

# Análise Quantitativa de Dados em Linguística

Questões prévias à análise de dados

---

**Ronaldo Lima Jr.**

`ronaldojr@letras.ufc.br`

`ronaldolimajr.github.io`

Universidade Federal do Ceará

1. Análise Quantitativa de Dados na Linguística
2. Objetivos da Análise Quantitativa de Dados
3. Antes da Coleta de Dados
4. Antes da análise dos dados

# **Análise Quantitativa de Dados na Linguística**

---

- Séc. XX → Estruturalismo e Gerativismo
  - Objeto da Linguística como estrutura invariável, sistema de oposições, inato, discreto e categórico
  - Levou a introspecção e dados/exemplos autocunhados

*“Large groups of people make up all their utterances out of the same stock of lexical forms and grammatical constructions. A linguistic observer therefore can describe the speech-habits of a community without resorting to statistics”* (Bloomfield 1935: 37)

*“I think we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure”* (Chomsky 1957: 17)

- Abordagens baseadas no uso, variação e mudança
- Dados de corpora, questionários, experimentos, correlatos físicos, construtos da psicologia, etc.
- psicolinguística, linguística de corpus, fonética/fonologia, sociolinguística, linguística aplicada, aquisição de L2, linguística forense, aprendizagem de máquinas, etc.
  - e mesmo subáreas “tradicionais” que não utilizavam (sintaxe, fonologia, etc.)
- Como saber se uma diferença de 3%, 5%, 8%, 10%, etc. na fala de pessoas de dois estados, de duas faixas etárias, etc. é aleatória ou efeito?

# **Objetivos da Análise Quantitativa de Dados**

---

# Objetivos da análise quantitativa de dados

- Descrever
- Explicar
- Prever

## **Antes da Coleta de Dados**

---



Problema → Pergunta(s) → Hipótese(s)  
→ Verificar/Observar/Testar → Inferência/Conclusão

- Ex. 1: encontrar o melhor preço para um produto
- Ex. 1: motivos de lentidão em um novo percurso

“Pesquisa mostra que correr pode ser prejudicial a saúde”

# Passos

- identificar/caracterizar o problema
  - estudar a literatura (teoria, construtos, métodos, resultados, variáveis, lacunas, etc.)
  - observar o fenômeno e raciocínio dedutivo por parte do pesquisador
  - compilar variáveis passíveis de influência
  - estipular hipóteses
    - falsiáveis e testáveis
    - H1 e H0
      - somadas devem abarcar todos possíveis resultados
      - generalizáveis à população e outras amostras
      - direcional / não direcional
    - identificar/vislumbrar covariáveis (*confounding variables*)
  - pensar em como operacionalizar as variáveis (observar, medir, contar)
- medir errado não tem conserto!
- pensar em amostra equilibrada e o mais randomizada possível

→ Tudo isso **ANTES** de coletar os dados!

- HARKing
- p-hacking
- crises de replicabilidade
- ciência aberta
- Pre-registration
- OSF, Github
  - Palestra do Brian Nosek no Interab12

# Coleta: operacionalizando as variáveis

Exemplo 1:



The image shows a screenshot of a website. At the top left is a logo for '10 anos acontecendo aqui' with a colorful '10' and the text 'anos acontecendo aqui' below it. To the right of the logo is a dark grey rectangular area. Below the logo and the dark area is a horizontal navigation bar with several buttons: 'inicio', 'comunicação', 'design', 'propaganda', 'digital', 'marketing' (which is highlighted in white), 'variedades', 'colunas', and 'entre'. Below the navigation bar is the main content area. On the left side of this area is a vertical sidebar with three elements: a box with the number '7', a 'Tweeter' button with a Twitter icon, and a 'Like' button with a thumbs-up icon and the number '18'. To the right of the sidebar is the main article. The article title is 'Ação com gêmeos idênticos mostra que mascar chiclete dá uma boa impressão'. Below the title is the date '28 de novembro de 2013'. To the right of the date are two small buttons labeled 'a-' and 'A+'. Below the date and buttons is the article text: 'A marca argentina de chicletes Beldent Gum, uma versão do Trident, realizou uma ação para mostrar que, ao contrário do que se pensa, mascar chiclete dá uma boa impressão às pessoas.'

10 anos  
acontecendo  
aqui

inicio comunicação design propaganda digital marketing variedades colunas entre

7  
Tweeter  
18  
Like

## Ação com gêmeos idênticos mostra que mascar chiclete dá uma boa impressão

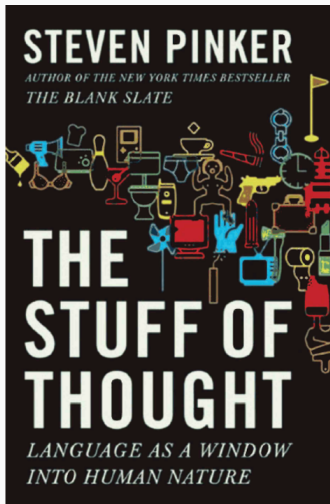
28 de novembro de 2013

A- A+

A marca argentina de chicletes Beldent Gum, uma versão do Trident, realizou uma ação para mostrar que, ao contrário do que se pensa, mascar chiclete dá uma boa impressão às pessoas.

## Coleta: operacionalizando as variáveis

Exemplo 2:



## **Antes da análise dos dados**

---

# Organizar os dados

- Primeira linha com nomes da variáveis
- Cada linha com apenas uma observação
- Cada coluna contendo uma variável
- Dados faltantes com NA
- Sem formatação, sem espaços e sem caracteres especiais

**Perguntas?**