

CONHECIMENTOS ESPECÍFICOS

Considere que o tempo de duração (X , em horas) de uma viagem por via ferroviária seja uma variável aleatória com função de densidade de probabilidade expressa por $f(x) = 2e^{-2(x-5)}$ em que $x \geq 5$ horas. Com base nessas informações, julgue os próximos itens.

- 51 Na situação em questão, é impossível observar o evento $[X < 2]$.
- 52 Se uma amostra aleatória simples for retirada dessa distribuição, então o coeficiente de assimetria produzida por essa amostra será negativo.
- 53 Se $\{X_1, X_2, \dots, X_{100}\}$ forem cópias estocásticas independentes de X , então a mediana amostral desse conjunto será igual a $0,5 \ln 2$.
- 54 A moda da distribuição X é superior a 6 horas.
- 55 O tempo médio de duração da viagem em questão é de 5,5 horas.

Considere que a função geradora de momentos de uma variável aleatória discreta X seja dada pela relação $M_X(q) = (0,8 + 0,2e^q)^2$, em que $q \in \mathbb{R}$. Com base nessas informações, julgue os itens a seguir.

- 56 A variância de X é igual a 0,32.
- 57 A probabilidade de ocorrência do evento $[X = 0]$ é igual a 0,64.

Suponha que X e Y sejam variáveis aleatórias de um mesmo espaço amostral e que $E(X|Y=y) = Var(X|Y=y) = 4y^2$ em que Y segue uma distribuição normal com média zero e desvio padrão 1. Com base nessas informações, julgue os seguintes itens.

- 58 As variáveis aleatórias X e Y são dependentes e possuem correlação linear estritamente positiva.
- 59 O valor esperado da variável aleatória X é inferior a 2.
- 60 O desvio padrão de X é igual a 6.

Considere que a distribuição das velocidades v dos veículos (em km/h) em uma via seja uma variável aleatória V com função de distribuição acumulada $F(v) = \frac{1}{1 + e^{-(v-100)}}$. A partir dessas informações, julgue os itens que se seguem.

- 61 O desvio padrão da distribuição das velocidades dos veículos nessa via é superior a 20 km/h.
- 62 A função de densidade da distribuição de V é igual a $f(v) = \frac{e^{-|v-100|}}{2}$.
- 63 A média das velocidades dos veículos nessa via é de 100 km/h.

Considere que a quantidade de carga perdida (Y , em kg) em operações de transbordo seja uma variável aleatória com função de densidade de probabilidade $f(y) = \frac{2(y-a)}{(b-a)^2}$ em que $0 < a \leq y \leq b$. Considere, ainda, que Y_1, Y_2, \dots, Y_n representa uma amostra aleatória simples retirada dessa distribuição e que $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ representam suas estatísticas de ordem. Com base nessas informações, julgue os itens de 64 a 67.

- 64 As estatísticas de ordem $Y_{(1)}$ e $Y_{(n)}$ são os estimadores de máxima verossimilhança dos parâmetros a e b , respectivamente.

- 65 A moda amostral é um estimador do parâmetro b .
- 66 A média amostral (\bar{Y}) é um estimador não tendencioso da distribuição de perdas Y . Nesse caso, o valor esperado de \bar{Y} é igual a $\frac{a+b}{2}$.
- 67 Com base apenas na média amostral, não é possível obter estimativas dos parâmetros a e b pelo método dos momentos.

RASCUNHO

RASCUNHO

Os testes estatísticos são bastante úteis na etapa de diagnósticos do processo de modelagem estatística de dados, pois permitem avaliar aspectos como independência, normalidade, homogeneidade e aderência dos dados, entre várias outras hipóteses. Considerando que X e Y representam variáveis quantitativas e que A e B denotam variáveis qualitativas, julgue os seguintes itens, a respeito de testes de hipóteses.

- 68 Pode-se testar a normalidade de uma variável X , por meio de diversos testes, como, por exemplo, o de Jarque-Bera, o de Anderson-Darling, o de Cramér-von Mises, o de Lilliefors, o de Kolmogorov-Smirnov e o de Shapiro-Wilk.
- 69 Caso os resultados do cruzamento entre as variáveis A e B encontrem-se em uma tabela de contingência 3×2 com totais marginais fixos, a hipótese de independência poderá ser testada pelo método da razão de verossimilhança.
- 70 Não é possível aplicar a estatística qui-quadrado de Pearson para testar a hipótese de independência entre X e Y .
- 71 O teste bilateral de Kolmogorov-Smirnov é um método não paramétrico que permite avaliar a hipótese de independência entre as variáveis X e Y .

