Universidade Federal de Viçosa
Departamento de Estatística

Prova Seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - dezembro/2012

Nome: ____

Assinatura: ____.

Número do CPF ou RG: _______________________________.

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São 6 questões em páginas numeradas de 1 a 8, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

1 (10%). Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população com a seguinte função densidade de probabilidade,

$$f(x|\gamma) = \frac{1}{\Gamma(\theta+1)\gamma^{\frac{\theta+1}{2}}}\; x^{\theta}\; e^{\frac{-x}{\sqrt{\gamma}}}$$

em que $\theta > -1$ é uma constante conhecida e $\gamma > 0$ é um parâmetro desconhecido. Pede-se: obtenha o estimador de máxima verossimilhança de γ com base na amostra $x_1, x_2, \ldots, x_n$.

2 (15%). De acordo com a literatura clássica de Probabilidade e Inferência, a Lei Fraca dos Grandes Números é de suma importância para a Inferência Estatística. Formalmente, a lei diz:

Teorema 1 (Weak law of large numbers) *Let $f(.)$ be a density with mean μ and finite variance $\sigma^2 < \infty$, and let $\bar{X}$ be the sample mean of a random sample of size n from $f(.)$. Let ε and δ be any two specified numbers satisfying $\varepsilon > 0$ and $0 < \delta < 1$. If n is any integer greater than $\frac{\sigma^2}{\varepsilon^2 \delta}$, then*

$$P[-\varepsilon < \bar{X} - \mu < \varepsilon] \geq 1 - \delta.$$

Pergunta-se:

a.(5%) Em outras palavras, o que esse teorema está dizendo, em termos de sua importância para a Inferência Estatística?

b.(5%) Demonstrar esse teorema usando o Teorema e Corolário da Desigualdade de Chebyshev, dados por:

Teorema 2 *Seja X uma v.a. e $g(.)$ uma função não-negativa com domínio na reta real. Então*

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}, \forall k > 0.$$

Corolário 1 (Desigualdade de Chebyshev) *Se X é uma v.a. com variância finita*

$$P[|X - \mu_X| \geq r\sigma_X] = P[(X - \mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2}, \forall r > 0.$$

c.(5%) Qual o tamanho da amostra para 99% de certeza que $\bar{X}$ divirja em 0.5σ de μ?

Solução da questão 2:

3 (5%). Considere o seguinte modelo,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

em que $\mathbf{X}$ é uma matriz de dimensão $N \times p$ com posto $r < p$, $\boldsymbol{\beta}$ é um vetor de dimensão $p \times 1$ cujos elementos são parâmetros desconhecidos de efeitos fixos e $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$.
Pede-se: Enuncie o teorema de Gauss-Markov.

4 (5%). A tabela a seguir mostra os valores da idade (X_1), da escolaridade (X_2) e da renda mensal (Y) para uma amostra aleatória de 8 pessoas.

X_1	X_2	Y
20	12	16
20	8	10
26	10	13
32	12	20
32	4	10
38	6	19
44	8	20
44	4	20

Considerando o modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

a.(10%) Obtenha as estimativas não-tendenciosas de variância mínima para β_0, β_1 e β_2.

b.(10%) Mostre como testar a hipótese $H_0 : \beta_1 = 0$.

Dicas: considere os seguintes resultados:

FV	GL	SQ
Modelo	2	114,00
Resíduo	5	24,00
Total	7	138,00

$$\left(X^t X\right)^{-1} = \begin{bmatrix} 5,6012 & -0,1012 & -0,2798 \\ & 0,0022 & 0,0037 \\ & & 0,0201 \end{bmatrix}$$

5 (20%). Considere a seguinte função densidade de probabilidade conjunta,

$$f(x,y) = \begin{cases} x+y & , \quad 0 \leq x \leq 1 \ \text{ e } \ 0 \leq y \leq 1 \\ \\ 0 & , \quad \text{outros valores} \end{cases}$$

a.(5%) Calcule a probabilidade conjunta: $P\,(X < 0,5\ ,\ Y > 0,25)$.

b.(5%) Calcule $E(2X - \frac{1}{6})$.

c.(10%) Calcule a probabilidade condicional: $P\,(X \geq 0,8 \mid Y = 0,5)$.

6 (30%). Em um experimento com salsinha foram testados 2 tratamentos (T_1, o controle sem pré-resfriamento e T_2, o uso de pré-resfriamento com água gelada) e 7 épocas ou tempos de avaliação (TE = 0, 1, 2, 3, 4, 5 e 6 dias após o tratamento). O delineamento experimental foi o inteiramente casualizado no esquema de parcelas subdivididas, com 4 repetições. Foi analisada a variável dependente "teor relativo de água", submetido inicialmente à análise de variâncias (ANOVA) conforme o quadro a seguir.

a.(15%) Complete o quadro da ANOVA.

b.(15%) Indique os testes F de interesse com as respectivas hipóteses de nulidade.

FV	GL	SQ	QM	F_c
Tratamentos (T)		176,8612		
Resíduo (a)		30,8443		
TE		4954,7409		
Interação T × TE		111,3411		
Resíduo (b)				
Total		5654,1774		

Universidade Federal de Viçosa
Departamento de Estatística

Prova Seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - Junho/2013

Nome: ____

Assinatura: ____.

Número do CPF ou RG: _______________________________.

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São 5 questões em páginas numeradas de 1 a 6, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

1 (15%). Mood, Graybill and Boes (1974) say "for a given probability space $(\Omega, \mathbb{A}, P[.])$, let A and B be two events in $\mathbb{A}$. Events A and B are defined to be independent if and only if any one of the following conditions is satisfied:

(i) $P(A \cap B) = P(A) . P(B)$

(ii) $P(A|B) = P(A)$, if $P(B) > 0$

(iii) $P(B|A) = P(B)$, if $P(A) > 0$"

Mostre a equivalência entre as 3 condições, ou seja, mostre que (i) implica em (ii), (ii) implica em (iii) e (iii) implica em (i).

2 (20%). Seja $X_1, ..., X_n$ uma amostra aleatória do modelo exponencial (λ),

$$f(x) = \lambda\ e^{-\lambda x}, \quad 0 \leq x < \infty, \quad 0 < \lambda$$

Pede-se: Obtenha o estimador de máxima verossimilhança para λ.

3 (15%). Para avaliar hábitos e atitudes de estudo existe um índice denominado SSHA (Survey of Study Habits and Attitudes Brown, W. F. e Holtzman, W. H. (1967)) cujos valores oscilam entre 0 e 200. Estudantes de ambos os sexos foram pesquisados e os respectivos valores de SSHA são apresentados abaixo. Acredita-se que a 5% o SSHA médio do sexo Feminino (F) seja superior ao do sexo Masculino (M) .

F: 154; 109; 137; 115; 152; 140; 154; 178; 101 e 148.
M: 108; 140; 114; 91; 115; 126; 92; 169; 146 e 104.

Para F: $\sum_i X_i = 1388$ e $\sum_i X_i^2 = 197800$
Para M: $\sum_i Y_i = 1205$ e $\sum_i Y_i^2 = 150799$

(a) Sugira uma sequência de testes de hipóteses,"coerentes segundo a teoria", para verificar essa suposição do problema.

(b) Os testes de normalidade para "F" e "M" foram executados no programa R, conforme resultados abaixo. A hipótese testada é: H_0 : a amostra é de uma população normal. Interprete-os CONFORME o conhecimento do significado do "p-value" que em português é traduzido como valor-p ou nível crítico ou probabilidade de significância.
Shapiro-Wilk normality test statistics = W
data F: W = 0.9388, p-value = 0.5394
data M: W = 0.9387, p-value = 0.5388

(c) Conclua sobre a suposição do problema. Apresente os cálculos e a decisão caso estivessem disponíveis os valores tabelados necessários.

(d) Que tipo de erro pode estar sendo cometido e com que probabilidade? Detalhe o máximo possível.

(e) Como fica a questão do Poder do Teste nesse caso? Detalhe o máximo possível.

4 (30%). Considere a tabela a seguir com os resultados obtidos em um experimento instalado no delineamento experimental inteiramente casualizado, com quatro repetições. Os tratamentos foram definidos no esquema fatorial com três níveis do fator A e dois níveis do fator B.

Totais de tratamentos			
	Fator B		
Fator A	B_1	B_2	Total
A_1	225	167	
A_2	325	227	
A_3	290	320	
Total			

Sabendo-se que,

$$SQ_{\text{Total}} = 4.228; \quad SQ_A = 1.792 \quad \text{e} \quad SQ_B = 1.944$$

$F_{0,05}(6;12) = 3,00$; $F_{0,05}(3;17) = 3,20$; $F_{0,05}(2;23) = 3,42$; $F_{0,05}(2;18) = 3,55$, faça a ANOVA e conclua em nível de 5% de significância se os fatores A e B atuam de maneira independente.

5 (20%). Seja (X, Y) uma variável aleatória contínua bidimensional com a seguinte função densidade de probabilidade conjunta,

$$f(x,y) = \begin{cases} e^{-(x+y)} & , \quad 0 \leq x \quad \text{e} \quad 0 \leq y \\ \\ \quad 0 & , \quad \text{para outros valores } x \text{ e } y \end{cases}$$

Pede-se: Encontre a função densidade de probabilidade de $U = X + Y$.

Universidade Federal de Viçosa

Departamento de Estatística

Prova Seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 22/nov/2013

Nome: ____

Assinatura: ____.

Número do CPF ou RG: ________________________________.

