Webconferência
Estatística e Probabilidade

Prof. Marcio Eisencraft

Estatística e Probabilidade – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

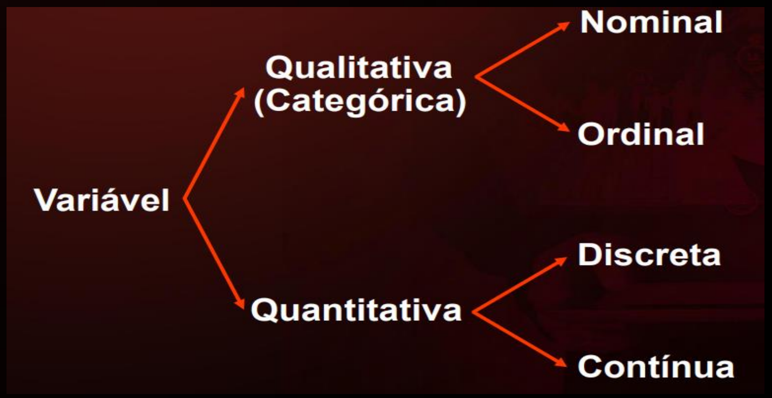
Semana 1: Introdução à estatística descritiva

Estatística descritiva: emprega estatísticas descritivas, como medidas-resumo (média, variância, recordes) e dispositivos gráficos (histogramas, *boxplots*) derivados dos dados para descrevê-los

Estatística inferencial: busca inferir propriedades de uma população ou distribuição de probabilidades a partir dos dados, através de estimativas matemáticas e testes de hipóteses

População: coleção de todos os resultados, respostas, medições ou contagens que são de interesse

Amostra: um subconjunto ou parte de uma população



Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

Semana 2: Organização e apresentação de dados

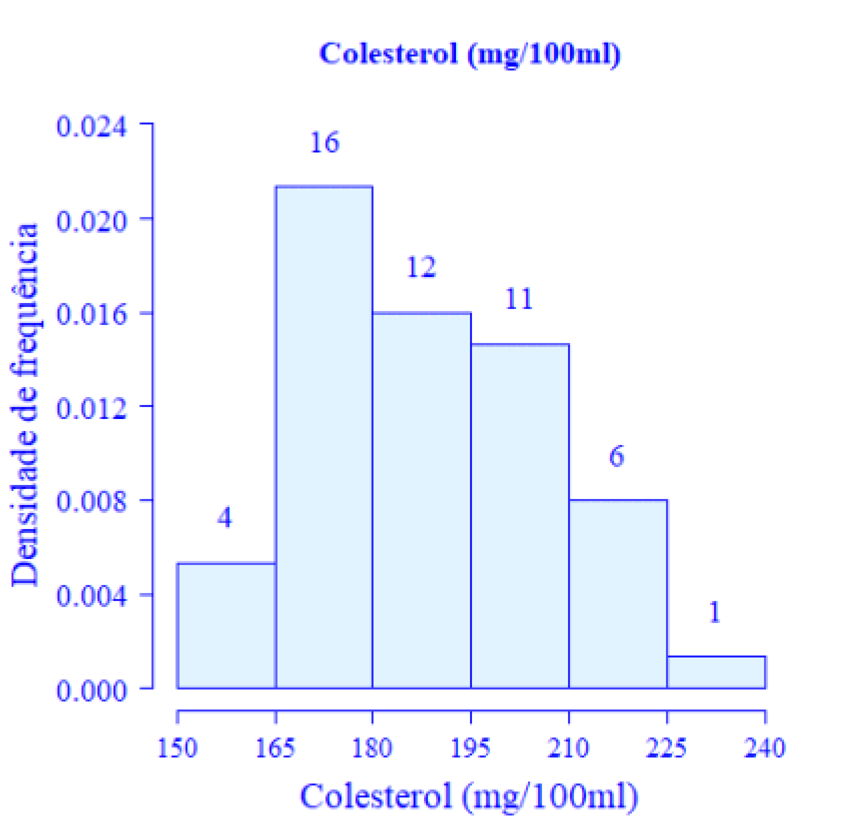
DISTRIBUIÇÃO DE FREQUÊNCIAS

Uma distribuição de frequências é uma tabela na qual os valores dos dados são agrupados em classes ou intervalos juntamente com a contagem do número de incidências em cada classe ou intervalo

