

CONHECIMENTOS ESPECÍFICOS

turma	média aritmética	variância amostral
A	7,2	4,0
B	6,8	3,6

RASCUNHO

Um estudo foi realizado em determinada escola para se avaliar o efeito, no desempenho dos estudantes, do uso de computadores em sala de aula. Para esse estudo, foram selecionados aleatoriamente 60 alunos de determinado ano escolar, separando-os em duas turmas A e B, cada uma com 30 alunos. Ao longo de um semestre letivo, um método de ensino com auxílio de computadores foi aplicado na turma A, enquanto, nesse mesmo período, outro método sem auxílio de computadores foi aplicado na turma B. Ao final desse semestre, o mesmo teste foi aplicado para os 60 alunos participantes desse estudo. O quadro acima mostra algumas estatísticas acerca das notas obtidas pelos alunos de ambas as turmas.

Considerando essas informações, acerca de probabilidade, inferência e amostragem, julgue os itens a seguir.

- 51 Considerando-se que os resultados entre as turmas A e B são estatisticamente independentes, é correto afirmar que a variância da diferença entre as médias foi inferior a 0,26.
- 52 Sob a hipótese nula de que as distribuições das notas de ambas as turmas são normais e identicamente distribuídas, a estatística do teste t para a comparação dessas duas médias amostrais segue uma distribuição t de Student com 58 graus de liberdade.
- 53 A variância amostral das notas da turma B utilizou um denominador igual a 29.
- 54 Considerando que as notas da turma A estão ordenadas da menor nota para a maior nota, então a mediana dessas notas ocupa a 15.^a posição nesse rol de dados ordenados.
- 55 Considere que 600 estudantes tenham cursado o ano escolar pertinente a esse estudo na ocasião da sua realização. Nessa situação, a probabilidade de seleção da amostra para o estudo foi inferior a 0,15 e superior a 0,05.
- 56 Os alunos da turma B apresentaram desempenho mais homogêneo que os alunos da turma A, pois a variância amostral da turma B foi inferior a 4,0.
- 57 O erro padrão da média das notas dos alunos da turma A foi superior a 0,40.
- 58 Considere que a prova seja composta por 50 questões e que a

nota obtida por determinado aluno seja igual a $\sum_{q=1}^{50} 0,2 \times y_q$, em

que $y_q = 1$, se o aluno acertou a questão q , e $y_q = 0$, se o aluno não acertou a questão q . Nessa situação, a probabilidade de um aluno obter exatamente a nota 7,5 é superior a 0,01.

A		B	
8	012	4	122
9	4489	5	0018
10	001125	6	012456
11	4566	7	1235
12	78	8	099
13	0		

RASCUNHO

O diagrama A de ramos e folhas acima mostra a distribuição do número de livros destruídos (Y) nas 20 escolas inundadas por causa das fortes chuvas em determinada cidade. O diagrama B mostra a distribuição dos tempos de duração dessas chuvas (X , em minutos) nos dias em que essas 20 escolas foram inundadas.

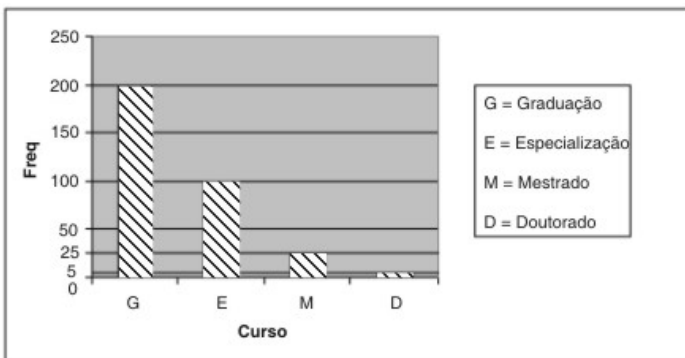
Com base nessas informações e considerando que o valor 100 é representado nesses diagramas como 10|0, julgue os itens que se seguem.