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São 5 questões em páginas numeradas de 1 a 9, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

1 (20%). Os dados abaixo se referem a medidas de umidade relativa do ar de secagem de sementes (X) e do percentual de germinação das mesmas (Y), de 10 amostras de uma determinada cultura.

$X(\%)$	12	15	16	20	25	30	35	42	47	50
$Y(\%)$	80	83	82	84	85	91	92	90	95	97

Considere o seguinte modelo: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Pede-se:

a.(2%) Represente o modelo matricialmente.

b.(6%) Obtenha matricialmente a solução de mínimos quadrados $\widehat{\boldsymbol{\beta}} = \left[\hat{\beta}_0 \ \hat{\beta}_1\right]^t$.

c.(4%) Estime o percentual de germinação das sementes para uma umidade relativa do ar de secagem de 55%. Comente sobre esta estimativa.

d.(8%) Proceda à análise de variância da regressão, considerando-se um nível de significância de 1%. Informe o que está sendo testado.
Valores de F ao nível de 1% de probabilidade: $P\left(F \geq F(n_1, n_2)\right) = 0,01$.
Número de graus de liberdade: n_1 = numerador e n_2 = denominador.
$\boxed{F(1,8) = 11,26 \quad F(1,9) = 10,56 \quad F(1,10) = 10,04}$
$\boxed{F(2,8) = \ 8,65 \quad F(2,9) = \ 8,02 \quad F(2,10) = \ 7,56}$.

2 (20%). Considere uma amostra aleatória $Y_1, Y_2, \ldots, Y_n$ do seguinte modelo,

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1} & , \quad 0 < y < 1 \;\text{ e }\; \theta > 0 \\ \\ 0 & , \quad \text{para outros valores } y \end{cases}$$

Pede-se: obtenha o estimador de máxima verossimilhança de θ.

3 (20%). Considere o modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ em que:

$\boldsymbol{Y}$: vetor (aleatório) contendo os valores $y_1, y_2, \ldots, y_n$ de uma variável resposta
$\boldsymbol{X}$: matriz $(n \times p)$ de incidência
$\boldsymbol{\beta}$: vetor (constante) contendo os parâmetros $\beta_1, \beta_2, \ldots, \beta_p$
$\boldsymbol{\varepsilon}$: vetor (aleatório) contendo os erros $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$

a. (8%) Quais são as suposições que o vetor $\boldsymbol{\varepsilon}$ tem que satisfazer para que o modelo seja considerado um modelo linear de Gauss-Markov-Normal?

b. (6%) Suponha que se deseje estimar a função $\boldsymbol{C}\boldsymbol{\beta}$. Quais são as condições que a matriz $\boldsymbol{C}$ deve satisfazer para que $\boldsymbol{C}\boldsymbol{\beta}$ seja estimável?

c. (6%) O valor de $\widehat{\boldsymbol{C}\boldsymbol{\beta}}$ depende da escolha de $\widehat{\boldsymbol{\beta}}$? Explique.

4 (20%). Seja $Y_1, Y_2, \ldots, Y_n$ uma amostra aleatória do modelo Bernoulli(p), portanto $f(y_i) = p^{y_i}(1-p)^{1-y_i}$. Admita que a distribuição a priori de p seja beta(α, β), portanto $\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$. Pede-se: Com base no referencial teórico apresentado a seguir, encontre a distribuição a posteriori de p. Dica: $\Gamma(n+1) = n!$ para n inteiro.

Referencial teórico: Na inferência sob o enfoque Bayesiano admite-se que a função de verossimilhança da amostra aleatória, $L(\theta) = L(\theta | y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f(y_i|\theta)$, e a distribuição a priori do parâmetro θ, $\pi(\theta)$, forneçam a distribuição marginal (função densidade de probabilidade ou função de probabilidade) de $Y_1, Y_2, \ldots, Y_n$ por,

$$m(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(\theta | y_1, y_2, \ldots, y_n)\ \pi(\theta)\ d\theta.$$

Finalmente, a distribuição a posteriori de θ é obtida por,

$$\pi(\theta | y_1, y_2, \ldots, y_n) = \frac{L(\theta | y_1, y_2, \ldots, y_n)\ \pi(\theta)}{m(y_1, y_2, \ldots, y_n)}.$$

A distribuição a posteriori sumariza toda a informação à respeito de θ valendo-se de ambas as informações, a priori contida em $\pi(\theta)$ e dos dados contida em $L(\theta)$.

Atenção: continue a solução da questão 4 nesta página.

5 (20%). Em um estudo sobre a germinação de sementes de Pinhão Manso, deseja-se avaliar três tipos de embalagens (saco de papel, saco de ráfia e saco plástico). Deseja-se ainda, avaliar com maior precisão, quatro épocas de armazenamento das sementes (0, 30, 60 e 90 dias). Como há diferenças nos tamanhos das sementes, estas devem ser divididas em 3 lotes (sementes pequenas, médias e grandes). Cada unidade experimental será composta por 100 sementes. A variável resposta será a porcentagem de sementes germinadas. Pede-se:

a. (5%) Qual deve ser o delineamento e o esquema experimental, mais adequados para este experimento? Justifique.

b. (5%) Apresente o modelo estatístico e defina cada um de seus componentes.

c. (5%) Monte um resumo do quadro da análise de variância (ANOVA), mostrando a decomposição dos graus de liberdade.

d. (5%) Após a ANOVA, qual deve ser o procedimento adequado para a análise do fator época de armazenamento? Faça todas as considerações que julgar necessárias.

Universidade Federal de Viçosa
Departamento de Estatística

Prova Seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - novembro/2014

Nome: ____

Assinatura: ____.

Número do CPF ou RG: _______________________________.

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente. UTILIZE DUAS CASAS DECIMAIS nos cálculos.
- São 5 questões em páginas numeradas de 1 a 9, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

1 (20%). Visando estudar a poluição de um rio, um pesquisador mensurou a concentração de um determinado composto orgânico (Y, ppm) e a precipitação pluviométrica na semana anterior (X, mm) obtendo um coeficiente de correlação estimado de $r_{XY} = 0,887$. Além disso, foram obtidas as seguintes estatísticas:

$$\overline{Y} = 1,717 \quad \overline{X} = 2,023 \quad S_X = 1,233 \quad \text{e} \quad S_Y = 1,206$$

Pede-se:

a.(6%) Sabendo que,

$$\widehat{\beta}_1 = \frac{SPD_{XY}}{SQD_X}, \quad r_{XY} = \frac{SPD_{XY}}{\sqrt{SQD_X \; SQD_Y}} \quad \text{e} \quad S_W = \sqrt{\frac{SQD_W}{n-1}},$$

verifique se a seguinte igualdade é verdadeira,

$$\widehat{\beta}_1 = r_{XY} \frac{S_Y}{S_X}$$

b.(8%) Ajuste uma equação de regressão linear simples (RLS) e interprete a estimativa do coeficiente de regressão. (Dado que $\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$)

c.(6%) Considerando a equação de RLS ajustada, qual proporção da variabilidade da concentração do composto orgânico foi explicada pela precipitação pluviométrica na semana anterior?

2 (20%). Considere que $\boldsymbol{C\beta}$ seja estimável para o seguinte modelo,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad \text{com} \quad \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{I}_n\sigma^2\right)$$

em que:

$\boldsymbol{Y}$: vetor (aleatório) contendo os valores $y_1, y_2, \ldots, y_n$ de uma variável resposta
$\boldsymbol{X}$: matriz $(n \times p)$ de incidência
$\boldsymbol{\beta}$: vetor (constante) contendo os parâmetros $\beta_1, \beta_2, \ldots, \beta_p$
$\boldsymbol{\varepsilon}$: vetor (aleatório) contendo os erros $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$

a. (10%) Podemos afirmar que $\boldsymbol{C\widehat{\beta}}$ apresenta distribuição normal multivariada? Justifique a sua resposta.

b. (10%) Quais são os parâmetros (média e variância) da distribuição de $\boldsymbol{C\widehat{\beta}}$?

3 (20%). Para A e B, dois eventos pertencentes a um mesmo espaço amostral, a desigualdade de Bonferroni estabelece o seguinte limite para a probabilidade da interseção,

$$P(A \cap B) \geq P(A) + P(B) - 1. \tag{1}$$

Note que se $P(A) + P(B) < 1$ a desigualdade (1) é verdadeira porém inútil. A desigualdade de Boole, válida para qualquer sequência de eventos $\{A_i\}$ pertencentes a um mesmo espaço amostral, finita ou infinita, é dada por (caso finito, n eventos),

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) \tag{2}$$

Casella & Berger (2002) concluem que (1) e (2) são essencialmente o mesmo resultado. Esta conclusão é obtida trabalhando-se a desigualdade (2) até a obtenção de uma terceira desigualdade, a desigualdade (3) mostrada a seguir, que é uma versão mais geral da desigualdade (1).

