

# Princípios básicos da Análise Exploratória de Dados

## UFCD: 10807

CARGA HORÁRIA: 25 horas

ALEXANDRA CAMPOS VIDAL DE SOUZA  
FORMADORA



# Conteúdos

## Objetivos e tarefas da estatística e análise de dados

### Mentiras, ambiguidades e más utilizações da estatística

Variáveis estatísticas e tipos de dados

Métodos de análise de dados para uma variável

Distribuições de frequências e histogramas

Medidas de tendência central

Medidas de dispersão

Medidas de localização

Outros indicadores

Métodos de análise de dados para duas variáveis

Covariância e correlação

R de Spearman

Medidas de concentração

Números índice

Princípios básicos sobre probabilidades

Princípios básicos sobre amostragem e metodologia de recolha de dados

Princípios da análise de componentes principais

Projeto de análise exploratória de dados



# Mentiras e Ambiguidades na Estatística

Concordas com essa afirmação?

A estatística é uma ferramenta poderosa para entender e comunicar informações, mas, quando mal utilizada ou interpretada de forma incorreta, pode levar a **mentiras, ambiguidades e más utilizações**



# Por Que a Estatística é Tão Mal Utilizada?

## Falta de Conhecimento

Muitas pessoas não entendem os princípios básicos da estatística, levando a interpretações errôneas.

## Intenção Maliciosa

Alguns indivíduos usam a estatística para enganar ou manipular outros, geralmente para fins de lucro ou influência.



# Detectando Erros, Exageros e Mentirinhas Leves

- **Erros honestos** – Podem ocorrer devido a enganos não intencionais, algo que pode acontecer com qualquer pessoa.
- **Exageros e ajustes** – Muitas vezes, estatísticas são manipuladas para favorecer um ponto de vista, seja aumentando valores ou distorcendo a forma como os dados são apresentados.



# Detectando Erros, Exageros e Mentirinhas Leves

- **Dados fabricados** – O pior cenário envolve números completamente inventados, que não podem ser reproduzidos e nunca existiram de fato.



# Detectando Erros, Exageros e Mentirinhas Leves

- **Questionar a estatística** – Nunca aceite um número sem questioná-lo. Muitos erros podem ocorrer na coleta, resumo e interpretação dos dados.
- **Erros aritméticos** – Verifique se as somas estão corretas, especialmente em gráficos de pizza e tabelas. A falta de coerência numérica pode indicar erros.



# Exemplo 1

O **USA Today** relatou o resultado de uma pesquisa de opinião feita pela Tupperware com relação ao hábito de aquecer as sobras de alimento no forno microondas. A notícia dizia que **28% das pessoas entrevistadas disseram que aqueciam diariamente** as sobras de alimentos no forno micro-ondas, **43% disseram que usavam de duas a quatro vezes por semana** o forno microondas para aquecer as sobras de alimentos e **15% disseram fazer isso apenas uma vez por semana**. Considerando que todos entrevistados deveriam se enquadrar em uma dessas categorias, a soma das porcentagens deveria ser de 100% ou o mais próximo possível.

- **A soma de 28% + 43% + 15% = 86%. não bate!**





# Detectando Erros, Exageros e Mentirinhas Leves

- **Erro por omissão** – Dados podem ser apresentados sem informações essenciais, dificultando a interpretação correta dos números.
- **Falta de informação sobre amostras** – Quando uma pesquisa não informa quantos participantes foram entrevistados, sua confiabilidade fica comprometida.

# Detectando Erros, Exageros e Mentirinhas Leves

- **Amostras pequenas** – Estatísticas baseadas em poucos entrevistados podem ser enganosas e não representar a realidade.
- **Falta de totalização** – Sempre verifique se a soma das categorias é próxima de 100%, pois discrepâncias podem indicar falhas na estatística.

# Case Trident

Comercial do Trident que dizia que **“Quatro em cada cinco dentistas recomendam Trident para seus pacientes”**. Esse comercial é um pouco antigo, mas, recentemente, ele foi lembrado por uma série de comerciais divertidos em que se perguntava o que teria acontecido com o quinto dentista e, depois, mostrava alguns incidentes que poderiam ter ocorrido, impedindo que ele ou ela apertasse o botão do “sim”. Mas, aqui vai a verdadeira pergunta: Quantos dentistas realmente foram entrevistados? Você não sabe, pois a pesquisa não informou.





Os gráficos são ótimas formas de se mostrar de maneira clara e rápida o ponto em que você quer chegar, desde que sejam feitos de maneira correta e honesta.

