Exercícios e Complementos ME706 - 2017

González-López, V. A.*

May 10, 2017

1. Nos seguintes bancos de dados determine (i) o esquema dos cinco números, (ii) o boxplot, (iii) as observações discrepantes (outliers).

Table 1: Observações do índice Dow Jones. Período: 16/08/2001-29/09/2001

10.39	10.24	10.32	10.17	10.28	10.23	10.42	10.38	10.22	10.1
9.92	9.95	9.99	10.03	9.84	9.61	8.92	8.90	8.76	8.38

Table 2: Observações da cidade de New York, ano 1973.

Ozônio	Radiação Solar	Vento	Temperatura
91	189	4.6	93
47	95	7.4	87
32	92	15.5	84
20	252	10.9	80
23	220	10.3	78
21	230	10.9	75
24	259	9.7	73
44	236	14.9	81
21	259	15.5	76
28	238	6.3	77
9	24	10.9	71
13	112	11.5	71
46	237	6.9	78
18	224	13.8	67
13	27	10.3	76
24	238	10.3	68
16	201	8.0	82
13	238	12.6	64
23	14	9.2	71
36	139	10.3	81
7	49	10.3	69

*University of Campinas, Brazil. E-mail: veronica@ime.unicamp.br

Table 3: Medições, em minutos de espera até a próxima erupção do gêiser Old Faithful em Yellowstone National Park, Wyoming, USA.

79	51	20	78	69	74	83	55	76	78	79	73	77	66	80	74	52	48	80	59	90	80

2. Compute o *Interquantile Range* (IR) e o *Normalized Interquantile Range* (IRN) das seguintes distribuições:

   i. $\text{N}(\mu, \sigma^2)$,

   ii. $\text{Laplace}(\mu, b)$,

   iii. $\text{Cauchy}(\mu, \gamma)$.

   Onde μ é o parâmetro de locação em todos os casos.

3. Consistência. Defina (i) consistência em probabilidade e (ii) consistência uniforme, apresente exemplos para tais conceitos. Prove:

   i. Dada $\{Z_n\}_n$ uma sequência de variáveis e Z_0 um valor constante. Se $Z_n \to^P Z_0$, quando $n \to \infty$, g é uma função contínua em Z_0, então $g(Z) \to^P g(Z_0)$, quando $n \to \infty$. Ou seja a consistência em probabilidade é preservada por continuidade.

   ii. Seja X uma variáverl aleatória, não negativa, $f : \mathbb{R} \to \mathbb{R}$, crescente tal que $\mathbb{E}(f(X)) < \infty$. Então, $\forall a \in \mathbb{R}$, $f(a)P(X \geq a) \leq \mathbb{E}(f(X))$.

      a. Identifique a desigualdade em questão se: $f(x) = |x|^k, a > 0$.

      b. Identifique a função f e a variável sobre a qual é considerada a desigualdade de modo a obter a desigualdade de Chebyshev: $P(|X - \mathbb{E}(X)| \geq a) \leq \frac{Var(X)}{a^2}$.

      c. Identifique a função f que aplicada na variável X de média μ e variância σ^2, sobre a qual é considerada a desigualdade, garante a relação: $P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$.

4. Seja $\mathbb{P} = \left\{p = (p_1, \cdots, p_k) : 0 \leq p_j \leq 1, 1 \leq j \leq k, \sum_{j=1}^k p_j = 1\right\}$. Suponha que $p_j = Prob(X_1 = x_j)$ onde x_j é a categoria j entre as possíveis k categorias. Considere uma amostra de tamanho n, e seja N_j o total de objetos observados na categoria j, assim $\hat{p}_j = \frac{N_j}{n}, j = 1, \cdots, k$. Dada uma aplicação $q : \mathbb{P} \to R^s$, contínua. Então, $\hat{q}_n = q(\hat{p}_n)$ é uniformemente consistente para $q(p)$, onde $\hat{p}_n = (\hat{p}_1, \cdots, \hat{p}_k)$.

5. Mostre que as seguintes distribuições são membros da família exponencial e ache as suas formas canônicas.

   a. $\text{N}(\mu, \sigma^2)$: (i) μ conhecido e σ^2 desconhecido. (ii) σ^2 conhecido e μ desconhecido.

   b. $\text{Gama}(p, \lambda)$: (i) p conhecido e λ desconhecido. (ii) λ conhecido e p desconhecido.

   c. $\text{Beta}(r, s)$: (i) r conhecido e s desconhecido. (ii) s conhecido e r desconhecido.

