# Repositório de Estatistica e Probabilidade

UFSC Joinville - EMB5010
Artur Gemaque

24 de novembro de 2024

**Resumo**

Este documento tempo como principal funcionalidade registrar os contudos ensinados em sala de aula pelo professor Jaimes, ademais servirá como fonte de estudo para as provas referentes há Matéria.

## 1 Análise Exploratória de Dados

Vamos começar com as definições mais simplórias da matéria que são resultado do cálculo dos dados.

### 1.1 Média

A média aritmética de um conjunto de n números é a soma desses números dividida por n.

$$\bar{x} = \sum_{i=1}^{n} \frac{X_i}{n}$$

### 1.2 Variância

A variância é um conceito fundamental em estatística que mede a dispersão de um conjunto de dados em relação à sua média.

$$(S)^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}$$

### 1.3 Desvio padrão

O desvio padrão é uma medida de dispersão estatística que indica o quão afastados os dados de um conjunto estão da sua média. Ele é a raiz quadrada da variância, o que o torna mais interpretável.

$$S = \sqrt[2]{S^2}$$

### 1.4 Valores de análise

Partindo agora para valores que são obtidos apartir dos dados ordenados em formato crescente. Observação que as seguintes fórmulas serão para que possamos obter as posição dos referidos valores.

### 1.5 Mediana

Mediana é o número no centro de um grupo de números.

$$Md = X_{\left(\frac{n+1}{2}\right)}$$

### 1.6 Quartis

Os quartis são valores que dividem uma amostra de dados em quatro partes iguais e são usados para avaliar a dispersão e a tendência central de um conjunto de dados. Eles são os valores contidos nas posições de $n * 25\%$ entre outras porcentagens, caso "n"dê um valor quebrado o Quartil vai ser a média entre os dois valores.

$$Q_1 = X_{\left(\frac{n+1}{4}\right)}$$
$$Q_3 = X_{\left(\frac{3(n+1)}{4}\right)}$$

## 1.7 Moda

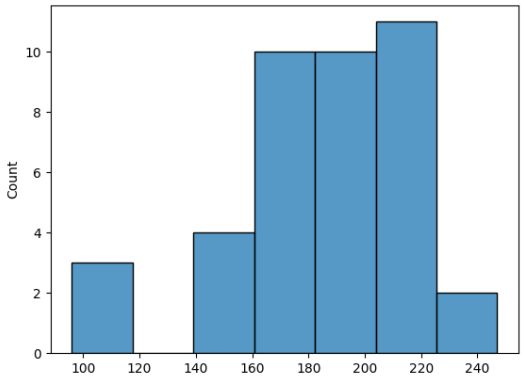
A moda é o valor que mais aparece em um conjunto de dados, ou seja, o valor que tem maior frequência.

## 1.8 Amplitude

A amplitude de um conjunto de dados é a diferença entre o maior e o menor valor. Para calcular a amplitude, subtrai-se o menor valor do maior.

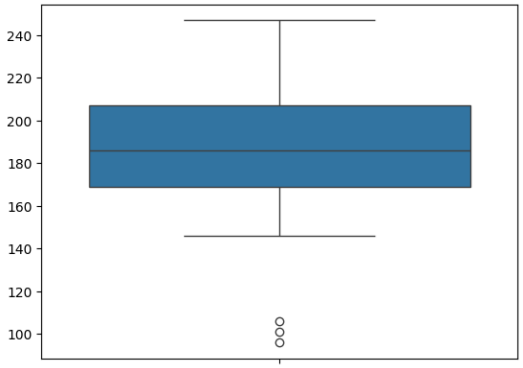
## 1.9 Histograma

Um histograma é uma espécie de gráfico de barras que demonstra uma distribuição de frequências. No histograma, a base de cada uma das barras representa uma classe e a altura representa a quantidade ou frequência absoluta com que o valor de cada classe ocorre.




## 1.10 Boxplot

O Box Plot, que estudamos no curso Green Belt, é uma ferramenta gráfica que ajuda a identificar a existência de possíveis outliers no conjunto de dados. Em um boxplot são apresentadas 5 estatísticas: o mínimo, o primeiro quartil (Q1), a mediana, o terceiro quartil (Q3) e o máximo.



