

Estatística e Inferência Causal

Uma breve introdução

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

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

Café do DEST - UFPR, 2020



O que é inferência causal?

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.
- ▶ Como estatísticos, epidemiologistas, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.
- ▶ Como estatísticos, epidemiologistas, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.
 - ▶ Suponha que um estudo encontre uma associação entre a “o pai possuir gravata de seda” e a “mortalidade infantil”.

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.
- ▶ Como estatísticos, epidemiologistas, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.
 - ▶ Suponha que um estudo encontre uma associação entre a “o pai possuir gravata de seda” e a “mortalidade infantil”.
 - ▶ Com base nisso, o governo implementa um programa no qual 5 gravatas de seda são dadas a todos os homens com idade entre 18 e 45 anos, com o objetivo de reduzir a mortalidade infantil.

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.
- ▶ Como estatísticos, epidemiologistas, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.
 - ▶ Suponha que um estudo encontre uma associação entre a “o pai possuir gravata de seda” e a “mortalidade infantil”.
 - ▶ Com base nisso, o governo implementa um programa no qual 5 gravatas de seda são dadas a todos os homens com idade entre 18 e 45 anos, com o objetivo de reduzir a mortalidade infantil.
- ▶ Nós todos concordamos que isso é uma bobagem!

O que é inferência causal?

- ▶ A inferência causal é a ciência de **inferir a presença e a magnitude das relações de causa e efeito a partir dos dados**.
- ▶ Como estatísticos, epidemiologistas, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.
 - ▶ Suponha que um estudo encontre uma associação entre a “o pai possuir gravata de seda” e a “mortalidade infantil”.
 - ▶ Com base nisso, o governo implementa um programa no qual 5 gravatas de seda são dadas a todos os homens com idade entre 18 e 45 anos, com o objetivo de reduzir a mortalidade infantil.
- ▶ Nós todos concordamos que isso é uma bobagem!
- ▶ Isso porque entendemos a **diferença entre associação e causalidade**.

O que é inferência causal?

O roteiro da inferência causal consiste em (pelo menos) três etapas:

1. Uma **linguagem formal** para definir inequivocamente conceitos causais.
 - ▶ Desfechos potenciais, contrafactuals, operador *do()*
2. **Suposições causais** para a **identificação** dos efeitos causais.
 - ▶ **Diagramas causais (DAGs)** são uma ferramenta para exibir nossas suposições causais
3. **Métodos de análise** (isto é, métodos estatísticos) que podem nos ajudar a tirar conclusões causais mais confiáveis a partir dos dados disponíveis.

Um pouco de dor de cabeça!

Um pouco de dor de cabeça!

Um exemplo

- ▶ 12 senhoras estão sofrendo de **dor de cabeça**.
- ▶ Algumas tomam **aspirina**; outras não.
- ▶ Uma hora depois, perguntamos para cada uma delas se a dor de cabeça **sumiu (desapareceu)**.

Os dados observados

	Z (tomou aspirina?)	R (dor de cabeça sumiu?)
Mary	0	0
Anna	1	0
Emma	1	1
Elizabeth	0	0
Minnie	0	1
Margaret	1	0
Ida	1	0
Alice	0	0
Bertha	0	1
Sarah	0	0
Annie	0	1
Clara	1	1

Os dados observados

	Z (tomou aspirina?)	R (dor de cabeça sumiu?)
Mary	0	0
Anna	1	0
Emma	1	1
Elizabeth	0	0
Minnie	0	1
Margaret	1	0
Ida	1	0
Alice	0	0
Bertha	0	1
Sarah	0	0
Annie	0	1
Clara	1	1

Os dados observados

- ▶ Emma tomou aspirina ($Z = 1$) e a sua dor de cabeça passou ($R = 1$).
- ▶ A aspirina **causou** o desaparecimento da sua dor de cabeça?
- ▶ Esta é uma questão causal! Assim como:

Os dados observados

- ▶ Emma tomou aspirina ($Z = 1$) e a sua dor de cabeça passou ($R = 1$).
- ▶ A aspirina **causou** o desaparecimento da sua dor de cabeça?
- ▶ Esta é uma questão causal! Assim como:
 - ▶ O tabaco **causou** a presença de câncer de pulmão?

Os dados observados

- ▶ Emma tomou aspirina ($Z = 1$) e a sua dor de cabeça passou ($R = 1$).
- ▶ A aspirina **causou** o desaparecimento da sua dor de cabeça?
- ▶ Esta é uma questão causal! Assim como:
 - ▶ O tabaco **causou** a presença de câncer de pulmão?
 - ▶ As políticas de restrição a propaganda de cigarro **causaram** a diminuição das taxas de câncer de pulmão?

Os dados observados

- ▶ Emma tomou aspirina ($Z = 1$) e a sua dor de cabeça passou ($R = 1$).
- ▶ A aspirina **causou** o desaparecimento da sua dor de cabeça?
- ▶ Esta é uma questão causal! Assim como:
 - ▶ O tabaco **causou** a presença de câncer de pulmão?
 - ▶ As políticas de restrição a propaganda de cigarro **causaram** a diminuição das taxas de câncer de pulmão?
 - ▶ O diploma de ensino superior **causou** a promoção no emprego?

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.
- ▶ Z_i é o tratamento atribuído: tomou aspirina?

