

EPI13 - Seminários de Doutorado III

Oficina de DAGs

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

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA E PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

Vitória da Conquista, 2019



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**.
- ▶ Como epidemiologistas, estatísticos, sociólogos, etc., e de fato como seres humanos, é algo sobre o qual **sabemos muito**.
 - ▶ Suponha que um estudo encontre uma associação entre a propriedade paterna de 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 de 18-45 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?

A área de inferência causal consiste em (pelo menos) três partes:

1. Uma **linguagem formal** para definir inequivocamente conceitos causais.
2. **Diagramas causais**: uma ferramenta para exibir claramente 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 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 (passou)**.

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?

Relembrando

A estrutura da inferência causal:

- ▶ Z é o tratamento atribuído: tomou aspirina?
- ▶ R é o desfecho/resposta: dor de cabeça sumiu?
- ▶ r_C e r_T representam as **respostas potenciais**.
 - ▶ r_C é a resposta que teria sido observada caso a aspirina NÃO tivesse sido tomada.
 - ▶ r_T é a resposta que teria sido observada caso a aspirina tivesse sido tomada.
- ▶ Uma destas respostas é observada: se $Z = 0$, r_C é observada; se $Z = 1$, r_T é observada (ou seja, $R = Z \times r_T + (1 - Z) \times r_C$).
- ▶ A outra é **contrafactual**.

Os dados ideais

	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

Os dados ideais

Com o par de repostas 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?

Os dados ideais

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

O problema fundamental da inferência causal

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

Efeitos causais populacionais

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

$$\bar{\delta} = \bar{r}_T - \bar{r}_C.$$

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

$$\bar{\delta} = \Pr(r_T = 1) - \Pr(r_C = 1).$$

Efeitos causais populacionais

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

- $\bar{r}_T = \Pr(r_T = 1) = 4/12$ e $\bar{r}_C = \Pr(r_C = 1) = 4/12$, e portanto,

$$\bar{\delta} = \bar{r}_T - \bar{r}_C = \frac{4}{12} - \frac{4}{12} = 0.$$

- 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** $\Pr(r_T = 1)$ como a proporção de todos os indivíduos com $r_T = 1$.
- ▶ Da mesma forma, **não podemos simplesmente estimar** $\Pr(r_C = 1)$ como a proporção de todos os indivíduos com $r_C = 1$.
- ▶ Assim, **não podemos estimar** facilmente $\bar{\delta} = \Pr(r_T = 1) - \Pr(r_C = 1)$ 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 contrafactuais não observáveis.

Quando associação é igual a causalidade?

- ▶ O que pode ser um bom substituto para $\Pr(r_T = 1)$?
 - ▶ E quanto a $\hat{r}_T = \Pr(R = 1|Z = 1)$?
 - ▶ Esta é a proporção de “dor de cabeça desapareceu” entre aquelas que realmente tomaram a aspirina.
 - ▶ Isso é o mesmo que $\Pr(r_T = 1)$?
 - ▶ Somente se aquelas que tomaram a aspirina forem **intercambiáveis** com aquelas que não o fizeram.
- ▶ Este seria o caso se a escolha de tomar a aspirina fosse feita ao acaso.
- ▶ É por isso que **experimentos aleatórios** 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
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

- $\hat{r}_T = \Pr(R = 1|Z = 1) = 2/5$ e $\hat{r}_C = \Pr(R = 1|Z = 0) = 3/7$, e portanto,

$$\hat{\delta} = \hat{r}_T - \hat{r}_C = \frac{2}{5} - \frac{3}{7} = -\frac{1}{35}.$$

- Se assumirmos “associação = causalidade”, concluiremos que a aspirina foi, em média, prejudicial.

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”!

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 (permutabilidade) condicional** (dado X).
- ▶ Sob intercambiabilidade condicional dada X , “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 .

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** podem nos ajudar a usar esse conhecimento para determinar se a intercambiabilidade condicional é válida ou não.

