

CIÊNCIA DE DADOS (BIG DATA)

ANÁLISE ESTATÍSTICA

Professor curador: Mário Olímpio de Menezes



Mackenzie



TRILHA 1

PARTE A – INTRODUÇÃO À ANÁLISE ESTATÍSTICA

ANÁLISE ESTATÍSTICA

PARTE A – INTRODUÇÃO À ANÁLISE ESTATÍSTICA



Mackenzie



ESTATÍSTICA vs. PROBABILIDADE

- Estatística é a disciplina científica que provê métodos que nos ajudam a dar sentido aos dados.
- Probabilidade é a área da teoria estatística responsável por lidar com aleatoriedade e incerteza.
 - Probabilidade é uma medida da possibilidade de ocorrência de um evento.

ESTATÍSTICA vs. PROBABILIDADE

Probabilidade lida com a previsão da probabilidade de eventos futuros, enquanto a Estatística envolve a análise de frequência de eventos passados.

“Sem dados, você é simplesmente mais uma pessoa com uma opinião”

VARIABILIDADE

Precisamos entender a variabilidade para:

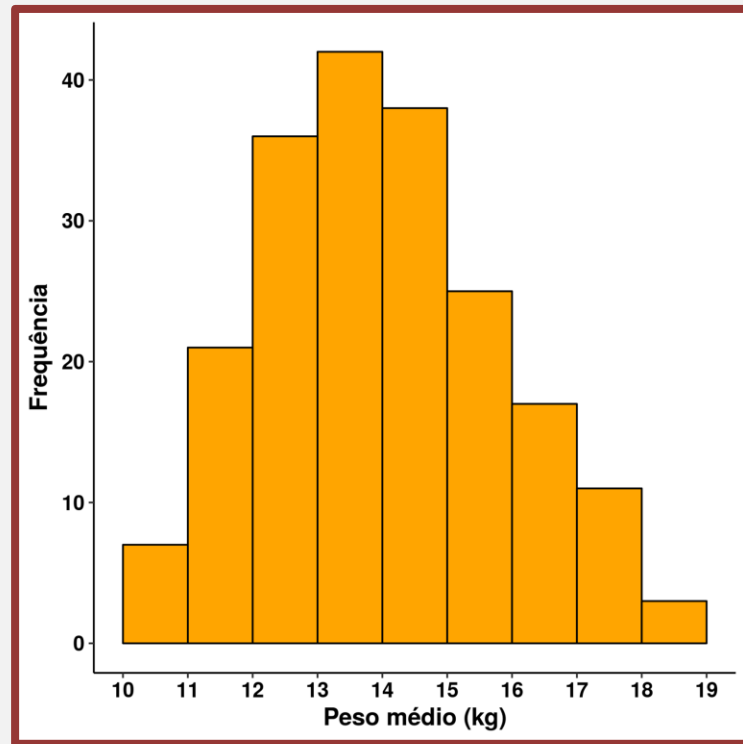
- Coletar.
- Descrever.
- Analisar.
- Derivar conclusões dos dados de modo sensato.

Se todas as medidas fossem idênticas para todos os indivíduos, a análise estatística seria desnecessária e chata!

**Entender a variabilidade nos
permite distinguir entre
eventos usuais e não usuais!**

VARIABILIDADE

- Gráfico ao lado apresenta distribuição do peso médio diário (200 dias) de 5 espécimes de um criadouro de peixes.
- Durante 10 dias, a quantidade de ração foi aumentada por descuido.
- Um mês após este evento, 5 espécimes apresentaram peso médio de 15,5kg.
- Considerando a variação no peso antes do evento, será este valor uma evidência de que o peso dos peixes aumentou?



ETAPAS DO PROCESSO DE ANÁLISE DE DADOS

1. Entendimento da natureza do problema

- Objetivo, questão a responder, direção a seguir.

2. Decidindo o que medir e como medir

- Que informação precisamos para responder à questão de pesquisa?

ETAPAS DO PROCESSO DE ANÁLISE DE DADOS

3. Coleta de Dados

- Fonte dos dados, planejamento e metodologia de coleta; identificação da população alvo e seleção da amostra.

4. Análise Exploratória

- Resumos numéricos, tabulações e gráficos.

ETAPAS DO PROCESSO DE ANÁLISE DE DADOS

5. Análise Formal

- Seleção dos métodos apropriados, tipos de modelos de regressão, classificação etc.

6. Interpretação dos Resultados

- O que aprendemos com os dados? Que conclusões podem ser tomadas? Respondemos às questões de pesquisa?

TIPOS DE DADOS

Características dos indivíduos de qualquer população podem ser estudadas.

Estas características variam de indivíduo para indivíduo!

- Esta variação nos permite utilizar a estatística para estudá-los.

Chamamos de **variáveis** ou **atributos** as características dos indivíduos

CONJUNTO DE

Variáveis

Núm Registro	Atributo 1	Atributo 2	Atributo 3	Atributo 4
1	Sim	Solteiro	Valor	Casa
		Casado	150.000,00	Apartamento

Observações

TIPOS DE ATRIBUTOS

Os atributos ou as variáveis podem ser classificados em:

- Categóricos (ou qualitativos) se os seus valores são categorias ou classes.
- Numéricos (ou quantitativos) se os seus valores são números.

VARIÁVEL DISCRETA vs. VARIÁVEL CONTÍNUA

- Variável numérica **discreta** – valores possíveis correspondem a pontos isolados em uma reta numerada.
- Variável numérica **contínua** – valores possíveis formam um intervalo completo na reta numerada.







ESCALA DE MENSURAÇÃO

Os atributos são classificados também com relação à sua escala, isto é, à forma como medem o que medem.

- **Nominal:** cor dos olhos, CEP, nomes de bairros etc.
- **Ordinal:** ordem de preferência de um produto, nota de alunos (A, B, C...), estatura (alto, médio, baixo).
- **Intervalar:** datas de calendário, temperaturas °C ou °F.
- **Razão:** comprimento, tempo, massa, contagens etc.

ESCALA DE MENSURAÇÃO

A escala de mensuração implica também, para cada atributo, quais propriedades e quais operações são permitidas:

- Nominal  
- Ordinal  
- Intervalar  
- Razão 