Estatística aplicada à epidemiologia II
Modelos para desfecho binário

Leo Bastos – leonardo.bastos@fiocruz.br

PROCC – Fundação Oswaldo Cruz

https://github.com/lsbastos/eae2

FIOCRUZ

Atenção

- NÃO ESTUDEM PELOS SLIDES!
- Referências utilizadas:

  1. VanderWeele, T.J. (2015) *Explanation in causal inference: methods for mediation and interaction*. Oxford University press.
  2. Richiardi et al. (2013) Mediation analysis in epidemiology: methods, interpretation and bias. *IJE*
  3. VanderWeele, T.J. (2015) Mediation Analysis: A Practitioner's Guide. *Annual Review of Public Health*
  4. Lange, T., et al. (2017) Applied mediation analyses: a review and tutorial. *Epidemiology and Health*. (Open access)
  5. Steen, J., et al. (2017) Medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. *Journal of Statistical Software*.

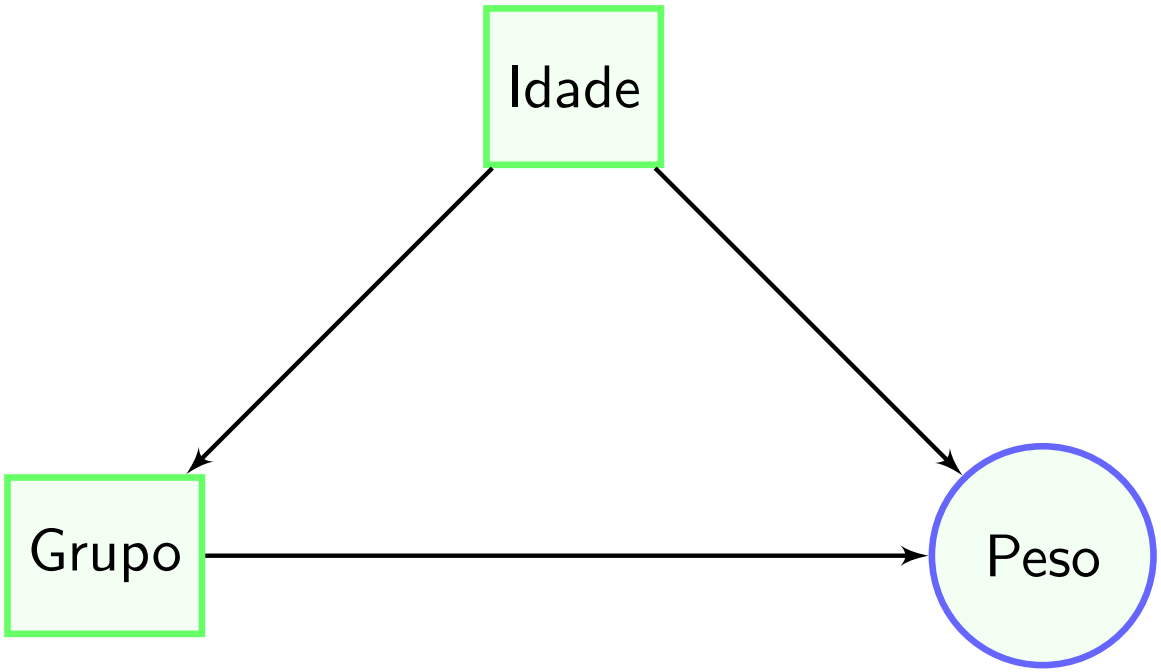
FIOCRUZ

Exemplo hipotético

- Suponha que estamos interessados em avaliar a diferença de peso de meninos divididos um dois grupos de atividades física {Grupo A, Grupo B}.
- Suponha também que esses dois grupos se diferem na idade.
- Como a idade influencia no peso, a diferença do peso médio entre os grupos iria refletir não apenas o efeito do grupo, mas também o efeito da idade.
- Para isolar o efeito do grupo, precisamos comparar o peso dos garotos de mesma idade.

FIOCRUZ

DAG do exemplo



FIOCRUZ

Exemplo

- A idade é uma variável confundidora
- Nesse caso, a solução é simples, basta incluirmos a variável confundidora no modelo como uma variável explicativa.
- Note que o efeito do grupo no peso pode ser diferente entre diferentes idades, então é necessário verificarmos a presença de interação.
- Lembrem-se do exemplo do Titanic, onde o efeito da idade e sexo era diferente entre as diferentes classes sociais:
    - As chances de um homem na terceira classe sobreviver no Titanic eram menores que as chances de um homem de primeira classe, que por sua vez eram menores que de uma mulher na primeira classe.

FIOCRUZ

Confundimento: Regra geral

- A variável confundidora está associada de forma causal com o desfecho
- A variável confundidora está associada com a exposição, ou ela causa a exposição ou ela está associada a algo que causa a exposição.
- A variável confundidora NÃO está no caminho causal entre a exposição e o desfecho.
- Se a variável estiver no caminho causal entre exposição e desfecho, ela é **mediadora**.

