

# MÓDULO 1

# INTRODUÇÃO A MÉTODOS QUANT.

PEDRO H. G. FERREIRA DE SOUZA  
SERGEI S. D. SOARES

INSTITUTO DE PESQUISA ECONÔMICA APLICADA  
MESTRADO PROFISSIONAL EM POLÍTICAS PÚBLICAS E DESENVOLVIMENTO

4ª TURMA

SET-DEZ 2019

**ipea**

Toda pesquisa científica ambiciona produzir **conhecimento sistemático** sobre o mundo.

- Análise de regularidades socialmente relevantes por meio do confronto entre teoria e evidência conforme procedimentos estruturados.
- **Rigor** é a palavra-chave: sem improvisos, romper com senso comum.

Que tipo de análise?

- Descrição
- Causalidade
- Predição

Qual a pergunta de pesquisa de vocês?

# QUALITATIVO VS. QUANTITATIVO

Escolha da metodologia: qual a melhor forma de responder à pergunta de pesquisa?

- **Qualitativa:** pesquisa exploratória, indutiva; compreender motivações e opiniões; descobrir novos padrões; formular novas hipóteses e teorias; explorar individualidade dos casos.
- **Quantitativa:** pesquisa descritiva ou causal, dedutiva; testar hipóteses e teorias; avaliar regularidades sociais; foco na abstração, mensuração, generalização e replicabilidade.
- (Estudos comparativos, teóricos etc.)

Lógica da inferência tanto em pesquisas qualitativas quanto quantitativas. Nesse último caso, base explícita é probabilidade e estatística.

**Validade interna** diz respeito ao grau de consistência ou controle da análise: **isolar** as relações que nos interessam para garantir que conclusões não serão espúrias.

- Se queremos saber se X tem efeito causal sobre Y, precisamos “**controlar**” por todas outras possíveis fontes de influência sobre Y.

**Validade externa** diz respeito à possibilidade de **generalizar** os resultados para um determinado universo.

- Garantir que resultados obtidos para nossa amostra também se aplicam à população mais ampla que realmente nos interessa.

**Probabilidade:** a partir de um modelo gerador, quais eventos devemos observar?

- “Preditivo”: regras que quantificam a incerteza ou **aleatoriedade**
- Modelo  $\rightarrow$  dados

**Estatística:** a partir de eventos observados, qual o modelo gerador subjacente?

- “Retrospectivo”: identificar padrões em meio à **aleatoriedade**
- Dados  $\rightarrow$  modelo

Ferramentas e regras formais para assegurar validade interna e externa.

# TRÊS NÍVEIS DO DISCURSO

**Teoria:** modelos estatísticos e distribuições de probabilidade

**Universo:** objeto de interesse, parâmetros a descobrir via inferência estatística

**Amostra:** dados observados

Aleatoriedade intrínseca ao mundo e na seleção de amostras!

$$y = \alpha + \beta \cdot x \quad \text{vs.} \quad y = \alpha + \beta \cdot x + \epsilon$$

# O QUE É MAIS COMPLICADO?

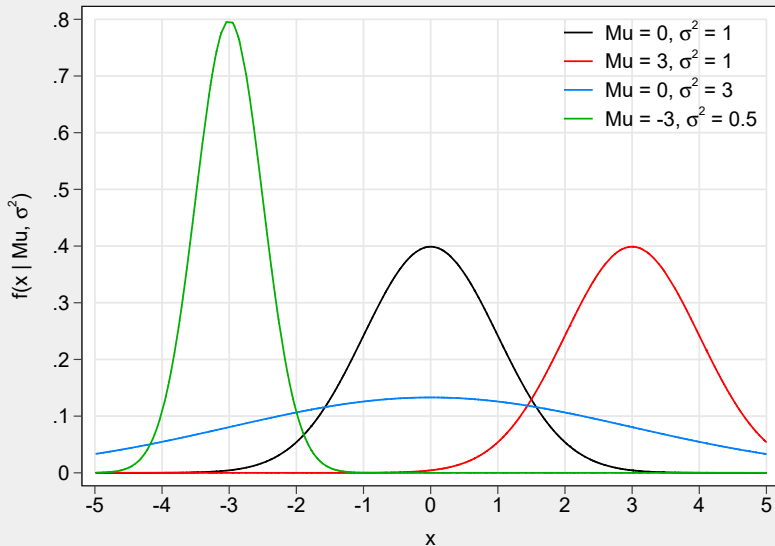
James Joyce, “Finnegans Wake” (1939)

