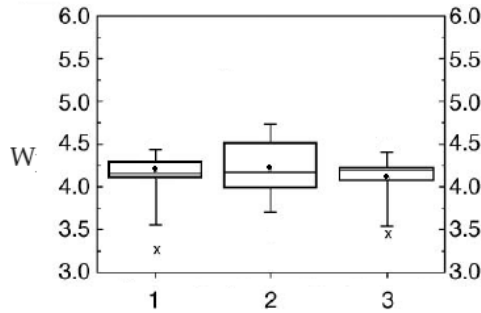


## CONHECIMENTOS ESPECÍFICOS



C. Brabec, J. Hauch, P. Schilinsky e C. Waldauf. **Production aspects of organic photovoltaics and their Impact on the commercialization of devices.** MRS Bulletin, 30, 2005 (com adaptações).

RASCUNHO

A figura acima apresenta os diagramas esquemáticos (*box-plot*) referentes às distribuições das eficiências (*W*) proporcionadas por três diferentes tipos de células fotovoltaicas. Nesses diagramas, o ponto no interior de cada caixa (*box*) refere-se à média amostral da distribuição. Com base nessas informações, julgue os itens que se seguem.

- 51 Em cada *box-plot*, a caixa ou a área retangular do diagrama corresponde a 75% da distribuição das eficiências.
- 52 Os diagramas correspondentes aos tipos 1 e 3 apresentam valores atípicos (ou *outliers*) na cauda inferior, o que sugere que ambas as distribuições possuem assimetria negativa.
- 53 O primeiro quartil de cada distribuição é superior a 3,75 e inferior a 4,25.
- 54 Em cada distribuição, o desvio padrão é inferior a 2,0.

ano	produto 1		produto 2	
	preço	quantidade	preço	quantidade
2009 (ano-base)	10	4	5	8
2010	12	5	5	8
2011	15	4	5	10
2012	20	5	6	10

A tabela acima mostra os preços (em R\$) e as quantidades vendidas (em mil unidades) de dois produtos, no período de 2009 a 2012. A partir desses dados, julgue os itens seguintes, acerca de números índices.

- 55 O índice de Laspeyres permite relacionar o custo de aquisição dos produtos 1 e 2 em 2010 — nas quantidades vendidas em 2010 — com o custo de aquisição desses mesmos produtos no ano-base 2009. Portanto, o valor desse índice em 2010 foi superior a 120%.
- 56 Tendo como base o ano de 2009, o valor do índice de Paasche em 2011 foi superior a 120% e inferior a 124%.
- 57 Tendo como base o ano de 2009, o preço relativo ao produto 1 em 2012 foi igual a R\$ 10,00.
- 58 O índice agregativo de preços (ou índice simples de preços agregados) para esses produtos em 2011, tendo 2009 como ano-base, é superior a 130%.
- 59 Com relação ao ano de 2010, o produto do índice de preços de Fisher e do índice de quantidades de Fisher foi igual ao índice relativo de valor de vendas.

## RASCUNHO

A quantidade diária de reclamações —  $X$  — recebidas em uma central de atendimento ao consumidor segue uma distribuição de Poisson, em que  $P(X = 0) = 0,5$ . Nesse sentido, julgue os próximos itens.

**60** A probabilidade condicional  $P(X = 1 | X > 0)$  é igual à variância de  $X$ .

**61** O valor esperado de  $X$  é igual a  $\ln 2$ .

Para cada  $t = 0, 1, 2, \dots, n$ , em que  $n \geq 1$ ,  $X_t$  representa o estado de um sistema de comunicação no instante  $t$ . Se  $X_t = 0$ , o sistema está inoperante no instante  $t$ , e se  $X_t = 1$ , o sistema esteja operante no instante  $t$ . Considerando que as variáveis aleatórias que formam a sequência de estados  $\{X_0, X_1, \dots, X_n\}$  sejam independentes e identicamente distribuídas, e que  $P(X_t = 1) = 0,9$  e  $P(X_t = 0) = 0,1$ , julgue os seguintes itens acerca da soma  $Y_n = X_0 + X_1 + \dots + X_n$ .