Colesterol (mg/100ml)	n_i	$f_i = n_i/n$	$d_i = f_i/\Delta$	$F_i = \Sigma_{j\leq i} f_j$
$150 \vdash 165$	4	0,080	0,0053	0,080
$165 \vdash 180$	16	0,320	0,0213	0,400
$180 \vdash 195$	12	0,240	0,0160	0,640
$195 \vdash 210$	11	0,220	0,0147	0,860
$210 \vdash 225$	6	0,120	0,0080	0,980
$225 \vdash 240$	1	0,020	0,0013	1,000
Total	50	1,000	$\Sigma_i d_i \Delta = 1$	1,000

Regra de Sturge para o número de classes:

$$K = \lfloor 1 + \log_2 n \rfloor = \lfloor 1 + 3{,}322 \log_{10} n \rfloor$$



Semana 2: Organização e apresentação de dados

MEDIDAS DE POSIÇÃO

MÉDIA

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

MEDIANA

Se n é um número ímpar, a mediana vale

$$\text{med}(x) = x_{\left(\frac{n+1}{2}\right)}$$

Se n é um número par, a mediana vale

$$\text{med}(x) = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$$

MODA

- Valor que aparece com mais frequência nos dados.
- Não tem fórmula.
- Nem sempre existe.

Semana 2: Organização e apresentação de dados

MEDIDAS DE DISPERSÃO

DESVIO PADRÃO

- Desvio padrão populacional σ, dado pela raiz quadrada da variância

$$\mathrm{Var}(x) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Desvio padrão amostral s, dado pela raiz quadrada de

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

Repare na diferença de denominador, $1/n$ ou $1/(n-1)$

AMPLITUDE TOTAL

$$R = \max(x) - \min(x) = x_{(n)} - x_{(1)}$$

Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

Semana 3: Probabilidades

ESPAÇO AMOSTRAL

Um espaço amostral Ω é um conjunto que contém todos os possíveis resultados do experimento aleatório

Axioma 1: $P(A) \geq 0$ para todo $A \subseteq \Omega$

Axioma 2: $P(\Omega) = 1$

Axioma 3: $P(A \cup B) = P(A) + P(B)$ se $A \cap B = \emptyset$

FUNÇÃO PROBABILIDADE

Uma função probabilidade $P(\omega)$ é uma função que associa a cada evento elementar $\omega \in \Omega$ um número real

- $P(\emptyset) = 0$
- $0 \leq P(A) \leq 1$ para todo $A \subseteq \Omega$
- $P(\bar{A}) = 1 - P(A)$
- Se $A \subseteq B$ então $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ para dois eventos A e B quaisquer de Ω

Semana 3: Probabilidades

PROBABILIDADE CONDICIONAL

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(A)P(B)$$

dizemos que A e B são eventos independentes

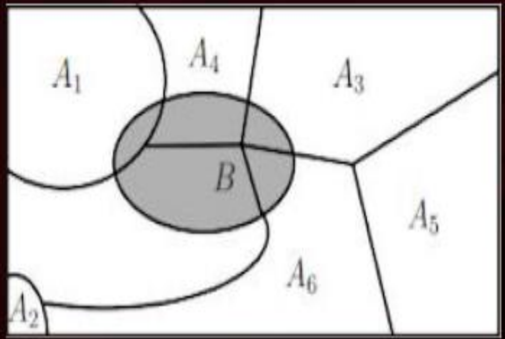
O TEOREMA DE BAYES

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Thomas Bayes

PROBABILIDADE TOTAL

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$



Semana 3: Dica - Cálculos com fatoriais

Suponha que você tenha que calcular $\binom{72}{3} = \frac{72!}{3!(72-3)!} = \frac{72!}{3!(69)!}$.