Pede-se: mostre que a desigualdade (2) aplicada a $\{A_i^c\}$ fornece a seguinte desigualdade,

$$P(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1). \tag{3}$$

4 (20%). Considere uma variável aleatória X com a seguinte função densidade de probabilidade (em que ln é o logaritmo neperiano, ou na base e):

$$f(x) = \begin{cases} e^{-2x} & , \ 0 \leq x \leq \ln(2) \\ \\ \dfrac{15}{304}\sqrt{x+1}, & 3 \leq x \leq 8 \\ \\ 0 & , \ \text{caso contrário} \end{cases}$$

Pede-se:

a. (8%) Determine $P(X \leq \dfrac{1}{2})$.

b. (12%) O valor esperado ou a esperança matemática de uma variável aleatória X é o valor médio calculado de acordo com o modelo de probabilidade associado a X, sendo denotada por $E(X)$. Para o caso em que X é uma variável aleatória contínua, temos que $E(X) = \int\limits_{-\infty}^{+\infty} x\ f(x)\ dx$, sendo $f(x)$ a função densidade de probabilidade da variável aleatória X. Mostre que o valor esperado de X é:

$$E(X) = \frac{1133}{304} - \frac{\ln(2)}{8}.$$

5 (20%). Foi realizado um experimento com o objetivo de comparar quatro clones de cana-de-açúcar (x_1, x_2, x_3 e x_4) em relação ao percentual de manifestação de uma determinada doença (Y), sob o Delineamento em Quadrado Latino com quatro repetições por tratamento. No campo, as linhas (l_i) e as colunas (c_i) foram instaladas para controlar as possíveis diferenças entre a fertilidade e a umidade do solo, respectivamente. É sabido que quanto menores a média e a variância paramétricas para Y e menor o custo de aquisição, melhor é o clone.

	Coluna				
Linha	c_1	c_2	c_3	c_4	Total
l_1	44,3 (x_4)	26,8 (x_2)	25,3 (x_1)	36,2 (x_3)	132,6
l_2	45,8 (x_3)	34,5 (x_1)	55,1 (x_4)	35,7 (x_2)	171,1
l_3	43,9 (x_1)	56,7 (x_3)	46,2 (x_2)	64,9 (x_4)	211,7
l_4	56,1 (x_2)	75,3 (x_4)	65,6 (x_3)	55,8 (x_1)	252,8
Total	190,1	193,3	192,2	192,6	768,2

	x_1	x_2	x_3	x_4
Totais (T_i)	159,5	164,8	204,3	239,6
s_i	0,51	0,27	0,54	0,28
Custo (R$)	90	85	80	75

Teste	valor-p	Teste	valor-p	$q_{5\%}$	dms do Tukey
Bartlett	0,563	Kolmogorov-Smirnov	$> 0,15$	4,9	$q\sqrt{\frac{QMRes}{r}}$

FV	GL	SQ	QM	F_{cal}	F_{tab}
Fertilidade					4,76
Umidade		1,43			4,76
Clone					4,76
Resíduo					
Total		3069,30			

a. (10%) Complete o quadro da Análise de Variância (ANOVA) apresentado anteriormente.

b. (10%) De acordo com os dados fornecidos e os resultados da ANOVA e da aplicação do teste de Tukey ($\alpha = 0,05$), qual o melhor clone? justifique sua resposta.

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 12/Junho/2015

Nome: ____

Assinatura: ____

Número do CPF ou RG: ______________________________

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São CINCO (5) questões em páginas numeradas de 1 a 9, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

1) (20%). Seja X uma variável aleatória contínua tal que $X \sim Gama(\alpha, \beta)$. Logo sua função densidade de probabilidade é dada por,

$$f\left(x\right) = \frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}},$$

sendo $\alpha > 0$, $\beta > 0$ e $0 < x < +\infty$. Pede-se:

a) Mostre que a função geradora de momentos de X é $M_X(t) = \left(\dfrac{1}{1-\beta t}\right)^{\alpha}$

b) Com base na função geradora de momentos, obtenha $E(X)$ e $V(X)$.

2) (20%). Considere o seguinte intervalo de confiança, para o parâmetro θ:

$$IC_{(1-\alpha)} : \widehat{\theta} \pm t_{\alpha/2} \frac{S(\widehat{\theta})}{\sqrt{n}}$$

a) Por que a denominação "Intervalo de Confiança" e não "Intervalo de probabilidade"?

b) Na Inferência Bayesiana obtém-se o intervalo de credibilidade, e não o intervalo de confiança. Explique qual é a diferença entre estes dois tipos de intervalos, em termos de afirmações probabilísticas com relação a θ.

3) (20%). Na teoria dos modelos lineares, escritos matricialmente como $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, são abordados tópicos sobre análise de *modelos de posto completo* e de *posto incompleto*. Referente a esse assunto, responda às questões abaixo:

a) Explique o que significa ser de posto completo e posto incompleto.

b) Monte um exemplo hipotético de um modelo de posto completo com 6 observações. Apresente o modelo estatístico com a descrição dos seus componentes e represente as matrizes $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}$ associadas ao modelo.

c) Monte um exemplo hipotético, de um modelo de posto incompleto com 6 observações. Apresente o modelo estatístico com a descrição dos seus componentes e represente as matrizes $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}$ associadas ao modelo.

4) (20%). Instalou-se um experimento para avaliar o efeito de 4 tipos de farinhas de trigo e de 3 conservantes, sobre a qualidade de pães. Visando-se reduzir custos no experimento e considerando-se a capacidade de 7,5 kg do misturador utilizado, cada tipo de farinha foi utilizado para preparar três porções de massa de 7,5 kg no misturador. Em seguida, cada porção de 7,5 kg de massa preparada foi dividida em 3 partes de 2,5 kg e em cada parte foi incorporado um dos conservantes testados.

a) Identifique o esquema e o delineamento experimental utilizado.

b) Apresente o quadro parcial da ANOVA com as fontes de variação e respectivos graus de liberdade.

5) (20%). Sejam X_1 e X_2 variáveis aleatórias contínuas conjuntamente distribuídas com função densidade de probabilidade $f_{X_1,X_2}(x_1,x_2)$ e, $\mathfrak{X} = \{(x_1,x_2) \mid f_{X_1,X_2}(x_1,x_2) > 0\}$. Assuma que:

i) $y_1 = g_1(x_1,x_2)$ e $y_2 = g_2(x_1,x_2)$ é uma transformação biunívoca de $\mathfrak{X}$ em $\mathfrak{D}$, sendo $\mathfrak{D}$ um subconjunto do plano $y_1 \times y_2$;

ii) As derivadas parciais de $x_1 = g_1^{-1}(y_1,y_2)$ e $x_2 = g_2^{-1}(y_1,y_2)$ são contínuas em $\mathfrak{D}$;

iii) O Jacobiano da transformação (J) é diferente de zero para $(y_1,y_2) \in \mathfrak{D}$.

Então a densidade conjunta de $Y_1 = g_1(X_1,X_2)$ e $Y_2 = g_2(X_1,X_2)$ é dada por

$$f_{Y_1,Y_2}(y_1,y_2) = |J|\, f_{X_1,X_2}\left(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2)\right) I_{\mathfrak{D}}(y_1,y_2),$$

sendo $I_{\mathfrak{D}}(y_1,y_2) = \begin{cases} 1, & (y_1,y_2) \in \mathfrak{D} \\ 0, & (y_1,y_2) \notin \mathfrak{D} \end{cases}$ e, $J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$.

Este consiste no chamado método do jacobiano, muito utilizado quando desejamos encontrar a distribuição conjunta de duas funções de variáveis aleatórias que satisfazem as condições deste método.

Sejam X_1, e X_2 variáveis aleatórias independentes e identicamente distribuídas, ambas com distribuição normal padrão, $N(0,1)$. Considere $Y_1 = X_1 + X_2$ e $Y_2 = \dfrac{X_1}{X_2}$

a) Utilizando o método do Jacobiano, determine $f_{Y_1,Y_2}(y_1,y_2)$;

b) Obtenha a distribuição marginal de Y_2, isto é, $f_{Y_2}(y_2)$

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 20/Novembro/2015

Nome: ______________________________

Assinatura: ______________________________

Número do CPF ou RG: ______________________________

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São CINCO (5) questões em páginas numeradas de 1 a 7, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Formulário

$$P(Y \geq r) \leq \frac{E(Y)}{r}$$

$$P(|Y - \mu_Y| < k\sigma_Y) \geq 1 - \frac{1}{k^2}$$

$$P(Y \geq \mu_Y + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

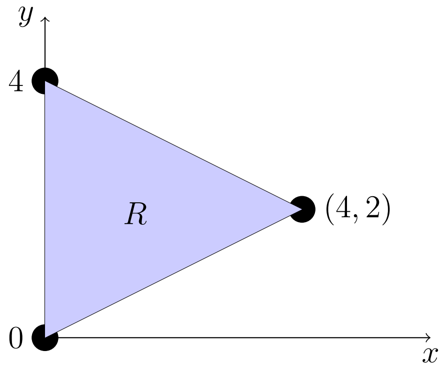
Tabela 1: Qui-quadrado com n graus de liberdade e $P\left[\chi^2 \geq \chi^2_{tab}\right] = \alpha$.

n	α					
	0,99	0,975	0,90	0,75	0,50	0,10
1	0,00	0,00	0,02	0,10	0,45	2,71
2	0,02	0,10	0,21	0,58	1,39	4,61
3	0,11	0,35	0,58	1,21	2,37	6,25
4	0,30	0,71	1,06	1,92	3,36	7,78

Tabela 2: t de Student com ν graus de liberdade e $P\left[-t_c \leq T \leq t_c\right] = 1 - \alpha$.

n	Nível de probabilidade bilateral				
	0,20	0,10	0,04	0,02	0,01
1	3,08	6,31	15,89	31,82	63,66
2	1,89	2,92	4,85	6,96	9,92
3	1,64	2,35	3,48	4,54	5,84
4	1,53	2,13	3,00	3,75	4,60

1) (20%). Considere a variável aleatória bidimensional conjuntamente contínua (X, Y) com função densidade de probabilidade conjunta $f_{X,Y}(x, y) = \frac{1}{8} I_R(x, y)$, sendo R a região dada na figura a seguir e, $I_R(x, y) = \begin{cases} 1, & \text{se } (x, y) \in R \\ 0, & \text{se } (x, y) \notin R \end{cases}$.