**62** A distribuição condicional  $Y_3 | \{X_0 = 0, X_1 = 0, X_2 = 0\}$  é uma distribuição de Bernoulli cuja probabilidade de sucesso é  $0,9$ .

**63** A variável aleatória  $Y_n$  segue uma distribuição binomial cuja média é igual a  $0,9 \times (n + 1)$ .

Considerando que a distribuição dos tempos de reparo —  $Y$  — de um sistema de telecomunicações siga uma distribuição exponencial com média igual a 1 hora, julgue os itens a seguir.

**64** A função  $F(y) = e^{-y}$  é a função de distribuição acumulada da variável aleatória  $Y$ .

**65** Considerando o evento  $A_n = [Y > n]$ , em que  $n = 0, 1, \dots$ , é correto afirmar que  $P(A_1 | A_2) = P(A_1)$ , o que permite caracterizar a distribuição  $Y$  como “sem memória”.

**66** O desvio padrão dos tempos de reparo desse sistema de telecomunicações é superior a 1,5 hora.

O ruído gerado por equipamentos na transmissão de um sinal de rádio é uma variável aleatória  $R$  que segue uma distribuição normal, com média zero e desvio padrão igual a 2. O ruído gerado por outras fontes (como aqueles devido às condições meteorológicas) é uma variável aleatória normal  $S$ , com média zero e desvio padrão igual a 5. Com base nessas informações, julgue os itens que se seguem, considerando que  $R$  e  $S$  sejam independentes.

**67** A média e a variância da distribuição  $S^2$  são, respectivamente, iguais a 25 e 1,625.

**68** A soma  $R + S$  segue uma distribuição normal com média 0 e desvio padrão superior a 6.

Suponha que  $X$  seja uma variável aleatória uniforme no intervalo  $[0, 2]$  e que  $Y$  seja uma variável aleatória geométrica, tal que  $P(Y = k) = 0,4 \times 0,6^k$ , em que  $k \geq 0$ . Considerando que  $X$  e  $Y$  sejam independentes, julgue os próximos itens.

**69** Se  $B_n = [X < 1 + 1/n]$  e  $C_n = [Y = n]$ , em que  $n = 0, 1, \dots$ , então  $P(B_n \cap C_n) = 0,2 \times (1 + 1/n) \times 0,6^n$ .

**70** O valor esperado do produto  $XY$  é inferior a 1,4.

Uma amostra aleatória de números telefônicos foi selecionada para se averiguar a qualidade do serviço de Internet 3G móvel oferecida pelas operadoras de telefonia celular no Brasil. Nessa situação, julgue os itens subsecutivos.

RASCUNHO

- 71** Caso os números telefônicos tenham sido selecionados por amostragem sistemática e estejam dispostos em ordem crescente e caso cada usuário tenha recebido seu número telefônico de forma aleatória, é correto afirmar que o erro amostral associado a esse tipo de amostragem é equivalente ao erro amostral proporcionado pela amostragem aleatória simples.
- 72** Considerando que cada usuário entrevistado deva responder se está ou não satisfeito com os serviços prestados pela sua operadora e que  $P(Z < 1,96) = 0,975$ , em que  $Z$  represente a distribuição normal padrão, se o nível de confiança for de 95% — simétrico e conservador —, o erro amostral for de até 5% e o plano de amostra aleatória simples com reposição, então o tamanho da amostra deve ser de, pelo menos, 385 números telefônicos.
- 73** Se o erro amostral da estimativa do percentual de satisfeitos for fixado em até 5% para cada operadora, então o plano de amostragem aleatória estratificada, em que cada operadora constitui um estrato, permite estabelecer essa taxa de erro amostral por operadora.
- 74** Considere uma pequena operadora com apenas 500 números telefônicos cadastrados e uma amostragem aleatória simples sem reposição, com margem de erro de até 10% e 95% de confiança. Nesse caso, o tamanho mínimo da amostra — mediante correção para populações finitas — será inferior a 80 números telefônicos.
- 75** Para a aplicação do estimador do tipo razão com a finalidade de se estimar o percentual de usuários satisfeitos, é possível utilizar as informações do IBGE acerca do tamanho da população brasileira como covariável.
- 76** Se o planejador da pesquisa utilizasse o plano amostral aleatório simples sem reposição, o efeito desse planejamento ( $Deff$ ) seria inferior a 1.
- 77** Se a amostragem for estratificada segundo o gasto do usuário com o serviço 3G de Internet móvel, a variância do estimador do percentual  $P$  de clientes satisfeitos será superior à variância desse percentual  $P$  associado ao plano de amostragem aleatória simples.

