



LABORATÓRIO 2

Introdução à Regressão Linear Simples

1. Dados contendo medidas morfológicas de 104 gambás das montanhas, que foram capturados através de armadilhas em sete localizações na Inglaterra, estão disponíveis no arquivo “possum_red.csv”. As variáveis contidas nesse arquivo referem-se ao sexo (**sex**), largura do crânio (**skullw**) e comprimento total (**totlngth**) desses animais. Com base nestes dados:
 - (a) Descreva o comportamento de cada uma dessas variáveis. Sumarize as medidas morfológicas segundo o sexo. Apresente *boxplots* para essas variáveis segundo o sexo dos animais.
 - (b) Represente graficamente a distribuição do sexo (utilize gráficos apropriados).
 - (c) Construa um histograma para as variáveis morfológicas (**skullw** e **totlngth**).
 - (d) Discuta a relação entre as variáveis morfológicas através do uso de diagramas de dispersão. Qual a direção da relação? Estime o coeficiente de correlação de Pearson.
 - (e) Parecem haver pontos atípicos nessa relação entre variáveis morfológicas? Caso positivo, identifique-os, retire-os do conjunto de dados e refaça a análise do item (d).
 - (f) O ajuste de um modelo linear de regressão parece adequado para descrição destes dados? Justifique sua resposta. Caso afirmativo, ajuste um modelo linear aos dados após a retirada dos pontos atípicos, se houverem. Interprete as estimativas obtidas.
 - (g) Se a resposta foi afirmativa no item (f), refaça o diagrama de dispersão da letra (e), incluindo no mesmo a reta ajustada pelo modelo definido em (f).
2. Considere dados sobre a eleição presidencial de 2000 nos Estados Unidos. Os dados disponíveis referem-se ao número de votos de cada um dos candidatos por condado no estado da Flórida. Deseja-se investigar a relação entre o número de votos que Bush recebeu em relação ao número de votos recebidos por Buchanan (outro candidato naquela eleição). Os dados estão disponíveis no arquivo “florida.csv”. Com os dados deste estudo:
 - (a) Discuta a relação entre os votos recebidos por Bush e por Buchanan através do uso de diagramas de dispersão. Qual a direção da relação?
 - (b) Existe um debate sobre os votos recebidos por Buchanan que poderiam ter sido de Gore (outro candidato em 2000) se o primeiro não estivesse no pleito. Como Bush e Gore foram os candidatos principais daquela eleição, é de interesse avaliar a relação entre os votos recebidos por Bush e Buchanan na Flórida, que é um Estado importante na corrida presidencial dos EUA. Para isto, ajuste um modelo de regressão linear no qual o número de votos de Bush é usado para prever o número de votos de Buchanan.
 - (c) Parecem haver pontos atípicos nesses dados? Caso positivo, retire-os do conjunto de dados e refaça as análises anteriores. Compare os resultados com aqueles obtidos anteriormente.
 - (d) Inclua no diagrama de dispersão a reta ajustada pelo modelo definido em (c). Se Bush receber 152.846 votos em determinado condado, quantos votos Buchanan espera receber naquela mesma área?

- (e) Programe para obter as estimativas de mínimos quadrados para o intercepto e o coeficiente angular do modelo de regressão linear simples (sem usar a função `lm` ou similares) do item (c) e compare-as com as obtidas com uso da função `lm` do R.

Observações:

- Apresente as conclusões sobre os resultados em forma de relatório (arquivo PDF). O arquivo deve ter texto corrido, sem inclusão de *outputs*.
- O *script* utilizado também deverá ser enviado pelo AVA Moodle.