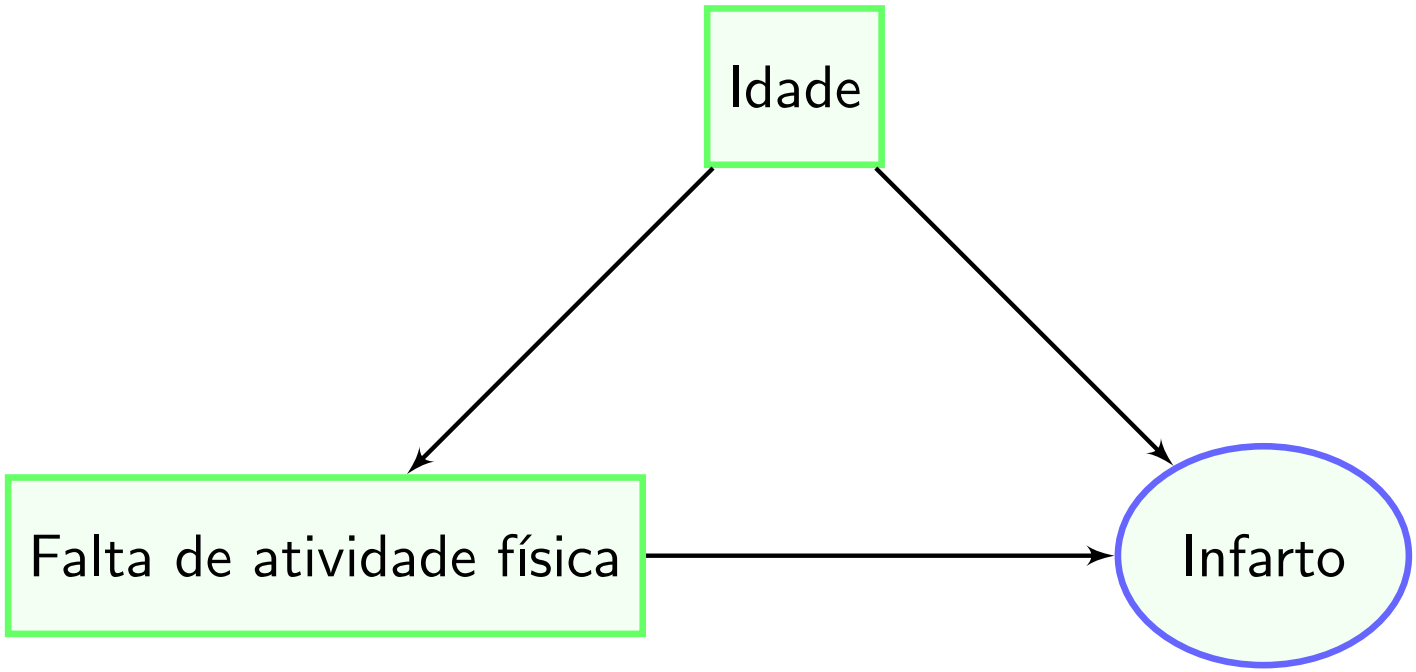
FIOCRUZ

Exemplo

- Suponha que nós estamos interessados em medir o efeito da falta de atividade física em provocar um infarto no miocárdio.
- Existem outras variáveis que têm efeito no infarto do miocárdio que possam estar assocaidas com a falta de atividade física.
- Por exemplo:
  - Idade
  - Hábitos de alimentação não saudáveis
  - Hipertensão

