UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Probabilidade e Estatística

Prova Final **08-03-2016**

Atenção: Não serão aceitas respostas sem justificativa.

1. Admita que cães e gatos são os únicos animais domésticos das famílias que moram em um condomínio residencial. Sejam X e Y, respectivamente, o número de cães e o número de gatos de uma família sorteada ao acaso nesse condomínio. A distribuição conjunta, juntamente com as marginais, das variáveis aleatórias discretas X e Y é dada pela tabela a seguir:

X	Y			$\Pr(X = x_i)$
	0	1	2	
0	0	p	q	0,2
1	p	r	p	0,6
2	q	p	0	0,2
$\Pr(Y = y_i)$	0,2	0,6	0,2	1

sendo p, q e r constantes a determinar. Sabendo que $\text{Cov}(X,Y) = -\,0,1$, calcule:

(a) Os valores de p, q e r;

(b) A probabilidade de que uma família tenha no máximo um animal doméstico, dado que essa família só tem animais de um mesmo tipo (ou seja, só cães ou só gatos).

2. Uma instituição de sauúde, preocupada com a má alimentação das crianças, deseja realizar uma pesquisa para investigar a obesidade delas numa determinada região, através do índice de massa corpórea (IMC). Sabe-se que o IMC, em kg/m^2, das crianças nesta região segue uma distribuição Normal com média 21,5 e desvio padrão de 2 e que indivíduos com IMC superior a 23 são considerados obesos.

(a) Com base nestas informações, calcule a probabilidade de que uma criança escolhida ao acaso não seja obesa.

(b) Em uma amostra aleatória de 20 crianças, qual a probabilidade de que exatamente 15 dessas crianças não sejam obesas.

3. Um aluno está interessado em estudar o tempo que ele gasta esperando o ônibus para retornar a sua casa. Admita que esse tempo seja descrito por uma variável aleatória com distribuição Uniforme no intervalo $(0, \theta)$ e que ele deseje estimar esse parâmetro θ. Para isso, usará o estimador $\hat{\theta} = 2\bar{X}$, onde $\bar{X}$ é a média de uma amostra de tempos de espera a ser coletada.:

(a) O estimador $\hat{\theta}$ é viesado (ou tendencioso) para estimar θ? Justifique.

(b) Calcule o erro quadrático médio do estimador $\hat{\theta}$.

(c) Considere que uma amostra com 12 tempos de espera foi coletada e os seguintes valores foram observados, em minutos: $7,2$; $3,5$; $11,0$; $0,8$; $5,8$; $9,1$; $2,6$; $28,3$; $15,7$; $6,0$; $12,8$; $8,5$. Construa um boxplot para esses dados e baseado nele responda: A hipótese dos tempos de espera terem distribuição uniforme parece coerente com os dados? Justifique.

4. O número médio diário de clientes de um posto de gasolina tem sido 250. Durante uma campanha de 25 dias, em que os clientes recebiam um brinde, o número médio de clientes foi 270, com um desvio padrão de 60, ambos amostrais. Suponha que o número de clientes se comporta segundo uma distribuição Normal, antes e após a campanha.

(a) Há razões para suspeitar que a campanha modificou, para mais ou para menos, o número médio de clientes do posto. Responda essa pergunta formulando o problema como um teste de hipóteses ao nível de significância de 5%?

(b) A partir de que nível de significância esta conclusão mudaria?

(c) Obtenha o intervalo de 95% de confiança para o número médio diário de clientes.

Boa prova!

1. **Gabarito:**

   (a) A determinação de p, q e r pode ser feita resolvendo um sistema de 3 equações a 3 incógnitas:

   - A soma das probabilidades conjuntas que estão na linha onde $X = 0$ deve ser igual a $\Pr(X = 0) = 0,2$. Então, $p + q = 0,2$. (I)
   - A soma das probabilidades conjuntas que estão na linha onde $X = 1$ deve ser igual a $\Pr(X = 1) = 0,6$. Então, $2p + r = 0,6$. (II)
   - Temos $\mathrm{Cov}(X,Y) = \mathrm{E}(XY)–\mathrm{E}(X)\mathrm{E}(Y) = -\,0,1$.
     Por outro lado, $\mathrm{E}(X) = \mathrm{E}(Y) = 0 \times 0,2 + 1 \times 0,6 + 2 \times 0,2 = 1$.
     Além disso, $\mathrm{E}(XY) = 1 \times 1 \times r + 1 \times 2 \times p + 2 \times 1 \times p = r + 4p$.
     Portanto, $r + 4p–1 \times 1 = -\,0,1$. Ou seja, $r + 4p = 0,9$. (III)

   A solução do sistema formado pelas equações (I), (II) e (III) é $p = 0,15$, $q = 0,05$ e $r = 0,30$.

