

Laboratório de Fonologia



Estatística para Linguística

Prof. Dr. Adelino Pinheiro Silva

Sumário

Introdução

Estatística descritiva

Encerramento

Referências



Sumário

Introdução

Estatística descritiva

Encerramento

Referências



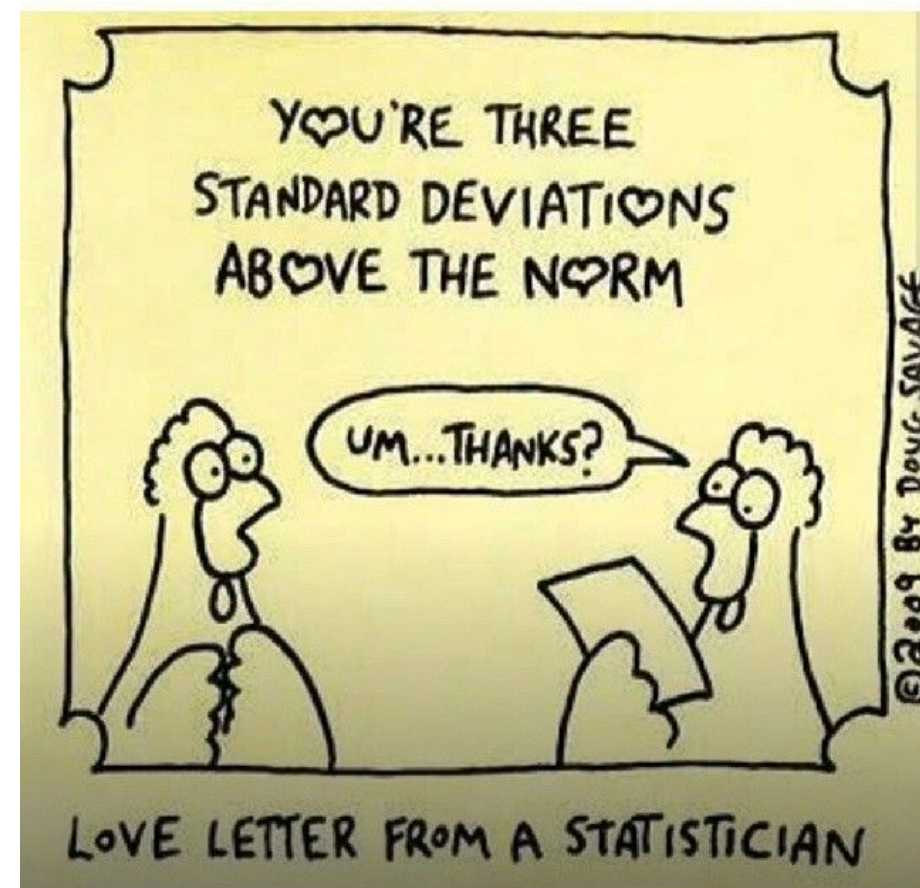
► In a hole in the ground there lived a...

Por que estudar estatística?

- Compreender **fatores** que afetam um resultado.
- Julgar de forma crítica as informações recebidas.
- Argumentar estatisticamente.

O que é estatística (Agresti, 2018)?

- Conjunto de métodos para se **obter** e **analisar** dados.
- Metodologia baseada na **ocorrência** para realizar **previsão**.



