ser estimado pelo método dos momentos.
- 69 Os indicadores X e Y possuem médias iguais a $\frac{\sqrt{\alpha}}{2}$.
- 70 O valor esperado de $\frac{[X - Y]^2}{2}$ é igual à variância de X .
- 71 Como $0 = f(0, 0) < f(1, 1) = \alpha$, é correto afirmar que X e Y se correlacionam positivamente.
- 72 A média de $\sqrt{X^2 + Y^2}$ é inferior a 1.



A figura acima mostra a forma do gráfico da função de densidade de uma distribuição t de Student com 16 graus de liberdade (T_{16}), com a indicação de alguns valores x e de duas probabilidades associadas a esses quantis. Com base no gráfico e nas informações, julgue os seguintes itens.

- 73 A sequência (T_n) , em que T_n é a distribuição t de Student com n graus de liberdade, converge em distribuição para a distribuição normal padrão à medida que n aumenta.
- 74 Se Y representa uma distribuição F de Snedecor com 1 grau de liberdade no numerador e 16 graus de liberdade no denominador, então $P(0 < Y < 2,12^2) = 0,95$.
- 75 $P(2,12 < T_{16} < 2,921) = 0,04$.
- 76 Se T_{20} representa uma distribuição t de Student com 20 graus de liberdade, então $P(T_{20} < 2,921) > 0,99$.

Uma amostra aleatória simples X_1, X_2, \dots, X_n foi retirada de uma distribuição contínua, em que θ é o parâmetro de interesse e $S_n = S(X_1, X_2, \dots, X_n)$ é o seu estimador. A respeito dessa amostra, julgue os próximos itens.

- 77 O teorema limite central trata da convergência em probabilidade do estimador S_n para o parâmetro θ .
- 78 Se o estimador S_n converge em norma L_1 para o parâmetro θ à medida que o tamanho da amostra aumenta, então S_n converge em probabilidade para θ .
- 79 Se S_n e θ forem as médias amostral e populacional, respectivamente, então — conforme a lei fraca dos grandes números — S_n converge quase certamente para θ , à medida que n cresce.
- 80 Considerando-se que a amostragem tenha sido feita sobre uma população normal com média μ e variância σ^2 , e que S_n seja a mediana, a distribuição amostral da estatística S_n é assintoticamente normal com média μ e variância igual a $\frac{\pi\sigma^2}{2n}$.

RASCUNHO

RASCUNHO

Considerando que x_1, \dots, x_n representa uma amostra aleatória simples retirada de uma distribuição contínua X cuja função densidade de probabilidade é $f(x) = \frac{\lambda^3 x^2 e^{-\lambda x}}{2}$, em que $x \geq 0$ e $\lambda > 0$, julgue os próximos itens, acerca da estimação de máxima verossimilhança do parâmetro λ .

- 81 O estimador de máxima verossimilhança do parâmetro λ é $\hat{\lambda} = \frac{3}{n\bar{x}}$, em que \bar{x} é a média amostral.
- 82 A função de verossimilhança é $L(\lambda) = 2^{-n} \lambda^{3n} e^{-\lambda n\bar{x}} (\prod_{i=1}^n x_i)^2$, em que \bar{x} é a média amostral.

Suponha que x_1, \dots, x_n seja uma sequência de cópias independentes retiradas de uma distribuição com função densidade de probabilidade $f(x) = \alpha x e^{-\frac{\alpha x^2}{2}}$, em que $x \geq 0$ e $\alpha > 0$ é seu parâmetro. Com base nessas informações, julgue os itens a seguir.

- 83 Supondo que $(x_1, \dots, x_5) = (3, 4, 4, 6, 6)$, a estimativa de máxima verossimilhança do parâmetro α é inferior a $\frac{1}{10}$.
- 84 A função de log-verossimilhança para a estimação do parâmetro α é $\ln f(x) = n \ln \alpha + \sum_{i=1}^n \ln x_i - \frac{\alpha}{2} \sum_{i=1}^n x_i$.

indivíduo (i)	altura (x_i), em cm	peso (y_i), em kg
1	150	55
2	160	65
3	170	75
4	180	75
5	190	90

Considerando que a tabela acima mostra as alturas e as massas corporais de cinco pessoas participantes de um estudo nutricional, e que $\bar{x} = \sum_{i=1}^5 \frac{x_i}{5} = 170$ e $\bar{y} = \sum_{i=1}^5 \frac{y_i}{5} = 72$ são a altura média e o peso médio, respectivamente, $\sum_{i=1}^5 x_i^2 = 145.500$; $\sum_{i=1}^5 y_i^2 = 26.600$; $\sum_{i=1}^5 x_i y_i = 62.000$, julgue os seguintes itens acerca do modelo de regressão linear simples $y_i = a + bx_i + \varepsilon_i$, em que ε_i é um erro aleatório com média nula e variância constante, e a e b são os objetos da estimação.