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.
- ▶ Z_i é o tratamento atribuído: tomou aspirina?
 - ▶ $Z_i = 0$, se não tomou aspirina; $Z_i = 1$, se tomou aspirina.

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.
- ▶ Z_i é o tratamento atribuído: tomou aspirina?
 - ▶ $Z_i = 0$, se não tomou aspirina; $Z_i = 1$, se tomou aspirina.
- ▶ Utilizaremos $\pi_i = \Pr(Z_i = 1)$ para designar a probabilidade do indivíduo i ser atribuído ao grupo tratado (tomou aspirina).

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.
- ▶ Z_i é o tratamento atribuído: tomou aspirina?
 - ▶ $Z_i = 0$, se não tomou aspirina; $Z_i = 1$, se tomou aspirina.
- ▶ Utilizaremos $\pi_i = \Pr(Z_i = 1)$ para designar a probabilidade do indivíduo i ser atribuído ao grupo tratado (tomou aspirina).
- ▶ R_i é o desfecho/resposta: dor de cabeça sumiu?

Estrutura

A estrutura da inferência causal:

- ▶ X_i representa um vetor de covariáveis **observadas**: idade, faz uso de medicamentos, tem pressão alta, etc.
- ▶ u_i representa um vetor de covariáveis **não observadas**: variante de um gene, fator ambiental, etc.
- ▶ Z_i é o tratamento atribuído: tomou aspirina?
 - ▶ $Z_i = 0$, se não tomou aspirina; $Z_i = 1$, se tomou aspirina.
- ▶ Utilizaremos $\pi_i = \Pr(Z_i = 1)$ para designar a probabilidade do indivíduo i ser atribuído ao grupo tratado (tomou aspirina).
- ▶ R_i é o desfecho/resposta: dor de cabeça sumiu?
 - ▶ $R_i = 0$, se dor de cabeça não sumiu; $R_i = 1$, se dor de cabeça sumiu.

Estrutura (continuação)

- r_{C_i} e r_{T_i} representam os **desfechos potenciais**.

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $\delta_i = r_{T_i} - r_{C_i}$ de o **efeito causal na escala da diferença** (ou a **diferença do efeito causal**).

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $\delta_i = r_{T_i} - r_{C_i}$ de o **efeito causal na escala da diferença** (ou a **diferença do efeito causal**).
 - ▶ Se $\delta_i = 0$ então o efeito causal é **nulo**; se $\delta_i \neq 0$ então existe um efeito causal (benéfico ou prejudicial).

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $\delta_i = r_{T_i} - r_{C_i}$ de o **efeito causal na escala da diferença** (ou a **diferença do efeito causal**).
 - ▶ Se $\delta_i = 0$ então o efeito causal é **nulo**; se $\delta_i \neq 0$ então existe um efeito causal (benéfico ou prejudicial).
- ▶ Um destes desfechos potenciais é observado: se $Z = 0$, r_C é observada; se $Z = 1$, r_T é observada.

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $\delta_i = r_{T_i} - r_{C_i}$ de o **efeito causal na escala da diferença** (ou a **diferença do efeito causal**).
 - ▶ Se $\delta_i = 0$ então o efeito causal é **nulo**; se $\delta_i \neq 0$ então existe um efeito causal (benéfico ou prejudicial).
- ▶ Um destes desfechos potenciais é observado: se $Z = 0$, r_C é observada; se $Z = 1$, r_T é observada.
 - ▶ Ou seja, $R_i = Z_i \times r_{T_i} + (1 - Z_i) \times r_{C_i}$.

Estrutura (continuação)

- ▶ r_{C_i} e r_{T_i} representam os **desfechos potenciais**.
 - ▶ r_C é o desfecho que teria sido observado caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é o desfecho que teria sido observado caso a aspirina tivesse sido tomada.
- ▶ O par de desfechos potenciais (r_{C_i}, r_{T_i}) será descrito como o **efeito causal** (individual).
 - ▶ Chamaremos $\delta_i = r_{T_i} - r_{C_i}$ de o **efeito causal na escala da diferença** (ou a **diferença do efeito causal**).
 - ▶ Se $\delta_i = 0$ então o efeito causal é **nulo**; se $\delta_i \neq 0$ então existe um efeito causal (benéfico ou prejudicial).
- ▶ Um destes desfechos potenciais é observado: se $Z = 0$, r_C é observada; se $Z = 1$, r_T é observada.
 - ▶ Ou seja, $R_i = Z_i \times r_{T_i} + (1 - Z_i) \times r_{C_i}$.
 - ▶ O outro é **contrafactual**.

Estrutura: observações

1. Consideramos apenas dois níveis de tratamento por uma questão de simplicidade. Esta ideia pode ser generalizada para múltiplos níveis de tratamento e para outros regimes de tratamento mais gerais.
2. Em um **experimento aleatorizado** $\pi_i = 1/2$.
3. Na estatística, a ideia de definir **efeitos causais** como comparações de **desfechos potenciais** foi introduzida por **Neyman** no contexto de experimentos aleatorizados¹. Posteriormente, **Rubin** generalizou esta ideia para o contexto de experimentos não-aleatorizados (estudos observacionais)². Alguns autores se referem a esta abordagem como o **modelo causal de Rubin**³.

¹Splawa-Neyman, J., Dabrowska, D.M., Speed, T.P. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science* 5:465-472, 1990.

²Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701, 1974.

³Holland, P. Statistics and Causal Inference. *JASA*, 81:945-960, 1986.

