

MAT02018 - Estatística Descritiva

Conceitos Básicos

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

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

Porto Alegre, 2022

Conceitos básicos

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”.

Dados e variáveis

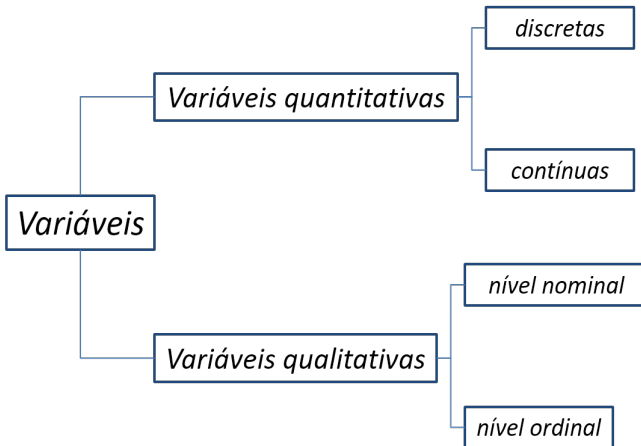
Variável

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.

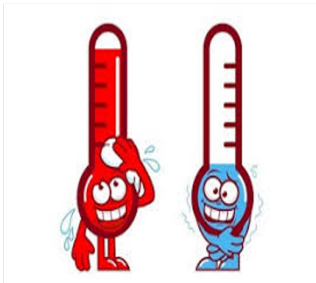
Tipos de variáveis

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



Variáveis quantitativas

- ▶ A **variável quantitativa** é expressa por números.



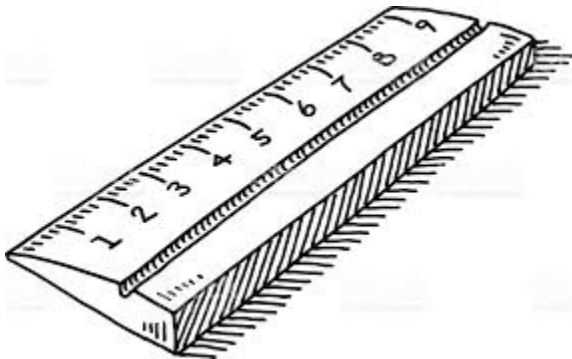
Variáveis quantitativas discretas

- ▶ A **variável discreta** resulta do processo de contagem.



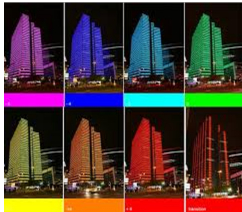
Variáveis quantitativas contínuas

- ▶ A **variável contínua** resulta do processo de medição.



Variáveis qualitativas

- A **variável quantitativa** é expressa por palavras.



Variáveis qualitativas ordinais

- ▶ A **variável ordinal** tem duas ou mais categorias que são, necessariamente, organizadas segundo uma lógica.



Variáveis qualitativas nominais

- ▶ A **variável nominal** tem duas ou mais categorias, que podem ser apresentadas em qualquer ordem.



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
- ▶ 60 anos

Variáveis qualitativas

- ▶ Inglaterra
- ▶ Rock

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.

População

- ▶ 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 e 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 despende 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)¹.

¹Vargas, J. B. *Estatística: uma linguagem para dialogar com a incerteza*, Cadernos de matemática e estatística. 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



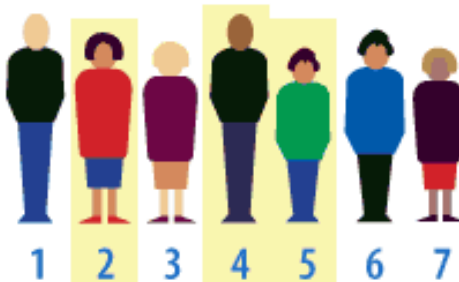
- Ver a discussão sobre **representatividade da amostra** na apresentação do **Prof. Chris Fonnesbeck**.

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**.



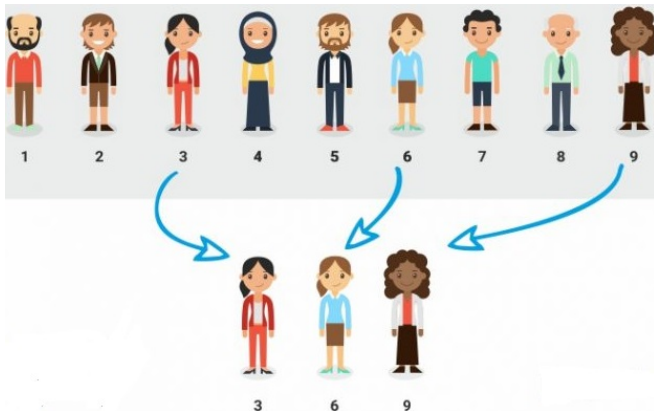
Amostra estratificada

- ▶ 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 de torcedores do América Mineiro, esse valor constituiria uma estimativa do parâmetro “percentual de torcedores do América Mineiro matriculados naquela disciplina”.

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)?

Para casa

- ▶ Assistir o vídeo: **Statistical Thinking for Data Science** (no Moodle ou em <https://youtu.be/TGGGDpb04Yc>; você pode configurar o vídeo para apresentar legendas traduzidas para o português).
 - ▶ Leia a notícia sobre **quedas de gatos** na íntegra (Moodle).

Próxima aula

- ▶ Fases do levantamento estatístico.

Bons estudos!