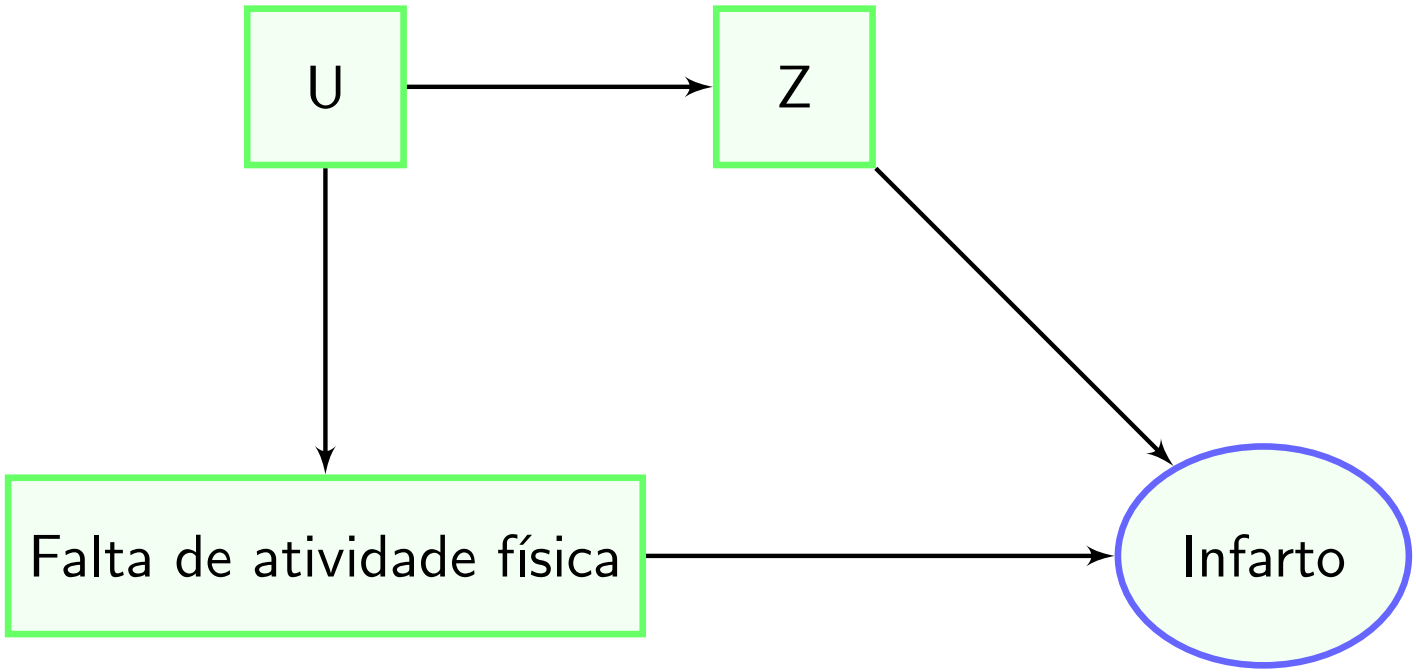
FIOCRUZ

Idade



FIOCRUZ

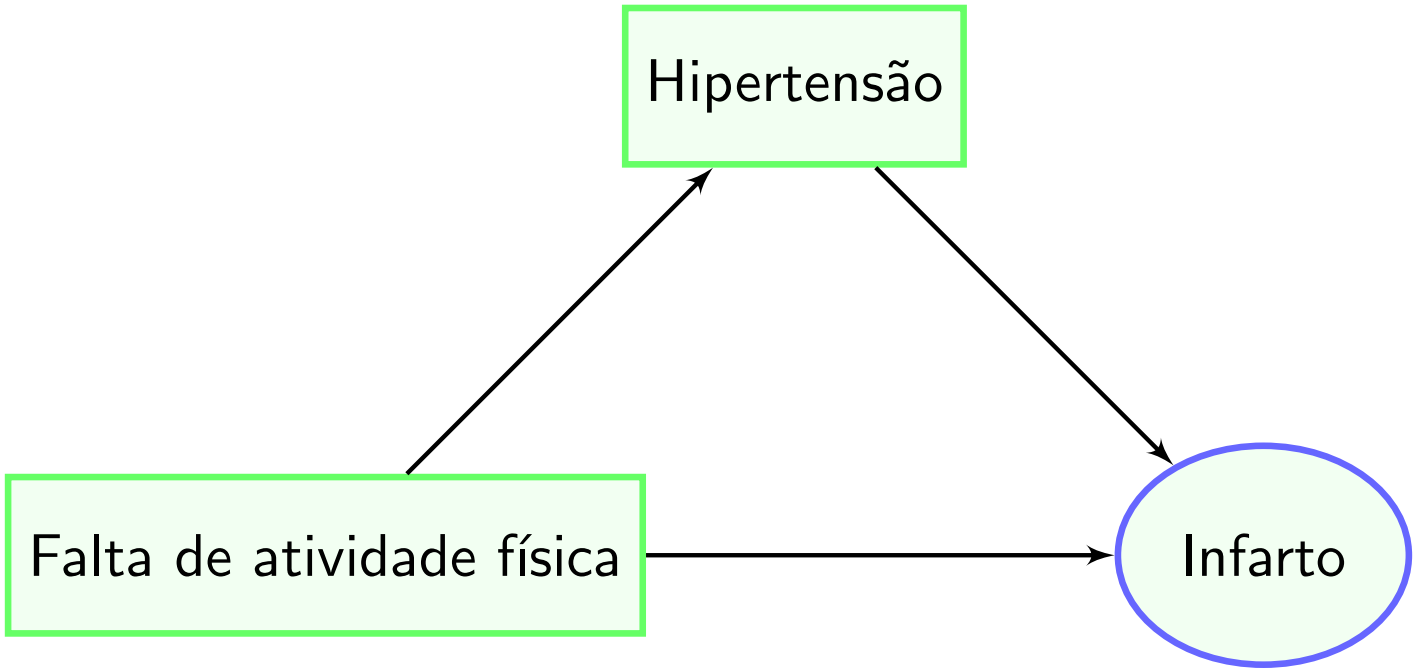
Idade



U = Falta de preocupação com a saúde.
Z = Hábito alimentar não saudável.

FIOCRUZ

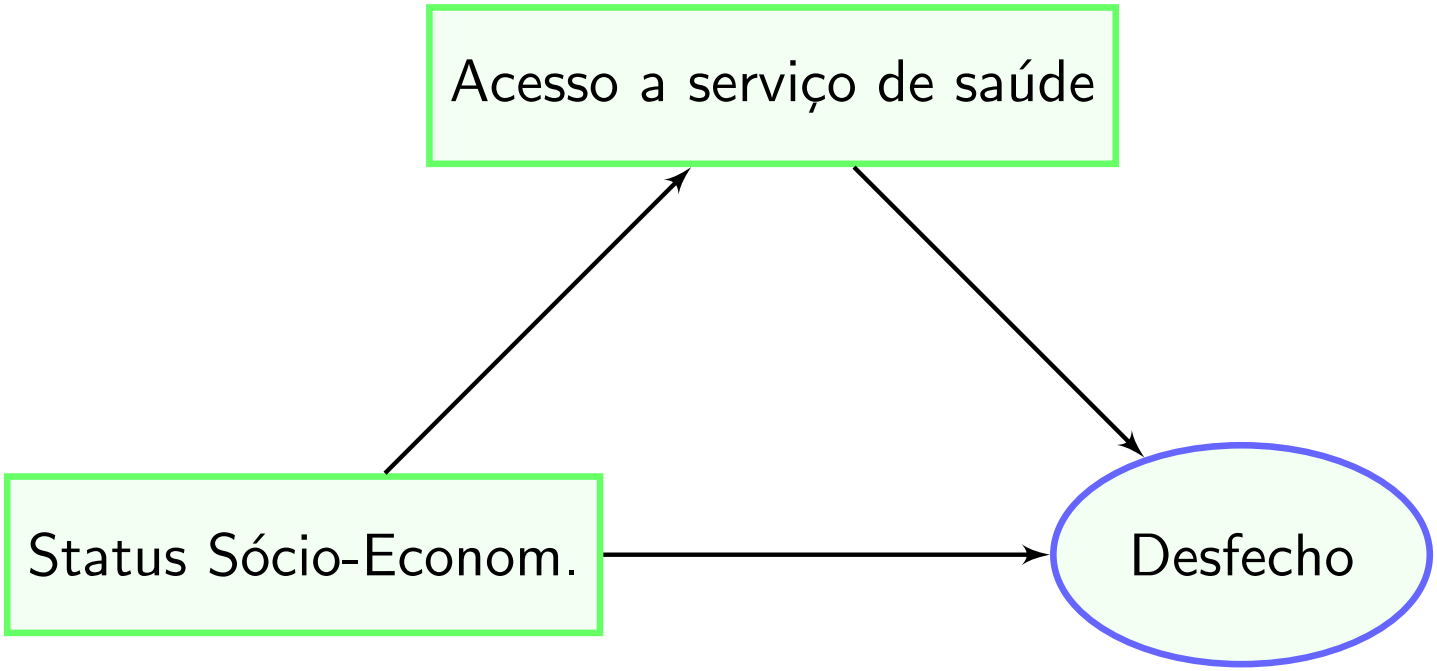
Hipertensão



Hipertensão é mediadora, pois é consequência da falta de atividade física, e pode ser causa do infarto.

