

# MAT02018 - Estatística Descritiva

## Conceitos Básicos

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



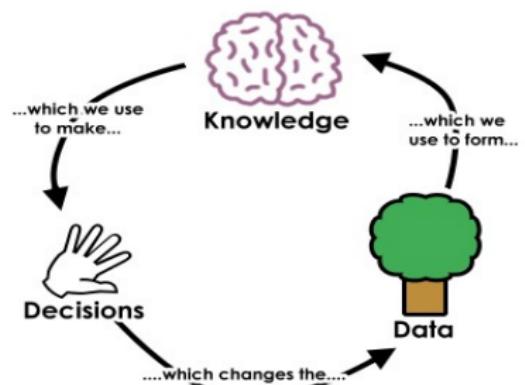
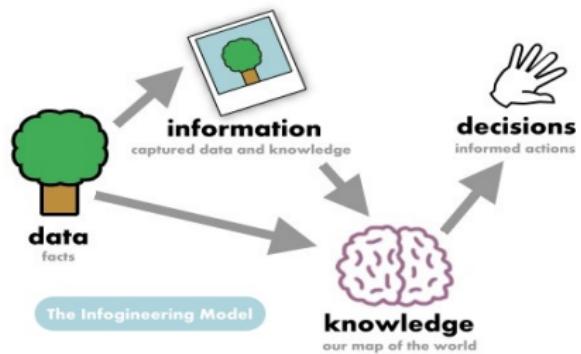
# Introdução

# Dados ↵ Conhecimento

- ▶ Em alguma fase de seu trabalho, o pesquisador depara-se com o problema de **analisar e entender** um **conjunto de dados** relevante ao seu particular objeto de estudos.
- ▶ Ele necessitará trabalhar os dados para **transformá-los em informações**, para compará-los com outros resultados, ou ainda para **julgar sua adequação a alguma teoria**.

