Ministério da Saúde

CADERNO DE PROVAS - PARTE II

CONHECIMENTOS ESPECÍFICOS

CARGO 12

ESTATÍSTICO

CONCURSO PÚBLICO **Nível Superior**

LEIA COM ATENÇÃO AS INSTRUÇÕES ABAIXO.

1 Nesta parte II do seu caderno de provas, confira atentamente os seus dados pessoais e os dados identificadores de seu cargo transcritos acima com o que está registrado em sua **folha de respostas**. Confira também o seu nome, o nome e número de seu cargo no rodapé de cada página numerada desta parte II de seu caderno de provas. Caso o caderno esteja incompleto, tenha qualquer defeito, ou apresente divergência quanto aos seus dados pessoais ou aos dados identificadores de seu cargo, solicite ao fiscal de sala mais próximo que tome as providências cabíveis, pois não serão aceitas reclamações posteriores nesse sentido.

2 Quando autorizado pelo chefe de sala, no momento da identificação, escreva, no espaço apropriado da folha de respostas, com a sua caligrafia usual, a seguinte frase:

A natureza oferece um caminho diferente a cada um.

cespeUnB
Centro de Seleção e de Promoção de Eventos

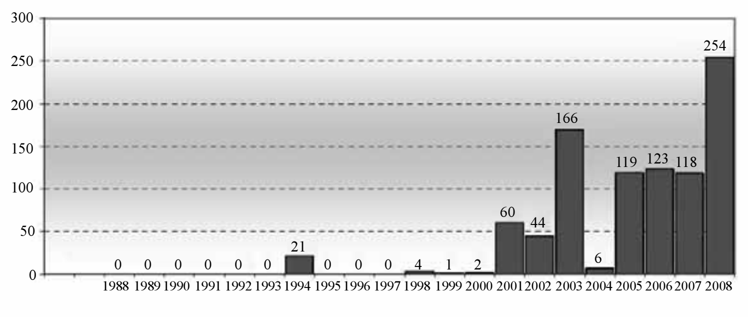
Universidade de Brasília

Ministério da Saúde
BRASIL UM PAÍS DE TODOS
GOVERNO FEDERAL

OBSERVAÇÕES

- Não serão objeto de conhecimento recursos em desacordo com o estabelecido em edital.
- Informações adicionais: telefone 0(XX) 61 3448-0100; Internet — www.cespe.unb.br.

CONHECIMENTOS ESPECÍFICOS



A figura acima apresenta os totais anuais de casos de febre hemorrágica da dengue, de 1988 a 2008, em Fortaleza, cidade em que a doença foi confirmada pela primeira vez em 1994. A partir de 1998, verifica-se a ocorrência anual da enfermidade, iniciando em um patamar de baixa incidência (1998 a 2000) e seguindo para um patamar elevado que varia de 44 a 254 casos, com exceção de 2004.

Secretaria Municipal da Saúde de Fortaleza. **Plano de contingência para o controle da dengue no município de Fortaleza em 2009**, (com adaptações).

Com base nas informações acima, considerando que a variável X representa o total anual de casos de febre hemorrágica da dengue em Fortaleza, julgue os itens a seguir.

51 A média aritmética de X no triênio 2001-2003 foi igual a 75% da média aritmética de X no triênio 2005-2007.

52 Considerando o período de 1988 a 2008, a moda da variável X foi igual a 254.

53 De 1988 a 2008, a mediana amostral de X foi superior a 3.

54 Construindo-se o diagrama *boxplot* usual, com relação à variável X e com os dados do ano 2001 em diante, é correto afirmar que a exceção observada em 2004 não deve ser considerada como um valor atípico.