Estrutura: observações

4. Efeitos causais são comparações de desfechos potenciais sob tratamentos alternativos.
 - ▶ Para um indivíduo específico é perguntado:
 - ▶ O que aconteceria com esse indivíduo sob o primeiro tratamento?
 - ▶ O que aconteceria com o indivíduo sob o segundo tratamento?
 - ▶ O indivíduo se sairia melhor sob o primeiro, em vez do segundo tratamento?
 - ▶ O desfecho seria o mesmo sob os dois tratamentos?
5. **Problema fundamental da inferência causal⁴:** nós nunca vemos o efeito causal porque o efeito causal é a comparação de dois desfechos potenciais que o indivíduo teria exibido sob os dois tratamentos alternativos.
 - ▶ A inferência causal é difícil porque é sobre algo que nunca podemos ver.
6. Em verdade, $R_i = Z_i \times r_{T_i} + (1 - Z_i) \times r_{C_i}$ é **uma suposição**. Chamaremos esta **suposição de consistência** (se $Z_i = 0 \Rightarrow R_i = r_{C_i}$, e se $Z_i = 1 \Rightarrow R_i = r_{T_i}$).

⁴Holland, P. Statistics and Causal Inference. *JASA*, 81:945-960, 1986.

Estrutura: observações

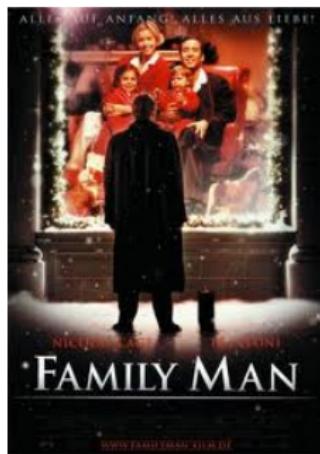
7. **Notações alternativas:** o par de desfechos potenciais (r_{C_i}, r_{T_i}) muitas vezes é denotado por ($Y_i(0), Y_i(1)$).
8. No exemplo, o desfecho é dicotômico. Mas, o mesmo poderia ser um desfecho qualquer (discreto ou contínuo).
9. Alguns autores utilizam os termos “desfechos potenciais” e “variáveis contrafactualas” de forma alternada⁵. Já outros autores fazem a distinção entre estes dois termos⁶, pois no primeiro, antes da atribuição ao tratamento, os dois desfechos são possíveis (potenciais) de ocorrer, enquanto que o segundo se refere ao momento posterior a atribuição ao tratamento (fato e contrário ao fato).

⁵VanderWeele, T. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, 2015.

⁶Imbens, G. W., Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Desfechos potenciais no senso comum

The Family Man (2000)



Jack Campbell é um investidor de Wall Street jovem e solteiro vivendo uma vida de rico em Nova Iorque. Ele se surpreende quando sua ex-namorada, Kate, tentou ligar para ele após anos sem se verem. Após uma conversa com o seu mentor na empresa, Jack resolve pensar se responderia a esta chamada no dia seguinte. Naquela noite de Natal, porém, ele resolve ir a pé até sua casa, passando por uma loja de conveniências no caminho e convencendo para que um vencedor da loteria, irritado, chamado Cash, não atirasse no vendedor. Ele oferece ajuda à Cash antes de ir dormir em sua cobertura.

Tudo muda num passe de mágica quando na manhã seguinte ele acorda em um quarto no subúrbio de Nova Jersey com Kate, a sua atual esposa, com quem anteriormente ele havia deixado de se casar e ainda com duas crianças que ele sequer conhecia. Jack percebe então que esta é justamente **a vida que ele teria** se não tivesse se transformado em um investidor financeiro quando jovem. Ao invés disso, ele tem uma vida modesta, onde ele é um vendedor de pneus e Kate é uma advogada não-remunerada.

Desfechos potenciais no senso comum

Rodrigo do violão



O “mundo ideal”

	r_C	r_T
Mary	0	0
Anna	1	0
Emma	0	1
Elizabeth	0	0
Minnie	1	1
Margaret	0	0
Ida	0	0
Alice	0	0
Bertha	1	0
Sarah	0	0
Annie	1	1
Clara	0	1

O “mundo ideal”

Com o par de desfechos potenciais, podemos responder as seguintes perguntas:

- ▶ A aspirina causou o desaparecimento da dor de cabeça de Emma?
 - ▶ E de Margaret?
 - ▶ E de Clara?
 - ▶ E de Alice?

O “mundo ideal”

i	r_{C_i}	r_{T_i}	$\delta_i = r_{T_i} - r_{C_i} \neq 0$? (existe efeito causal?)
Mary	0	0	0 (Não)
Anna	1	0	-1 (Sim, prejudicial)
Emma	0	1	1 (Sim, benéfico)
Elizabeth	0	0	0 (Não)
Minnie	1	1	0 (Não)
Margaret	0	0	0 (Não)
Ida	0	0	0 (Não)
Alice	0	0	0 (Não)
Bertha	1	0	-1 (Sim, prejudicial)
Sarah	0	0	0 (Não)
Annie	1	1	0 (Não)
Clara	0	1	1 (Sim, benéfico)

O mundo real!

i	r_{C_i}	r_{T_i}	Z_i	R_i
Mary	0	?	0	0
Anna	?	0	1	0
Emma	?	1	1	1
Elizabeth	0	?	0	0
Minnie	1	?	0	1
Margaret	?	0	1	0
Ida	?	0	1	0
Alice	0	?	0	0
Bertha	1	?	0	1
Sarah	0	?	0	0
Annie	1	?	0	1
Clara	?	1	1	1

O mundo real!

