-- CONHECIMENTOS ESPECÍFICOS --

Com respeito ao conjunto de dados {0, 0, 1, 1, 1, 3}, julgue os itens que se seguem.

51 Com base na medida $\kappa = \frac{\mu_4}{(\mu_2)^2} - 3$, em que μ_r denota o r-ésimo momento amostral centrado na média, é correto afirmar que a forma do conjunto de dados em tela é considerada platicúrtica.

52 Se esse conjunto de dados fosse representado por um diagrama de *box-plot*, então os valores 0 e 3 seriam chamados valores exteriores, ou, ainda, discrepantes, atípicos ou *outliers*.

53 Como a média amostral é igual à mediana amostral, a distribuição em tela pode ser considerada como simétrica em torno da média.

54 Se μ_3 representa o terceiro momento amostral centrado na média, então $\mu_3 > 0$, o que sugere que a distribuição seja assimétrica à direita.

55 O coeficiente de variação é igual ou superior a 1,2.

x_1	x_2	x_3	x_4	x_5
10,6	9,4	10,6	10,4	9,0

Suponha que o conjunto de dados mostrados no quadro acima seja uma realização de uma amostra aleatória simples de tamanho $n = 5$ que foi retirada de uma população cuja função de densidade de probabilidade é dada por

$$f(x) = \frac{\theta e^{-\theta|x-\mu|}}{2},$$

na qual $x \in \mathbb{R}$, e $\theta > 0$ e $\mu \in \mathbb{R}$ são parâmetros desconhecidos.

Com base nessas informações, julgue os itens subsequentes.

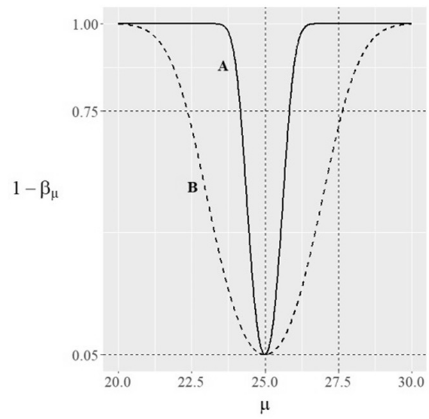
56 A estimativa de máxima verossimilhança para o parâmetro μ é igual a 10,4.

57 A estimativa de máxima verossimilhança da moda populacional é igual a 10,6.

58 De acordo com o método dos mínimos quadrados ordinários, a estimativa do parâmetro μ é igual a 10.

59 Se X for definida como uma variável aleatória que representa a distribuição populacional em tela e se $p = P\,(X = 10{,}6)$, então a estimativa dessa probabilidade será $\hat{p} = 2/5$.

60 A estimativa de máxima verossimilhança do desvio padrão populacional é igual a $\frac{1}{\hat{\theta}}$, em que $\hat{\theta}$ representa a estimativa de máxima verossimilhança do parâmetro θ.



Considerando que a figura acima mostra as curvas de poder referentes a dois testes de hipóteses — A (linha contínua) e B (linha tracejada) — para a média populacional μ, julgue os itens a seguir.

61 β_μ é denominada probabilidade de significância ou nível descritivo do teste.

62 Os tamanhos dos testes de hipóteses A e B são coincidentes.

63 O teste de hipóteses A é uniformemente mais poderoso que o teste de hipóteses B.

64 Com respeito ao teste de hipóteses B, se $\mu = 27{,}5$, então a probabilidade de se rejeitar a hipótese nula será inferior a 0,75.

65 Os testes de hipóteses A e B são bilaterais, com $H_0 : \mu = 25$ e $H_1 : \mu \neq 25$.

O quadro abaixo mostra a realização de uma amostra aleatória simples u_1, u_2, u_3, u_4, que foi retirada de uma distribuição uniforme contínua no intervalo $[0, a]$.

u_1	u_2	u_3	u_4
7,53	3,02	5,69	1,22

Considerando que $\hat{a}$ representa a estimativa de máxima verossimilhança do parâmetro a, julgue os itens seguintes.

66 A estimativa de máxima verossimilhança para a média da distribuição em tela é igual a 4,365.

67 A estimativa não viciada para o parâmetro a é dada pela expressão $1{,}25 \times \hat{a}$.

68 $[\hat{a}, \hat{a}(0{,}05)^{-0{,}25}]$ representa um intervalo de 95% de confiança para o parâmetro a.

O quadro abaixo mostra o resultado de uma pesquisa de opinião acerca de certo assunto que foi aplicada a dois públicos distintos, I e II.

público	opinião		total
	favoráveis	desfavoráveis	
I	120	30	150
II	30	20	50
total	150	50	200

Com respeito a essa situação hipotética, julgue os próximos itens.