Determine:

a) (**06%**) $F_{X,Y}(1/4, 9/4)$, isto é, o valor da função distribuição acumulada avaliada em $(1/4, 9/4)$;

b) (**14%**) Se $W = Y - 2X - 1$ e $V = 3X - \frac{Y}{2}$, determine $\text{Cov}(W, V)$;

2) (20%). Considere a seguinte definição sobre o Método da Máxima Verossimilhança (Fisher): "O estimador $\widehat{\theta}\,(\underline{X})$" é o estimador de máxima verossimilhança (EMV) de θ, quando $P\left[\underline{X}, \widehat{\theta}\,(\underline{X})\right] = \max P\,(\underline{X}, \theta)\,, \forall \theta \in \Theta$.

Suponha que θ pode assumir os valores 0 (zero) ou 1 (um) ou 2 (dois) e que $P(x, \theta)$ é dada pelo quadro a ser completado abaixo:

x	θ		
	0	1	2
-1		0,30	0,15
0	0,20		0,10
1	0,70	0,10	

a) Qual é o estimador EMV de θ quando $x = 0$? Por que?

b) Qual é o estimador EMV de θ quando $x = 1$? Por que?

3) (20%). Seja X a variável aleatória que expressa os números possíveis ao lançarmos um dado honesto.

a) (4%) Calcule $E(X)$ e $Var(X)$;

b) (4%) Calcule o valor exato de $p = P(X \geq 6)$;

c) (4%) Calcule $p = P(X \geq 6)$ pela desigualdade de Markov;

d) (4%) Calcule $p = P(X \geq 6)$ pela desigualdade de Chebyshev;

e) (4%) Os resultados de c) e d), contradizem b)? Justifique.

4) (20%). Sejam X e Y variáveis aleatórias independentes com distribuição $N(0,5)$. Pede-se:

a) $P\left(X^2 + Y^2 \leq 2,9\right)$

b) $P\left[(X+Y)^2 < 4,5\right]$

5) (20%). Considere os resultados parciais obtidos de um experimento fatorial, no qual foram estudados três níveis do fator A (variedades de trigo) e dois níveis do fator B (aplicações de fungicidas às sementes), segundo o delineamento em blocos casualizados e com quatro repetições por tratamento para a avaliação da produção de grãos (Y). É sabido que quanto maior a média e menor a variância de Y, em termos paramétricos (populacionais), além do menor custo de aquisição do tratamento, melhor. Se uma empresa insiste em aplicar o fungicida b_2, por ser mais barato do que o b_1, qual é a variedade de trigo (a_i) a ser combinada com ele (b_2) de acordo com os testes F e de Tukey apropriados, sabendo-se dos custos de a_1 (R$11,00), a_2 (R$12,00) e a_3 (R$13,00) e considerando-se satisfeitas as pressuposições da análise de variância ($\alpha = 0,05$)? justifique sua resposta.

Totais

	b_1	b_2	
a_1	225	167	392
a_2	325	227	552
a_3	290	230	520
	840	624	1464

$$SQ_{\text{Tratamento}} = 3863$$

$q_A = q_{A/B} = 3,67$	$q_B = q_{B/A} = 3,01$	$\Delta = q\sqrt{\frac{QMRes}{r}}$	r = número de repetições que deram origem às médias comparadas

FV	GL	SQ	QM	f_{cal}	f_{tab}
Bloco		2552			3,29
A					3,68
B					4,54
AxB					3,68
Resíduo		177			
Total		6592			

FV	GL	SQ	QM	f_{cal}	f_{tab}
A/b_1		1287,50			3,68
A/b_2		631,5			3,68
B/a_1		420,5			4,54
B/a_2		1200,50			4,54
B/a_3		450			4,54
Resíduo					

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 7/Novembro/2016

Nome: ____

Assinatura: ____

Número do CPF ou RG: ____

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São SEIS (**6**) questões em páginas numeradas de 1 a 11, total de 100 pontos.
- Nas páginas 10 e 11 temos formulário e tabela.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Nota

1) (20%) Sejam X_1 e X_2 variáveis aleatórias contínuas com função densidade de probabilidade conjunta dada por $f_{X_1,X_2}(x_1, x_2)$. Sejam $Y_1 = X_1 + X_2$ e $Y_2 = X_1 - X_2$. Pede-se:

a) (8%) Encontre a função densidade de probabilidade conjunta de Y_1 e Y_2 em termos de $f_{X_1,X_2}(x_1, x_2)$.

b) (6%) Se X_1 e X_2 são variáveis aleatórias independentes e possuem distribuição exponencial com parâmetros λ_1 e λ_2, respectivamente, então, qual é a função densidade de probabilidade conjunta de Y_1 e Y_2?

Nota

c) (6%) Se X_1 e X_2 são variáveis aleatórias independentes e possuem distribuição uniforme $(0, 1)$, qual é a função densidade de probabilidade conjunta de Y_1 e Y_2?

Nota

2) (20%) A estatística do teste da razão de verossimilhanças para testar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ é dada por:

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x})}{\sup\limits_{\theta \in \Theta} L(\theta \mid \boldsymbol{x})},$$

em que Θ é o espaço paramétrico irrestrito e Θ_0 é o espaço paramétrico restrito à hipótese nula. Denotando $\widehat{\theta}$ como o estimador de máxima verossimilhança de θ sobre o espaço Θ e $\widehat{\theta}_0$ o estimador de máxima verossimilhança de θ sobre o espaço paramétrico Θ_0, a estatística do teste da razão de verossimilhanças pode ser escrito como:

$$\lambda(\boldsymbol{x}) = \frac{L\left(\widehat{\theta}_0 \mid \boldsymbol{x}\right)}{L\left(\widehat{\theta} \mid \boldsymbol{x}\right)},$$

Sobre este teste descreva: qual sua utilização, apresentando exemplos; em que condições é empregado; qual distribuição tem a estatística do teste; graus de liberdade, etc.

Nota

3) (20%) A distribuição normal representa um papel importante em um grande conjunto de análises estatísticas. Os principais fatores para isto, é que a distribuição normal é tratável analiticamente, pode ser aplicada em muitos modelos de população e, em condições moderadas, a distribuição normal pode aproximar uma grande variedade de distribuições de probabilidade em grandes amostras (Teorema Central do Limite).

Considere que $X \sim N(\mu, \sigma^2)$ cuja função de densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

sendo $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+^*$. A distribuição normal apresenta algumas características, dentre elas:

a) $x = \mu$ é um ponto de máximo de $f(x)$.

b) $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f(x)$.

Prove, utilizando ferramentas do cálculo, as informações a) e b).

Nota

4) (20%) Considere um experimento instalado no delineamento em blocos casualizados, com 4 repetições, no esquema de parcelas subdivididas. Avaliou-se o efeito de 5 tipos de irrigação (fator primário) e de 2 variedades (fator secundário) sobre a produção de cana de açúcar (t/ha).
OBS: utilize 2 casas decimais nos cálculos.

Tabela 1: Análise de variância para dados de produção de cana de açúcar

FV	GL	SQ	QM	F	F(5%)
Bloco		782,18			3,49
Irrigação		11.524,38			3,26
Resíduo (a)		893,85	74,49		
(Parcelas)		(13.200,41)			
Variedade		405,13		22,60	4,54
I × V		105,40			3,06
Resíduo (b)		269,02	17,93		
Total		13.979,96			

a) Complete o quadro da ANOVA.

b) A respeito da interação entre os fatores tipo de irrigação e variedade (I × V), conclua a 5% de significância.

c) Aplique o teste de Tukey a 5% de significância para verificar se o contraste entre médias dos tipos de irrigação 1 e 3 é estatisticamente nulo.

Dado:

$\widehat{C} = \widehat{m}_1 - \widehat{m}_3 = 94,3 - 68,9 = 25,4$

$\Delta = q\sqrt{\frac{\text{QMres}}{r}}$

$q_{5\%} = 4,51$

Nota

d) Baseado no resultado do teste de Tukey do item c), é possível concluir a respeito da significância do contraste C, pelo teste de Duncan a 5%? **Sim ou não e justifique sua resposta.**

e) Com as informações disponíveis, é possível concluir se o contraste entre médias das variedades 1 e 2 é estatisticamente, nulo a 5% de significância pelo teste de t de Student? **Sim ou não e justifique sua resposta.**

Nota

5) (10%) Considere o modelo de regressão (Gauss-Markov) com duas variáveis explicativas, uma variável resposta, e quatro observações:

$$Y_i = a + b_1 X_{i1} + b_2 X_{i2} + e_i.$$

a) (6%) Apresente este modelo na forma matricial $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Represente todas as matrizes em suas dimensões apropriadas.

b) (4%) Dê dois exemplos de funções lineares dos parâmetros desse modelo que seriam estimáveis.

Nota

6) (10%).Considere o modelo linear (Gauss-Markov) referente ao delineamento em blocos casualizados com dois tratamentos ($i = 1, 2$) e três repetições ($j = 1, 2, 3$):

$$Y_{ij} = a + t_i + b_j + e_{ij}.$$

a) (6%) Escreva o modelo na forma matricial $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Represente todas as matrizes em suas dimensões apropriadas.

b) (4%) Dê dois exemplos de funções lineares dos parâmetros desse modelo que seriam estimáveis.

