

# MAT02018 - Estatística Descritiva

## Apresentações

Rodrigo Citton P. dos Reis  
[citton.padilha@ufrgs.br](mailto:citton.padilha@ufrgs.br)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2020



O professor

Olá!



Olá!



# Graduação

- ▶ **Estatística - UFRGS**

- ▶ 2002 - 2007



- ▶ **Orientador:** Prof. Alberto Cargnelutti

- ▶ **“Métodos de agrupamento para o estudo da divergência genética em cultivares de feijão”**



# Mestrado

- ▶ **Estatística - UFMG**

- ▶ 2008 - 2010



- ▶ **Orientadores:** Prof. Enrico A. Colosimo e Prof.<sup>ª</sup> Maria do Carmo P. Nunes
- ▶ **“Técnicas estatísticas para avaliação de novos marcadores de risco”**



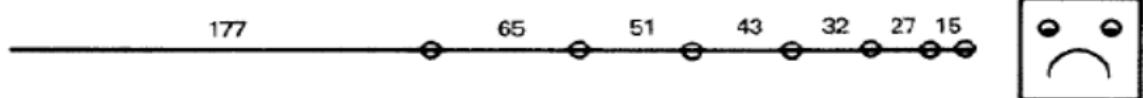
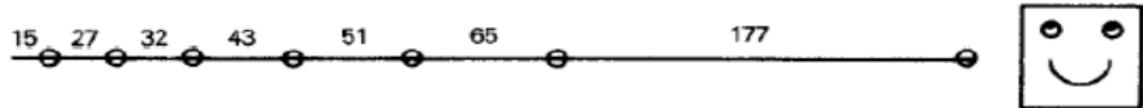
- ▶ **PPG em Infectologia e Medicina Tropical**
- ▶ Cardiopatia em pacientes com doença de Chagas

# Doutorado

- ▶ Estatística - UFMG
  - ▶ 2010 - 2014



- ▶ Orientadores: Prof. Enrico A. Colosimo e Prof. Gustavo L. Gilardoni
- ▶ “Análise hierárquica de múltiplos sistemas reparáveis”



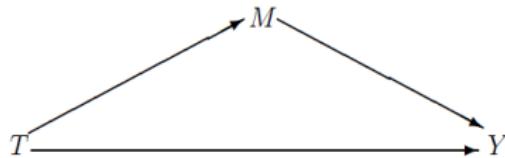
- ▶ ELSA-Brasil

# Pós-Doutorado

- ▶ Saúde Pública - UFMG
  - ▶ 2015 - 2017



- ▶ Supervisora: Prof.<sup>a</sup> Sandhi M. Barreto
- ▶ “Análise de mediação causal”



- ▶ ++ ELSA-Brasil

# Um caminho potencial



# Atividades atuais

- ▶ **Departamento de Estatística da UFRGS**
  - ▶ **Renan Baiocco.** Modelos espaço-temporais para o mapeamento da prevalência de HIV (Orientação de trabalho de conclusão de curso)
  - ▶ **Bruna Silveira.** Testes não-paramétricos de independência condicional em DAGs (Orientação de trabalho de conclusão de curso)
- ▶ **Pós-Graduação em Epidemiologia da UFRGS**
  - ▶ **Letícia Menegotto.** Análise de mediação em modelos de curva de crescimento (Orientação de mestrado em conjunto com a **Prof.<sup>a</sup> Vanessa Leotti**)
  - ▶ **Luísa Back.** Associação entre trajetória de peso corporal da infância ao incio da vida adulta e diabetes tipo 2: o Estudo Longitudinal de Saúde do Adulto (Orientação de mestrado em conjunto com a **Prof. Alvaro Vigo**)
  - ▶ **Juliana Souza.** Métodos de estimação em delineamentos duas fases: uma aplicação à subamostra de exames laboratoriais da Pesquisa Nacional de Saúde (Orientação de mestrado em conjunto com a **Prof. Márcia Barbian**)
- ▶ +++++++ **ELSA-Brasil**
- ▶ + PNS 2013 + PNS 2019

# ELSA-Brasil



Estudo Longitudinal de Saúde do Adulto



- ▶ O **Estudo Longitudinal de Saúde do Adulto** é uma investigação multicêntrica de coorte composta por 15 mil funcionários de seis instituições públicas de ensino superior e pesquisa das regiões Nordeste, Sul e Sudeste do Brasil.
- ▶ A pesquisa tem o propósito de investigar a **incidência** e os **fatores de risco** para **doenças crônicas**, em particular, as cardiovasculares e o **diabetes**.
- ▶ Em cada centro integrante do estudo, os sujeitos da pesquisa — com idade entre 35 e 74 anos — fazem exames e entrevistas nas quais são avaliados aspectos como condições de vida, diferenças sociais, relação com o trabalho, gênero e especificidades da dieta da população brasileira.