FV	gl	SQ	QM	F
regressão			810	
resíduo	98			
total			80	

O quadro acima mostra parte de uma tabela de análise de variância (ANOVA), que resultou da regressão linear simples do tempo que um usuário permanece conectado à Internet 3G —  $Y$ , em minutos — sobre a renda —  $X$  — declarada por esse usuário. Os dados utilizados nesse ajuste pelo método de mínimos quadrados ordinários foram selecionados por amostragem aleatória simples de um cadastro de usuários. Com base nessas informações e no quadro apresentado, julgue os itens seguintes acerca dos conceitos de análise de regressão, correlação e amostragem.

- 78** O método de mínimos quadrados ordinários para a estimação dos coeficientes do modelo requer a suposição de normalidade dos resíduos.
- 79** O tamanho da amostra é de 99 usuários.
- 80** A variância de  $Y$  é superior a 75.
- 81** O coeficiente de correlação de Pearson é superior a 0,25.
- 82** O valor da estatística  $t$  para o coeficiente angular do modelo é inferior a 4.
- 83** O teste  $t$  utilizado para testar se a significância do coeficiente angular é igual a zero possui 99 graus de liberdade.
- 84** A variância residual é maior que a variância de  $Y$ .
- 85** O gráfico dos resíduos permite diagnosticar a hipótese de sua normalidade, ou seja, com base nesse gráfico, é possível efetuar uma análise confirmatória.
- 86** No estimador de máxima verossimilhança da variância residual, o denominador é igual a 100.

Um auditor, em atendimento a reclamações de usuários, decidiu verificar se as velocidades da Internet disponibilizadas estavam de acordo com as respectivas velocidades contratadas. Para isso, ele coletou uma amostra de 500 usuários. Com referência a essa situação, julgue os itens de **87** a **94**.

- 87** O intervalo de confiança para a média requererá o uso da distribuição qui-quadrado com 1 grau de liberdade para o caso em que a variância populacional for desconhecida.
- 88** Considere que os usuários tenham sido classificados segundo as velocidades contratadas de 1 MB, 5 MB e 10 MB, que esses três grupos de usuários correspondem aos estratos da pesquisa e que a distribuição percentual de usuários nesses estratos seja de 50%, 30% e 20%, respectivamente. Nessa situação, em uma amostragem aleatória estratificada com alocação proporcional, deveriam ser selecionados 100 usuários do grupo dos contratantes da velocidade de 5 MB.
- 89** Suponha que, com referência aos usuários que haviam contratado velocidade de 1 MB, o auditor observou que a velocidade média amostral foi de 0,85 MB com erro padrão da média de 0,1 MB. Nesse caso, considerando-se que a distribuição das velocidades seja normal, é correto inferir, com 95% de confiança, que a velocidade medida está de acordo com a velocidade contratada.
- 90** Um teste de comparações múltiplas será o indicado para a amostra coletada devido a finalidade do trabalho de auditoria.
- 91** Se a variância populacional da variável de interesse for desconhecida, não será possível aplicar o teste paramétrico recomendado para a análise desses dados.

- 92** O erro padrão da mediana amostral será igual a  $\sqrt{\frac{\sigma^2 \pi}{n 2}}$ , em que  $\sigma$  é o desvio padrão populacional e  $n$  é o tamanho da amostra.
- 93** O teste para proporção — em que é medida a fração do tempo que a velocidade observada se encontra abaixo da velocidade contratada — apresenta menor poder estatístico na verificação da adequação da velocidade que o teste para média, haja vista que o teste para proporção não permite captar a magnitude das diferenças entre a velocidade observada e a contratada.
- 94** O teorema central do limite (TCL) garante que, se  $n$  é grande, então a estatística menos sua média, dividida pelo erro padrão tende para a distribuição normal padrão.

