

# EPI02 - Fundamentos de Bioestatística

## Levantamentos estatísticos

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



# Delineamento de pesquisa

# Introdução

- ▶ A realização de pesquisas empíricas levanta questões relativas ao **plano de observação da realidade** e à **escolha de métodos** para processamento e análise de dados.
- ▶ Estudar essas questões significa buscar opções lógicas e operacionais para a **composição de delineamentos** que assegurem o valor científico das informações obtidas e das conclusões alcançadas.

# Introdução

- ▶ Entende-se por **delineamento** o **plano estratégico de observação da realidade** que orienta o detalhamento posterior dos métodos e técnicas necessários à execução da pesquisa.
- ▶ De acordo com a natureza lógica e estrutural desse plano, as pesquisas podem ser reconhecidas como **experimentos, quase experimentos** ou **estudos observacionais**.

# Experimentos

- ▶ **Experimento** é o tipo de pesquisa em que o investigador **controla** a ocorrência das variáveis independentes (causa-fator-tratamento) com a finalidade de observar seus efeitos sobre as variáveis consideradas dependentes (efeito-resposta-doença).
- ▶ A **validade** dos efeitos detectados é assegurada por meio da **alocação aleatória** das unidades experimentais aos níveis definidos para as variáveis independentes.
- ▶ O **princípio da aleatorização** (também chamado de **princípio da randomização**) representou uma significativa contribuição da estatística às ciências empíricas, indicando caminhos para o desenvolvimento de planos de observação resistentes às diferentes fontes de erro, que podem comprometer a validade dos resultados.

# Experimentos

## Exemplo



**Do this**

...snap out of it the way doctors proved *faster*. Take Bromo-Seltzer! It does *more* for you. Not only relieves headache pain, but settles the stomach—calms the nerves—ALKALIZES! Soon you're smiling ...

**And feel like this**

...head clear—that painful pounding gone. Your stomach is on an even keel—your nerves relaxed and calm. Bromo-Seltzer leaves you refreshed and peppy, too. You feel keener, more set for the day.

Get Bromo-Seltzer at drugstores and soda fountains everywhere. Keep it handy at home, too.

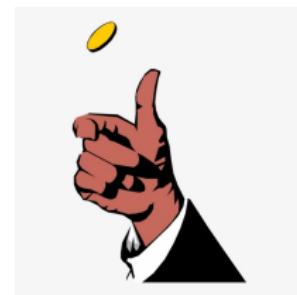
**BROMO-SELTZER**

Makes you feel fit FASTER

# Experimentos

## Exemplo

- ▶ 12 pessoas sofrendo de dor cabeça.
- ▶ Algumas tomam aspirina ( $Z_i = 1$ ); outras não ( $Z_i = 0$ ).
  - ▶ Em um **delineamento experimental**, os grupos de **tratamento** e **controle** são aleatorizados. Por exemplo, cada indivíduo é alocado ao grupo de tratamento com  $\Pr(Z_i = 1) = 1/2$ .
- ▶ Uma hora depois, perguntamos para cada uma delas se a dor de cabeça sumiu ( $Y_i = 0$  se a dor não sumiu,  $Y_i = 1$  se a dor sumiu).