# ELSA-Brasil

## Centro de investigação Minas Gerais (Hospital Borges da Costa)



# ELSA-Brasil

Centro de investigação Rio Grande do Sul (Centro de Pesquisas Clínicas)



# A disciplina

# Objetivos

- ▶ Trabalhar o instrumental da **Estatística Descritiva** evidenciando sua importância como primeira abordagem na análise de dados, explorando sua potencialidade no campo da aplicação às demais ciências.
- ▶ Aprofundar os conceitos fundamentais da **Estatística Descritiva**, estabelecendo, via **procedimentos computacionais**, o vínculo do aluno com a **linguagem estatística**.
- ▶ Iniciar o estudante nas técnicas da construção de **indicador** de preços e de indicador de concentração, bem como nos procedimentos clássicos de **análise de séries temporais**.
- ▶ Desenvolver a **análise exploratória de dados** enfatizando sua importância como ferramenta inicial na interpretação de dados.

# Organização

- ▶ **Disciplina:** Estatística Descritiva
- ▶ **Turma:** U
- ▶ **Modalidade:** Presencial com apoio online ([Moodle](#))
- ▶ **Professor:** Rodrigo Citton Padilha dos Reis
  - ▶ e-mail: [citton.padilha@ufrgs.br](mailto:citton.padilha@ufrgs.br) ou [rodrigocpdosreis@gmail.com](mailto:rodrigocpdosreis@gmail.com)
  - ▶ Sala: B215 do Instituto de Matemática e Estatística
- ▶ Sala F115 do Instituto de Química - Salas de Aula - Campus do Vale - 43123
- ▶ **Horário:** Terças e quintas, das 10hs 30min às 12hs 10min

# Aulas e material didático

- ▶ **Aulas** (teóricas e práticas)
  - ▶ Exposição e **discussão** dos conteúdos
    - ▶ **Faremos leituras semanais de artigos e capítulos de livros**
  - ▶ Exemplos
- ▶ **Notas de aula**
- ▶ **Canais de comunicação:**
  - ▶ Durante as aulas
  - ▶ Moodle: aulas, materiais, listas de exercícios
  - ▶ Sala de aula virtual: notas das avaliações
  - ▶ e-mail do professor

# Aulas e material didático



- ▶ Exemplos e exercícios com o apoio do computador:
  - ▶ R e RStudio

```
# draw the histogram with the specified number of bins  
hist(x, breaks = bins, col = 'darkgray', border = 'white')
```

# Conteúdo programático

## “Área 1”

- ▶ Semana 1 (05/03)
  - ▶ Variáveis e medidas
- ▶ Semanas 2 a 4 (12/03)
  - ▶ Tabelas de distribuição de frequências
- ▶ Semanas 5 a 6 (02/04)
  - ▶ Análise gráfica de dados
- ▶ Semanas 7 a 8 (16/04)
  - ▶ Medidas de tendência central
- ▶ Semanas 9 a 10 (30/04)
  - ▶ Medidas de Variabilidade
- ▶ Semana 11 (14/05)
  - ▶ Medidas de forma

# Conteúdo programático (continuação)

## “Área 2”

- ▶ Semana 12 (21/05)
  - ▶ Distribuições bidimensionais
- ▶ Semanas 13 a 14 (28/05)
  - ▶ Números índices
- ▶ Semana 15 (11/06)
  - ▶ Taxas de crescimento populacional
- ▶ Semanas 16 a 18 (18/06)
  - ▶ Séries temporais
- ▶ Semana 19 (09/07)
  - ▶ Atividades de recuperação

