Causal Thinking

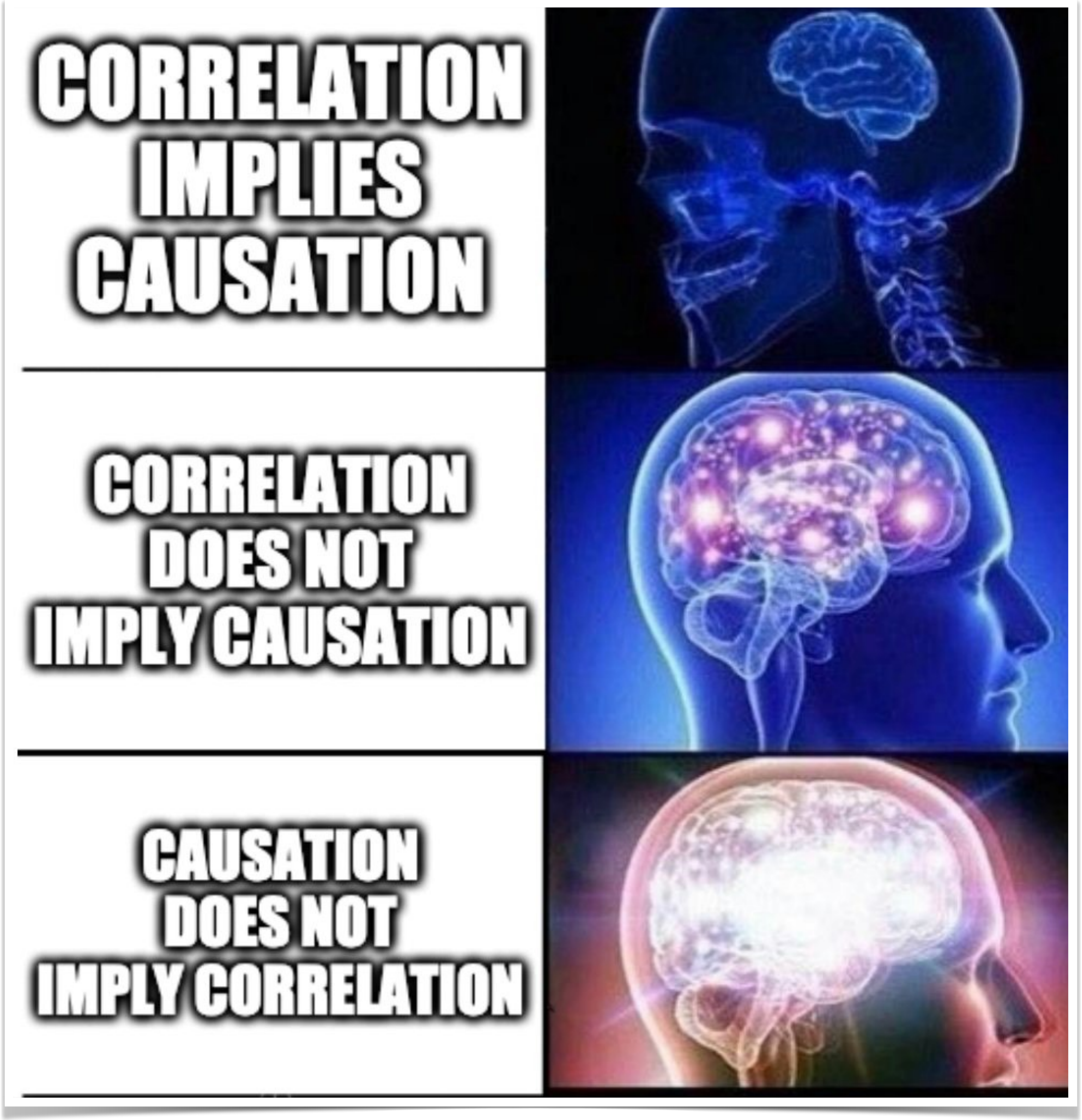
Usando conhecimento científico para gerar modelos estatísticos

