



Data Science Academy

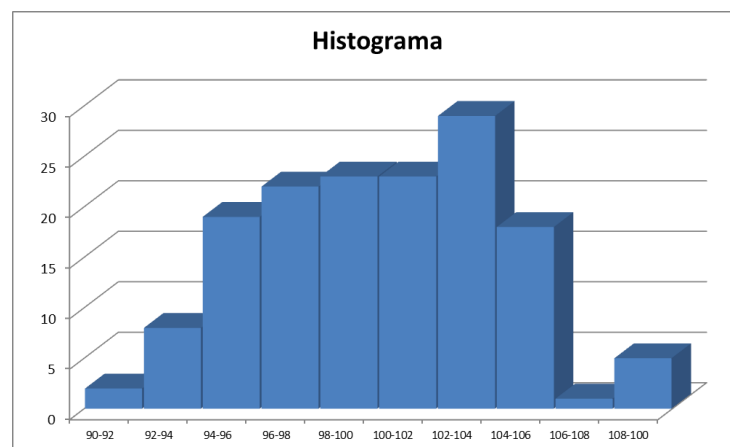
www.datascienceacademy.com.br

Microsoft Power BI Para Data Science

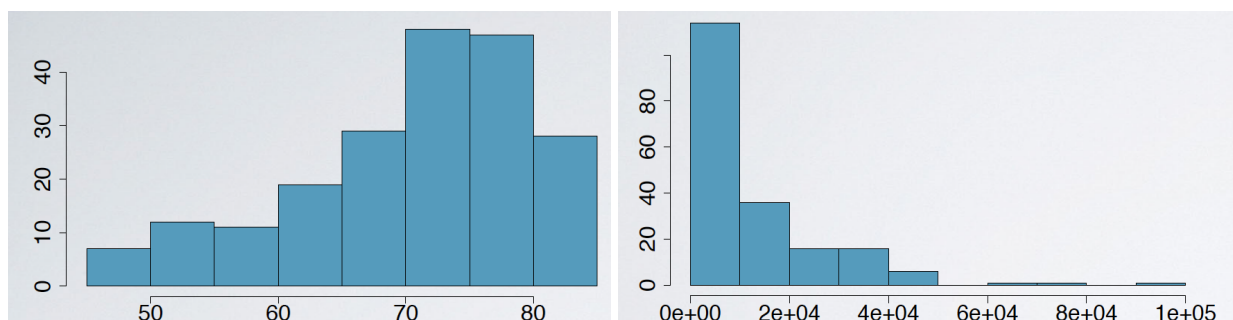
Histograma

O propósito de um histograma, é oferecer uma descrição geral sobre os dados e não sobre os dados individualmente. Um histograma pode resumir as características dos dados numéricos. O histograma se parece com um gráfico de barras, mas possui algumas diferenças. O objetivo é visualizar de que forma os dados se distribuem pelos diversos valores diferentes observados (onde é mais comum, onde é mais raro).

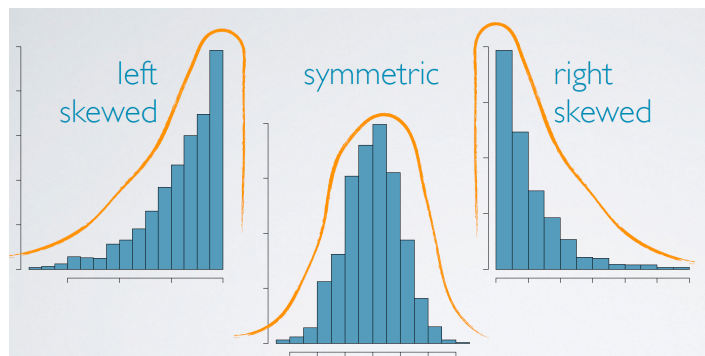
Um histograma pode ser desenhado para uma distribuição de frequências, uma distribuição de frequências relativas ou uma distribuição de porcentagens. Para desenhar um histograma, primeiramente marcamos as classes no eixo horizontal e as frequências (ou frequências relativas, ou porcentagens) no eixo vertical. Depois disso, desenhamos uma barra para cada uma das classes, de maneira que a respectiva altura represente a frequência daquela classe. As barras em um histograma são desenhadas de forma adjacente, uma em relação à outra, sem nenhum espaço entre elas. Um histograma é chamado de histograma de frequências, histograma de frequências relativas ou histograma de porcentagens, dependendo do fato de estarem marcadas no eixo vertical as frequências, as frequências relativas ou as porcentagens.



O histograma é uma ótima forma de visualizar a densidade dos dados...