69 Caso o objetivo da pesquisa em questão seja avaliar se as distribuições das opiniões seriam as mesmas para ambos os públicos, testando-se a hipótese nula $H_0 : p_{\mathrm{I}} = p_{\mathrm{II}}$ contra a hipótese alternativa $H_1 : p_{\mathrm{I}} \neq p_{\mathrm{II}}$, em que p_{I} e p_{II} representam, respectivamente, as proporções populacionais de indivíduos dos públicos I e II que se posicionam favoráveis, então, para essa situação, os valores corretos esperados sob H_0 para a aplicação do teste χ^2 serão aqueles mostrados na tabela abaixo.

público	opinião		total
	favoráveis	desfavoráveis	
I	110	40	150
II	40	10	50
total	150	50	200

70 Caso o objetivo da pesquisa em apreço seja testar se a variável opinião é independente da variável público, então a estatística do teste χ^2 para esse propósito possuirá três graus de liberdade.

Considerando que a função de densidade conjunta do par de variáveis aleatórias (X, Y) seja dada por

$$f(x,y) = \begin{cases} \dfrac{3(1-x^2)}{4}, & \text{se } |x| \leq 1 \text{ e } 0 \leq y \leq 1; \\ 0, & \text{se caso contrário,} \end{cases}$$

julgue os próximos itens.

71 $P(|X| \leq y | Y = y) = \frac{y(3-y^2)}{2}$, em que $0 \leq y \leq 1$.

72 $\mathrm{Var}(Y) = \frac{1}{12}$.

73 $P(Y = y | |X| \leq y) = y$, em que $0 \leq y \leq 1$.

74 $\mathrm{E}(X) > 0$.

75 A correlação linear entre as variáveis X e Y é positiva.

Considerando que $X_1, X_2, \dots X_n$ seja uma sequência de variáveis aleatórias independentes e identicamente distribuídas, tais que

$$\mathrm{P}(X_k = x) = p(1-p)^x,$$

em que $x \in \{0, 1, 2, 3, \dots\}$, $0 < p \leq 1$ e $k \in \{1, 2, \dots, n\}$, julgue os itens a seguir.

76 Se $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, então, segundo a lei fraca dos grandes números, $\bar{X}_n$ converge em probabilidade para $\frac{1}{p}$.

77 $\mathrm{Var}(\sum_{k=1}^{n} X_k) = \frac{n(1-p)}{p^2}$.

78 $\mathrm{P}(\sum_{k=1}^{n} X_k = s) = \binom{n}{s} p^{n-s}(1-p)^s$.

79 Se $X_{(1)} = \min\{X_1, \dots, X_n\}$, então

$$\mathrm{P}\big(X_{(1)} \leq x\big) = 1 - [(1-p)^{x+1}]^n.$$

80 Se $Y_n = \sum_{k=1}^{n} 0{,}5^k X_k$, então, mediante a aplicação do teorema central do limite, é correto concluir que $Y_n \xrightarrow{D}$ Normal.

Supondo que

$$P(Y = y | M = m) = \frac{e^{-m} m^y}{y!},$$

para $y \in \{0, 1, 2, 3 \dots\}$, em que $m > 0$, e M é uma variável aleatória contínua cuja função de densidade é dada por $f_M(m) = e^{-m}$, julgue os itens a seguir.

81 $\mathrm{Var}(Y = y | M = m) = m$.

82 $\mathrm{P}(Y = y) = \frac{1}{2^{y+1}}$, para $y \in \{0,1,2,3 \dots\}$.

83 Y e M são variáveis aleatórias independentes.

84 $\mathrm{P}(Y > 0 | M = m) = \mathrm{P}(M \leq m)$.

Considerando que uma amostra aleatória simples $U_1, \dots, U_n$ seja retirada de uma distribuição uniforme contínua no intervalo $[0,1]$, em que n é número ímpar, e considerando que $\bar{U}_n$ denote a média amostral e $\tilde{U}_n$ represente a mediana amostral, julgue os itens a seguir.

85 $\mathrm{E}[\tilde{U}_n] = 0{,}5$.

86 Para todo n suficientemente grande, $\mathrm{Var}[\tilde{U}_n] > \mathrm{Var}[\bar{U}_n]$.

87 $12n(\bar{U}_n - 0{,}5)$ converge para uma distribuição normal padrão.

O quadro a seguir mostra as estimativas de mínimos quadrados ordinários dos coeficientes de um modelo de regressão linear simples na forma $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, em que $i \in \{1, \dots, 6\}$ e ε_i representa o erro aleatório com média zero e variância σ^2.

coeficiente	estimativa	erro padrão	razão t
β_0	0,9	0,10	9
β_1	0,2	0,05	4

Considerando essas informações e sabendo que $\hat{\sigma}^2 = 0{,}01$, julgue os itens seguintes.