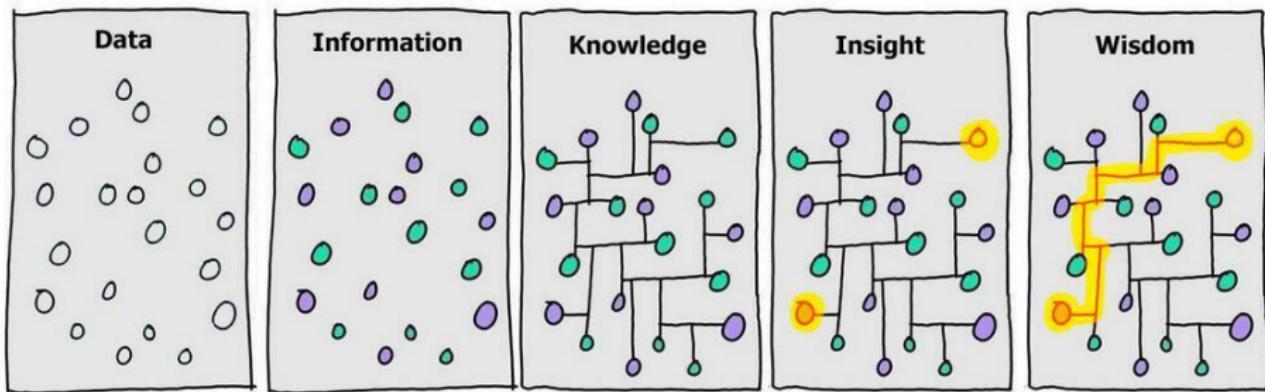
# Dados ↵ Conhecimento

## Uma representação



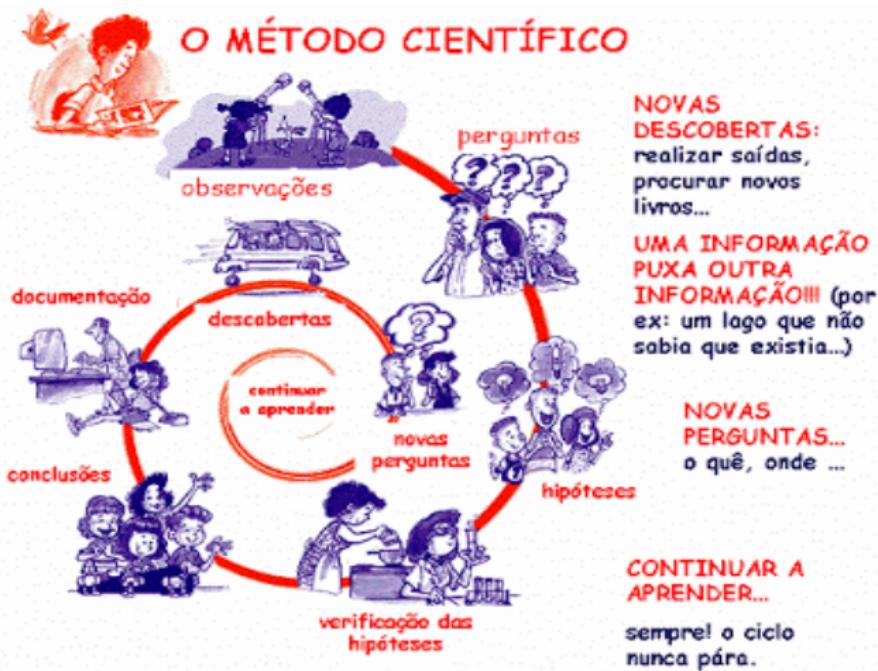
# Dados $\rightsquigarrow$ Conhecimento

Uma representação mais ousada!



# O Método Científico

- De modo bem geral, podemos dizer que a essência da Ciência é a **observação** e que seu objetivo básico é a **inferência**.



# O Método Científico

- De modo bem geral, podemos dizer que a essência do Aprendizado (da Evolução).



# Conceitos básicos

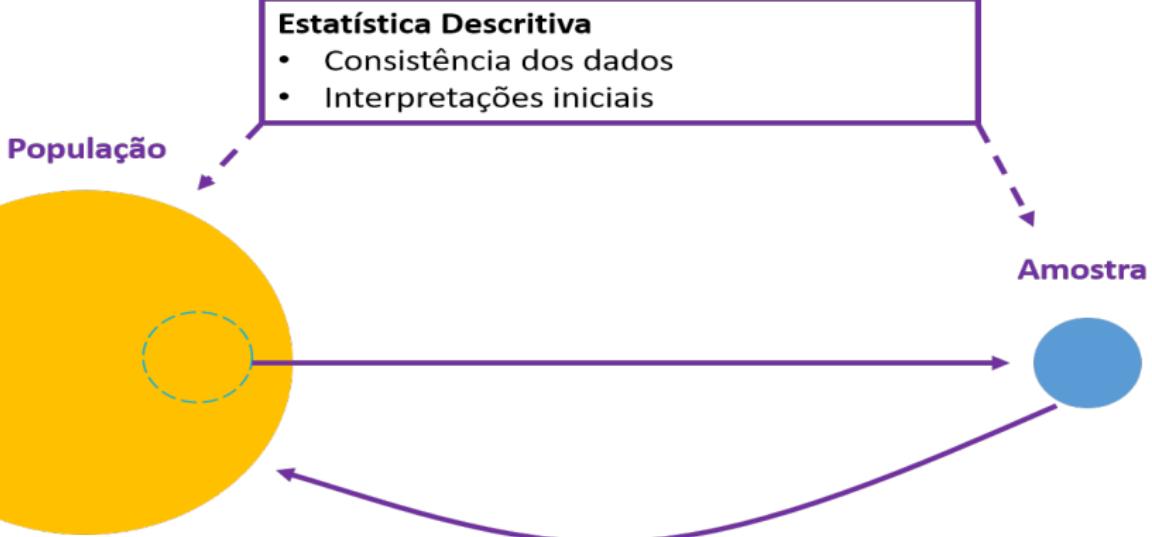
# O que é a estatística?