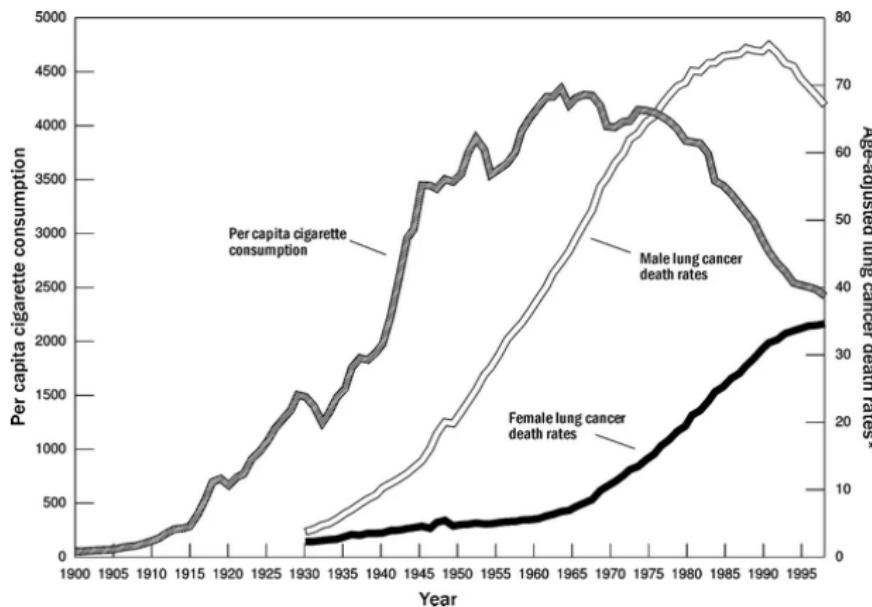


## Estudos observacionais

Nas pesquisas em que a unidade de estudo é o ser humano (ou seres vivos), os investigadores, por questões éticas e práticas, não podem controlar a ocorrência da variável independente ou tratamento nem proceder à alocação aleatória das unidades observadas aos grupos de comparação, optando, assim, por delineamentos considerados **não experimentais (estudos observacionais)**.

- ▶ Estudos de coorte;
- ▶ Estudos do tipo caso-controle;
- ▶ Estudos de cortes transversais ou levantamentos.

## Exemplo: “onde há fumaça ...”

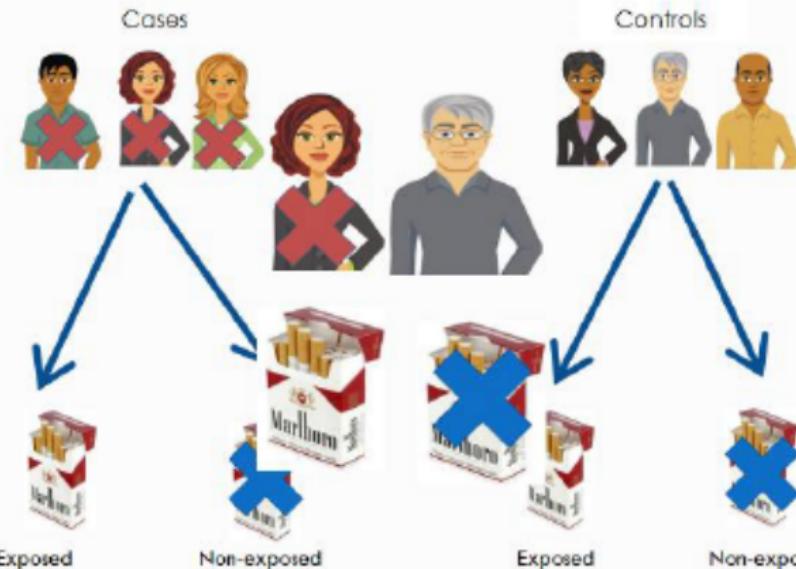


**Figura 1:** Uso de tabaco e mortalidade por câncer de pulmão, EUA, 1900-1997.  
Asterisco (\*) Por 100.000 e ajustado por idade para a população padrão dos EUA em 1970.