- Forma ruim ☹ :

$$\frac{72!}{3!(69)!} = \frac{6{,}123\times10^{103}}{6\times1{,}78\times10^{98}} = 59640.$$

- Exige calculadora potente
- Erros de precisão.
- Possibilidade de errar

- Forma bem melhor ☺ :

$$\frac{72!}{3!\,(69)!} = \frac{72\times71\times70\times\cancel{69!}}{6\times\cancel{69!}}$$
$$= 12\times71\times70$$
$$= 59640\ .$$

- Dá para fazer quase de cabeça.
- Sem números gigantescos.
- Melhor fazer assim toda vez que houver divisão de fatoriais.

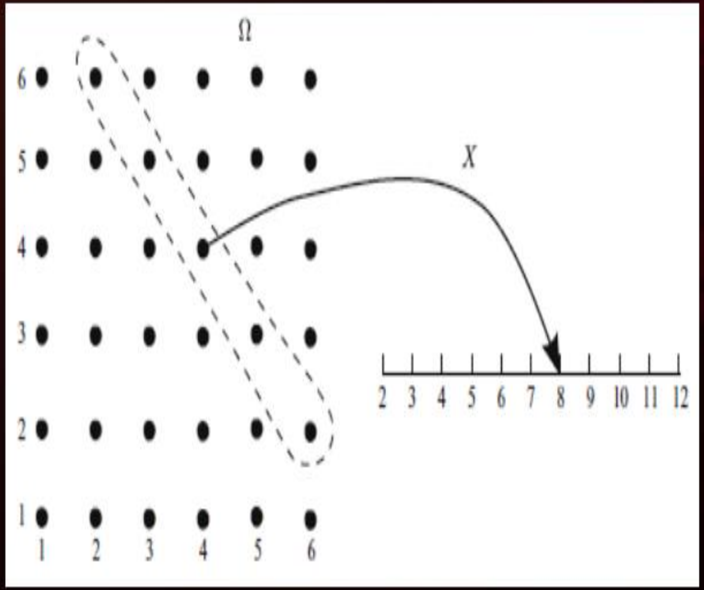
Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

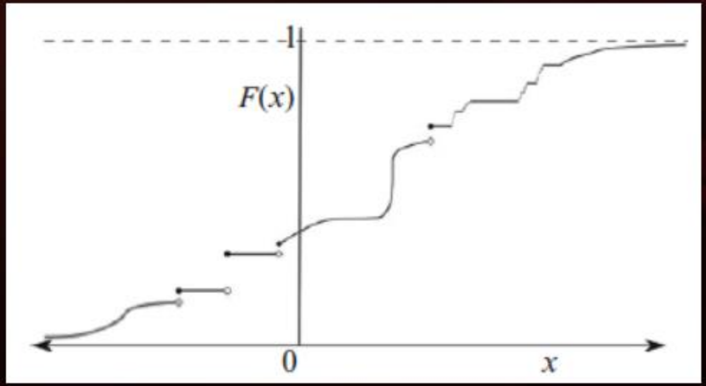
Semana 4: Variáveis aleatórias discretas

Matematicamente, uma variável aleatória é uma função

$X: \Omega \to \mathbb{R}$



▪**Distribuição (cumulativa) de probabilidades**: função $F: \mathbb{R} \to [0,1]$ dada por $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$