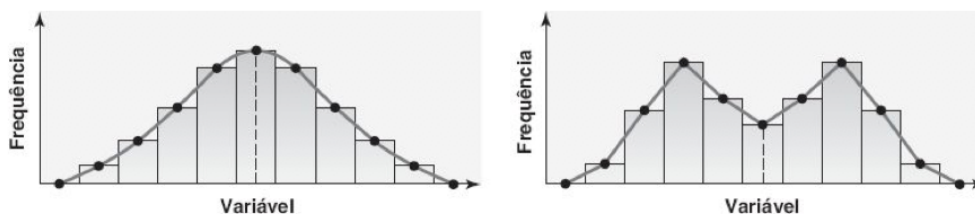
...e muito útil para descrever o formato de uma distribuição.



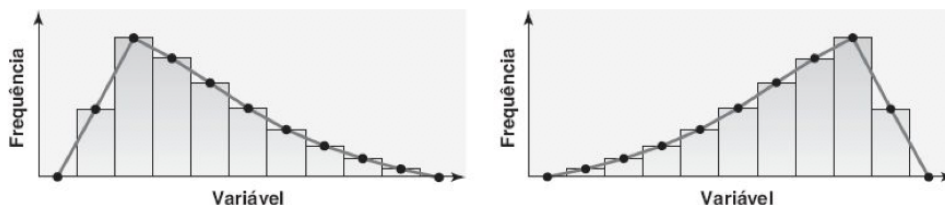
Um histograma pode assumir qualquer formato dentre um grande número de formatos. Os mais comuns entre esses formatos são:

1. Simétrico
2. Assimétrico
3. Uniforme ou retangular

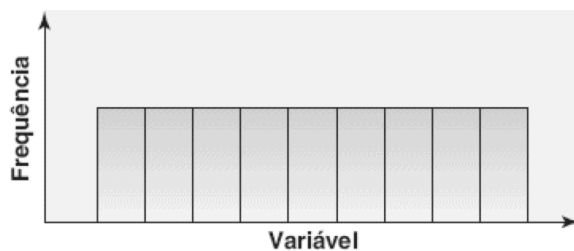
Um histograma simétrico é idêntico em ambos os lados, em relação a seu ponto central. Os histogramas mostrados na figura abaixo são simétricos em torno das linhas tracejadas, que representam seus pontos centrais.



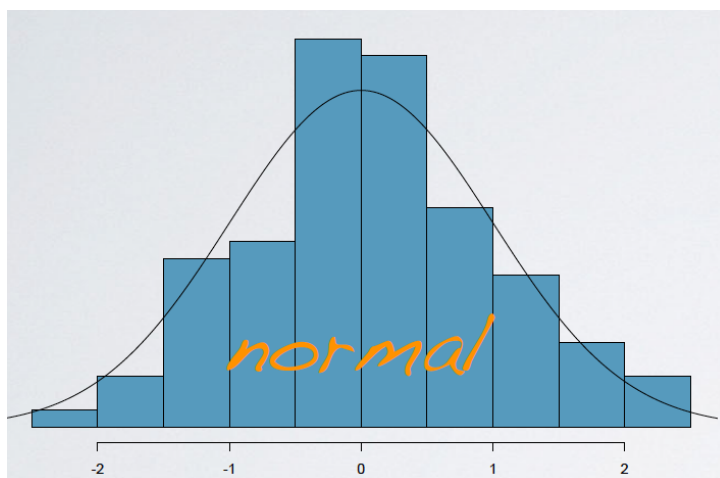
Um histograma assimétrico tem formato não simétrico. Para um histograma assimétrico, a cauda em um dos lados é mais longa do que a cauda do outro lado. Um histograma assimétrico à direita possui uma cauda mais longa em seu lado direito. Um histograma assimétrico à esquerda possui uma cauda mais longa em seu lado esquerdo.



Um histograma retangular ou uniforme possui a mesma frequência para cada classe.



A distribuição normal será a que você usualmente irá trabalhar com mais frequência. Vários algoritmos de Machine Learning requerem que os dados apresentem uma distribuição normal e se não apresentarem, precisaremos converter os dados antes da modelagem preditiva.



Analizando um Histograma


- Basicamente, o histograma pode ser usado para obter três características principais dos dados numéricos: o **modo** como os dados se **distribuem** (simétrico, distorcido para a direita, distorcido para a esquerda), a **variabilidade** encontrada nos dados e onde fica o **centro** dos dados.
- Examine a escala utilizada para o eixo vertical (frequência ou frequência relativa) e tome cuidado com resultados que pareçam exagerados ou subestimados devido ao uso de escalas inadequadas.
- Observe a escala utilizada para os grupos de variáveis numéricas no eixo horizontal. Se a variação para cada grupo for muito pequena, os dados podem parecer mais suaves do que realmente são.

Confira o Curso Visualização de Dados e Design de Dashboards para mais detalhes sobre este tipo de gráfico!

Visualização de Dados e Design de Dashboards

Seja muito bem-vindo à nossa Academia.

Matricule-se já!



<https://www.datascienceacademy.com.br/pages/curso-visualizacao-de-dados>