UNIVERSIDADE FEDERAL DE LAVRAS
DEPARTAMENTO DE ESTATÍSTICA
Programa de Pós-Graduação em Estatística e Experimentação Agropecuária

Prova do Processo Seletivo para Doutorado 2018-1

Nº de inscrição do candidato: _______________

- Utilizar APENAS o número de inscrição para identificar a sua prova;
- A interpretação das questões é parte da avaliação;
- Indique todos os cálculos organizadamente;
- São DEZ (10) questões, valendo UM (1) ponto cada, totalizando 10 pontos;
- O tempo máximo para a realização desta prova é de 4 horas;
- A prova pode ser feita a lápis ou caneta (azul ou preta) e é permitido utilizar somente a calculadora.
- BOA SORTE !!

1) Mostre que a função $f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, se\ (x,y) \neq (0,0) \\ 0, \quad se\ (x,y) = (0,0) \end{cases}$ é descontinua na origem.

Justifique sua resposta.

Resolução **(1,0)**

Considere a função passando pela origem

a) $y = x$

$$f(x,y) = \frac{x^2}{2x^2} = \frac{1}{2}\,, \qquad \lim_{x\to 0}\frac{1}{2} = \frac{1}{2}\,;$$

b) $y = x^2$

$$f(x,x^2) = \frac{x^3}{x^2+x^4} = \frac{x^3}{x^2(1+x^2)} = \frac{x}{1+x^2}\,, \qquad \lim_{x\to 0}\frac{x}{1+x^2} = 0\,.$$

Como não existe o limite em (0,0) $\Rightarrow$ a função é descontínua na origem.

2) Determine os pontos críticos da função $f(x,y) = x^3 - 3xy + y^3$ e diga se máximo, mínimo ou sela.

Resolução

$$\frac{\partial f}{\partial x} = 3x^2 - 3y = 0 \quad \Rightarrow \quad y = x^2$$

$$\frac{\partial f}{\partial x} = -3x + 3y^2 = 0 \quad \Rightarrow \quad -3x + 3x^4 = 0$$

$$x(-1+x^3) = 0 \quad \Rightarrow \quad \begin{cases} x = 0 \\ x^3 = 1 \quad \Rightarrow \quad x = 1 \end{cases}$$

para $x = 0$, ponto $(0,0)$
para $x = 1$, ponto $(1,1)$ **(0,5)**

$$\frac{\partial^2 f}{\partial x^2} = 6x; \quad \frac{\partial^2 f}{\partial y^2} = 6y; \quad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = -3$$

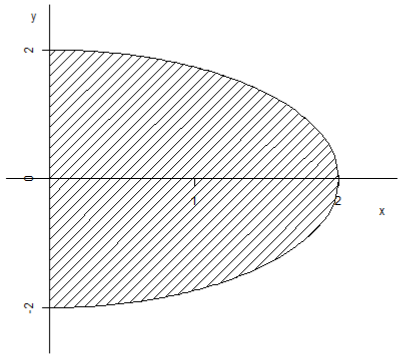
$$H(x_0,y_0) = \begin{vmatrix} 6x & -3 \\ -3 & 6y \end{vmatrix} = 36xy - 9 \qquad \textbf{(0,5)}$$

Para o ponto $(0,0)$ $\Rightarrow$ $H(0,0) = -9 < 0$ $\Rightarrow$ ponto de sela;

Para o ponto $(1,1)$ $\Rightarrow$ $H(1,1) = 25 > 0$ e $\frac{\partial^2 f}{\partial x^2} > 0$ $\Rightarrow$ ponto de mínimo.

3) Calcule a integral $\int_0^2 \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \frac{x+y}{x^2+y^2}\,dy\,dx$.

Resolução



$$x = r\cos\theta \qquad dxdy = r\,drd\theta; \quad -\frac{\pi}{2} < \theta < \frac{\pi}{2}\,, \quad 0 < r < 2 \qquad \mathbf{(0{,}4)}$$
$$y = r\,sen\,\theta$$

$$\int_0^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{r(\cos\,\theta + \text{sen}\,\,\theta)}{r^2} r\,d\theta dr = \int_0^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta + \text{sen}\,\theta) d\theta dr = r\Big|_0^2 \;\; \text{sen}\,\theta - \cos\theta\,\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} =$$

$$2\left\{\text{sen}\frac{\pi}{2} - \cos\frac{\pi}{2} - \left[\text{sen}\left(-\frac{\pi}{2}\right) - \cos\left(-\frac{\pi}{2}\right)\right]\right\} = 2(1+1) = 4 \quad \mathbf{(0{,}6)}$$

4) Se a variável aleatória X tem densidade Normal$\left(\mu,\sigma^2\right)$, sua função geradora de momentos é dada por

$$m_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

Sabendo que Z tem distribuição Normal Padrão e que $Y = e^Z$, determinar a esperança $E[Y]$.