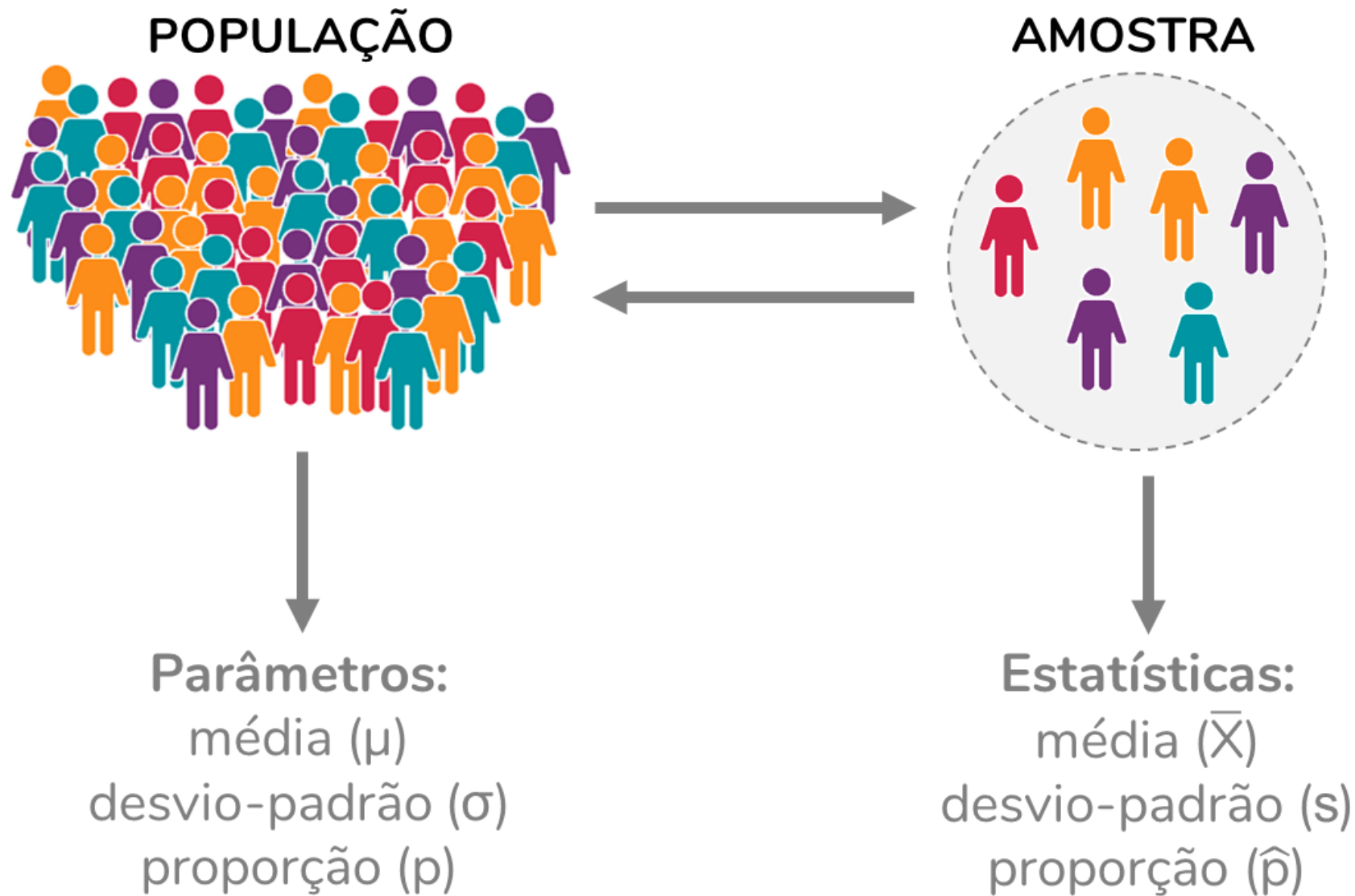
► I Have the High Ground

Alguns termos para começar

- **Dado:** Observação obtida sobre o objeto de interesse.
- **Observação:** Medida, ou informação coletada (sujeita a ruído e erros).
- **Base de dados:** Conjunto de dados, e.g., *general social survey*.
- **População:** Conjunto total dos elementos (desconhecido, inacessível).
- **Amostra:** subconjunto da população, dados (medidas) coletados.
- **Parâmetro:** Fator (resumo) numérico da população (dica: letras gregas).
- **Estatística:** Valor obtido da amostra !!!!!
- **Ferramental:** R-studio



I Have the High Ground



Maneiras de extrair informações de interesse.

