

O que influenciou a sobrevivência dos passageiros do Titanic?

<https://www.kaggle.com/competitions/titanic>

P/ o tratamento de dados:

1. Tratamento de dados ausentes (substituir ausentes por média, mediana ou moda)
2. Transformação de dados categóricos em numéricos (LabelEncoder - scikitlearn)
3. Padronização de dados (Standard Scaler - scikitlearn)

P/ a análise descritiva:

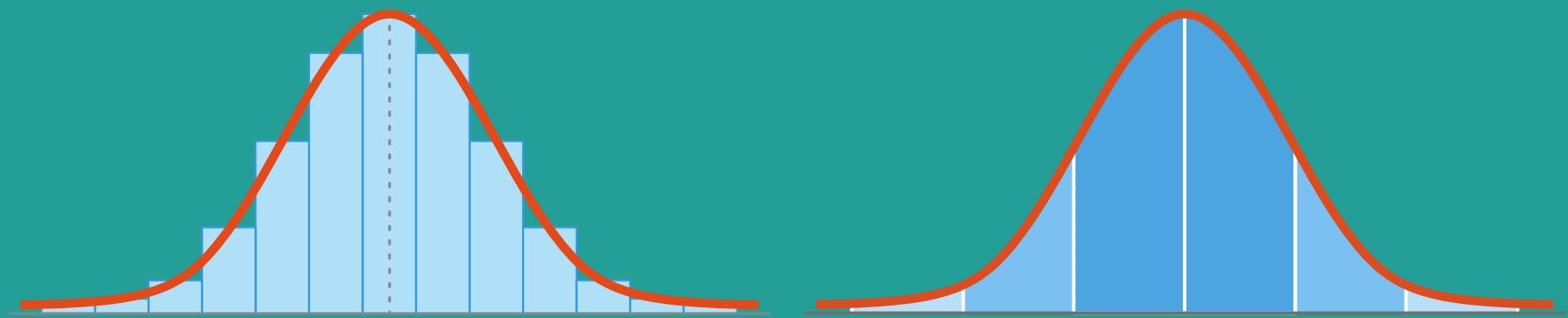
Média; Moda; Mediana; Desvio padrão; frequências.

P/ a visualização de dados:

Histograma – gráfico de distribuição;
Dispersão dos dados, gráfico de pizza, caixas.



Gráfico de distribuição dos dados



Comparar grupos



Gráfico de caixas c/ bigode
(diferença entre grupos, encontrar outliers)

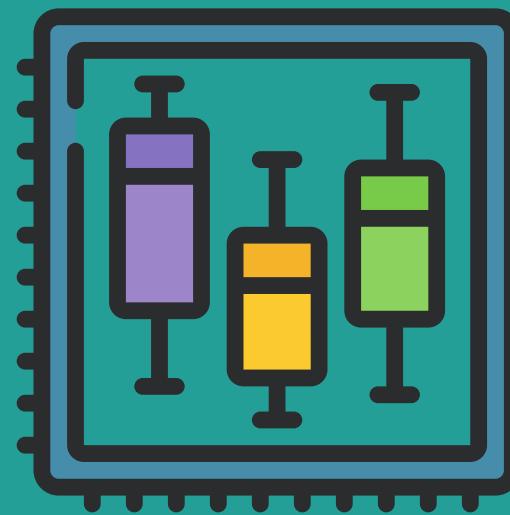


Gráfico de dispersão
(correlação entre variáveis)



Gráficos de frequência relativa
(percentuais)