Mas cuidado!!



## Exemplo 3: Número de Vezes que cada Número foi Sorteado (Loteria Pick 3 de Kansas em 15/03/97)

- A loteria do estado de Kansas frequentemente mostra os últimos resultados da Loteria Pick 3. Uma das estatísticas demonstrada é o número de vezes que cada número (de 0 a 9) é sorteado entre os três números vencedores. A tabela 2-3 mostra o número de vezes que cada número foi sorteado em 15 de março de 1997 (durante 1.613 jogos da loteria Pick 3, para um total de **4.839 números sorteados**). Dependendo de como você escolhe enxergar esses resultados, mais uma vez você poderá fazer com que a estatística conte histórias muito diferentes.



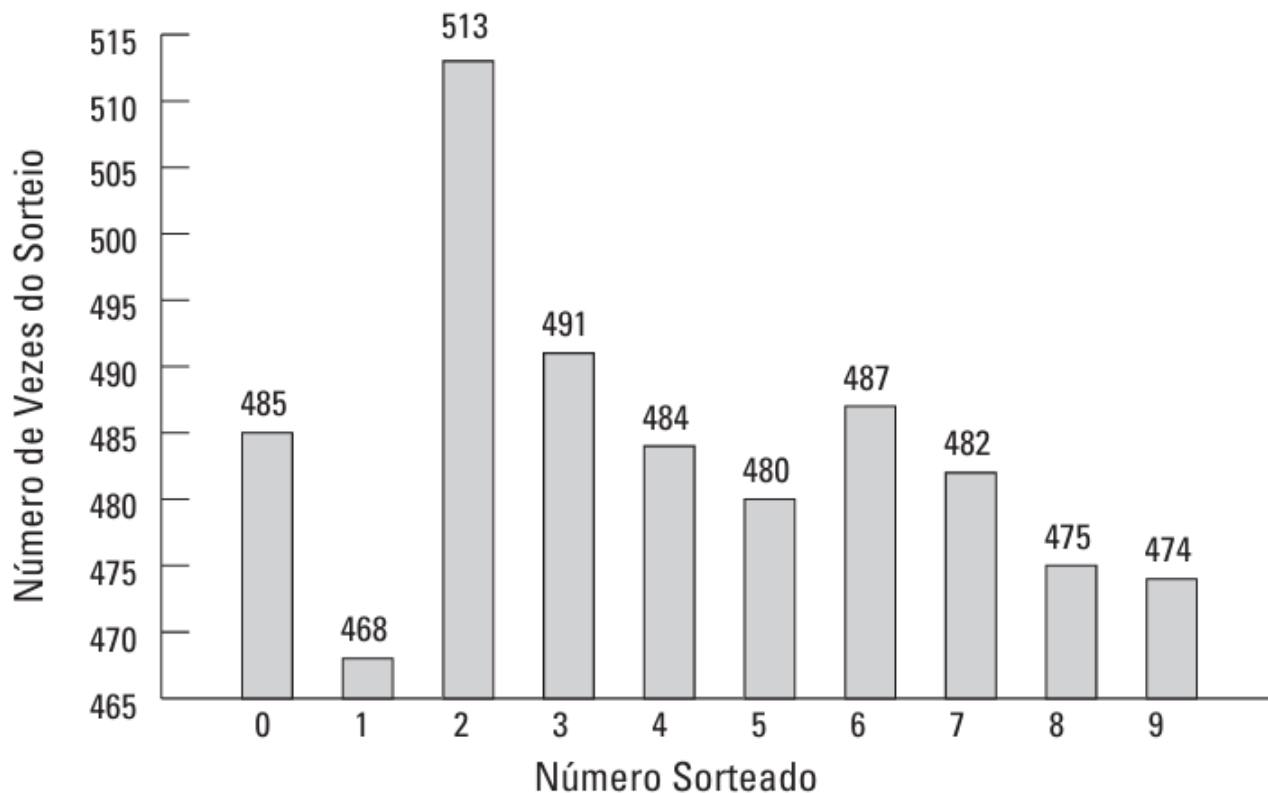
# Tabela mostra o numero de vezes que o numero foi sorteado.