   (b)

$$\begin{aligned}\Pr(X+Y \leqslant 1|\{X=0\}\cup\{Y=0\}) &= \frac{\Pr(X=0,Y=0)+\Pr(X=1,Y=0)+\Pr(X=0,Y=1)}{\Pr(\{X=0\}\cup\{Y=0\})} \\ &= \frac{2p}{2p+2q} = \frac{p}{p+q} = \frac{0,15}{0,15+0,05} = 0,75,\end{aligned}$$

   pois

$$\begin{aligned}\Pr(\{X=0\}\cup\{Y=0\}) &= \Pr(X=0,Y=0)+\Pr(X=1,Y=0)+\Pr(X=2,Y=0) \\ &+ \Pr(X=0,Y=1)+\Pr(X=0,Y=2).\end{aligned}$$

---

2. **Gabarito:**

   (a) Seja X uma variável aleatória que representa o IMC das crianças nesta região, tal que $X \sim \mathrm{Normal}(21,5\ ;\ 2^2)$. Deseja-se:

$$\Pr(X < 23) = \Pr\left(\frac{X-21,5}{2} < \frac{23-21,5}{2}\right) = \Pr(Z < 0,75) = 0,7734 \approx 0,77,$$

   pela tabela da Normal padrão.

   (b) Seja Y uma variável aleatória que representa o número de crianças não obesas nesta amostra. Tem-se que $Y \sim \mathrm{Binomial}(20\ ;\ 0,77)$. Deseja-se:

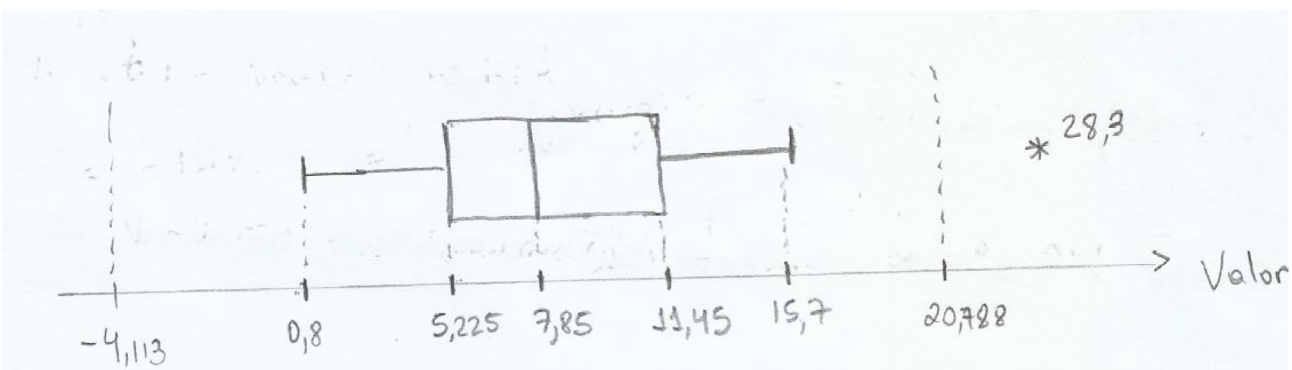
$$\Pr(Y = 15) = \binom{20}{15} \times 0,77^{15} \times 0,23^{5} = 0,1979.$$

3. **Gabarito:**

(a) $\mathrm{E}(\widehat{\theta}) = \mathrm{E}(2\overline{X}) = 2\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = 2\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i)$. Como $X \sim \text{Uniforme}(0,\theta)$, temos que $\mathrm{E}(X_i) = \frac{\theta}{2}$. Daí, $\mathrm{E}(\widehat{\theta}) = \mathrm{E}(2\overline{X}) = \frac{2}{n}\sum_{i=1}^{n}\frac{\theta}{2} = \theta$. Logo, como $\mathrm{E}(\widehat{\theta}) = \theta$, temos que $\widehat{\theta}$ é um estimador não viesado (ou não tendencioso) para θ.

(b) $\mathrm{EQM}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}) + [\mathrm{B}(\widehat{\theta})]^2$. Como $\widehat{\theta}$ é um estimador não viesado (ou não tendencioso) para θ, temos que $\mathrm{B}(\widehat{\theta}) = \mathrm{E}(\widehat{\theta}) - \theta = 0$. Por outro lado, $\mathrm{Var}(\widehat{\theta}) = \mathrm{Var}(2\overline{X}) = 4\mathrm{Var}(2\overline{X}) = 4\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)$. Como $X \sim \text{Uniforme}(0,\theta)$, temos que $\mathrm{Var}(X_i) = \frac{\theta^2}{12}$. Daí, $\mathrm{Var}(\widehat{\theta}) = \mathrm{Var}(2\overline{X}) = \frac{4}{n^2}\sum_{i=1}^{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}$. Logo, $\mathrm{EQM}(\widehat{\theta}) = \frac{\theta^2}{3n}$.

