UNIVERSIDADE FEDERAL DE LAVRAS
DEPATAMENTO DE CIÊNCIAS EXATAS
Programa de Pós-Graduação em Estatística e Experimentação Agropecuária

Prova do Processo Seletivo para Mestrado 2016-1

GABARITO

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. Ache a área total entre a curva $y = 1 - x^2$ e o intervalo [0,2] no eixo x.

** QUESTÃO ANULADA **

Todos os candidatos receberam 1,0 ponto pela questão

2. Suponha que $f(1) = 2, f(4) = 7, f'(1) = 5, f'(4) = 3$ e que f'' seja contínua. Calcule o valor de $\int_1^4 xf''(x)dx$. **(1,0 pt)**

$$\int_1^4 x\, f''(x)\, dx = x\, f'(x)\Big]_1^4 - \int_1^4 f'(x)\, dx$$

$$\int_1^4 x\, f''(x)\, dx = \big(4f'(4) - f'(1)\big) - f(x)\Big]_1^4 = 7 - 5 = 2.$$

3. A função de distribuição logística pode ser escrita como ($s > 0$)

$$F(x;\ \mu, s) = \frac{e^{\frac{x-\mu}{s}}}{1+e^{\frac{x-\mu}{s}}} = \frac{e^{\frac{x-\mu}{s}}}{1+e^{\frac{x-\mu}{s}}}.$$

Encontra a função densidade *f(x)* (função derivada) e a moda (ponto de densidade máxima) dessa distribuição.

Dado $F(u;\ \mu, s) = \frac{e^u}{1+e^u}$, com $u = \frac{x-\mu}{s}$, tem-se $\frac{du}{dx} = \frac{1}{s}$.

Derivada primeira ou $f(x)$: **(0,4 pts)**

$$f(x) = F'(x) = F'(u)\frac{du}{dx} = \frac{dF(u)}{du}\cdot\frac{1}{s} = \frac{e^u(1+e^u) - e^u e^u}{(1+e^u)^2}\cdot\frac{1}{s} = \frac{e^u}{s(1+e^u)^2}$$

Assim, $f(x) = \frac{e^{\frac{x-\mu}{s}}}{s\left(1+e^{\frac{x-\mu}{s}}\right)^2}$.

Derivada segunda de $F(x)$ ou $f'(x)$: **(0,4 pts)**

$$f'(x) = \frac{e^u \cdot s(1+e^u)^2 \cdot \frac{1}{s} - e^u \cdot 2s(1+e^u) \cdot e^u \cdot \frac{1}{s}}{[s(1+e^u)^2]^2}$$

$$= \frac{e^u \cdot (1+e^u)^2 - e^u \cdot 2(1+e^u) \cdot e^u}{[s(1+e^u)^2]^2}$$

$$= \frac{e^u \cdot (1+e^u) \cdot (1+e^u - 2e^u)}{[s(1+e^u)^2]^2} = \frac{e^u \cdot (1+e^u) \cdot (1-e^u)}{[s(1+e^u)^2]^2}$$

$$= \frac{e^u \cdot (1-e^{2u})}{[s(1+e^u)^2]^2} = \frac{e^{\frac{x-\mu}{s}} \cdot \left(1 - e^{2\frac{x-\mu}{s}}\right)}{\left[s(1+e^{\frac{x-\mu}{s}})^2\right]^2}$$

Assim, $f'(x) = \frac{e^{\frac{x-\mu}{s}} \cdot \left(1 - e^{2\frac{x-\mu}{s}}\right)}{\left[s(1+e^{\frac{x-\mu}{s}})^2\right]^2}$.

Solução para a moda (numerador igual a zero) **(0,2 pts)**

$$f'(x) = \frac{e^{\frac{x-\mu}{s}} \cdot \left(1 - e^{2\frac{x-\mu}{s}}\right)}{\left[s(1+e^{\frac{x-\mu}{s}})^2\right]^2} = 0$$

$$1 - e^{2\frac{x-\mu}{s}} = 0 \;\; \Rightarrow \;\; 2\frac{x-\mu}{s} = ln(1) = 0 \;\; \Rightarrow \;\; x = \mu$$

Portanto, a moda da distribuição logística é μ.

