

EPI02 - Fundamentos de Bioestatística

Conceitos e definições

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

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

Porto Alegre, 2024



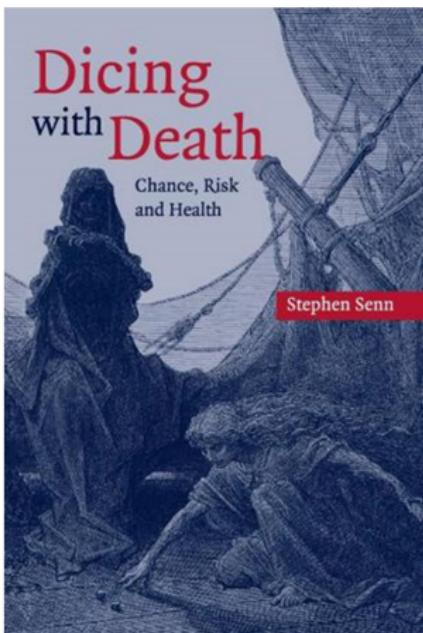
Relembrando

A mensuração do Diabetes

- ▶ O que é necessário para medir a ocorrência do diabetes em uma população ou em um grupo dentro da população?
 - ▶ Quais os elementos necessários?
 - ▶ Como podemos caracterizar uma população?
 - ▶ Com relação ao diabetes, como podemos proceder com a medição do diabetes na população brasileira?
 - ▶ Vamos medir a população de forma completa?
 - ▶ Qual o instrumento de medida será adotado?
 - ▶ Existe uma “amostra adequada” (ou amostras não adequadas) para análise?
 - ▶ Como selecionar uma amostra da população?
 - ▶ Se repetirmos o processo de seleção de uma amostra, o que é esperado? Os mesmos indivíduos irão compor amostras distintas? No que isto impacta em nossos resultados?

A Estatística e o seu papel

Dados ↵ Conhecimento



"Os estatísticos estão engajados em uma luta exaustiva, mas emocionante, com o maior desafio que a filosofia coloca à ciência: como traduzimos informação em conhecimento? A estatística nos diz como avaliar as evidências, como planejar experimentos, como transformar dados em decisões, quanto crédito deve ser dado a quem, o quê e por quê, como calcular as chances e quando aproveitá-las."

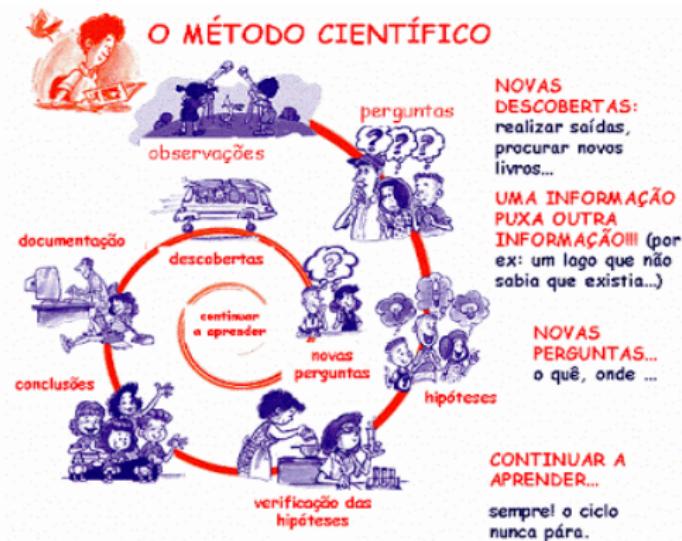
Stephen Senn – Dicing with Death, 2003

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**.

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**.
 - ▶ Os cientistas (sociais ou físicos) geralmente fazem uso do **método científico** nas suas tentativas de compreender o mundo.