FIOCRUZ

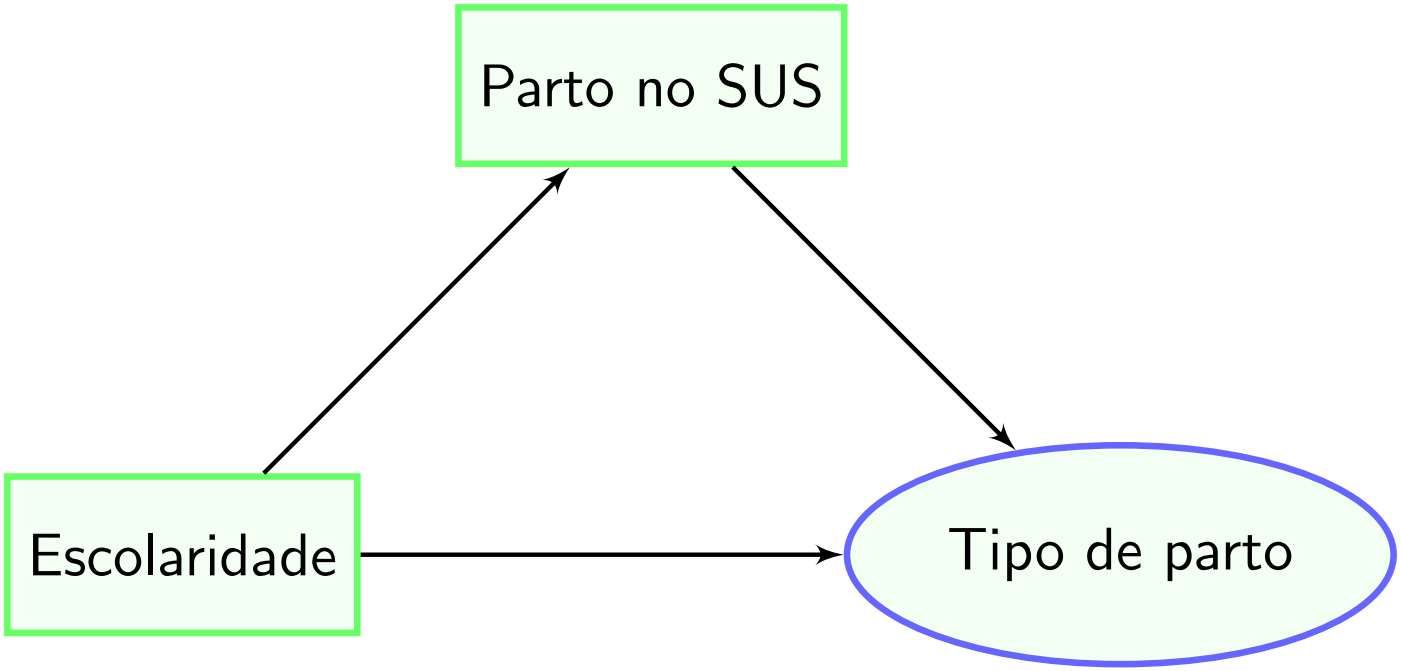
Outros exemplos



Acesso a serviço de saúde funciona como mediador de vários desfechos.

FIOCRUZ

Outros exemplos



Acesso a serviço de saúde funciona como mediador de vários desfechos.

FIOCRUZ

Interpretação

- Quando uma variável M é incluída no modelo junto com a variável explicativa X, temos o seguinte modelo

$$Y \sim f(\cdot)$$

com valor esperado dado por

$$g(\mathbb{E}[Y]) = \alpha + \beta_x X + \beta_z Z$$

- Se Z está associada com X e Y, não é possível distinguirmos estatísticamente se Z é confundidora ou mediadora.
- Por outro lado, a interpretação de β_x é diferente se Z for confundidora ou mediadora.

FIOCRUZ

Z confundidora

- Se Z é confundidora, e assumindo que não existam outras confundidoras, então β_x é o efeito da exposição no desfecho.
- Ao incluirmos Z no modelo, a estimativa de β_x está mais próxima do verdadeiro efeito causal do que seria se a confundidora Z não fosse incluída.
- Para um mesmo valor de Z, β_x representa a mudança no desfecho quando aumenta-se X em uma unidade.

FIOCRUZ

Z mediadora

- Se Z é mediadora, β_x **não** representa a mudança no desfecho quando aumenta-se X em uma unidade.
- Na verdade β_x representa só parte do efeito de X no desfecho, a parte do efeito que não passa por Z.
- Na literatura de mediação, esse efeito é chamado de **efeito direto** de X no desfecho.
- O efeito de X no desfecho é dado pelo **efeito direto**, medido por β_x, somado ao **efeito indireto** que é o efeito de X que passa pelo mediador.

FIOCRUZ

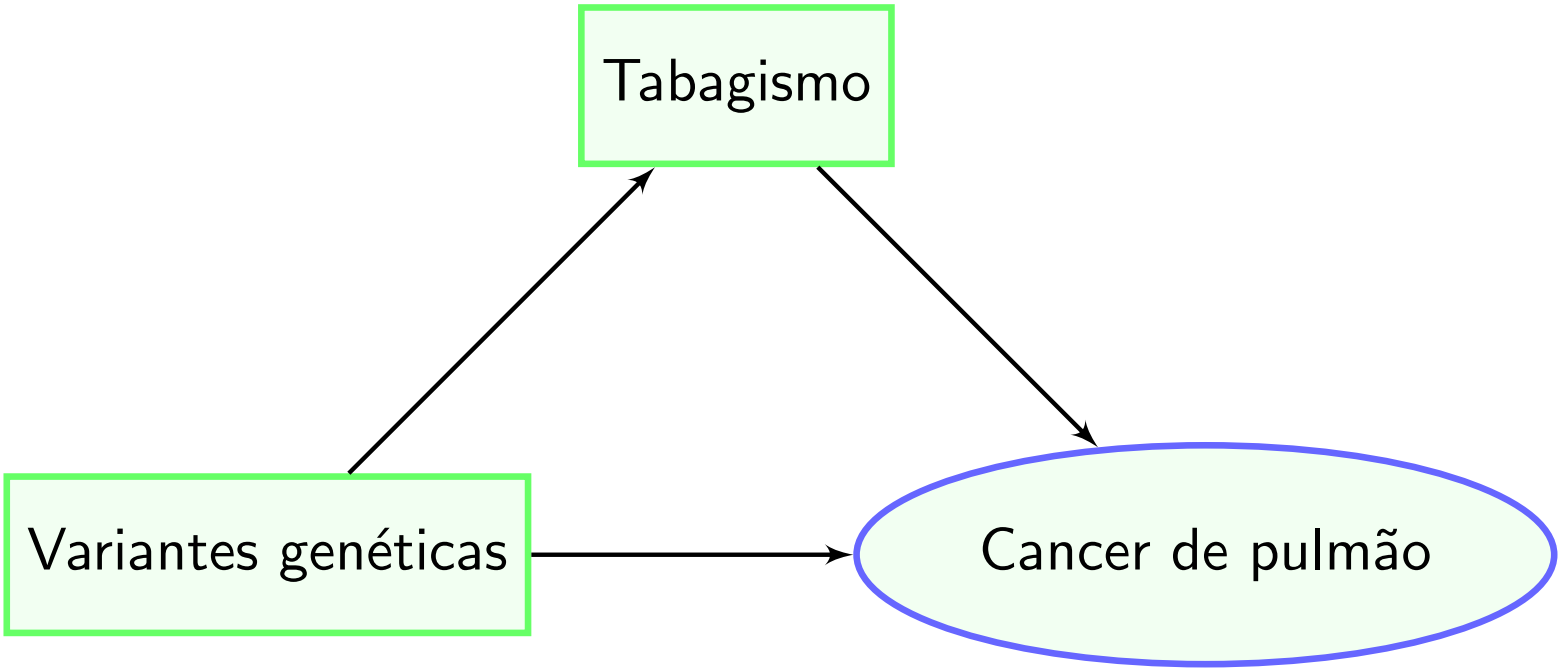
Análise de mediação

- De uma forma geral, a análise de mediação consiste em uma coleção de ferramentas e formas de pensamento que visam ajudar os pesquisadores a identificar, formalizar, e quantificar possíveis mecanismos que ligam a causa a um efeito. (Caminhos causais).
- A descoberta das bactérias como fator causal de certas infecções por Louis Pasteur, tem como linha de pensamento a mediação.
- A análise de mediação tem duas linhas de pensamento:
  - A mediação tradicional, proposta por Baron and Kenny (1986)
  - A mediação causal, que estende a mediação tradicional no contexto da lógica contra-factual.