- 85 A soma dos quadrados total é igual a 680.
- 86 A soma dos quadrados dos erros entre as massas corporais observadas e esperadas segundo o modelo ajustado é igual a 40.
- 87 A soma dos quadrados de regressão é igual a 650.
- 88 O coeficiente de determinação do modelo (R^2) é igual a $\frac{15}{17}$.
- 89 Os estimadores para os coeficientes a e b pelo método de mínimos quadrados ordinários satisfazem ao sistema de equações normais a seguir.

$$\begin{cases} \bar{y} = a + b\bar{x} \\ \sum_{i=1}^5 x_i y_i = 5a\bar{x} + b \sum_{i=1}^5 x_i^2 \end{cases}$$

- 90 As estimativas de máxima verossimilhança para os coeficientes a e b , na hipótese de os erros aleatórios serem normais, são $\hat{a} = -81$ e $\hat{b} = 0,8$.
- 91 Os resíduos são dados pela expressão $r_i = y_i - \hat{a} - \hat{b}x_i$. Em particular, $r_1 = -6$.

$t =$ tempo (em anos, a partir do início do estudo)	$y_t =$ tamanho da população
0	500.000
10	550.000
20	610.000
30	680.000
40	750.000

A tabela acima mostra os resultados de um estudo demográfico em que se analisou o crescimento da população de determinada cidade ao longo do tempo. Considerando os dados da tabela e uma curva de crescimento exponencial $y = \varepsilon \alpha e^{\beta t}$, em que ε representa um erro aleatório com média unitária, julgue os itens subsequentes.

92 Calculando-se as derivadas parciais S_α e S_β da soma dos erros quadrados, as equações $S_\alpha = 0$ e $S_\beta = 0$ fornecem as seguintes equações normais:

$$\alpha \sum_{i=1}^n e^{2\beta t_i} = \sum_{i=1}^n y_i e^{\beta t_i} \text{ e } \alpha \sum_{i=1}^n t_i e^{\beta t_i} = \sum_{i=1}^n t_i y_i e^{\beta t_i}$$

93 É correto linearizar o modelo com a reparametrização $a = \ln \alpha$ e a transformação da variável dependente $z = \ln y$. Dessa forma resulta o modelo $z = a + \beta t + \ln \varepsilon$.

94 É correto estimar os parâmetros α e β pelo método dos mínimos quadrados ordinários mediante linearização do modelo, isto é, minimizando-se a soma dos erros quadrados

$$S(\alpha, \beta) = \sum_{k=1}^5 (\ln \alpha + \beta t_k - \ln y_k)^2.$$

A população de uma cidade divide-se em três estratos: classe baixa (CB), com 25% da população; classe média (CM), com 60%; e classe alta (CA), com 15%. O desvio padrão dos salários mensais das classes é R\$ 400,00 R\$ 600,00 e R\$ 2.800/3, respectivamente. A fim de se estimar o salário mensal médio da população, escolhe-se uma amostra de tamanho n . Com base nessas informações, julgue os itens subsequentes acerca da amostragem.

95 A probabilidade de que uma amostra aleatória simples de tamanho $n = 10$ não contenha pessoas da CB é superior a 0,1%.

96 Na amostragem aleatória estratificada com alocação proporcional aos tamanhos dos estratos, uma amostra de $n = 400$ pessoas deve contemplar 60 pessoas da CB, 240 da CM e 100 da CA.

97 Considerando uma amostra estratificada de tamanho $n = 600$ com alocação ótima de Neyman, é correto afirmar que do estrato CM devem ser amostradas 360 pessoas.

Julgue os itens a seguir, a respeito de amostragem por conglomerados, considerando uma população $U = \{1, \dots, 6\}$ com conglomerados $C_1 = \{1\}$, $C_2 = \{2, 3\}$ e $C_3 = \{4, 5, 6\}$ e o vetor de dados associado $D = (15, 10, 4, 5, 8, 6)$.

98 Se dois dos três conglomerados — C_1, C_2, C_3 — da população U forem escolhidos em seguida para formar a amostra, as possíveis amostras serão: $\{1, 2, 3\}$, $\{1, 4, 5, 6\}$, $\{2, 3, 1\}$, $\{2, 3, 4, 5, 6\}$, $\{4, 5, 6, 1\}$, $\{4, 5, 6, 1, 2, 3\}$.

