Estatística Básica

Tipos

Tipos de Variáveis Estatísticas

**População**: É o conjunto completo de elementos ou indivíduos que queremos estudar. Por exemplo, todas as pessoas de uma cidade.

**Amostragem**: É um grupo menor selecionado da população, usado para representar o todo. Por exemplo, 100 pessoas escolhidas para uma pesquisa na cidade.

**Classificação das Variáveis**

Na estatística, há dois tipos principais de variáveis:

1. **Qualitativas**: Descrevem características ou categorias.
   * **Nominais**: Sem ordem (ex.: cor dos olhos).
   * **Ordinais**: Com ordem (ex.: nível de escolaridade).
2. **Quantitativas**: Representam números ou valores.
   * **Discretas**: Valores inteiros (ex.: número de filhos).
   * **Contínuas**: Qualquer valor dentro de um intervalo (ex.: altura em metros).

**Amostragem**

**Amostragem** é o processo de selecionar uma parte da população para representar o todo em um estudo.

**Tipos de amostragem:**

1. **Probabilística**: Todos têm chance conhecida de ser escolhidos.
   * **Aleatória Simples**: Sorteio puro.
   * **Sistemática**: Seleciona a cada intervalo fixo (ex.: a cada 10 pessoas).
   * **Estratificada**: Divide em grupos (estratos) e sorteia de cada.
   * **Por Conglomerados**: Sorteia grupos inteiros (ex.: bairros).
2. **Não Probabilística**: Não usa sorteio, a escolha é intencional.
   * **Conveniência**: Seleciona os mais acessíveis.
   * **Intencional**: Escolhe com base em critérios específicos.
   * **Por Quotas**: Seleciona seguindo proporções definidas da população.

**Distribuição de frequências**

* **Frequência**

É o número de vezes que um valor ou intervalo aparece em um conjunto de dados. Por exemplo, se "azul" foi escolhido 5 vezes, sua frequência é 5.

* **Frequência absoluta**

É o valor exato de vezes que um dado ocorre, sem transformações. Exemplo: "Azul" apareceu 5 vezes.

* **Classe**:

É um intervalo de valores usados para agrupar dados contínuos ou extensos. Exemplo: Idades agrupadas em classes como 10-20, 21-30.

* **Frequência relativa percentual**:

É a proporção de uma frequência em relação ao total, expressa em porcentagem. Fórmula:

Exemplo: Se "azul" foi escolhido 5 vezes em um total de 20 respostas, sua frequência relativa percentual é (5/20) \* 100 = 25%

* **Largura da classe**

É a diferença entre o limite superior e o limite inferior de uma classe (intervalo). Exemplo: Na classe "10-20", a largura é 20−10=1020 - 10 = 1020−10=10. Ela mostra o tamanho de cada intervalo.

* **Amplitude total**

É a diferença entre o maior e o menor valor do conjunto de dados. Exemplo: Se os dados variam de 5 a 45, a amplitude total é 45−5= 40

**Por que servem?**

* **Largura da classe**: Ajuda a organizar os dados em intervalos iguais, facilitando análises e construção de tabelas ou gráficos.
* **Amplitude total**: Define o alcance dos dados e é usada para calcular o número de classes e a largura delas.

**Medidas de Tendência Central**

* Média

É a soma de todos os valores dividida pelo número total de valores. É usada para encontrar o "valor médio" de um conjunto de números.  
Exemplo: Para os números 2, 4 e 6, a média é (2 + 4 + 6) ÷ 3 = 4.

* Mediana

É o valor central de um conjunto de dados ordenado. Se o número de valores for ímpar, é o do meio; se for par, é a média dos dois centrais.  
Exemplo:

* Para 1, 3, 7, a mediana é 3.
* Para 2, 4, 6, 8, a mediana é (4 + 6) ÷ 2 = 5.
* Moda

É o valor que mais se repete em um conjunto de dados. Pode haver mais de uma moda ou nenhuma (se todos os valores forem diferentes).  
Exemplo: Para 2, 3, 3, 4, 5, a moda é 3.

**Por que usar as três medidas?**

* A **média** pode ser distorcida por Valores muito altos ou baixos.
* A **mediana** dá uma visão mais equilibrada, pois não é afetada por outliers.
* A **moda** mostra o valor mais frequente, útil para entender o valor padrão.

**Análise de Outliers**

Um outlier é um valor que foge muito do padrão dos outros dados em um conjunto. Ele pode ser muito maior ou muito menor que a maioria dos valores e pode indicar algo inesperado, como erros, eventos raros ou dados excepcionais.

**Por exemplo:** em uma turma com idades entre 20 e 30 anos, alguém com 80 anos seria um outlier.

**Probabilidade**

Probabilidade é a chance de algo acontecer. É um número entre 0 e 1, onde 0 significa impossível e 1 significa certeza. Por exemplo, ao jogar uma moeda, a probabilidade de dar cara é 50%, ou 0,5, porque há duas opções igualmente possíveis: cara ou coroa.

Principais Termos da probabilidade:

**1. Espaço Amostral (S):**

* Conjunto de **todos os resultados possíveis** de um experimento.
* Exemplo: Jogar um dado → S= {1,2,3,4,5,6}

**2. Experimento Determinístico:**

* Experimento cujo **resultado é sempre o mesmo**.
* Exemplo: Apertar um interruptor → ou a luz acende, ou não.

**3. Experimento Aleatório:**

* Experimento cujo **resultado não pode ser previsto com certeza**.
* Exemplo: Jogar uma moeda → o resultado pode ser **caro** ou **coroa**.

**4. Evento:**

* **Subconjunto do espaço amostral**, ou seja, um resultado ou grupo de resultados que queremos observar.
* Exemplo: No experimento de jogar um dado, o evento "sair um número par" é:  
  A = {2,4,6}

A **probabilidade de não ocorrer um evento (A)** é chamada de **evento complementar (Aᶜ)**. É calculada como:

Uma imagem contendo objeto, relógio

Descrição gerada automaticamente

Ou seja, basta subtrair a probabilidade de A acontecer do total (1).

**Exemplo:**  
Qual a probabilidade de cair de obter a soma 11 jogando dois dados

Conjunto = {5, 6} , {6, 5}

A chance de acontecer se torna então 2/36 = 0,05

Logo, a chance de não acontecer é igual a: 1 – 0,05 = 95,5%

**Métricas de desempenho**

**Erro** -> Erro

**Versionamento**

|  |  |  |
| --- | --- | --- |
| **Data Alteração** | **Versão** | **Detalhamento Alteração** |
| **19/01/2025** | **01** | * **Conteúdo de probabilidade** |
|  |  |  |