FIOCRUZ

Análise de mediação

Exemplo:

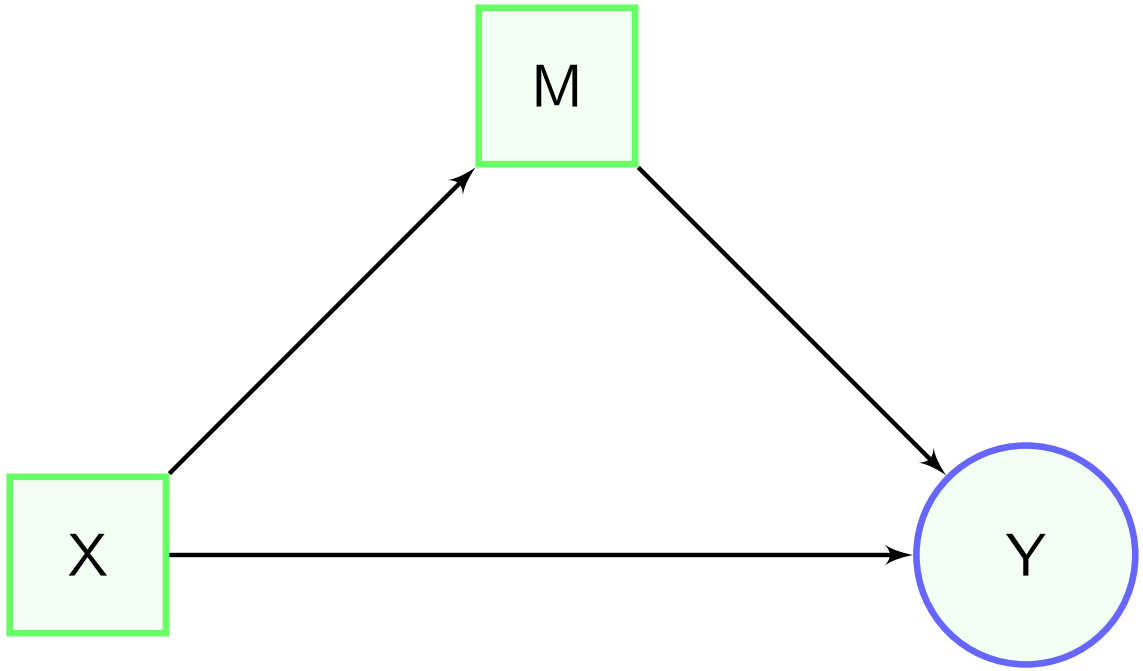


Thorgeirsson et al. (2008, Nature), Fisher (1958, Nature), VanderWeele et al. (2012, AJE)

FIOCRUZ

Análise de mediação

De forma geral (mantendo a notação do livro do VanderWeele):



- X: variável exposição
- M: variável mediadora

FIOCRUZ

Análise de mediação tradicional

A abordagem proposta por Baron & Kenny (1988), chamada de mediação tradicional, consiste no ajuste de 3 modelos:

- Modelo: Desfecho x exposição

$$\mathbb{E}[Y|X = a] = \alpha_0 + \alpha_1 x$$

- Modelo: Desfecho x exposição e moderadora

$$\mathbb{E}[Y|X = x, M = m] = \theta_0 + \theta_1 x + \theta_2 m$$

- Modelo: Moderadora x exposição

$$\mathbb{E}[M|X = x] = \beta_0 + \beta_1 x$$

FIOCRUZ

Análise de mediação

- A proposta de Baron & Kenny é chamada de metodo do produto:
  - O efeito direto de X é dado por θ_1
  - O efeito indireto de X é dado por $\beta_1\theta_2$
- Limitações: Esse método não se aplica para modelos não-lineares e nem na presença de interações.
- Para isso, usaremos a mediação causal usando a lógica contrafactual.
  - Seja $Y_i(x, m)$ o possivel valor do desfecho (potential outcome) para o indivíduo i sob a exposição x e variável mediadora m.

FIOCRUZ

Contrafatos e decomposição dos efeitos

- $Y_i(1, m)$ seria o valor do desfecho do indivíduo i caso a exposição fosse $X = 1$
- $Y_i(0, m)$ seria o valor do desfecho do indivíduo i caso a exposição fosse $X = 0$
- O efeito individual causal direto da exposição no desfecho controlado pela mediadora

$$Y_i(1, m) - Y_i(0, m)$$

- O efeito direto médio controlado é definido por

$$CDE(m) = \mathbb{E}[Y(1, m) - Y(0, m)]$$

FIOCRUZ

Análise de mediação: Mediação causal

- Robins and Greenland (1992) proporam o contrafato composto, $Y(x, M(x^*))$, que leva ao efeito direto natural

$$NDE(0) = \mathbb{E}[Y(1, M(0)) - Y(0, M(0))]$$

que seria o efeito da exposição no desfecho quando o mediador é fixado no valor natural caso a exposição não ocorresse.

- VanderWeele and Vansteelandt (2009) definem o efeito total esperado

$$\mathbb{E}[Y(1, M(1)) - Y(0, M(0))]$$

- E por diferença efeito natural indireto esperado

$$NIE(1) = \mathbb{E}[Y(1, M(1)) - Y(1, M(0))]$$

que é a diferença no desfecho se todos fossem expostos ($X = 1$) mas o mediador tivesse mudado para o cenário onde não tivesse sido exposto.