O Método Científico

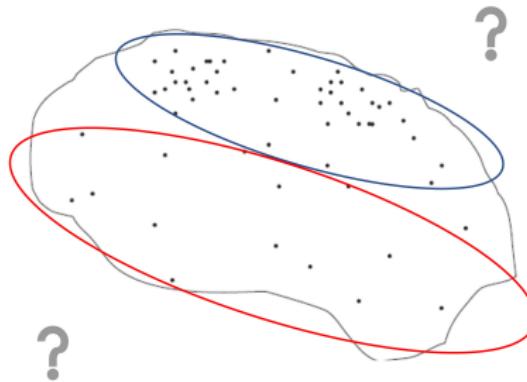
Suponha que estamos interessados em descrever e explicar o **padrão dos casos de câncer** em uma área metropolitana.

1. Observação/registro dos casos;
2. Descrição do padrão por meio da apresentação dos casos em um mapa.



O Método Científico

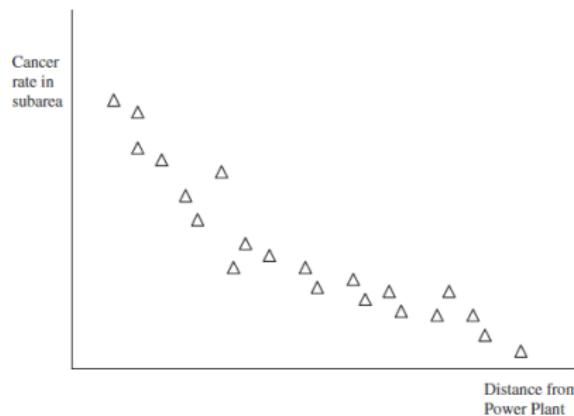
3. O que explica o resultado observado?



O Método Científico

4. Formulação de hipóteses/modelos explicativos.

(Exemplo) *Hipótese:* o padrão de casos de câncer está relacionado à distância das usinas de energia locais.



O Método Científico

5. Os dados fornecem evidências para avaliarmos as hipóteses formuladas.
 6. A partir desta avaliação, modelos e hipóteses podem ser reconsiderados, ou conclusões teóricas podem ser elaboradas.
- Os métodos estatísticos ocupam um papel central no método científico, como visto no exemplo dos casos de câncer, pois nos permitem **sugerir** e **testar** hipóteses.

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 apenas coletariam dados (e não os analisam) para uma profissão que busca parceria com cientistas em todas as etapas da investigação, do planejamento à análise.

- ▶ O indício mais remoto do que tenha sido um censo parece ter ocorrido na China, por volta de **2.238 a.C.**, quando o imperador Yao mandou fazer a **contagem da população** e das **lavouras cultivadas**.

Uma perspectiva histórica

- ▶ 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.

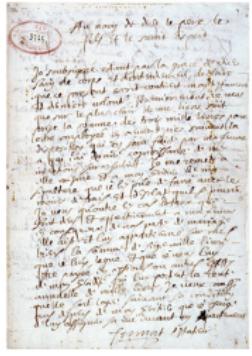
Uma perspectiva histórica

Consolidação dos estados nacionais (e dos seus órgãos da burocracia), Industrialização, Revolução Científica.

- ▶ Houve, no período, um **aperfeiçoamento da teoria das probabilidades e estatística inferencial**, com a aplicação da **teoria da distribuição normal** por **Aldophe Quetelet (1796-1874)** no estudo de diversas características das populações humanas: altura, peso, natalidade, mortalidade, renda, etc.
- ▶ A **teoria da estimativa** na estatística também recebeu contribuições nos procedimentos de preencher espaços vazios entre os dados coletados, por meio das **técnicas de previsão, retrospectiva e interpolação**.