4. Uma amostra de n = 60 animais produtores de leite de uma população normal apresentou média de 12 litros/dia/animal e variância de 13,6 em uma região produtora. Construir os intervalos de 95% e de 99% de confiança para μ. Verificar e discutir o efeito provocado pelo aumento do coeficiente de confiança.

Dados: $t_{0,025;\nu=59}$ = 2,001 e $t_{0,005;\nu=59}$=2,662.

Para coeficiente de confiança de 95%:

$$IC_{1-\alpha}(\mu): \; \bar{X} \pm t_{\alpha/2;\nu=n-1}\frac{S}{\sqrt{n}} = 12 \pm 2,001 \times \frac{3,6878}{7,7460} = 12 \pm 0,9527$$

$$= [11,05; 12,95]$$

(0,35 pts)

Para o coeficiente de confiança de 99%:

$$IC_{1-\alpha}(\mu): \ \bar{X} \pm t_{\alpha/2;\nu=n-1}\frac{S}{\sqrt{n}} = 12 \pm 2,662 \times \frac{3,6878}{7,7460} = 12 \pm 1,2674$$
$$= [10,73; 13,27]$$

(0,35 pts)

Houve um aumento do comprimento do intervalo quando se passou de 95% para 99%. Isso ocorre sempre, pois o preço que se paga usando-se os mesmos recursos amostrais para se ter uma maior confiança (probabilidade de cobertura) é de intervalo de confiança mais amplo (maior comprimento). Nesse caso, embora a amplitude no segundo caso tenha sido maior, ela não foi expressivamente maior. **(0,30 pts)**

5. Determinar o tamanho da amostra *n* necessário para se estimar a proporção de empresas que utilizam comércio eletrônico na internet, considerando um erro máximo igual a 5% e confiança de 95%. Como proceder para ***tentar*** diminuir o valor de *n* necessário, sem alterar o erro máximo e a confiança?

Dado: $z_{0,025} = 1,96$.

Como não foi apresentado o resultado de uma amostra piloto, a fórmula a ser utilizada e seu resultado são:

$$n = \frac{Z^2_{\alpha/2}}{4e^2} = \frac{1,96^2}{4 \times 0,05^2} = 384,16$$
$$\approx 385.$$

Assim, são necessárias 385 empresas para se ter uma estimativa com erro máximo de 5 pontos percentuais para mais e para menos com 95% de confiança.
(0,6 pts)

Como não se pode mexer na confiança e nem no erro máximo (margem de erro) da pesquisa e como se utilizou uma fórmula em que o *n* foi dimensionado para a pior situação, ou seja, com $p=0,5$, então a única saída para tentar reduzir o tamanho amostral utilizado é trabalhar com uma amostra piloto e com a fórmula de dimensionamento dada por

$$n = \frac{Z^2_{\alpha/2}\hat{p}(1-\hat{p})}{e^2}$$

que será menor tanto quanto a estimativa de *p* se afastar de 0,5. Entretanto, não há garantias da redução, pois esse fato só acontecerá se a população a ser amostrada tiver *p* diferente de 0,5, o que não dá para garantir para população nenhuma. Ademais, a qualidade da estimativa na amostra piloto pode prejudicar a precisão da estimativa de *n* necessária. **(0,4 pts)**

6. Em um teste de hipótese existem dois tipos de erro envolvidos. Defina os dois erros e discuta sobre suas probabilidades e sobre os mecanismos que o pesquisador pode utilizar para controlá-los ou reduzi-los.