FIOCRUZ

Análise de mediação

- Ajustando dois modelo lineares, um para Y outro para M temos que

$$\begin{aligned} \mathbb{E}[Y|X = x, M = m] &= \theta_0 + \theta_1 x + \theta_2 m \\ \mathbb{E}[M|X = x] &= \beta_0 + \beta_1 x \end{aligned}$$

- Ignorando outras confundidoras e possíveis interações
- O efeito direto natural:

$$NDE(0) = \theta_1$$

- O efeito indireto natural:

$$NIE = \beta_1 \theta_2$$

FIOCRUZ

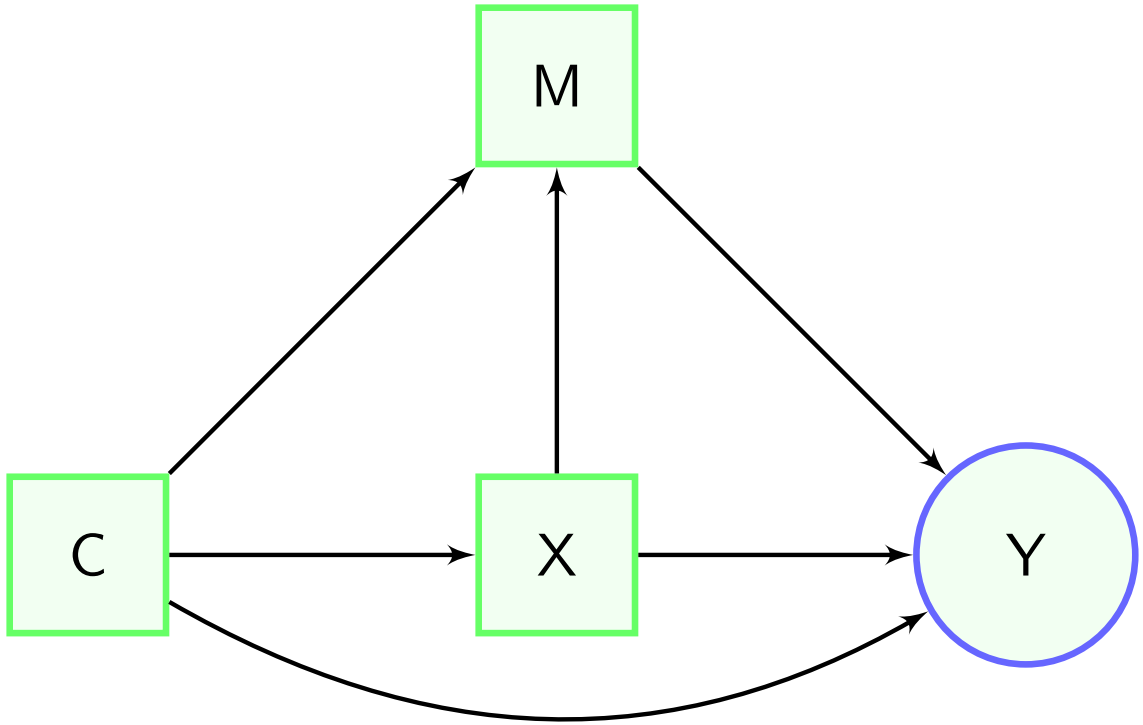
Análise de mediação

Os efeitos diretos e indiretos só são válidos se as suposições abaixo são verdadeiras:

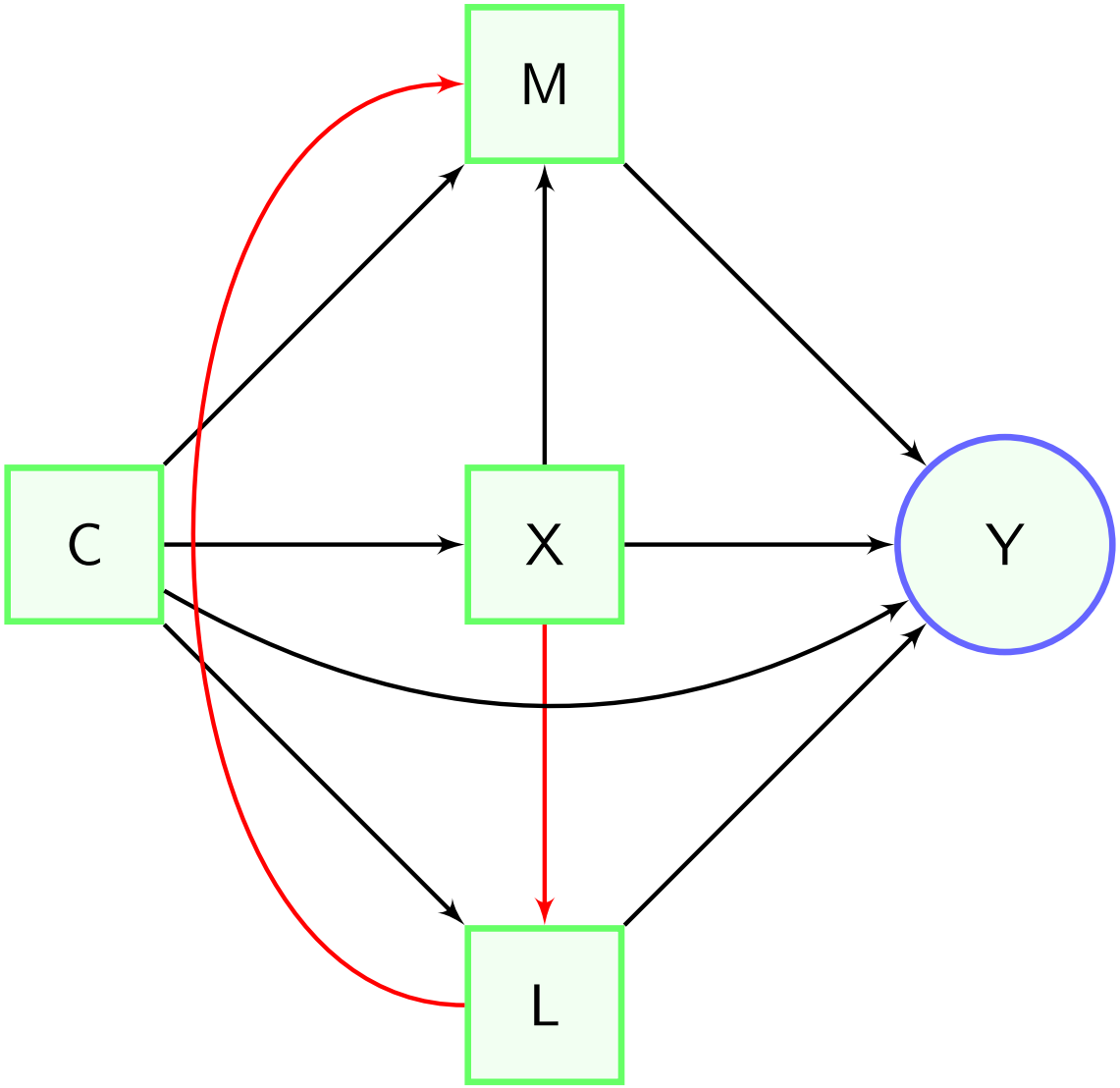
- A1 Estamos controlando o confundimeto na relação exposição-desfecho.
- A2 Estamos controlando o confundimeto na relação mediador-desfecho.
- A3 Estamos contolando o confundimento na relação exposição-mediador.
- A4 Nenhum confundidor da relação mediador-desfecho pode ser afetado pela exposição.