RASCUNHO

RASCUNHO

antes	800	750	900	950	900	810	780	800	920	950
depois	980	890	1.000	1.000	960	890	950	1.000	990	1.000

O quadro acima mostra os resultados de um estudo realizado para avaliar o efeito provocado pela intensificação da fiscalização em serviços de comunicação de dados. Esses dados são relativos a velocidades de tráfego de dados (em bates) observadas em contratos com velocidades nominais de 1 MB. Das observações realizadas, dez correspondem ao período anterior à intensificação da fiscalização e as outras dez correspondem ao período posterior à intensificação. Com base nessas informações, considerando válida a hipótese de normalidade dos dados, julgue os itens a seguir.

- 95 Para que seja testado se as duas amostras possuem variâncias iguais, é correto utilizar o teste  $F$  com 9 graus de liberdade em todos os parâmetros necessários.
- 96 Independentemente de os dados serem pareados ou não, a hipótese nula do teste a ser utilizado para a comparação entre as médias é  $H_0: \mu_D - \mu_A = 0$ , em que  $\mu_D$  e  $\mu_A$  são as médias posterior e anterior, respectivamente.
- 97 Se os dados forem pareados, será recomendável a aplicação de um teste para amostras independentes.
- 98 Em caso de pareamento, o valor mínimo do vetor de diferenças  $D = Y_D - Y_A$  será inferior a 200, em que  $Y_D$  é a velocidade observada depois da intensificação da fiscalização e  $Y_A$  é velocidade observada antes da intensificação.
- 99 A estatística do teste de comparação de médias entre duas amostras pareadas possui 9 graus de liberdade.
- 100 Considerando que a duas amostras sejam independentes e que as variâncias populacionais sejam conhecidas, é correto empregar o teste  $t$  com 9 graus de liberdade para a comparação das suas médias.

Com relação à análise multivariada, julgue os itens seguintes.

- 101 Na análise de agrupamento hierárquico, é correto determinar a quantidade de grupos a serem formados com base em uma medida de dissimilaridade.

- 102 Considerando a matriz de covariância  $\Sigma = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$ , em que

$ab > c$ , e que a variância do primeiro fator seja unitária, é correto afirmar que o modelo fatorial ortogonal associado a essa matriz possui solução única.

- 103 Suponha que a matriz de covariância observada em um

conjunto de dados com três variáveis seja  $\Sigma = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & d \end{bmatrix}$ , e que

se  $\lambda_1, \lambda_2$  e  $\lambda_3$  forem seus autovalores, então  $\lambda_1 + \lambda_2 + \lambda_3 = 9$ . Nesse caso, a variância da terceira variável do conjunto de dados,  $d$ , é inferior a 1,5.

- 104 Considere que, na análise de um conjunto de dados tridimensionais, o intervalo de confiança para o vetor de médias  $\mu$  seja expresso por  $n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \leq 6 \times F(3; \eta)$ , em que  $S$  representa a estimativa da matriz de covariâncias,  $F(3; \eta)$  é o quantil de ordem  $1 - \alpha$  de uma distribuição  $F$  de Snedecor com 3 e  $\eta$  graus de liberdade. Nesse caso, é correto afirmar que o tamanho da amostra será superior a 10.

Acerca das variáveis aleatórias multidimensionais, julgue os próximos itens.

RASCUNHO

105 Considere  $X' = [X_1 \ X_2 \ X_3] \sim N(\mu, \Sigma)$  o vetor aleatório

gaussiano, em que  $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$  e  $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ . Nesse caso,

é correto afirmar que as variáveis  $Y_1 = X_1$  e  $Y_2 = \frac{1}{3}(X_1 + X_2 + X_3)$  seguem uma distribuição normal bivariada e, se todos os elementos de  $\Sigma$  forem positivos, essas variáveis serão independentes.

106 Considere que a matriz de covariâncias da distribuição

conjunta do vetor aleatório  $X' = [X_1 \ X_2]$  seja  $\Sigma = \begin{bmatrix} \sigma_{11} & \lambda \\ \lambda & \sigma_{22} \end{bmatrix}$ .

Nesse caso, se  $\lambda = 0$ ,  $X_1$  e  $X_2$  serão independentes.