Os dois erros são o Erro do tipo I e o Erro do tipo II. O erro do tipo I ocorre quando uma hipótese verdadeira é rejeitada e o Erro tipo II é o erro que se comete ao não se rejeitar uma hipótese falsa. **(0,6 pts)**

As propriedades básicas dos dois tipos de erro são: a) o erro tipo I possui probabilidade de ser cometido igual a α que pode ser fixada pelo pesquisador *a priori*; b) o erro tipo II possui probabilidade de ocorrência igual a β e, em geral, não é controlada pelo pesquisador diretamente; c) as probabilidades dos dois tipos de erro são inversamente proporcionais; d) a probabilidade de se cometer o erro tipo II depende da distância dos valores hipotético e paramétrico; entre outras características. **(0,2 pts)**

Quando se fala em mecanismos de controle ou redução dos erros, abusa-se um pouco da linguagem, pois o controle na verdade é das probabilidades dos mesmos. De qualquer forma, os mecanismos de controle (redução das probabilidades dos erros) podem ser amostras maiores, maior qualidade do plano amostral, fixação a priori de uma diferença mínima significativa a ser detectada e utilização de testes melhores e mais eficientes, que os dados coletados atendam aos pressupostos requeridos para suas aplicações. **(0,2 pts)**

7. A correlação entre teor de alumínio e teor de cálcio em um tipo específico de solo foi estimada em uma amostra de tamanho igual a $n = 120$ e resultou no seguinte valor amostral, $r = -0{,}87$. Teste a hipótese de que o coeficiente de correlação linear populacional é nulo, assumindo normalidade bivariada entre as duas variáveis, ou seja, testar a hipótese $H_0: \rho = 0$ e retire as conclusões de interesse.

Dado: $t_{0{,}025;\nu=118}$=1,980.

** QUESTÃO ANULADA **

Todos os candidatos receberam 1,0 ponto pela questão

8. Um experimento foi conduzido para avaliar o efeito de diferentes cultivares na produção de grãos (em t/ha) de milho, sendo utilizado o delineamento em blocos casualizados, com quatro repetições. Sabendo-se que o quadrado médio do erro foi igual à $QMErro$ = 21,40 e que as médias de tratamentos são:

Tratamentos	Br10	Br15	U1	U2	H5	H6
Médias	14,5	10,5	9,4	17,0	5,7	9,5

Dado: $DMS = \Delta = q_{(5\%;\, I;\, GLErro)}\sqrt{\frac{QMErro}{r}}$; $DMS = S = \sqrt{(I-1)F_\alpha \frac{QMErro}{r}\sum_{i=1}^{I} c_i^2}$

$F_{(5\%;\,1,\,15)} = 4{,}54$; $F_{(5\%;\,2,\,15)} = 3{,}68$; $F_{(5\%;\,3,\,15)} = 3{,}29$; $F_{(5\%;\,4,\,15)} = 3{,}06$; $F_{(5\%;\,5,\,15)} = 2{,}90$; $F_{(5\%;\,1,\,18)} = 4{,}41$; $F_{(5\%;\,2,\,18)} = 3{,}55$; $F_{(5\%;\,3,\,18)} = 3{,}16$; $F_{(5\%;\,4,\,18)} = 2{,}93$; $F_{(5\%;\,5,\,18)} = 2{,}77$; $q_{(5\%;\,5,\,15)} = 4{,}37$; $q_{(5\%;\,6,\,15)} = 4{,}60$; $q_{(5\%;\,5,\,18)} = 4{,}28$; $q_{(5\%;\,6,\,18)} = 4{,}50$.

Pede-se:

a) Complete a tabela de análise da variância: **(0,2 pts)**

F.V.	G.L.	S.Q.	Q.M.	F_{cal}	F_{tab}
Blocos	3	-	-	-	-
Tratamentos	5	325,36	65,07	3,04*	2,90
Erro	15	321,00	21,40		
Total	23	1021,93			

$$SQTrat = \frac{1}{4}(58{,}0^2 + 42{,}0^2 + 37{,}6^2 + 68^2 + 22{,}8^2 + 38^2) - \frac{266{,}4^2}{24} = 325{,}36$$

