

MAE0217 - Estatística Descritiva - Primeiro semestre de 2017
Professora: Márcia D'Elia Branco

LISTA 1

1) Os dados representam velocidades do vento (km/h) num determinado aeroporto para os primeiros 15 dias de dezembro de 2008.

Dia	1	2	3	4	5	6	7	8
Velocidade	22,2	61,1	13,7	27,8	22,7	7,4	8,7	6,3
Dia	9	10	11	12	13	14	15	
Velocidade	20,4	25,6	23,2	11,1	13,0	7,2	14,8	

(a) Calcule a média, a mediana, o desvio padrão e os quartis da velocidade e desenhe o gráfico *boxplot*. Faça manualmente.

(b) Existe algum valor atípico? Em caso afirmativo, qual? Remova esse valor e refaça o item anterior. Comente as diferenças encontradas.

2) Considere os dados da planilha “dadosdomiciliosCEA15P02.xlsx”. Use o R para responder os itens a seguir.

(a) Construa os gráficos *boxplot* para a variável “consumo de gás anual domiciliar per capita (em Kg)” , segundo o tipo de domicílio (Casa, Apartamento, Comodo).

(b) Obtenha as medidas resumos (média, mediana, min, max, Q1, Q3 e dp) da variável “consumo de gás anual domiciliar per capita (em Kg)” , segundo o tipo de domicílio (Casa, Apartamento, Comodo). Apresente esses resultados em uma tabela.

(c) Analise os resultados (comente) obtidos com base nos resultados de (a) e (b). Compare os grupos em relação a tendência central, dispersão e existência de valores discrepantes.

(d) Construa os gráficos *boxplot* para a variável “consumo de gás anual domiciliar per capita (em Kg)” , segundo a presença de rede geral de energia elétrica. Compare os grupos em relação a assimetria dos dados.

(e) Determine os coeficientes de assimetria e curtose, baseados em momentos, da variável “consumo de gás anual domiciliar per capita (em Kg)” , para cada grupo (segundo a presença de rede geral de energia elétrica).

(f) Determine o histograma para a variável “renda domiciliar per capita mensal (em reais)”. A suposição de normalidade para essa variável é razoável? Por que?

(g) Categorize a variável renda. Considere as seguintes classes: até 1 salário mínimo(s.m.), de 1 a 2 s.m., de 2 a 4 s.m. e mais que 4 s.m. . Para essa nova variável, obtenha um gráfico de setores (pizza).

(h) Construa os gráficos *boxplot* para a variável “consumo de gás anual domiciliar per capita (em Kg)” , segundo a variável renda categorizada. Obtenha as medidas resumos por categoria e faça uma análise dos resultados (comente).

(i) Determine o gráfico de probabilidades normal (qqnorm) para variável “renda domiciliar per capita mensal (em reais)”. Interprete.

3) Duas turmas de alunos foram submetidas a um exame e as notas são apresentadas na tabela a abaixo.

Turma		Média	Variância	n
A	5 7 3 5 4 6	5	1.67	6
B	2 5 4 3 6	4	2.00	5

(a) Somente para as notas da turma A, desenhe o gráfico de frequências acumuladas e também o gráfico alisado. Utilize o gráfico para obter o percentil de ordem $p = 0.3$ da variável.

(b) Construa o gráfico de *quantis x quantis* (qqplot) para as notas das turmas A e B.

(c) Como ficaria a função distribuição acumulada para um conjunto de dados com observações repetidas? Considere os dados numa tabela de frequência

X	x_1	x_2	\dots	x_k
Freq.	n_1	n_2	\dots	n_k

Proponha uma maneira geral para obter a função acumulada alisada (função de distribuição empírica F_E). Ilustre o uso desta função através de um exemplo (desenhe a função e calcule alguns percentis).

4) Uma matéria sobre desigualdade de renda no Brasil do Jornal Folha de São Paulo de 13/11/2015, faz a seguinte afirmação: ”‘O índice de Gini (medida de distribuição de renda) do rendimento do trabalho recuou de 0.495 em 2013 para 0.490 em 2014 - quanto mais próximo de zero, mais igualitária é a distribuição da renda no país.’” Explique como o índice de Gini é calculado. Explique também o que é a Curva de Lorenz.

5) A edição de fevereiro de 2017 da revista *Pesquisa FAPESP* apresentou o estudo da professora Carolina Araújo, do Instituto de Filosofia e Ciências Sociais da Universidade Federal do Rio de Janeiro (UFRJ), sobre a situação da mulher na área de Filosofia. Os dados quantitativos mostraram que as mulheres são minoria entre estudantes e docentes dessa área do conhecimento. O texto da revista pode ser acessado em:

<http://revistapesquisa.fapesp.br/2017/02/09/corredor-estreito>

Como seria a presença da mulher na pós-graduação nas áreas de conhecimento do Instituto de Matemática e Estatística (IME) da USP? Para descrever a situação, explore as bases de dados da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), fundação vinculada ao Ministério da Educação (MEC) que atua na pós-graduação. As informações públicas estão disponíveis na página de dados abertos da capes:

<http://dadosabertos.capes.gov.br>

Para fazer o *download*, é necessário realizar um cadastro simples e aguardar o envio de senha por e-mail. Baixe os dados mais atuais sobre docentes e discentes nos programas de pós-graduação do país (lembre-se de registrar o ano dessas informações).

Busque as informações para construir gráficos de barras por sexo, no estilo dos apresentados na revista *Pesquisa FAPESP*, para discentes e docentes permanentes (incluindo os aposentados) dos programas de pós-graduação da USP (áreas de conhecimento: Ciência da Computação, Matemática Pura, Matemática Aplicada e Estatística).

Entrega: 06 de abril de 2017.