

Quinta Lista de Exercícios: Regressão Linear

Exercício 1. Na estação de pesquisa Palmer na [Antártica 1](#), pesquisadores fizeram medições em três espécies diferentes de pinguins: Adélie, Chinstrap e Gentoo. Os dados obtidos estão no arquivo **penguins_size.csv**. Vamos inicialmente conhecer o conjunto de dados. Para isso, comece utilizando o comando `str(penguins)`. A partir da saída desse comando, já é possível afirmar que o conjunto de dados possui 344 observações e cada observação possui 7 variáveis, como exemplo, sexo, espécie, ilha onde o pinguim habita, peso em gramas (`body_mass_g`), tamanho da asa em milímetros (`flipper_length_mm`), dentre outras. Agora que já compreendemos um pouco melhor o nosso conjunto de dados, podemos dar início a nossa análise estatística. Os primeiros exercícios referem-se à análise descritiva dos dados. Os últimos exercícios referem-se à análise de regressão.

- (a) Determine o peso médio e o comprimento médio da asa dos pinguins. Se você encontrou algum erro ao utilizar a função `mean()`, talvez seja porque há dados faltantes! utilize o argumento `na.rm = TRUE` para corrigir esse erro (esse argumento exclui os dados faltantes para, em seguida, realizar os cálculos da função). Agora, calcule o desvio padrão de cada uma dessas variáveis (função `sd()`).
- (b) Refaça a parte (a), mas agora calculando a média e o desvio padrão do peso e do comprimento da asa para cada uma das espécies. Comente os resultados encontrados.
- (c) Refaça a parte (a), mas agora calculando a média e o desvio padrão do peso e do comprimento da asa para cada um dos sexos. Comente os resultados encontrados.
- (d) Plote o gráfico de `flipper_length_mm` versus `body_mass_g`.
- (e) Utilize a função `cor()` para calcular a correlação entre as variáveis `flipper_length_mm` e `body_mass_g`.
- (f) A partir das duas respostas anteriores, responda: há alguma relação entre `flipper_length_mm` e `body_mass_g`? Quão forte é essa relação? A relação é positiva ou negativa?
- (g) Utilize a função `lm()` para determinar a reta do modelo de regressão linear simples. Considere `flipper_length_mm` como a variável explanatória (x) e `body_mass_g` como a variável resposta (y).
- (h) Explique o coeficiente angular da reta encontrada em (g).
- (i) A partir do modelo linear encontrado em (g), qual seria o peso médio de um pinguim que possui uma asa de 204 mm? Você poderia utilizar esse modelo para estimar o peso médio de um pinguim que tivesse uma asa de 168 mm? Justifique sua resposta.