107 Considere que um vetor aleatório siga a distribuição normal bidimensional cuja densidade é expressa por

$$f(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right\},$$

em que  $\Sigma = \begin{bmatrix} \sigma_1^2 & \lambda \\ \lambda & \sigma_2^2 \end{bmatrix}$  é a matriz de covariância. Nesse caso, se

as variáveis que compõem o vetor aleatório forem independentes, a função de densidade poderá ser representada como o produto das funções de densidade de duas distribuições normais unidimensionais.

108 Considere que a matriz de covariância de um vetor aleatório que segue uma distribuição normal bivariada seja expressa por

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \tau \\ \tau & 1 \end{bmatrix}, \tau > 0, \text{ e que, no contorno dessa distribuição, o}$$

eixo maior da elipse seja igual a 3 vezes o eixo menor. Nesse caso, é correto afirmar que  $\tau < 1,5$ .

109 Na distribuição normal  $p$ -variada, a estatística  $n(\bar{x} - \mu)' S^{-1}(\bar{x} - \mu)$  segue a distribuição  $T_2$  de Hotelling. Em particular, no caso em que  $p = 1$ , essa distribuição equivale à distribuição  $t$  de Student com  $n - 1$  graus de liberdade.

110 Se  $X \sim N_p(\mu, \Sigma)$ , é correto afirmar que a forma quadrática  $n\bar{x}' \Sigma^{-1} \bar{x}$  segue a distribuição qui-quadrado com  $p$  graus de liberdade.

Julgue os itens subsecutivos acerca de lógica de programação em estatística computacional.

- 111 Os programas compilados em *bytecode* e em linguagem de máquina nativa dependem da arquitetura de máquina para a qual foram desenvolvidos.
- 112 Em uma linguagem de programação compilada em *bytecodes*, a depuração somente será possível se o código fonte estiver disponível.
- 113 Nos dois fragmentos de pseudocódigo mostrados abaixo, é correto afirmar que a função `Fatorial` será executada o mesmo número de vezes e para a mesma sequência numérica.

fragmento 1:

```
k = 0
Enquanto k < 10 faça
  Fatorial(k)
  k = k + 1
Fim-enquanto
```

fragmento 2:

```
k = 0
Repita
  k = k + 1
  Fatorial(k)
Até que k == 10
```

Com relação ao *software* R, julgue os itens que se seguem.

- 114 O comando

```
read.csv("arquivo.csv", header = TRUE)
```

permite carregar os dados exportados por uma planilha Excel em português, sem que seja necessário qualquer manipulação no arquivo CSV.

- 115 A instrução `paste(c("X", "Y"), 1:9, sep="")` permite gerar a sequência ["X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "Y1", "Y2", "Y3", "Y4", "Y5", "Y6", "Y7", "Y8", "Y9"].
- 116 O pacote `foreign`, do R, permite abrir arquivo de dados exportado por *softwares* cujo idioma não seja o inglês.
- 117 Considere que X e Y sejam vetores numéricos de mesma dimensão e que Y contenha dados omissos (ou perdidos). Nesse caso, é correto afirmar que os comandos `(X == NA) | (Y == NA)` e `is.na(X) | is.na(Y)` são equivalentes.
- 118 Entre os dois fragmentos de código R mostrados abaixo, é correto afirmar que o fragmento 1 é mais eficiente que o fragmento 2, no que diz respeito ao gerenciamento de memória do sistema.

fragmento 1:

```
X = numeric(1000)
for(k in 1:1000) X[k] = k
```

fragmento 2:

```
X = 0
for(k in 1:1000) X[k] = k
```

Acerca do sistema estatístico brasileiro, julgue os itens a seguir.

- 119 Na realização do censo demográfico brasileiro, o IBGE aplica dois modelos de questionários, o simplificado e o completo. O questionário completo se aplica em um esquema de amostragem sistemática.
- 120 A Pesquisa Nacional por Amostra de Domicílios (PNAD) é uma pesquisa anual do IBGE, realizada com questionário que apresenta um conjunto variável de questões.

RASCUNHO



**cespeUnB**

Centro de Seleção e de Promoção de Eventos