FIOCRUZ

DAG



DAG



FIOCRUZ

Análise de mediação

- Suponha agora um desfecho binário e um mediador também binário.

$$Y(x, m) \equiv Y|X = x, M = m \sim Bernoulli(\pi(x, m))$$

onde

$$logit(\pi(x, m)) = \theta_0 + \theta_1 x + \theta_2 m$$

- Se M confundidor, então e^{θ_1} seria a OR da exposição.
- Se M mediador, então e^{θ_1} mede o efeito direto de X em Y na escala de OR
  - Mas não seria o efeito total de X em Y.

FIOCRUZ

Análise de mediação

- O efeito natural total (NTE)

$$OR^{(NTE)} = \frac{odds(Y(1,1))}{odds(Y(0,0))}$$

- O efeito natural direto (NDE)

$$OR^{(NDE)} = \frac{odds(Y(1,0))}{odds(Y(0,0))} = e^{(\theta_1)}$$

- O natural efeito indireto (NIE)

$$OR^{(NIE)} = \frac{odds(Y(1,1))}{odds(Y(1,0))}$$

- Notem que:

$$OR^{(NTE)} = OR^{(NDE)} \times OR^{(NIE)}$$

FIOCRUZ

Estimando os efeitos naturais

Passo a passo apresentado em Lange et al. (2017) e implementado no pacote medflex.

1. Ajustar o modelo com exposição e mediadora;

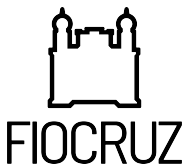
$$Y \sim X + M + Z_1 + \cdots + Z_p$$

2. Expandir o dado segundo a lógica contrafactual
3. Imputar numericamente o valores potenciais
4. Para cada amostra imputada ajustar um modelo com a exposição observada, com a exposição do contrafato, e se houver outras covariáveis.
5. Com as amostras dos coeficientes estimar a incerteza associada. (Como se fosse um bootstrap)

FIOCRUZ

Dados

i	X_i	Y_i	M	Z_1	Z_2
1	1	Y_1	m_1	z_{11}	z_{12}
2	0	Y_2	m_2	z_{21}	z_{22}
3	0	Y_3	m_3	z_{31}	z_{32}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$



Dados estendidos (incorporando o contrafato)

i	X_i	x	x^*	$Y_i(x, M_i(x^*))$	M	Z_1	Z_2
1	1	1	1	Y_1	m_1	z_{11}	z_{12}
1	1	**0**	1	.	m_1	z_{11}	z_{12}
2	0	0	0	Y_2	m_2	z_{21}	z_{22}
2	0	**1**	0	.	m_2	z_{21}	z_{22}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

FIOCRUZ

Dados estendidos (incorporando o contrafato)

i	X_i	x	x^*	$Y_i(x, M_i(x^*))$	M	Z_1	Z_2
1	1	1	1	Y_1	m_1	z_{11}	z_{12}
1	1	**0**	1	$\hat{Y}_1(\mathbf{0}, M_1)$	m_1	z_{11}	z_{12}
2	0	0	0	Y_2	m_2	z_{21}	z_{22}
2	0	**1**	0	$\hat{Y}_2(\mathbf{1}, M_2)$	m_2	z_{21}	z_{22}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

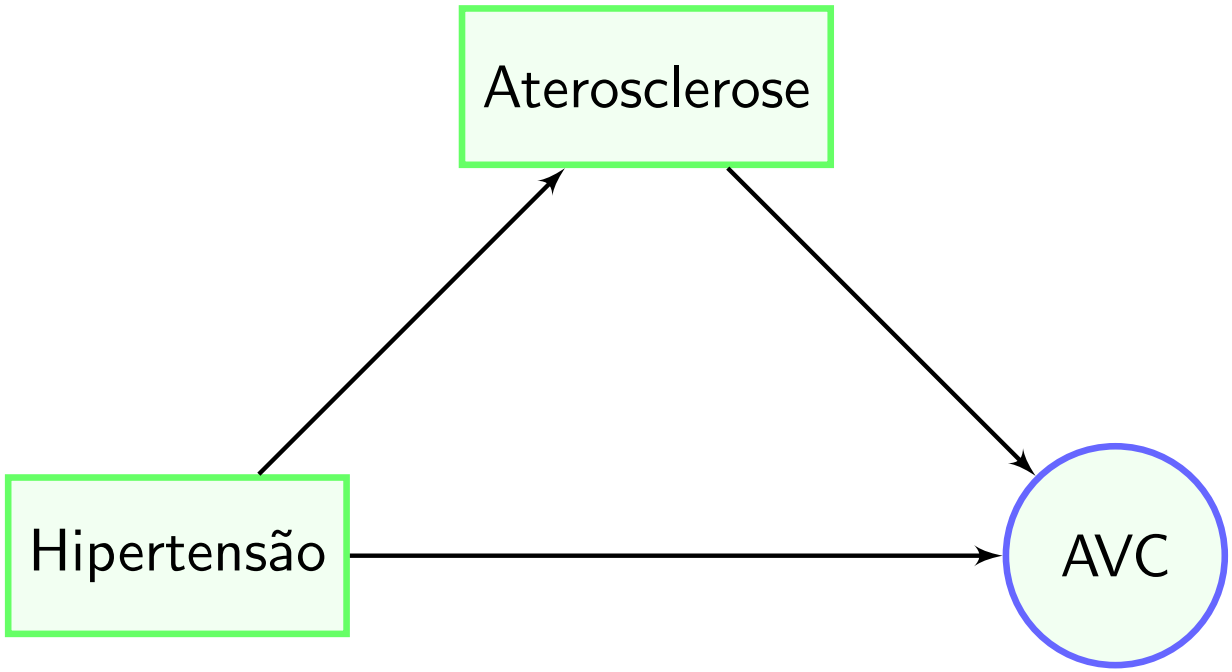
FIOCRUZ

Exemplo: Hipertensão e AVC

- O objetivo é avaliar o efeito da presença de hipertensão na ocorrência de acidente vascular cerebral (AVC)
- Suponha que acompanhamos 1000 pessoas em uma coorte.
  - Desfecho: AVC
  - Exposição: hipertensão
  - Outra variável: aterosclerose

FIOCRUZ

Exemplo: Hipertensão e AVC



FIOCRUZ

Exemplo hipotético

- Efeito da hipertensão

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.3340	0.0897	-14.87	0.0000
X1	1.1309	0.1562	7.24	0.0000

Resultando em uma OR =3.1.