Resolução

Utilizando a função geradora de momentos

$$m_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\} \qquad = E\left[\exp\{tX\}\right]$$

$$\Rightarrow \qquad E\left[\exp\{X\}\right] = E\left[\exp\{tX\}\right]\Big|_{t=1}$$

$$= \exp\left\{\mu + \frac{1}{2}\sigma^2\right\}$$

$$\Rightarrow \qquad m_Z(t) = \exp\left\{\frac{1}{2}t^2\right\} \qquad \mathbf{(0{,}3)}$$

$$\Rightarrow \qquad E[Y] = E\left[\exp\{Z\}\right] \qquad = \exp\left\{0 + \frac{1}{2}1^2\right\} = \exp\left\{\frac{1}{2}\right\} = \sqrt{e} \qquad \mathbf{(+\ 0{,}7)}$$

<u>Outra resolução</u>

Calculando a densidade de Y e a sua esperança:

$$Y = e^Z \qquad \Rightarrow \qquad Y > 0,\ z = \ln(y) \quad \text{e} \quad \frac{dz}{dy} = \frac{1}{y}$$

$$\Rightarrow \qquad f_Z(z) \qquad = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\}$$

$$\Rightarrow \qquad f_Y(y) = f_Z(z)\big|_{z=\ln(y)} \left|\frac{dz}{dy}\right| = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\ln^2(y)\right\}\left|\frac{1}{y}\right|$$

$$= \frac{1}{y\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\ln^2(y)\right\} \qquad \mathbf{(0{,}4)}$$

$$\Rightarrow \qquad E[Y] = \int_0^\infty y\, f_Y(y)\, dy = \int_0^\infty y \frac{1}{y\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\ln^2(y)\right\} dy$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\ln^2(y)\right\} dy \text{ ,fazendo } \begin{cases} w = \ln(y) \Leftrightarrow y = e^w \\ dy = e^w\, dw \end{cases}$$

$$E[Y] = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} w^2\right\} e^w\, dw, \text{ completando quadrados}$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(w^2 - 2w + 1\right)\right\} e^{1/2}\, dw$$

$$= e^{1/2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(w-1)^2\right\} dw$$

$$= e^{1/2} = \sqrt{e} \qquad \text{, pois } \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(w-1)^2\right\} dw = 1 \ \text{(é}$$

a integral da uma $Normal(1,1)$). **(+0,6)**

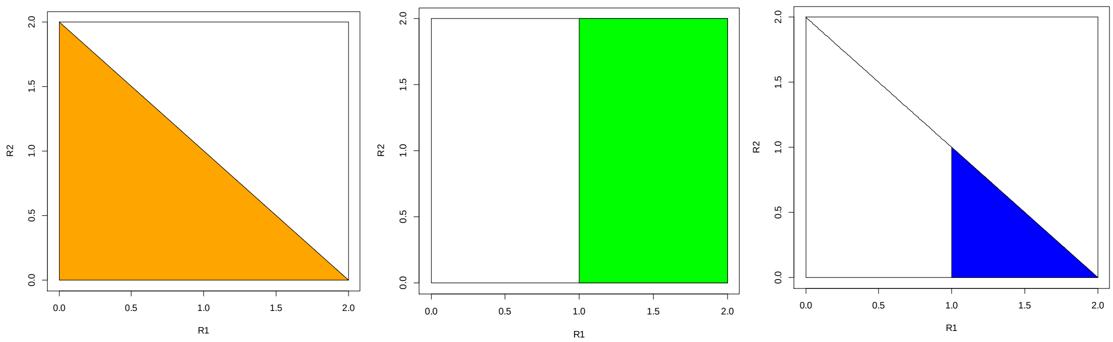
5) Suponha R_1 e R_2 variáveis aleatórias independentes, ambas com densidade uniforme no intervalo $(0,2)$. Determinar a probabilidade de $R_1 > 1$, dado que $R_1 + R_2 < 2$.