O problema fundamental da inferência causal

- ▶ É impossível de se observar o valor de r_{T_i} e r_{C_i} na mesma unidade e, portanto, é impossível observar o efeito δ_i .
- ▶ Um problema de **dados ausentes**.

Efeitos causais populacionais

- ▶ $\delta_i = r_{T_i} - r_{C_i} = ?$, para todo indivíduo i , pois um dos desfechos potenciais nunca é observado.
- ▶ Um objetivo menos ambicioso é focar no **efeito causal médio** (ou em nível populacional):

$$\bar{\delta} = E[r_{T_i} - r_{C_i}] = E[r_{T_i}] - E[r_{C_i}].$$

- ▶ No caso em que a resposta é dicotômica, temos

$$\bar{\delta} = \Pr(r_{T_i} = 1) - \Pr(r_{C_i} = 1).$$

Efeitos causais populacionais

i	r_{C_i}	r_{T_i}	δ_i
Mary	0	0	0
Anna	1	0	-1
Emma	0	1	1
Elizabeth	0	0	0
Minnie	1	1	0
Margaret	0	0	0
Ida	0	0	0
Alice	0	0	0
Bertha	1	0	-1
Sarah	0	0	0
Annie	1	1	0
Clara	0	1	1
$E[r_{C_i}] = \Pr(r_{C_i} = 1) =$		$E[r_{T_i}] = \Pr(r_{T_i} = 1) =$	
$4/12$		$4/12$	
		$\bar{\delta} = 4/12 - 4/12 = 0$	

- Ou seja, **não existe** efeito causal em nível populacional.

Efeitos causais populacionais

- ▶ Em verdade, **não sabemos** r_T para cada indivíduo, então **não podemos simplesmente estimar** $\Pr(r_{T_i} = 1)$ como a proporção de todos os indivíduos com $r_T = 1$.
- ▶ Da mesma forma, **não podemos simplesmente estimar** $\Pr(r_{C_i} = 1)$ como a proporção de todos os indivíduos com $r_C = 1$.
- ▶ Assim, **não podemos estimar** facilmente
$$\bar{\delta} = \Pr(r_{T_i} = 1) - \Pr(r_{C_i} = 1)$$
pelo mesmo motivo que não podemos estimar $\delta_i = r_{T_i} - r_{C_i}$.
- ▶ A inferência causal é **toda sobre a escolha de quantidades dos dados observados** (isto é, envolvendo Z , R e outras variáveis observadas) **que representam substitutos razoáveis** para quantidades hipotéticas tais como $\bar{\delta}$, que envolvem contrafactuals não observáveis.

Quando associação é igual a causalidade?

- ▶ O que pode ser um bom substituto para $\Pr(r_{Ti} = 1)$?
 - ▶ $\hat{r}_T = \Pr(R = 1 | Z = 1)$?
 - ▶ Esta é a proporção de “dor de cabeça que desapareceu” entre aquelas senhoras que realmente tomaram a aspirina.
 - ▶ Isso é o mesmo que $\Pr(r_{Ti} = 1)$?

Ignorabilidade

- ▶ Somente se aquelas senhoras que tomaram a aspirina forem **intercambiáveis** com aquelas que não o fizeram.
- ▶ Esta condição (suposição) é conhecida como **ignorabilidade da atribuição ao tratamento** ($(r_C, r_T) \perp\!\!\!\perp Z$, **ausência de confundimento**).
- ▶ Este seria o caso se a escolha de tomar a aspirina fosse feita **ao acaso**.
- ▶ É por isso que **experimentos aleatorizados** são o **padrão-ouro** para inferir efeitos causais.

Quando associação é igual a causalidade?

i	r_{C_i}	r_{T_i}	Z_i	R_i
Mary	0	?	0	0
Elizabeth	0	?	0	0
Minnie	1	?	0	1
Alice	0	?	0	0
Bertha	1	?	0	1
Sarah	0	?	0	0
Annie	1	?	0	1

$\hat{r}_C = \Pr(R = 1 | Z = 0) = 3/7$

Quando associação é igual a causalidade?

i	r_{C_i}	r_{T_i}	Z_i	R_i
Anna	?	0	1	0
Emma	?	1	1	1
Margaret	?	0	1	0
Ida	?	0	1	0
Clara	?	1	1	1

$\hat{r}_T = \Pr(R = 1 | Z = 1) = 2/5$

Quando associação é igual a causalidade?

- ▶ Portanto,

$$\hat{\delta} = \Pr(R = 1|Z = 1) - \Pr(R = 1|Z = 0) = \frac{2}{5} - \frac{3}{7} = -\frac{1}{35}.$$

- ▶ Se assumirmos “associação = causalidade”, concluiremos que **existe efeito causal** e que a aspirina foi, em média, prejudicial.
- ▶ **Teste exato de Fisher:** a “base racional para inferência” ($H_0 : \delta_i = 0, i = 1, \dots, I$).
 - ▶ No capítulo 2 de seu **Design of Experiments**, Fisher falou da aleatorização em experimentos como a “**base racional da inferência**” nos experimentos.

Mas, e se ...