- ▶ A **estatística**<sup>1</sup> é a ciência que tem por objetivo orientar a coleta, o resumo, a apresentação, a análise e a interpretação de dados.
- ▶ Podem ser identificadas duas grandes áreas de atuação desta ciência:
  - ▶ a **estatística descritiva**, envolvida com o resumo e a apresentação dos dados.
  - ▶ a **estatística inferencial**, que ajuda a concluir sobre conjuntos maiores de dados (populações) quando apenas partes desses conjuntos (as amostras) foram estudadas.

---

<sup>1</sup>Do grego *statistós*, de *statízo*, “estabelecer”, “verificar”, acrescido do sufixo *ica*.

# O que é a estatística?



# O que é a estatística descritiva?

- ▶ 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.
- ▶ Estas técnicas podem ser utilizadas em pelo menos dois contextos
  - ▶ Análise da **consistência dos dados**.
  - ▶ **Análise Exploratória de Dados** (*Exploratory Data Analysis - EDA*)<sup>2</sup>.
- ▶ Tais métodos tanto podem ser gráficos como envolver análise computacional.

---

<sup>2</sup>Tukey, J. W. *Exploratory data analysis*, Reading: Addison-Wesley, 1977.

# Estatística descritiva: alguns exemplos

## Descriptive Statistics

**tobacco**

N: 1000

	age	BMI	cigs.per.day	samp.wgts
<b>Mean</b>	49.60	25.73	6.78	1.00
<b>Std.Dev</b>	18.29	4.49	11.88	0.08
<b>Min</b>	18.00	8.83	0.00	0.86
<b>Q1</b>	34.00	22.93	0.00	0.86
<b>Median</b>	50.00	25.62	0.00	1.04
<b>Q3</b>	66.00	28.65	11.00	1.05
<b>Max</b>	80.00	39.44	40.00	1.06
<b>MAD</b>	23.72	4.18	0.00	0.01
<b>IQR</b>	32.00	5.72	11.00	0.19
<b>CV</b>	0.37	0.17	1.75	0.08
<b>Skewness</b>	-0.04	0.02	1.54	-1.04
<b>SE.Skewness</b>	0.08	0.08	0.08	0.08
<b>Kurtosis</b>	-1.26	0.26	0.90	-0.90
<b>N.Valid</b>	975.00	974.00	965.00	1000.00
<b>Pct.Valid</b>	97.50	97.40	96.50	100.00

# Estatística descritiva: alguns exemplos

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
<b>F</b>	489	50.00	50.00	48.90	48.90
<b>M</b>	489	50.00	100.00	48.90	97.80
<b>&lt;NA&gt;</b>	22			2.20	100.00
<b>Total</b>	1000	100.00	100.00	100.00	100.00

# Estatística descritiva: alguns exemplos

População/1 milhão ● 250 ● 500 ● 750 ● 1000 ● 1250    Continente ● Africa ● Americas ● Asia ● Europe ● Ocea



# Unidades experimentais e observacionais

- ▶ **Unidade experimental** ou **unidade de observação** é a menor unidade a fornecer informação.
  - ▶ Ex: alunos, pacientes, animais, plantas, carros, hospitais, escolas, cidades, universidades, países, tweets, etc.

## Crash course de inferência causal

### Qual o melhor tratamento para o choque séptico?

Dois tipos de estudo podem ser conduzidos para responder a esta questão de pesquisa:

1. Em um **experimento aleatorizado** (*randomized trial*), uma moeda justa é lançada repetidamente para designar o tratamento de cada paciente.
2. Um **estudo observacional** é uma investigação empírica em que o objetivo é elucidar relações de causa e efeito, em que não é factível o uso de experimentação controlada, no sentido de ser capaz de impor procedimentos ou tratamentos cujos os efeitos se deseja descobrir.