“Heraduma noite, tarde, munto timplo atroz, numa antaiga  
erdade das perdas, quando Adão socavava e sua madãominha  
tessia cedas d’água, quandomem montenote era todimundo e a  
premeira leal costeladra que jamais ouve osseu em-fim  
todomigo com seus olhos plenamormorejantes e todomiro vivia  
solamante com todamina amais e Conte Dom Cabeço metia a  
testada tostada benhalta no farol queimorava, impondo mão  
fria assi mesmo.”

Distribuição normal:  $N(\mu, \sigma^2)$

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# DISTRIBUIÇÃO NORMAL: $N(\mu, \sigma^2)$





Suponha que vocês já partiram de uma ideia geral, especificaram um problema bem delimitado, revisaram a literatura, decidiram o escopo do estudo e elaboraram suas hipóteses de trabalho.

A próxima etapa é escolher o **desenho de pesquisa** mais apropriado para testar suas hipóteses e responder à pergunta de pesquisa.

Essa etapa é particularmente importante para estudos sobre **causalidade**.

Efeito causal de um tratamento  $T_i$  para uma unidade  $i$  dado por:

$$Y_i^1 - Y_i^0$$

O problema é que só observamos:

$$Y_i = \begin{cases} Y_i^0 & \text{se } T_i = 0 \\ Y_i^1 & \text{se } T_i = 1 \end{cases}$$

Solução: grupo de controle como contrafactual para o grupo de tratamento. Mas como garantir que os grupos são, em média, idênticos em tudo exceto pelo tratamento?

**Desenhos experimentais:** maximizam *validade interna*

- Experimentos puros

- ▶ Equivalência inicial entre grupos de tratamento e controle obtida por **alocação aleatória**
- ▶ Manipulação intencional do tratamento

- Quase experimentos

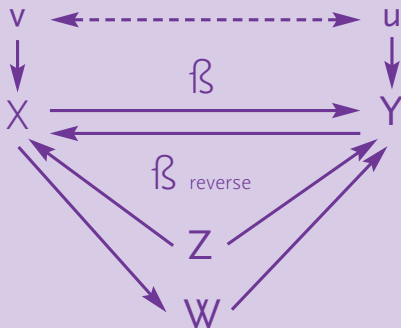
- ▶ Sem aleatorização
- ▶ Manipulação exógena (mas não intencional) do tratamento

**Desenhos não experimentais:** em geral, melhores para garantir *validade externa*

- Dados transversais (“cross-section”)
- Dados longitudinais
- Séries temporais

# POR QUE ISSO IMPORTA?

**Figure 2:**  
The Real World



Fonte: Dowd & Town 2002, p. 4.

## EXEMPLOS: VIESES

**Viés de seleção** → enquetes em sites de notícias

**Viés de variável omitida** → comparação de desempenho entre escolas públicas e privadas

**Causalidade reversa** → receber programas sociais causa pobreza

Qual o efeito causal da **educação** sobre a **renda do trabalho** dos indivíduos?

