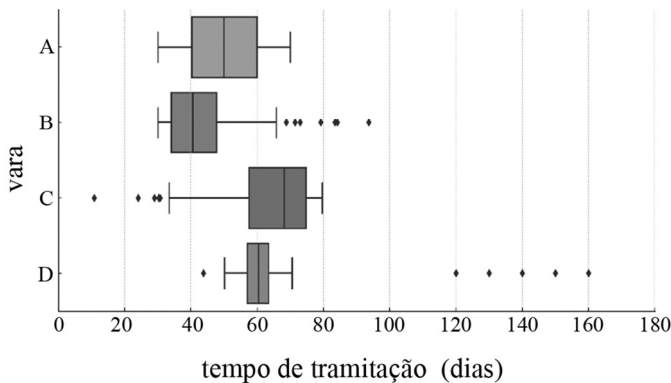


-- CONHECIMENTOS ESPECÍFICOS --

Em um estudo estatístico sobre o tempo médio de tramitação de processos judiciais em quatro varas cíveis de certo tribunal, considerou-se uma base de dados referentes ao tempo (em dias) de tramitação de 1.200 processos. A amostra foi estratificada por vara, e os dados foram tratados e analisados por meio de gráficos, tabelas e medidas descritivas, com o objetivo de identificar padrões, desvios e oportunidades de melhoria na gestão processual. A seguir, são apresentados diagramas *boxplots* elaborados para cada vara, bem como as medidas de tendência central, dispersão e curtose.



medida	vara A	vara B	vara C	vara D
média	50	44	64	64
mediana	50	40,5	68	60
desvio padrão	12	13,5	14,5	18,5
curtose	-1,2	2,0	1,7	16,0

Com base nessas informações, julgue os próximos itens.

- 51 A curtose negativa observada na vara A indica que a distribuição dos tempos de tramitação nessa vara é menos achatada do que uma distribuição normal, apresentando maior concentração em torno da média.
- 52 Na vara B, o fato de a média ser superior à mediana, associado à curtose positiva, sugere que a distribuição dos tempos de tramitação possui cauda mais pesada à sua esquerda.
- 53 A distribuição dos tempos de tramitação na vara C apresenta assimetria negativa.
- 54 Apesar de apresentarem médias iguais, as varas C e D diferem quanto à forma da distribuição: a curtose bastante elevada na vara D reflete a presença de pontos exteriores indicados no seu diagrama *boxplot*.
- 55 Entre as quatro varas analisadas, a vara C apresenta o menor coeficiente de variação.

Em um estudo sobre uma grande população de decisões judiciais proferidas em primeira instância, um analista avaliou a relação entre o tipo de processo (cível ou criminal) e o tempo decorrido entre o ajuizamento e a sentença. Sabendo que não é possível classificar um processo, ao mesmo tempo, como cível e criminal, o analista observou que 60% dos processos eram cíveis (evento C) e que 30% do total dos processos foram decididos em menos de 6 meses (evento M). Porém, considerando somente os processos cíveis, o analista observou que o percentual de processos decididos em menos de 6 meses diminuiu para 25%.

Com respeito a essa situação hipotética, julgue os itens que se seguem.

- 56 $P(M \cap C) = 0,6 \times 0,3$.

- 57 A probabilidade de um processo ser resolvido em 6 meses ou mais é igual a 0,7.
- 58 Se K representa o evento “processo criminal” e se C denota o evento “processo cível”, então K e C são eventos independentes.
- 59 Denotada por $P(C|M)$, a probabilidade condicional de um processo ter sido cível, dado que foi julgado em menos de 6 meses, é $P(C|M) = 0,5$.
- 60 Se 10 processos forem selecionados ao acaso, a probabilidade de se observar exatamente 6 processos cíveis e 4 processos criminais nessa amostra é 0,6.

Considerando duas variáveis aleatórias discretas X e Y , tais que $P(X = x, Y = y) = \frac{0,2 \cdot e^{-0,8} \cdot 0,8^{x+y}}{x!}$, em que $x \in \{0, 1, 2, 3, \dots\}$ e $y \in \{0, 1, 2, 3, \dots\}$, julgue os itens subsequentes.

- 61 $E[X] > E[Y]$.
- 62 X e Y são variáveis aleatórias independentes.
- 63 $E[X|Y = y] = E[X] + y$.
- 64 $P(Y \geq 3|X = 0) = 0,512$.
- 65 A variância de Y é inferior a 18.

Se V e W forem duas cópias independentes de uma distribuição exponencial com variância 1, então é correto afirmar que

- 66 a diferença $V - W$ segue uma distribuição exponencial com variância igual a 2.
- 67 a forma $2 \cdot (V + W)$ se distribui conforme uma distribuição exponencial com variância igual a 8.
- 68 a razão $\frac{V}{V+W}$ segue uma distribuição uniforme contínua.
- 69 o mínimo entre V e W segue uma distribuição exponencial com média igual a 0,5.
- 70 $\exp(-V)$ e $\exp(-W)$ são cópias independentes de uma distribuição uniforme contínua no intervalo $[0,1]$.