b) Escreva as hipóteses envolvidas no teste F para tratamentos; **(0,1 pts)**

H_0: As cultivares de milho não se diferem
H_a: Existe ao menos uma diferença entre as cultivares
ou
H_0: $\mu_{Br10} = \mu_{Br15} = \cdots = \mu_{H6}$
H_a: Existe ao menos uma diferença

c) Qual o modelo linear apropriado para analisar este experimento? Escreva o modelo e o significado de seus componentes; **(0,2 pts)**

$$y_{ij} = \mu + t_i + b_j + e_{ij},$$

em que

y_{ij} representa os valores observados;
μ constante inerente à todas as observações;
t_i efeito da i-ésima cultivar, $i = 1, 2, 3, 4, 5, 6$;
b_j efeito do j-ésimo bloco, $j = 1, 2, 3, 4$;
e_{ij} erro experimental associado à observação y_{ij}, tal que $e_{ij} \sim N(0; \sigma^2)$.

d) Considerando o nível de 5% de probabilidade e as hipóteses elaboradas em (b), interprete os resultados obtidos de a análise da variância; **(0,2 pts)**

Como F_c = 3,04 > F_t = 2,90, então existem evidências para se rejeitar H_o, ao nível de 5% de significância, indicando que existe alguma diferença significativa entre os tratamentos. (ou entre as cultivares de milho com relação à produção de grãos)

e) Faça a comparação múltipla das médias de tratamentos utilizando o teste de Tukey ($\alpha = 5\%$) e interprete os resultados. **(0,2 pts)**

$$DMS = \Delta = q_{(5\%;\,6,\,15)}\sqrt{\frac{QMErro}{r}} = 4{,}60\sqrt{\frac{21{,}40}{4}} = 10{,}6$$

$\hat{m}_{U2} = 17{,}0\ a$
$\hat{m}_{Br10} = 14{,}5\ ab$
$\hat{m}_{Br15} = 10{,}5\ ab$
$\hat{m}_{H6} = 9{,}5\ ab$
$\hat{m}_{U1} = 9{,}4\ ab$
$\hat{m}_{H5} = 5{,}7\ \ b$

Tabela de médias

Cultivares	Médias
Br10	14,5 ab
Br15	10,5 ab
U1	9,4 ab
U2	17,0 a
H5	5,7 b
H6	9,5 ab

A cultivar U2 difere da cultivar H5, sendo U2 mais produtiva que H5. As cultivares Br10, Br15, U1, U2 e H6 não se diferem.

f) Utilizando-se o teste de Scheffé (α = 5%), teste o seguinte contraste, interpretando o resultado: **(0,1 pts)**

Y: as cultivares Br10 e Br15 contra as cultivares H5 e H6

$$Y = \mu_{Br10} + \mu_{Br15} - \mu_{H5} - \mu_{H6}$$

$\hat{Y} = 14{,}5 + 10{,}5 - 5{,}7 - 9{,}5 = 9{,}8$ t/ha

$$DMS = S = \sqrt{(I-1)F_{\alpha}\frac{QMErro}{r}\sum_{i=1}^{I} c_i^2}$$

$$= \sqrt{(6-1)\cdot F_{(5\%;\,5,15)}\cdot\frac{21{,}40}{4}\cdot(1+1+(-1)^2+(-1)^2)}$$

$$= \sqrt{5\cdot 2{,}90\cdot 21{,}40} = 17{,}6$$

Como $|\hat{Y}| < S$, não se rejeita H_0, logo o contraste é não significativo, indicando que as cultivares Br's não diferem das cultivares H's.

9. Com o propósito de verificar a eficiência de diferentes doses de hormônio em cordornas, em uma escala de 5 a 20 µg/kg, um experimento foi realizado e os resultados obtidos, encontram-se ilustrados na Figura 1.