Nota

Formulário

$X \sim Exp(\lambda) \Longrightarrow f_X(x) = \lambda \mathrm{e}^{-\lambda x} I_{(0,+\infty)}(x), \lambda \in \mathbb{R}_+^*$
$X \sim U(0,1) \Longrightarrow f_X(x) = I_{(0,1)}(x)$

Tabela 2: Tabela da distribuição t de Student para $P\left[-t_c \leq T \leq t_c\right] = 1 - \alpha$, na primeira coluna temos os graus de liberdade (ν) e na primeira linha o nível de significância.

ν	Nível de probabilidade bilateral								
	$0,20$	$0,10$	$0,05$	$0,04$	$0,02$	$0,01$	$0,005$	$0,002$	$0,001$
1	3,0777	6,3138	12,7062	15,8945	31,8205	63,6567	127,3213	318,3088	636,6192
2	1,8856	2,9200	4,3027	4,8487	6,9646	9,9248	14,0890	22,3271	31,5991
3	1,6377	2,3534	3,1824	3,4819	4,5407	5,8409	7,4533	10,2145	12,9240
4	1,5332	2,1318	2,7764	2,9985	3,7469	4,6041	5,5976	7,1732	8,6103
5	1,4759	2,0150	2,5706	2,7565	3,3649	4,0321	4,7733	5,8934	6,8688
6	1,4398	1,9432	2,4469	2,6122	3,1427	3,7074	4,3168	5,2076	5,9588
7	1,4149	1,8946	2,3646	2,5168	2,9980	3,4995	4,0293	4,7853	5,4079
8	1,3968	1,8595	2,3060	2,4490	2,8965	3,3554	3,8325	4,5008	5,0413
9	1,3830	1,8331	2,2622	2,3984	2,8214	3,2498	3,6897	4,2968	4,7809
10	1,3722	1,8125	2,2281	2,3593	2,7638	3,1693	3,5814	4,1437	4,5869
11	1,3634	1,7959	2,2010	2,3281	2,7181	3,1058	3,4966	4,0247	4,4370
12	1,3562	1,7823	2,1788	2,3027	2,6810	3,0545	3,4284	3,9296	4,3178
13	1,3502	1,7709	2,1604	2,2816	2,6503	3,0123	3,3725	3,8520	4,2208
14	1,3450	1,7613	2,1448	2,2638	2,6245	2,9768	3,3257	3,7874	4,1405
15	1,3406	1,7531	2,1314	2,2485	2,6025	2,9467	3,2860	3,7328	4,0728
16	1,3368	1,7459	2,1199	2,2354	2,5835	2,9208	3,2520	3,6862	4,0150
17	1,3334	1,7396	2,1098	2,2238	2,5669	2,8982	3,2224	3,6458	3,9651
18	1,3304	1,7341	2,1009	2,2137	2,5524	2,8784	3,1966	3,6105	3,9216
19	1,3277	1,7291	2,0930	2,2047	2,5395	2,8609	3,1737	3,5794	3,8834
20	1,3253	1,7247	2,0860	2,1967	2,5280	2,8453	3,1534	3,5518	3,8495
21	1,3232	1,7207	2,0796	2,1894	2,5176	2,8314	3,1352	3,5272	3,8193
22	1,3212	1,7171	2,0739	2,1829	2,5083	2,8188	3,1188	3,5050	3,7921
23	1,3195	1,7139	2,0687	2,1770	2,4999	2,8073	3,1040	3,4850	3,7676
24	1,3178	1,7109	2,0639	2,1715	2,4922	2,7969	3,0905	3,4668	3,7454
25	1,3163	1,7081	2,0595	2,1666	2,4851	2,7874	3,0782	3,4502	3,7251
26	1,3150	1,7056	2,0555	2,1620	2,4786	2,7787	3,0669	3,4350	3,7066
27	1,3137	1,7033	2,0518	2,1578	2,4727	2,7707	3,0565	3,4210	3,6896
28	1,3125	1,7011	2,0484	2,1539	2,4671	2,7633	3,0469	3,4082	3,6739
29	1,3114	1,6991	2,0452	2,1503	2,4620	2,7564	3,0380	3,3962	3,6594
30	1,3104	1,6973	2,0423	2,1470	2,4573	2,7500	3,0298	3,3852	3,6460
31	1,3095	1,6955	2,0395	2,1438	2,4528	2,7440	3,0221	3,3749	3,6335
32	1,3086	1,6939	2,0369	2,1409	2,4487	2,7385	3,0149	3,3653	3,6218
33	1,3077	1,6924	2,0345	2,1382	2,4448	2,7333	3,0082	3,3563	3,6109
34	1,3070	1,6909	2,0322	2,1356	2,4411	2,7284	3,0020	3,3479	3,6007
35	1,3062	1,6896	2,0301	2,1332	2,4377	2,7238	2,9960	3,3400	3,5911
36	1,3055	1,6883	2,0281	2,1309	2,4345	2,7195	2,9905	3,3326	3,5821
37	1,3049	1,6871	2,0262	2,1287	2,4314	2,7154	2,9852	3,3256	3,5737
38	1,3042	1,6860	2,0244	2,1267	2,4286	2,7116	2,9803	3,3190	3,5657
39	1,3036	1,6849	2,0227	2,1247	2,4258	2,7079	2,9756	3,3128	3,5581
40	1,3031	1,6839	2,0211	2,1229	2,4233	2,7045	2,9712	3,3069	3,5510
50	1,2987	1,6759	2,0086	2,1087	2,4033	2,6778	2,9370	3,2614	3,4960
60	1,2958	1,6706	2,0003	2,0994	2,3901	2,6603	2,9146	3,2317	3,4602
120	1,2886	1,6577	1,9799	2,0763	2,3578	2,6174	2,8599	3,1595	3,3735
$+\infty$	1,2816	1,6449	1,9600	2,0537	2,3263	2,5758	2,8070	3,0902	3,2905

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 31/Maio/2017

Nome: ____

Assinatura: ____

Número do CPF ou RG: ____

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São SEIS (**6**) questões em páginas numeradas de 1 a 12, total de 100 pontos.
- Nas página 12 temos formulário.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Nota

1) (20%) Em situações em que temos uma única variável resposta, porém com várias variáveis explicativas, dentro do contexto de modelos lineares, o pesquisador poderá estar interessado em selecionar algumas variáveis visando obter um modelo mais parcimonioso. Assim, seu objetivo é selecionar, dentre vários, um possível melhor modelo de regressão linear envolvendo um subconjunto dessas várias variáveis explicativas iniciais. Há vários procedimentos na literatura que podem ser usados para selecionar variáveis explicativas dentro desse contexto de modelos lineares de posto completo. Um desses procedimentos é o chamado “stepwise regression”. Pede-se:

a) Explique o que vem a ser modelo de posto completo. Ilustre matricialmente um modelo desse tipo. Também apresente, matricialmente, um exemplo de modelo de posto incompleto com as justificativas.

Nota

b) Explique, em no máximo duas páginas, como funciona o procedimento "stepwise regression". Para ilustrar seu raciocínio, idealize uma situação, inclusive apresentando quadros de ANOVA, e os passos que levariam a um eventual leitor (leigo) de sua resposta o perfeito entendimento do procedimento apresentado.

Nota

2) (20%) Um processo de mortalidade em tempo contínuo, a uma taxa constante de $\lambda = 0,5$ por ano, segue uma distribuição exponencial. Assim, faça o que se pede:

a) Obtenha a função de distribuição acumulada (f.d.a.);

b) Usando a f.d.a., calcule a probabilidade de sobrevivência de um ano ou mais;

c) Qual a probabilidade de sobrevivência de um ano e meio ou mais?

d) Qual a probabilidade de sobrevivência por meio ano ou mais?

Nota

e) Qual a probabilidade de sobrevivência por um ano e meio ou mais, dado que sobreviveu-se até um ano?

f) Qual a característica da distribuição exponencial que fica evidente a partir dos itens anteriores? Qual expressão matemática que simboliza esta característica?

Nota

3) (20%) A distribuição de Poisson é uma distribuição discreta de probabilidade amplamente aplicada em diversos fenômenos e é, frequentemente, usada para modelar o número de ocorrências de um evento por um certo período de tempo ou por um determinado espaço (volume, área). Uma variável aleatória discreta X segue a distribuição de Poisson com parâmetro λ ($\lambda > 0$), se sua função de probabilidade for dada por:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \cdots$$

em que λ é a taxa média de ocorrência do evento.

Utilizando séries de Taylor e propriedades de somatórios, mostre detalhadamente que:

a) $\sum\limits_{x=0}^{+\infty} P(X = x) = 1$

Nota

b) $\mathrm{var}(X) = \lambda$; (Sugestão utilize a seguinte expressão para o cálculo da variância: $\mathrm{var}\left(X\right) = E\left(X^2\right) - \left[E\left(X\right)\right]^2 = E\left(X^2 - X\right) + E\left(X\right) - \left[E\left(X\right)\right]^2$.)

Nota

4) (20%) Considere um experimento instalado para avaliar o efeito de cinco tratamentos com cinco repetições. Apresente o quadro parcial da Anova com as fontes de variação e os respectivos graus de liberdade para cada delineamento indicado a seguir.

a) Delineamento Inteiramente Casualizado

b) Delineamento em Blocos Casualizados

Nota

c) Delineamento em Quadrado Latino

d) Para cada delineamento, indique as pressuposições para que o resultado da Anova seja válida.

Nota

5) (10%) Sejam X e Y duas variáveis aleatórias independentes com distribuição uniforme contínua no intervalo $(-1, 1)$ e $W = X + Y$. Determine a função densidade de probabilidade de W.