# Avaliação

- ▶ Serão realizadas trabalhos e listas de exercícios, teóricos e práticos, ( $MT$  é a média dos trabalhos) e duas provas ( $P_1$  e  $P_2$ ) como atividades de avaliação
  - ▶ Alguns trabalhos e listas de exercícios serão realizados em grupo
  - ▶ A prova é individual e sem consulta
- ▶ Datas das provas:
  - ▶ Prova 1: 19/05
  - ▶ Prova 2: 07/07
- ▶ Cada atividade de avaliação vale 10 pontos
- ▶ Será realizada uma prova escrita como atividade de recuperação ( $P_R$ ) no dia 14/07
  - ▶ Somente para os alunos que não atingirem o conceito mínimo e com frequência mínima de 75%

# Avaliação

$$NF = 1/3 \times MT + 1/3 \times P_1 + 1/3 \times P_2$$

- ▶ **A:**  $9 \leq NF \leq 10$
- ▶ **B:**  $7,5 \leq NF < 9$
- ▶ **C:**  $6 \leq NF < 7,5$
- ▶ Se  $NF < 6$  e frequência  $\geq 75\%$ , o aluno poderá realizar a prova recuperação e neste caso

$$NF' = NF \times 0,4 + P_R \times 0,6$$

- ▶ **C:**  $NF' \geq 6$
- ▶ **D:**  $NF' < 6$
- ▶ O aluno que não obtiver o mínimo de 75% de frequência receberá o conceito **FF**

# Referências bibliográficas



- ▶ **Fernandez, D. W. X.** *Estatística descritiva II*, Cadernos de matemática e estatística. Série B, 1994. (**SABi**)
- ▶ **Fernandez, D. W. X.** *Estatística descritiva I*, Cadernos de matemática e estatística. Série B, 1994. (**SABi**)
- ▶ **Fernandez, D. W. X.** *Números índices*, Cadernos de matemática e estatística. Série B, 1992. (**SABi**)
- ▶ **Morettin, P. A. e Bussab, W. O.** *Estatística Básica*, Saraiva, 2010.
- ▶ **Tutorial para acesso fora do Campus UFRGS – Proxy e CAFé:**  
<https://www.ufrgs.br/bibliotecacentral/pesquisa/acesso-ao-proxy/>

# A Estatística

# O que é Estatística?

- ▶ Essa pergunta já vem sendo feita (e diversas vezes) há muito tempo.
- ▶ A persistência da pergunta e a variedade das respostas durante os anos sugerem que a Estatística não se caracteriza como um objeto singular.
- ▶ Ainda, a Estatística apresenta diferentes faces para diferentes áreas da ciência.

# Uma perspectiva histórica

- ▶ A Estatística mudou drasticamente desde os primeiros dias até o presente, passando de uma profissão que reivindicou uma objetividade extrema que os estatísticos coletariam apenas dados (não os analisam) para uma profissão que busca parceria com cientistas em todas as etapas da investigação, do planejamento à análise.
- ▶ Primeiros censos ocorrem por volta do ano zero da era cristã.
  - ▶ Por muito tempo, o aspecto descritivo da Estatística manteve-se como a única faceta desta ciência .
- ▶ No século XVII, ocorrem as primeiras interpretações de dados.
  - ▶ Em 1693, foram publicados pela primeira vez, em Londres, os totais anuais de falecimentos, estratificados por sexo.
  - ▶ Primeiros estudos formais da teoria das probabilidades.

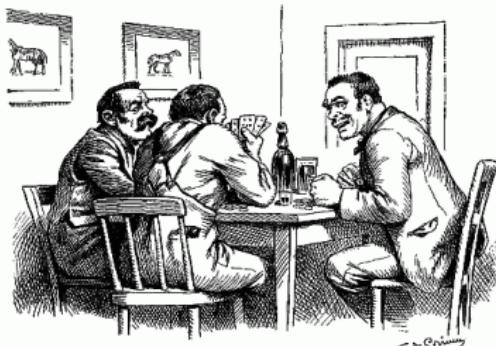
# Uma perspectiva histórica

Pascal (1623-1662)



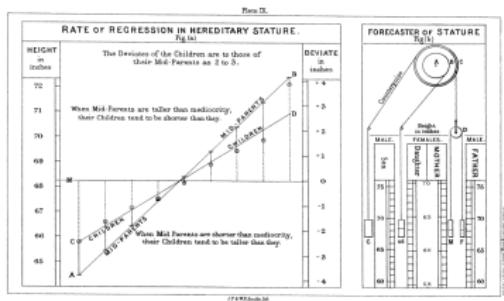
# Uma perspectiva histórica

Fermat (1601-1665)