- 59 O ramo e folhas é uma ferramenta exploratória que, além de servir para ordenar um pequeno conjunto de dados, permite destacar a forma da sua distribuição.
- 60 Caso a soma dos tempos X seja igual a 1.261 minutos, então a média aritmética da distribuição dos tempos X será inferior a 1 hora.
- 61 A distribuição dos tempos X não possui moda nem mediana.
- 62 Em algumas escolas desse conjunto de dados, os tempos X foram inferiores a 1/4 de hora.
- 63 Considerando que a correlação entre X e Y seja igual a 0,95, então a variação da distribuição dos tempos X explica pelo menos 94% da variação da distribuição da variável Y .
- 64 Suponha que a regressão linear simples da variável Y em X resulte no modelo ajustado por mínimos quadrados ordinários $Y = aX + b + e$, em que b representa a estimativa do intercepto. Nessa situação, o coeficiente b foi superior a 40 e inferior a 45.
- 65 Considerando o teste t bilateral para a significância do coeficiente angular de um modelo de regressão linear simples na forma $Y = aX + b + e$, em que e representa o erro aleatório, quanto maior for o valor absoluto da razão t , maior será a significância estatística desse coeficiente.
- 66 Na regressão linear simples de X em Y , considere os testes t de significância para o intercepto e o coeficiente angular e suponha que o erro padrão do intercepto foi maior que o erro padrão do coeficiente angular. Com base nessas informações, é possível concluir inequivocamente que o valor do teste t de Student relativo ao intercepto foi superior ao valor do teste t de Student relativo ao coeficiente angular.
- 67 O coeficiente de correlação linear de Pearson é uma estatística adequada para medir a correlação linear entre X e Y .
- 68 Computacionalmente, geralmente os *softwares* SAS, SPSS e Microsoft Excel produzem os mesmos resultados acerca da estimação dos coeficientes de um modelo de regressão linear simples.
- 69 A mediana da distribuição do número de livros destruídos pelas chuvas foi igual a 101.

Para orientar os investimentos em educação em certo município, um analista foi contratado para criar um *ranking* das escolas públicas desse município. Para cada escola, as variáveis disponíveis são a quantidade de turmas, a quantidade de alunos, a quantidade de professores, a nota da Prova Brasil e a área do terreno.

A partir dessa situação, julgue os itens subsequentes.

- 70 Suponha que a distribuição da quantidade de turmas por escola siga uma distribuição de Poisson. Nessa situação, o modelo que descreve essa distribuição pode ser escrito como $P(X = k) = \lambda e^{-\lambda k}$, em que $k > 0$ e $\lambda > 0$ representa a média de turmas por escola.
- 71 A análise de componentes principais é uma técnica da análise multivariada que permite formar um *ranking* dessas escolas com base nesses dados.
- 72 Independentemente da distribuição das notas da Prova Brasil, caso seja necessário simular as notas dessa prova para permitir a aplicação de teorias assintóticas, é recomendável a aplicação do método de Monte Carlo, considerando-se que as notas da referida prova seguem distribuição normal.
- 73 A análise discriminante é técnica multivariada equivalente a um modelo de regressão linear múltiplo.
- 74 A mineração de dados (*data mining*) é um método computacional que permite extrair informações a partir de grande quantidade de dados.
- 75 Considere que as áreas de todas as escolas desse município sejam distintas e que cada escola tenha obtido uma nota diferente na prova Brasil. Nessa situação, os modelos de probabilidade para variáveis aleatórias discretas são adequados para representar a distribuição de todas as variáveis analisadas por esse analista.
- 76 O método de reamostragem *bootstrap* é paramétrico se uma distribuição de probabilidade teórica for especificada; e é não paramétrico se for considerada uma distribuição empírica dos dados em estudo.



A qualificação dos professores é de grande importância para a qualidade da formação dos estudantes. Considerando que a figura acima apresenta a distribuição do número de professores em uma faculdade, segundo a formação acadêmica (curso), julgue os itens de 77 a 85.

- 77 A variável curso é qualitativa nominal.
- 78 Nessa faculdade, a quantidade de professores com mestrado representa 25% dos professores com especialização.

- 79 O total de professores dessa faculdade é igual a 330.
- 80 É correto afirmar que o percentual de professores que possui graduação ou especialização é superior a 85%.
- 81 A porcentagem dos professores dessa faculdade que possui o título de mestre ou doutor é maior que 10%.
- 82 Caso a distribuição mostrada na figura fosse apresentada em um gráfico de setores, o ângulo do setor correspondente à categoria especialização seria superior a 100°.
- 83 A representação dos dados mostrados na figura em um gráfico de *pizza* permite apresentar apenas o percentual de professores em cada nível da variável curso.
- 84 A figura apresenta um gráfico de barras horizontais.
- 85 A distribuição do número de professores segundo a sua formação acadêmica poderia ser adequadamente representada por um gráfico de *pizza* ou por um histograma.

RASCUNHO

Uma fábrica produz parafusos de índices $C_p = 2,0$ e $C_{pk} = 0,5$, que se referem aos diâmetros (em mm) desses parafusos. O controle do processo é do tipo 6 sigma, e sabe-se que o desvio padrão da distribuição dos diâmetros é igual a $\frac{1}{6}$ mm. Com base nessas

informações, julgue os itens a seguir, acerca de controle estatístico de qualidade.