Nota

6) (10%) Na Inferência Bayesiana, utilizando o teorema de Bayes sob o enfoque de variáveis aleatórias, podemos escrever:

$$\underbrace{P(\theta|X)}_{\substack{\text{Probabilidade} \\ \text{a posteriori}}} \propto \underbrace{P(X|\theta)}_{\substack{\text{Função de} \\ \text{Verossimilhança}}} \times \underbrace{P(\theta)}_{\substack{\text{Distribuição} \\ \text{a priori}}}$$

a) Descreva resumidamente o que é uma distribuição a priori e como ela é incorporada em uma análise estatística?

b) O que representa a distribuição a posteriori e qual a diferença fundamental em relação a inferência clássica (frequentista)?

Nota

Formulário

$X \sim Exp(\lambda) \Longrightarrow f_X(x) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x), \lambda \in \mathbb{R}_+^*$	
$X \sim U(a,b) \Longrightarrow f_X(x) = \dfrac{1}{b-a} I_{(a,b)}(x)$	
$E[X^n] = \int_{-\infty}^{+\infty} x^n f(x) dx$	$E[X^n] = \sum_{i=1}^{n} x_i^n P[X = x_i]$
$\text{var}[X] = E[X^2] - (E[X])^2$	
$Z = X + Y \Longrightarrow f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx = \int_{-\infty}^{+\infty} f_{X,Y}(z-y, y) dx$	

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 8/Novembro/2017

Número de inscrição: ______________________________

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São CINCO (**5**) questões em páginas numeradas de 1 a 7, total de 100 pontos.
- Na página 7 temos formulário.
- É permitida a utilização de calculadora científica.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Nota

1) (20%) No contexto de modelos lineares de posto completo é discutido sobre multicolinearidade. Pede-se: (se necessário, dê exemplos hipotéticos, que possam contribuir para melhor esclarecer suas ideias).

a) **(05%)** Conceito de multicolinearidade.

b) **(05%)** Formas de detecção.

c) **(05%)** Efeito da multicolinearidade sobre a qualidade de ajuste nesses modelos.

d) **(05%)** Formas de remediar os problemas causados pela multicolinearidade.

Nota

2) (20%) Seja X uma variável aleatória discreta com a seguinte função de probabilidade,

x_i	0	1	2	3	Total
$P(x_i \mid \theta)$	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{(1-\theta)}{3}$	1,00

Pede-se: Obtenha uma estimativa de máxima verossimilhança para θ com base na seguinte amostra aleatória observada de valores x_i para $i = 1, 2, \ldots, 10$,

$$X = \{3, 0, 2, 1, 3, 2, 1, 0, 2, 1\}.$$

Nota

3) (20%) Sejam X e Y duas variáveis aleatórias independentes com $E[X] = \mu_X$, $E[Y] = \mu_Y$, $\text{var}[X] = \sigma_X^2$ e $\text{var}[Y] = \sigma_Y^2$, todos finitos e seja $Z = XY$.

a) **(10%)** Determine $\text{var}[Z]$ em termos de μ_X, μ_Y, σ_X^2 e σ_Y^2;

b) **(10%)** Sejam $X \sim N(\mu_X; \sigma_X^2)$, $Y \sim N(\mu_Y; \sigma_Y^2)$ e $Z = XY$, sendo $\mu_X \in \mathbb{R}$, $\mu_Y \in \mathbb{R}$, $\sigma_X^2 \in \mathbb{R}_+^*$ e $\sigma_Y^2 \in \mathbb{R}_+^*$. Utilizando o item a) determine condições para μ_X, μ_Y, σ_X^2 e σ_Y^2 tais que

$$\text{var}[Z] = \text{var}[X]\,\text{var}[Y].$$

Nota

4) (20%) Sejam X e Y duas variáveis aleatórias contínuas e independentes, ambas uniformemente distribuídas no intervalo $[0, 1]$. Se $W = XY$, determine $P\left(W < \frac{1}{2}\right)$.

Nota

5) (10%) Considere um experimento instalado no delineamento em blocos casualizados, com 4 repetições, no esquema de parcelas subdivididas. Neste experimento avaliou-se o efeito de 5 tipos de irrigação (fator primário) e de 2 variedades (fator secundário).

a) **(10%)** Apresente o quadro parcial da ANOVA com as Fontes de Variação (FV) e os respectivos Graus de Liberdade (GL).

b) **(10%)** Neste experimento, explique como deve ser feito o estudo dos fatores tipo de irrigação e variedade, nas duas situações abaixo, informando qual o resíduo adequado:

b_1)**(05%)** Interação não significativa;

b_2)**(05%)** Interação significativa.

Nota

Formulário

$$\boxed{T = XY \Longrightarrow f_T(t) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{t}{x}\right) dx = \int_{-\infty}^{+\infty} |y| f_{X,Y}\left(\frac{t}{y}, y\right) dy}$$

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 11/junho/2018

Nome: ____

Assinatura: ____

Número do CPF ou RG: ____

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- É permitida a utilização de calculadora;
- São CINCO (**5**) questões em páginas numeradas de 1 a 7, total de 100 pontos.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Nota

1) (20%) Seja $Y_1, Y_2, \ldots, Y_{20} \overset{iid}{\sim} f(y|p) = p^y(1-p)^{1-y}$ com $y = 0$ ou $y = 1$ e $0 \leq p \leq 1$. Observe que $n = 20$ é o tamanho da amostra. Admita que a distribuição a *priori* para p seja $\pi(p) = 1$, isto é, $p \sim$ uniforme contínua $[0, 1]$. Pede-se: Se foi observado $\sum_{i=1}^{20} y_i = 12$, obtenha a estimativa da média a *posteriori* para p. Ou seja, do valor esperado de p com base em $\pi(p|\boldsymbol{y})$.

Dica: Se $X \sim \text{beta}(\alpha, \beta)$, então $f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, com $0 \leq x \leq 1$, $\alpha > 0$ e $\beta > 0$. $E(X) = \frac{\alpha}{\alpha+\beta}$.

Nota

2) (20%) Suponha que foi realizado um experimento para avaliar o efeito de um fator quantitativo sobre uma variável resposta.

Nível do fator quantitativo (x_i)	Variável resposta (y_i)
5	3
10	5
15	10
20	11
25	13

Suponha que uma análise de regressão, baseando-se num Modelo Linear de Gauss-Markov Normal (MLGMN), para os dados desse experimento retornou os seguintes resultados

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0,60 \\ 0,52 \end{bmatrix}$$

a) (7%) Apresente a matriz $\mathbf{X}'\mathbf{X}$ referente ao MLGMN utilizado na análise de regressão.

Nota

b) (6%) Obtenha $Var(\mathbf{y})$. Apresente todos os elementos dessa matriz indicando a ordem da mesma.

c) (7%) Expresse a hipótese de nulidade para o teste da significância do efeito do fator quantitativo em termos de

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$$

em que $\mathbf{C}\boldsymbol{\beta}$ é uma função estimável.

Nota

3) (20%)

Croqui do experimento

	Bloco1			Bloco2			Bloco3		
	(10,25)	(9,47)	(9,88)	(9,97)	(10,77)	(8,56)	(9,57)	(10,23)	(10,51)
	a_2b_3	a_1b_1	a_3b_2	a_3b_1	a_2b_3	a_1b_2	a_3b_2	a_2b_1	a_1b_3
	(9,61)	(8,93)	(9,87)	(11,07)	(11,54)	(8,49)	(10,16)	(10,40)	(9,55)
	a_2b_1	a_1b_3	a_3b_3	a_3b_2	a_2b_2	a_1b_3	a_3b_1	a_2b_3	a_1b_1
	(10,22)	(9,48)	(10,58)	(11,06)	(9,24)	(10,95)	(10,28)	(9,22)	(10,88)
	a_2b_2	a_1b_2	a_3b_1	a_3b_3	a_2b_1	a_1b_1	a_3b_3	a_2b_2	a_1b_2
Totais	30,08	27,88	30,33	32,10	31,55	28,00	30,01	29,85	30,94

a) (10%) De acordo com o croqui experimental de um Delineamento em Blocos Casualizados (DBC) para avaliar os fatores A e B, identifique o esquema experimental (fatorial ou parcelas subdivididas) e apresente o quadro parcial da Anova com as fontes de variação e respectivos graus de liberdade.

b) (5%) Calcule a soma de quadrados para o fator A.

c) (5%) Coloque os testes (Duncan, Dunnett, SNK, Scheffé, t de Student e Tukey) em ordem decrescente de sensibilidade. Discuta a respeito da relação entre sensibilidade do teste e erro tipo I.

Nota

4) (20%) Sejam $X \sim \text{Exp}\,(\lambda_1)$ e $Y \sim \text{Exp}\,(\lambda_2)$, sendo $\lambda_1 > 0$, $\lambda_2 > 0$ e X e Y variáveis aleatórias independentes. Determine a função densidade de probabilidade de W, sendo $W = |X - Y|$.

Nota

5) (20%) Seja X uma variável aleatória mista com função de distribuição

$$F_X\left(x\right) = \begin{cases} 0, & \text{se } x \leq 0 \\ x, & \text{se } 0 < x \leq \frac{1}{2} \\ 1, & \text{se } x \geq \frac{1}{2}. \end{cases}$$

Obtenha a função acumulada da parte discreta de X e a função densidade de probabilidade da parte contínua de X.

