

Métodos Estatísticos Básicos

Aula 1 - Conceitos Introdutórios

Regis A. Ely

Departamento de Economia
Universidade Federal de Pelotas

22 de junho de 2020

Conteúdo

Importância da estatística

Por que economistas precisam estudar estatística?

Por que a estatística na economia é tão complexa?

Ramos da estatística

Definições e Conceitos

Tipos de variáveis estatísticas

Exemplo no R

Como instalar o R

Instalar e carregar pacotes no R

Exemplo: salários e bebidas alcoólicas

Plotando um gráfico no R

Por que economistas precisam estudar estatística?

Principal instrumento de análise da economia aplicada

Através do uso da estatística, o economista pode:

1. Interpretar e analisar dados
2. Identificar relações entre variáveis
3. Avaliar o impacto de ações governamentais ou corporativas
4. Prever e construir cenários para o futuro

Em resumo, através da estatística, o economista pode condensar informação para auxiliar na tomada de decisão

Por que a estatística na economia é tão complexa?

Como você estimaria

- a relação entre horas de sono e rendimento escolar?
- a relação entre democracia e crescimento econômico?

A ciência econômica, em geral, é uma ciência não-experimental

Variáveis econômicas são extremamente interdependentes entre si

Ramos da estatística

Podemos citar 6 grandes ramificações da estatística

- **Probabilidade:** base fundamental da teoria estatística
- **Amostragem:** estuda os tipos de amostra e os procedimentos para obtermos amostras representativas de uma população (ex: *PNAD, PME, Questionários*)
- **Inferência:** base teórica para o estudo da relação entre variáveis e a elaboração de testes de hipótese estatísticos

Ramos da estatística

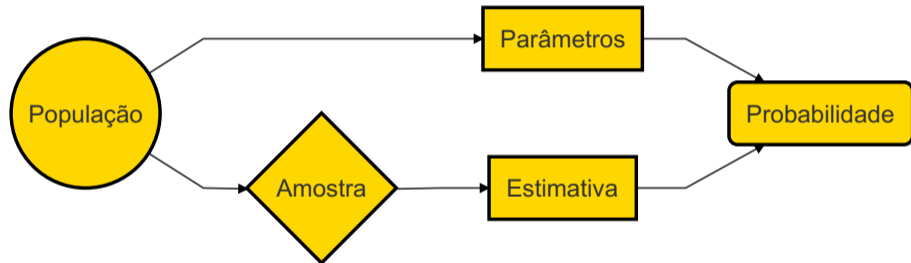
- **Econometria:** conjunto de métodos para estimar a relação entre variáveis, especialmente em estudos não-observacionais (*extensão da inferência*)
- **Séries Temporais:** conjunto de métodos para estudar a dinâmica das variáveis no tempo
- **Aprendizagem estatística:** conjunto de algoritmos capazes de aprender com os dados e melhorar suas previsões automaticamente

Definições e Conceitos

Termos importantes que serão utilizados neste curso

- **População:** conjunto total de elementos portadores de, pelo menos, uma característica comum
- **Amostra:** parcela representativa da população que é examinada com o propósito de tirarmos conclusões sobre a população
- **Parâmetros:** são valores singulares que existem na população e servem para caracterizá-la (*ex: média, variância*)
- **Estimativa:** valor aproximado de um parâmetro que é calculado através de uma amostra da população
- **Probabilidade:** conjunto de resultados possíveis de um fenômeno, descrito por uma função matemática que utiliza os parâmetros da população ou da amostra

Definições e Conceitos



Tipos de variáveis estatísticas

Os dados podem se apresentar em diversos formatos

- **Variável qualitativa:** valores são expressos por atributos (ex: *sexo, cor, gênero*)
- **Variável quantitativa:** valores possuem uma estrutura numérica e essas variáveis podem ser
 - *Discretas:* valores são expressos em números inteiros (ex: *nº de filhos, nº de alunos*)
 - *Contínuas:* valores são expressos em números reais (ex: *temperatura, salário*)

Como instalar o R

Durante este curso usaremos o R para exercícios e exemplos¹

- O R é uma software estatístico gratuito e de código aberto
- O R pode ser instalado em <http://cran.r-project.org/>
- É recomendável também instalar o RStudio Desktop em <https://rstudio.com/products/rstudio/>
- O RStudio é uma interface gráfica que facilita o uso do R
- Você pode encontrar muito material sobre o R online, um bom exemplo é o curso em <https://r4ds.had.co.nz/>

¹*Curiosidade:* Todo este curso, incluindo os slides, são feitos no R.