... as senhoras com uma dor de cabeça **mais forte** (grave) fossem **mais propensas** a tomarem a aspirina?

- ▶ Neste caso, “associação \neq causalidade”! Ou $(r_C, r_T) \not\perp\!\!\!\perp Z$.

Levando em conta a gravidade

- ▶ Suponha que perguntamos a cada uma das 12 senhoras no início do estudo: “**sua dor de cabeça é forte?**”.
 - ▶ Então, poderíamos propor que, depois de levar em conta a gravidade, a decisão de tomar ou não a aspirina **fosse efetivamente tomada de forma aleatória (?!)**.
- ▶ Suponha que X denota a gravidade. Então, sob essa suposição, **dentro dos estratos** de X , os indivíduos expostos e não expostos podem ser **intercambiáveis**.
- ▶ Isso é chamado de **intercambiabilidade condicional**
 $((r_C, r_T) \perp\!\!\!\perp Z | X \text{ e } 0 < \pi_i = \Pr(Z_i = 1 | X_i) < 1)$.
- ▶ Sob intercambiabilidade condicional, “associação = causalidade” **dentro dos estratos** de X .

Levando em conta a gravidade

i	r_{C_i}	r_{T_i}	Z_i	R_i	X_i
Mary	0	0	0	0	1
Anna	1	0	1	0	0
Emma	0	1	1	1	0
Elizabeth	0	0	0	0	1
Minnie	1	1	0	1	0
Margaret	0	0	1	0	1
Ida	0	0	1	0	1
Alice	0	0	0	0	0
Bertha	1	0	0	1	1
Sarah	0	0	0	0	0
Annie	1	1	0	1	0
Clara	0	1	1	1	1

Estratificando por gravidade

No estrato $X = 0$:

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 1/2$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 2/4$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{1}{2} - \frac{2}{4} = 0.$$

No estrato $X = 1$:

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 1/3$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 1/3$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{1}{3} - \frac{1}{3} = 0.$$

- Ou seja, dentro dos estratos **não existe** efeito causal.

Exemplo da dor de cabeça: breves conclusões

- ▶ De maneira mais geral, se **existe** um efeito causal de Z em R , mas também **uma associação não-causal** devido a X , então o efeito causal será estimado com **viés**, a menos que estratifiquemos em X .
 - ▶ Este viés será chamado de **viés de confundimento** (ou viés de confusão) e X será chamada de **variável confundidora**.

Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.

Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- ▶ Assim, precisamos identificar, se possível, um conjunto de (co)variáveis X_1, X_2, \dots , de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.

Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- ▶ Assim, precisamos identificar, se possível, um conjunto de (co)variáveis X_1, X_2, \dots , de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.
- ▶ Na vida real, pode haver muitas variáveis candidatas X .

Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- ▶ Assim, precisamos identificar, se possível, um conjunto de (co)variáveis X_1, X_2, \dots , de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.
- ▶ Na vida real, pode haver muitas variáveis candidatas X .
- ▶ Estes podem ser **causalmente inter-relacionados** de uma maneira **muito complexa**.

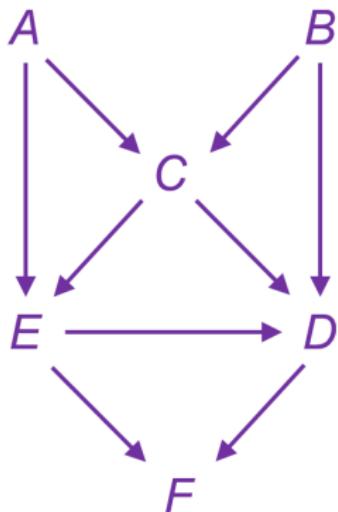
Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- ▶ Assim, precisamos identificar, se possível, um conjunto de (co)variáveis X_1, X_2, \dots , de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.
- ▶ Na vida real, pode haver muitas variáveis candidatas X .
- ▶ Estes podem ser **causalmente inter-relacionados** de uma maneira **muito complexa**.
- ▶ Decidir se os indivíduos expostos e o não expostos são condicionalmente intercambiáveis, dado X_1, X_2, \dots , **requer conhecimento detalhado do assunto**.

Exemplo da dor de cabeça: breves conclusões

- ▶ A **intercambiabilidade condicional** é o principal critério que nos permite fazer declarações causais usando **dados observacionais**.
- ▶ Assim, precisamos identificar, se possível, um conjunto de (co)variáveis X_1, X_2, \dots , de tal forma que a intercambiabilidade condicional é válida, dado este conjunto de variáveis.
- ▶ Na vida real, pode haver muitas variáveis candidatas X .
- ▶ Estes podem ser **causalmente inter-relacionados** de uma maneira **muito complexa**.
- ▶ Decidir se os indivíduos expostos e o não expostos são condicionalmente intercambiáveis, dado X_1, X_2, \dots , **requer conhecimento detalhado do assunto**.
- ▶ Os **diagramas causais** (DAGs) podem nos ajudar a usar esse conhecimento para determinar se a intercambiabilidade condicional é válida ou não (**critério back-door**, variáveis não observadas).

DAGs: um exemplo



Grafo acíclico dirigido

- ▶ Este é um exemplo de um **grafo acíclico dirigido** (DAG) causal (diagrama causal).
- ▶ É **dirigido**, pois cada aresta é uma seta de ponta única.
- ▶ É **causal**, pois as setas representam nossas suposições a respeito da direção da influência causal.
- ▶ É **acíclico**, pois não contém ciclos: nenhuma variável causa a si mesma.