<i>Número Sorteado</i>	<i>Número de Vezes do Sorteio</i>
0	485
1	468
2	513
3	491
4	484
5	480
6	487
7	482
8	475
9	474



## Exemplo 3: Número de Vezes que cada Número foi Sorteado (Loteria Pick 3 de Kansas em 15/03/97)

**Figura 2-1:**  
Gráfico  
de barras  
mostrando o  
número de  
vezes que cada  
número foi  
sorteado.



# Porcentagem de Vezes que Cada Número Foi Sorteado

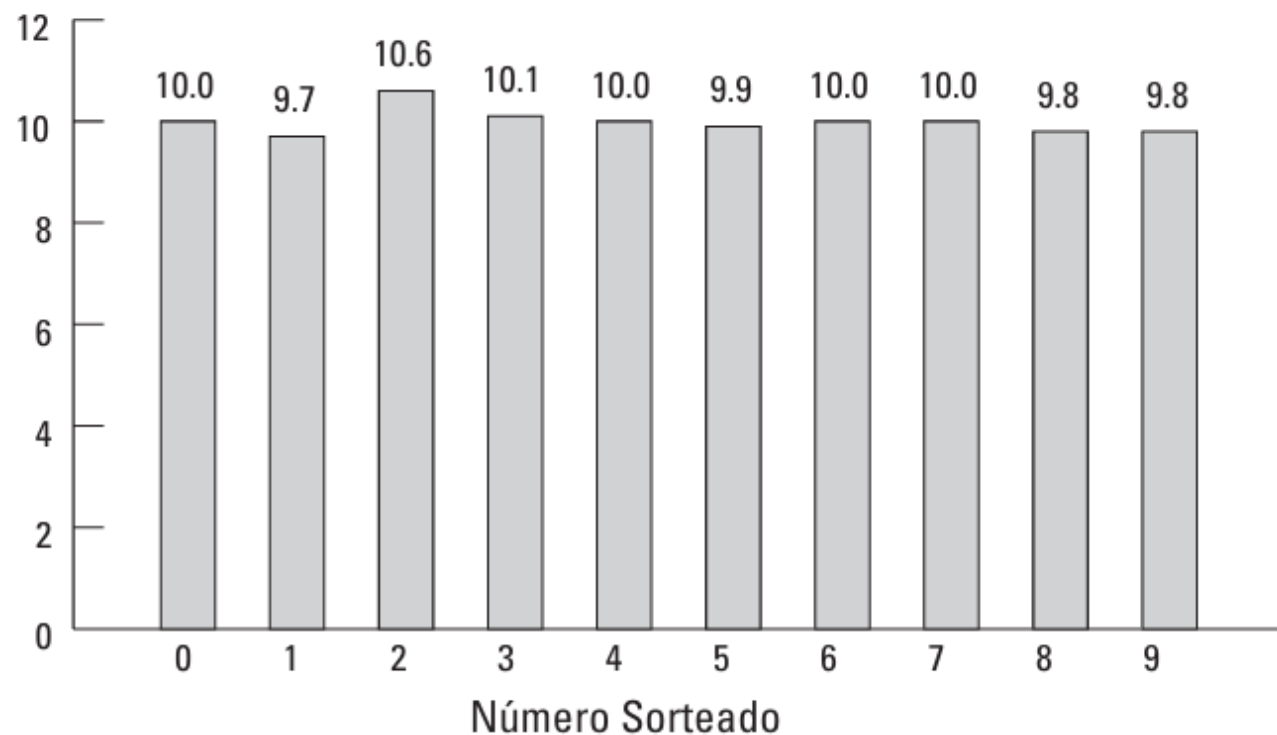
<i>Número Sorteado</i>	<i>Número de Vezes do Sorteio</i>	<i>Porcentagem de Vezes do Sorteio</i>
0	485	$10,0\% = 485 \div 4.839$
1	468	$9,7\% = 468 \div 4.839$
2	513	$10,6\% = 513 \div 4.839$
3	491	$10,1\% = 491 \div 4.839$
4	484	$10,0\% = 484 \div 4.839$
5	480	$9,9\% = 480 \div 4.839$
6	487	$10,0\% = 487 \div 4.839$
7	482	$10,0\% = 482 \div 4.839$
8	475	$9,8\% = 475 \div 4.839$
9	474	$9,8\% = 474 \div 4.839$





# Escala mais realista

**Figura 2-2:**  
Gráfico  
de barras  
mostrando a  
porcentagem  
de vezes que  
cada número foi  
sorteado



# Como podemos observar esses erros e conhecendo evitamo-los



# Exemplos de Más Práticas Estatísticas

## Amostragem Tendenciosa

Ocorre quando a amostra não é representativa da população em estudo, levando a conclusões erróneas. Por exemplo, realizar um inquérito sobre hábitos de consumo apenas em centros comerciais de luxo não representa toda a população.

