

-- CONHECIMENTOS ESPECÍFICOS --

Em relação a tipos de variáveis e seus níveis de mensuração, julgue os itens a seguir.

- 51** Para dados de uma variável expressa em escala de razão, podem ser aplicadas somente algumas estatísticas descritivas.
- 52** A variável dicotômica é um tipo de variável qualitativa à qual podem ser atribuídos valores numéricos.
- 53** Suponha que, em uma pesquisa realizada para verificar a aceitação de um produto, o respondente deva marcar a faixa salarial de usuários à qual ele pertence. Nessa situação, esse tipo de variável deve ser classificado como métrica.
- 54** Uma variável qualitativa expressa por três categorias pode ser representada por uma variável *dummy*.
- 55** Para o cálculo da média, a variável deve ser necessariamente do tipo métrica.

Em um lote de 30 itens de um produto, 5 apresentam defeitos. Desse lote, 3 itens serão retirados ao acaso, sendo irrelevante a ordem deles.

A partir dessa situação hipotética, julgue os itens seguintes.

- 56** A seleção amostral de elementos sem reposição, quer a ordem importe (arranjo), quer a ordem não importe (combinação), é exemplo de aplicação do princípio multiplicativo.
- 57** A situação apresentada constitui um exemplo de arranjo.
- 58** O número de diferentes combinações de 3 itens que podem ser retirados do lote é igual a 4.060.
- 59** A probabilidade de serem retirados exatamente 2 itens defeituosos do lote é superior a 0,05.
- 60** A probabilidade de exatamente 2 itens retirados do lote não serem defeituosos é inferior a 0,4.
- 61** A probabilidade de todos os três itens retirados do lote não serem defeituosos é superior a 0,5.

Em uma fábrica de teclados, cada teclado produzido tem 2% de chance de estar com defeito. Será realizado um teste de qualidade no qual cada um dos teclados produzidos será analisado até que se encontre um teclado com defeito. A variável aleatória X representa o número de teclados testados até se encontrar o primeiro com defeito, ou seja, se o primeiro teclado analisado for defeituoso, então $X = 1$.

Tendo como referência essa situação hipotética, julgue os próximos itens.

- 62** O desvio padrão de X é superior a 50.
- 63** X tem distribuição hipergeométrica.
- 64** O valor esperado de X é igual a 50.
- 65** A probabilidade de o terceiro teclado testado ser o primeiro com defeito é inferior a 0,05.

A respeito de amostragem, julgue os itens que se seguem.

- 66** Um dos objetivos da estratificação da amostra é a homogeneização da variância dentro de cada estrato em relação à principal variável de interesse.
- 67** Suponha que, de um total de 35 departamentos de uma empresa, sejam sorteados, aleatoriamente, 10 departamentos e que, em cada um deles, seja entrevistada uma pequena amostra da sua população. Nessa situação, é utilizada amostragem por conglomerados em dois estágios.
- 68** A partir dos dados correspondentes a uma população de 30 indivíduos, é possível extrair 900 amostras aleatórias simples com reposição composta por 2 pessoas.
- 69** Uma estatística é uma característica da amostra; um parâmetro é uma medida utilizada para descrever uma característica da população.
- 70** Para uma amostra IID de variáveis com variância finita, à medida que o tamanho da amostra aumenta, a distribuição amostral da média aproxima-se cada vez mais de uma distribuição normal para dados de algumas distribuições específicas.
- 71** Na amostragem sistemática, supõe-se haver uma listagem das unidades populacionais.

Espaço livre

Julgue os próximos itens, referentes a inferência estatística.