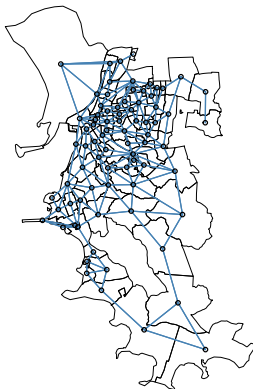
Introdução

Introdução

- ▶ Pesquisa epidemiológica e relações entre variáveis.
- ▶ Modelos estatísticos.
 - ▶ Propósitos dos modelos estatísticos: *predição versus* explicação (exploratório, causalidade).
 - ▶ Modelos estatísticos: o papel das variáveis, e a relação de dependência entre estas.
- ▶ Na estatística, especificamos relações de dependência por meio de distribuições de probabilidades.

Introdução

- **Grafos:** na matemática e ciência da computação, grafos são usados para modelar objetos, representados por vértices (nós), e as suas relações, representadas por arestas.



Introdução

- ▶ **DAGs:** Grafos acíclicos dirigidos (*Directed acyclic graphs*)
 - ▶ Diagramas causais
 - ▶ Modelos causais
 - ▶ Modelos gráficos
 - ▶ Modelos de equações estruturais não paramétricos
 - ▶ Modelos causais estruturais

Preliminares

Como duas variáveis podem estar associadas na população?



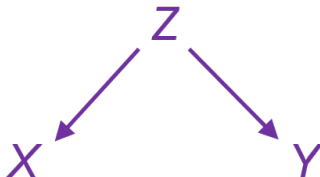
- Duas variáveis X e Y serão **associadas** na população se X causa Y .

Como duas variáveis podem estar associadas na população?



- ▶ X e Y serão associadas na população se Y causa X.

Como duas variáveis podem estar associadas na população?



- Por fim, X e Y serão associadas na população se existir alguma variável Z que causa **ambas** X e Y .

Como duas variáveis podem estar associadas na população?



- ▶ X e Y não podem ser associadas na população por qualquer outra razão.
- ▶ Se X e Y são associadas na população, então **pelo menos uma** das situações acima deve ser verdade.

O que queremos dizer com associação “na população”?

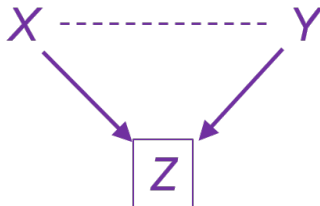
- ▶ Na terminologia estatística, X e Y são associadas “na população” significa que estas variáveis são **marginalmente associadas**.
- ▶ Se X e Y são marginalmente associadas, então, para um indivíduo em particular, saber a respeito de X nos dá alguma informação sobre o valor provável de Y , e vice-versa.
- ▶ Suponha, por simplicidade, X e Y dicotômicas. Se X e Y são marginalmente associadas, então

$$\Pr(Y = 1|X = 1) \neq \Pr(Y = 1|X = 0)$$

e

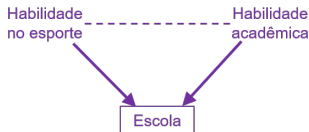
$$\Pr(X = 1|Y = 1) \neq \Pr(X = 1|Y = 0)$$

Como duas variáveis podem estar associadas em uma subpopulação?



- ▶ Suponha que Z é um **efeito** tanto de X como de Y.
- ▶ Então, X e Y serão **associadas dentro do estrato de Z**, mesmo se na população estas variáveis forem independentes.
- ▶ X e Y serão condicionalmente associadas (dado Z), mesmo que sejam marginalmente independentes (não associadas).
- ▶ A caixa ao redor de Z denota que estamos estratificando (condicionando) em Z.
- ▶ A reta tracejada denota a associação condicional induzida.

Como duas variáveis podem estar associadas em uma subpopulação?



- ▶ Suponha que uma escola aceita alunos ou porque são “bons” no **esporte**, ou porque são “bons” **academicamente**; ou ainda, alunos que são “bons” nos dois.
- ▶ Suponha que a habilidade acadêmica e a habilidade no esporte sejam **independentes** na população.
- ▶ **Dentro da escola**, existirá uma associação (negativa) entre habilidade acadêmica e habilidade no esporte.
- ▶ Por que? Suponha que escolhemos um aluno ao acaso e percebemos que ele não tem habilidade nos esportes. Então, este deve ser “bom” academicamente.

