**DEPARTAMENTO DE ESTATÍSTICA - UFMG**

**PROVA DE ESTATÍSTICA & PROBABILIDADES**
**SELEÇÃO – MESTRADO/UFMG – 2012/2013**

## Instruções gerais sobre a Prova:

(a) Cada questão respondida ***corretamente*** vale **1 (um) ponto**.
(b) Cada questão respondida ***incorretamente*** vale **-1 (menos um) ponto**.
(c) Cada questão deixada em branco vale 0 (zero) pontos (neste caso marque TODAS as alternativas).
(d) Pelo menos 9 (nove) questões devem ser respondidas pelo candidato.
(e) A nota final será a soma dos pontos (negativos e positivos) de todas as questões.
(f) As opções escolhidas devem ser assinaladas na folha de respostas no final da prova.

**A prova tem duração de 3 horas**

**É proibido: usar celular; consultar referências bibliográficas diferentes das que estão estabelecidas no edital de seleção; emprestar calculadoras e/ou livros para consulta de outros candidatos durante a prova.**

**DEPARTAMENTO DE ESTATÍSTICA - UFMG**
**PROVA DE ESTATÍSTICA & PROBABILIDADES**
**SELEÇÃO – MESTRADO/UFMG – 2012/2013**

**Nome do (a) candidato (a):**________________________________________________

**Questão 1 –** Considere duas urnas. A urna 1 contém 2 bolas pretas e 3 bolas vermelhas; a urna 2 contém 3 bolas pretas e 2 bolas vermelhas. Uma moeda honesta é lançada para decidirmos de qual urna retirar uma bola. Suponha que a primeira bola retirada é preta e é devolvida à urna de origem. Uma segunda bola será retirada da mesma urna escolhida no passo anterior. Qual a probabilidade de que a segunda bola retirada também seja preta?

(a) 0,48
(b) 0,52
(c) 0,56
(d) 0,60

**Questão 2 –** Suponha que o abastecimento de água de uma cidade dependa de 3 reservatórios. Suponha que cada um destes reservatórios tenha capacidade máxima diária de 50.000 litros e que suas respectivas demandas diárias sejam independentes e tenham densidades exponenciais dadas respectivamente por:

Reservatório 1: 0,00002 exp(-0,00002 $\boldsymbol{x}$) , $\boldsymbol{x} \in (0,\infty)$ .
Reservatório 2: 0,00004 exp(-0,00004 $\boldsymbol{x}$) , $\boldsymbol{x} \in (0,\infty)$ .
Reservatório 3: 0,00005 exp(-0,00005 $\boldsymbol{x}$) , $\boldsymbol{x} \in (0,\infty)$ .

Qual a probabilidade aproximada de que a cidade fique sem água em um dia específico?

(a) 0,004
(b) 0,005
(c) 0,006
(d) 0,007

**Questão 3 –** Escolha um ponto uniformemente distribuído no intervalo (0,1) e o chame de $X$. Em seguida escolha um ponto uniformemente no intervalo (0, $X$ ) e o chame de $Y$. Qual é a função de densidade de $Y$ ?

(a) $\frac{1}{x} I_{\{0<y<x\}}$

(b) $1 \; I_{\{0<y<1\}}$

(c) $y \; I_{\{0<y<\sqrt{2}\}}$

(d) $-\ln y \; I_{\{0<y<1\}}$

**Questão 4 –** O número de pessoas que entram em um elevador no térreo de um prédio é uma variável aleatória com distribuição de Poisson com parâmetro $\lambda, \lambda > 0$. Suponha que há $N$ andares acima do térreo e que cada pessoa escolhe com igual probabilidade, independentemente das demais, em qual andar irá sair do elevador. Defina $Y$ como o número de paradas que o elevador fará até que todas as pessoas tenham saído. Seja E[$Y$ ] a esperança matemática de $Y$. Nesse caso, é correto dizer que a E[$Y$ ] é igual a

(a) $N(1 - e^{-\frac{\lambda}{N}})$

(b) $(1 - e^{-\frac{\lambda}{N}})$

(c) $N(1 + e^{\frac{\lambda}{N}})$

(d) $e^{-\frac{\lambda}{N}}$

**Questão 5 –** Um revendedor de componentes eletrônicos compra caixas com 10 componentes cada. É seu costume inspecionar 2 itens de cada caixa e decidir aceitá-la somente se os dois componentes avaliados estiverem funcionando. Se 30% das caixas tem 4 componentes defeituosos, 50% tem 1 defeituoso e 20% tem todos os componentes perfeitos, qual a proporção de caixas rejeitadas pelo revendedor?

