

# ACH3553 - Estatística I

Curso de Gestão de Políticas Públicas

## Aula 2

Alexandre Ribeiro Leichsenring  
alexandre.leichsenring@usp.br

- 1 Organização dos dados
  - Distribuição de frequência de dados nominais
  - Comparação de distribuições
  - Distribuições de frequências agrupadas de dados intervalares
  - Proporções e porcentagens
  - Razões e taxas

- 1 Organização dos dados
  - Distribuição de frequência de dados nominais
  - Comparação de distribuições
  - Distribuições de frequências agrupadas de dados intervalares
  - Proporções e porcentagens
  - Razões e taxas

# Organização dos dados

- O que os cientistas sociais podem fazer para organizar a miscelânea de números brutos que coletam de seus objetos de estudos?
- O que podem fazer para transformar a massa de dados brutos em uma forma de fácil compreensão?
- 1º passo: construir distribuição de frequência em forma de tabela.

# Distribuição de frequência de dados nominais

## Exemplo

Suponha que estejamos estudando o desempenho das escolas de São Paulo no Enem.

Vamos examinar uma base de dados do R e construir tabelas de frequência para variáveis nominais.

Em particular, vamos analisar o número de escolas que oferecem EJA.

Acessar os dados em: <https://rstudio.cloud/project/448060>.

# Comparação de distribuições

## Exemplo

Suponha que em seguida se queira comparar se as escolas com indicador de nível socioeconômico (INSE) mais baixos são mais propensas a oferecer EJA.

Determinar a distribuição das escolas por oferecimento de EJA e INSE.

# Apresentação dos resultados

- Variáveis qualitativas nominais são rotuladas (e não admitem ordenação natural)
- Categorias de distribuições em termos nominais não precisam ser listadas nenhuma ordem particular (qualquer ordenação é aceitável)

Distribuição do estado civil: qualquer ordem é aceitável

| <b>Estado civil</b> | $n_i$ |
|---------------------|-------|
| Casado(a)           | 30    |
| Solteiro(a)         | 20    |
| Divorciado(a)       | 10    |
| Total               | 60    |
| Ok!                 |       |

| <b>Estado civil</b> | $n_i$ |
|---------------------|-------|
| Solteiro(a)         | 20    |
| Casado(a)           | 30    |
| Divorciado(a)       | 10    |
| Total               | 60    |
| Ok!                 |       |

| <b>Estado civil</b> | $n_i$ |
|---------------------|-------|
| Divorciado(a)       | 10    |
| Solteiro(a)         | 20    |
| Casado(a)           | 30    |
| Total               | 60    |
| Ok!                 |       |

# Apresentação dos resultados

- Variáveis qualitativas ordinais devem ser apresentadas na sua ordenação

Distribuição de frequências de atitudes em relação às cotas de entrada na USP pelo ENEM:

| Atitude                    | $n_i$ |
|----------------------------|-------|
| Concordo parcialmente      | 15    |
| Discordo totalmente        | 15    |
| Nem concordo, nem discordo | 30    |
| Discordo parcialmente      | 25    |
| Concordo totalmente        | 45    |
| Total                      | 100   |

Inadequado

| Atitude                    | $n_i$ |
|----------------------------|-------|
| Concordo totalmente        | 45    |
| Concordo parcialmente      | 15    |
| Nem concordo, nem discordo | 30    |
| Discordo parcialmente      | 25    |
| Discordo totalmente        | 15    |
| Total                      | 100   |

Adequado

# Distribuições de frequências agrupadas de dados intervalares

- Variáveis quantitativas (escalares) muitas vezes estão dispersas em uma ampla extensão de valores (do valor mais baixo ao mais alto);
- Distribuição de frequência nesse caso pode ficar longa e difícil de ler (pouco informativa!);
- Poucos casos se enquadram em cada valor possível
- Alternativa: **distribuição de frequências agrupadas.**

## Exemplo

Vamos estudar a distribuição das notas das escolas em Língua Portuguesa.

# Proporções e porcentagens

- Quando um pesquisador estuda populações do mesmo tamanho, os dados podem ser usados diretamente para fazer comparações entre grupos;
- Esse raramente é o caso.
- Para um uso mais geral, precisamos de um método de *padronização* das distribuições em termos dos tamanhos das populações: **Proporções e porcentagens**.