55 A figura apresentada é um histograma da variável X.

56 Considerando-se as observações no período de 1995 a 1999, a variância amostral de X foi igual a 3.

O número de casos diários de desordens musculoesqueléticas (W) em certa empresa é uma variável aleatória discreta que segue uma distribuição condicional na forma $\mathrm{P}(W = k \mid Y \geq y) = \dfrac{(20y+2)^k}{k!} \times exp(-20y)\,exp(-2)$, em que Y é uma variável aleatória contínua tal que $\mathrm{P}(Y \geq 0) = 1$, $k = 0, 1, 2, \ldots$ e $exp(-2) = 0{,}137$.

R. Quintana e I. Pawlowitz. **Safety Science**, 32, 1999, p. 19-31 (com adaptações).

Com base nessas informações, julgue os itens subsequentes.

57 A probabilidade $\mathrm{P}(W = 1)$ é superior a 0,12 e inferior a 0,16.

58 Se $m > 0$ for a mediana da variável Y, então $\mathrm{P}(W = 0 \mid Y < m) = \dfrac{2 + exp(-20m)}{2}$.

59 O desvio padrão da variável aleatória W é igual a 1.

RASCUNHO

Um estudo mostrou que o tempo de ocupação de um leito hospitalar — T —, em horas, segue uma distribuição cuja função de densidade é expressa por $f(t) = 0{,}25at^{-0{,}75} \times exp(-at^{0{,}25})$, em que $a > 0$ é um parâmetro fixo e $t > 0$.

M.D. Banks *et alli*. **Clinical Nutrition**, 2009, p. 1-7 (com adaptações).

Com base nessas informações, julgue os itens seguintes.

60 A função de densidade $f(t)$ assume apenas valores entre 0 e 1.

61 $P(T > 16) = exp(-2a)$.

62 O tempo médio é igual a $\frac{1}{4a}$ horas.

63 O desvio padrão da distribuição do tempo T é igual a $\frac{1}{2a}$ horas.

64 A moda e a mediana da distribuição da variável aleatória T são, ambas, iguais a $\frac{1}{4a}$ horas.

65 O primeiro momento central da variável aleatória T é igual a zero, enquanto o segundo momento central dessa mesma distribuição corresponde à variância dos tempos.

João foi submetido a um teste de laboratório para o diagnóstico de uma doença rara. A probabilidade de essa doença se desenvolver em um indivíduo como o João é igual a 0,001. Sabe-se que esse teste pode resultar em "falso positivo", ou seja, indicar que João possui essa doença, quando na verdade ele não a tem. Ou, o teste pode resultar em "falso negativo", isto é, indicar que João não possui a doença, quando na verdade ele está doente. A probabilidade de o teste resultar em falso positivo é igual a 0,05 e a probabilidade de o teste resultar em falso negativo é igual a 0,02.

Com base nas informações dessa situação hipotética, julgue os itens subsequentes.

66 Se qualquer indivíduo como João submeter-se ao teste, então a probabilidade de o teste produzir um resultado negativo é superior a 0,94 e é inferior a 0,98.

67 Se o teste ao qual João foi submetido der resultado positivo, então a probabilidade de ele estar de fato com a doença é inferior a 0,02.

68 Se quatro indivíduos que possuem essa doença forem selecionados ao acaso e submetidos ao referido teste de laboratório, e se os resultados forem independentes entre si, então a probabilidade de ocorrerem exatamente dois resultados negativos e dois resultados positivos é inferior a 0,005.

Um laboratório farmacêutico produz certo medicamento em três locais diferentes: A, B e C. Do total produzido, 40% têm origem em A; 35% em B e o restante, 25%, tem origem em C. As probabilidades de que haja defeitos no produto final variam segundo o local de origem e são iguais a 0,01, 0,02 e 0,03 para os locais A, B e C, respectivamente. A produção desse laboratório é reunida em certo local D para ser vendida, de maneira que os medicamentos são misturados ao acaso, fazendo com que a identificação da sua origem (A, B ou C) seja impossível.

Considerando essa situação hipotética, julgue o item abaixo.

69 Se um comprador adquire um medicamento defeituoso no local D, é mais provável que sua origem seja de A.

RASCUNHO

Com respeito às distribuições Z (normal padrão), t de Student, χ^2 (quiquadrado) e F de Snedecor, julgue os itens que se seguem.