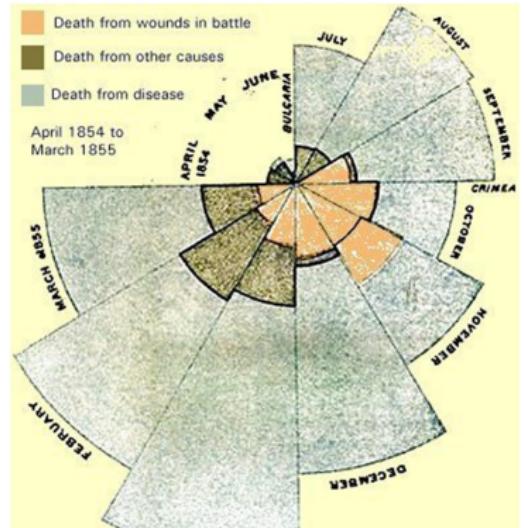
Uma perspectiva histórica

Pascal-Fermat (Séc. XVII)



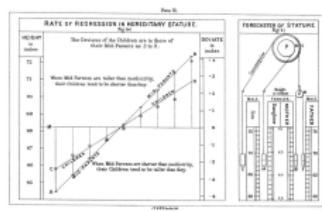
Uma perspectiva histórica

Florence Nightingale (Séc. XIX)



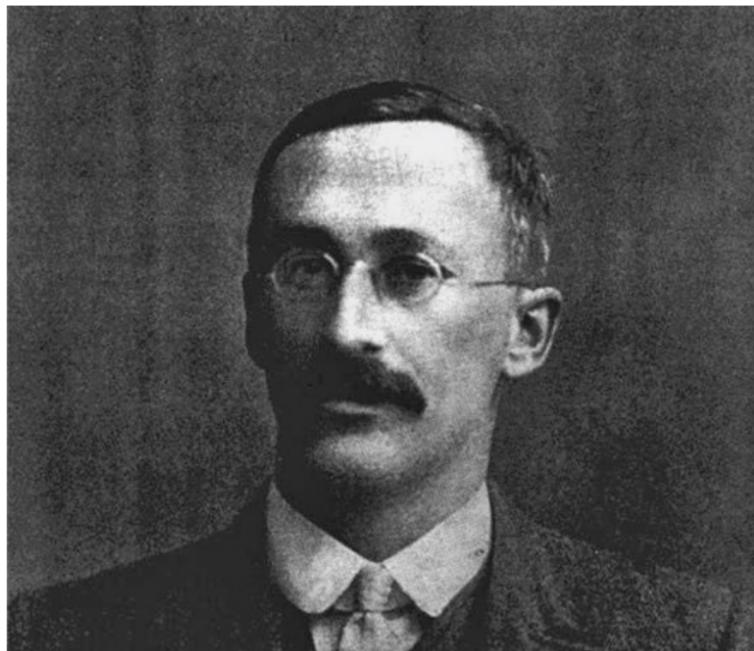
Uma perspectiva histórica

Galton-Pearson (Séc. XIX e XX)



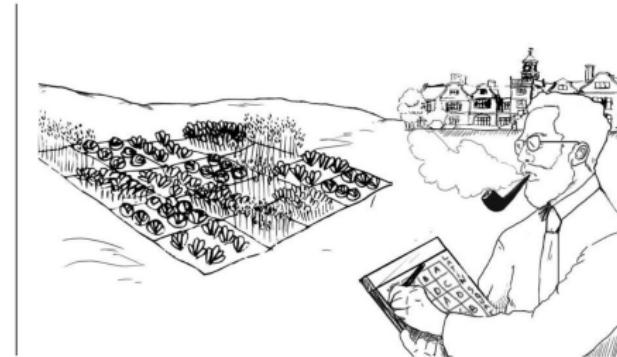
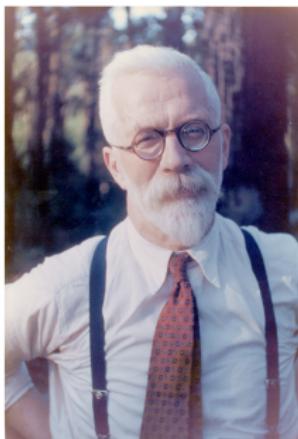
Uma perspectiva histórica