# Exemplo: “onde há fumaça ...”

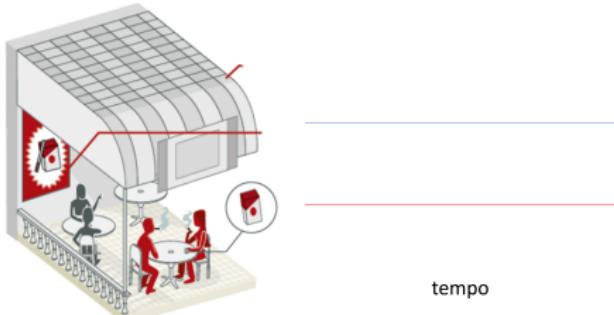
## Estudo caso-controle

### Steps in a case-control study

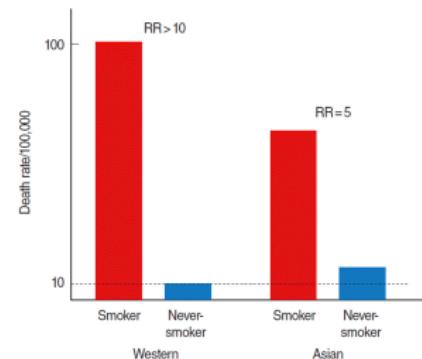


# Exemplo: “onde há fumaça ...”

## Estudo de coorte

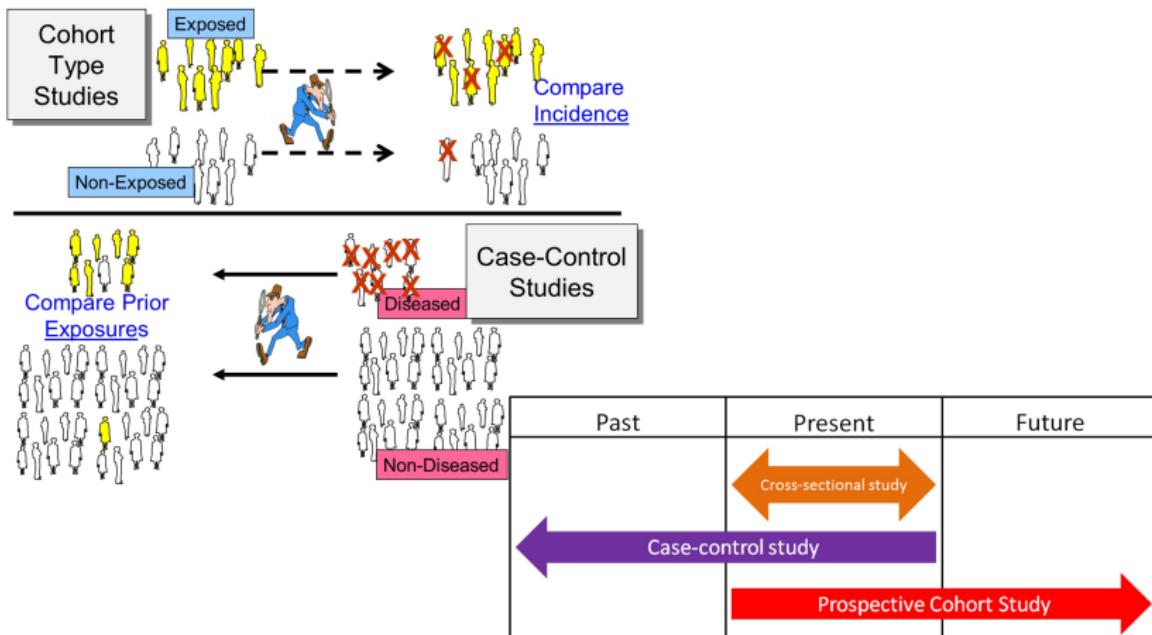


## Estudos de coorte



Fonte: Jung KJ, Jeon C, Jee SH. The effect of smoking on lung cancer: ethnic differences and the smoking paradox. *Epidemiol Health*. 2016;38:e2016060. doi: 10.4178/epih.e2016060

# Estudos observacionais



# Estudos observacionais

## Exemplo

- ▶ 12 pessoas sofrendo de dor cabeça.
- ▶ Algumas tomam aspirina ( $Z_i = 1$ ); outras não ( $Z_i = 0$ ).
  - ▶ Em um **delineamento observacional**, os grupos de **tratamento** e **controle** **não** são aleatorizados. Os indivíduos podem ser “alocados” ao grupo de tratamento devido a influência de covariáveis pré-tratamento ( $X_i$ ). Por exemplo, a **gravidade da dor de cabeça** pode influenciar na decisão de tomar ou não aspirina.
- ▶ Uma hora depois, perguntamos para cada uma delas se a dor de cabeça sumiu ( $Y_i = 0$  se a dor não sumiu,  $Y_i = 1$  se a dor sumiu).
  - ▶ Note que as covariáveis que afetam o tratamento, também podem influenciar o desfecho ( $Y$ ).