Métodos de estimação

- ▶ Suponha que, aplicando o critério *back-door*, nosso diagrama causal nos diga que o conjunto $X = \{X_1, X_2, \dots, X_p\}$ (**sexo, idade, gravidade da dor, uso de álcool, ...**) é suficiente para controlar para confusão.
- ▶ **Como** analisamos os dados para estimarmos o efeito causal médio da exposição (ou tratamento, o $\bar{\delta}$)?
 - ▶ Estratificação.
 - ▶ Ajuste por covariável na análise de regressão.
 - ▶ Escore de propensão.
 - ▶ Pareamento.
 - ▶ Ponderação.

Métodos de estimação: estratificação

- ▶ Se o número de covariáveis de ajuste é pequeno e estes são categóricos/dicotômicos, podemos criar **estratos** com base nestas covariáveis.
- ▶ Em seguida, calculamos o efeito de interesse em cada estrato e então **combinamos** os resultados (ajuste direto e Mantel-Haenszel).

Métodos de estimação: estratificação

$$\begin{aligned}\bar{\delta} &= \bar{r}_T - \bar{r}_C \\&= E[r_{T_i}] - E[r_{C_i}] \\&= E[E[r_{T_i}|X_i]] - E[E[r_{C_i}|X_i]] \\&= E[E[r_{T_i}|Z_i = 1, X_i]] - E[E[r_{C_i}|Z_i = 0, X_i]] \\&= E[E[R_i|Z_i = 1, X_i]] - E[E[R_i|Z_i = 0, X_i]] \\&= E[E[R_i|Z_i = 1, X_i] - E[R_i|Z_i = 0, X_i]] \\&= \sum_x (E[R_i|Z_i = 1, X_i = x] - E[R_i|Z_i = 0, X_i = x]) \Pr(X_i = x).\end{aligned}$$

Métodos de estimação: estratificação

Table 5.6. How many strata are there with C covariates each at L levels?

<i>Number of covariates (C)</i>	$L = 2$ levels	$L = 3$ levels
1	2	3
2	4	9
5	32	243
10	1,024	59,049
20	~1 million	~3.5 billion
30	~1 billion	$\sim 2.1 \times 10^{14}$
50	$\sim 1.1 \times 10^{15}$	$\sim 7.2 \times 10^{23}$
75	$\sim 3.8 \times 10^{22}$	$\sim 6.1 \times 10^{35}$

Métodos de estimação: regressão

- ▶ Se o nosso conjunto suficiente para controle de confusão contém muitos confundidores (e possivelmente contínuos), teremos muitos e pequenos estratos, perdendo precisão nas estimativas.
- ▶ Uma alternativa natural é ajustar para X em um **modelo de regressão**.
 - ▶ Desenhe o DAG
 - ▶ Identifique o conjunto suficiente de confundidores X
 - ▶ Inclua estes confundidores **apropriadamente** em um modelo de regressão
 - ▶ É importante avaliar relações não-lineares, termos de interação, suposições distribucionais, etc.

Métodos de estimação: regressão

- ▶ Considere o caso em que R , a variável resposta/desfecho, seja contínua.
- ▶ O modelo de regressão linear supõe:

$$\mathbb{E}[R|Z = z, X_1, \dots, X_p] = \beta_0 + \beta_Z z + \beta_1 x_1 + \dots + \beta_p x_p.$$

- ▶ Assim,

$$\begin{aligned}
 \hat{\delta} &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\mathbb{E}}[R_i | Z_i = 1, x_1, \dots, x_p] - \hat{\mathbb{E}}[R_i | Z_i = 0, x_1, \dots, x_p] \right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ (\hat{\beta}_0 + \hat{\beta}_Z + \sum_{k=1}^p \hat{\beta}_k x_{ik}) - (\hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k x_{ik}) \right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\beta}_Z \right\} = \frac{1}{n} \times n \hat{\beta}_Z = \hat{\beta}_Z.
 \end{aligned}$$

Métodos de estimação: regressão

Se

1. o conjunto suficiente de confundidores Z foi corretamente selecionado a partir do DAG corretamente especificado;
2. o modelo de regressão foi corretamente especificado;

então pode ser dada uma interpretação causal a $\hat{\beta}_Z$.

- ▶ **Intervalos de confiança e testes de hipóteses** podem ser construídos a partir das propriedades de $\hat{\beta}_Z$.

Métodos de estimação: regressão

- ▶ Agora, suponha que R , a variável resposta/desfecho, seja dicotômica.
- ▶ O modelo de regressão logística (uma possibilidade de modelo para uma variável resposta dicotômica) supõe:

$$\begin{aligned} E[R|Z = z, X_1, \dots, X_p] &= \Pr[R = 1|Z = z, X_1, \dots, X_p] \\ &= \frac{\exp\{\beta_0 + \beta_Z z + \sum_{k=1}^p \beta_k x_k\}}{1 + \exp\{\beta_0 + \beta_Z z + \sum_{k=1}^p \beta_k x_k\}}. \end{aligned}$$