70 A distribuição t de Student, com k graus de liberdade, é definida pela razão $\frac{Z}{\sqrt{Q/k}}$, em que Z é a distribuição normal padrão e Q é a distribuição quiquadrado com $k > 0$ graus de liberdade, com Z e Q independentes.

71 A variância de uma distribuição t de Student, com 10 graus de liberdade, é inferior a 1.

72 A média de uma distribuição t de Student é igual a zero.

73 A distribuição F de Snedecor é definida pela razão de duas distribuições quiquadrado independentes.

74 A média de uma distribuição F de Snedecor depende de dois parâmetros: o número de graus de liberdade do denominador e o número de graus de liberdade do numerador.

75 A variância de uma distribuição quiquadrado é quatro vez maior do que a sua média.

grupos	uso recente		total
	sim	não	
caso	9	12	21
controle	33	390	423
total	42	402	444

Shapiro, Slone e Rosenberg. **Oral contraceptive use in relation to myocardial infarction**. The Lancet, v. 1, 1979, p. 743-7 (com adaptações).

A tabela acima apresenta os resultados de um estudo relativo à associação entre infarto cardíaco e a utilização de contraceptivos orais. A partir dos arquivos de um hospital, foi levantada uma amostra consistindo de 444 pacientes com idade entre 30 e 34 anos, dividida nos grupos caso — pacientes que apresentaram histórico de infarto do miocárdio — e controle — pacientes com perfis semelhantes, mas sem histórico de infarto do miocárdio. Assumindo-se um nível de significância de 1%, foi aplicado um teste quiquadrado, obtendo-se uma estatística igual a $\chi^2 = 28{,}7068$ com probabilidade de significância igual a $8{,}42 \times 10^{-8}$.

Tendo como referência as informações acima, julgue os itens de **76** a **83**.

76 A hipótese testada é H_0: $\chi^2 = 0$ *versus* H_0: $\chi^2 > 0$, que corresponde ao teste de homogeneidade entre os grupos caso e controle.

77 A probabilidade de significância pode ser interpretada como sendo muito baixa a probabilidade de se obter um valor da estatística χ^2 superior a 28,71, assim é correto inferir que a proporção das pacientes do grupo caso difere da proporção das pacientes do grupo controle.

78 O poder para o teste em questão é igual a 0,9, considerando o nível de significância mencionado e o tamanho da amostra igual a 444.

79 A estatística χ^2 tem distribuição quiquadrado com $r \times c$ graus de liberdade, em que r é o número de linhas da tabela e c é o número de colunas.

80 Considerando os dados informados, ao ser aplicada sob a hipótese nula, a estatística χ^2 tem uma distribuição com aproximação deficiente para a distribuição quiquadrado, uma vez que um pressuposto básico foi violado. Um teste de hipótese mais adequado é o teste exato de Fisher.

81 Se o estudo fosse aplicado a um número muito superior a 444 pacientes, a distribuição amostral de χ^2 tenderia a se comportar como uma distribuição normal com média 21 e variância $\frac{20}{444}$.

82 Os valores da distribuição de probabilidade condicional do grupo caso e do grupo de controle, considerando a utilização recente de contraceptivo oral, são, respectivamente, 0,21 e 0,79.

83 Se a amostra estudada representar adequadamente a população em questão, então a probabilidade de uma paciente ter infarto do miocárdio, considerando que ela não usava contraceptivos orais, está compreendida entre 1% e 5%.

RASCUNHO

Em uma campanha de vacinação, 1.000 empregados de uma grande indústria receberam a vacina contra gripe. Destes, 100 apresentaram alguma reação alérgica de baixa intensidade. A esse respeito, julgue os próximos itens.

