

Especialização em Ciências de Dados - Tarefa 10

11 de maio de 2019

Deverá ser entregue um script com os comandos utilizados e um arquivo (pdf) contendo:

- um pequeno relato das etapas realizadas;
 - e as suas conclusões dos questionamento apresentados durante a atividade.
1. Importe o banco de dados `Beisebol.xlsx`. O banco de dados contém as observações das seguintes variáveis, para cada time de beisebol durante uma temporada:
 - `Vitorias`: total de vitórias durante a temporada;
 - `Derrotas`: total de derrotas durante a temporada;
 - `Proporcao_de_vitorias`: proporção de vitórias na temporada;
 - `Runs_Feitos`: total de corridas completas durante a temporada (ataque);
 - `Runs_Permitidos`: total de corridas completas tomadas durante a temporada (defesa);
 - `Saldo` = `Runs_Feitos` - `Runs_Permitidos`.
 2. Considere as variáveis `Proporcao_de_vitorias` e `Runs_Feitos`.
 - (a) Construa o gráfico de dispersão entre as variáveis.
 - (b) Encontre o coeficiente de correlação linear e determine se há evidência suficientes para apoiar uma afirmativa da existência de uma relação linear entre as variáveis.
 - (c) Encontre o ajuste da regressão linear com a variável resposta representando a proporção de vitória e a variável explicativa representando o total de *runs* feitos.
 - (d) Faça a verificação das suposições do modelo (Análise de Resíduos).
 3. Considere as variáveis `Proporcao_de_vitorias` e `Runs_Permitidos`.
 - (a) Construa o gráfico de dispersão entre as variáveis.
 - (b) Encontre o coeficiente de correlação linear e determine se há evidência suficientes para apoiar uma afirmativa da existência de uma relação linear entre as variáveis.
 - (c) Encontre o ajuste da regressão linear com a variável resposta representando a proporção de vitória e a variável explicativa representando o total de *runs* permitidos (tomados).
 - (d) Faça a verificação das suposições do modelo (Análise de Resíduos).
 4. Considere as variáveis `Proporcao_de_vitorias` e `Saldo`.
 - (a) Construa o gráfico de dispersão entre as variáveis.
 - (b) Encontre o coeficiente de correlação linear e determine se há evidência suficientes para apoiar uma afirmativa da existência de uma relação linear entre as variáveis.
 - (c) Encontre o ajuste da regressão linear com a variável resposta representando a proporção de vitória e a variável explicativa representando a diferença entre o total de *runs* feitos e permitidos.
 - (d) Faça a verificação das suposições do modelo (Análise de Resíduos).
 5. Compare os ajustes lineares precedentes.
 - (a) Qual é o melhor ajuste? Explique.
 - (b) Qual parece ser mais eficaz para se ganhar um jogo de beisebol: uma defesa forte ou um ataque forte? Explique.
 6. Encontre uma predição para a proporção de vitórias quando o total de *runs* feitos e permitidos são 844 e 722, respectivamente.