## Experimentos: exemplo

- ▶ “O chá servido sobre o leite parecia ficar com gosto diferente do que apresentava ao receber o leite sobre ele”<sup>3</sup>.



---

<sup>3</sup>Salsburg, D. *Uma senhora toma chá ... como a estatística revolucionou a ciência no século XX*, Zahar, 2009.

## Estudos observacionais: exemplo

- ▶ “O Ministério da Saúde adverte: fumar pode causar câncer de pulmão”<sup>4</sup>.



<sup>4</sup>Salsburg, D. *Uma senhora toma chá ... como a estatística revolucionou a ciência no século XX*, Zahar, 2009.

# Exercício

1. Elabore uma questão de pesquisa de seu interesse (anote a sua questão em algum lugar).
2. Discuta a respeito da sua questão de pesquisa com os colegas.

# Dados e variáveis

## Dados

São as informações obtidas de uma unidade experimental ou observacional.

- **Ex:** “Vitor tem 25 anos e é fumante”. Os dados são “25 anos” e “fumante”.

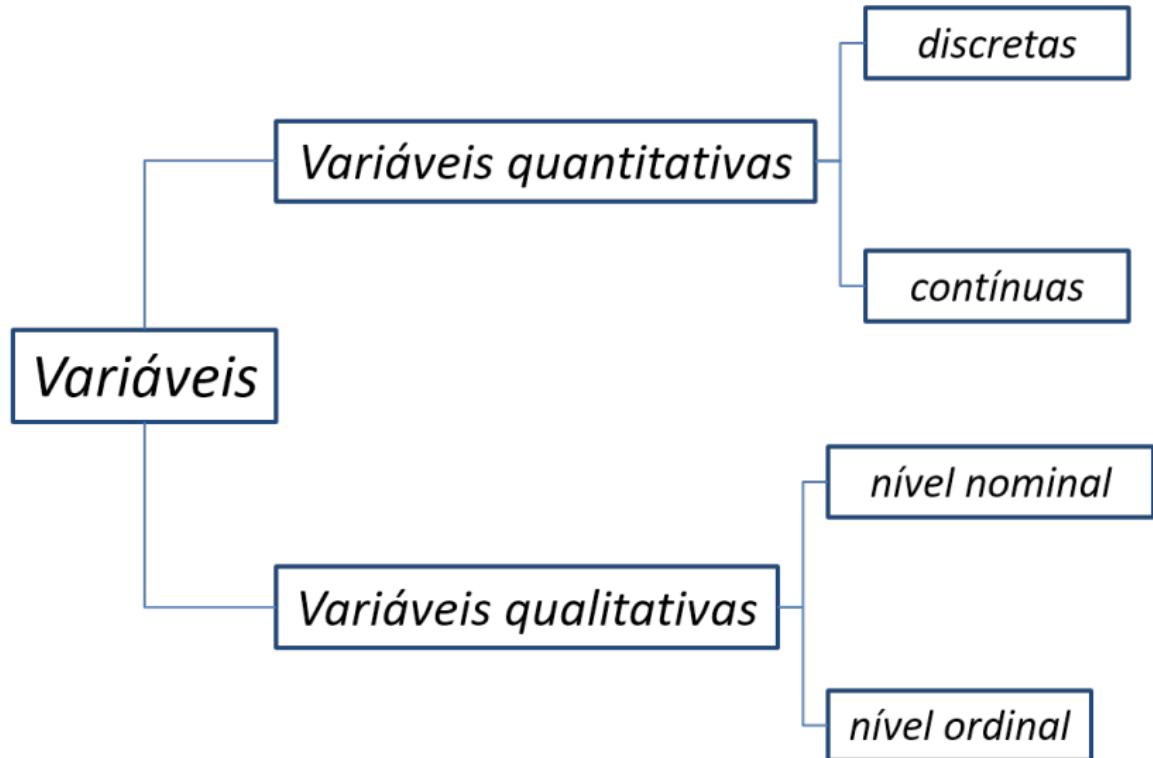
## Variável

É toda característica que, observada em uma unidade (experimental ou observacional), pode variar de um indivíduo para outro.

