

ESTATÍSTICA : Distribuição de Probabilidade

E TESTES TESTES ESTATÍSTICA

DIA 23/02



Distribuição de Probabilidade

↳ São a base da inferência estatística

↳ As distribuições de probabilidade são modelos matemáticos que ajudam a entender como os dados se comportam

Para variáveis contínuas usamos a **Função de Densidade de Probabilidade (PDF)**, que mostra a chance de um valor específico ocorrer.

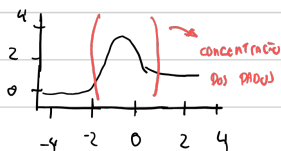
ou, a **Função de Distribuição Acumulada (FDA)**, que indica a probabilidade de a variável assumir um valor menor ou igual a um ponto

• **Distribuição normal**: Descreve fenômenos contínuos, onde a maioria dos valores está concentrada ao redor da média

• **Distribuição Binomial**: Usada para eventos com dois resultados possíveis, como sucesso ou fracasso.

• **Distribuição t de Student**: Aplicada quando temos amostras pequenas e não conhecemos a variação da população.

Distribuição normal

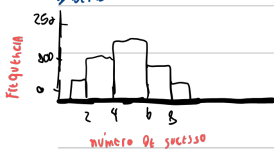


→ É conhecida como "curva em sino", é simétrica em torno da média com maior concentração em torno dela

→ Características:

- Simétrica em torno da média
- Média, mediana e moda são iguais

Distribuição Binomial



→ Descreve números de sucesso em experimentos com dois resultados possíveis (sucesso ou fracasso). É comum em situações como a contagem de vezes em que um evento ocorre em uma série de tentativas

DISTRIBUIÇÃO T STUDENT

HISTOGRAMA T STUDENT



→ UTILIZADA EM AMOSTRAS PEQUENAS E QUANDO A VARIÂNCIA POPULACIONAL É DESCONHECIDA

CARACTERÍSTICAS:

- ADEQUADA PARA AMOSTRAS PEQUENAS
- APRESENTA MAIOR VARIABILIDADE
- AMPLAMENTE USADA EM TESTES DE HIPÓTESE

QUANDO USAR CADA UMA DELAS

- **DISTRIBUIÇÃO NORMAL**: USA QUANDO OS DADOS FOREM CONTÍNUOS E APROXIMADAMENTE EM TORNO DA MÉDIA
 - **Exemplo**: ANALISAR A ALTURA DA POPULAÇÃO ADULTA
- **DISTRIBUIÇÃO BINOMIAL**: APLICAR EM CONTÁVEIS DE SUCESSO, EM EXPERIMENTOS COM DOIS RESULTADOS POSSÍVEIS
 - **Exemplo**: CONTAR O NÚMERO DE VEZES QUE UM DADO HONESTO MOSTRA O NÚMERO 6 EM 20 LANÇAMENTOS
- **DISTRIBUIÇÃO DE T STUDENT**: UTILIZE A O TRABALHO COM AMOSTRAS PEQUENAS E VARIÂNCIA POPULACIONAL DESCONHECIDA
 - **Exemplo**: DETERMINAR SE MÉDIAS DE 35 ALUNOS EM UM TESTE É SIGNIFICATIVAMENTE DIFERENTE DE 70

Passo a Passo

Passo 1: Definir População

↳ 50.000 UNIDADES PESSOAS QUARENTENADAS

Passo 2: Determinar tamanho amostral

$$n = \frac{z^2 \cdot p(1-p)}{e^2} \Rightarrow n = \frac{1.96^2 \cdot p(1-0.05)}{0.05} \Rightarrow n = 382,25$$

Passo 3 \Rightarrow AJUSTE PARA A POPULAÇÃO FINITA

Como a população é finita (90.000), ajustamos o tamanho da amostra

$$n_{ajustado} = \frac{n \cdot n}{n + (n - 1)} = 123 //$$

Distribuição em R

\hookrightarrow INTERVALO DE CONFIANÇA: Indica um intervalo onde a verdadeira média da população provavelmente se encontra com certa segurança estatística

\hookrightarrow `set.seed(123)` Escolhe dados ALEATORIAMENTE

\hookrightarrow `mean(amostra)` Média no R

\hookrightarrow `sd(amostra)` Desvio padrão no R

\hookrightarrow Erro padrão é calculado na seguinte forma

$$\hookrightarrow \text{erro_padrão} \leftarrow \text{Desvio_padrão} / \sqrt{\text{length(amostra)}}$$

\hookrightarrow A função `qt` serve para obter o valor crítico na distribuição T de Student, necessário para calcular

o Intervalo de Confiança

$$t_{\text{crítico}} \leftarrow qt(0.975, df = \text{length(amostra)} - 1)$$

$$\text{Limite inferior} \leftarrow \text{Média amostra} - t_{\text{crítico}} * \text{Erro padrão}$$

$$\text{Limite superior} \leftarrow \text{Média amostra} + t_{\text{crítico}} * \text{Erro padrão}$$



O INTERVALO DE CONFIANÇA ESTÁ ENTRE O LIMITE INFERIOR E O LIMITE SUPERIOR

SIMULAÇÃO DE AMOSTRAGEM

- ↳ Amostragem simples = todos tem a mesma chance de serem escolhidos
- ↳ Amostragem estratificada : é dividido em grupos, e retirado uma parte de cada grupo
- ↳ Amostragem sistemática : escolho um a cada 10, 200, 3000 elemento
- ↳ Amostragem conglomerado : dividido em grupos, e seleciono 1 grupo tirando todos elementos

DELE

- ↳ QUANTO maior a amostra maior, próximos ficamos da população
- ↳ DADO A AMOSTRA, PRECISAMOS INFERIR SOBRE A POPULAÇÃO

TESTES ESTATÍSTICOS

↳ teste A/B

• É um teste bem usado em MARKETING E VENDAS, ONDE GRUPOS fazem WEBSITES DIFERENTES, E TESTAMOS SE A PÁGINA A, PERFORMA MELHOR QUE A PÁGINA B (POR EXEMPLO: PÁGINA A, VENDE MAIS QUE A B)

↳ MÉTODO FISHER, NEYMAN-PERSON

.

BACKTITE E BOOSTING



Resposta e incluir AcimaAAA

Teste De Hipótese

↳ É USADA QUANDO TEMOS DUAS TEORIAS

EXISTEM DOIS TIPOS DE TESTES DE HIPÓTESE

- BILATERAL: USADO QUANDO VOCÊ QUER TESTAR SE ALGUM PARÂMETRO É DIFERENTE DE ALGUM VALOR ESPECÍFICO
- UNILATERAL: É USADO QUANDO VOCÊ QUER TESTAR SE ALGUM PARÂMETRO É MAIOR OU MENOR DO QUE ALGUM

VALOR ESPECÍFICO

Passo 1: DETERMINAR QUAIS HIPÓTESES QUERO TESTAR.

Passo 2: DECIDIR QUAIS AS HIPÓTESES SERÃO USADAS PARA TESTAR AS HIPÓTESES.

μ : MÉDIA AMOSTRAL OU μ : MÉDIA POPULACIONAL

Passo 3: FIXAR OS ERROS BÁSICOS

	NÃO REJEITAR H_0	REJEITAR H_0
H_0 VERDADEIRO	SEM ERRO	α ERRO TIPO I
H_0 FALSA	β ERRO TIPO II	SEM ERRO

- O ERRO MAIS BAIXO É O ERRO DO TIPO I
- POIS É A MUDA DE OPINIÃO E CRENÇAS MILHÃO DE PESSOAS