- 86 Considere que, em uma amostra aleatória de parafusos, tenha se observado que a amplitude do intervalo de 95% de confiança foi igual a 2,5 mm. Nessa situação, é correto inferir que todos os parafusos produzidos possuíam diâmetros que atendiam às especificações do processo.
- 87 Se o limite superior de especificação para os diâmetros dos parafusos for igual a 10 mm, então a média do processo será igual ou superior a 8,5 mm.
- 88 Considere que o processo de produção seja alterado para elevar o valor do índice C_{pk} , mantendo-se o índice $C_p = 2,0$. Nessa situação, não é possível obter um novo índice C_{pk} superior a 2,0.

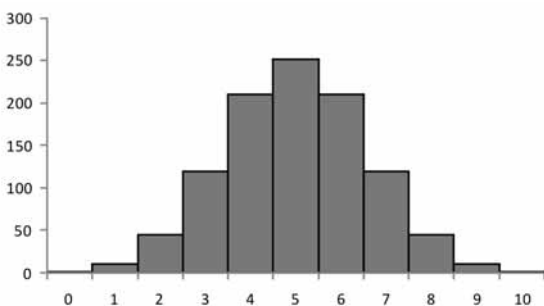
Considere que a chegada e o atendimento de estudantes em determinada fila para matrícula em uma escola possam ser modelados segundo um passeio aleatório simples em tempo discreto $t = 0, 1, 2, 3 \dots$ e que, em cada instante t , apenas dois eventos sejam possíveis: ou um novo estudante entra na fila com probabilidade p ou um estudante na fila é atendido com probabilidade $1 - p$. Suponha, ainda, que, no instante inicial $t = 0$, a quantidade de estudantes na fila não seja nula e grande o suficiente para que a fila não fique vazia em pouco tempo, e que, a cada instante t , no máximo um estudante pode ser atendido.

Com base nessa situação, julgue os itens a seguir.

- 89 Na situação em que se consideram apenas os dois instantes iniciais, sendo $t = 1$ e $t = 2$, e $p = 0,6$, é mais provável que a fila não cresça.
- 90 A fila somente atinge seu tamanho original em instantes de tempo t que são múltiplos de 2.

Com relação à análise de dados discretos, julgue os itens de 91 a 93.

- 91 Se duas amostras $\{X_1, \dots, X_n\}$ e $\{Y_1, \dots, Y_m\}$ são provenientes de distribuições de baixas contagens, então a comparação entre as médias dessas distribuições pode ser feita por um teste t de Student.
- 92 O gráfico ilustrado abaixo é inadequado para representar a distribuição de uma variável aleatória binomial.



93

i	0	1	2	3	4	5	6	7	8	9	10
O_i	0	1	1	4	8	17	30	20	15	3	1
E_i	0,0	0,2	1,1	4,3	11,2	20,1	25,1	21,5	12,1	4,0	0,6

Considere que os valores na tabela acima correspondam aos valores observados – O_i – e aos valores esperados – E_i – segundo determinada hipótese nula de um teste de aderência

cujas estatística do teste tem a forma $Q = \sum_{i=0}^{10} \frac{(O_i - E_i)^2}{E_i}$. Nessa

situação, sob a hipótese nula, é correto afirmar que Q segue uma distribuição F de Snedecor.

RASCUNHO

Julgue os itens subsecutivos, acerca de análise multivariada.

RASCUNHO

94 O método de agrupamento conhecido por *single linkage* usa como medida de distância entre *clusters* $d_{(U)W} = \min \{d_{UW}, d_{VW}\}$, em que d_{UW} e d_{VW} são as distâncias entre os vizinhos mais próximos, respectivamente, nos *clusters* U e W e nos *clusters* V e W . Por outro lado, o método *complete linkage* usa como medida de distância entre *clusters* $d_{(U)W} = \max \{d_{UW}, d_{VW}\}$, em que d_{UW} e d_{VW} são as distâncias entre os vizinhos mais distantes, respectivamente, nos *clusters* U e W e nos *clusters* V e W . Logo, no método *complete linkage*, os elementos em cada *cluster* mantêm similaridade mínima.

95 As medidas de similaridade de Canberra, $d(x, y) = \sum_{i=1}^p \frac{|x_i - y_i|}{x_i + y_i}$, e de Czekanowski, $d(x, y) = 1 - \frac{2 \sum \min(x_i, y_i)}{\sum x_i + y_i}$, podem ser aplicadas nas situações em que $-\infty < x_i, y_i < +\infty$.