# Proporções e porcentagens

## Proporção

- Compara o número de casos  $n_i$  em uma determinada categoria com o tamanho total da amostra (ou da população),  $n$ .
- A partir da frequência observada  $n_i$ , podemos obter a *frequência relativa*  $f_i$  correspondente, que nada mais é do que a proporção:

$$f_i = \frac{n_i}{n}$$

# Proporção

## Exemplo

Calcular as proporções correspondentes às frequências calculadas no exemplo anterior (proporção de Escolas que oferecem EJA).

# Porcentagem

## Porcentagem

- Apesar da utilidade das proporções, muitas pessoas preferem indicar o tamanho relativo de uma subpopulação em termos de *porcentagem*;
- A **porcentagem** é a frequência de ocorrência de uma determinada categoria por 100 casos;
- Para calcular uma porcentagem, simplesmente multiplicamos um proporção dada por 100:

$$100 \times f_i$$

# Porcentagem

## Exemplo

Calcular as porcentagens correspondentes às frequências calculadas no exemplo anterior.

# Proporções e porcentagens

## Exercício

Para ilustrar a utilidade de porcentagens em comparações entre distribuições grandes e desiguais, examine os gêneros de estudantes de engenharia de duas Universidades, conforme a tabela abaixo. Interprete os resultados.

Gênero de estudantes de engenharia nas universidades A e B.

| Gênero do estudante | Universidade A |   | Universidade B |   |
|---------------------|----------------|---|----------------|---|
|                     | <i>f</i>       | % | <i>f</i>       | % |
| Masculino           | 1082           |   | 146            |   |
| Feminino            | 270            |   | 37             |   |
| Total               |                |   |                |   |

# Razões e taxas

- A *razão* é outro método de padronização de quantidades com relação ao tamanho;
- Compreende o número de casos que se enquadram em uma categoria (por exemplo, homens) com o número de casos enquadrados em outra categoria (por exemplo, mulheres).

## Razão

Seja  $n_1$  a frequência na categoria 1 e  $n_2$  a frequência na categoria 2. Então:

$$\text{Razão} = \frac{n_1}{n_2}$$

# Razão

## Exemplo: *Razão de raça*

Se estivéssemos interessados em determinar a razão de negros para brancos, compararíamos o número de entrevistados negros ( $n_1 = 100$ ) com o número de entrevistados brancos ( $n_2 = 150$ ).

$$\text{Razão de raça} = \frac{n_1}{n_2} = \frac{100}{150} = \frac{2}{3}$$

Há 2 negros para cada 3 brancos.

## *Razão de gênero*

Conceito bastante empregado por demógrafos. Se, por exemplo, a relação de homens para mulheres é 150/50, há 150 homens para cada 50 mulheres. Para obter a versão convencional de razão de gênero, multiplicamos a razão anterior por 100.

$$\text{Razão de gênero} = \frac{n_{\text{homens}}}{n_{\text{mulheres}}} \times 100 = \left( \frac{150}{50} \right) \times 100 = 300$$

# Taxa

- Outro tipo de razão: *Taxa*
- Mais amplamente utilizada por pesquisadores;
- Sociólogos frequentemente analisam populações em termos de taxas (de reprodução, morte, criminalidade, desemprego, divórcio, casamento, etc);
- Comparam número de casos de um subgrupo com número de casos de outro subgrupo;

## Taxas

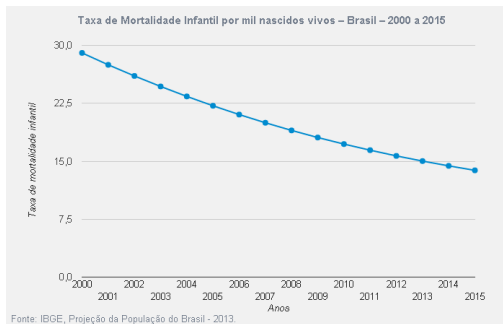
Indicam comparações entre o número de casos efetivos (*reais*) com o número de casos *em potencial*.

$$\text{Taxa} = \frac{n_{\text{casos reais}}}{n_{\text{casos potenciais}}}$$

# Taxa

Exemplo: taxa de mortalidade infantil (por mil habitantes)

$$\text{Taxa de mortalidade infantil} = \frac{n_{\text{crianças mortas antes de completar 1 ano de idade}}}{n_{\text{nascidos vivos}}} \times 1.000$$



## Outros exemplos:

- Taxa de fecundidade
- Taxa de natalidade
- Taxa de analfabetismo
- Taxa de homicídios
- Taxa de suicídio
- gasto *per capita*
- etc

## Consultar:

- <http://brasilemsintese.ibge.gov.br/populacao.html>
- <http://ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/conceitos.shtm>