# Estudos experimentais vs. observacionais

## Exemplo (cont.): delinamentos experimentais vs. observacionais

- ▶ Entre as pessoas que tomaram aspirina, a proporção de “dor de cabeça sumiu” foi  $1/12$ ; entre as que não tomaram aspirina, a proporção foi  $4/12$ .
- ▶ **Quais conclusões podemos elaborar a respeito do tratamento “tomar aspirina”?**

# Levantamentos por amostragem

# Levantamentos por amostragem

- ▶ “Levantamento por amostragem” é a tradução ao português da expressão em língua inglesa *survey sampling* ou *sample surveys*.
- ▶ Com a finalidade de produzir “instantâneos” das realidades estudadas, os levantamentos por amostragem reúnem as seguintes características:
  1. Aplicam-se a **conjuntos reais e finitos** (na maioria dos casos), compostos de elementos, denominados **população de estudo**.
  2. Os **elementos** podem ser seres humanos, animais, árvores, fichas, prontuários, domicílios, áreas ou objetos.
  3. As características ou atributos são observados em cada elemento e, posteriormente, agregados por meio de medidas estatísticas chamadas **parâmetros** ou **valores populacionais**.
  4. Os dados são coletados em **amostras** das populações de estudo, e as medidas calculadas (estimativas) passam a ser a informação disponível para os valores populacionais desconhecidos.

## Levantamentos descritivos vs. analíticos

Os levantamentos podem ter a finalidade **descritiva**, limitando-se a estimar frequências de elementos com determinada propriedade ou médias e variâncias de características quantitativas.

- ▶ **EX.:** prevalência de diabetes na população brasileira de adultos (**proporção, frequência, erro padrão, intervalo de confiança**).

Outros levantamentos, denominados **analíticos (ou investigações)**, definem grupos de comparação e, além de estimar, procuram detectar relações entre as características, buscando aumentar as explicações para o objeto pesquisado.

- ▶ **EX.:** associação entre índice de massa corporal (IMC) e prevalência de diabetes (**razão de prevalências [RP] entre grupos de IMC, intervalo de confiança para RP**).
- ▶ Na prática, a maioria dos levantamentos cumpre as duas finalidades por meio de métodos de estimação e de técnicas de análise estatística multivariada.



## Vantagens do método das amostragens

## Uma pequena digressão

### Como obter uma boa amostragem e como tirar conclusões adequadas dos resultados?

Isso não importa, desde que o material do qual retiramos as amostras seja **homogêneo**, pois qualquer tipo de amostragem dá quase o mesmo resultado. Os diagnósticos de laboratório sobre nosso estado de saúde são feitos com apenas algumas gotas de sangue. Esse processo fundamenta-se na presunção de que o sangue em circulação está sempre bem misturado e que uma gota conta a mesma história que qualquer outra. Entretanto, quando o material está longe de ser homogêneo, **como acontece frequentemente**, o processo pelo qual se obtém a amostra se torna crítico e o estudo das técnicas que assegurem amostras dignas de confiança é importante.

## Vantagens do método das amostragens

- ▶ Custo reduzido;
- ▶ Maior rapidez;
- ▶ Maior amplitude;
- ▶ Maior exatidão.

