

Prova Prática

Pontuação total: 10

Prazo: 23/11/2017 - Definido em sala de aula

| | |
|-------|------------|
| Nome: | Matrícula: |
| Nome: | Matrícula: |
| Nome: | Matrícula: |

Considere os dados gerados pela função geradora de dados ([gerar_dados_v8.R](#) disponível na [página da disciplina](#)), como uma amostra aleatória estratificada de 2000 pessoas do município X.

As variáveis aleatórias de interesse são: **Y1** (medido em *un*), **Y2** (medido em *un*) e **Sexo**. Adicionalmente, assuma que **Y1** e **Y2** não podem assumir valores reais negativos. Os dados são fictícios e tem finalidades exclusivamente didáticas para fins de avaliação prática em análise de dados.

Realizar a análise exploratória dos dados com respostas às seguintes questões:

1 AED: Apresentações tabulares e gráficas (2.5)

1.1 Diagrama de caixa (boxplot) para Y1 e Y2 (1.5)

1. (0.75) Antes e após a eliminação de possíveis outliers¹;
2. (0.75) Após a eliminação de possíveis outliers².

1.2 Y1 (1.0)

1. (0.5) Uma apresentação tabular contendo apenas as frequências: absoluta (F_i), relativa (Fr , %) e acumulada (Fac , %), nessa ordem²;
2. (0.5) Histograma e o polígono de frequência acumulada dos dados².

2 AED: Medidas estatísticas básicas (3.0)

Para as variáveis Y1 e Y2 elaborar apresentações tabulares² contendo as seguintes estimativas:

1. (1.0) Tendência central: Média e mediana;
2. (1.0) Posição: Quartis e decis;
3. (1.0) Dispersão: Amplitude total, variância, desvio padrão e coeficiente de variação.

¹Não distinguindo sexo

²Para cada sexo: M seguido de F

3 AED: Medidas estatísticas de associação e regressão linear (3.5)

3.1 Associação (1.5)

1. (0.5) Estimativas: covariância e correlação linear simples²;
2. (0.5) Diagramas de dispersão dos dados²³;
3. (0.5) Um estudo semelhante foi realizado em um outro município, por outras pessoas. Contudo, as unidades de medida usadas foram: $Y1$ ($100 * un$) e $Y2$ ($100 * un$).

Para comparar associações entre as variáveis de ambos os estudos, qual seria a medida estatística recomendada? Justifique.

3.2 Regressão linear (2.0)

Considere os dados gerados pela função geradora de dados ([gerar_dados_rl_v1.R](#) disponível na [página da disciplina](#)), como uma amostra de um estudo da influência de X sobre Y .

As variáveis aleatórias de interesse são Y (medido em un) e X (medido em un). Os dados são fictícios e tem finalidades exclusivamente didáticas para fins de avaliação prática em análise quantitativa de dados.

1. (0.75) Ajuste aos dados dois modelos de regressão linear: polinômios de grau I e II;
2. (0.75) Apresente um diagrama de dispersão dos dados⁴ com os dois modelos sobrepostos;
3. (0.5) Qual dos dois modelos melhor explica o fenômeno em estudo? Justifique.

4 Contextualização (1.0)

Localize um artigo científico (periódico Qualis A ou B) em área de seu interesse no qual a análise exploratória de dados (AED - possivelmente com medidas de associação e uso de regressão linear como modelo explicativo) teve papel preponderante. Discuta o artigo com ênfase nos recursos da AED usados e também na adequação das normas básicas das apresentações gráficas e tabulares adotada pelo periódico.

Observações:

- Para possibilitar a correção, anexe esta prova devidamente preenchida na primeira página das respostas.
- As normas para apresentações gráficas e tabulares são obrigatórias, serão observadas e corrigidas.
- Sugere-se (mas não é obrigatório) o uso do ambiente R na resolução das questões propostas.
- Cada hora de atraso na entrega da avaliação implica na perda de 25%. Portanto, após 4 horas não entregue.
- Não anexe cópia do artigo ao final da prova, indique apenas o link.

³Considere $Y2$ no eixo das ordenadas e $Y1$ no eixo das abscissas

⁴Considere Y no eixo das ordenadas e X no eixo das abscissas