- Efeito da hipertensão controlado pela aterosclerose

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.1617	0.1306	-16.55	0.0000
X1	0.3342	0.1828	1.83	0.0676
M1	2.2260	0.1750	12.72	0.0000

Resultando em uma OR =1.4.

FIOCRUZ

medflex

- Usando o pacote medflex

```
> library(medflex)
> # Ajustando modelo completo
> modelo <- glm(Y ~ X + M, family = binomial(),
+               data = dados)
> # Duplicando o banco para imputacao
> expDados <- neImpute(object = modelo)
> # Ajustando estimando os efeitos naturais
> output <- neModel(formula = Y ~ X0 + X1,
+                   family = binomial("logit"),
+                   expData = expDados)
```

FIOCRUZ

Output

- A saída é dada por

```
> summary(output)
Natural effect model
with standard errors based on the non-parametric bootstrap
---
Exposure: X
Mediator(s): M
---
Parameter estimates:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.33575    0.08628 -15.482   <2e-16 ***
X01          0.26917    0.14717   1.829   0.0674 .
X11          0.85989    0.08588  10.013   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Output

```
> summary(neEffdecomp(output))
Effect decomposition on the scale of the linear predictor
with standard errors based on the non-parametric bootstrap
---
with x* = 0, x = 1
---
                          Estimate Std. Error z value Pr(>|z|)
natural direct effect      0.26917    0.14717   1.829   0.0674 .
natural indirect effect    0.85989    0.08588  10.013  < 2e-16 *
total effect               1.12906    0.15190   7.433 1.06e-13 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Univariate p-values reported)
```

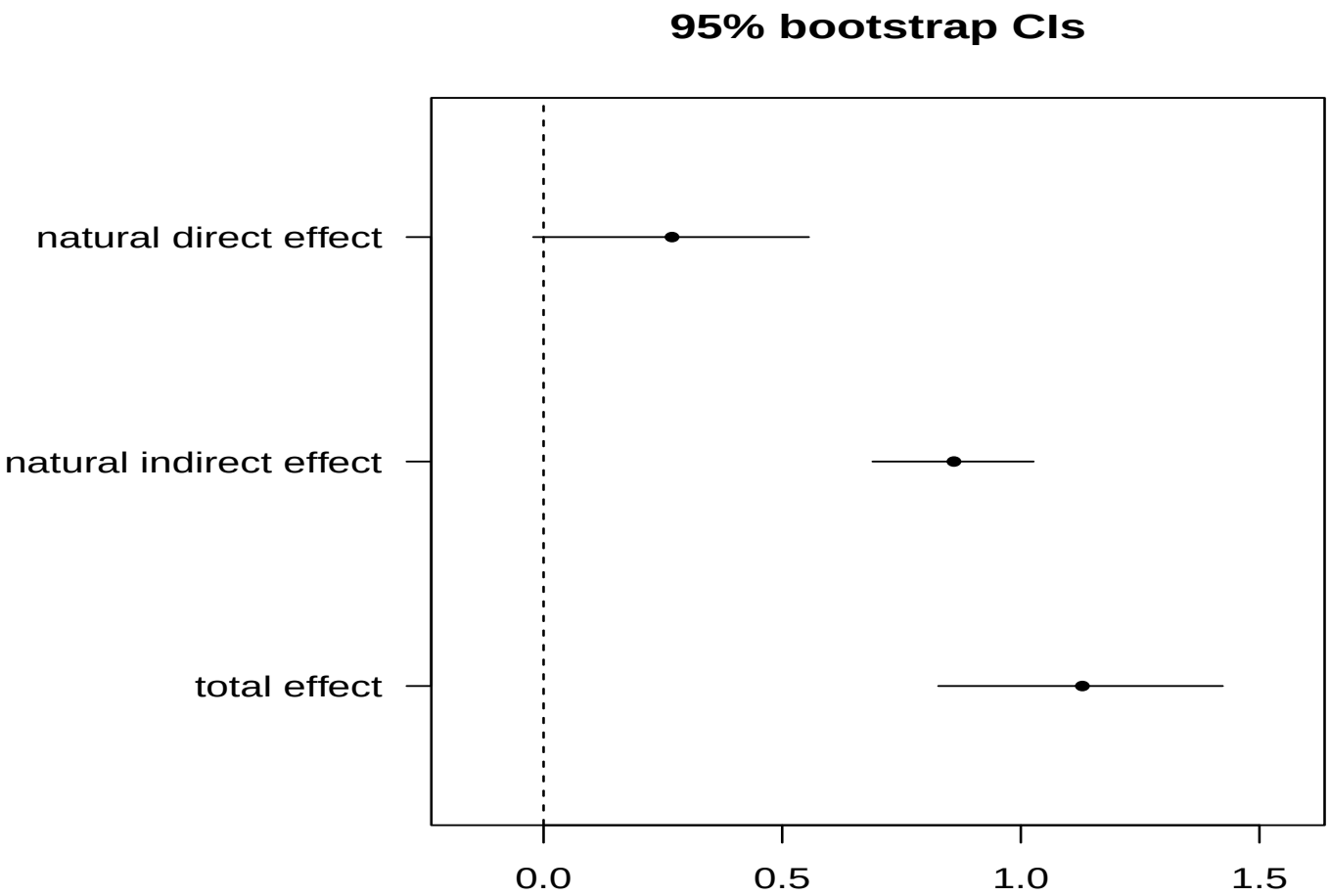
Efeitos naturais

- Os efeitos naturais estimados

	OR	95% LCL	95% UCL
(NDE)	1.31	0.98	1.74
(NIE)	2.36	1.99	2.79

FIOCRUZ

Plot


95% bootstrap CIs
natural direct effect
natural indirect effect
total effect
0.0
0.5
1.0
1.5

Aula prática

- Replicar o exemplo dos slides (exemploSimulado.R)
- Lista de exercicios
- Dados da lista (SINASC de 2016) e lista do CNES (Data/)

FIOCRUZ