Instalar e carregar pacotes no R

Após instalar o **R** e o **RStudio** você deverá instalar e ler dois pacotes para o exemplo desta aula

- A função `install.packages()` instala pacotes no R
- A função `library()` lê pacotes no ambiente do R
- O pacote `HistData` contém a base de dados que iremos trabalhar
- O pacote `tidyverse` contém várias ferramentas para ler e analisar dados, incluindo o pacote `ggplot2`, com funções para a produção de gráficos

Instalar e carregar pacotes no R

Só é preciso instalar uma vez cada pacote:

```
install.packages("HistData")  
install.packages("tidyverse")
```

Sempre que iniciar o R é necessário ler os pacotes que for utilizar:

```
library(HistData)  
library(tidyverse)
```

Exemplo: salários e bebidas alcoólicas

Vamos verificar as primeiras 5 linhas da base de dados DrinksWages

```
head(DrinksWages, 5)
```

	class	trade	sober	drinks	wage	n
1	A	papercutter	1	1	24.00000	2
2	A	cabmen	1	10	18.41667	11
3	A	goldbeater	2	1	21.50000	3
4	A	stablemen	1	5	21.16667	6
5	A	millworker	2	0	19.00000	2

Descrição da base de dados

Base de dados sobre salários, profissões e consumo de álcool

- Para obter uma descrição detalhada da base de dados no R digite `?DrinksWages`
- Para acessar todas as bases de dados disponíveis no R digite `data()`

Volte para o slide anterior e responda:

- Quais variáveis são qualitativas, quantitativas, discretas e contínuas?

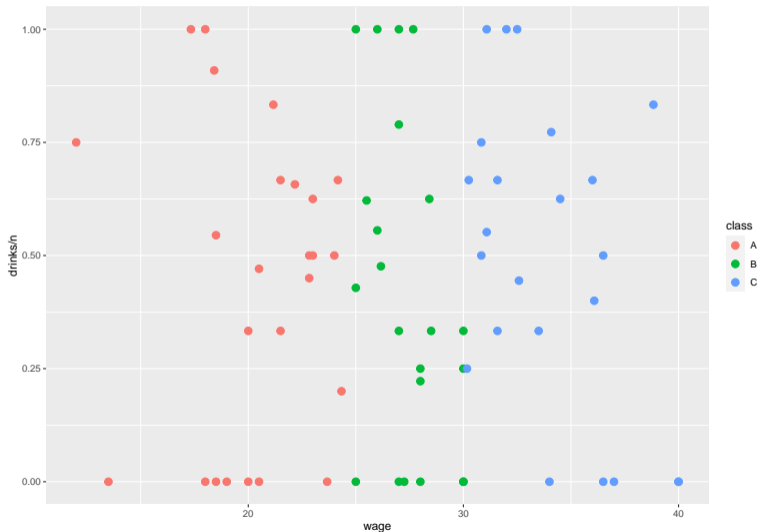
Plotando um gráfico no R

Vamos investigar a relação entre o salário dos filhos e o consumo de álcool dos pais (*Elderton e Pearson, 1910*)

- Note que as observações estão agrupadas por profissão dos pais
- Variável `wage` é o salário semanal médio dos filhos
- Variável `drinks` é o número de pais que bebem álcool
- Variável `n` é o número de pais em cada profissão
- Vamos plotar as variáveis `wage` contra `drinks/n`, colorindo os pontos conforme a classe salarial dos filhos (`class`):

```
qplot(x = wage, y = drinks/n, color = class,  
      data = DrinksWages)
```

Relação entre salário e consumo de álcool dos pais



Relação entre salário e consumo de álcool dos pais

Entre as três classes salariais dos filhos não parece haver relação significativa entre o número médio de pais que consomem bebidas alcoólicas em cada profissão e o salário semanal médio dos filhos

Referências

Elderton, M. E. e Pearson, K. *A first study of the influence of parental alcoholism on the physique and ability of their offspring*. Eugenics Laboratory Memoirs, n. 10, 1910.