- Desenho não experimental

- ▶ Com base em dados observacionais, estimar, por exemplo,  
$$\ln(w) = \alpha + \beta_1 \cdot educ + \beta_2 \cdot exper + \beta_3 \cdot exper^2 + \epsilon$$

- Desenho quase experimental

- ▶ Aproveitar variações exógenas: por ex., regras do sistema educacional e loteria para Guerra do Vietnã

- Desenho experimental (hipotético)

- ▶ Dividir uma amostra de crianças em  $n$  grupos e sortear para cada grupo quantos anos de estudo serão obrigatórios

## Características

- Como foram coletadas as informações?
- Qual a unidade de análise?
- Qual o período de cobertura?
- Como foi selecionada a amostra?
- Qual o formato de divulgação?
- etc

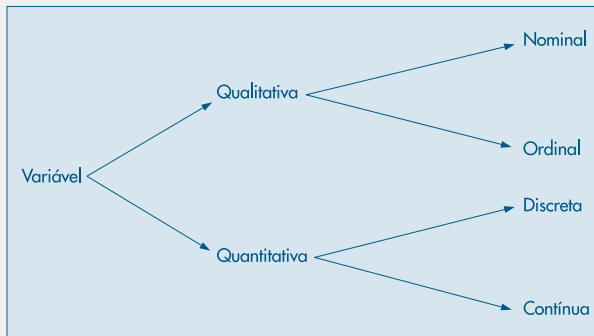
## Estrutura

- Dados transversais (“cross-section”)
- Séries temporais
- Dados longitudinais ou em painel

# TIPOS DE VARIÁVEIS

**Variáveis** são características que variam entre unidades de análise e cujos valores são quantificados de acordo com uma escala de mensuração.

**Figura 2.1:** Classificação de uma variável.



Fonte: Bussab e Morettin (2010)



# MENSURAÇÃO DE VARIÁVEIS

Medir ou mensurar → processo de vincular conceitos abstratos com indicadores empíricos, seguindo regras explícitas.

- Definição conceitual → teórica; definição operacional → empírica

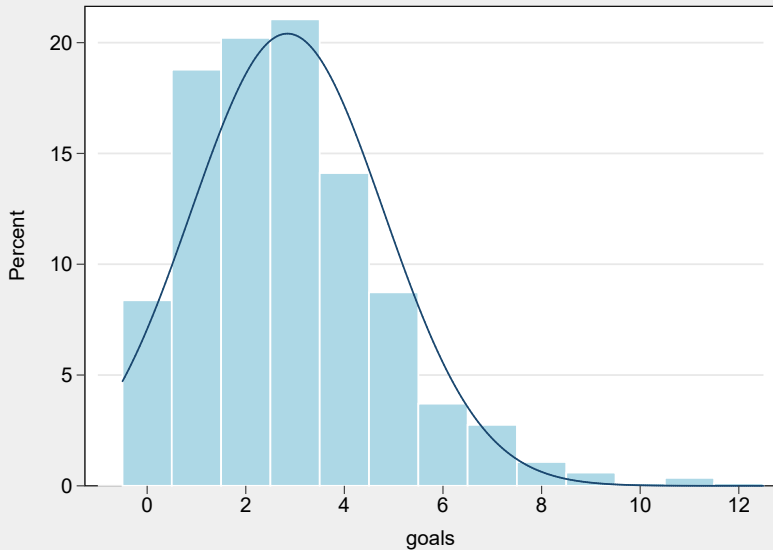
Instrumento de mensuração deve ter:

- **Confiabilidade:** aplicações repetidas com resultados iguais
- **Validade:**
  - ▶ **Conteúdo:** instrumento reflete adequadamente o domínio de conteúdo daquilo que medimos; abarca as principais dimensões
  - ▶ **Critério:** instrumento congruente com outros critérios externos que medem a mesma coisa
  - ▶ **Constructo:** instrumento reflete de fato o constructo teórico subjacente; mede o que queremos que ele esteja medindo
- **Objetividade:** não influência de vieses e tendências dos pesquisadores