Resolução

Usando probabilidade condicional

$$P\left[R_1 \geq 1 \middle| R_1 + R_2 \leq 2\right] = \frac{P\left[\left(R_1 \geq 1\right) \cap \left(R_1 + R_2 \leq 2\right)\right]}{P\left[R_1 + R_2 \leq 2\right]} = \frac{\frac{1}{8}\,\frac{1}{4}}{\frac{1}{2}\,\frac{1}{4}} = \frac{1}{8}\,\frac{2}{1} = \frac{1}{4} \quad \textbf{(1)}$$

ou usando a geometria:



$R_1 + R_2 \leq 2$ $R_1 \geq 1$ $(R_1 \geq 1) \cap (R_1 + R_2) \leq 2$

Caso o candidato tenha utilizado integral dupla e errado a probabilidade condicional. **(0,4)**

6) a) Se foi observada uma amostra $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ de uma população Normal$\left(\mu, \sigma^2\right)$, demonstre que as estimativas de máxima verossimilhança de μ e σ^2 são, respectivamente:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

(**<u>Observação:</u>** não é necessário mostrar que o ponto crítico da função de verossimilhança é um ponto de máximo e $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$.

b) O conceito de ***estatística suficiente*** é importante em estatística no sentido de se poder concentrar a informação sobre o parâmetro. Formalmente tem-se: se $X_1, X_2, ..., X_n$ é uma amostra de uma densidade $f(x;\theta)$, em que θ é um vetor de parâmetros, as estatísticas $S_1, S_2, ..., S_r$ são conjuntamente suficientes se, e somente se, a distribuição conjunta de $X_1, X_2, ..., X_n$, dadas $S_1 = s_1, S_2 = s_2, ..., S_r = s_r$, não depende de θ.

A aplicação desta definição para se provar que um conjunto de estatísticas é conjuntamente suficiente é, em geral, difícil. Um critério algébrico mais fácil de ser aplicado é dado pelo Teorema da Fatoração, dado por:

Seja $X_1, ..., X_n$ uma amostra de uma densidade $f(x;\theta)$ em que θ é um vetor de parâmetros. As estatísticas $S_1, ..., S_r$ são conjuntamente suficientes se, e somente se, a densidade conjunta pode ser fatorada como:

$$\begin{aligned} f_{X_1,...,X_n}(x_1,...,x_n;\theta) &= g(s_1(x_1,...,x_n),...,s_r(x_1,...,x_n);\theta)h(x_1,...,x_n) \\ &= g(s_1,...,s_r;\theta)h(x_1,...,x_n) \ , \end{aligned}$$

em que $h(x_1,...,x_n)$ é não negativa e não depende do parâmetro θ e a função $g(s_1,...,s_r;\theta)$ é não negativa e depende de $x_1,...,x_n$ somente através das funções $s_1,...,s_r$.

Utilizando o Teorema da Fatoração, prove que as estatísticas $S_1 = \sum_{i=1}^{n} X_i$ e $S_2 = \sum_{i=1}^{n} X_i^2$ são conjuntamente suficientes para a $normal(\mu, \sigma^2)$.

Resolução

a) $$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \qquad \textbf{(0,2)}$$

$$l(\mu,\sigma^2) = Ln\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\right]$$

$$= Ln\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}l(\mu,\sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{x}_n \quad \textbf{(0,2)}$$

$$\frac{\partial}{\partial\sigma^2}l(\mu,\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 \qquad \textbf{(0,2)}$$

b) $$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\mu,\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2)\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right)\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(s_2-2\mu s_1+n\mu^2)\right\}$$

Portanto: $$g(s_1,s_2;\mu,\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(s_2-2\mu s_1+n\mu^2)\right\} \quad \text{e}$$

$$h(x_1,\ldots,x_n) = 1 \quad \textbf{(0,4)}$$

7) Sejam $X_1, X_2, ..., X_n$ uma amostra de uma população Normal$\left(\mu, \sigma^2\right)$, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. São grandezas pivotais:

(I) $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ (supondo que σ^2 seja um valor conhecido)

(II) $\frac{\bar{X}_n - \mu}{S/\sqrt{n}}$

(III) $\frac{(n-1)S^2}{\sigma^2}$

(a) Qual é a distribuição de cada uma das quantidades pivotais (I), (II) e (III) , apresentadas anteriormente. Basta citar o nome da distribuição.

(b) Define-se x_α como o α-quantil de uma densidade $f_X(x)$ se $P[X \geq x_\alpha] = \alpha$. Obtenha, em termos dos quantis apropriados, intervalos de confiança a 95% para os parâmetros presentes nas grandezas pivotais (I), (II) e (III) , apresentadas anteriormente.