   Em cada caso acima apresentado, considere uma amostra $x_1, \cdots, x_n$, identifique a estatística suficiente em função da amostra $T^* = T(x_1, \cdots, x_n)$ e determine a sua distribuição. Ache o valor esperado $\mathbb{E}(T^*)$ e a variância $Var(T^*)$.

6. Considere a família de distribuições $\text{N}(\mu, \sigma^2)$.

   a. Mostre que ela pertence à família exponencial.

   b. Apresente a sua representação canônica.

   c. Considere uma amostra $x_1, \cdots, x_n$ e identifique a estatística suficiente para (μ, σ^2) em termos da amostra. Determine média e variância-covariância da estatística identificada.

7. Discuta e exemplifique o problema de identificabilidade.

8. Sejam $x_1, \cdots, x_n$ i.i.d. $\mathrm{N}(\mu, \sigma^2)$, $T_1 = \frac{\sum_{i=1}^n x_i}{n}$ e $T_2 = \frac{\sum_{i=1}^n x_i^2}{n}$. Identifique a distribuição assintótica de $\sqrt{n}(\overline{X} - \mu, \hat{\sigma}^2 - \sigma^2)$, onde $\hat{\sigma}^2 = T_2 - (T_1)^2$.

9. Famílias de Locação- Escala.

   a. Mostre que as seguintes distribuições são famílias de *Locação*: (i)$\mathrm{N}(\mu, \sigma^2)$, σ^2 conhecido; (ii) Cauchy(μ,1) onde a densidade é $f(x;\mu) = \frac{1}{\pi} \frac{1}{(1+(x-\mu)^2)}$; (iii) Unif($\theta - \frac{1}{2}$, $\theta + \frac{1}{2}$).

   b. Mostre que as seguintes distribuições são famílias de *Escala*: (i) N(0,σ^2); (ii) Exp(λ); (iii) Unif(θ, 2θ).

   c. Mostre que as seguintes distribuições são famílias de *Locação-Escala*: (i) $\mathrm{N}(\mu, \sigma^2)$; (ii) E2($\mu, \sigma$) onde a densidade é dada por $f(x;\mu,\sigma) = \frac{1}{2\sigma} e^{-|\frac{x-\mu}{\sigma}|}$; (iii) Cauchy($\mu$,$\sigma^2$) onde a densidade é $f(x;\mu,\sigma) = \frac{1}{\pi\sigma} \frac{1}{1+\frac{(x-\mu)^2}{\sigma^2}}$.

10. Considere o modelo $x_i = \mu + \sigma u_i$ onde $u_i, i = 1, \cdots, n$ são i.i.d. com distribuição F, e F é dada pela seguinte forma: $F = (1-\epsilon)\Phi + \epsilon F^*$, sendo Φ a distribuição acumulada da N(0,1).

    a. Fixe diferentes valores de ϵ e fixe a F^* na Exponencial (λ), variando a taxa λ. Simule valores de x_i. Se o objetivo é estimar μ compare o desempenho de dois estimadores: $\overline{x}$ e a $mediana\{x_1, \cdots, x_n\}$.

    b. Fixe diferentes valores de ϵ e fixe a F^* na Uniforme, variando o intervalo. Simule valores de x_i. Se o objetivo é estimar μ compare o desempenho de dois estimadores: $\overline{x}$ e a $mediana\{x_1, \cdots, x_n\}$.

11. Assuma o seguinte modelo $x_i = \mu + \sigma u_i, i = 1, \cdots, n$. Compute o M-estimador de locação, de Huber em cada caso apresentado no exercício 1, explore os seguintes valores de k, $k = 1, 1.5, 2, 2.5$. Determine o valor do Median Absolute Deviation MAD para cada situação e inspecione seus resultados com base no boxplot dos dados.

12. No contexto do exercício anterior compare o M-estimador de locação de Huber com o estimador de máxima verossimilhança e com a mediana.

13. Determine a função característica das seguintes distribuições:

    i. Distribuição de Laplace $f(x) = \frac{e^{-|x|}}{2}$, $-\infty < x < \infty$.

    ii. Distribuição de Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

    iii. Distribuição Normal $f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$, $-\infty < x < \infty$.