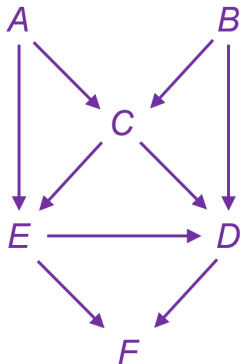
Resumindo



- ▶ X e Y serão **associadas na população** se:
 - ▶ X causa Y .
 - ▶ Y causa X .
 - ▶ existe uma Z que é causa comum de X e Y .
- ▶ X e Y serão **associadas em subpopulações definadas por Z** se Z é um **efeito** de X e Y .

DAGs: uma introdução mais formal

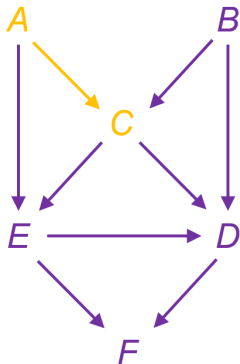
Um exemplo



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.

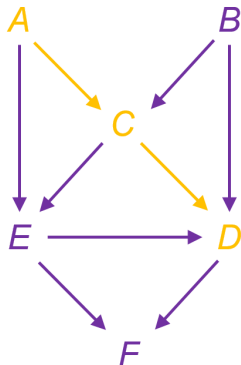
Terminologia



Pais e filhos

- ▶ A é **pai** (ou **mãe**) de C.
- ▶ C é **filho** (ou **filha**) de A.

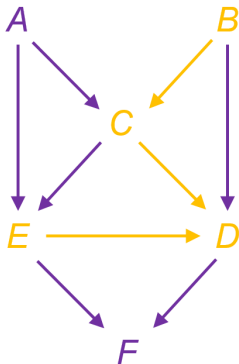
Terminologia



Ancestrais e descendentes

- ▶ A é um **ancestral** de D.
 - ▶ D é **descendente** de A.
 - ▶ A também é um **ancestral** de C.
 - ▶ C também é um **descendente** de A.
- ▶ Ou seja, pais são ancestrais, e filhos são descendentes.

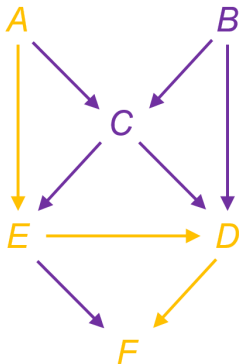
Terminologia



Caminho

- ▶ Este é um **caminho** de E para B .

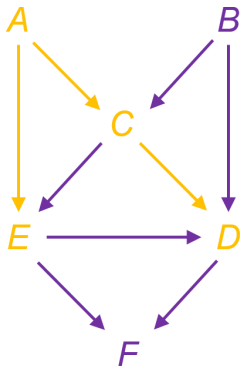
Terminologia



Caminho dirigido

- ▶ Este é um **caminho dirigido** de A para F (todas as setas no caminho apontam “para frente”).

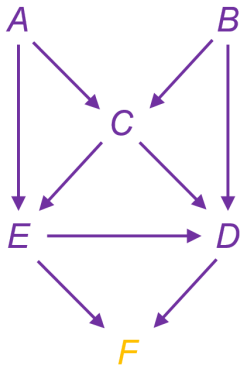
Terminologia



Caminho back-door

- Este é um **caminho porta dos fundos** de E para D (o caminho começa com uma seta **chegando em** E).

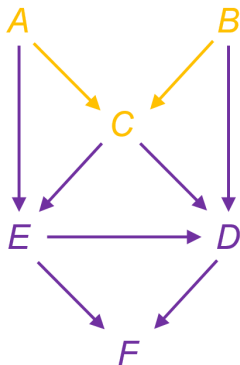
Terminologia



Collider

- F é um **colisor** desde que duas pontas de setas se encontram em F .

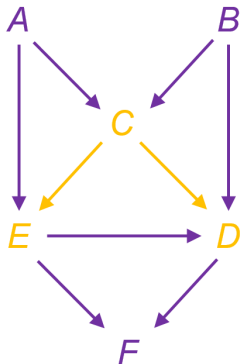
Terminologia



Nota

- Note que C é um colisor no caminho $A \rightarrow C \leftarrow B$.