<u>Resolução</u>

a)

(I) $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ Normal padrão

(II) $\frac{\bar{X}_n - \mu}{S/\sqrt{n}}$ t de Student com $n-1$ graus de liberdade

(III) $\frac{(n-1)S^2}{\sigma^2}$ qui-quadrado com $n-1$ graus de liberdade

1 resposta certa (0,2); 2 certas (0,3); 3 certas (0,4); a falta dos parâmetros (-0,1)

b)

(I) Fazendo $\dfrac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}=Z$

$$\left(-z_{0,975}\le \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\le z_{0,975}\right)\Leftrightarrow\left(\bar{X}_n-\frac{\sigma}{\sqrt{n}}z_{0,975}\le\mu\le\bar{X}_n+\frac{\sigma}{\sqrt{n}}z_{0,975}\right)$$ **(0,2)**

(II) Fazendo $\dfrac{\bar{X}_n-\mu}{S/\sqrt{n}}=T_{n-1}$

$$\left(-t_{n-1;0,975}\le \frac{\bar{X}_n-\mu}{S/\sqrt{n}}\le t_{n-1;0,975}\right)\Leftrightarrow\left(\bar{X}_n-\frac{S}{\sqrt{n}}t_{n-1;0,975}\le\mu\le\bar{X}_n+\frac{\sigma}{\sqrt{n}}t_{n-1;0,975}\right)$$ **(0,2)**

(III) Fazendo $\dfrac{(n-1)S^2}{\sigma^2}=\chi^2_{n-1}$

$$\left(\chi^2_{n-1;0,025}\le\frac{(n-1)S^2}{\sigma^2}\le\chi^2_{n-1;0,975}\right)\Leftrightarrow\left(\frac{(n-1)S^2}{\chi^2_{n-1;0,975}}\le\sigma^2\le\frac{(n-1)S^2}{\chi^2_{n-1;0,025}}\right)$$ **(0,2)**

Obs.: Os intervalos de confiança não são determinados unicamente. Outros intervalos, se corretamente construídos, serão considerados. A questão também é considera correta caso o candidato tenha utilizado a definição usual de quantil:

O q-ésimo quantil de uma de uma densidade $f_X(x)$, com notação ξ_q, como o menor número ξ que satisfaz $F_X(\xi)\ge q$.

8) Um estudante de pós-graduação executou um experimento, utilizando o delineamento em quadrado latino, com o objetivo de comparar a produtividade de 5 cultivares de soja (A, B, C, D e E). Ele estava precisando terminar as análises estatísticas, para apresentar um trabalho em um congresso da área e, como ele ainda não cursou a disciplina de Estatística Experimental e soube que você entende do assunto, resolveu te procurar. Ele lhe mostrou os seguintes resultados, fornecidos pelo orientador dele, que já havia iniciado as análises:

- Totais dos tratamentos, em Kg/parcela:

$$T_A = 46{,}0; \quad T_B = 58{,}8; \quad T_C = 57{,}3; \quad T_D = 61{,}7; \quad T_E = 74{,}5;$$

- $CV = 14{,}85\%$;
- Não houve parcela perdida no experimento, ou seja, o experimento é balanceado;
- Teste F da análise de variância para cultivares foi significativo.

Suponha que o interesse do estudante seja aplicar o teste de Student-Newman-Keuls (SNK), pois seu orientador havia pedido. Somente com as informações apresentadas é possível aplicar esse teste? Se sua resposta for *SIM*, aplique o teste, considerando $\alpha = 5\%$, concluindo corretamente e recomende a(s) cultivar(es) mais produtiva(s). Caso sua resposta seja *NÃO*, justifique sua resposta e indique qual ou quais informações estão faltando para que você possa aplicar o teste de SNK.

Dados: $DMS_k = q_{(5\%;\, k;\, \text{GLErro})}\sqrt{\frac{QMErro}{r}}$; $CV = \frac{\sqrt{QMErro}}{\hat{m}} \cdot 100$;

$q_{(5\%;\, 2,\, 12)} = 3{,}08$; $q_{(5\%;\, 3,\, 12)} = 3{,}77$; $q_{(5\%;\, 4,\, 12)} = 4{,}20$; $q_{(5\%;\, 5,\, 12)} = 4{,}51$;

$q_{(5\%;\, 2,\, 20)} = 2{,}95$; $q_{(5\%;\, 3,\, 20)} = 3{,}58$; $q_{(5\%;\, 4,\, 20)} = 3{,}96$; $q_{(5\%;\, 5,\, 20)} = 4{,}23$.