Métodos de estimação: regressão

► Assim,

$$\begin{aligned}
 \hat{\delta} &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{E}[R_i | Z_i = 1, x_1, \dots, x_p] - \hat{E}[R_i | Z_i = 0, x_1, \dots, x_p] \right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{Pr}[R = 1 | Z = 1, X_1, \dots, X_p] - \hat{Pr}[R = 1 | Z = 0, X_1, \dots, X_p] \right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\exp\{\hat{\beta}_0 + \hat{\beta}_Z + \sum_{k=1}^p \hat{\beta}_k x_k\}}{1 + \exp\{\hat{\beta}_0 + \hat{\beta}_Z + \sum_{k=1}^p \hat{\beta}_k x_k\}} - \frac{\exp\{\hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k x_k\}}{1 + \exp\{\hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k x_k\}} \right\}.
 \end{aligned}$$

Métodos de estimação: regressão

- ▶ **Observação:** fora do modelo linear, $\hat{\beta}_Z$ não é a estimativa do δ .
- ▶ Métodos de reamostragem (por exemplo, o **bootstrap**) podem ser utilizados na construção de intervalos de confiança e testes de hipóteses para o δ .

Métodos de estimação: o escore de propensão

- ▶ No caso em que o conjunto suficiente possui muitos confundidores, o ajuste destes confundidores na análise de regressão pode implicar na falta de precisão das estimativas.
- ▶ Uma alternativa é ajustar por uma função (resumo) dos confundidores, por exemplo, o **escore de propensão**.
- ▶ O escore de propensão $p(X)$ é a probabilidade condicional de $Z = 1$ dado $X = (X_1, \dots, X_p)$

$$p(X) = \Pr(Z = 1|X).$$

Métodos de estimação: o escore de propensão

- ▶ $p(X)$ é um escalar, independentemente da dimensão de X .
- ▶ $p(X)$ pode ser estimado por regressão logística.
 - ▶ Abordagens mais modernas têm utilizado métodos de **aprendizagem de máquina** para estimar $p(X)$.
- ▶ **Intuição:** Se dois indivíduos, um exposto e outro não-exposto, têm o mesmo valor do escore de propensão, 0,25, por exemplo, são igualmente propensas a serem expostas (receberem o tratamento).
- ▶ **Resultado teórico:** intercambiabilidade condicional dado $p(X)$.

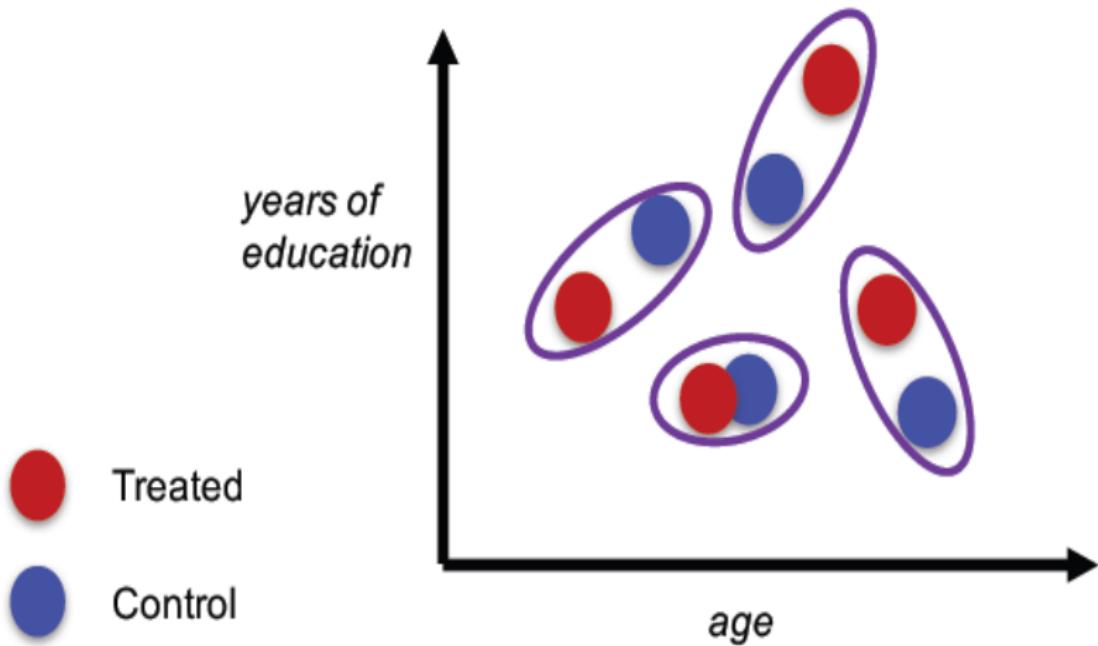
Métodos de estimação: o escore de propensão

- ▶ Existem diferentes maneiras de incorporar o escore de propensão na análise para a estimação do efeito causal de interesse:
 1. estratificação (faixas do escore de propensão como estratos)
 2. ajuste na regressão (no lugar dos confundidores, se utiliza apenas o escore de propensão)
 3. pareamento (escore de propensão como uma medida de distância entre indivíduos)
 4. ponderação (o inverso do escore de propensão como uma ponderação).
- ▶ Estes métodos são válidos somente se os confundidores corretos X são incluídos no conjunto de ajuste e se $p(X)$ é modelado corretamente.

Métodos de estimação: pareamento

- ▶ **Métodos de pareamento:** controlam os confundidores por uma etapa pré-análise.
 - ▶ Encontram pares de indivíduos (um exposto e um não-exposto) **similares** com respeito aos confundidores.
 - ▶ Métricas (distâncias) são utilizadas para estabelecer a similaridade entre indivíduos.
 - ▶ O escore de propensão pode ser utilizado como medida de similaridade, assim como a distância Euclidiana, ou a distância de Mahalanobis.