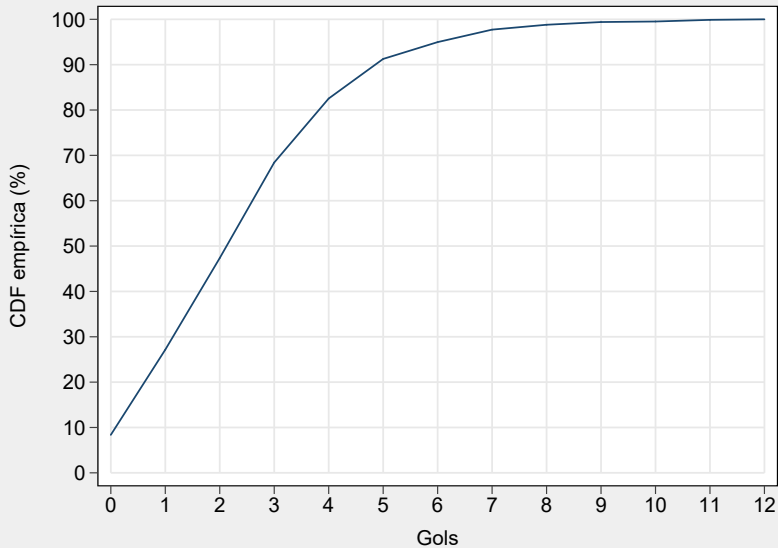
# EXEMPLO - JOGOS DE COPA DO MUNDO

id	date	round	stadium	city	country	team1	g1	team2	g2
1	07/13/1930	Group 1	Pocitos	Montevideo	Uruguay	France	4	Mexico	1
2	07/13/1930	Group 4	Pq. Central	Montevideo	Uruguay	USA	3	Belgium	0
3	07/13/1930	Group 2	Pq. Central	Montevideo	Uruguay	Yugoslavia	2	Brazil	1
4	07/13/1930	Group 3	Pocitos	Montevideo	Uruguay	Romania	3	Peru	1
5	07/15/1930	Group 1	Pq. Central	Montevideo	Uruguay	Argentina	1	France	0
6	07/16/1930	Group 1	Pq. Central	Montevideo	Uruguay	Chile	3	Mexico	0
7	07/17/1930	Group 2	Pq. Central	Montevideo	Uruguay	Yugoslavia	4	Bolivia	0
8	07/17/1930	Group 4	Pq. Central	Montevideo	Uruguay	USA	3	Paraguay	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
829	07/04/2014	Qtr-finals	Maracana	Rio de Janeiro	Brazil	France	0	Germany	1
830	07/04/2014	Qtr-finals	Castelao	Fortaleza	Brazil	Brazil	2	Colombia	1
831	07/05/2014	Qtr-finals	Est. Nacional	Brasilia	Brazil	Argentina	1	Belgium	0
832	07/05/2014	Qtr-finals	Fonte Nova	Salvador	Brazil	Netherlands	0	Costa Rica	0
833	07/08/2014	Semis	Mineirao	Belo Horizonte	Brazil	Brazil	1	Germany	7
834	07/09/2014	Semis	Arena SP	Sao Paulo	Brazil	Netherlands	0	Argentina	0
835	07/12/2014	3 <sup>rd</sup> place	Est. Nacional	Brasilia	Brazil	Brazil	0	Netherlands	3
836	07/13/2014	Final	Maracana	Rio de Janeiro	Brazil	Germany	1	Argentina	0