William S. Gosset, o *Student* (Séc. XX)



Uma perspectiva histórica

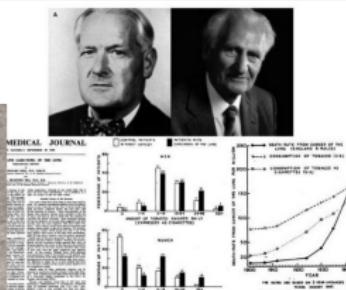
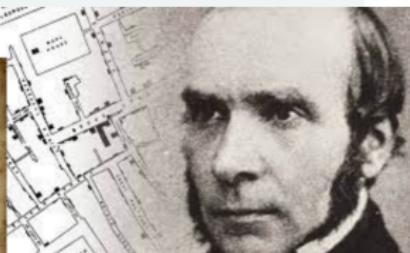
Ronald A. Fisher (Séc. XX)



Uma perspectiva histórica

FORMAÇÃO HISTÓRICA DA EPIDEMIOLOGIA

Fundamenta-se em três eixos: um **saber clínico** naturalizado, racionalista e moderno; uma base metodológica, a **Estatística**; e um substrato político-ideológico, a **Medicina Social**.²

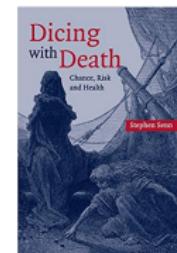
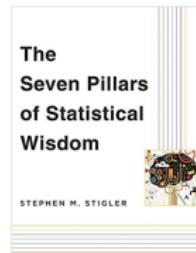
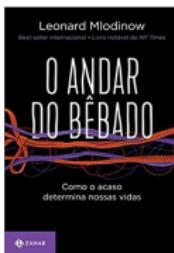
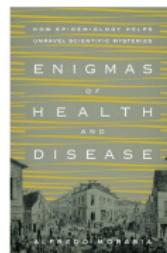
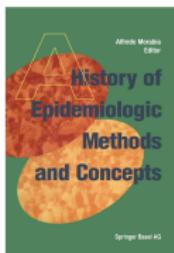


² ALMEIDA FILHO, N. de. Uma Breve História da Epidemiologia. In: Rouquayrol, M. Z.; Almeida Filho, N. de. Epidemiologia & Saúde. 5^a ed. Rio de Janeiro: MEDSI, 1999. p. 01-13.

Uma perspectiva histórica

Essa história continua!

Para saber mais, veja:



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.

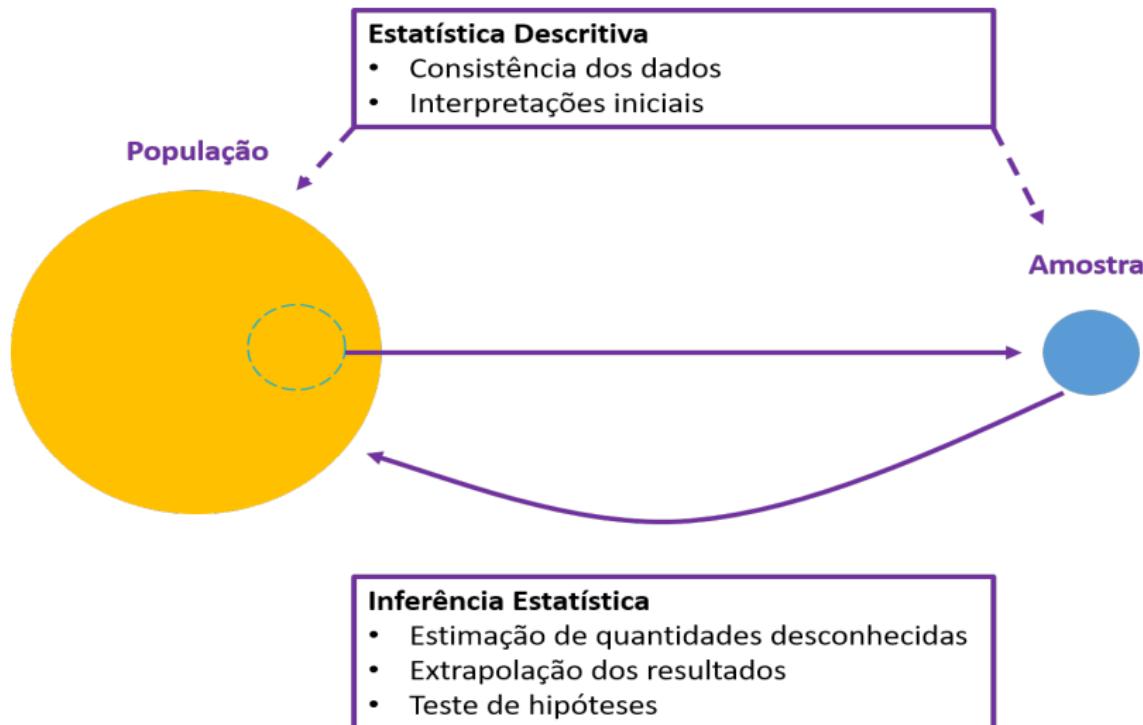
O que é a estatística?