O processo MA(2) é escrito como $z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ e a respectiva função de autocorrelação é expressa por

$$\begin{cases} \rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \\ \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_k = 0; \quad \forall k \geq 3 \end{cases}$$

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

96 Se $\{W_t\}$ e $\{Z_t\}$ são processos MA(2) não correlacionados, então o processo $\{W_t + Z_t\}$ é um processo MA(2).

97 Se $\theta_1 = 0,5$ e $\rho_2 = 3\rho_1$, então é correto afirmar que $z_t = a_t - \frac{1}{3}a_{t-1} - \frac{1}{2}a_{t-2}$.

RG(*)	gênero	grau de instrução	hiperatividade
35805684	M	3	N
21355706	F	5	N
43674475	M	2	S
2305386	F	3	N
⋮	⋮	⋮	⋮
98814652	M	4	N

(*) números fictícios gerados por computador.

A tabela acima contém um conjunto de dados formado por quatro variáveis: RG; gênero (M = masculino; F = feminino); grau de instrução (1 = analfabeto; 2 = fundamental incompleto; 3 = fundamental completo; 4 = médio incompleto; 5 = médio completo ou superior); e hiperatividade (S = sim; N = não). Com base nessa tabela, julgue os itens seguintes.

98 Suponha que uma tabela de dupla entrada tenha sido produzida com base nas variáveis gênero e hiperatividade, e que um teste qui-quadrado tenha sido realizado para se efetuar inferências sobre essa tabela. Nesse caso, esse teste deve ter sido, obrigatoriamente, um teste de independência entre essas variáveis.

99 As variáveis mostradas na tabela são qualitativas.

100 Com relação à variável dicotômica hiperatividade, atribuindo-se valor 1 para uma categoria e 0 para a outra, a média dessa variável binária representa a frequência relativa de observações que receberam valor 1.

101 A média da variável grau de instrução possui significado estatístico; mas não o cálculo da média da variável RG.

aluno (<i>i</i>)	matemática (X_i)	física (Y_i)
1	5,59	5,63
2	5,61	5,80
3	6,36	6,85
4	6,37	6,27
5	6,06	6,03
6	6,74	6,80
7	4,58	5,57
8	5,44	5,85
9	7,06	6,16
10	5,49	5,38

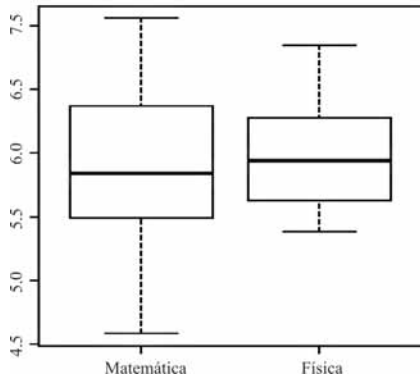
Acerca da tabela acima que mostra as notas de uma amostra de 10 alunos em provas de matemática e física, julgue os próximos itens.

102 Supondo que a tabela abaixo mostra as médias das notas segundo o gênero – masculino ou feminino – desses alunos, é correto afirmar que, pelas medidas abaixo, há indícios de interação entre as variáveis gênero e disciplina.

Disciplina	gênero	
	feminino	masculino
matemática	6,00	5,86
física	6,12	5,95

103 Considerando-se as estatísticas $\sum_{i=1}^{10} X_i = 59,30$, $\sum_{i=1}^{10} Y_i = 60,34$, $\sum_{i=1}^{10} X_i Y_i = 360,3$, é correto afirmar que a correlação entre as notas em matemática e em física foi positiva.

104 A partir dos diagramas de *boxplot* mostrados abaixo, é possível inferir que não existem diferenças significativas entre as médias das notas em matemática e em física na população de alunos.



A menção final *Z* de cada aluno de determinada turma será calculada na forma de nota padronizada: $Z = (X - m)/s$, em que *X* é a nota bruta desse aluno, *m* é a média das notas brutas dos alunos da turma e *s* é o desvio padrão dessas notas. Cada nota padronizada será comparada com os quantis da distribuição normal, de acordo com a tabela abaixo.

menção	ponto de corte	probabilidade normal
A	1,2816	0,10
B	0,4538	0,25
C	0,0000	0,50
reprovado	valor negativo	0,50

Com base nessas informações, julgue os itens de 105 a 108.

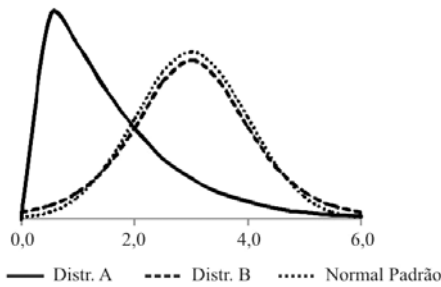
105 A proporção esperada de alunos aprovados na disciplina será maior se a média da turma diminuir.