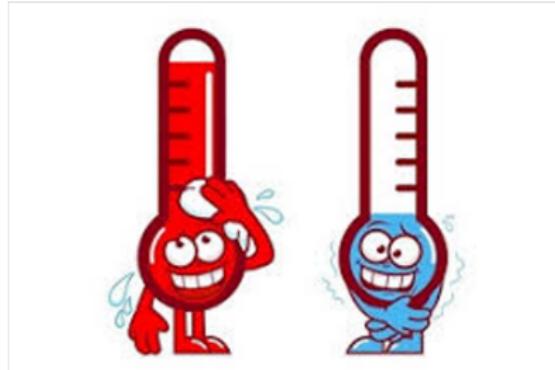
- **Ex:** idade, sexo, altura, nível de hemoglobina no sangue, espaçamento entre plantas, doses de um medicamento, tipo de medicamento, cultivares, número de caracteres, velocidade da rede, tempo gasto na rede social, nível de monóxido de carbono em emissões do escape de automóveis, etc.

É importante **identificar que tipo de variável** está sendo estudada, uma vez que são recomendados **procedimentos estatísticos diferentes** em cada situação.

# Tipos de variáveis



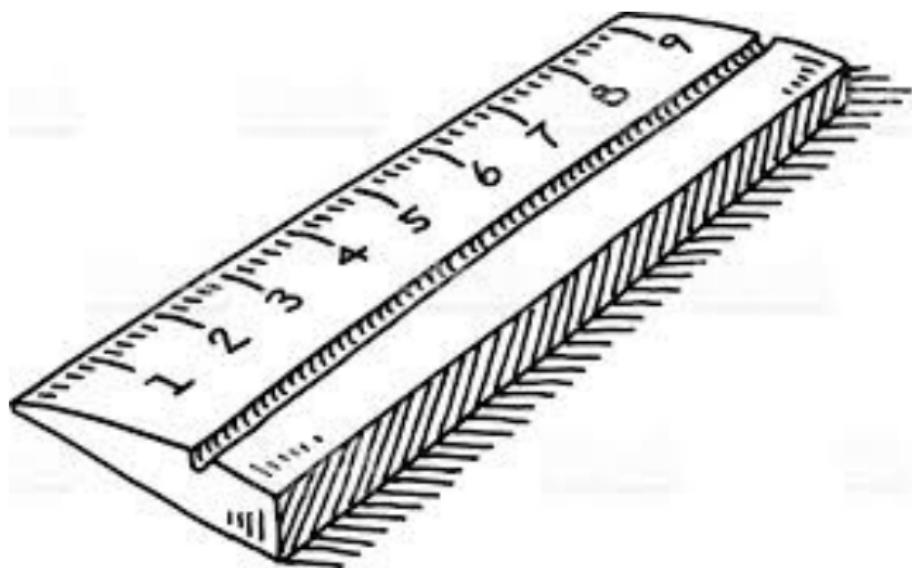
# Variáveis quantitativas



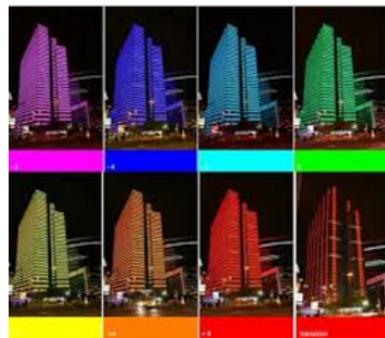
# Variáveis quantitativas discretas



# Variáveis quantitativas contínuas



# Variáveis qualitativas



## Variáveis qualitativas ordinais



# Variáveis qualitativas nominais



## Exemplos (1)



# Exemplos (1)

## Variáveis quantitativas

- ▶ 3 andares
- ▶ 14,85 metros de altura

## Variáveis qualitativas

- ▶ Multicolorido
- ▶ Cheira “bem”