A **estatística**¹ 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.

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

O que é a estatística?



A Estatística Descritiva

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*)².
- ▶ Tais métodos tanto podem ser gráficos como envolver análise computacional.

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

Alguns exemplos

Descriptive Statistics

tobacco

N: 1000

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

Conceitos básicos

Unidades experimentais e observacionais

Unidade

(experimental ou observacional) é a menor unidade a fornecer informação.

- ▶ **EX.:** alunos, pacientes, participantes de um estudo, animais, plantas, carros, hospitais, escolas, cidades, universidades, países, *tweets*, etc.

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.

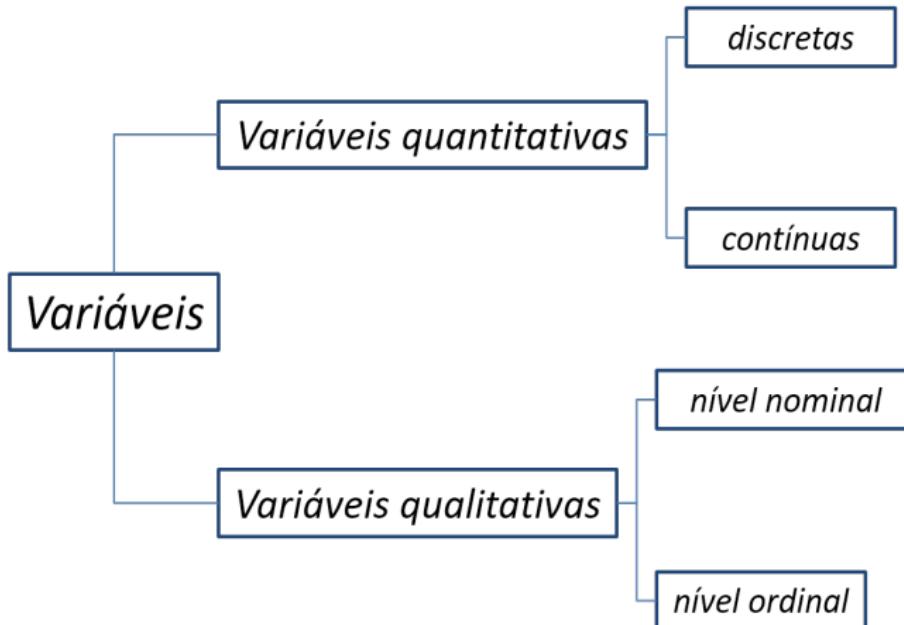
- ▶ **EX.:** idade, sexo, altura, nível de hemoglobina no sangue, espaçamento entre plantas, doses de um medicamento, tipo de medicamento, cultivares, número de caracteres, velocidade da rede, tempo gasto na rede social, nível de monóxido de carbono em emissões do escape de automóveis, etc.

