Base Teórica da Estatística Gaussiana

Thalita Carvalho Routh

1 Distribuição Normal (Gaussiana)

A **distribuição normal**, também chamada de **distribuição de Gauss**, é definida por sua função de densidade de probabilidade (PDF - *Probability Density Function*):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Onde:

- x é a variável aleatória contínua;
- μ é a **média** da distribuição;
- σ^2 é a **variância**;
- σ é o **desvio padrão** ($\sigma = \sqrt{\sigma^2}$);
- e é a base do logaritmo natural ($e \approx 2.718$);
- π é a constante matemática ($\pi \approx 3.1416$).

Essa função descreve a famosa **curva em forma de sino**, onde a maioria dos valores está próxima à média, e a probabilidade diminui à medida que nos afastamos dela.

2 Distribuição Normal Padrão (Z-score)

A **distribuição normal padrão** é uma normalização da distribuição normal, onde a média é $\mu = 0$ e o desvio padrão é $\sigma = 1$. Para converter uma variável aleatória x para a forma padronizada z, utilizamos:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

Onde z representa o **Z-score**, que indica o número de desvios padrão que x está afastado da média.

3 Distribuição Normal Multivariada

A **distribuição normal multivariada** é uma generalização da normal univariada para múltiplas variáveis aleatórias. Sua função de densidade de probabilidade é:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right) \tag{3}$$

4 Intervalo de Confiança

O **intervalo de confiança** para a média de uma população normalmente distribuída é dado por:

$$IC = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{4}$$

5 Testes de Hipóteses

5.1 Teste Z

Usado quando a variância populacional é conhecida:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{5}$$

5.2 Teste T de Student

Quando a variância populacional é desconhecida:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{6}$$

6 Correlação e Covariância

6.1 Covariância

Mede a relação linear entre duas variáveis X e Y:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{7}$$

6.2 Correlação de Pearson

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{8}$$

7 Teorema Central do Limite

O **Teorema Central do Limite** afirma que, para um número grande de amostras n, a média das amostras de qualquer distribuição tende a seguir uma distribuição normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{9}$$

8 Conclusão

A **estatística gaussiana** é essencial para modelagem de incertezas e tomada de decisões. Com conceitos como distribuição normal, intervalos de confiança, testes de hipóteses e correlação, podemos analisar dados de forma rigorosa. Seu uso é amplamente aplicado em ciência de dados, aprendizado de máquina, finanças e engenharia.

Explicando de forma mais acessível, Imagine que você está tentando prever a temperatura de amanhã com base nos dias anteriores. Você sabe que a temperatura pode variar, mas também sabe que ela não muda de forma totalmente aleatória. Existem padrões, como as estações do ano, o clima da sua cidade e outras influências. O modelo matemático ajuda a fazer previsões levando em conta esses padrões e incertezas.

Basicamente, ele é um conjunto de números (ou variáveis) que seguem uma distribuição normal, aquela famosa "curva em forma de sino". Esse modelo é muito usado em inteligência artificial, ciência de dados e estatística porque consegue prever eventos futuros considerando informações do passado e da relação entre os dados.