88 $S_{xx} = \sum_{i=1}^{6}(x_i - \bar{x})^2 = 4$, em que $\bar{x} = \sum_{i=1}^{6} x_i/6$.

89 A covariância entre a variável resposta (y) e a variável explicativa (x) é igual ou superior a 0,2.

90 O coeficiente de determinação do modelo (R^2) é igual a 0,8.

91 $SQ_{RESÍDUOS} = \sum_{i=1}^{6}(\hat{y}_i - \bar{y})^2 = 0{,}04$, em que $\hat{y}_i = 0{,}9 + 0{,}2x_i$.

92 $SQ_{TOTAL} = \sum_{i=1}^{6}(y_i - \bar{y})^2 = 0{,}2$.

Considerando que $\hat{y}_k$ denote o valor ajustado — pelo método de mínimos quadrados ordinários — da variável resposta y_k de um modelo de regressão linear múltipla na forma $y_k = \beta_0 + \beta_1 x_{1,k} + \beta_2 x_{2,k} + \varepsilon_k$, para $k \in \{1, \dots, 10\}$; que, nesse modelo, $\{\varepsilon_1, \dots, \varepsilon_{10}\}$ seja um conjunto de erros aleatórios independentes com médias iguais a zero e variâncias iguais a σ^2; e que cada resíduo produzido pelo ajuste seja escrito como $r_k = y_k - \hat{y}_k$, julgue os próximos itens.

93 A distância D de Cook representa uma medida da influência. Em particular, essa medida é dada por $\sum_{k=1}^{10} \frac{\hat{y}_{k(i)} - \hat{y}_k}{3\hat{\sigma}^2}$, na qual $\hat{y}_{k(i)}$ denota o valor ajustado para y_k, omitindo-se o elemento i da amostra no cálculo das estimativas dos coeficientes do modelo.

94 Os valores da sequência $r_1, \dots, r_{10}$ são mutuamente independentes.

95 A estatística $\sum_{k=2}^{10}(r_k - r_{k-1})^2 / r_k^2$ é uma estatística qui-quadrado que permite avaliar a falta de ajuste (*lack-of-fit*) do modelo ajustado.

96 A razão $\frac{r_k}{\hat{y}_k}$ é denominada resíduo padronizado.

97 $\hat{\sigma}^2 = \sum_{k=2}^{10} \frac{r_k^2}{7}$.

A tabela ANOVA a seguir se refere ao ajuste de um modelo de regressão linear simples escrito como $y = a + bx + \varepsilon$, cujos coeficientes foram estimados pelo método da máxima verossimilhança, com $\varepsilon \sim \mathrm{N}(0, \sigma^2)$. Os erros em torno da reta esperada são independentes e identicamente distribuídos.

fonte de variação	graus de liberdade	soma de quadrados	quadrado médio
modelo	1	10	10
erro	99	990	10
total	100	1.000	10

Com base nessas informações, julgue os itens a seguir.

98 O coeficiente de explicação do modelo é igual a 0,99.

99 A variância amostral da variável dependente é inferior a 12.

100 O R^2 ajustado é maior ou igual a 0,05.

101 $\widehat{\sigma}^2 = 10$.

102 Para se testar a hipótese nula $H_0: y = a + \varepsilon$ contra a hipótese alternativa $H_1: y = a + bx + \varepsilon$, a estatística do teste F proporcionada pela tabela ANOVA é igual ou superior a 2.

número de coeficientes no modelo	$R^2_{ajustado}$	C_p de Mallows	BIC
1	0,6	3.200	−700
4	0,9	220	−1.760
8	0,92	17	−1.920
10	0,92	13	−1.915
12	0,92	16	−1.905

Considerando as informações apresentadas no quadro precedente, julgue os itens subsequentes, acerca de modelos de regressão linear.

103 O melhor modelo candidato apontado pelo critério BIC possui 8 coeficientes.

104 O melhor modelo candidato não necessariamente apresenta maior $R^2_{ajustado}$.

105 A vantagem da medida C_p de Mallows em relação às outras medidas para a modelagem dos dados por regressão linear é sua robustez frente a presença de muitos pontos influentes na amostra.

amostragem aleatória simples		tamanho da amostra (n)
I	com reposição	6
II	sem reposição	5

Suponha que determinada população de tamanho $N = 100$ seja constituída pelos elementos $x_1, ..., x_{100}$. Para a realização de um levantamento amostral sobre essa população, cogitam-se duas possibilidades mostradas no quadro anterior, ambas pelo método de amostragem aleatória simples. Se o tipo I for o escolhido, então a amostragem será com reposição com $n = 6$. No entanto, se o escolhido for o tipo II, então a amostra será sem reposição com $n = 5$.