Nota

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 19/Novembro/2018

Número de inscrição:

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São SEIS (**6**) questões em páginas numeradas de 1 a 11, total de 100 pontos.
- É permitida a utilização de calculadora científica.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Revisão Total

1) (20%) Sejam X e Y variáveis aleatórias contínuas. Assumindo que as esperanças condicionais existem, bem como suas propriedades, faça o que se pede:

a) **(10%)** Se X e Y são independentes, prove que $E[X \mid Y] = E[X]$.

b) **(10%)** Considere a função densidade de probabilidade de X e Y

$$f(x, y) = \begin{cases} 4xy, & \text{se } 0 \leq x \leq 1 \text{ e } 0 \leq y \leq 1 \\ 0, & \text{para os demais valores de } x \text{ e } y \end{cases}$$

Mostre numericamente que esta propriedade é verdadeira.

Nota

2) (10%) Considere o modelo linear (Gauss-Markov) referente ao delineamento em blocos casualizados com dois tratamentos ($i = 1, 2$) e três repetições ($j = 1, 2, 3$):

$$Y_{ij} = a + t_i + b_j + e_{ij}.$$

a) **(05%)** Escreva o modelo na forma matricial $y = X\beta + \varepsilon$. Represente todas as matrizes em suas dimensões apropriadas.

b) **(05%)** Podemos estimar o parâmetro t_1 nesse modelo? Justifique e comente sua resposta.

Nota

3) (10%) Formas quadráticas são comumente apresentadas e discutidas na disciplina de modelos lineares. Pede-se:

a) **(04%)** Dê um pequeno exemplo de uma forma quadrática, primeiramente em sua forma matricial, e posteriormente na correspondente forma expandida. Dica: use um vetor y de tamanho 3.

b) **(03%)** Discuta sobre a importância do estudo das formas quadráticas na disciplina de modelos lineares, principalmente quando tratamos sobre a análise de variância (ANOVA).

c) **(03%)** Considere o modelo da questão **2)**. Escreva a SQTotal (não corrigida para a média) na sua forma quadrática matricial.

Nota

4) (20%) Considere um experimento instalado para avaliar o efeito de cinco tratamentos com cinco repetições. Apresente o modelo estatístico com a descrição dos fatores incluídos no modelo para cada delineamento indicado a seguir.

a) **(05%)** Delineamento Inteiramente Casualizado.

b) **(05%)** Delineamento em Blocos Casualizados.

Nota

c) **(05%)** Delineamento em Quadrado Latino.

d) **(05%)** Para cada delineamento citado anteriormente, descreva como é feita a casualização.

Nota

5) (20%) Considere que as variáveis aleatórias $\boldsymbol{Y} = \{Y_1, Y_2, \ldots Y_n\}$ satisfaçam ao modelo,

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

em que $x_1, x_2, \ldots, x_n$ são constantes conhecidas e $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ são variáveis aleatórias independentes e identicamente distribuídas $N(0, \sigma^2)$, com σ^2 um valor desconhecido. Dica: se $W \sim N(\mu, \delta^2)$, então a função densidade de probabilidade é dada por,

$$f_W(w) = \frac{1}{\delta\sqrt{2\pi}} \exp\left\{\frac{-(w-\mu)^2}{2\delta^2}\right\}, \quad \delta > 0, \quad -\infty < \mu < \infty \quad \text{e} \quad -\infty < w < \infty.$$

Pede-se:

a) **(05%)** Encontre uma estatística bidimensional $T(\boldsymbol{Y})$ suficiente para (β, σ^2).

Nota

b) **(05%)** Encontre $\widehat{\beta}_{MV}$ o estimador de máxima verossimilhança para β.

Nota

c) **(05%)** Seja $\widehat{\beta} = \dfrac{\sum Y_i}{\sum x_i}$, mostre que ambos, $\widehat{\beta}_{MV}$ e $\widehat{\beta}$, são estimadores não viesados para β.

d) **(05%)** Compare os dois estimadores quanto à eficiência, calcule a eficiência relativa $\mathrm{ER}\left(\widehat{\beta}_{MV}, \widehat{\beta}\right)$ e conclua.

Nota

6) (20%) Seja (X, Y) uma variável aleatória bidimensional contínua com função densidade de probabilidade conjunta dada por

$$f_{X,Y}(x, y) = k, \qquad \text{para } |x| + |y| \leq k,$$

sendo $k > 0$.

a) **(03%)** Quais condições a função $f_{X,Y}(x, y)$ deve satisfazer para ser uma função densidade de probabilidade conjunta?

b) **(07%)** Determine o valor de k para que $f_{X,Y}(x, y)$ seja uma função densidade de probabilidade conjunta.

Nota

c) **(04%)** Sabendo que as variáveis X e Y não são independentes, se efetuássemos o cálculo da covariância poderíamos obter $\text{cov}\left[X, Y\right]$ igual a zero? justifique.

d) **(06%)** Determine $P\left[Y > \frac{1}{10}\right]$.

Nota

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 11/Junho/2019

Número de inscrição:

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São CINCO (**5**) questões em páginas numeradas de 1 a 7, total de 100 pontos.
- É permitida a utilização de calculadora científica.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE !!**

Revisão Total

1) (20%) Suponha que foi realizado um experimento no delineamento inteiramente casualizado (DIC) com duas repetições para cada um dos três níveis de um fator em estudo. Supondo o modelo estatístico (1) e o Modelo Linear de Gauss Markov Normal associado (2), responda as questões apresentadas.

$$y_{ik} = \mu + \alpha_i + e_{ik} \text{ para } i = 1, 2, 3 \text{ e } k = 1, 2, \tag{1}$$

em que

•y_{ik}: $k-$ésimo valor observado para o $i-$ésimo fator em estudo;

•μ: média geral;

•α_i: efeito do $i-$ésimo fator em estudo, tal que

$$\alpha_i = \mu_i - \mu \quad \text{e} \quad \sum_{i=1}^{3} \alpha_i = 0;$$

•e_{ik}: efeito do erro experimental associado ao valor observado y_{ik}.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}; \quad \mathbf{e} \sim N(\mathbf{0}; \sigma^2\mathbf{I}) \tag{2}$$

a) (06%) Apresente de forma matricial o valor esperado de $\mathbf{y}$, ou seja, $E(\mathbf{y})$.

Nota

b) (07%) A função $\mathbf{C}\boldsymbol{\beta} = \alpha_1 + \alpha_2$ é estimável? **Justifique a sua resposta.**

c) (07%) Expresse as hipóteses para comparar as médias dos níveis do fator em estudo da ANOVA em termos de uma função estimável $\mathbf{C}\boldsymbol{\beta}$.

Nota

2) (20%) Um procedimento de controle da qualidade utilizado em uma indústria, determina que uma máquina deverá ser tirada de operação e regulada, se ela estiver produzindo mais do 10% de itens defeituosos. Se uma amostra aleatória de 100 itens obtidos dessa máquina apresentar 15 itens defeituosos, qual deverá ser a decisão. Adote $\alpha = 1\%$ como nível de significância e responda com base em um teste de hipóteses. Pede-se:

a) (05%) Hipóteses estatísticas.

b) (05%) Valor calculado.

c) (05%) Conceitue valor-p de um teste de hipóteses e faça um desenho ilustrativo para apenas indicar (ou ilustrar) qual é o valor-p deste teste.

d) (05%) Considere o valor de α adotado e a seguinte informação para tomar uma decisão. Se $Z \sim \text{N}(0, 1)$ então $P(Z > 2,33) = 0,01$. Explique se a máquina deve ou não ser parada para regulagem.

Nota

3) (20%) Suponha que X tenha a seguinte função densidade de probabilidade:

$$f(x) = \lambda e^{-\lambda(x-a)}, \qquad x \geq a.$$

a) (10%) Determine a função geradora de momentos (f.g.m) de X;

b) (10%) Utilizando a f.g.m., determine $E[X]$.

Nota

4) (20%) Um pesquisador tem interesse em verificar se há diferenças entre as médias populacionais do fosfato na água (Y) de somente três e específicas praias (a_1, a_2 e a_3). Na pesquisa, ele escolheu aleatoriamente quatro lagoas ($b_{(i)1}$, $b_{(i)2}$, $b_{(i)3}$ e $b_{(i)4}$) de cada praia e, depois, sorteou cinco pontos amostrais de cada lagoa para realizar as avaliações. De acordo com a análise de variância e considerando-se as suas pressuposições satisfeitas, qual(is) é(são) a(s) praia(s) que mais acumula(m) fosfato na água? Se necessário, aplique o teste de Tukey ($\alpha = 0,05$). Por quê?

a_1				a_2				a_3			
$b_{(1)1}$	$b_{(1)2}$	$b_{(1)3}$	$b_{(1)4}$	$b_{(2)1}$	$b_{(2)2}$	$b_{(2)3}$	$b_{(2)4}$	$b_{(3)1}$	$b_{(3)2}$	$b_{(3)3}$	$b_{(3)4}$
6,4	4,8	4,9	5,3	6,8	5,6	5,0	7,8	9,2	9,5	6,8	7,5
6,9	5,8	5,5	6,2	5,7	5,5	5,2	6,5	7,9	8,8	7,8	6,9
5,9	5,5	5,2	4,8	5,2	5,9	4,8	6,8	7,3	9,2	7,0	7,7
6,1	5,8	5,6	5,0	5,3	6,6	5,8	7,0	7,6	9,3	7,3	7,6
6,7	5,6	4,8	5,5	5,5	5,9	4,2	6,9	8,0	9,2	7,6	7,8

FV	GL	SQ	QM	f_{cal}	f_{tab}
A		678,630			
B/A		255,335			
Resíduo		112,720			
Total		1046,685			

gl_{num}	gl_{den}	f_{tab}
2	9	4,26
2	48	3,19
9	48	2,08

nº médias	gl	q_{tab}	Δ
3	9	3,95	$q_{tab}\sqrt{\frac{QM(?)}{(?)}}$
3	48	3,42	

Nota

5) (20%) Sejam X e Y variáveis aleatórias com função densidade de probabilidade conjunta $f(x,y) = cxyI_{(0,4)}(x)\,I_{(1,5)}(y)$ em que $I_{(a,b)}(x) = \begin{cases} 1, & \text{se } x \in (a,b) \\ 0, & \text{caso contrário} \end{cases}$,

a) (05%) Quais condições uma função $f(x,y)$ deve satisfazer para ser uma função densidade de probabilidade conjunta?

b) (07%) Determine o valor de c de tal forma que $f(x,y)$ seja uma função densidade de probabilidade conjunta.

c) (08%) Determine a função densidade de probabilidade de $U = X + 2Y$.

