

## Relatório 5 – Prática: Estatística p/ Aprendizado de Máquina (I)

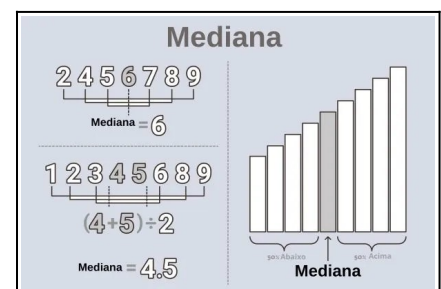
Aluno: Leonardo José Reis Pinto  
RA: 2554097

No card 5, foram mostrados os principais cálculos estatísticos para machine learning, também foi mostrado algumas bibliotecas de visualização gráfica em python tais como matplotlib.

**Média:** é a soma de um conjunto qualquer de valores dividido pela quantidade de valores, gerando assim uma tendência central.

$$\bar{X} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

**Mediana:** É um valor central de um conjunto de dados ordenados, muito utilizada quando tem valores muito discrepantes dentro do conjunto de dados, nesse caso a media não faz sentido, então se usa a mediana.



**Moda:** É o valor que ocorre mais vezes em um conjunto de dados, muito utilizada em dados qualitativos onde não faz sentido usar media e mediana.

**Variância:** a variância é a dispersão dos dados em torno de um valor central, muito utilizada para análise de risco e previsões.

**Desvio padrão:** O desvio padrão é uma medida que indica a dispersão de um conjunto de dados em relação a uma media. Muito utilizado, um exemplo é para ver o desempenho de diversas campanhas de marketing.

**Função de densidade de probabilidade:** a função de densidade de probabilidade é uma ferramenta essencial para modelar e entender como as probabilidades estão distribuídas em variáveis aleatórias contínuas, um exemplo é uma empresa querendo ver como seus salários estão distribuídos na empresa, isso ajuda a identificar padrões de desigualdade salarial.

**Função massa de probabilidade:** Uma função que fala a probabilidade de uma variável aleatória assumir cada um dos seus possíveis valores, um exemplo é entender quais momentos um grupo de clientes da netflix estão propensos em cancelar o plano.

**Distribuição normal:** É uma distribuição de dados simétrico em relação a media, muito utilizado em machine learning para tentar visualizar a distribuição normal dos erros.

**Distribuição binomial:** A distribuição binomial é uma ferramenta fundamental para analisar experimentos com dois resultados possíveis, sucesso ou fracasso. Um exemplo é para analisar qual a probabilidade de um produto de uma fábrica ter defeito.

**Distribuição de poisson:** A distribuição de Poisson é um modelo estatístico utilizado para descrever a probabilidade de um determinado número de eventos ocorrer em um intervalo fixo de tempo ou espaço.

**Percentil:** Ele indica a porcentagem de valores em um conjunto de dados que são menores ou iguais a um determinado valor, um exemplo é um cientista de dados que calcula x salários de uma determinada empresa e tem como resultado que 90 por cento dos salários se concentram em torno de 1500.

**Assimetria:** são dados que não são simétricos em relação a media, existem assimetria positiva e negativa, um exemplo é a distribuição da renda familiar em um país

**Curtose:** A curtose é uma medida que complementa a assimetria na descrição da forma de uma distribuição. Ao analisar a curtose, podemos obter insights sobre a concentração dos dados

**Covariância:** covariância é uma ferramenta fundamental em ciência de dados para entender a relação entre duas variáveis numéricas. Muito utilizada em análise de risco, em resultado de marketing.

**Correlação:** A correlação mostra se duas variáveis tendem a variar juntas de forma consistente. Podendo ser positiva(quando as duas variáveis aumentam ou diminuem juntas), negativa(quando uma aumenta e a outra diminui) e nula.

**Probabilidade Condicional:** A probabilidade condicional é a probabilidade de um evento ocorrer, dado que outro evento já aconteceu.

**Teorema de Bayes:** ajuda a calcular a probabilidade de uma hipótese ser verdadeira, considerando as informações que já temos.

**Conclusão:** Nesse card 5, foi possível absorver muitos tópicos importantes para ciências de dados em geral, com muitos exemplos em python e conseguindo visualizar gráficos de forma simples com matplotlib.