(c) Ordenar a amostra 0,8; 2,6; 3,5; 5,8; 6,0; 7,2; 8,5; 9,1; 11,0; 12,8; 15,7; 28,3.
$q_1 = x_{((n+3)/2)} = x_{(15/4)} = x_{(3{,}75)} = 0,25x_{(3)} + 0,75x_{(4)} = 0,25 \times 3,5 + 0,75 \times 5,8 = 5,225$.
$q_2 = x_{((n+1)/2)} = x_{(13/2)} = x_{(6{,}5)} = 0,5x_{(6)} + 0,5x_{(7)} = 0,5 \times 7,2 + 0,5 \times 8,5 = 7,85$.
$q_3 = x_{((3n+1)/4)} = x_{(37/4)} = x_{(9{,}25)} = 0,75x_{(9)} + 0,25x_{(10)} = 0,75 \times 11,0 + 0,25 \times 12,8 = 11,45$.
$\mathrm{DIQ} = q_3 - q_1 = 11,45 - 5,225 = 6,225$.
$\mathrm{CI} = q_1 - 1,5\mathrm{DIQ} = 5,225 - 1,5 \times 6,225 = -4,113$.
$\mathrm{CS} = q_3 + 1,5\mathrm{DIQ} = 11,45 - 1,5 \times 6,225 = 20,788$.



A distribuição Uniforme não parece razoável. Temos assimetria à direita, que não ocorre na distribuição Uniforme. Ainda mais, teríamos $\theta > 28,3$ e no intervalo [16; 28] não temos nenhuma observação, enquanto que em [4; 16] (mesmo comprimento) temos 9 das 12 observações. Esse padrão de comportamento não condiz com uma distribuição Uniforme.

4. **Gabarito:**

(a) Seja X uma variável aleatória que representa o número diário de clientes de um posto de gasolina. Sabe-se que $X \sim \text{Normal}(250, \sigma^2)$. Após a campanha, deseja-se verificar se a o número médio de clientes alterou-se, a partir de uma amostra de tamanho $n = 25$. Logo, o problema constitui-se de um teste de hipóteses para a média μ com variância σ^2 desconhecida, o qual está descrito nos seguintes passos:

(i) $H_0 : \mu = 250$ vs $H_1 : \mu \neq 250$

(ii) Estatística de teste: $T = \dfrac{\overline{X} - 250}{S/\sqrt{n}}$, a qual sob H_0 apresenta distribuição $\mathrm{t}_{n-1} = t_{24}$.

(iii) Pela tabela da t_{24}, tem-se que rejeita-se H_0 se $T < -2,06$ ou $T > 2,06$, ou equivalentemente, se $\overline{X} < 225,28$ ou $\overline{X} > 274,72$.

(iv) Estatística de teste calculada para a amostra: $t_{obs} = \dfrac{270 - 250}{60/\sqrt{25}} = 1,67$.

(v) Decisão: como $t_{obs} = 1,67$ não pertence a região de rejeição, então não rejeita-se H_0.

Portanto, não há razões para suspeitar que a campanha modificou o número médio de clientes do posto ao nível de significância de 5%.

Outra maneira de resolver é substituindo o passo (iii) pelo cálculo do p-valor, da seguinte forma: a estatística de teste observada para a amostra é $t_{obs} = 1,67$. Logo, p-valor $= 2P(T > 1,67) = 2 \times 0,054 = 0,108 = 10,8\%$, o qual é maior que o nível de significância de 5%, portanto obtemos a mesma conclusão do teste em (a).

(b) Como p-valor $= 2P(T > 1,67) = 2 \times 0,054 = 10,8\%$, então para qualquer nível de significância maior ou igual a 10,8% obtemos uma conclusão diferente do item (a), portanto nos levaria a rejeição de H_0.

(c) O intervalo de confiança de 95% para o número médio diário de clientes é dado por:

$$
\begin{aligned}
\mathrm{IC}_{95\%}(\mu) &= \left(\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}};\ \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right) \\
&= \left(270 - 2,06 \times \frac{60}{\sqrt{25}};\ 270 + 2,06 \times \frac{60}{\sqrt{25}}\right) \\
&= (245,28;\ 294,72).
\end{aligned}
$$