# 2 Probabilidade

## 2.1 Conceitos Introdutórios

O conceito de "Modelo Probabilístico"consiste contém variáveis aleatórias, dadas as entradas não se há certeza das saídas. Em contra partida ao modelo deterministíco que contém dado os valores de entrada se tem certeza dos valores de saída. Exemplo:

- $F = m * a$
- $v = \frac{\Delta x}{\Delta t}$
- $M_f = M_i(1+i)^n$

O "Experimento Aleatório"pode ser entendido como o experimento que pode fornecer diferentes resultados, embora seja repetido da mesma maneira, este por sua vez é feito dentro de um conjuntos de possibilidades que podemos chamar de "Espaço Amostral", ou conjunto de todos os possíveis resultados de um experimento aleatório, podendo ser classificado em "Espaço amostral discreto"quando conjunto finito ou "Espaço amostral contínuo"quando os possíveis resultados representam um intervalo de números reais. Por fim, entendemos por "Evento"um subconjunto do espaço amostral, um resultado ou combinação de resultados do experimento aleatório.

Probabilidade condicional é quando um evento tem uma condição fixa para que aconteça, ou seja, para um dado espaço amostral $\Omega$ em que dois dados sejam lançados aleatoriamente uma probabilidade condicional seria que todos os eventos que buscamos sejam quando a soma do D1 com D2 seja sempre 8, isto representaria uma 5 de 36 possibilidades.

- $P(D1 + D2 = 8) = \frac{N-de-Eventos}{Espaco-Amostral} = \frac{5}{36}$

## 2.2 Axiomas da Probabilidade

Sejam $A_i$ num espaço amostral $\Omega$ :

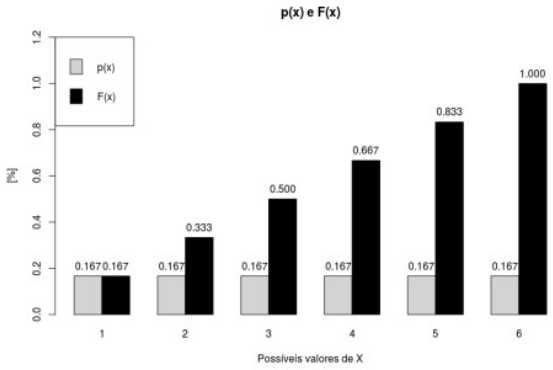
- $0 \leqslant P(A_i) \leqslant 1$
- $P(\Omega) = 1$
- $P(A_1 \bigcup A_2) = P(A_1) + P(A_2)$
  $\because P(A_1 \cap A_2) = 0$

Algumas propriedades decorrentes dos axiomas:

- $P(\varnothing) = 0$
- $P(A) + P(\bar{A}) = 1 \ \rightarrow P(\bar{A}) = 1 - P(A)$
- $P(A_1 \bigcup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

Função probabilidade $P(x)$ e Função probabilidade Acumulada $F(x)$

| $x_i$ | $p(x_i)$ | $F(x_i)$ |
|---|---|---|
| 1 | 1/6 | 1/6 |
| 2 | 1/6 | 2/6 |
| 3 | 1/6 | 3/6 |
| 4 | 1/6 | 4/6 |
| 5 | 1/6 | 5/6 |
| 6 | 1/6 | 6/6 |



# 3 Variáveis Aleatórias Discretas

Dentro dos conceitos definidos anteriormente de VAD (Variáveis Aleatórias Discretas), exitem as principais distribuições discretas como Bernoulli, que compreende a probabilidade de sucesso como $P$ e o fracasso como $1 - P$ além das distribuições Binomial e de Poisson.

## 3.1 Distribuição Binomial

Existem algumas condições para aplicação de tal distribuição

- Os ensaios sejam independentes
- Caso haja somente dois resultados possíveis: Sucesso ou fracasso
- A probabilidade de sucesso (p) seja constante
- X = Número de ensaios que resultam em sucesso
- n = Número de ensaios realizados

Definição como distribuição Binomial:

$$p(x) = \left( \frac{n!}{(n-x)!x!} \cdot p^x \cdot (1-p)^{n-x} \right)$$

Calculo de $\mu$ e $\sigma^2$:

- $\mu = n \cdot p$
- $\sigma^2 = n \cdot p \cdot (1-p)$

## 3.2 Distribuição de Poisson

A variável aleatória X, que é igual ao número de ocorreências no intervalo

$$p(x) = \frac{e^{-\lambda T} \cdot (\lambda T)^x}{x!}$$

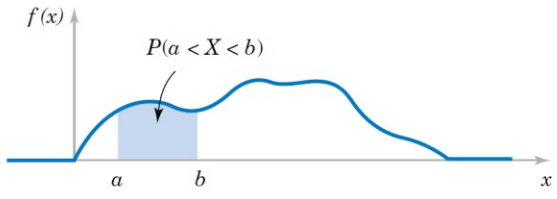
- $x$ = número de ocorreências no intervalor de tamanho $T$
- $\lambda$ = taxa de ocorreências (constante)
- $e$ = número de Euler (constante)