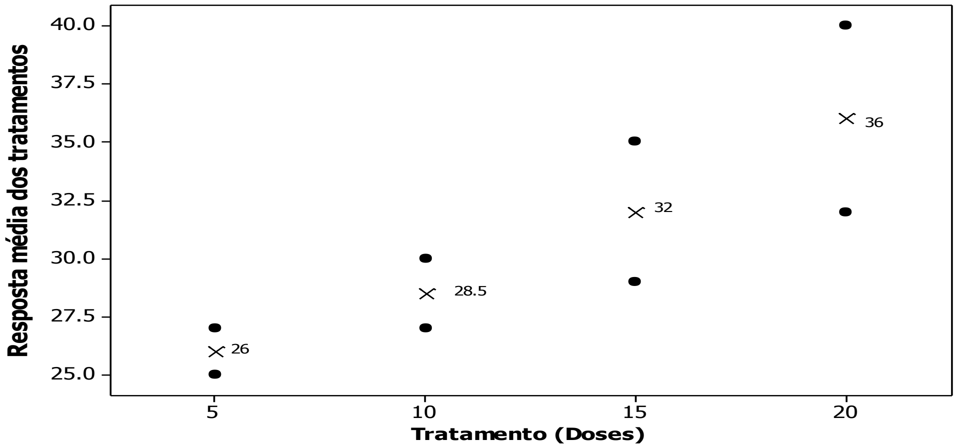

Figura 1 – Resultados experimentais obtidos para comparar diferentes dosagens de hormônios em codornas, em que × correspondem às respostas médias das doses e • correspondem aos valores observados.

Pede-se:

a) Quantas repetições cada tratamento apresenta? Considere que não há sobreposição de símbolos na figura. **(0,25 pts)**

Duas repetições, pois cada ponto cheio (●) representa um valor observado.

b) Qual tratamento proporcionou resposta mais precisa? **(0,25 pts)**

O tratamento ou dose de 5 µg/kg, pois os valores observados são menos heterogêneos, ou seja, que apresentam menor variabilidade.

c) Considerando a estimativa da variância do erro experimental dada por $S^2 = 14{,}123$, calcule o coeficiente de variação e interprete seu resultado. **(0,25 pts)**

$$CV = \frac{S}{\bar{\bar{X}}} \cdot 100$$

em que $S = \sqrt{S^2}$ e $\bar{X} = \frac{26{,}0+28{,}5+32{,}0+36{,}0}{4} = 30{,}625$. Assim,

$$CV = \frac{\sqrt{14{,}123}}{30{,}625} \cdot 100 = 12{,}27\%$$

d) Ajuste um modelo de regressão linear simples considerando a resposta média dos tratamentos em função das doses e interprete-o corretamente. Observe que, sendo o modelo dado por $Y_i = \beta_0 + \beta_1 X_i + e_i$, tem-se:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \text{ e } \hat{\beta}_1 = \frac{SPXY}{SQX} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}.$$ **(0,25 pts)**

$$\sum_{i=1}^{4} Y_i = 26{,}0 + 28{,}5 + 32{,}0 + 36{,}0 = 122{,}5; \quad \sum_{i=1}^{4} X_i = 5 + 10 + 15 + 20 = 50;$$

$$\bar{Y} = \frac{122{,}5}{4} = 30{,}625; \ \bar{X} = \frac{50}{4} = 12{,}5; \ \sum_{i=1}^{4} X_i^2 = 5^2 + 10^2 + 15^2 + 20^2 = 750;$$

$$\sum_{i=1}^{4} X_i Y_i = 5 \cdot 26{,}0 + 10 \cdot 28{,}5 + 15 \cdot 32{,}0 + 20 \cdot 36{,}0 = 1615{,}0$$

$$\hat{\beta}_1 = \frac{1615{,}0 - \frac{50 \cdot 122{,}5}{4}}{750 - \frac{50^2}{4}} = 0{,}67; \quad \hat{\beta}_0 = 30{,}625 - 0{,}67 \cdot 12{,}5 = 22{,}25$$

$$\hat{Y}_i = 22{,}25 + 0{,}67 X_i$$

Portanto, para um aumento de 1 µg/kg de hormônio, no intervalo de 5 a 20 µg/kg, espera-se um aumento de 0,67 na variável resposta.