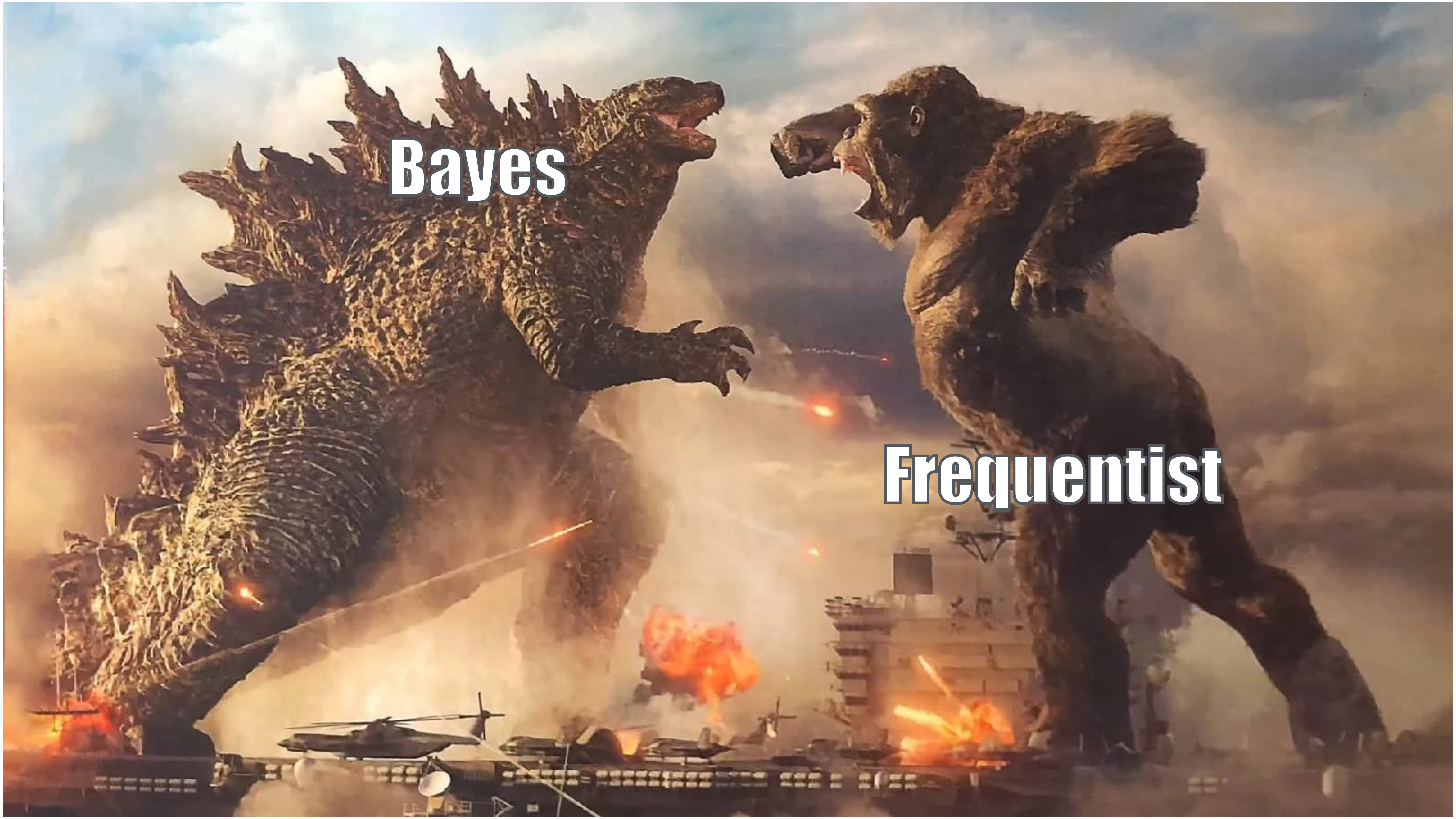
Diogo Melo

Para que servem os modelos?