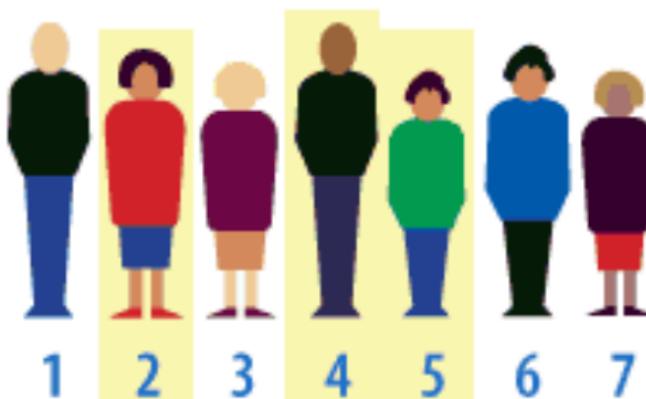
# Amostragem

# 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

- ▶ Talvez o **plano de amostragem** mais conhecido é **amostra casual simples** (ou aleatória simples).
  - ▶ Os indivíduos (unidades) da amostra são selecionados ao acaso, **com ou sem reposição** a partir de uma listagem completa da população.



## Amostra aleatória simples

- ▶ **OBS.:** a amostragem aleatória simples é simples na (elaboração de uma) teoria, mas menos simples na prática.

## Amostra estratificada

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



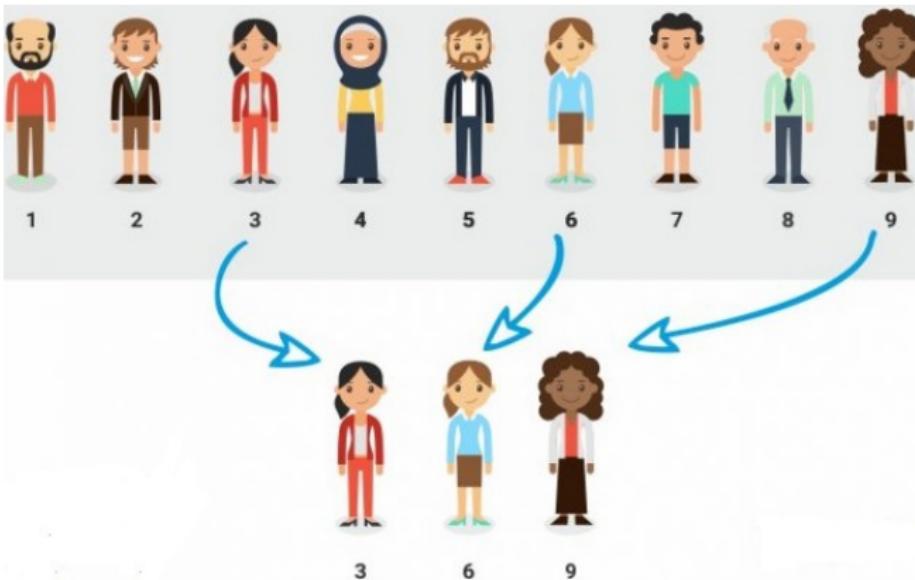
## Amostra estratificada

- ▶ O aspecto fundamental da amostragem estratificada é que os estratos são amostrados separadamente e os resultados apropriadamente combinados na análise.
- ▶ Se a maior parte da variabilidade no que estamos medindo está entre estratos, podemos usar estratificação para reduzir a variância devido à amostragem.

## 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.

## Amostra sistemática



## Amostra sistemática

Suponha que queiramos obter uma amostra sistemática de uma lista cronológica de todas as internações hospitalares durante um ano.

- ▶ Se o objetivo da amostragem é estimar a **proporção** de todas as **internações** hospitalares durante o ano devido a **doenças infecciosas**, é bastante provável que uma amostra sistemática de internações seja preferível a uma amostra aleatória simples.

## Amostra sistemática

Considere que as internações por doenças infecciosas podem ter **picos sazonais** e que por amostragem sistemática a partir de uma lista cronológica temos a certeza de obter  $(1/r)$  das internações em cada estação.