## Exemplos (2)



# Exemplos (2)

## Variáveis quantitativas

- ▶ 4 integrantes
- ▶ 56 anos

## Variáveis qualitativas

- ▶ Inglaterra
- ▶ Rock

# Exercício

1. Com base na questão de pesquisa elaborada no exercício anterior, liste variáveis que você teria interesse em coletar e analisar para responder a sua questão de pesquisa.
2. Classifique as variáveis de acordo com a classificação discutida anteriormente.
3. Discuta a respeito das suas variáveis com os colegas.

# População

- ▶ **População ou universo:** esse termo é usado em estatística com um sentido bem mais amplo do que na linguagem coloquial.
- ▶ É entendido aqui como o **conjunto de todos os elementos** que apresentam uma ou mais características **em comum**.
- ▶ **Exemplo 1:** a população de colegiais de oito anos de Belo Horizonte.
  - ▶ Estes colegiais têm em comum a idade e o local onde vivem.
- ▶ **Exemplo 2:** a população de indústrias brasileiras.
  - ▶ Estas indústrias têm em comum o fato de que foram criadas no Brasil.
- ▶ Este conjunto por vezes é denominado por *U* (de **conjunto universo**).
- ▶ **O tamanho da população** é a sua quantidade de elementos, que anotamos por *N*.
- ▶ Uma população pode ser **finita** (limita em tamanho;  $N < \infty$ ) ou **infinita** ( $N = \infty$ ).
  - ▶ **Exemplo de pop. finita:** torcedores do São Raimundo de Santarém, residentes de Porto Alegre.
  - ▶ **Exemplo de pop. infinita:** equipamentos (de um certo tipo) fabricados em série.

# Censo e amostra

- ▶ Quando o estudo é realizado com toda a população de interesse, chamemos este estudo de **censo**.
- ▶ Por motivos de tempo, custo, logística, entre outros, geralmente não é possível realizar um censo.
  - ▶ Nestes casos, estudamos apenas uma parcela da população, que chamamos de **amostra**.

## Censo vs. amostra

À primeira vista, uma coleta de dados realizada em toda a população é preferível a uma realizada apenas numa parte da população. Na prática, entretanto, o oposto é frequentemente verdadeiro porque:

1. Um censo é impossível quando a população é infinita.
2. Os ensaios (testes) podem ser destrutivos (**como nos testes de segurança dos carros**).
3. Rapidez: estudar toda a população pode despender de muito tempo, não sendo compatível com a urgência do estudo (**como quando estudamos os casos de um surto de uma nova doença**).

Para uma consideração mais completa ver Vargas (2000)<sup>a</sup>.

<sup>a</sup>Vargas, J. B. *Estatística: uma linguagem para dialogar com a incerteza*, Cadernos IADE, Rio de Janeiro, Série B, 2000.

# Amostra

- ▶ **Amostra** é qualquer fração de uma população.
  - ▶ Como sua finalidade é representar a população, deseja-se que a amostra escolhida apresente as mesmas características da população de origem, isto é, que seja uma amostra “**representativa**” ou “**não-tendenciosa**”.
- ▶ Tanto o número de indivíduos selecionados para a amostra quanto a técnica de seleção são extremamente importantes para que os resultados obtidos no estudo sejam generalizados para a população.