(a) 28%
(b) 29%
(c) 30%
(d) 31%

**Questão 6 –** Considere 10 lançamentos independentes de uma moeda com probabilidade de sair cara igual a 0,4. Cada moeda que mostra cara é lançada novamente. Seja $Y$ o número de caras obtidas na segunda rodada de lançamentos. Nesse caso, $P[Y=2]$ é aproximadamente igual a:

(a) 0,2266
(b) 0,2631
(c) 0,2792
(d) 0,2854

**Questão 7 –** Deseja-se estimar o tempo médio diário (em minutos) que os candidatos a um concurso gastam estudando um determinado tópico do programa. Supondo que o tempo de estudo diário tem distribuição normal, escolha, dentre as alternativas apresentadas, aquela que apresenta o tamanho da amostra necessário para se estimar o tempo médio com uma margem de erro de 20 minutos. Suponha que se deseja um nível de confiança de 90% e que um estudo piloto tenha demonstrado que o desvio padrão do tempo de estudo é estimado em 100 minutos.

(a) 11
(b) 41
(c) 68
(d) 97

**Questão 8 –** Os dados a seguir referem-se a uma amostra aleatória de tamanho n=5 da variável aleatória $\boldsymbol{X}$: número de colônias de bactérias por 10 ml de água de um lago. Suponha que a distribuição de $\boldsymbol{X}$ seja Poisson $(\theta), \theta > 0$, desconhecido. Seja $g(\theta)$ a probabilidade de que seja encontrada uma colonia de bactérias a cada 10 ml de água do lago. Sejam $\hat{g}_1(\theta)$ e $\hat{g}_2(\theta)$ os estimadores UMVU e de máxima verossimilhança de $g(\theta)$ . Uma amostra aleatória de tamanho n=5 da variável aleatória $\boldsymbol{X}$ resultou nos seguintes valores:

$$x_1=4, x_2=7 , x_3=5 ,x_4=4, x_5=3$$

Com base nos dados amostrais pode-se afirmar que

(a) a estimativa de máxima verossimilhança de $g(\theta)$ é igual a UMVU.
(b) a estimativa de máxima verossimilhança de $g(\theta)$ é maior que a UMVU.
(c) a estimativa de máxima verossimilhança de $g(\theta)$ é menor que a UMVU.
(d) não há relação entre as estimativas pois a estimativa UMVU não existe.

**Questão 9 –** Deseja-se verificar, ao nível de 5% de significância, se existe diferença na perda média de peso (em kg) de indivíduos submetidos a três diferentes dietas. Cada dieta foi aplicada independentemente, a quatro indivíduos diferentes e parte dos resultados utilizada na análise estatística é apresentada na Tabela 1.

Tabela 1. Resultados da Questão 9

| **Fonte de Variação** | **Soma de Quadrados** |
|---|---|
| Dieta | 31,60 |
| Erro | 9,55 |
| Total | 41,15 |

Supondo que o peso perdido em cada dieta tenha distribuição normal de mesma variância, verifique se cada uma das afirmativas seguintes é verdadeira **(V)** ou falsa **(F)**.

**(I)** A hipótese nula assume que a perda média de peso das três dietas é igual.

**(II)** O valor aproximado da estatística de teste é 14,9.

**(III)** Conclui-se que não há evidências suficientes para afirmar que há diferença na perda média de peso das três dietas.