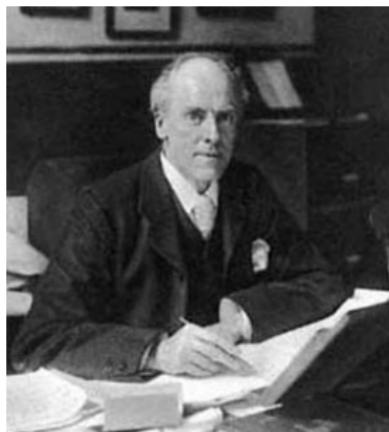
# Uma perspectiva histórica

Galton (1822-1911)



# Uma perspectiva histórica

Pearson (1857-1936)



- ▶  $\chi^2$  de Pearson
- ▶ Coeficiente de correlação linear
- ▶ Laboratório de Biometrika

# Uma perspectiva histórica

Gosset, o *Student* (1876-1937)



# Uma perspectiva histórica

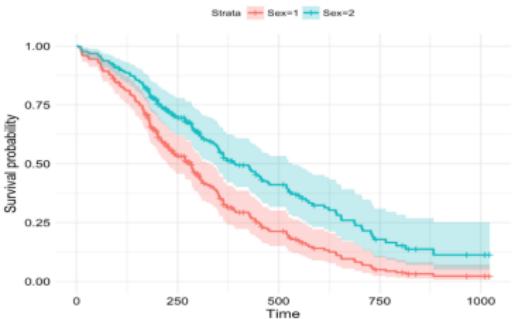
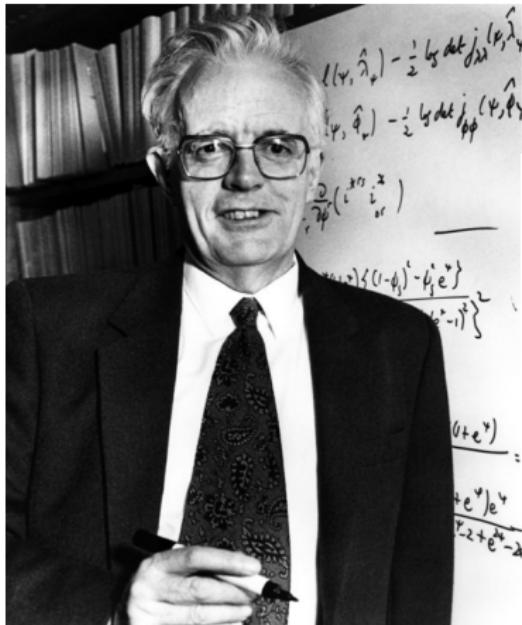
Fisher (1890-1962)



- ▶ "On the mathematical foundations of theoretical statistics"
- ▶ Princípios da experimentação
- ▶ Planejamento de experimentos
- ▶ Análise de variância (**ANOVA**)

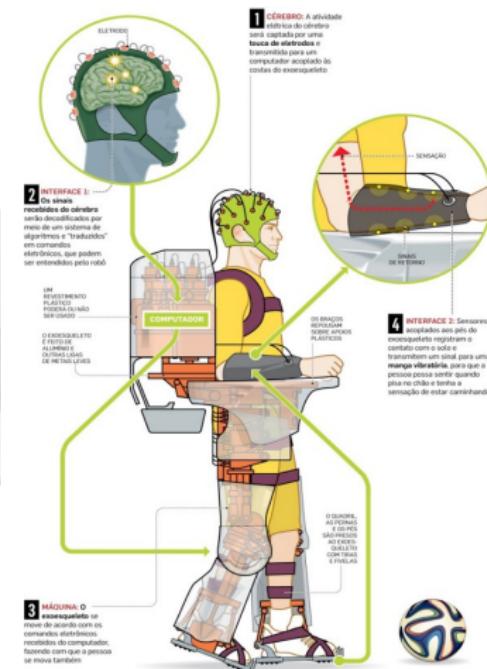
# Uma perspectiva histórica

Cox (1924)



# Uma perspectiva histórica

Nicolelis (1961)



# O papel da estatística na pesquisa

- ▶ No **planejamento**, auxilia na escolha de situações experimentais e na determinação da quantidade de indivíduos a serem examinados.
- ▶ Na **análise**, indica técnicas para resumir e apresentar as informações, bem como para comparar as situações experimentais.
- ▶ Na **elaboração das conclusões**, os vários métodos estatísticos permitem generalizar a partir dos resultados obtidos.
- ▶ De modo geral, não existe certeza sobre a correção das conclusões científicas; no entanto, os métodos estatísticos permitem determinar a margem de erro associada às conclusões, com base no conhecimento da variabilidade observada nos resultados.