De acordo com determinada pesquisa, o tempo médio de espera de um passageiro em uma rodoviária, desde a chegada ao terminal até o embarque no ônibus, é de 20 minutos. O desvio padrão dos tempos de espera é, também, de 20 minutos e o tamanho da amostra dessa pesquisa é $n = 900$. Considerando que a pesquisa tenha sido feita por amostragem aleatória simples, julgue os itens a seguir.

- 72 O fato de o tempo médio amostral ser igual ao valor do desvio padrão amostral permite concluir que a distribuição dos tempos de espera é exponencial.
- 73 Considere que, para testar a hipótese nula $H_0: \mu = 15$ minutos, em que μ representa o tempo médio de espera da população, o valor da estatística do teste t seja igual a 7,5. Nesse caso, assumindo-se que sejam válidas todas as condições exigidas para a aplicação desse teste, não há evidências estatísticas contra a hipótese nula.
- 74 Nessa situação, o erro padrão do estimador do tempo médio populacional é inferior a 0,5 minutos.
- 75 Com 95% de confiança, a estimativa intervalar do tempo médio de espera de um passageiro na população equivale a 20 minutos \pm 20 minutos.

parâmetro	estimativa	erro padrão	razão t	p -valor
β_0	60	6,0	10,0	0,00000
β_1	0,8	0,2	4,0	0,00007
β_2	3,6	2,0	1,8	0,07218
β_3	- 0,10	0,05	- 2,0	0,04578

R. Cordeiro, E. Lima Filho, F. M. Fischer e D. C. Moreira Filho. Associação da pressão arterial diastólica com o tempo acumulado de trabalho entre motoristas e cobradores. *In: Rev. Saúde Pública*, 1993, vol. 27, n.º 5, p. 363-372 (com adaptações).

Um estudo para investigar a associação da pressão arterial diastólica com o tempo acumulado de trabalho dos motoristas de ônibus em determinada cidade considerou o modelo de regressão linear na forma $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$, em que y_i representa a pressão arterial diastólica (mmHg) do motorista i , X_{1i} é a idade (em anos) do motorista i , X_{2i} denota o logaritmo natural do tempo de trabalho (em meses) do motorista i e ε_i representa o erro aleatório com média nula e variância σ^2 . Esse estudo foi realizado com base em uma amostra aleatória de 1.000 motoristas de ônibus. A tabela acima apresenta a estimativa de cada parâmetro β_i ($i = 0, 1, 2, 3$) obtida pelo método de mínimos quadrados ordinários, o erro padrão, a razão t e o p -valor correspondentes.

Com base nessas informações e na tabela apresentada, julgue os itens a seguir.

- 76 Considerando-se o nível de significância de 5%, não se rejeita a hipótese $H_0: \beta_2 = 0$.
- 77 Para se obter a estimativa de um coeficiente do modelo pelo método de mínimos quadrados ordinários, exige-se que o erro aleatório ε_i siga uma distribuição normal com média 0 e variância σ^2 .
- 78 O produto $X_{1i} X_{2i}$, que se denomina interação, permite representar o efeito multiplicativo da idade e do logaritmo natural do tempo de trabalho na pressão arterial diastólica média de um motorista.
- 79 O estimador do coeficiente β_1 segue uma distribuição t de Student com 995 graus de liberdade.
- 80 Por meio do método estatístico análise de variância (ANOVA), é possível testar, por exemplo, a hipótese nula $\beta_1 = \beta_2 = \beta_3 = 0$.

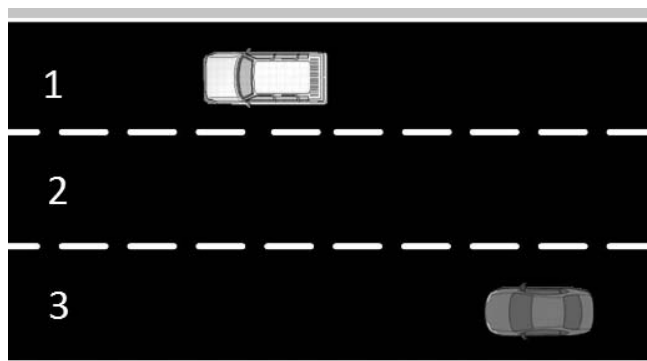
Com o objetivo de se estimar o percentual P de pessoas que utilizam, predominantemente, o sistema de transporte público (trens, metrô e ônibus) em uma grande região metropolitana, realizou-se uma pesquisa domiciliar com 2.500 domicílios selecionados por amostragem aleatória simples, registrando-se as informações acerca da utilização de transporte público de todos os residentes nesses domicílios, o que perfazia o total esperado de 10 mil pessoas. A partir dessas informações, julgue os itens seguintes.

