



Iris

Análise do banco de dados Iris do software R.

Introdução

Sendo um dos banco de dados padrão do R, o data set apresenta as medidas em centímetros do comprimento e da largura das partes referentes as sépalas e as pétalas de 3 tipos diferentes de flores, sendo as mesmas, *Iris Setosa*, *Iris Versicolor* e *Iris Virginica*. O banco de dados foi primeira-mente introduzido pelo biólogo e estatístico britânico Ronald Fisher no ano de 1936, também é chamado de “Dataset de Edgar Anderson”, pois o mesmo foi responsável por coletar os dados para quantificar a variação morfológica das flores das 3 espécies.

Para realizar o estudo dos dados serão utilizados métodos de análise de regressão simples, bem como testes de regressão linear, estatísticas descritivas e representações gráficas, visando gerar uma melhor interpretação sobre o conjunto.

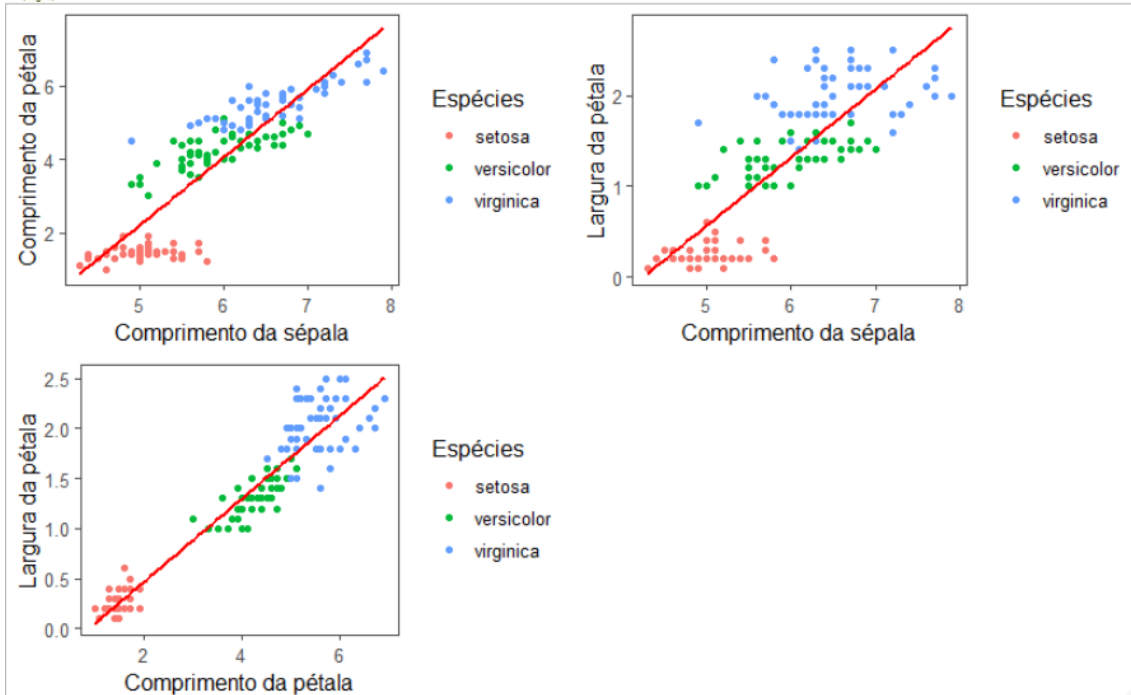
Resultados

A média geral da largura das pétalas foi de 1.2, enquanto a do comprimento apresentou valor 3.8, na análise referente a sépala das flores os valores respectivos foram 3.0 e 5.8. O maior desvio padrão apresentado esteve presente no conjunto do comprimento das pétalas, com valor 1.77, enquanto o restante foi menor que 0.9, além disso as sépalas apresentaram assimetria positiva tanto no comprimento como na largura.

Referente à correlação linear de Pearson, e considerando o nível de confiança dos intervalos como 95%, as maiores correlações crescentes encontradas foram entre a largura das pétalas e seu comprimento com valor de 0.96 (IC 0.95 - 0.97), entre a largura das pétalas e das sépalas com valor de 0.87(IC 0.82-0.90), e entre a largura das pétalas e comprimento das sépalas com valor 0.82 (IC 0.76-0.86)

Discussão

Para o início da discussão é interessante mostrar o seguinte gráfico das regressões lineares previamente apresentadas nos resultados:



Referente ao comprimento da pétala e da sua largura é notável que a reta se ajusta melhor é interessante perceber que o valor máximo do comprimento da pétala de uma setosa no banco não passou de 2 centímetros, e o comportamento no geral dos dados é interessante pois revela que quanto maior o comprimento das pétalas maior sua largura, a mesma interpretação se repete para as outras duas correlações analisadas referentes ao comprimento da pétala e da sépala, bem como entre a largura da pétala e o comprimento da sépala.

Ao analisar separadamente as categorias no gráfico referente ao comprimento da sépala e pétala é possível perceber que as setosas não se ajustam bem à reta de regressão linear, mesmo comportamento que se repete no gráfico de comprimento da sépala e largura da pétala, porém mesmo assim ao analisar os três gráficos essa tendência crescente e o ajustamento da reta se apresenta melhor na espécie versicolor.

Conclusão

A partir do estudo realizado é possível concluir que a correlação linear de Pearson muito forte entre o comprimento e a largura das pétalas dessas 3 espécies se dá de forma crescente, ou seja quanto maior o comprimento da pétala de uma dessas 3 espécies maior sua largura, porém a causa de tal forte correlação não foi revelada no estudo.

Bem como foi revelado que nas outras duas categorias analisadas as setosas não se ajustam muito bem nessa interpretação, já as das flores da espécie virginica não se ajustam muito bem a análise entre o comprimento da sua sépala e a largura de sua pétala, tornando difícil tirar conclusões sobre essas duas espécies presentes no banco ao utilizar o método de regressão linear de Pearson simples.