10. Em um experimento fatorial, no delineamento inteiramente casualizado foram combinados dois níveis do fator A (A_1 e A_2) com três níveis do fator B (B_1, B_2 e B_3), ambos qualitativos, com três repetições. Os valores obtidos para cada repetição nos tratamentos avaliados são dados abaixo.

	B_1			B_2			B_3		
A_1	12	14	16	15	17	18	12	11	13
A_2	14	13	16	11	12	11	12	12	13

Dado: $SQtotal = 78{,}44$; $SQTratamentos = 57{,}78$;

$F_{(5\%;\,1,\,12)} = 4{,}75$; $F_{(5\%;\,2,\,12)} = 3{,}89$; $F_{(5\%;\,3,\,12)} = 3{,}49$;

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

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

Pede-se:

a) Verificar se os dois fatores (A e B) atuam independentemente. Considere $\alpha = 5\%$. **(0,5 pts)**

Quadro auxiliar (totais)

	B_1	B_2	B_3	Totais
A_1	$42^{(3)}$	50	36	$128^{(9)}$
A_2	43	34	37	114
Totais	$85^{(6)}$	84	73	$242^{(18)}$

$$SQA = \frac{1}{9}(128^2 + 114^2) - \frac{(242)^2}{18} = 10{,}89$$

$$SQB = \frac{1}{6}(85^2 + 84^2 + 73^2) - \frac{(242)^2}{18} = 14{,}78$$

$$SQAxB = SQTrat - SQA - SQB = 32{,}11$$

$$SQErro = SQtotal - SQTrat = 78{,}44 - 57{,}78 = 20{,}66$$

Tabela da Análise de Variância

FV	GL	SQ	QM	F_c
A	1	10,89		
B	2	14,77		
A x B	2	32,11	16,06	9,34
Erro	12	20,66	1,72	
Total	17	78,44		

Como o valor do F calculado (F_{cal} = 9,34) para a interação A x B é maior que o valor do F tabelado (F_{tab} = 3,89), concluímos que os fatores A e B atuam de forma dependente sobre a variável resposta em estudo.

b) Faça um estudo completo acerca dos níveis do fator A, concluindo corretamente. Considere α = 5% e aplique o teste de Tukey, se necessário. **(0,5 pts)**

Tabela da Análise de Variância do desdobramento de A dentro de B

FV	GL	SQ	QM	F_c
$A(B_1)$	1	0,17	0,17	$0{,}099^{ns}$
$A(B_2)$	1	42,66	42,66	$24{,}80^{*}$
$A(B_3)$	1	0,17	0,17	$0{,}099^{ns}$
Erro	12	20,66	1,72	

* Significativo a 5%; ^{ns} Não significativo.

$$SQA(B_1) = \frac{1}{3}(42^2 + 43^2) - \frac{(85)^2}{6} = 0{,}17$$
$$SQA(B_1) = \frac{1}{3}(50^2 + 34^2) - \frac{(84)^2}{6} = 42{,}66$$
$$SQA(B_1) = \frac{1}{3}(36^2 + 37^2) - \frac{(73)^2}{6} = 0{,}17$$

O $F_{tab} = F_{5\%;\,1,12}$ = 4,75 e, assim, tem-se que existe diferença significativa entre os níveis do fator A, apenas dentro de B_2. Dentro dos demais níveis de B, não existe diferença significativa entre os níveis do fator A.

Como o fator A apresenta apenas dois níveis, não é necessário aplicar o teste de Tukey, pois o teste F é conclusivo. Portanto, a tabela com as médias pode ser dada por:

Fator A	Fator B*		
	B_1	B_2	B_3
A_1	14,0 a	16,7 a	12,0 a
A_2	14,3 a	11,3 b	12,3 a

* Médias seguidas de mesmas letras nas colunas não diferem entre si ao nível de 5% de probabilidade, pelo teste F.

Assim, pode-se concluir que a média de A_1 é superior à média de A_2, dentro de B_2. Dentro dos demais níveis de B, não existe diferença significativa entre as médias de A_1 e A_2.