# Amostra representativa



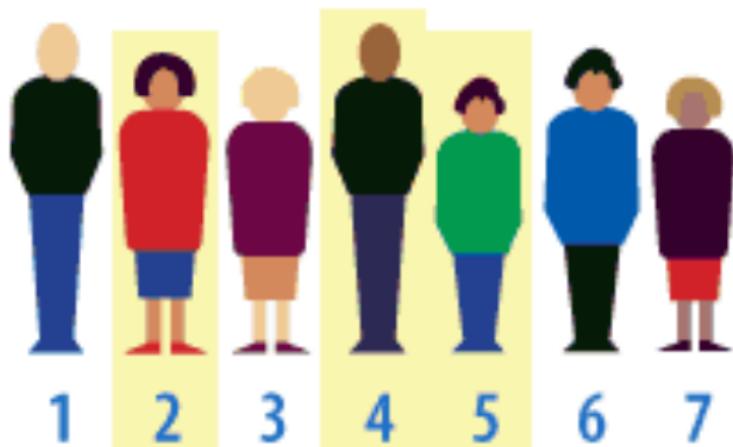
- ▶ Ver a discussão sobre **representatividade da amostra** na apresentação do **Prof. Chris Fonnesbeck**.

# Amostragem

- ▶ A seleção da amostra pode ser feita de várias maneiras.
- ▶ Esta dependerá:
  - ▶ Do grau de conhecimento que temos da população.
  - ▶ Da quantidade de recursos disponíveis.
- ▶ A seleção da amostra tenta fornecer um subconjunto de valores o **mais parecido possível** com a população que lhe dá origem.
  - ▶ **Amostra representativa** da população.

# Amostra aleatória simples

- ▶ A amostragem mais usada é a **amostra casual simples** (ou aleatória simples).
  - ▶ Os indivíduos (unidades) da amostra são selecionados ao acaso, **com** ou **sem reposição**.



# Amostra estratificada

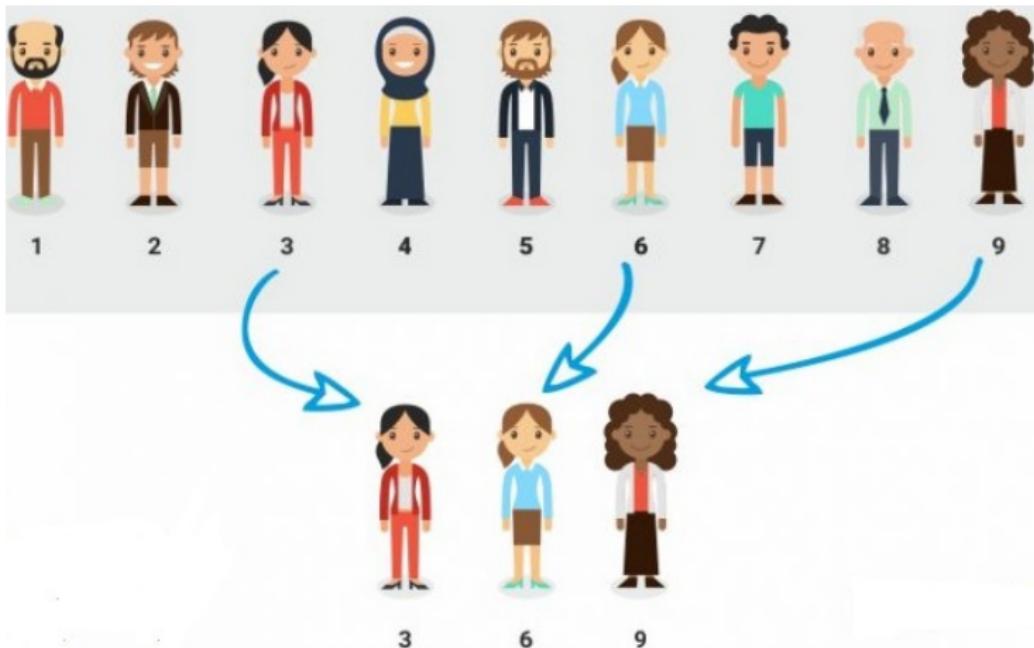
- Eventualmente, se tivermos informações adicionais a respeito da população de interesse, podemos utilizar outros esquemas de amostragem mais sofisticados.