Resolução: Sim, é possível.

$$QMErro = \left(\frac{CV \cdot \hat{m}}{100}\right)^2$$

$$\hat{m} = \frac{46{,}0 + 58{,}8 + 57{,}3 + 61{,}7 + 74{,}5}{25} = 11{,}932$$

$QMErro = \left(\frac{14{,}85 \cdot 11{,}932}{100}\right)^2 = 3{,}14$ **(0,2)**

$$\begin{cases} H_0: m_i = m_{i'} \\ H_a: m_i \neq m_{i'} \end{cases} \quad \text{para } i \neq i'$$

<u>Médias ordenadas e letras</u>: **(0,2)**

$$\hat{m}_E = \frac{74{,}5}{5} = 14{,}90 \text{ a}$$

$$\hat{m}_D = \frac{61{,}7}{5} = 12{,}34 \quad \text{b}$$

$$\hat{m}_B = \frac{58{,}8}{5} = 11{,}76 \quad \text{b}$$

$$\hat{m}_C = \frac{57{,}3}{5} = 11{,}46 \quad \text{b}$$

$$\hat{m}_A = \frac{46{,}0}{5} = 9{,}20 \quad \text{b}$$

<u>Cálculo das DMS</u>: **(0,3)**

$$DMS_5 = q_{(5\%;\,5;\,12)}\sqrt{\frac{QMErro}{r}} = 4{,}51\sqrt{\frac{3{,}14}{5}} = 3{,}57$$

$$DMS_4 = q_{(5\%;\,4;\,12)}\sqrt{\frac{QMErro}{r}} = 4{,}20\sqrt{\frac{3{,}14}{5}} = 3{,}33$$

$$DMS_3 = q_{(5\%;\,3;\,12)}\sqrt{\frac{QMErro}{r}} = 3{,}77\sqrt{\frac{3{,}14}{5}} = 2{,}99$$

$$DMS_2 = q_{(5\%;\,2;\,12)}\sqrt{\frac{QMErro}{r}} = 3{,}08\sqrt{\frac{3{,}14}{5}} = 2{,}44$$

<u>Conclusão</u>: Médias seguidas pela mesma letra não diferem entre si ao nível de 5% de probabilidade pelo teste de SNK. **(0,2)**

<u>Resposta:</u> A mais produtiva foi a cultivar *E*. **(0,1)**

9) Um experimento de competição de cultivares de mandioca foi realizado utilizando-se o delineamento inteiramente casualizado com cinco repetições. As cultivares comparadas foram IAC-5, IAC-7, IAC-11, Iracema e Mantiqueira. As produtividades obtidas, em t/ha, foram:

	Repetições					
Cultivares	1	2	3	4	5	Totais
IAC-5	38,9	25,4	20,3	25,7	29,3	139,6
IAC-7	20,9	26,2	32,3	28,3	28,7	136,4
IAC-11	28,1	27,0	25,8	26,9	22,3	130,1
Iracema	38,7	43,2	41,7	39,0	40,3	202,9
Mantiqueira	47,8	47,8	44,7	50,5	56,4	247,2

(a) Apresente o modelo estatístico referente a esse experimento, descrevendo adequadamente cada termo do modelo.

(b) Quais são as pressuposições que o modelo exige para que os resultados da análise de variância e dos testes estatísticos sejam válidos?

(c) Efetuar a Análise de Variância e concluir corretamente.

DADOS:

$$\boldsymbol{SQTotal = 2509,46}$$

$F_{(5\%;\,4,\,16)} = 3{,}01$; $F_{(5\%;\,5,\,16)} = 2{,}85$; $F_{(5\%;\,4,\,20)} = 2{,}87$; $F_{(5\%;\,5,\,20)} = 2{,}71$;

$F_{(1\%;\,4,\,16)} = 4{,}77$; $F_{(1\%;\,5,\,16)} = 4{,}44$; $F_{(1\%;\,4,\,20)} = 4{,}43$; $F_{(1\%;\,5,\,20)} = 4{,}10$;

(d) Formule um contraste para comparar a produtividade média do grupo de cultivares IAC com as demais (*IAC vs demais*) e aplique o teste de Scheffé a este contraste, concluindo corretamente.

(e) Apresente o modelo linear de Gauss-Markov na forma matricial, descrevendo adequadamente cada vetor e matriz do modelo, desse experimento. Apresente também os vetores e a matriz desse modelo.

(f) Apresente as matrizes e os vetores do sistema de equações normais (*SEN*) $\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{X'y}$, referente a esse experimento.