Métodos de estimação: pareamento

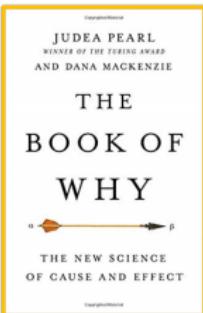
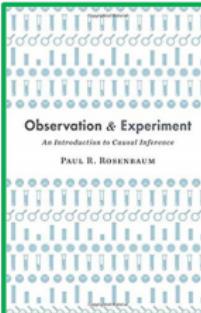
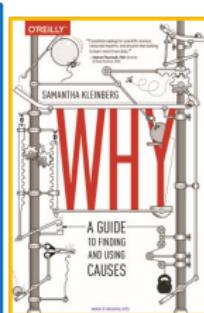
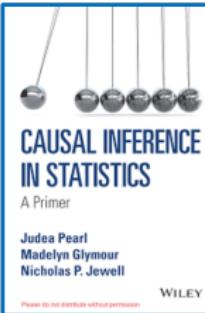
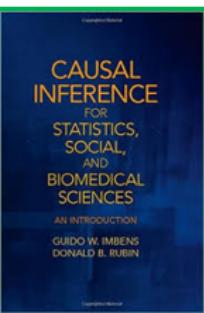
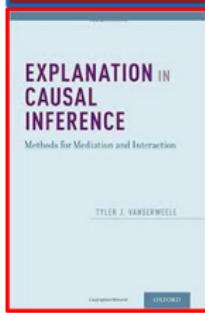
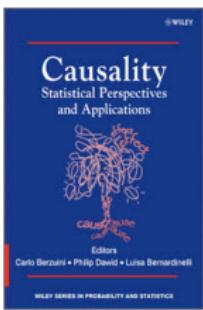
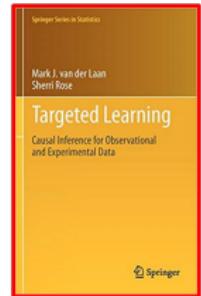
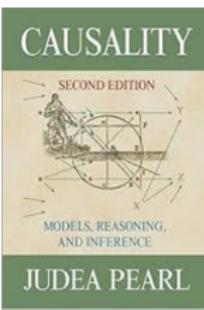
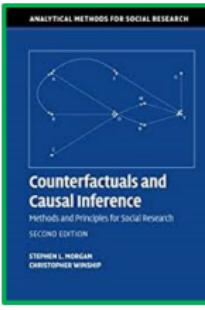
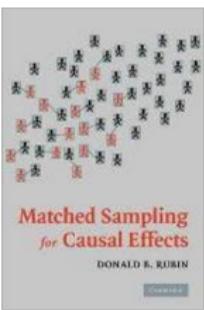
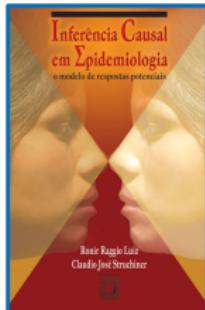


Comentários finais

Comentários finais

- ▶ Nesta introdução apresentamos os conceitos básicos da abordagem de desfechos potenciais para inferência causal.
 - ▶ Definição dos efeitos causais.
 - ▶ Identificação dos efeitos causais.
 - ▶ Estimação dos efeitos causais.
- ▶ **O que vem depois?**
 - ▶ Análise de sensibilidade.
 - ▶ Análise de mediação: decomposição do efeito total em efeitos direto e indireto.
 - ▶ Análise de interação: modificação do efeito.
 - ▶ Confundidores intermediários.
 - ▶ Confundidores tempo-dependentes.
 - ▶ Desfechos: tempo até evento; séries temporais; dados longitudinais.

Algumas referências



- Ainda: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

Atividades atuais

Departamento de Estatística da UFRGS:

- ▶ **Introdução à Inferência Causal.** Fundamentos da inferência causal em estudos aleatorizados e não-aleatorizados (<https://github.com/rdosreis/MAT02010>).
- ▶ **Bruna Silveira.** Avaliação e aplicação de testes não-paramétricos em DAGs (orientação de TCC).
- ▶ **Luana Krindges.** Análise de mediação causal com múltiplos mediadores (orientação de TCC).

Atividades atuais

Pós-Graduação em Epidemiologia da UFRGS:

- ▶ **Seminários de Doutorado II e III.** Conceitos e métodos relacionados à inferência causal em epidemiologia mediante abordagens de desfechos potenciais e grafos (em conjunto com o **Prof. Ricardo Kuchenbecker**).
- ▶ **Letícia Menegotto.** Análise de mediação com dados longitudinais (orientação de mestrado em conjunto com a **Prof.^a Vanessa Leotti**).
- ▶ **Luísa Back.** Efeito mediador da trajetória de índice de massa corporal na relação peso ao nascer e desenvolvimento de diabetes na vida adulta: ELSA-Brasil (orientação de mestrado em conjunto com o **Prof. Alvaro Vigo**).
- ▶ **Juliana Souza.** Métodos de estimação em delineamentos duas fases (orientação de mestrado em conjunto com a **Prof.^a Márcia Barbian**).

Muito obrigado!

caffeine causality loop