FUNÇÃO MASSA DE PROBABILIDADE

$$f_X(x_i) = P(X = x_i)$$

VALOR ESPERADO DE UMA VARIÁVEL ALEATÓRIA

$$E(g(X)) = \sum_{i=1}^{n} g(x_i) P(X = x_i)$$

Semana 4: Variáveis aleatórias discretas

DISTRIBUIÇÃO DE BERNOULLI

Qual é a probabilidade de obter sucesso em um experimento que só pode resultar em "sucesso" ($X = 1$) ou "fracasso" ($X = 0$)? Um experimento desse tipo é conhecido como ensaio de Bernoulli

$$P(X = \text{sucesso}) = p, \quad P(X = \text{fracasso}) = 1 - p$$

$$X \sim \text{Bernoulli}(p)$$

DISTRIBUIÇÃO BINOMIAL

Quais são as chances de obter k sucessos em n ensaios de Bernoulli independentes, cada um com probabilidade p de sucesso?

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad k = 0, 1, \dots, n$$

$$X \sim \text{Bin}(n, p)$$

DISTRIBUIÇÃO DE POISSON

Se um evento ocorre aleatoriamente a uma taxa média de λ eventos por unidade de tempo, qual é a probabilidade de observar k eventos em um período unitário?

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \qquad k = 0, 1, 2, \dots$$

Semana 4: Variáveis aleatórias discretas

DISTRIBUIÇÃO GEOMÉTRICA

Qual é a probabilidade de precisar realizar k ensaios independentes de Bernoulli, cada um com probabilidade de sucesso p, até obter o primeiro sucesso?

$$P(X = k) = (1 - p)^{k-1}p, \qquad k = 0, 1, 2, \ldots$$

$$X \sim \text{Geom}(p)$$

DISTRIBUIÇÃO HIPERGEOMÉTRICA

Em uma população finita de N indivíduos contendo K indivíduos com determinada característica, qual é a probabilidade de observar k indivíduos com a característica em uma amostra de tamanho n obtida sem reposição?

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \max(0, n + K - N) \leq k \leq \min(K, n)$$

$$X \sim \text{Hiper}(N, K, n)$$

Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

Semana 5: Variáveis aleatórias contínuas

Qualquer função $f: \mathbb{R} \to \mathbb{R}^+$ (isto é, $f(x) \geq 0$)) tal que:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

define uma variável aleatória continua, para a qual:

$$P(a \leq X < b) = \int_a^b f(x)dx$$

VALOR ESPERADO DE UMA V.A. CONTÍNUA

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

Semana 5: Variáveis aleatórias contínuas

DISTRIBUIÇÃO UNIFORME

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & x \notin (a,b) \end{cases}$$