Espaço livre

Julgue os próximos itens, considerando que a distribuição condicional $Y|R=r$ segue uma distribuição gama com parâmetro de forma r e parâmetro de escala 1, e supondo que R segue uma distribuição geométrica com probabilidade de sucesso 0,5, tal que $r \in \{1, 2, 3, \dots\}$.

- 71 $E[Y|R=r] = 1/r$.
- 72 O valor esperado de Y é igual a 2.
- 73 A variância de Y é igual a 4.
- 74 A distribuição de R condicionalmente a um valor observado de Y , denotado como $R|Y=y$, segue uma distribuição binomial.
- 75 A distribuição do par (Y, R) pode ser representada por uma função de densidade absolutamente contínua $f(y, r)$, na qual (y, r) representa um ponto do suporte da distribuição de (Y, R) .

Acerca de estimadores pontuais, julgue os itens a seguir.

- 76 O método dos momentos e o método da máxima verossimilhança fornecem os mesmos estimadores.
- 77 Um estimador suficiente retém todas as informações relevantes da amostra a respeito do parâmetro populacional.
- 78 O estimador pontual da média populacional independe do tamanho da amostra.
- 79 Um estimador pontual consistente aproxima-se do parâmetro populacional quando o tamanho da amostra cresce.
- 80 Para cada parâmetro populacional, há um único estimador pontual.
- 81 O estimador pontual da média populacional é dado pela média aritmética da amostra.

Julgue os itens subsequentes, no que se refere a intervalos de confiança e à credibilidade de um parâmetro.

- 82 O intervalo de confiança não permite medir a magnitude de erro (ou erro de precisão) com relação ao parâmetro populacional.
- 83 Um intervalo de confiança permite a aproximação do parâmetro populacional por meio de intervalos.
- 84 O coeficiente de confiança mede a percentagem de intervalos de confiança construídos que contêm o parâmetro populacional.
- 85 Dada uma amostra, vários intervalos de confiança podem ser construídos.
- 86 A situação ideal na construção de intervalos de confiança é a obtenção de intervalos amplos que apresentem alta probabilidade de conter o parâmetro populacional.
- 87 Um intervalo de credibilidade é baseado em informações prévias.
- 88 A construção de um intervalo de credibilidade independe da distribuição de probabilidade.

Julgue os próximos itens, relativos a testes de hipóteses.

- 89 Testar hipóteses consiste, essencialmente, em avaliar a hipótese nula.
- 90 O nível de significância de um teste assume valores altos, tais como 90%, 95% e 99%.
- 91 A definição do nível de significância de um teste é baseada na probabilidade do erro do tipo II.
- 92 A escolha do nível de significância tem influência direta na potência de um teste.
- 93 A definição da potência de um teste é baseada na probabilidade do erro do tipo I.
- 94 O teste t de Student é um procedimento que compara as médias de duas amostras.
- 95 O teste qui-quadrado é um procedimento que permite analisar a associação entre variáveis numéricas.

Considerando que um modelo de regressão linear apresenta a forma $y = X \cdot \beta + \varepsilon$, em que y é o vetor resposta, X é a matriz de covariáveis, β é o vetor de parâmetros e ε é o erro do modelo, julgue os próximos itens acerca do estimador de mínimos quadrados (EMQ) e do estimador de máxima verossimilhança (EMV) para esse modelo.

- 96 Sob a suposição de erros com distribuição normal e variância constante, o EMQ e o EMV para β são idênticos.
- 97 O EMQ minimiza a soma do quadrado dos resíduos, independentemente da distribuição dos dados.
- 98 Para o EMV existir para uma regressão linear, a variável resposta deve seguir uma distribuição normal.
- 99 Quando os erros são homoscedásticos e normalmente distribuídos, os estimadores EMQ e EMV são não viesados e atingem o limite inferior de Cramér-Rao.
- 100 O EMQ requer menos suposições sobre distribuições que o EMV para ser consistente.

A respeito da inferência do modelo de regressão linear múltipla da forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$, julgue os itens subsequentes, considerando que valem as suposições clássicas, como linearidade nos parâmetros, independência entre observações, homoscedasticidade, normalidade e ausência de multicolinearidade perfeita.