## Correlação vs. Causalidade

Confundir correlação com causalidade é um erro comum. Duas variáveis podem estar correlacionadas sem que uma cause a outra. Por exemplo, o aumento das vendas de gelados e o aumento de afogamentos no verão estão correlacionados, **mas não há relação causal direta.**



# Mentiras e Ambiguidades na Estatística

## Manipulação de Gráficos

Gráficos podem ser manipulados para exagerar ou minimizar diferenças, alterando escalas ou omitindo informações cruciais. Um gráfico de barras com eixo Y truncado pode fazer pequenas diferenças parecerem significativas.

## Seleção Seletiva de Dados

Escolher apenas os dados que suportam uma conclusão desejada, ignorando informações contraditórias. Por exemplo, uma empresa farmacêutica que publica apenas os estudos favoráveis ao seu medicamento.



# Como as Pessoas Manipulam os Dados?

## **Influência Financeira**

Empresas podem distorcer estatísticas para aumentar seus lucros ou atrair investidores.

## **Propaganda Política**

Políticos podem usar estatísticas para defender suas posições ou atacar seus oponentes.

## **Mídia Sensacionalista**

Alguns meios de comunicação distorcem dados para gerar manchetes e aumentar a audiência.



# Alguns Vieses Estatísticos

## **Viés de Seleção**

Quando a amostra não representa a população-alvo.

## **Viés de Confirmação**

Procurar evidências que confirmem uma hipótese ou crença pré-existente, ignorando dados que possam contrariá-la.

## **Viés de Sobrevivência**

Focar apenas nos casos de sucesso e ignorar os fracassos.



# Más Utilizações da Estatística

1

## Extrapolação

Aplicar conclusões de um estudo a situações muito diferentes das originais. Por exemplo, generalizar resultados de um estudo realizado com estudantes universitários para toda a população adulta.

2

## Omissão de Contexto

Apresentar estatísticas sem o contexto necessário para a sua interpretação adequada. Por exemplo, reportar um aumento de 100% nos casos de uma doença rara sem mencionar que o número absoluto passou de 1 para 2 casos.

3

## Uso de Médias

Utilizar médias quando a distribuição dos dados é muito assimétrica ou contém valores extremos. A média salarial de uma empresa pode ser inflacionada por alguns salários muito altos, não representando a realidade da maioria dos trabalhadores.

4

## Falsa Precisão

Apresentar resultados com um nível de precisão injustificado pelos dados ou métodos utilizados. Por exemplo, afirmar que 74,62% das pessoas preferem um produto, quando o tamanho da amostra não suporta tal precisão.



A má utilização da estatística pode levar a consequências significativas em diversas áreas

**Caso de Lucia de Berk (Holanda):** Lucia de Berk, uma enfermeira holandesa, foi condenada por múltiplos assassinatos com base em evidências circunstanciais e análises estatísticas que sugeriam uma correlação entre sua presença e mortes de pacientes. Posteriormente, verificou-se que as análises estatísticas utilizadas eram falhas, resultando na condenação injusta de Lucia. **Após revisão, ela foi absolvida de todas as acusações.**