Resolução

(a) $y_{ij} = \mu + \alpha_i + e_{ij}$,

em que

y_{ij} é a produtividade observada na repetição j da cultivar i, com $i = 1, \dots, 5$ e $j = 1, \dots, 5$;

μ é uma constante associada a cada observação;

α_i é o efeito da i-ésima cultivar;

e_{ij} é o erro associado a cada observação, sendo $e_{ij} \sim N(0, \sigma^2)$. **(0,1)**

(b) Aditividade dos termos do modelo, homogeneidade de variâncias, independência e normalidade dos erros. **(0,1)**

(c)

Tabela 1 – Análise de variância da produtividade **(0,25)**

FV	GL	SQ	QM	F
Cultivar	4	2135,94	533,98	28,59
Erro	20	373,52	18,68	
Total	24	2509,46		

Teste F:

(0,05) $\begin{cases} H_0: m_1 = m_2 = m_3 = m_4 = m_5 = m \\ H_a: \text{Existe pelo menos uma diferença entre as médias das cultivares} \end{cases}$

$F_{(5\%;\,4,\,20)} = 2{,}87$; $F_{(1\%;\,4,\,20)} = 4{,}43$

$F_{calc} > F_{1\%}$ $\Rightarrow$ Rejeita-se H_0 ao nível de 1% de probabilidade pelo teste F. Portanto, existe diferença significativa na produtividade das cultivares de mandioca. **(0,1)**

(d) $Y = 2m_1 + 2m_2 + 2m_3 - 3m_4 - 3m_5$ **(0,1)**

$\begin{cases} H_0: Y = 0 \\ H_a: Y \neq 0 \end{cases}$

$\hat{Y} = 2 \cdot 27{,}92 + 2 \cdot 27{,}28 + 2 \cdot 26{,}02 - 3 \cdot 40{,}58 - 3 \cdot 49{,}5 = -107{,}80$ **(0,02)**

$DMS = \sqrt{(I-1) \cdot F_\alpha \cdot \frac{QMErro}{r} \cdot \sum_i c_i^2} = \sqrt{(5-1) \cdot 4{,}43 \cdot \frac{18{,}68}{5} \cdot 30} = 44{,}57$ **(0,03)**

$|\hat{Y}| > DMS$ $\Rightarrow$ Rejeita-se H_0 ao nível de 1% de probabilidade pelo teste de Scheffé. O contraste Y é estatisticamente diferente de zero, ou seja, as cultivares Iracema e Mantiqueira apresentam, em média, maior produtividade que as cultivares IAC. **(0,05)**

(e) $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$,

em que

$\boldsymbol{Y}$ é o vetor de observações, de ordem $25x1$;

$\boldsymbol{X}$ é a matriz de incidência, de ordem $25x6$;

$\boldsymbol{\beta}$ é o vetor de efeitos do modelo, de ordem $6x1$;

$\boldsymbol{\varepsilon}$ é o vetor de erros, com $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2)$. **(0,05)**

$$\boldsymbol{Y} = \begin{bmatrix} 38,9 \\ 25,4 \\ 20,3 \\ 25,7 \\ 29,3 \\ 20,9 \\ 26,2 \\ 32,3 \\ 28,3 \\ 28,7 \\ 28,1 \\ 27,0 \\ 25,8 \\ 26,9 \\ 22,3 \\ 38,7 \\ 43,2 \\ 41,7 \\ 39,0 \\ 40,3 \\ 47,8 \\ 47,8 \\ 44,7 \\ 50,5 \\ 56,4 \end{bmatrix}; \ \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \ \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}; \quad \boldsymbol{\varepsilon} = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \\ e_{25} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{34} \\ e_{35} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{44} \\ e_{45} \\ e_{51} \\ e_{52} \\ e_{53} \\ e_{54} \\ e_{55} \end{bmatrix}$$

Matrizes e vetores **(0,05)**

(f)

$$X'X = \begin{bmatrix} 25 & 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 0 & 0 & 0 & 0 \\ 5 & 0 & 5 & 0 & 0 & 0 \\ 5 & 0 & 0 & 5 & 0 & 0 \\ 5 & 0 & 0 & 0 & 5 & 0 \\ 5 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}; \ \textbf{(0,05)}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}; \quad X'Y = \begin{bmatrix} 856,2 \\ 139,6 \\ 136,4 \\ 130,1 \\ 202,9 \\ 247,2 \end{bmatrix}; \ \textbf{(0,05)}$$

10) Um experimento foi instalado em parcela subdividida com o objetivo de comparar a produtividade de cinco variedades de cana-de-açúcar (A, B, C, D, E) cultivadas em dois espaçamentos diferentes (1 e 2). O delineamento utilizado foi em blocos casualizados com quatro repetições. Os resultados obtidos para a produtividade, dada em t/ha, forneceram a seguinte análise de variância:

Tabela da Análise de Variância

Fontes de Variação	GL	SQ	QM	F
Espaçamento (E)	1	0,21	0,21	$0,02^{ns}$
Blocos	3	782,18	260,73	$18,11^{*}$
Erro(a)	3	43,20	14,40	
(Parcelas)	(7)	(775,59)		
Variedades (V)	4	11054,60	2763,65	$59,24^{**}$
Interação E x V	4	620,32	155,08	$3,32^{*}$
Erro(b)	24	1119,67	46,65	
Total	39	13620,18		

** significativo a 1%; * significativo a 5%; ^{ns} não significativo.

(a) Apresente o modelo estatístico referente a esse experimento, descrevendo adequadamente cada termo do modelo.