# Amostra sistemática

Copa coincide com queda sazonal da dengue

Veja variação semanal de casos desde 2009

LEGENDA

- 2009
- 2010
- 2011
- 2012
- 2013
- 2014

Números  
de casos

120.000

100.000

80.000

60.000

40.000

20.000

0

1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51

Semana epidemiológica

| JAN | FEV | MAR | ABR | MAI | JUN | JUL | AGO | SET | OUT | NOV | DEZ |

Periodo da Copa do Mundo:  
de 12/6 a 13/7 de 2014

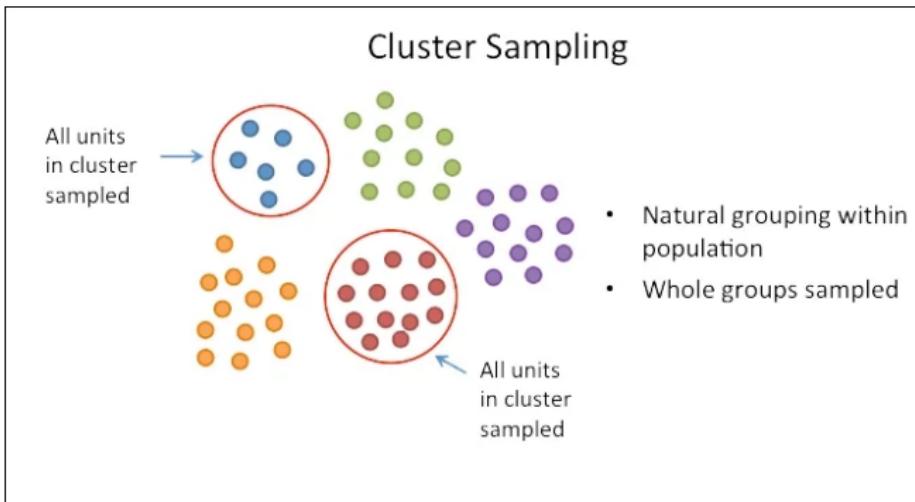
## Amostra sistemática

Na amostragem aleatória simples, no entanto, algumas das amostras possíveis incluiriam internações apenas de uma estação específica e, portanto, refletiriam inadequadamente a proporção de internações anuais por doenças infecciosas.

- ▶ Tal como no exemplo anterior, se a lista for ordenada de forma relacionada com o objetivo do estudo, a amostragem sistemática é, em alguns aspectos, análoga à amostragem estratificada.
- ▶ Se a lista estiver em ordem aleatória, a amostragem sistemática será análoga à amostragem aleatória simples.

# Amostra por conglomerados

- Também conhecida como amostragem por *cluster*.



- Apresenta mais benefícios práticos.

# Amostragem

- ▶ Outros esquemas de amostragem poderiam ser citados e todos fazem parte da chamada **teoria da amostragem**, cujos detalhes não serão aprofundados.
- ▶ Uma outra questão importante nos estudos epidemiológicos é aquela relacionada ao **tamanho da amostra**.
  - ▶ Esta está diretamente relacionada aos objetivos do estudo, precisão desejada das estimativas e variabilidade das medidas.

# Organização dos dados

# Coleta, consistência e processamento

- ▶ **Coletar, conferir e processar** os dados são atividades que podem e devem ser realizadas conjunta e progressivamente.
- ▶ A **crítica e tratamento** dos dados coletados é uma etapa indispensável para permitir a eliminação de erros grosseiros na massa de dados coletados (idade > 200 anos; indivíduo do sexo masculino com tempo de gestação > 0), os quais podem distorcer significativamente os resultados da pesquisa.
  - ▶ É preciso ter formas de detectar inconsistências e definir o tratamento para a correção dos dados individuais.
- ▶ **Codificação, criação de variáveis derivadas e cálculo de pesos amostrais** fazem parte da elaboração dos **arquivos de dados** para análises.
  - ▶ **Dicionários de variáveis** também são produzidos nesta etapa. Veja um exemplo de dicionário de variáveis da PNS 2013.

## Para casa

- ▶ Explore o conjunto de dados do Estudo de Framingham no R ou em algum outro software.
  - ▶ Compartilhe os seus achados no Fórum Geral do Moodle.

## Próxima aula

- ▶ Descrição dos dados por resumos numéricos
- ▶ A **Tabela 1!**