**(IV)** O valor crítico obtido da distribuição a ser utilizada para o teste é 4,256.

Escolha a opção correta da seqüência de **‘V’** e **‘F’** para as quatro afirmativas anteriores:

(a) V V F V
(b) F V F F
(c) V V F F
(d) V F V V

**Questão 10 –** Seja $X$ uma variável aleatória com distribuição Uniforme no intervalo (0, $\theta$), $\theta > 0$. Considere as seguintes hipóteses nula e alternativa: $\boldsymbol{H_0 : \theta \leq 5}$ e $\boldsymbol{H_A : \theta > 5}$. O teste UMP (uniformemente mais poderoso) foi construído para testar a hipótese nula, considerando um nível de significância igual a 5% e um tamanho de amostra igual a 25. Se o valor verdadeiro de $\boldsymbol{\theta}$ for igual a 5,3, o poder do teste será

(a) 0,6432
(b) 0,7784
(c) 0,9415
(d) 0,9900

**Questão 11 –** Um distribuidor de carne bovina deseja determinar se há relação entre a área geográfica e o tipo de carne que os consumidores preferem. Para isto ele observou uma amostra aleatória de 500 consumidores provenientes da região norte e sul da localidade em estudo sendo que cada consumidor foi questionado em relação a sua preferência pelos tipos A, B e C de carne. Os resultados obtidos estão na Tabela 2.

Tabela 2. Dados da Questão 11

| | | Região geográfica | | |
|---|---|---|---|---|
| | | Norte | Sul | Total |
| Tipo de carne | A | 100 | 50 | 150 |
| | B | 150 | 125 | 275 |
| | C | 50 | 25 | 75 |
| Total | | 300 | 200 | 500 |

Nesse caso, é **correto** afirmar que

(a) ao nível de significância de 2,5% não existe associação significativa entre preferência pelo tipo de carne e região geográfica. No entanto, ao nível de significância de 1% a associação é significativa.

(b) ao nível de significância de 1% não existe associação significativa entre preferência pelo tipo de carne e região geográfica. No entanto, ao nível de significância de 2,5% a associação é significativa.

(c) tanto ao nível de significância de 1% quanto ao de 2,5%, não existe associação significativa entre preferência pelo tipo de carne e região geográfica.

(d) tanto ao nível de significância de 1% quanto ao de 2,5% existe associação significativa entre preferência pelo tipo de carne e região geográfica.

**Questão 12 –** Um experimento foi conduzido para determinar se a temperatura de queima e/ou a posição da fornalha afetam a densidade de um anodo de carbono. Foram analisadas três temperaturas de queima (800ºC, 825ºC, 850ºC), duas posições da fornalha (A e B) e três réplicas em cada combinação temperatura x posição. Suponha que a densidade do anodo de carbono em cada tratamento tenha distribuição normal de mesma variância desconhecida $\sigma^2$. Parte dos resultados utilizada para análise estatística é apresentada na Tabela 3.

Tabela 3. Resultados da Questão 12

| **Fonte de Variação** | **Soma de Quadrados** |
|---|---|
| Temperatura | 2.671 |
| Posição | 7.160 |
| Temperatura x Posição | 818 |
| Erro | 5.371 |
| Total | 16.020 |

Escolha, dentre as alternativas a seguir, aquela que é verdadeira.

(a) ao nível de significância de 10%, conclui-se que há efeito de interação entre temperatura de queima e posição da fornalha na densidade do anodo de carbono.

(b) ao nível de significância de 10%, conclui-se que apenas a temperatura de queima influencia na densidade do anodo de carbono.

(c) ao nível de significância de 5%, conclui-se que tanto a temperatura de queima quanto a posição da fornalha influenciam na densidade do anodo de carbono, mas não há efeito de interação entre esses fatores.

(d) ao nível de significância de 5%, conclui-se que apenas a posição da fornalha influencia na densidade do anodo de carbono.

**Questão 13 –** Duas variedades de milho (A e B) foram comparadas para verificar se havia alguma diferença significativa no tempo de maturação. Sementes da variedade A foram plantadas em 10 subáreas distintas e sementes da variedade B foram plantadas em outras 10 áreas distintas. As médias amostrais observadas para as variedades A e B foram respectivamente: 95 e 74, sendo a diferença entre as médias igual a 21. O intervalo de 95% de confiança construído para a diferença populacional de tempos médios de maturação das duas variedades, considerando-se aproximação pela distribuição normal, foi igual a: (15,9; 26,1). Com base neste intervalo poder-se-ia concluir ao nível de 5% de significância que as duas variedades:

(a) diferem estatisticamente em relação ao tempo médio de maturação, pois o intervalo de confiança contém valores possíveis de diferenças acima de 21.

(b) diferem estatisticamente em relação ao tempo médio de maturação sendo que a variedade A tem um tempo médio de maturação maior que a variedade B.

(c) não diferem estatisticamente em relação ao tempo médio de maturação mas a variedade B é melhor que a variedade A.