99 Se dois dos três conglomerados — C_1, C_2, C_3 — da população U forem escolhidos para formar a amostra, a média amostral assume os seguintes possíveis valores: $29/3$, $17/2$ e $33/5$, de modo que cada um desses valores ocorre com a mesma probabilidade, isto é, $1/3$.

100 Na amostragem por conglomerados, a população é subdividida em grupos com tamanhos não necessariamente iguais. A amostra obtida consiste da união dos conglomerados escolhidos e geralmente o tamanho dessa amostra é uma variável aleatória.

RASCUNHO

contrato i	1	2	3	4	5	6	7	8	9	10
valor x_i	10	20	14	1	5	18	4	10	10	5

Na tabela acima, resultante da aplicação de uma amostragem aleatória simples, cada observação representa o valor (em R\$ milhões) do contrato i para a prestação de determinado serviço a um órgão público. Considerando que a distribuição populacional desses valores seja normal com variância desconhecida e que $\sum_{i=1}^{10} x_i^2 = 1.287$ e $\sum_{i=1}^{10} x_i = 97$, julgue os itens a seguir.

- 101 Considerando-se que $P(Z < 1,96) = 0,975$, em que Z representa a distribuição normal padrão, o intervalo (simétrico) de 95% de confiança para a média populacional apresenta quantil que multiplica o erro padrão da média amostral superior a 2.
- 102 Considere que, no intervalo de confiança para a variância populacional, L represente o valor do quantil da distribuição amostral que corresponde ao limite inferior desse intervalo. Nesse caso, o valor do quantil referente ao limite superior é $-L$.
- 103 A estimativa pontual para a mediana da população de valores contratados por esse órgão público é maior que R\$ 10,5 milhões.
- 104 A estimativa de máxima verossimilhança para a variância dessa população é superior a 36.
- 105 Ao se aplicar o teste t para a média populacional (μ), verifica-se que a estatística do teste cuja hipótese alternativa é $H_1: \mu \geq$ R\$ 5 milhões apresenta 9 graus de liberdade.
- 106 Caso se adote o método dos momentos, a estimativa da variância populacional será inferior a 34.

quantidade de processos	0	1	2	3	4
frequência	5	8	4	2	1

Considerando a tabela acima, que apresenta o registro das quantidades anuais de processos abertos contra autoridades públicas nas duas últimas décadas, julgue os itens de 107 a 113.

- 107 Considerando-se que esses dados sejam originários de uma distribuição de Poisson com média unitária e as aproximações $e^0 = 1$, $e^1 = 2,73$, $e^2 = 7,45$, então o número esperado de anos em que não foram registrados processos contra autoridades públicas é superior a 4.
- 108 Caso se aplique o teste de aderência desses dados para a distribuição de Poisson com taxa estimada igual a \bar{x} , em que \bar{x} representa a média amostral, a estatística desse teste apresentará 5 graus de liberdade.
- 109 Em qualquer teste qui-quadrado, a estatística do teste é calculada utilizando-se a diferença entre valores observados e valores esperados.
- 110 Utilizando-se o teste de aderência desses dados à distribuição de Poisson com parâmetro igual a 1, a estatística qui-quadrado apresentará dois graus de liberdade.

- 111 Ao se testar a hipótese nula de a média populacional ser igual a 2 mediante a aplicação do teste t , verifica-se que a estatística do teste apresenta valor positivo.
- 112 Caso o teste qui-quadrado de aderência a uma distribuição de Poisson apresente p -valor superior a 0,10, é recomendável rejeitar a hipótese nula de normalidade e aplicar o teste de Wilcoxon para o teste de valor central.
- 113 Na situação hipotética em questão, espera-se que o teste de Wilcoxon apresente maior poder estatístico que o teste t de Student.

RASCUNHO

RASCUNHO

Na avaliação do processo de informatização de um sistema jurídico de determinado órgão, observou-se redução do tempo médio (em meses) gasto para a análise de um processo. Na primeira etapa da avaliação, considerou-se uma amostra de tempos de 49 processos originados anteriormente à informatização do sistema e constataram-se os seguintes resultados: tempo médio de análise igual a 15 meses e desvio padrão dos tempos igual a 7 meses. Na segunda etapa, considerou-se uma amostra de tempos de 81 processos originados após a informatização do sistema e constataram-se os seguintes resultados: tempo médio de análise igual a 10 meses e desvio padrão dos tempos igual a 6 meses.

Com base nessas informações, julgue os itens subsecutivos, relativos aos testes de comparação entre médias.