84 Considere o seguinte teorema: Se X_1, X_2, ..., X_n for uma sequência de variáveis aleatórias independentes e identicamente distribuídas, cada uma tendo média μ e variância σ^2, então a distribuição formada por $n^{-1/2}\sigma^{-1}(X_1 + X_2 + ... + X_n - n \times \mu)$ tende para a distribuição normal padrão, quando $n \to \infty$. Na situação em questão, assumindo-se que n = 1.000 seja grande o suficiente, é correto afirmar que o referido teorema dá o embasamento probabilístico para a utilização de um teste para a hipótese alternativa H_1: $\mu < 15$.

85 Se a distribuição binomial for aproximada por uma distribuição normal e o erro-padrão da média for igual a 0,3, então há uma probabilidade de 95% de que o número médio de empregados da indústria com alguma reação alérgica à vacina esteja entre 99,4 e 100,6.

86 Se as variáveis aleatórias X e Y seguem distribuição binomial com parâmetros (n, p) e (m, p), respectivamente, e se a função geradora de momentos para a variável aleatória $Z = X + Y$ é dada por $M_z(t) = (pe^t + 1 - p)^{m+n}$, então é correto concluir que X e Y não são independentes uma da outra.

87 A estimativa de máxima verossimilhança para a raiz quadrada do número médio de empregados da indústria com reação alérgica à vacina é superior a 9.

Uma população de plantas contém 3 diferentes genótipos: A, B e C, com as respectivas proporções: θ_1, θ_2 e θ_3. Em um estudo em que 100 plantas dessa população foram registradas no cerrado, observou-se o número de plantas associadas a cada genótipo: 32, 57 e 11. De acordo com a literatura científica da área, as proporções esperadas são iguais a 30%, 50% e 20%.

Considerando essas informações, julgue os itens que se seguem.

88 Para se obterem os estimadores de máxima verossimilhança para θ_1, θ_2 e θ_3, deve-se maximizar a função $L(\theta_1, \theta_2, \theta_3) = \theta_1^{n_1} \times \theta_2^{n_2} \times (1-\theta_1-\theta_2)^{n-n_1-n_2}$, em que n_1, n_2 e n_3 são os números de plantas da amostra que pertencem aos genótipos A, B e C, respectivamente, e $n_1 + n_2 + n_3 = n$.

89 As estimativas de máxima verossimilhança para θ_1, θ_2 e θ_3 são, respectivamente, 0,32, 0,5 e 0,11.

90 A estatística do teste de aderência apresenta valor inferior a 10.

91 Se o percentil de 5% superior da distribuição quiquadrado com 2 graus de liberdade for igual a 5,99, então é correto inferir que há fraca evidência amostral para assumir que as proporções amostrais observadas diferem das proporções verificadas.

92 Os estimadores de máxima verossimilhança são sempre viciados, porém, consistentes.

RASCUNHO

Em um estudo oncológico, foi registrado o tempo, em semanas, de sobrevida de pacientes com leucemia aguda. Na data do diagnóstico da patologia, registrou-se também o número de glóbulos brancos, em escala logarítmica. Por meio de uma análise exploratória de dados, assumiu-se que os tempos de sobrevida t_i, $i = 1, ..., n$, em que n é o tamanho da amostra, seguem distribuição exponencial. A tabela a seguir apresenta medidas-resumo, calculadas por meio de um *software* estatístico, na qual o tempo de sobrevida dos pacientes está em unidade de tempo apropriada, e o número de glóbulos brancos está em logaritmo neperiano (ln).

descrição	tempo de sobrevida dos pacientes	ln do número de glóbulos brancos
mínimo	1	2,875
primeiro quartil	16	3,732
mediana	56	4
média	62,47	4,111
terceiro quartil	108	4,716
máximo	156	5

A partir dessas informações, julgue os itens a seguir.

93 A função densidade de probabilidade para a distribuição exponencial, utilizando estatísticas calculadas sobre a amostra, pode ser expressa por $f_T(t) = 62{,}47 \times e^{-62{,}47\,t}$, com esperança matemática $E(T) = 62{,}47$.