Dados e variáveis

- ▶ Em geral, na literatura estatística, as variáveis são denotadas por letras latinas maiúsculas:
 - ▶ $X, Y, Z;$
 - ▶ $X_1, X_2, \dots, X_p.$
- ▶ Já os valores que as variáveis assumem são denotados, de forma genérica, por letras latinas minúsculas:
 - ▶ $X = x, Y = y, Z = z.$

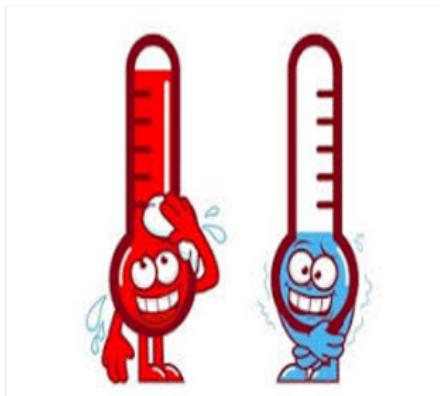
Tipos de variáveis

- 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.



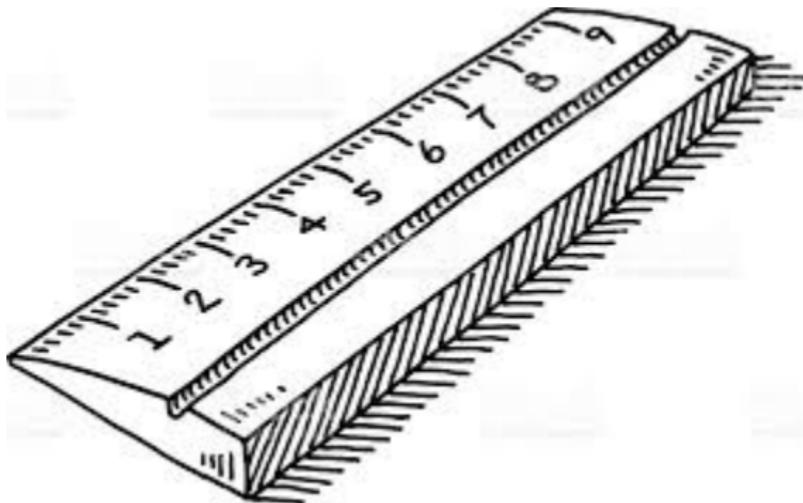
Variáveis quantitativas discretas

- ▶ A variável discreta resulta do processo de contagem.



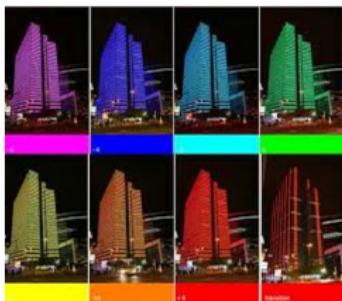
Variáveis quantitativas contínuas

- ▶ A variável contínua resulta do processo de medição.



Variáveis qualitativas

- A variável qualitativa 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.



Variáveis dicotômicas

Um “subtipo” de variável, muito comum nas investigações epidemiológicas, é a **variável dicotônica** (ou variável binária).

- ▶ Quando uma variável qualitativa apresenta apenas dois níveis (duas categorias), esta é chamada de variável dicotônica.
- ▶ **Ex.:** possuir níveis altos de catecolaminas (**Sim** ou **Não**).

Usualmente, se codificam os dois níveis da variável dicotônica por **0 (não/ausente)** e **1 (sim/presente)**.

- ▶ Na literatura ainda veremos *expostos* (E) e *não-expostos* (\bar{E}) e *casos* (C) e *não-casos (controles)* (\bar{C}).