(b) Avalie a precisão do experimento.

(c) Faça o desdobramento da interação entre os fatores estudados, uma vez que a mesma foi significativa. Aplique o teste F e o teste de Tukey, quando for o caso, concluindo corretamente. Indique as variedades e os espaçamentos que garantiram as maiores produtividades. Considere $\alpha = 5\%$.

DADOS: (continua na página seguinte)

Quadro auxiliar (totais)

Espaçamentos	Variedades					Totais
	A	B	C	D	E	
1	409,2	355,2	277,2	244,7	234,0	1520,3
2	361,3	406,6	278,0	241,9	229,6	1517,4
Totais	770,5	761,8	555,2	486,6	463,6	3037,7

DADOS: (continuação)

$QMErro_{combinado} = 1085{,}46$; $GLErro_{combinado} = 27$;

$$DMS = q_{(5\%;\, I;\, \text{GLErro})}\sqrt{\frac{QMErro}{r}}\ ;$$

$q_{(5\%;\, 2,\, 3)} = 4{,}50$; $q_{(5\%;\, 2,\, 24)} = 2{,}92$; $q_{(5\%;\, 2,\, 27)} = 2{,}91$;

$q_{(5\%;\, 5,\, 3)} = 7{,}50$; $q_{(5\%;\, 5,\, 24)} = 4{,}17$; $q_{(5\%;\, 5,\, 27)} = 4{,}13$.

$F_{(5\%;\, 1,\, 3)} = 10{,}13$; $F_{(5\%;\, 1,\, 24)} = 4{,}26$; $F_{(5\%;\, 1,\, 27)} = 4{,}21$;

$F_{(5\%;\, 4,\, 3)} = 9{,}12$; $F_{(5\%;\, 4,\, 24)} = 2{,}78$; $F_{(5\%;\, 4,\, 27)} = 2{,}73$.

Resolução

(a) $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ij} + \gamma_k + \alpha\gamma_{ik} + \varepsilon_{ijk}$,

em que

y_{ijk} é a produtividade na parcela do bloco j que recebeu a k-ésima variedade plantada como o i-ésimo espaçamento, com $i = 1, 2$; $j = 1, \dots, 4$ e $k = 1, \dots, 5$;

μ é uma constante associada a cada observação;

α_i é o efeito do i-ésimo espaçamento;

β_j é o efeito do j-ésimo bloco;

ϵ_{ij} é o erro aleatório da parcela;

γ_k é o efeito da k-ésimo variedade;

$\alpha\gamma_{ik}$ é o efeito da interação entre Espaçamento e Variedade;

ε_{ijk} é o erro aleatório associado a cada observação, sendo que $\varepsilon_{ijk} \sim N(0, \sigma^2)$. **(0,1)**

(b) Avalie a precisão do experimento.

$$CV_{Parcela} = \frac{\sqrt{QMErro(a)}}{\hat{m}} \cdot 100 = \frac{\sqrt{14{,}40}}{75{,}94} \cdot 100 = 5{,}0$$

$$CV_{Subparcela} = \frac{\sqrt{QMErro(b)}}{\hat{m}} \cdot 100 = \frac{\sqrt{46{,}65}}{75{,}94} \cdot 100 = 8{,}99$$

Como ambos os *CV* foram inferiores a 10%, pode-se concluir que o experimento teve alta precisão tanto em nível de parcela quanto de subparcela. **(0,1)**

(c)

1. <u>Desdobramento de Variedades dentro de Espaçamento</u>:

Tabela da Análise de Variância **(0,2)**

FV	GL	SQ	QM	F
Variedades (Esp.1)	4	5705,80	1426,45	30,58*
Variedades (Esp.2)	4	5969,12	1492,28	31,99*
Erro	24	1119,67	46,65	

* Significativo a 5%

1.1. Variedades dentro do espaçamento 1:

$$\begin{cases} H_0: m_A = m_B = m_C = m_D = m_E = m \\ H_a: \text{Existe pelo menos uma diferença entre as médias} \end{cases}$$

$F_{(5\%;\,4,\,24)} = 2{,}78$

$F_{calc} > F_{5\%} \Rightarrow$ Rejeita-se H_0 ao nível de 5% de probabilidade pelo teste F. Portanto, existe diferença significativa entre as médias das variedades quando se utilizou o espaçamento 1.