# Uma definição

“Estatística é a ciência de dados”<sup>1</sup>.

---

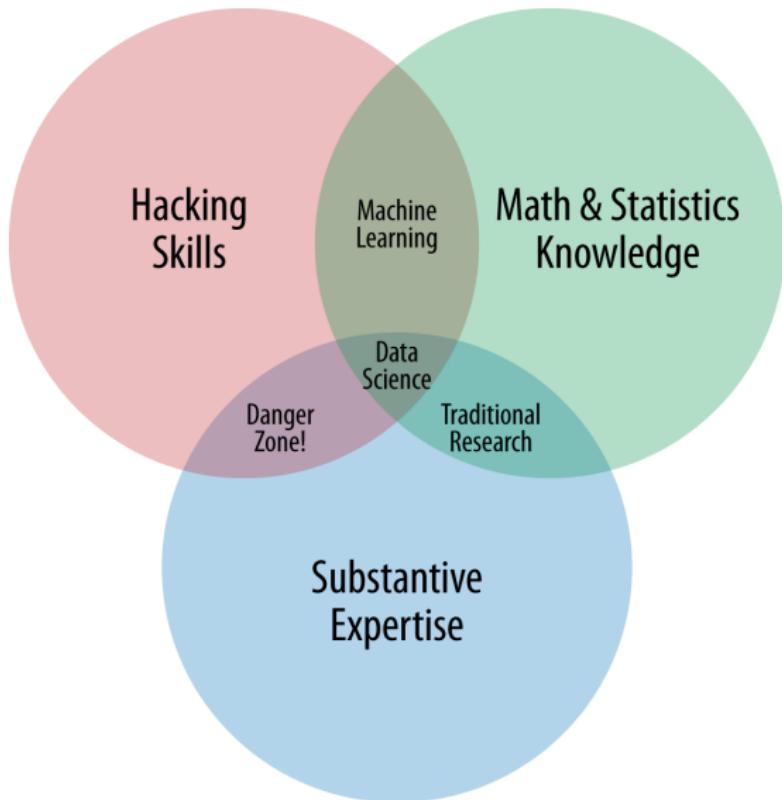
<sup>1</sup>Montgomery, D., Runger, G. Estatística aplicada e probabilidade para engenheiros. LTC, 2016.

# A Ciência de dados

# Data scientist



# Ciência dos dados



# Modern data scientist

## MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



## PROGRAMMING & DATABASE

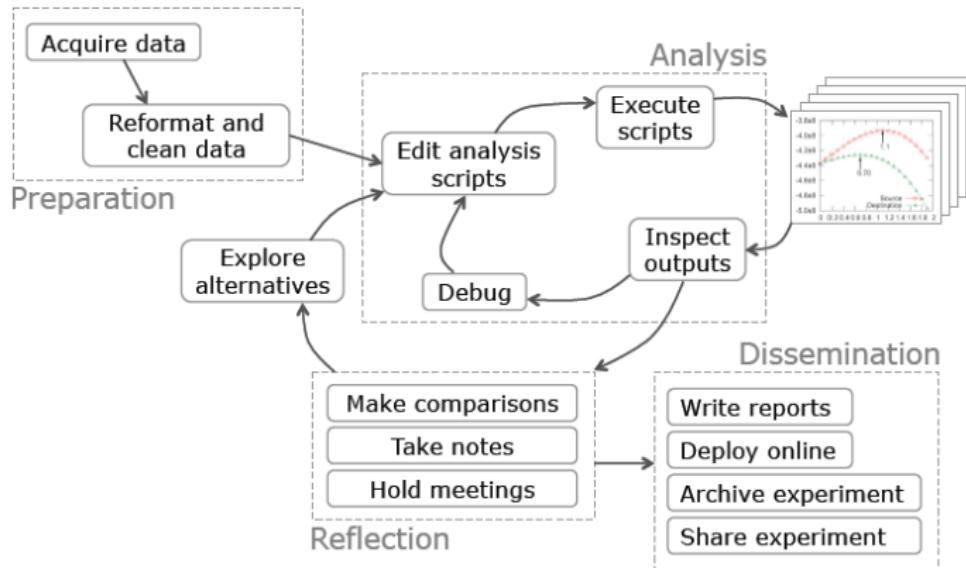
- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

## COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

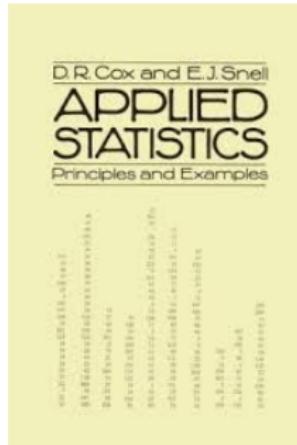
# Ciéncia dos dados

## Fluxo de trabalho



# Estatística . . . (ou ciéncia dos dados?)

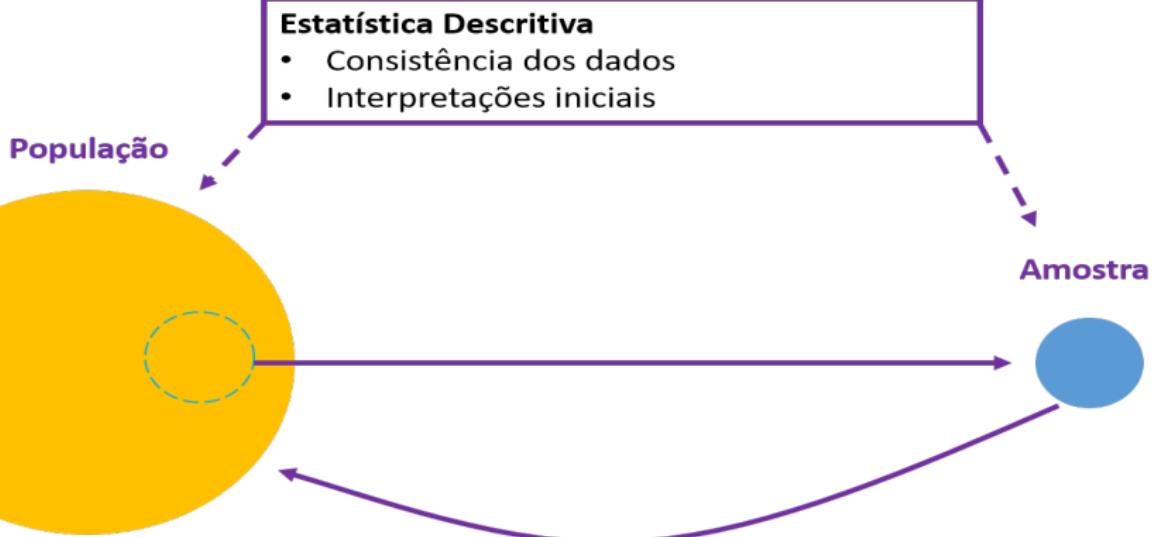
## Fases de análise



- 1. Manipulação inicial** dos dados.
  - ▶ Limpeza dos dados.
  - ▶ Criação, transformação e recodificação de variáveis.
- 2. Análise preliminar.**
  - ▶ Conhecimento dos dados, identificação de outliers, investigação preliminar.
- 3. Análise definitiva.**
  - ▶ Disponibiliza a base para as conclusões.
- 4. Apresentação das conclusões** de forma precisa, concisa e lúcida.

# A Estatística Descritiva

# A Estatística e suas divisões



# Uma definição

- ▶ A **Estatística Descritiva** corresponde aos procedimentos relacionados com a coleta, elaboração, tabulação, análise, interpretação e apresentação dos dados.
- ▶ Isto é, inclui as técnicas que dizem respeito à sintetização e à descrição de dados numéricos.
- ▶ Tais métodos tanto podem ser gráficos como envolver análise computacional.

## Próxima aula

- ▶ Conceitos básicos de Estatística.

# Por hoje é só! Sejam todos bem-vindos!