106 Sabendo-se que a média histórica na disciplina foi 5,09, se as menções A, B e C forem atribuídas com relação aos respectivos pontos de corte da nota bruta 9,0, 7,0 e 5,0, é correto afirmar que a proporção de alunos com menções A ou B tenderá a ser maior somente se o desvio padrão for menor que 3,0.

107 Se a distribuição das notas dos alunos for platicúrtica, então a proporção de alunos com menção A será maior que a proporção esperada de alunos com essa menção.

108 Considere que o coeficiente de assimetria da distribuição das notas dos alunos tenha sido igual a 1,80. Nessa situação, as proporções de menções A e B observadas foram menores que as proporções esperadas correspondentes.

RASCUNHO



Considerando as densidades de probabilidade ilustradas na figura acima, julgue os itens que se seguem a respeito dos momentos dessas distribuições.

- 109 Se a distribuição B for t de Student, então o seu coeficiente de assimetria será nulo.
- 110 Se a distribuição de B não for normal, então sua curtose é negativa.
- 111 Se a distribuição A for gama, com parâmetros 1,5 e 1,0, então a média e a moda dessa distribuição serão iguais.
- 112 Supondo que a distribuição B seja também normal, então o seu desvio padrão será inferior a 1,0.

$$P(X = x, Y = y) = \frac{e^{-\lambda} \lambda^y p^x (1-p)^{y-x}}{(y-x)! x!},$$

$$y \geq 0, \quad 0 \leq x \leq y, \quad \lambda > 0, \quad 0 \leq p \leq 1$$

A respeito da distribuição conjunta (XY), de variáveis aleatórias discretas, apresentada acima, julgue os itens a seguir.

- 113 A variância de X é menor que $\lambda \cdot p^4$.
- 114 O valor esperado de X é negativo.

X	Y			
	-1	0	1	
-1	0	1/5	0	1/5
0	1/5	1/5	1/5	3/5
1	0	1/5	0	1/5
	1/5	3/5	1/5	1

Julgue os próximos itens, considerando que o vetor aleatório (X, Y) possui distribuição conjunta de probabilidade conforme o quadro acima.

- 115 As variáveis aleatórias $Z = X + Y$ e $W = X - Y$ são dependentes.
- 116 A correlação linear entre X e Y é nula; disso se conclui que ambas são independentes.

Com respeito a distribuições conjuntas (X,Y), julgue os itens de 117 a 120.

- 117 A distribuição normal bivariada tem função de densidade dada por

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right]\right)$$

em que ρ representa a correlação linear entre X e Y.

- 118 A distribuição

$$f(x, y) = \frac{y^\alpha \cdot (1-y)^{x+\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq y \leq 1, x = 0, 1, 2, \dots$$

é uma combinação de uma variável aleatória geométrica com uma variável aleatória beta.

- 119 Considerando a função de densidade $f(x, y) = 1/4$, para $-1 \leq x \leq 1$ e $-1 \leq y \leq 1$, e o evento $A = \{(x, y): (x, y) \in C\}$, em que C é o círculo de raio 1 e centro (0, 0), é correto afirmar que $P(A) = \pi/4$.

- 120 A função $f(x, y) = \frac{e^{x+y} - 1}{e - 1}$, $0 \leq x \leq 1$, $0 \leq y \leq 1$

é uma função de distribuição acumulada.

RASCUNHO

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando 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 além da 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.

Suponha que o rendimento escolar y_i de certo aluno, $i = 1, \dots, n$, e o tempo total x_i que esse aluno i dedicou para os estudos são variáveis que se relacionam segundo um modelo na forma $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, em que β_0 é o coeficiente linear, β_1 é o coeficiente angular e $\varepsilon_1, \dots, \varepsilon_n$ representa uma sequência de erros aleatórios independentes e identicamente distribuídos conforme uma distribuição normal com média zero e variância σ^2 .

Com base na situação hipotética acima, redija um texto dissertativo que atenda, necessariamente, aos seguintes aspectos:

- ▶ explique porque (ou como) o modelo proposto na forma linear é adequado ao problema;
- ▶ mostre que no modelo de regressão linear simples, os estimadores para os coeficientes linear e angular são os mesmos tanto quando obtidos pelo método de mínimos quadrados ordinários, quanto pelo método da máxima verossimilhança;
- ▶ explique os significados dos coeficientes linear e angular.

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	