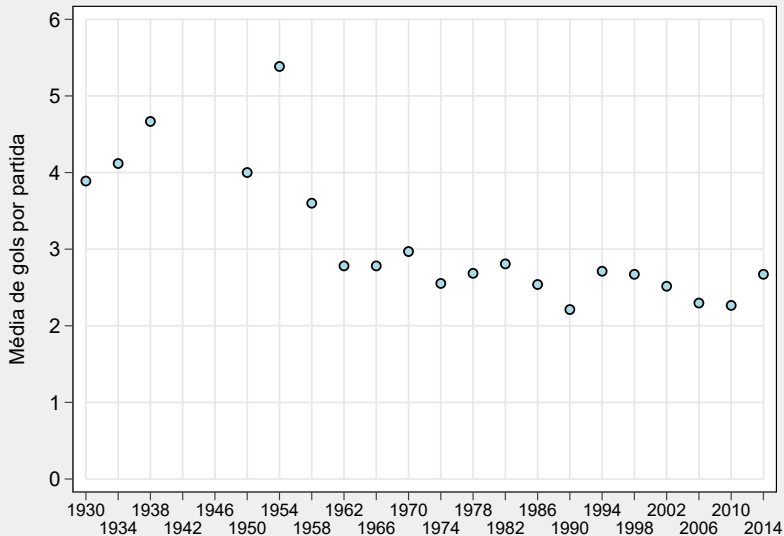
## EXEMPLO - PDF DE GOLS POR JOGO



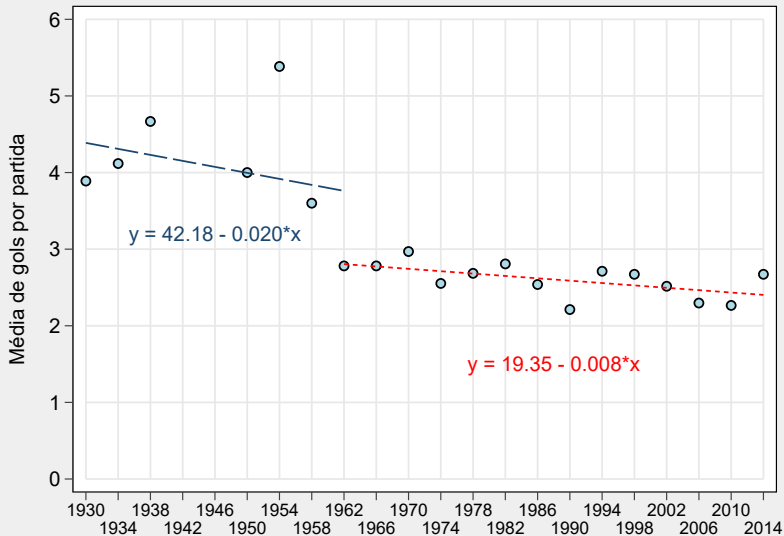
## EXEMPLO - CDF DE GOLS POR JOGO



# EXEMPLO - MÉDIA DE GOLS POR JOGO



# EXEMPLO - MÉDIA DE GOLS POR JOGO



## Definição informal

Uma variável aleatória é uma variável que assume um valor numérico único, determinado pelo “acaso”, para cada resultado de um procedimento.

Uma distribuição de probabilidade é uma descrição das probabilidades de ocorrência de cada valor da variável aleatória.

## Definição um pouco menos informal

Uma variável aleatória é uma função que associa os elementos do espaço amostral a valores numéricos de acordo com uma determinada distribuição de probabilidade.

# LEITURAS RECOMENDADAS

## Metodologia de pesquisa

Ragin C.; Amoroso, M. (2011) *Constructing Social Research*. Londres, Los Angeles: SAGE, Pine Forge Press.

Sampieri, R., Collado C.; Lucio M. (2013). *Metodologia de Pesquisa*. Porto Alegre: McGraw Hill, Penso. (Capítulos 2-10)

## Causalidade

Dowd B.; Town R. (2002) *Does X really cause Y?*. Washington, DC: AcademyHealth.

Goldthorpe J. (2001) Causation, Statistics, and Sociology. *European Sociological Review*, v. 17, n. 1, p. 1-20.

Pearl J. (2018) *The Book of Why*. Nova York: Basic Books.

Rubin D. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, v. 66, n. 5, p. 688-701.

## Mensuração de variáveis

Bussab W.; Morettin P. (2010): *Estatística Básica*. São Paulo: Saraiva. (Capítulo 2)

Sampieri, R., Collado C.; Lucio M. (2013). *Metodologia de Pesquisa*. Porto Alegre: McGraw Hill, Penso. (Capítulo 9)

## Efeitos da educação sobre a renda

Soares S. (2011) *O conhecimento paga bem? Habilidades cognitivas e rendimentos do trabalho no Brasil (e no Chile)*. Tese de Doutorado em Economia, Universidade de Brasília.



## Leituras recomendadas

- Agresti & Finlay, capítulo 3 (p. 31-72)
- Bussab & Morettin, capítulos 2-4 (p. 9-102)
- Lyman & Longnecker, capítulo 3 (p. 56-139)
- Triola, capítulo 3 (p. 82-135)