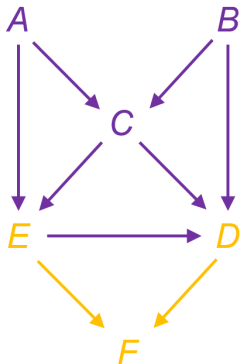
Terminologia



Nota

- ▶ No entanto, C NÃO É um colisor no caminho $E \leftarrow C \rightarrow D$.
- ▶ Assim, a definição de um colisor é em relação ao caminho que está sendo considerado.

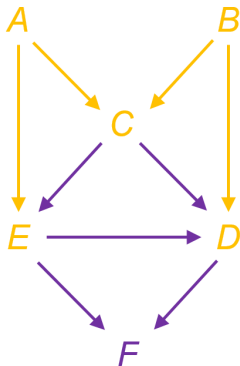
Terminologia



Caminho bloqueado

- ▶ O caminho $E \rightarrow F \leftarrow D$ é **bloqueado** desde que este contenha um colisor (F).

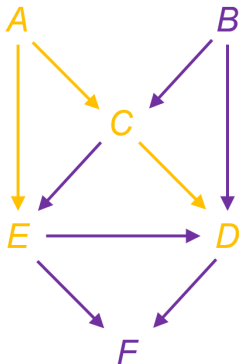
Terminologia



Caminho bloqueado

- ▶ Este caminho também é bloqueado (em C).

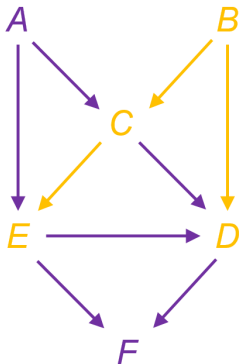
Terminologia



Caminho aberto

- Um caminho que não contém um colisor está **aberto**. Aqui temos um exemplo.

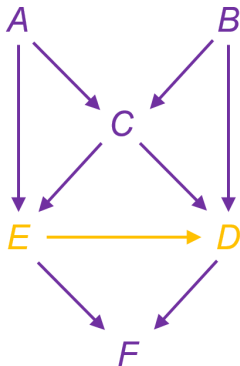
Terminologia



Caminho aberto

► E outro.

Terminologia



Caminho aberto

- ▶ E outro.

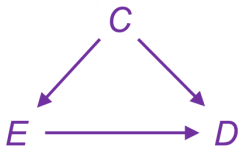
Construindo um DAG



Passo 1

- ▶ O primeiro passo na construção de um DAG para um problema particular é escrever a **exposição** e o **desfecho** de interesse, com uma seta da exposição para o desfecho.
- ▶ Esta seta representa o **efeito causal** que queremos estimar.

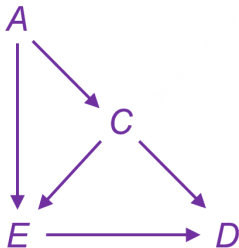
Construindo um DAG



Passo 2

- ▶ Se existir qualquer **causa comum** C de E e D , devemos colocá-lo no grafo, com setas de C para E e de C para D .
- ▶ Devemos incluir C no grafo, independentemente deste ter sido ou não mensurado em nosso estudo.

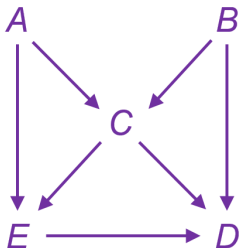
Construindo um DAG



Passo 3

- ▶ Continuamos assim, adicionando ao diagrama qualquer variável (observada ou não observada) que é uma **causa comum** de duas ou mais variáveis já existentes no diagrama.

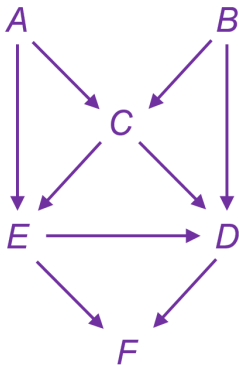
Construindo um DAG



Passo 3

- ▶ Continuamos assim, adicionando ao diagrama qualquer variável (observada ou não observada) que é uma **causa comum** de duas ou mais variáveis já existentes no diagrama.