Exercício (pequenos grupos)

1. Consulte o **questionário da PNS 2013** e, a partir das questões, apresente variáveis de acordo com a classificação discutida anteriormente.
 - ▶ Sugestão: utiliza os módulos M e N.
2. Discuta sobre a natureza do dado e a sua mensuração: dados originalmente contínuos podem ser categorizados na mensuração e/ou na análise? Ou seja, a partir da coleta de uma variável, podemos construir novas variáveis? Apresente exemplos destas situações.

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**.

- ▶ **EX.:** a população de adultos brasileiros.
 - ▶ Estes indivíduos têm em comum a nacionalidade brasileira e a idade ≥ 18 .

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:** os residentes de Porto Alegre em um momento específico.
 - ▶ **Exemplo de pop. infinita:** equipamentos (de um certo tipo) fabricados em série.

População

- ▶ A **população de estudo** é definida como **a parte identificável e acessível do conjunto (população-alvo)** para o qual idealmente se deseja a informação.
 - ▶ A distância entre ambas é decorrente quer das dificuldades da realidade, quer das limitações impostas pelo delineamento.

População

- ▶ **População alvo:** população sobre a qual se quer uma resposta (objetivo).
- ▶ **População de estudo:** população da qual a amostra é obtida.
- ▶ População alvo = População de estudo (**ideal**).
- ▶ População alvo \approx População de estudo (**comum; aceitável**).
- ▶ População alvo \neq População de estudo (**ruim**).

População

- ▶ O objeto dos estudos são sempre as populações, pois, somente assim, as conclusões dos trabalhos científicos não se restringem apenas às unidades neles estudadas.
- ▶ No entanto, como as populações são constituídas de um número muito grande de elementos, são estudadas por intermédio de alguns desses elementos, os quais constituirão o que se denomina uma **amostra**.

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. Um censo pode ser muito custoso.
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**).
4. É possível se aprender, com razoável confiança, sobre a população através de uma amostra.

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 (tamanho da amostra, n) quanto a técnica de seleção são extremamente importantes para que os resultados obtidos no estudo sejam generalizados para a população.

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.
- ▶ Por vezes, na literatura estatística, letras gregas são utilizadas para representar parâmetros:
 - ▶ $\theta, \mu, \alpha, \beta$, etc.
- ▶ Também podem ser utilizadas letras latinas (maiúsculas) e símbolos:
 - ▶ $\bar{X}, P, \text{Var}(X), DP(X), \text{Corr}(X, Y)$, etc.

Parâmetros, estatísticas e estimativas

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.
- ▶ Podem aparecer na literatura como letras latinas (minúsculas) e símbolos:
 - ▶ $\bar{x}, p, dp(x)$, etc.

Parâmetros, estatísticas e estimativas

Se consideramos que n indivíduos de uma amostra, com relação a uma variável quantitativa X , de tal forma que observamos (x_1, x_2, \dots, x_n) , então podemos tomar como exemplo as seguintes estatísticas:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n},$$

$$x_{(1)} = \min\{x_1, \dots, x_n\},$$

$$x_{(n)} = \max\{x_1, \dots, x_n\},$$

$$dp(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$$

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 do PPG em Epidemiologia da UFRGS, e 80% reportassem uma boa/muito boa percepção de saúde, esse valor constituiria uma estimativa do parâmetro “percentual de boa/muito boa percepção de saúde” na população de estudantes do PPG em Epidemiologia.

Parâmetros, estatísticas e estimativas

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.
- ▶ É muito comum o uso do “chapéu” para denotar uma estimativa com respeito a um parâmetro:
 - ▶ $\hat{\mu} = 173 \text{ cm}$ ou $\hat{X} = 173 \text{ cm}$.

Para casa



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

Próxima aula

- ▶ Delineamentos de pesquisa: estudos aleatorizados, não-aleatorizados e levantamentos
- ▶ Principais planos de amostragem probabilística
- ▶ Organização dos dados
- ▶ Descrição dos dados por resumos numéricos