- 81 Nesse caso, utiliza-se um estimador de razão para se fazerem inferências em relação ao percentual P .
- 82 Nessa pesquisa, cada residente participante representa uma unidade amostral primária.
- 83 O erro padrão do estimador do percentual P depende do número médio de residentes por domicílio, que, na situação em tela, corresponde a 4 pessoas por domicílio.

RASCUNHO

Considerando que um pesquisador, usando métodos computacionais, deseje estudar o impacto dos congestionamentos urbanos no consumo de combustível e no meio ambiente e que, para isso, deva gerar uma variável aleatória uniformemente distribuída no intervalo $[0,1]$ (U), uma variável aleatória normal padrão (Z) e uma variável aleatória exponencial com média unitária (Y), julgue os itens que se seguem.

- 84 A variável aleatória Y pode ser gerada pelo método da transformação integral, que produz a relação $Y = -\ln(1-U)$.
- 85 A variável Z pode ser obtida mediante a padronização da variável Y , ou seja, $Z = \frac{Y-\mu}{\sigma}$, em que μ e σ representam, respectivamente, a média e o desvio padrão de Y .
- 86 Uma realização da variável aleatória U pode ser gerada com base em um algoritmo computacional denominado Jackknife.



A figura acima apresenta um trecho de uma rodovia com três faixas de rolamento. Considere que $X(t)$ represente o número de veículos que passam nesse trecho durante um intervalo de tempo de duração t (em minutos) e que $X(t)$ siga um processo de Poisson com parâmetro $6t$, ou seja, $P(X(t) = x) = \frac{e^{-6t} (6t)^x}{x!}$. Suponha, ainda,

que, no intervalo t , cada veículo selecione aleatoriamente as faixas de rolamento 1, 2 e 3 com probabilidades 0,5; 0,3 e 0,2, respectivamente. Com base nessas informações, julgue os próximos itens.

- 87 O processo estocástico $X(t)$ é uma cadeia de Markov em tempo contínuo.
- 88 O número de veículos que passam nesse trecho pela faixa de rolamento 3 durante um intervalo de tempo de duração t (em minutos) segue um processo de Poisson com parâmetro $1,2t$.
- 89 O intervalo de tempo entre dois veículos sucessivos que passam pela faixa de rolamento 1 nesse trecho segue uma distribuição exponencial com média igual a 3 minutos.

X_t é uma variável aleatória dicotômica que sinaliza a ocorrência ($X_t = 1$) ou a não ocorrência ($X_t = 0$) de incidentes em certo terminal rodoviário de cargas no dia t , $t \in \{0,1,2, \dots\}$. Essa variável segue uma cadeia de Markov em tempo discreto, cuja probabilidade de transição é

$$P(X_{t+1} = a | X_t = b) = \frac{a^b + 1}{b + 2},$$

em que a e b podem assumir valores 0 ou 1.

Com base nessas informações e assumindo que $0^0 = 1$, julgue os itens a seguir.

- 90 No regime estacionário, o valor esperado da variável aleatória X_t é igual ou inferior a 0,5.
- 91 A esperança condicional $E(X_{t+1} | X_t = b) = \frac{1}{b+2}$ representa a reta de regressão de X_{t+1} em b .

Considerando que $W(t)$ represente um processo gaussiano com $E[W(t)] = 0$ e $Var[W(t)] = t$, em que $t \geq 0$, julgue os próximos itens.

- 92 Se $s < t$, então a função de covariância desse processo será $Cov[W(s), W(t)] = t - s$.
- 93 O processo $W(t)$ não é estacionário.
- 94 Dada uma malha temporal $0 < t_1 < t_2 < \dots < t_n$, é correto afirmar que as variáveis aleatórias $W(t_1), W(t_2), \dots, W(t_n)$ seguem, conjuntamente, uma distribuição normal multivariada.

RASCUNHO

Para avaliar o desempenho do transporte público por ônibus em determinada cidade, realizou-se um estudo estatístico mediante o uso de técnicas de análise multivariada de dados. Por meio desse estudo, foram identificados os grupos (homogêneos) de usuários, considerando-se a satisfação global dos serviços de transporte público, assim como os principais fatores que influenciam na opinião sobre esses serviços. O estudo identificou, por exemplo, aspectos como confiabilidade, segurança, tarifa e locais de parada como os mais importantes para se discriminar os usuários insatisfeitos daqueles que se consideram satisfeitos.