Nota

Universidade Federal de Viçosa
Departamento de Estatística

Prova seletiva para o Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Nível Doutorado - 11/Novembro/2019

Número de inscrição: | | | | | | | | | | |

- Interpretar corretamente as questões é parte da avaliação.
- Indique todos os cálculos organizadamente.
- São SEIS (**6**) questões em páginas numeradas de 1 a 11, total de 100 pontos.
- Na página 11 temos tabela.
- É permitida a utilização de calculadora científica.
- O tempo máximo para realização desta prova é de 4 horas.
- **BOA SORTE!!**

Revisão Total

1) (17%) Um físico fez 25 medidas independentes da gravidade específica de um corpo. Ele sabe que a limitação do equipamento é tal que o desvio-padrão de cada medida é de σ unidades.

a) (8%) Usando a desigualdade de Chebychev, encontre um limite inferior para a probabilidade de que a média das suas medidas difira da média da gravidade específica do corpo por menos de $\sigma/4$ unidades.

b) (9%) Usando o Teorema Central do Limite, encontre um valor aproximado para a probabilidade pedida em **a)**.

Nota

2) (17%) Seja a variável aleatória X o número de amostras de água que necessitam ser analisadas, de modo a detectar a presença de contaminação. Considere que a probabilidade de encontrar uma amostra contaminada é de 0,01 e que as amostras de água sejam independentes.

a) (5%) Apresente o espaço amostral do experimento em questão.

b) (6%) Determine a função de probabilidade da variável aleatória X.

c) (6%) Qual o tamanho ideal de amostras a serem coletadas, se desejamos uma probabilidade de, pelo menos 97% de que a contaminação não ocorrerá posteriormente à coleta?

Nota

3) (17%) Sejam Y_1, Y_2, Y_3, Y_4 variáveis aleatórias independentes e identicamente distribuídas com distribuição de Bernoulli, isto é, $Y_i \sim$Bernoulli(p), para $i = 1, 2, 3, 4$, com função de probabilidade dada por $P\left(Y_i = y_i | p\right) = f\left(y_i | p\right) = p^{y_i}(1-p)^{1-y_i}$, sendo $y_i = 0$ se o evento não ocorre (fracasso) ou $y_i = 1$ se o evento ocorre (sucesso) e $p \in \Theta$. Pede-se:

a) (5%) Seja $\boldsymbol{y} = \{y_1, y_2, y_3, y_4\}$ uma amostra aleatória observada. Obtenha a função de verossimilhança, geralmente denotada como $L\left(p | \boldsymbol{y}\right)$.

b) (6%) Seja $\hat{p}_{MV}$ o estimador de máxima verossimilhança de p. Explique como se define tal estimador.

Nota

c) (6%) Admita que o espaço paramétrico (Θ) seja dado por $\Theta = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$. Com base nas informações apresentadas na tabela a seguir, obtenha as estimativas de verossimilhança para p. Justifique sua resposta.

$\sum_{i=1}^{4} y_i$	$P(\boldsymbol{Y} = \boldsymbol{y}\|p)$		
	$p = \frac{1}{4}$	$p = \frac{1}{2}$	$p = \frac{3}{4}$
0	$\frac{81}{256}$	$\frac{16}{256}$	$\frac{1}{256}$
1	$\frac{27}{256}$	$\frac{16}{256}$	$\frac{3}{256}$
2	$\frac{9}{256}$	$\frac{16}{256}$	$\frac{9}{256}$
3	$\frac{3}{256}$	$\frac{16}{256}$	$\frac{27}{256}$
4	$\frac{1}{256}$	$\frac{16}{256}$	$\frac{81}{256}$

Nota

4) (17%) Supondo que $\boldsymbol{\omega}'\boldsymbol{A}\boldsymbol{\omega}$ seja uma forma quadrática de tal forma que:

i) $\boldsymbol{\omega}_{n\times 1}$ é um vetor aleatório com

- $\boldsymbol{\omega}_{n\times 1} \sim (\boldsymbol{\mu}; \boldsymbol{\Sigma})$
- $\boldsymbol{\Sigma}$ é uma matriz não singular.

ii) $\boldsymbol{A}_{n\times n}$ seja uma matriz de constantes sendo

- $\text{rank}(\boldsymbol{A}) = k$
- $\boldsymbol{A}\boldsymbol{\Sigma}$ é idempotente.

OBSERVAÇÃO: Para responder os itens abaixo, não é necessário demonstrar como a expressão é obtida. Apenas apresente a expressão matemática que permite fornecer o que é perguntado em cada item desta questão.

a) (9%) Obtenha o valor esperado da forma quadrática, ou seja, $E\left(\boldsymbol{\omega}'\boldsymbol{A}\boldsymbol{\omega}\right)$.

b) (8%) Qual é a distribuição de probabilidades da forma quadrática?

Nota

5) (16%) Um engenheiro agrônomo quer testar a seguinte hipótese: plantas de milho mais irrigadas, em menor densidade e com mais fertilizante produzem mais. Para isso, ele realizou um experimento fatorial 2^3, cujas combinações entre as lâminas de irrigação (A) (a_1 = baixa e a_2 = alta), as densidades de plantio (B) (b_1 = baixa e b_2 = alta) e as doses do fertilizante (C) (c_1 = baixa e c_2 = alta) foram casualizadas sob o DBC com duas repetições (d_1 e d_2). No total, foram obtidas 16 unidades experimentais. Considerando-se satisfeitas as pressuposições da análise de variância, o agrônomo conseguiu provar totalmente a sua hipótese $(\alpha = 0,05)$? Se sim, por quê? Se não, qual deve ser a recomendação atual para aumentar a produção (Y)?

$SQA = \frac{1}{ghr}\sum_{i=1}^{t} A_i^2 - \frac{G^2}{tghr}$	$SQA \times B = SQA,B - SQA - SQB$	$SQA,B = \frac{1}{hr}\sum_{i=1}^{t}\sum_{j=1}^{g} AB_{ij}^2 - \frac{G^2}{tghr}$
$SQB = \frac{1}{thr}\sum_{j=1}^{g} B_j^2 - \frac{G^2}{tghr}$	$SQA \times C = SQA,C - SQA - SQC$	$SQA,C = \frac{1}{gr}\sum_{i=1}^{t}\sum_{k=1}^{h} AC_{ik}^2 - \frac{G^2}{tghr}$
$SQC = \frac{1}{tgr}\sum_{k=1}^{h} C_k^2 - \frac{G^2}{tghr}$	$SQB \times C = SQB,C - SQB - SQC$	$SQB,C = \frac{1}{tr}\sum_{j=1}^{g}\sum_{k=1}^{h} BC_{jk}^2 - \frac{G^2}{tghr}$

	Bloco d_1			
	c_1		c_2	
	b_1	b_2	b_1	b_2
a_1	25,4	25,6	26,1	25,9
a_2	14,6	15,5	14,3	15,6
	Bloco d_2			
	c_1		c_2	
	b_1	b_2	b_1	b_2
a_1	32,9	34,7	35,2	35,8
a_2	25,2	24,7	25,5	24,8

FV	GL	SQ	QM	F_{cal}	F_{tab}
Bloco		359,1025			5,59
A					5,59
B					5,59
C					5,59
$A \times B$					5,59
$A \times C$					5,59
$B \times C$					5,59
$A \times B \times C$		0,2025			5,59
Resíduo		4,3775			
Total		781,1975			

Observação: a próxima folha encontra-se em branco caso seja necessário efetuar cálculos nesta questão.

Nota

Nota

6) (16%) A tabela a seguir informa os tempos, em dias, que normalmente são estabelecidos para a conclusão de uma auditoria interna de final de ano. Adicionalmente, nessa tabela, podem ser verificados o total de auditorias que foram concluídas com os respectivos tempos para uma amostra de 119 clientes da Sanderson e Clifford, uma pequena empresa de contabilidade pública.

Tempo	9	5	21	42	28	2	17
Total de auditorias	29	9	35	1	12	0	33

Pede-se:

a) (4%) Os tempos correspondentes ao primeiro e o terceiro quartis da amostra.

b) (4%) Explique o que significa o 92° percentil dessa amostra, isto é, $P_{92} = 28$.

Nota

c) (4%) A amplitude total e amplitude interquartil dos tempos.

d) (4%) Esboce o *box plot* dos tempos. Há algum indício inicial de que essa amostra possua *outliers*? Justifique sua resposta.

Nota

Tabela 1: Probabilidade na distribuição normal padrão de zero a um valor positivo z_c, ou seja, $P[0 \leq Z \leq z_c] = p$.

z_c	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Nota