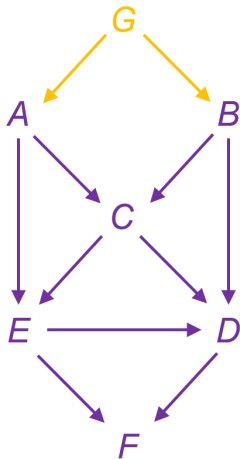
Construindo um DAG



Passo 3

- ▶ Se quisermos, podemos também incluir **outras variáveis**, mesmo que eles não sejam causas comuns de outras variáveis no diagrama.
- ▶ Por exemplo, *F*.
- ▶ Vamos supor que finalizamos nesse ponto. As variáveis e setas que **NÃO** estão em nosso grafo representam nossas **suposições causais**.

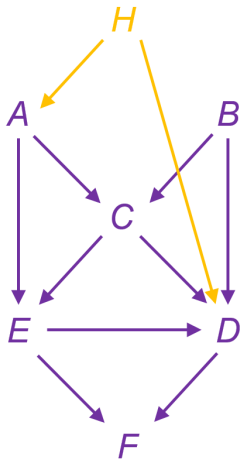
Construindo um DAG



Quais são as nossas suposições?

- Por exemplo, estamos fazendo a suposição que não há uma causa comum G de A e B .

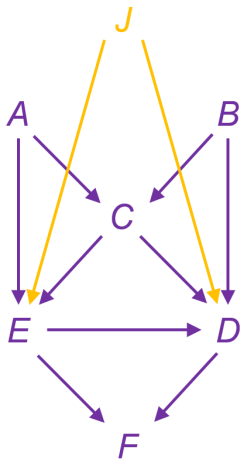
Construindo um DAG



Quais são as nossas suposições?

- E que não há uma causa comum H de A e D .

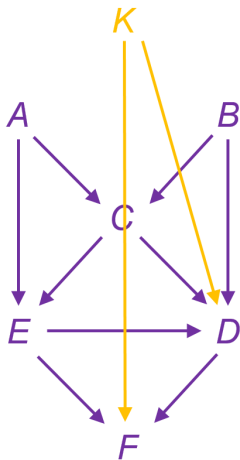
Construindo um DAG



Quais são as nossas suposições?

- E que A , B e C representam TODAS as causas comuns de E e D ; não há uma causa comum adicional J .

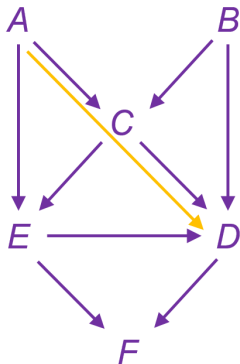
Construindo um DAG



Quais são as nossas suposições?

- E que não há uma causa comum adicional K de D e F .

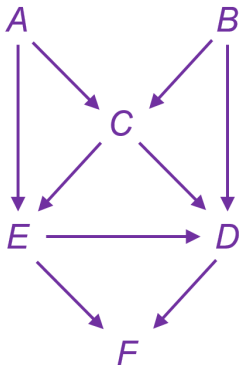
Construindo um DAG



Quais são as nossas suposições?

- ▶ Portanto, cada seta omitida também representa uma suposição.
- ▶ Por exemplo, estamos assumindo que todo o efeito de A em D atua por meio de C e E .

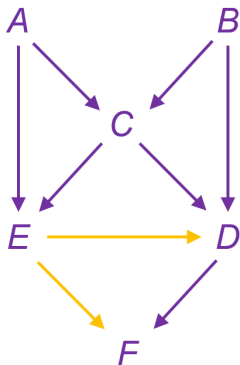
O critério back-door: existe confundimento?



Qual o próximo passo?

- ▶ **SE** acreditarmos em nosso diagrama causal, podemos proceder para determinar se a relação $E \rightarrow D$ está **confundida** ou não.
- ▶ Isto é feito utilizando o **critério porta dos fundos**.
- ▶ O critério porta dos fundos é aplicado em duas partes:
 1. a primeira parte define se existe ou não confundimento.
 2. se existir, a segunda parte determina se é possível controlar o confundimento.

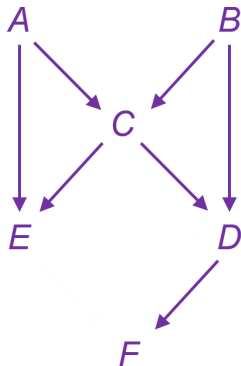
O critério back-door: existe confundimento?