- **Variável aleatória:** Característica que pode variar com os elementos da população ou amostra.
- **Escala de medição:** Extensão onde a variável aleatória pode ser medida. Exemplos:
 - Categóricas: (cara, coroa), (derrota, empate, vitória); ou
 - Quantitativas: $\{x \in \mathbb{R} | 0 \leq x \leq 1\}$, $[0, 1]$

Se caracteriza a variável aleatória como um resultado de uma experiência aleatória, que pode ser classificada como:

- **Categóricas:** valores aceitos dentro de um limite de categorias (qualitativos?).
- **Quantitativas:** valores numéricos de qualquer conjunto, e.g., \mathbb{N} , \mathbb{R} , \mathbb{C}

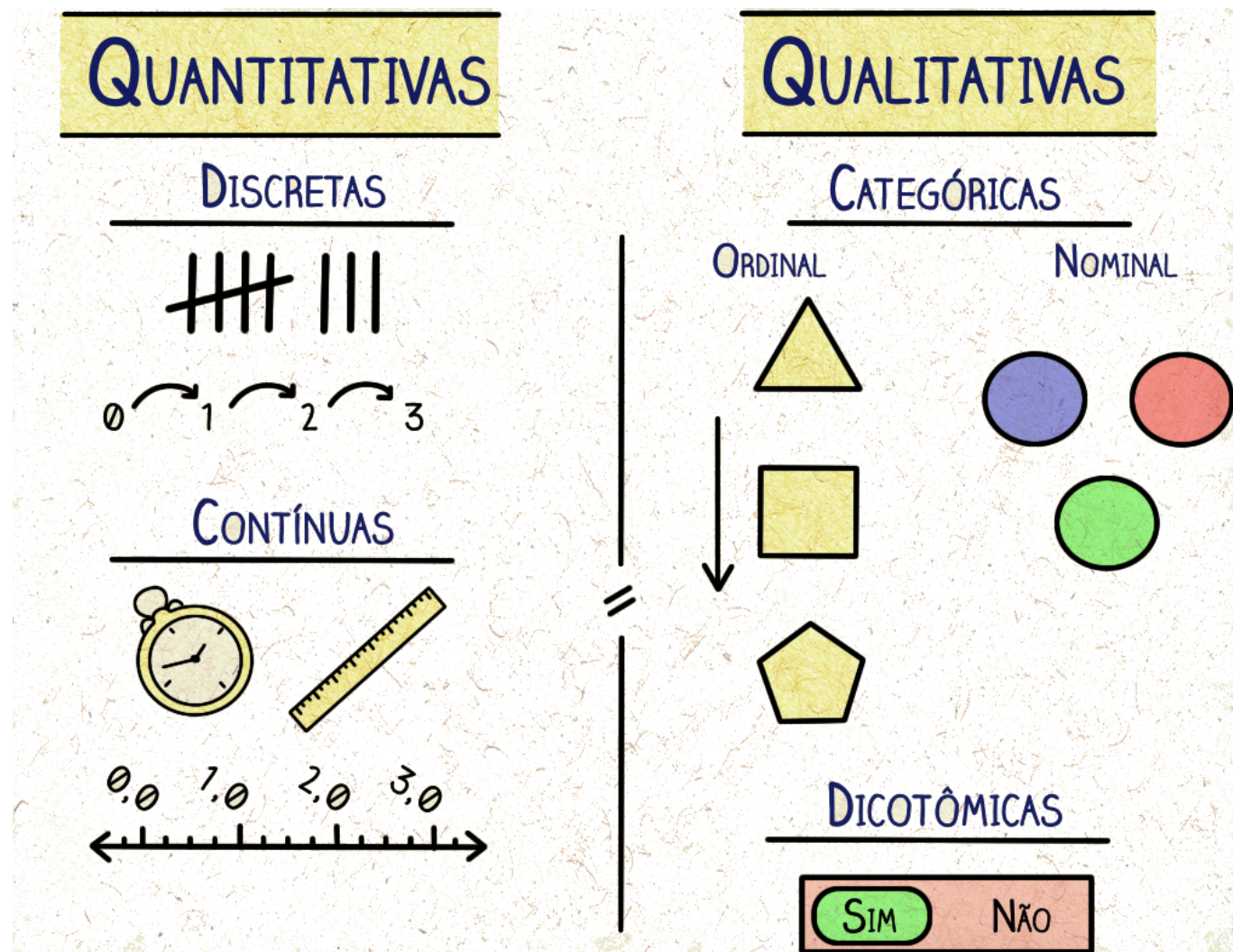
Escalas:

- **Intervalar:** delimitação numérica.
- **Nominal:** Nomes/categorias “não ordenáveis”, e.g., preferência de cores;
- **Ordenáveis:** Nomes/categorias que podem ser ordenadas em níveis, e.g., expectativa do curso (baixa, sem expectativa, alta).

Detalhe: Em escalas categóricas é muito difícil garantir uma homogeneidade dos intervalos, i.e., se os intervalos das categorias possuem escalas de mesmo tamanho.

Variáveis estatísticas

- **Amostra aleatória simples:** todas amostras de mesmo tamanho possuem a mesma “chance”. Seria um retrato da população(?).
- **Métodos de amostragem, *sample survey*:** Sistemática, estratificada, grupo (*cluster*), multiestágios.
- **Amostra enviesada:** alunos de uma sala de aula (?).



Estudo experimental

Experimento: Controlar variáveis independentes e observar a variação de variáveis dependentes para dar suporte ou refutar uma hipótese.

- Compara “tratamentos”.
- Unidades de testes.
- Grupos, pelo menos, “controle” e “tratamento”.
- Variáveis estranhas (predatórias).

Problemas experimentais

- Variação do instrumento (ou pessoa que conduz parte dele).
- Regressão analítica.
- Viés de seleção.
- Perda de unidade



Experimento

Efeitos do teste: principal e interativo

Soluções para experimentos:

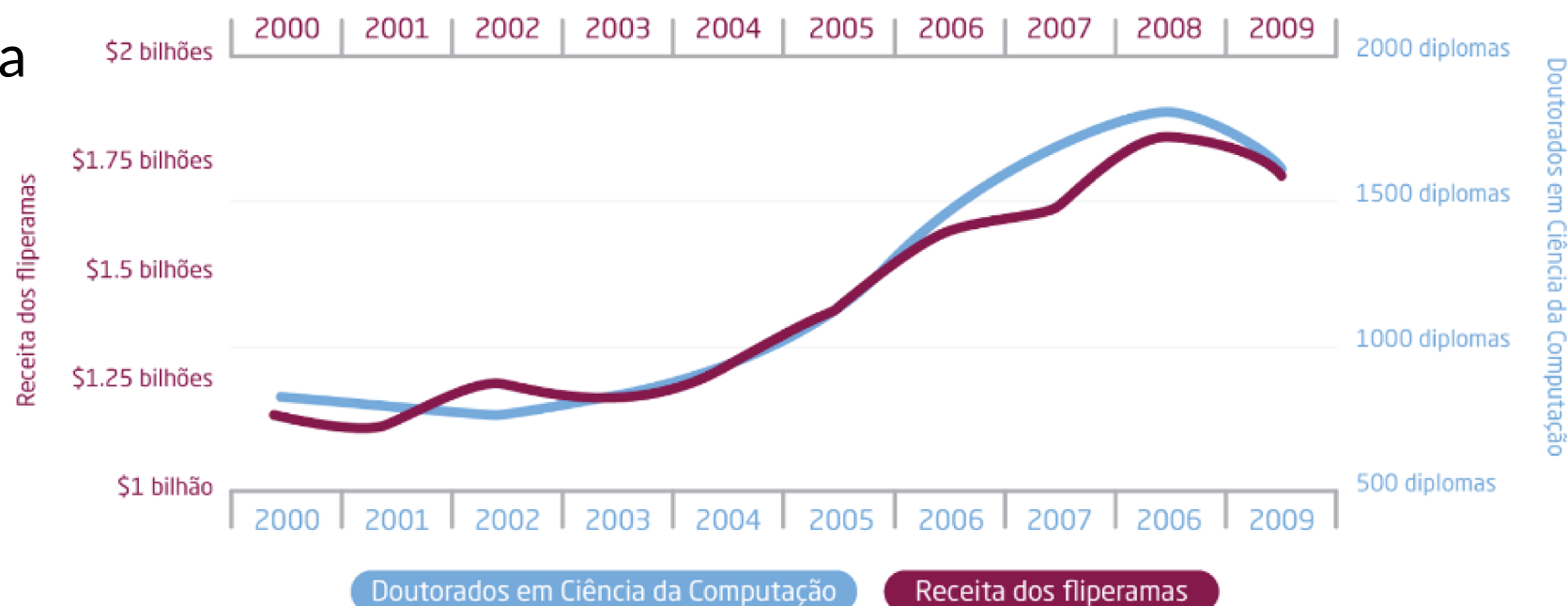
- Aleatorização.
- Emparelhamento.
- controle estatístico.
- Planejamento.
- Medições *a posteriori*.