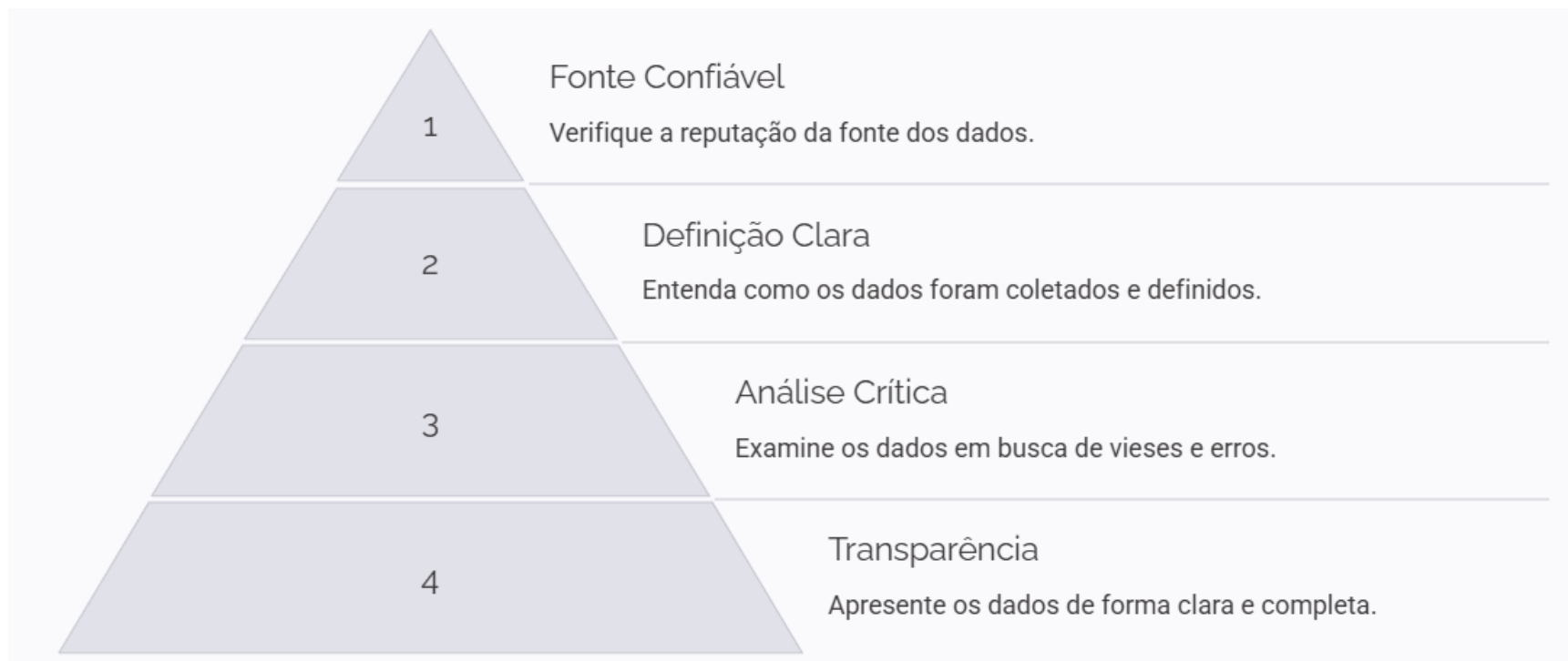


# cases

- A turma sera dividida em grupos para procurar cases que mostrem como a estatistica e a analise de dados mau realizada pode influenciar uma decisao erronea ou mudar o foco



# Melhores Práticas para Usar a Estatística de Forma Responsável



## Mentiras, mentiras deslavadas e estatística

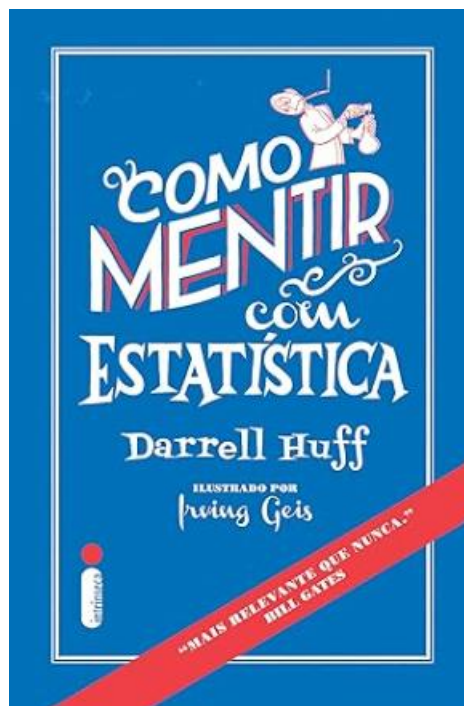
### Livro: Estatística: O que é, para que serve, como funciona

Charles Wheelan (Autor)

Mesmo nas melhores circunstâncias, a análise estatística raramente revela “a verdade”. Em geral construímos um caso circunstancial baseado em dados imperfeitos. Como resultado, há numerosas razões para que indivíduos intelectualmente honestos discordem acerca de resultados estatísticos ou suas implicações. No nível mais básico, podemos discordar sobre a questão que está sendo respondida. Fãs de esportes discutirão por toda a eternidade sobre quem é “o melhor jogador de beisebol de todos os tempos” porque não há definição objetiva de “melhor”. Estatísticas descritivas rebuscadas podem fornecer informações sobre essa questão, mas jamais a responderão em definitivo. Como será ressaltado no próximo capítulo, questões mais significativas socialmente viram presas do mesmo desafio básico. O que está acontecendo com a saúde econômica da classe média americana? A resposta depende de como se define “classe média” e “saúde financeira”.



# Leitura Recomendada



<https://www.youtube.com/watch?v=nJ4-K93ZaVk&t=84s>



# Obrigado!