- 101 A estatística t para testar $H_0: \beta_j = 0$ segue uma distribuição t de Student com $n - 1$ graus de liberdade.
- 102 Um intervalo de 95% de confiança para β_j que inclui o zero significa que se rejeita a hipótese $H_0: \beta_j = 0$ ao nível de 5% de significância.
- 103 A estatística F para testar $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ segue uma distribuição F com $(k, n - 1)$ graus de liberdade.
- 104 Se o erro padrão de β_j aumenta, mantendo as demais características constantes, o comprimento do intervalo de confiança para β_j diminui.
- 105 Para testar $H_0: \beta_1 + \beta_2 = 1$ usando um teste t , a estatística do teste é $\frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{EP(\hat{\beta}_1 + \hat{\beta}_2)}$, em que EP é o erro padrão, determinado por $EP(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$.
- 106 O p -valor representa a probabilidade de que a hipótese nula seja verdadeira condicional aos dados observados.

Espaço livre

Para um modelo de regressão linear múltipla na forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$, em que são válidas as suposições clássicas usuais, a tabela de ANOVA decompõe a variação total na resposta (y) conforme a seguir.

- SQT (soma de quadrados totais) = $\sum (y_i - \bar{y})^2$
- SQM (soma de quadrados do modelo) = $\sum (\hat{y}_i - \bar{y})^2$
- SQR (soma de quadrados do resíduo) = $\sum (y_i - \hat{y}_i)^2$

A partir dessas informações, e considerando que n denota o tamanho da amostra, julgue os itens seguintes acerca do uso da ANOVA para avaliar regressões lineares.

- 107** O coeficiente de determinação R^2 é igual à SQM/SQT, e incluir qualquer novo regressor (mesmo que estatisticamente insignificante) diminuirá o R^2 .
- 108** A estatística F para testar $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ é dada por $F = \frac{SQM/k}{SQR/(n-k-1)}$.
- 109** O erro quadrático médio é igual à SQR/($n - k - 1$) e provê um estimador não viesado da variância do erro, independentemente dos regressores terem efeitos significantes.
- 110** Quando comparados os modelos alinhados usando ANOVA, a estatística F para testar a significância de q regressores adicionais é $F = \frac{(SQR_{restrito} - SQR_{irrestrito})/q}{SQR_{irrestrito}/(n-k-1)}$, o qual segue uma distribuição $F(q, n - k - 1)$ sob a hipótese nula.

Para validar as suposições clássicas usuais e detectar potenciais violações de um modelo de regressão linear múltipla na forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$, será feita uma análise de resíduos, a partir dos dados a seguir.

- Resíduo: $e_i = y_i - \hat{y}_i$.
- Resíduos padronizados: $\frac{e_i}{\hat{\sigma}}$, em que $\hat{\sigma} = \sqrt{QMR}$ e QMR é o quadrado médio do resíduo.
- Resíduos estudatizados: $\frac{e_i}{\sqrt{QMR \cdot (1 - h_{ii})}}$, em que h_{ii} é o i -ésimo elemento da diagonal da matriz *hat*.

Considerando as informações precedentes, julgue os itens que se seguem.

- 111** Em um modelo de regressão bem-ajustado, os resíduos devem mostrar padrões sistemáticos quando plotados contra os valores ajustados $\hat{y}_i$.
- 112** A soma dos resíduos em um modelo de regressão com um intercepto é sempre igual a zero: $\sum e_i = 0$.
- 113** Se um gráfico de probabilidade normal (Q-Q *plot*) dos resíduos mostra pontos que seguem aproximadamente uma linha reta, isso indica que a suposição de normalidade não é satisfeita.
- 114** A heteroscedasticidade pode ser detectada plotando-se os resíduos contra os valores ajustados e observando-se se a dispersão dos resíduos aumenta ou decresce sistematicamente com relação aos valores ajustados.
- 115** Os resíduos estudatizados são preferidos aos resíduos ordinários para a detecção de *outliers*, pois consideram o efeito alavanca de cada observação e possuem variâncias mais uniformes entre as observações.
- 116** A estatística do teste de Durbin-Watson, calculada como $DW = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}$, é utilizada para testar as correlações de primeira ordem nos resíduos, com valores próximos de 2 indicando ausência de autocorrelação.

Julgue os itens a seguir a respeito de técnicas de amostragem.

- 117** Na amostragem estratificada, o tamanho amostral em cada estrato deve ser proporcional ao tamanho do estrato na população.
- 118** A amostragem por conglomerados sempre produz estimativas mais precisas que a amostragem aleatória simples quando os conglomerados são internamente homogêneos e heterogêneos entre os conglomerados.
- 119** Na amostragem sistemática com um elemento inicial escolhido ao acaso, o intervalo de seleção k deve sempre ser calculado como $k = \frac{N}{n}$ arredondado para o inteiro imediatamente anterior para evitar viés de seleção.
- 120** A amostragem em múltiplos estágios requer que cada estágio use o mesmo método amostral (por exemplo, todos os estágios devem usar a amostragem aleatória simples) para manter a estrutura probabilística.

Espaço livre