No que se refere aos métodos estatísticos de análise multivariada empregados na situação descrita acima, julgue os seguintes itens.

- 95 Empregando-se a análise discriminante, é possível separar estatisticamente os usuários insatisfeitos daqueles que se consideram satisfeitos, com base nas características do usuário. Essa técnica é uma forma especializada de regressão em que se ajusta a probabilidade de um indivíduo pertencer a um grupo ou a outro grupo com base no seu perfil (como, por exemplo, idade, gênero, renda e escolaridade).
- 96 A análise de conglomerados é a técnica que permite agrupar os usuários segundo o grau de satisfação com os serviços de transporte público.
- 97 Por meio da análise de correspondência, é possível representar as relações existentes em um conjunto de dados quantitativos com base em uma árvore de decisão. Essa técnica permite associar os aspectos confiabilidade, segurança, tarifa e locais de parada com o grau de satisfação dos usuários do serviço público de transporte.

Considere que uma série temporal $\{X_t\}$ seja gerada por

$$X_t = \left(\sum_{j=0}^{+\infty} \frac{\theta^j B^j}{j!} \right) Z_t,$$

em que B representa o operador de atraso (*backshift*), tal que $BZ_t = Z_{t-1}$, θ seja uma constante real e Z_t seja um ruído aleatório com média nula e variância unitária. Acerca da série temporal $\{X_t\}$, julgue os itens subsecutivos.

- 98 A média e a variância do processo $\{X_t\}$ são, respectivamente, iguais a 0 e 1.
- 99 Se $\theta = 5$, o processo $\{X_t\}$ não é estacionário.
- 100 A função de densidade espectral da série temporal $\{X_t\}$ é dada por $f(\omega) = \frac{e^{\theta\omega}}{2\pi}$ em que $|\omega| < \pi$.

Considerando que o modelo de regressão por mínimos quadrados ordinários (MQO) seja dado por $\hat{y}_t = \hat{a} + \hat{\beta} x_t$, em que \hat{y}_t é o valor estimado pelo modelo para a variável dependente y_t , x_t é a variável independente e \hat{a} e $\hat{\beta}$ são, respectivamente, os estimadores dos coeficientes linear a e angular β de um modelo de regressão linear simples, julgue os itens a seguir.

- 101 Ao se multiplicar a variável dependente por uma constante c qualquer, as estimativas de MQO são multiplicadas por c , isto é, \hat{a} e $\hat{\beta}$ são multiplicados por c .
- 102 O estimador $\hat{\beta}$ pode ser escrito da seguinte forma: $\hat{\beta} = \frac{Var(x_t)}{Cov(x_t, y_t)}$, em que $Var(x_t)$ é a variância de x_t e $Cov(x_t, y_t)$ é a covariância de x_t e y_t .
- 103 Na regressão pela origem $\hat{y}_t = \hat{\beta} x_t$, em que $\hat{a} = 0$, $\hat{\beta}$ é um estimador não viesado de β .

RASCUNHO

Acerca das propriedades dos estimadores de MQO em regressão linear simples, julgue os itens subsequentes.

- 104 Se $E(u|x) > 0$, em que u é o resíduo e x é a variável explicativa de um modelo de regressão linear simples, então as estimativas de MQO serão viesadas.
- 105 Para o coeficiente angular β , o estimador de MQO $\hat{\beta}$ apresenta uma componente não aleatória, β , e outra componente aleatória, a qual depende da covariância $Cov(x_i, u_i)$, tal que $\hat{\beta} = \beta + \frac{Cov(x_i, u_i)}{Var(x_i)}$, em que u_i é o resíduo da regressão e $Var(x_i)$ é a variância da variável independente x_i do modelo de regressão.
- 106 De acordo com a hipótese de consistência do estimador de MQO, à medida que o número de observações aumenta, o valor esperado do estimador converge para o valor do parâmetro a ser estimado e a variância do estimador converge para zero.
- 107 Se o estimador de MQO for não viesado e consistente, então ele será, necessariamente, eficiente.
- 108 A suposição de homocedasticidade é fundamental para mostrar que os estimadores de MQO são não viesados.

A respeito do método de estimação por MQO, julgue os itens que se seguem.