Passo 1

- Primeiro removemos todas as setas **saindo da exposição**.

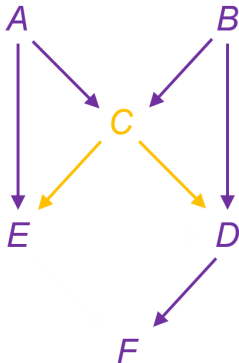
O critério back-door: existe confundimento?



Passo 2

- ▶ Em seguida, procuramos por caminhos abertos a partir da exposição até o desfecho.
- ▶ Relembrando: um caminho aberto não contém um colisor.

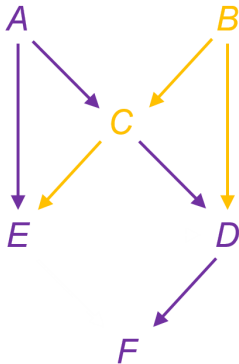
O critério back-door: existe confundimento?



Passo 2

- Este é um caminho aberto?

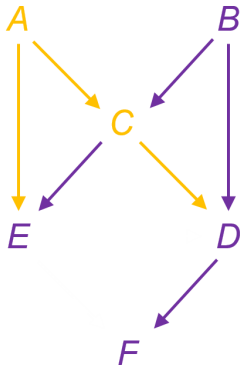
O critério back-door: existe confundimento?



Passo 2

- Este é um caminho aberto?

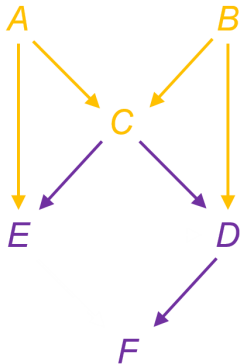
O critério back-door: existe confundimento?



Passo 2

- Este é um caminho aberto?

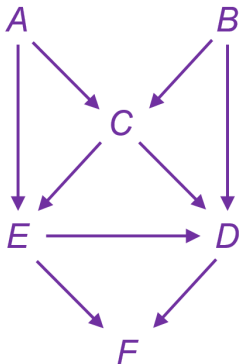
O critério back-door: existe confundimento?



Passo 2

- Este é um caminho aberto?

O critério back-door: existe confundimento?



Existe confundimento?

- ▶ Identificamos três caminhos porta dos fundos abertos de E para D . Assim, há confundimento.
- ▶ Próxima pergunta: podemos usar alguns ou todos de A , B , C , F para controlar esse confundimento?
- ▶ Existe um conjunto \mathcal{S} de variáveis tal que se estratificarmos (ajustarmos) por elas, podemos concluir que o efeito causal existe no estrato?

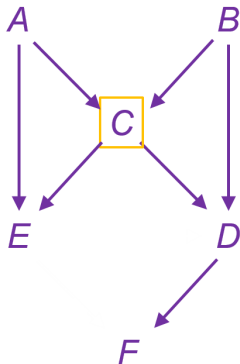
O critério back-door

A segunda parte do critério da porta dos fundos nos permite determinar, com base em nosso diagrama causal, se um conjunto de covariáveis candidato é ou não suficiente para controlar o confundimento:

O critério back-door

- (i) Primeiro, o conjunto candidato S não deve conter **descendentes da exposição**.
- (ii) Em seguida, removemos todas as setas que saem da exposição.
- (iii) Então, nós **juntamos com uma linha tracejada** quaisquer duas variáveis que compartilham um filho que esteja ela mesma em S ou que tenha um descendente em S .
- (iv) Existe um caminho aberto de E para D que não passa por um membro de S ? Se NÃO, então S é **suficiente** para controlar para confundimento.

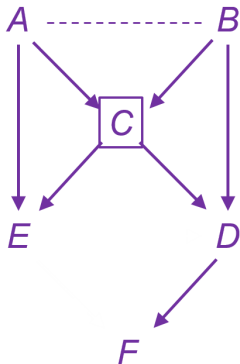
O critério back-door: podemos controlar o confundimento?



O critério back-door: passos (i) e (ii)

- ▶ C é suficiente?
- ▶ C não é um descendente de E , então o passo (i) é satisfeito.
- ▶ Todas as setas saindo da exposição já foram removidas (passo (ii)).