(d) não diferem estatisticamente em relação ao tempo médio de maturação pois o valor de diferença 21 encontra-se dentro do intervalo de confiança.

**Questão 14 –** Considere uma variável aleatória contínua cuja distribuição é simétrica em torno da mediana ($\mu_d$). Desejamos testar, ao nível de significância $\alpha$=0,05, se a média ($\mu$) dessa variável é maior que 33. Coletou-se então, uma amostra de tamanho 26 e o valor observado da estatística de teste (já normalizado para a distribuição conveniente), foi igual a 1,90.

Marque a alternativa **correta.**

(a) o teste aplicado é o teste *t* de Student e devemos concluir que há evidências suficientes de que a média é maior que 33.

(b) o teste aplicado é o teste Z e devemos concluir que há evidências suficientes de que a média é maior que 33.

(c) o teste aplicado é o teste de Wilcoxon e o valor aproximado da probabilidade de significância do teste é 0,029.

(d) o teste aplicado é o teste dos sinais e o valor aproximado da probabilidade de significância do teste é 0,058.

**Questão 15 –** Considere os modelos de regressão linear simples **(I)** e **(II)** apresentados a seguir, em que os $X_i's$ são fixos (não aleatórios).

**(I)** $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$**,** em que $\varepsilon_i$ são variáveis aleatórias independentes e identicamente distribuídas, com distribuição normal de média zero e variância constante $\sigma^2$.

**(II)** $Y_i = \alpha_0 + \alpha_1 \left( X_i - \bar{X} \right) + \tau_i$**,** em que $\tau_i$ são variáveis aleatórias independentes e identicamente distribuídas, com distribuição normal de média zero e variância constante $\delta^2$.

Para esses modelos, considerando uma amostra de tamanho *n*, os estimadores de máxima verossimilhança dos parâmetros de interesse são dados por:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X};\ \hat{\alpha}_0 = \bar{Y};\ \hat{\beta}_1 = \hat{\alpha}_1 = \frac{\sum_{i=1}^{n} Y_i \left( X_i - \bar{X} \right)}{\sum_{i=1}^{n} \left( X_i - \bar{X} \right)^2};\hat{\sigma}^2 = \hat{\delta}^2 = \frac{\sum_{i=1}^{n} \left( Y_i - \hat{Y}_i \right)^2}{n}.$$

Além disso, tem-se que

$$Var\left(\hat{\beta}_0\right) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum_{i=1}^{n} \left( X_i - \bar{X} \right)^2};Var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n} \left( X_i - \bar{X} \right)^2};Var\left(\hat{\alpha}_0\right) = \frac{\delta^2}{n};Var\left(\hat{\alpha}_1\right) = \frac{\delta^2}{\sum_{i=1}^{n} \left( X_i - \bar{X} \right)^2}$$

Verifique se cada uma das afirmativas seguintes é verdadeira **(V)** ou falsa **(F):**

**(I)** $COV\left(\hat{\beta}_0, \hat{\beta}_1\right) < 0$

**(II)** $COV\left(\hat{\alpha}_0, \hat{\alpha}_1\right) > 0$

**(III)** Os estimadores $\hat{\alpha}_0$ e $\hat{\beta}_0$ são não viciados.

**(IV)** Pode-se utilizar a distribuição *t* de Student com (*n-1*) graus de liberdade para testar hipóteses sobre $\beta_0,\ \beta_1,\ \alpha_0$ e $\alpha_1$.

Escolha a **opção correta** da seqüência de **'V'** e **'F'** para as quatro afirmativas anteriores:

(a) V F V V
(b) F F V F
(c) V V F V
(d) F V V F

**DEPARTAMENTO DE ESTATÍSTICA - UFMG**
**PROVA DE ESTATÍSTICA & PROBABILIDADES**
**SELEÇÃO – MESTRADO/UFMG – 2012/2013**

**20/11/2012**

***Instruções:***

*a) No quadro abaixo, assinale com um X a opção de resposta escolhida para cada questão*
*b) USE CANETA*

| Questão | Resposta | | | | Pontuação |
|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | |
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |

NOME COMPLETO: ______________________________________________________________

IDENTIDADE/PASSAPORTE Nº: _____________________________________________________

ASSINATURA: _____________________________________________________