- 109 Na análise de séries temporais, a suposição de ausência de autocorrelação serial dos resíduos deve sempre ser verificada para garantir que os estimadores de mínimos quadrados ordinários sejam não viesados e consistentes.
- 110 Um elevado valor da estatística R^2 em um modelo de regressão linear simples com uma variável independente x e uma variável dependente y implica, necessariamente, causalidade entre y e x .
- 111 A estatística R^2 é utilizada como critério de seleção para diferentes formas funcionais de estimação de uma variável dependente y , podendo-se, por exemplo, mediante essa estatística, comparar o desempenho do modelo $y_i = a + \beta x_i + u_i$ com o desempenho do modelo $\ln y_i = a + \beta \ln x_i + u_i$, em que x_i é a variável independente e u_i é uma variável aleatória de média igual a zero.

Julgue os itens seguintes, relativos à violação das suposições básicas dos modelos clássicos de regressão.

- 112 Uma vez detectada a presença de heterocedasticidade, é possível estimar o modelo por mínimos quadrados generalizados (MQG) para corrigir ou minimizar o problema, de tal forma que os estimadores de MQG sejam melhores que os estimadores de MQO.
- 113 Na presença de multicolinearidade, a variância e a covariância dos estimadores serão afetadas, sendo possível que sejam alterados tanto os sinais quanto a magnitude dos estimadores.
- 114 A violação da suposição de homocedasticidade dos resíduos afeta a distribuição de probabilidade dos estimadores sem afetar, contudo, o seu valor esperado.

Com relação aos modelos de séries temporais, julgue os próximos itens.

- 115 No modelo $ARMA(p, q)$, o teste de estacionariedade relaciona-se apenas à especificação $AR(p)$, sendo necessário verificar-se se o processo $MA(q)$ atende às propriedades de inversibilidade.
- 116 Um processo autorregressivo de ordem p , $AR(p)$, é estacionário somente se as p raízes da equação polinomial forem menores que um.
- 117 A suposição de estacionariedade não coloca restrição sobre a forma como x_t e x_{t-1} são relacionados entre si.
- 118 Para processos autorregressivos de ordem 1, $AR(1)$, a função de autocorrelação, $Cor(y_t, y_{t-k})$, apresenta decaimento exponencial à medida que k cresce, considerando-se que k é uma constante qualquer.

No que se refere aos aspectos econométricos relacionados à cointegração, julgue os itens subsequentes.

- 119 Os valores críticos do teste ADF (*Augmented Dickey-Fuller*) utilizados para verificar cointegração são os mesmos utilizados no teste de estacionariedade das séries temporais.
- 120 Duas séries não estacionárias $I(1)$ são ditas cointegradas se o resíduo da regressão $y_t = a + \beta x_t + u_t$ for integrado de primeira ordem, ou seja, $u_t \sim I(1)$ com $y_t \sim I(1)$ e $x_t \sim I(1)$.

RASCUNHO

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 serão avaliados fragmentos de texto escritos em locais indevidos**.
- Qualquer fragmento de texto que ultrapassar a 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 **20,00 pontos**, dos quais até **1,00 ponto** será atribuído ao quesito apresentação e estrutura textual (legibilidade, respeito às margens e indicação de parágrafos).

Uma determinada cidade será submetida a um levantamento estatístico com o objetivo de avaliar o sistema de transporte coletivo urbano, realizado por ônibus, com base em indicadores de qualidade dos serviços oferecidos pelos operadores do setor. A análise será constituída por opiniões de usuários e por referenciais quantitativos a respeito dos seguintes aspectos: número de linhas de ônibus, número de assentos disponíveis, tamanho da frota de veículos, frequência de acidentes e de falhas mecânicas e tempo de viagem. Os resultados dessa pesquisa devem fornecer um diagnóstico a respeito da qualidade do serviço de transporte coletivo oferecido nessa cidade, para subsidiar a formulação de propostas de melhoria dos serviços.

Considerando essa situação hipotética, redija um texto dissertativo acerca do seguinte tema.

A UTILIZAÇÃO DOS MÉTODOS ESTATÍSTICOS PARA DIAGNÓSTICOS NO SETOR DE TRANSPORTES

Ao elaborar seu texto, aborde, necessariamente, os seguintes aspectos:

- ▶ planejamento da pesquisa de opinião para avaliar o sistema de transporte coletivo; [valor: 6,50 pontos]
- ▶ coleta dos dados concernentes aos referenciais quantitativos do sistema de transporte coletivo; [valor: 6,50 pontos]
- ▶ técnicas estatísticas que podem ser empregadas para a elaboração de indicadores de qualidade. [valor: 6,00 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	