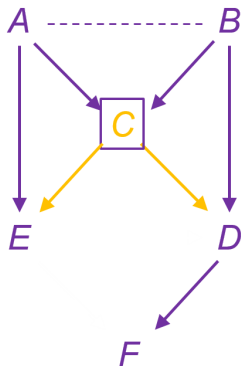
O critério back-door: podemos controlar o confundimento?



passo (iii)

- **Conectamos** A e B com uma linha tracejada, pois eles compartilham um filho (C) que está em nosso conjunto candidato (C).
- Nenhuma outra variável precisa ser conectada desta forma.

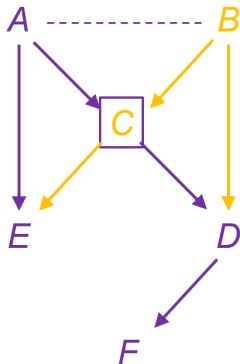
O critério back-door: podemos controlar o confundimento?



passo (iv)

- ▶ Agora procuramos por caminhos abertos de E para D e vemos se estes todos passam por C .
- ▶ Este está OK!

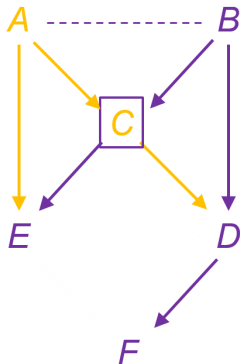
O critério back-door: podemos controlar o confundimento?



passo (iv)

► Este também!

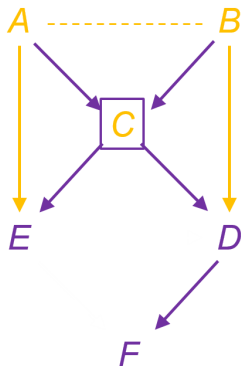
O critério back-door: podemos controlar o confundimento?



passo (iv)

- Este também!

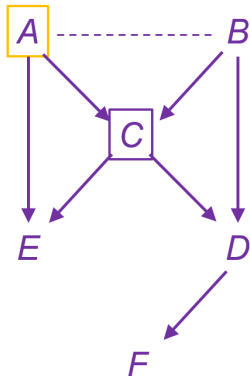
O critério back-door: podemos controlar o confundimento?



passo (iv)

- ▶ PORÉM, aqui está um caminho aberto de E para D que NÃO passa por C
- ▶ Assim, controlar apenas por C NÃO é suficiente.

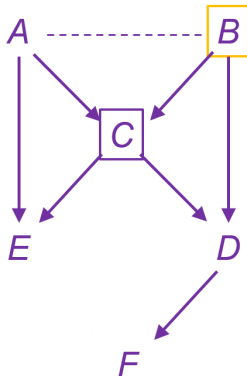
O critério back-door: podemos controlar o confundimento?



Qual é a solução?

- Devemos controlar adicionalmente para A.

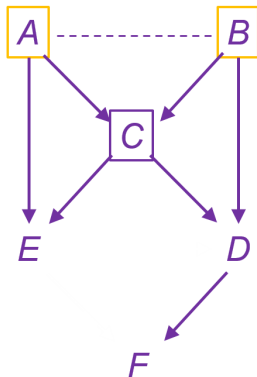
O critério back-door: podemos controlar o confundimento?



Qual é a solução?

► Ou B .

O critério back-door: podemos controlar o confundimento?

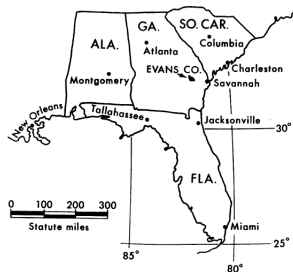


Qual é a solução?

- Ou ambos A e B para controlar para o confundimento.