- 72 A região crítica de um teste bilateral é representada por duas caudas de tamanhos iguais, respectivamente nas extremidades esquerda e direita da curva de distribuição, correspondendo cada uma delas ao nível de significância α .
- 73 O teste de Shapiro-Francia pode ser aplicado para amostras de tamanho $5 \leq n \leq 5.000$, sendo similar ao teste de Shapiro-Wilk para pequenas amostras.
- 74 Kolmogorov-Smirnov e Shapiro-Wilk são exemplos de testes para normalidade univariada.
- 75 São exemplos de testes para a homogeneidade de variâncias o teste C de Cochran e o teste $F_{\text{máx.}}$ de Hartley.

fonte de variação	graus de liberdade	quadrado médio	p-valor da razão F
regressão	2	14	0,002
erro	8	1	

Considerando o quadro precedente, que mostra parte de uma típica tabela de análise de variância (ANOVA) referente ao ajuste de um modelo de regressão linear que possui um intercepto e cujos coeficientes foram estimados pelo método de mínimos quadrados ordinários, julgue os itens a seguir.

- 76 Com nível de significância de 1%, o p-valor da tabela ANOVA indica que todos os coeficientes do modelo são estatisticamente significativos.
- 77 O valor da razão F é igual a 14.
- 78 O coeficiente de determinação do modelo é inferior a 80%.
- 79 A variância dos resíduos é inferior a 0,90.
- 80 O modelo é constituído por três coeficientes que foram estimados pelo método de mínimos quadrados ordinários.
- 81 A variância amostral da variável resposta (dependente) do modelo é igual a 14.

O quadro a seguir mostra os resultados de uma análise de regressão linear simples com base em uma amostra aleatória de tamanho $n = 10$. Os parâmetros desse modelo foram estimados com base no método da máxima verossimilhança sob erros aleatórios normais com média zero e variância V . A média amostral da variável dependente (resposta) é igual a 8.

	estimativa	razão t	p-valor
intercepto	2,5	3	0,008
coeficiente angular	0,8	1	0,170

A partir dessas informações, julgue os itens que se seguem.

- 82 Se a variável regressora (explicativa) for removida do modelo, restando apenas o intercepto, então, nesse caso, a estimativa do intercepto será superior a 7,6 e inferior a 9,3.
- 83 O coeficiente de determinação do modelo em questão é inferior a 12%.
- 84 A variância amostral da variável dependente é menor que a variância amostral da variável regressora (explicativa).
- 85 A distribuição amostral da razão t segue uma distribuição t de Student com 10 graus de liberdade.
- 86 Se o coeficiente angular do modelo for removido do modelo de regressão, então o R^2 do modelo resultante (sem coeficiente angular) deverá produzir um valor superior ao R^2 do modelo original (com coeficiente angular).
- 87 A média amostral da variável regressora (explicativa) é inferior a 7,5.

Julgue os próximos itens, supondo que $\mathbf{X} = (X_1, X_2)'$ represente um vetor aleatório que se distribui conforme uma normal bivariada tal que $E[\mathbf{X}] = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ e $Var[\mathbf{X}] = \begin{pmatrix} 9 & 2 \\ 2 & 4 \end{pmatrix}$.

- 88 Se $Z_1 = (X_1 - 2)/3$ e $Z_2 = (X_2 - 1)/2$, então $\mathbf{Z} = (Z_1, Z_2)'$ distribui-se conforme uma normal bivariada com $E[\mathbf{Z}] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $Var[\mathbf{Z}] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
- 89 A respeito da média condicional, é correto afirmar que $E[X_1|X_2 = x] = 1,5 + 0,5x$.
- 90 X_2 segue uma distribuição normal com média 1 e desvio padrão 2.
- 91 $Var[X_1 - X_2] \geq 10$.
- 92 A correlação linear entre X_1 e X_2 é igual a 0,5.

Julgue os itens seguintes, considerando que o número diário de petições iniciais com algum tipo de erro processual (Y) seja descrito por uma regressão de Poisson em função de duas variáveis explicativas, X_1 e X_2 .