94 O termo representado por $\sum_{i=1}^{n} t_i$ é uma estatística suficiente para estimar o parâmetro λ da distribuição exponencial.

95 Considerando que o tamanho da amostra seja pequeno, para se testar a hipótese H_0: $\mu = 50$ *versus* H_1: $\mu \neq 50$, deve-se utilizar o teste-t. Esse tipo de procedimento é adequado para modelar a distribuição amostral da média aritmética dos tempos de sobrevida dos pacientes.

96 Se os tempos de sobrevida dos pacientes seguirem distribuição exponencial, então é possível construir um intervalo de $100(1-\alpha)\%$ de confiança, utilizando o método da quantidade pivotal com estatística definida por $S = \sum_{i=1}^{n} t_i$, em que $1 - \alpha$ é o nível de confiança.

97 Se o número de glóbulos brancos seguisse distribuição lognormal, com parâmetros μ e σ, então o logaritmo do número de glóbulos brancos teria distribuição normal com média μ e desvio padrão σ.

98 A probabilidade de um paciente sobreviver mais de 30 semanas, considerando que ele se encontra vivo há mais de 10 semanas da data do diagnóstico, é igual à probabilidade de o mesmo paciente sobreviver mais de 20 semanas.

RASCUNHO

Uma concessionária de veículos estudou o preço de determinado tipo de veículo em função da idade (anos de uso). Os resultados encontram-se na seguinte tabela.

X (idade em anos)	Y (preço em R$)
0	80.000
1	75.000
2	55.000
3	48.000
4	42.000

Um estatístico ajustou o modelo de regressão linear simples $Y = a + bX + \varepsilon$ aos dados, em que ε representa um desvio aleatório. Com base nessas informações, julgue os itens a seguir.

99 Os parâmetros a e b são obtidos resolvendo-se o sistema de equações lineares a seguir

$$an + b\sum x_i = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i \,,$$

em que n representa o tamanho da amostra.

100 As estimativas dos parâmetros $\hat{a}$ e $\hat{b}$ são: $\hat{a} = 78.000$ e $\hat{b} = -10.300$.

101 O preço esperado de um veículo de 5 anos de idade é igual a R$ 30.100.

102 Comparando os preços observados da tabela com os preços esperados, o desvio absoluto entre esses valores será maior para o veículo com 2 anos de idade.

103 Se o desvio aleatório ε tiver distribuição $N(0, \sigma^2)$ com σ = R$ 2.000, então, considerando que $\Phi(0{,}85) = 0{,}8023$, em que Φ denota a função de distribuição do modelo normal padronizado, a probabilidade de que um veículo com 3 anos de idade tenha valor inferior a R$ 48.000 é inferior a 20%.

No modelo de regressão múltipla $Z = a + bX + cY + \varepsilon$, ε representa um desvio aleatório. Com referência a esse modelo, julgue os próximos itens.

104 As equações normais para esse problema são expressas pelas seguintes equações:

$$an + b\sum x_i + c\sum y_i = \sum z_i \,;$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i y_i = \sum x_i z_i \,;$$
$$a\sum y_i + b\sum x_i y_i + c\sum y_i^2 = \sum y_i z_i \,,$$

em que n representa o tamanho da amostra.

105 No caso específico em que $a = 0$, as equações normais são expressas por

$$b\sum x_i + c\sum y_i = \sum z_i \,;$$
$$b\sum x_i^2 + c\sum x_i y_i = \sum x_i z_i \,.$$

RASCUNHO

O modelo de regressão quadrática $Y = a + bX + cX^2 + \varepsilon$ deve ser ajustado aos dados da seguinte tabela.

X	Y
−2	3
−1	2
0	0
1	1
2	2

Nesse caso, é correto afirmar que

106 as equações normais são dadas por

$$an + b\sum x_i + c\sum x_i^2 = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum x_i y_i$$
$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum x_i y_i^2,$$

em que n representa o tamanho da amostra.

107 a parábola dos mínimos quadrados tem os parâmetros $\hat{a} = 0{,}6$, $\hat{b} = -0{,}3$, $\hat{c} = 0{,}5$.