- **Amostragem estratificada**



# Amostra sistemática

- Em outros casos, pode existir uma relação numerada dos itens da população que nos permitiria utilizar a chamada **amostragem sistemática** em que selecionamos os indivíduos de forma pré-determinada.



# Amostragem

- ▶ Outros esquemas de amostragem poderiam ser citados e todos fazem parte da chamada **teoria da amostragem**, cujos detalhes não serão aprofundados.

# Parâmetros, estatísticas e estimativas

- ▶ **Parâmetro** é um valor que resume, na população, a informação relativa a uma variável.
  - ▶ Ex: média populacional, prevalência populacional, coeficiente de variação populacional, taxa de mortalidade populacional, etc.
- ▶ **Estatística** (além de ser o nome da ciência/área do conhecimento) é a denominação dada a uma quantidade, calculada com base nos elementos de uma amostra, que descreve a informação contida nesse conjunto de dados.
  - ▶ Ex: A média, a porcentagem, o desvio padrão, o coeficiente de correlação, calculados em uma amostra, são estatísticas.

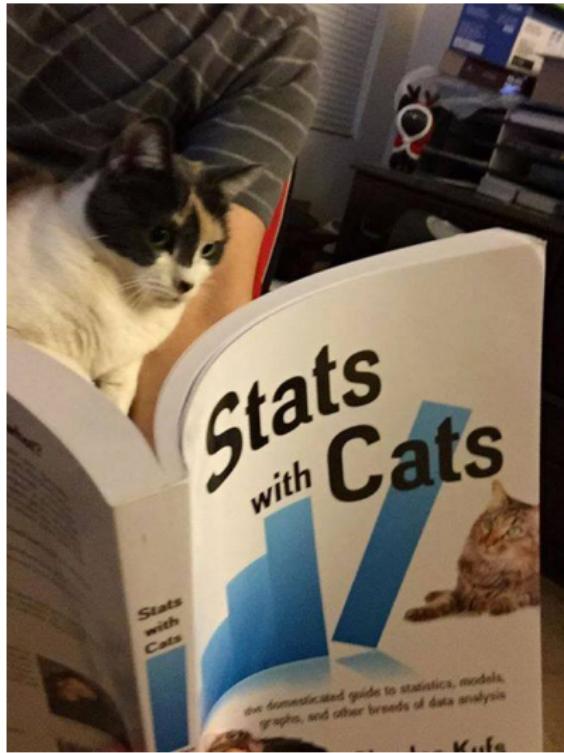
# Parâmetros, estatísticas e estimativas

- ▶ Os parâmetros são difíceis de se obter, pois implicam o estudo de toda a população e costumam ser substituídos por valores calculados em amostras representativas da população-alvo.
  - ▶ Se tivesse sido examinada uma amostra de 10 estudantes matriculados na disciplina MAT02218, e 40% fossem torcedores do América Mineiro, esse valor constituiria uma estimativa do parâmetro “percentual de torcedores do América Mineiro matriculados naquela disciplina”.
- ▶ A **estimativa** é um valor numérico de uma estatística, usado para realizar inferências sobre o parâmetro.
  - ▶ Da mesma forma, o valor numérico da média para a estatura desses 10 alunos, digamos 173 cm, é uma estimativa para a média de altura populacional.
- ▶ **P:** neste exemplo, quem é a população (alvo)?

## Próxima aula

- ▶ Organização dos dados

## Para casa



- ▶ Conhecer o Moodle da disciplina.
- ▶ Ler os Cap. 1 e 2 de "Estatística Descritiva I" de Fernandez.

# Por hoje é só! Bons estudos!