Atividade 1

Doença coronariana e níveis de catecolaminas



- ▶ Com base em experimentos com animais, a função de emergência da medula adrenal foi estudada.
- ▶ O sistema nervoso simpático é ativado quando um indivíduo é desafiado.
- ▶ Via o hipotálamo, o estresse psicológico estimula a medula adrenal para liberar as duas catecolaminas, epinefrina (adrenalina) e norepinefrina (noradrenalina) na corrente sanguínea.
 - ▶ Esta reação rápida de defesa prepara o corpo para a batalha.
- ▶ No entanto, nos dias atuais o sistema simpático é desafiado por ameaças de natureza social ou mental.
 - ▶ Elevação da pressão arterial e da frequência cardíaca e da liberação de ácidos graxos livres e glicose na corrente sanguínea numa rotina mentalmente estressante, mas sedentária vai ser prejudicial para o organismo, principalmente para o sistema cardiovascular.

Doença coronariana e níveis de catecolaminas

- ▶ Com este estudo de caso teremos a oportunidade de comparar a incidência de doença coronariana (CHD) em dois grupos de exposição (níveis alto e baixo de catecolaminas).

P1: Esta relação é causal?

P2: Quais as suposições necessárias para o estabelecimento de uma relação causal?

P3: É necessário a análise dos dados para responder as perguntas acima?

Dados

- ▶ Os dados considerados neste estudo de caso são derivados de um estudo de coorte – *Evans County Heart Disease Study* – realizado entre 1960 e 1969.
- ▶ Estes dados dizem respeito a uma coorte de 609 indivíduos do sexo masculino, brancos, com idades entre 40 e 76 anos, livres de **doença coronariana** e residentes no distrito de Evans, Georgia, em 1960.
- ▶ Após sete anos, toda a coorte foi reexaminada e os novos casos de **doença coronariana** foram identificados.
- ▶ **Nível de catecolaminas** é a variável exposição de interesse.
- ▶ Todas as variáveis, com exceção de CHD, foram medidas na linha de base (1960).

Variáveis

<i>Nome da variável</i>	<i>Descrição da variável</i>	<i>Código da variável</i>
CHD	Ocorrência de doença coronariana	0 = não caso 1 = novo caso
CAT	Nível sérico de catecolaminas	0 = baixo 1 = alto
SMK	Tabagismo	0 = nunca fumou 1 = fumante
ECG	Alterações do eletrocardiograma	0 = ECG normal 1 = qualquer alteração
OCC	Tipo de ocupação	0 = não agricultor 1 = agricultor
MAR	Estado civil	0 = não casado 1 = casado

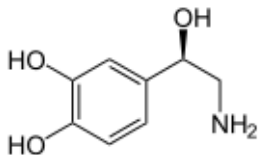
Variáveis

<i>Nome da variável</i>	<i>Descrição da variável</i>	<i>Código da variável</i>
AGE	Idade	Anos
CHL	Colesterol	mg/100 mL
SBP	Pressão arterial sistólica	mmHg
DBP	Pressão arterial diastólica	mmHg
QTI	Índice de Quételet	$100 \times [\text{peso (lb)}/\text{altura (in.)}^2]$
HEM	Hematócrito	Percentual
SES	Status socioeconômico	Índice de McGuire-White
PLS	Pulsção	Batidas/minuto

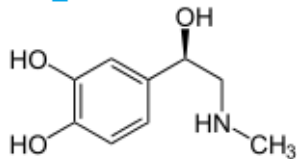
Dados (amostra grátis)

id	chd	cat	age	chl	smk	ecg	dbp	sbp	hpt	cc	ch
21	0	0	56	270	0	0	80	138	0	0	0
31	0	0	43	159	1	0	74	128	0	0	0
51	1	1	56	201	1	1	112	164	1	1	201
71	0	1	64	179	1	0	100	200	1	1	179
74	0	0	49	243	1	0	82	145	0	0	0
91	0	0	46	252	1	0	88	142	0	0	0

Conhecimento a priori



P: Faça uma discussão a respeito dos fatores psicossociais associados a liberação de catecolaminas. Quais os efeitos cardiovasculares são esperados destas catecolaminas?



Prática

- ▶ Construa o **diagrama causal** (DAG) para avaliar o efeito da exposição na resposta de interesse.

- ▶ Utilizando o critério *back-door* (**apresente os passos utilizados**) responda:

1. Existe confundimento? O que isso implica em termos de suposições de identificação do efeito causal?
2. Se existe confundimento, é possível controlar este confundimento?
3. Ainda, se existe confundimento, quais são as variáveis do conjunto suficiente para controle/ajuste?



Obrigado!