$$X \sim \mathrm{U}(a,b)$$

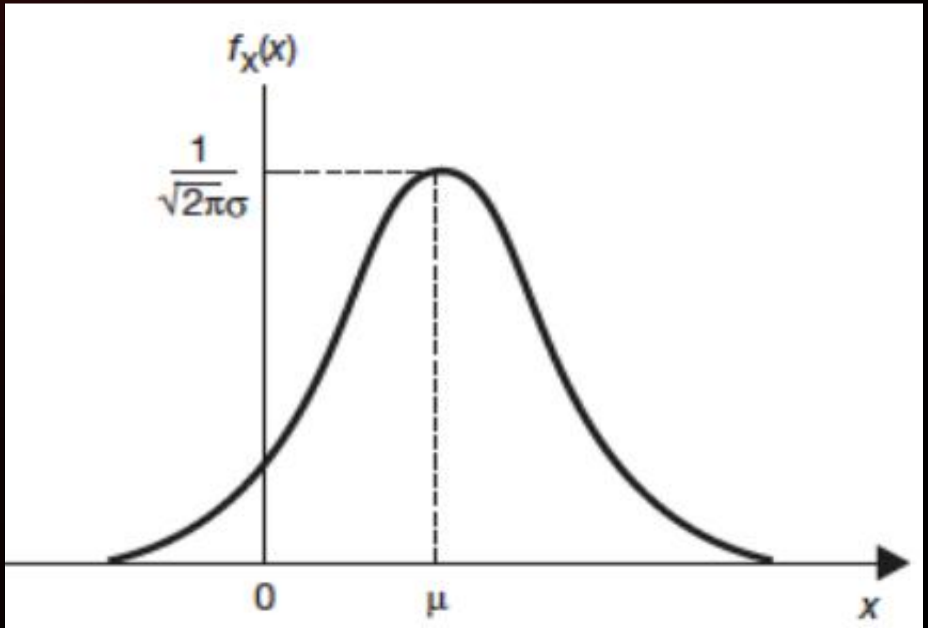
DISTRIBUIÇÃO EXPONENCIAL

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

$$X \sim \mathrm{Exp}(\lambda)$$

DISTRIBUIÇÃO NORMAL (GAUSSIANA)

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



Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

Semana 6: Teorema central do limite e leis dos grandes números

DESIGUALDADE DE MARKOV

$$P(X \geq t) \leq \frac{E[X]}{t}$$

Exemplo: se a média salarial de um país é 3 SM, qual fração da população ganha pelo menos 10 SM?

$$P(X \geq 10) \leq 3/10 = 30\%$$

DESIGUALDADE DE CHEBYSHEV

$$P(|X - E(X)| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}$$

LEI FRACA DOS GRANDES NÚMEROS

Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes e igualmente distribuídas com valor esperado $E(X)$ e variância σ_X^2 finitos → amostra aleatória de tamanho n

Sua média vale:

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

$$\lim_{n\to\infty} P(|\bar{X} - E(X)| \geq \varepsilon) = 0$$

Semana 6: Teorema central do limite e leis dos grandes números

TEOREMA CENTRAL DO LIMITE

O teorema central do limite (TCL) afirma que a soma de n variáveis aleatórias independentes pode ser aproximada por uma distribuição normal independentemente da distribuição particular das n variáveis aleatórias.

O teorema central do limite é, provavelmente, o mais importante teorema da probabilidade e estatística.

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independents, cada uma com valor esperado $E(X_i) = \mu_i$ e variância $\mathrm{Var}(X_i) = \sigma_i^2$ finitos e seja a variável aleatória:

$$Z = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

Então, a variável aleatória Z possui valor esperado $E(Z) = 0$, variância $\mathrm{Var}(Z) = 1$ e no limite de $n \to \infty$

$$Z \sim N(0,1)$$

Conteúdo semanal da disciplina – PES300

- Semana 1: Introdução à estatística descritiva
- Semana 2: Organização e apresentação de dados
- Semana 3: Probabilidades
- Semana 4: Variáveis aleatórias discretas
- Semana 5: Variáveis aleatórias contínuas
- Semana 6: Teorema central do limite e leis dos grandes números
- Semana 7: Introdução à inferência estatística

Semana 7: Introdução à inferência estatística

- O que você vai aprender na Semana 7:

  - Vamos usar todo o aprendizado de VAs para fazer inferências sobre dados amostrais: inferência estatística.

Suponha que temos uma população com determinada característica $X \sim f(x;\theta)$

O problema da inferência estatística seria estimar o valor de tt a partir de uma amostra $X_1, X_2, \ldots, X_n$. Para isso, usamos um estimador $\widehat{\theta} = \widehat{\theta}(X_1, X_2, \ldots, X_n)$

Conclusões e Dicas de Estudo

- Disciplina muito importante para a Computação, para a vida profissional e para o dia a dia.

- Importante entender bem:
  - Calcular medidas de posição e dispersão.
  - Como fazer histogramas.
  - Conceitos e tipos de variáveis aleatórias contínuas e discretas.
  - Desigualdades de Markov e Chebyshev.
  - Fundamentos da inferência estatística.

- Fazer muitos exercícios! ☺

Webconferência
Estatística e Probabilidade

Prof. Marcio Eisencraft