Regressão Linear

Entre duas ou mais variáveis, podemos desejar ver a relação que existe entre elas e medir sua força. Um exemplo de regressão linear seria medir um plano de saúde. Quanto maior a idade, mais caro ele é. Outro exemplo em altitude, quanto mais alto, mais frio.

À partir daí, poderíamos ter um gráfico de dispersão e prever à partir de novos valores, sendo que esses valores não poderiam fugir da ‘regularidade’ em que ocorrem, visto que prejudicariam o modelo por não corresponder a uma linearidade, cujo nome, a função recebe.

Uma função linear forte pode ser positiva (de baixo para cima) ou negativa (de cima para baixo) num gráfico linear.

Tomando o exemplo do plano de saúde, podemos considerar a idade e o custo como valores que refletem a projeção do gráfico. Em um plano cartesiano, colocamos a Idade no exio X e o custo no eixo Y. Se o custo do plano é menor para pessoas novas e cresce à medida que a idade aumenta, teremos uma projeção linear crescente. O mesmo poderia ocorrer em outra situação: a temperatura sobe, a venda de casacos, caem. Indicando que a projeção é fortemente decrescente. Isso também é uma regressão válida.

O coeficiente de determinação visa explicar o quão a variável dependente (a que desejamos prever) pode ser explicadas pelos valores contidos nas variáveis explanatórias (os valores que temos como base para gerar as previsões).

Um bom modelo de regressão linear não deve percorrer fielmente o a trilha tecida pelos cruzamentos do eixo X e Y em um plano cartesiano. Isso poderia levar a boas respostas em modelos de treino, porém incorreria em sérios erros em modelo de teste. O ideal é que o gráfico tenha um corte transversal na altura média em que os valores são lançados. Pertos ou afastados demais comprometeriam o modelo a ser desenvolvido.

Na construção do modelo é possível encontrarmos registros (pontos) num plano cartesiano que fogem bastante da curva. Esses pontos não devem ter forte interferências sobre os registros predominantes sob pena de comprometer a correlação entre eles. A esses elementos damos o nome de outliers ou extrapolação.

Outliers fogem completamente do padrão.

Extrapolação foge um pouco da tendência da projeção gráfica.

Os cálculos para regressão linear compreendem os seguintes elementos:

Correlação -> Inclinação -> Interceptação -> Previsão

[Arquivo Excel -XLSX] (Cálculos.xlsx)

<sub>Mostro como os cálculos são realizados. Observem as fontes dos valores em cada célula que contém as fórmulas.</sub>