X	Y
1	2
2	3
3	2
4	3
5	4

Considerando a tabela de valores acima, nas variáveis X e Y, julgue os itens subsequentes.

108 Se $Cov(X, Y)$ é a covariância entre X e Y, $V(X)$ é a varância de X e $V(Y)$ é a variância de Y, então é correto afirmar que o coeficiente de correlação linear, $\text{Corr}(X, Y) = \dfrac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}$, é inferior a 0,8.

109 Se o coeficiente de correlação linear entre as variáveis é igual a zero, então não existe nenhuma relação entre as variáveis X e Y.

Na amostragem aleatória simples, a relação entre o tamanho mínimo da amostra — n — e o tamanho da população — N — é dada por $n = \dfrac{1}{E_0^2 + 1/N}$, em que E_0 representa o erro amostral tolerável. Com base nessas informações, julgue os itens seguintes.

110 Se $N = 10.000$ e $E_0 = 0{,}04$, então $n > 600$.

111 Se $N = 5.000$ e $E_0 = 0{,}05$, então a amostra precisa conter mais de 7% da população.

112 Se $E_0 = 0{,}03$, então o tamanho $n = 1.000$ de uma amostra é suficientemente grande para qualquer tamanho N da população.

RASCUNHO

Suponha que X seja uma variável correspondente à altura de uma pessoa de determinada população. Uma amostra aleatória simples, considerando 5 pessoas de uma população de 100 pessoas, é representada pelas alturas (em cm): $x_1 = 160$, $x_2 = 165$, $x_3 = 170$, $x_4 = 172$, $x_5 = 178$. Com base nesses dados, julgue os itens a seguir.

113 O número de amostras aleatórias simples de tamanho 5 dessa população é igual a $\frac{100^5}{5!}$.

114 A estimativa para a altura média da população é igual a 169 cm.

115 Uma estimativa não viciada para a variância da altura da população é de 47 cm^2.

116 Se a altura da população tem distribuição $N(\mu, 100)$, então, considerando $\Phi(1{,}96) = 0{,}975$, em que Φ denota a função de distribuição do modelo normal padronizado, um intervalo de confiança de 95% para μ é dado por [162,2; 175,8].

Para estimar o salário médio mensal, os 5.000 empregados de uma empresa foram divididos em quatro estratos: homens com menos de 40 anos de idade, homens com mais de 40 anos de idade, mulheres com menos de 40 anos de idade e mulheres com mais de 40 anos de idade, conforme a tabela a seguir.

idade	homens	mulheres	totais
≤ 40 anos	1.200	1.400	2.600
> 40 anos	1.800	600	2.400
totais	3.000	2.000	5.000

Uma amostra estratificada proporcional de 200 empregados apresenta os seguintes salários médios observados nos estratos, em R$:

idade	homens	mulheres
≤ 40 anos	5.000	4.000
> 40 anos	8.000	7.000

De acordo com os dados acima, julgue os próximos itens.

117 A amostra consiste de 48 homens com menos de 40 anos, 72 homens com mais de 40 anos, 24 mulheres com menos de 40 anos, e 56 mulheres com mais de 40 anos.

118 O salário médio estimado dos empregados é superior a R$ 6.000.

Considere que uma pequena população seja composta de 12 pessoas com os seguintes pesos:

75 80 65 90 70 72 60 70 85 65 80 76

Com base nesses dados, julgue os itens que se seguem.

119 As amostras sistemáticas de tamanho 3 correspondem às colunas do seguinte esquema retangular:

$$\begin{matrix} 75 & 80 & 65 & 90 \\ 70 & 72 & 60 & 70 \\ 85 & 65 & 80 & 76 \end{matrix}$$

e o peso médio esperado de uma amostra sistemática é de 74 kg.

120 Na amostragem sistemática, o peso médio esperado de uma amostra corresponde sempre ao peso médio da população.

RASCUNHO