- 93 Se a variância de Y for maior que a média de Y , isso significará um problema de subdispersão no modelo de regressão de Poisson.
- 94 O modelo de regressão de Poisson segue a forma da expressão $\log(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, em que β_0 , β_1 e β_2 são os coeficientes do modelo.
- 95 Na regressão de Poisson, a *deviance* é um indicador que permite comparar dois modelos, e a diferença das *deviances* entre dois modelos aninhados segue, aproximadamente, uma distribuição qui-quadrado.

Espaço livre

Uma análise de componentes principais envolvendo 5 variáveis métricas proporcionou os seguintes resultados para a extração dos fatores componentes.

componente	autovalor
1	11
2	7
3	4
4	2
5	1

Com base nessas informações, julgue os próximos itens.

- 96** O desvio padrão da terceira componente principal é igual a 2.
- 97** A extração dos fatores componentes foi efetuada a partir de uma matriz de correlações.
- 98** A primeira componente principal explica 44% da variação total dos dados.
- 99** Os autovalores representam as cargas fatoriais das componentes principais.
- 100** A soma das variâncias das 5 variáveis métricas envolvidas no estudo foi igual ou inferior a 30.

Com relação à informática aplicada à estatística, julgue os itens a seguir.

- 101** Se, em uma simulação, for necessário preencher um vetor grande, será mais eficiente usar a linguagem R para alocar e preencher o vetor inteiro, em vez de criar um vetor unitário e redimensioná-lo.
- 102** Na linguagem Python, os arquivos com extensão `.pyc` permitem combinar, em um único arquivo fonte, códigos em Python e em C.
- 103** Caso `f1( )`, `f2( )` e `f3( )` sejam três funções booleanas, então, na condição `f1( ) and f2( ) or f3( )`, a função `f3( )` não será executada se `f1( ) = False`.
- 104** A quantidade mínima de execução é 0 na construção `while (cond) { ... }` e 1 na construção `do { ... } while (cond)`.
- 105** Os elementos de uma lista duplamente encadeada possuem uma referência para o próximo elemento e para o primeiro elemento da sequência.
- 106** Na biblioteca `magrittr` da linguagem R, os operadores `%>%` e `%<>%` realizam exatamente a mesma operação.
- 107** Na biblioteca Pandas da linguagem Python, se `df` é um *data-frame*, o comando `df.drop('telefone', axis=1)` irá remover a coluna `telefone` de `df`.
- 108** O código em Python a seguir está correto quanto à aplicação da técnica de recursão.

```
def f(x):
    f(x - 1)
```

- 109** Uma matriz esparsa pode ser representada por um dicionário.

Julgue os itens seguintes, a respeito de séries temporais.

- 110** Em uma série temporal que tem tendência de crescimento exponencial e na qual os valores iniciais são próximos de zero, a métrica mais adequada para avaliar o ajuste do modelo é a MAPE (*mean absolute percentage error*).
- 111** Uma função de autocorrelação que apresenta um decaimento exponencial indica a existência de um componente autoregressivo.
- 112** O modelo ARIMA com parâmetro $d = 1$ apresenta um componente de tendência linear.
- 113** Um periodograma é um gráfico utilizado para identificar tendências em uma série temporal.
- 114** SARIMA é um modelo para aplicação em dados com tendência e sazonalidade.
- 115** A suavização exponencial simples é equivalente a um modelo AR(p).

Acerca dos testes não paramétricos, julgue os itens que se seguem.

- 116** No teste Wald-Wolfowitz para aleatoriedade, é superior a 4 o número esperado de sequências para uma série com o número de sinais (–) igual a 3 vezes o número de sinais (+).
- 117** No teste qui-quadrado para independência, o número de graus de liberdade da estatística de teste é $(a - 1) \cdot (b - 1)$, em que a é o número de categorias da 1.ª variável e b é o número de categorias da 2.ª variável.
- 118** A estatística do teste da mediana segue uma distribuição qui-quadrado.
- 119** O teste de McNemar é usado para testar a concordância entre dois fatores dicotômicos.
- 120** A estatística do teste de Cox-Stuart segue uma distribuição t de Student.

Espaço livre