- 114 O teste t apropriado à situação em tela é o teste para dados pareados.
- 115 Para calcular a potência do teste, é suficiente realizar $1-\beta$, em que β é probabilidade de erro do tipo II e μ é o mesmo valor utilizado para calcular o teste com α (probabilidade de erro do tipo I).
- 116 Para a verificação da efetividade da informatização do sistema, a hipótese nula deve ser dada por $H_0: \mu_D - \mu_A = 0$, em que μ_D e μ_A são, respectivamente, a média logo após a informatização e a média antes da informatização, o que significa que o denominador da estatística do teste é dado, também, pela diferença das variâncias.
- 117 Considere que se pretenda calcular o intervalo de confiança clássico com 95% de confiança para a diferença dos tempos. Nessa situação, se fossem feitos 100 intervalos com base em amostras de mesmo tamanho, 95 desses conteriam o parâmetro, diferentemente do intervalo bayesiano de credibilidade, o que indica que a probabilidade de o parâmetro estar dentro do intervalo é de 95%.
- 118 Sabendo-se que as variâncias populacionais sejam iguais, é correto afirmar que a estatística do teste é superior a 2.
- 119 Considere que a expectativa fosse de que o tempo médio de análise dos processos cairia 25% após a informatização. Nesse caso, se $P(t_{80} > 1,6641) = 0,05$ e $P(t_{80} > 1,9901) = 0,025$, infere-se que o objetivo foi alcançado.
- 120 No caso da média populacional, o intervalo de confiança e o intervalo de credibilidade (usando-se uma priori não informativa) são numericamente iguais, mas com interpretações diferentes.

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando, caso deseje, 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 será avaliado fragmento de texto escrito em local indevido**.
- 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.
- Ao domínio do conteúdo serão atribuídos até **40,00 pontos**, dos quais até **2,00 pontos** serão atribuídos ao quesito apresentação (legibilidade, respeito às margens e indicação de parágrafos) e estrutura textual (organização das ideias em texto estruturado).

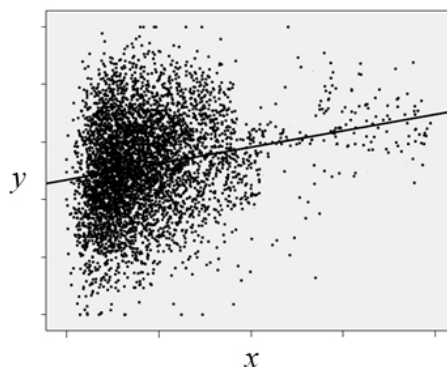
Um estudo sobre a qualidade da gestão pública em programas de assistência social considerou um modelo de regressão linear simples na forma $y_k = a + bx_k + \varepsilon_k$, em que y_k é um indicador que representa a capacidade administrativa do município k , a variável regressora x_k denota o valor do gasto público *per capita* do município k em programas de assistência social, e $k = 1, 2, \dots, 5.564$. Os coeficientes do modelo são representados por a e b , e ε_k representa o erro aleatório com média zero e variância constante. Sob a hipótese de que esses erros aleatórios sejam normais (independentes e identicamente distribuídos), a tabela 1 abaixo mostra as estimativas dos coeficientes — que foram obtidos pelo método dos mínimos quadrados ordinários — e a tabela 2 apresenta o resultado da análise de variância (ANOVA). A Figura 1 mostra o diagrama de dispersão entre as variáveis y_k e x_k e um esboço da reta ajustada.

tabela 1

coeficiente	estimativa	erro padrão	razão t	$P(T > t)$
a	40,0	10,0	4,0	< 0,001
b	0,2	0,05	4,0	< 0,001

tabela 2

fonte de variação	graus de liberdade	soma de quadrados	quadrado médio	razão F	p -valor
modelo	1	2.500	QM(modelo)	292,73	< 0,001
erro	5.562	47.500	8,54		
total	5.563	50.000	8,99		



Com base no conjunto de informações acima, redija um texto dissertativo que apresente análise estatística acerca dos resultados produzidos pelo referido estudo. Ao elaborar seu texto, faça, necessariamente, o que se pede a seguir.

- ▶ Descreva as distribuições amostrais dos estimadores dos coeficientes do modelo, de suas médias e variâncias. [valor: 13,00 pontos]
- ▶ Analise os resultados mostrados nas tabelas 1 e 2. [valor: 12,50 pontos].
- ▶ Apresente diagnóstico do modelo, com base na figura. [valor: 12,50 pontos]

RASCUNHO

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	