Média e variância

- $\mu = \lambda T$
- $\sigma^2 = \lambda T$

# 4 Variáveis Aleatórias Contínuas

## 4.1 Função de Densidade de Probabilidade




- $f(x) \geqslant 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \geqslant X \geqslant b) = \int_{a}^{b} f(x)dx$

Comparação entre a Média e a Variância

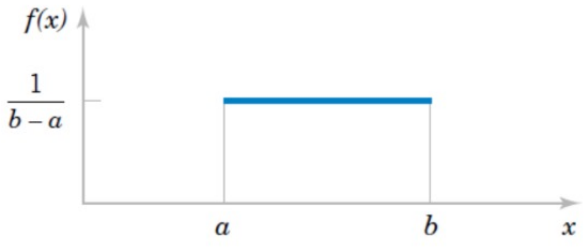
Média e Variância de uma VAD

- $E(x) = \sum_{i=1}^{n} x_i \cdot p_i$
- $V(x) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$

Média e Variância de uma VAC

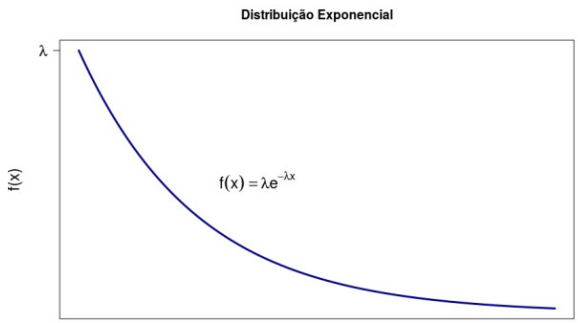
- $E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx$
- $V(x) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx$

## 4.2 Distribuição Uniforme




- $f(x) = \frac{1}{b-a}$
- $\mu = \frac{a+b}{2}$
- $\sigma^2 = \frac{(b-a)^2}{12}$

## 4.3 Distribuição Exponecial




- $f(x) = \lambda e^{-\lambda T}$
- $\mu = \frac{1}{\lambda}$
- $\sigma^2 = \frac{1}{\lambda^2}$

## 4.4 Distribuição Normal

$Z = \frac{X-\mu}{\sigma} \rightarrow$ O resto é tabela Tmj S2

# 5 Estimação de parâmetros

## 5.1 Intervalo de Confiança para $\mu$

Intervalos de confiança estabelecem limites dentro dos quais é altamente provável que se encontre o valor verdadeiro do parâmetro estimado, com um nível de confiaça especificado

- Desvio-padrão populacional conhecido ( ou $n \geqslant 40$):

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- Desvio-padrão populacional desconhecido ( estimado a partir da amostra ):

$$\bar{X} - t_{\frac{\alpha}{2}}, n-1 \frac{S}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + t_{\frac{\alpha}{2}}, n-1 \frac{S}{\sqrt{n}}$$

- Intervalor de confiança para a proporção:

$$IC(p, 1-\alpha) : p = \hat{\mathrm{p}} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\mathrm{p}}(1-\hat{\mathrm{p}})}{n}}$$

- Intervalor de confiança para a variância de uma população normal:

$$IC(\sigma^2, 1-\alpha) : \frac{(n-1)s^2}{X^2{}_{\frac{\alpha}{2}, n-1}} \leqslant \sigma^2 \leqslant \frac{(n-1)s^2}{X^2{}_{1-\frac{\alpha}{2}, n-1}}$$

- Tamanho da amostra para garantir determinado erro

$$n = (\frac{Z_{\frac{\alpha}{2}} \sigma}{E})^2$$

# 6 Introdução ao teste de Hipóteses

Tipos de hipôteses

- Hipótese nula

$$H_0 : \mu = \mu_0$$

- Hipótese alternativa bilateral

$$H_0 : \mu \neq \mu_0$$

- Hipótese alternativa unilateral

$$H_1 : \mu > \mu_0$$
$$H_1 : \mu < \mu_0$$

| | Realidade (desconhecida) | |
|---|---|---|
| | $H_0$ é verdadeira | $H_0$ é falsa |
| Rejeitar $H_0$ | Erro tipo I ($\alpha$) | Acerto |
| Não rejeitar $H_0$ | Acerto | Erro tipo II ($\beta$) |