Teste de Tukey:

$$\begin{cases} H_0: m_i = m_{i'} \\ H_a: m_i \neq m_{i'} \end{cases} \quad \forall\, i \neq i'$$

$$DMS = q_{(5\%;\,5;\,24)}\sqrt{\frac{QMErro}{r}} = 4{,}17\sqrt{\frac{46{,}65}{4}} = 14{,}24$$ **(0,05)**

$$\hat{m}_A = \frac{409{,}2}{4} = 102{,}30 \text{ a}$$ **(0,05)**

$$\hat{m}_B = \frac{355{,}2}{4} = 88{,}80 \text{ a}$$

$$\hat{m}_C = \frac{277{,}2}{4} = 69{,}30 \text{ b}$$

$$\hat{m}_D = \frac{244{,}7}{4} = 61{,}18 \text{ b}$$

$$\hat{m}_E = \frac{234{,}0}{4} = 58{,}50 \text{ b}$$

Médias seguidas pela mesma letra não diferem entre si ao nível de 5% de probabilidade pelo teste de Tukey. **(0,05)**

Com o uso do espaçamento 1, as variedades *A* e *B* foram as mais produtivas. **(0,05)**

1.2. Variedades dentro do espaçamento 2:

$$\begin{cases} H_0: m_A = m_B = m_C = m_D = m_E = m \\ H_a: \text{Existe pelo menos uma diferença entre as médias} \end{cases}$$

$F_{(5\%;\ 4,\ 24)} = 2{,}78$

$F_{calc} > F_{5\%} \Rightarrow$ Rejeita-se H_0 ao nível de 5% de probabilidade pelo teste F. Portanto, existe diferença significativa entre as médias das variedades quando se utilizou o espaçamento 2.

Teste de Tukey:

$$\begin{cases} H_0: m_i = m_{i'} \\ H_a: m_i \neq m_{i'} \end{cases} \quad \forall\, i \neq i'$$

$DMS = q_{(5\%;\ 5;\ 24)}\sqrt{\frac{QMErro}{r}} = 4{,}17\sqrt{\frac{46{,}65}{4}} = 14{,}24$ **(0,05)**

$\hat{m}_B = \frac{406{,}6}{4} = 101{,}65$ a **(0,05)**

$\hat{m}_A = \frac{361{,}3}{4} = 90{,}33$ a

$\hat{m}_C = \frac{278{,}0}{4} = 69{,}50$ b

$\hat{m}_D = \frac{241{,}9}{4} = 60{,}48$ b

$\hat{m}_E = \frac{229{,}6}{4} = 57{,}40$ b

Médias seguidas pela mesma letra não diferem entre si ao nível de 5% de probabilidade pelo teste de Tukey. **(0,05)**

Com o uso do espaçamento 2, as variedades *A* e *B* foram as mais produtivas. **(0,05)**

2. <u>Desdobramento de Espaçamento dentro de Variedades</u>:

Tabela da Análise de Variância **(0,1)**

FV	GL	SQ	QM	F
Espaçamento (A)	1	286,80	286,80	0,26^{ns}
Espaçamento (B)	1	330,25	330,25	0,30^{ns}
Espaçamento (C)	1	0,08	0,08	< 0,01^{ns}
Espaçamento (D)	1	0,98	0,98	< 0,01^{ns}
Espaçamento (E)	1	2,42	2,42	< 0,01^{ns}
Erro combinado	27	-	1085,46	

^{ns} Não significativo

<u>OBS</u>: O valor fornecido do *$QMErro_{combinado}$* está errado. Porém, foi considerado o valor fornecido na correção, o que não influencia na resolução da questão.

Para todos os testes, as hipóteses e o valor do $F_{tabelado}$ são os mesmos e dados por:

$$\begin{cases} H_0: m_1 = m_2 = m \\ H_a: \text{Existe pelo menos uma diferença entre as médias} \end{cases}$$

$F_{(5\%;\, 4,\, 24)} = 4,21$

As conclusões dos testes também são as mesmas, dado que todos os testes F foram não significativos. Assim, $F_{calc} < F_{5\%} \Rightarrow$ Não se rejeita H_0 ao nível de 5% de probabilidade pelo teste F. Portanto, não existe diferença significativa na produtividade média com o uso dos dois espaçamentos, para todas as variedades. **(0,1)**