Com base nessas informações, julgue os itens que se seguem.

106 Na amostragem do tipo I, a probabilidade de que o elemento da população x_{20} constitua a amostra de tamanho $n = 6$ é igual a 0,09.

107 Se o tipo II for aplicado, a probabilidade de que a amostra seja formada pelos elementos $x8$, $x27$, $x70$, $x77$, $x99$ é igual a $\binom{100}{5}^{-1}$.

108 Suponha que a variância populacional seja definida por

$$S^2 = \sum_{i=1}^{100} \frac{(x_i - \bar{x})^2}{99},$$

em que $\bar{x} = \sum_{i=1}^{100} x_i/100$. Nesse caso, se a média da amostra aleatória simples com reposição (tipo I) for representada por $\bar{X} = \sum_{i=1}^{6} X_i/6$, então $\mathrm{Var}(\bar{X}) = S^2/6$.

109 Na amostragem do tipo II, a fração amostral é igual a 0,05.

110 Suponha que $X1, X2, X3, X4, X5$ sejam variáveis aleatórias que representam a amostra a ser obtida pela amostragem do tipo II. Nesse caso, é correto afirmar que essas variáveis aleatórias são mutuamente independentes.

Considere uma população formada pelos elementos $x_1, ..., x_N$, cuja média populacional é representada por $\mu = \sum_{i=1}^{N} x_i/N$. A amostra aleatória de tamanho simples n retirada dessa população é denotada por $X_1, ..., X_N$ (com $1 < n < N$), tal que a média amostral seja definida por

$$\sum_{i=1}^{n} \frac{X_i}{n} = \sum_{i=1}^{N} \frac{a_i x_i}{n},$$

em que $\{a_1, ..., a_N\}$ forma uma sequência de variáveis aleatórias tais que $a_i \sim$ Bernoulli $\left(\frac{n}{N}\right)$ e $\sum_{i=1}^{N} a_i = n$. Considerando essas informações, julgue os próximos itens.

111 $\sum_{i=1}^{n} \frac{X_i}{n}$ é um estimador não viciado da média populacional μ.

112 $\{a_1, ..., a_N\}$ forma uma sequência de variáveis aleatórias independentes e identicamente distribuídas.

113 A situação em tela representa uma amostragem aleatória simples com reposição.

estrato h	N_h	s_h	$N_h s_h$	n_h
1	50	3	150	n_1
2	50	2	100	n_2
3	100	1	100	n_3
totais	200	---	350	n

A tabela precedente mostra informações para a determinação do tamanho amostral n referente a um levantamento por amostragem aleatória estratificada com alocação proporcional ao tamanho do estrato, em que N_h representa o tamanho do estrato h e s_h, o desvio padrão amostral no estrato h referente a uma variável de interesse X a ser estudada nesse levantamento. O objetivo do levantamento é produzir uma estimativa da média populacional de X com base no estimador usual $\bar{X}_{\mathrm{estrat}}$ da amostragem aleatória estratificada, cuja variância é representada por $V = \mathrm{Var}(\bar{X}_{\mathrm{estrat}})$. Tendo como referência essas informações, julgue os itens a seguir.

114 Considerando-se que $n = 80$, se V_0 for a variância do estimador $\bar{X}_{\mathrm{aas}}$ propiciado pela amostragem aleatória simples para a estimação da média populacional de X, então $V \leq V_0$.

115 Se $V = 0{,}03$, então $n < 80$.

116 Se $n = 100$, então $n_1 = n_2 = 25$ e $n_3 = 50$.

117 Se n_0 representa o tamanho da amostra obtido sem a correção para população finita (*finite population correction*), então é correto afirmar que $n_0 > n$.

Uma pesquisa de opinião foi realizada para se estimar o percentual de funcionários da empresa A que estão satisfeitos com certo serviço prestado por uma empresa terceirizada B. Cada funcionário atua em uma única equipe de trabalho, sendo que existem 500 equipes de trabalho na empresa A. Para essa pesquisa, 50 equipes foram selecionadas por amostragem aleatória simples. Todos os funcionários que constituem as equipes selecionadas foram entrevistados, perfazendo o total de 260 funcionários entrevistados. Desse total, 200 funcionários se manifestaram satisfeitos com o serviço.

Com respeito a essa situação hipotética, julgue os itens seguintes.

118 A técnica descrita no texto para a estimação do percentual de funcionários da empresa A que estão satisfeitos com o serviço prestado por B refere-se à amostragem aleatória simples.

119 Cada funcionário representa uma unidade amostral e, por isso, o tamanho da amostra foi igual a 260 funcionários.

120 Se P representa a estimativa do percentual de funcionários da empresa A que estão satisfeitos com o serviço prestado pela empresa B, então $P > 80\%$.

Espaço livre