Estudo de Observação

- Sem manipulação do objeto de estudo.
- Grupos desbalanceados, difícil de realizar uma comparação adequada.
- **Não permite estabelecer causa e efeito.**
- Pode indicar uma relação entre variáveis.
- Uma variável não medida pode ser responsável pelo padrão observado.

RECEITA TOTAL GERADA POR FLIPERAMAS CORRELACIONA COM DOUTORADOS EM CIÊNCIA DA COMPUTAÇÃO CONCEDIDOS NOS E.U.A



► Variabilidade amostral e viés

Erro de amostragem: erro ocorrido ao utilizar uma estatística da amostra para prever um parâmetro da população. Exemplo: Erro da pesquisa eleitoral com $n = 100$ de $+ \text{ ou } - 3\%$.

Viés: erro quando a amostra é enviesada, e.g., voluntários ou respostas de carta.

- **Viés de resposta** ocorre quando a pergunta é confusa, e.g., referendo do desarmamento;
- **viés de falha de dados** apenas uma fatia da amostra responde.



Fim da introdução - Dever de casa

Exercícios do livro Agresti (2018):

- Capítulo 1: 1.1, 1.3, 1.5-1.8, 1.14, 1.16;
- Capítulo 2: 2.2-2.10, 2.27, 2.35-2.37, 2.39

Preparação do terreno

- Instalar o R-studio.



Sumário

Introdução

Estatística descritiva

Encerramento

Referências



Dever de casa



Dever de casa



Dever de casa



Dever de casa



Sumário

Introdução

Estatística descritiva

Encerramento

Referências



Sobre este material

Esta obra está licenciada sob a licença *Creative Commons* CC BY-NC-SA 4.0 (mais detalhes neste *link*)

Favor fazer referência a este trabalho como:

Silva, A. P. (2022), *Notas de Aulas de Estatística para Linguística*. Online:
<https://github.com/adelinocpp/>

```
@Misc{Silva2022,  
  title={Notas de Aulas de Notas de Aulas de Estatística para Linguística},  
  author={Adelino Pinheiro Silva},  
  howPublished={\url{https://github.com/adelinocpp/}},  
  year={2022},  
  note={Version 1.0; Creative Commons BY-NC-SA 4.0.},  
}
```

Sumário

Introdução

Estatística descritiva

Encerramento

Referências



Referências I

Agresti, A. (2018). *Statistical methods for the social sciences*. Number 300.72 A3. Pearson.