Previsão vs Inferência

Porque correlação não implica causalidade?


CORRELATION IMPLIES CAUSATION
CORRELATION DOES NOT IMPLY CAUSATION
CAUSATION DOES NOT IMPLY CORRELATION



Joke stollen from Richard McElreath

Joke stollen from Richard McElreath

Representando modelos com Grafos

- Podemos usar um grafo para representar um modelo putativo
- Uma flecha ligando duas variáveis representa um possível efeito causal

$$x \longrightarrow y$$

Este é um grafo acíclico dirigido, um **DAG**

Elemental triads

- Todo DAG pode ser decomposto em 3 motivos básicos:
- O cano (pipe), a forquilha (fork) e o colisor (colider)
- Podemos usar esses motivos para sistematicamente informar a estrutura dos nossos modelos, dependendo da pergunta científica

The pipe:

$$x \longrightarrow z \longrightarrow y$$

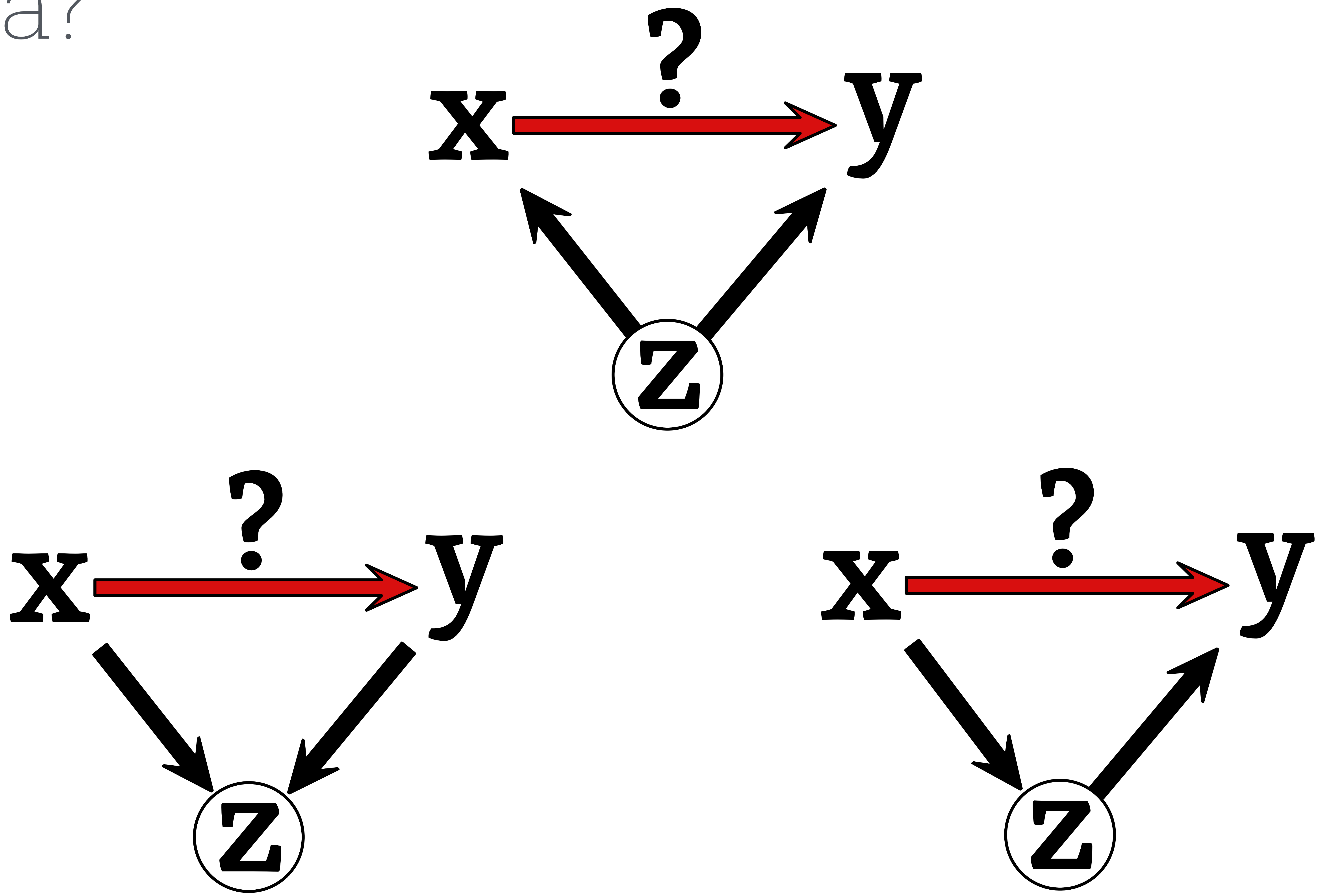
The fork:

$$x \longleftarrow z \longrightarrow y$$

The collider:

$$x \longrightarrow z \longleftarrow y$$

Como uma variável Z influencia nossa inferência?



The fork

Simulando uma Causa Compartilhada

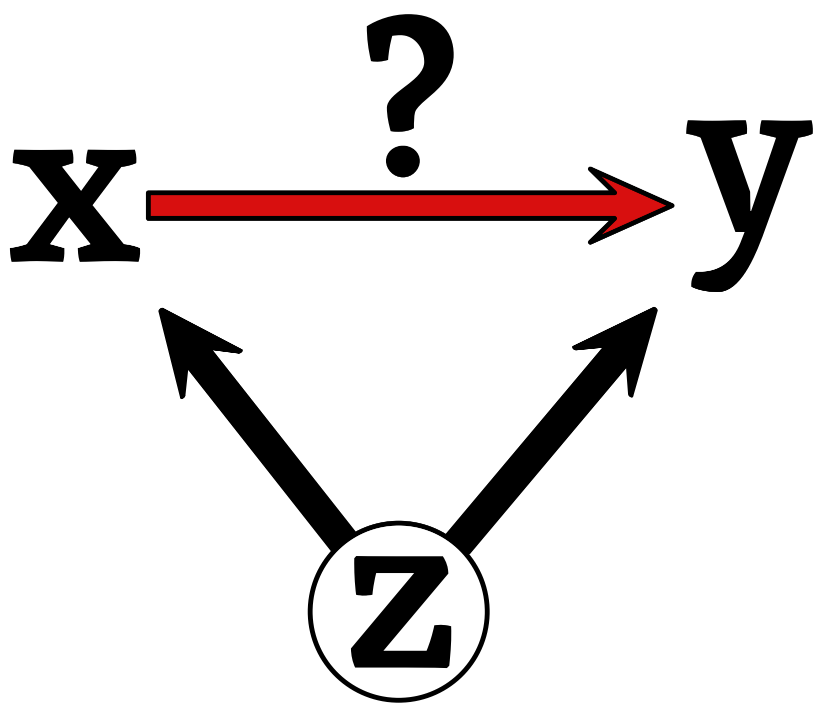
The Fork

- Todo DAG impõem uma relação causal entre as variáveis
- Podemos utilizar simulações para criar o modelo generativo induzido pelo DAG:

$$y \sim Normal(\alpha_y + \beta_{yx}x + \beta_{yz}z, \sigma_y)$$

$$x \sim Normal(\alpha_x + \beta_{xz}z, \sigma_x)$$

$$z \sim Bernoulli(p_z)$$



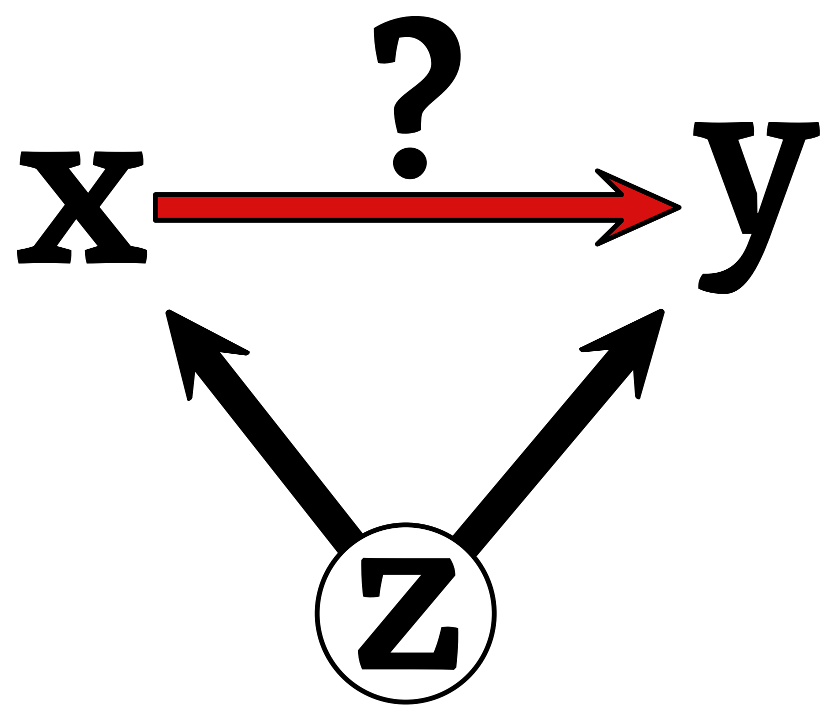
Simulating a shared cause

Math

$$y \sim Normal(\mu = 1 + 0.5x + 2z, \sigma = 1)$$

$$x \sim Normal(\mu = 1 + z, \sigma = 1)$$

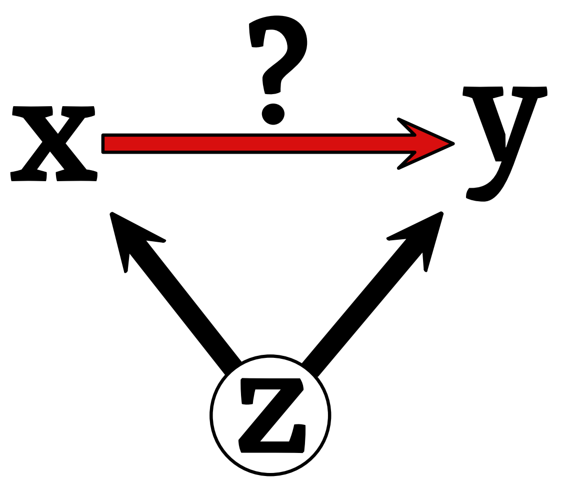
$$z \sim Bernoulli(p = 0.5)$$



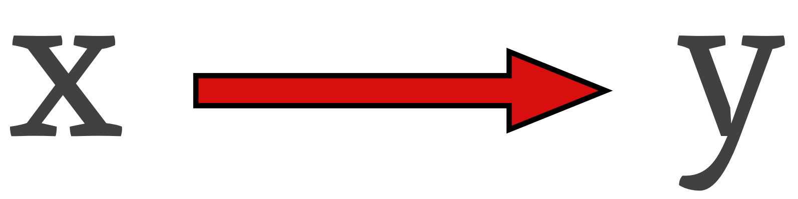
R Code

```
N = 200
z = rbinom(N, 1, 0.5)              # z ~ bernoulli(0.5)
x = rnorm(N, 1 + z)                # x ~ normal(1 + z, 1)
y = rnorm(N, 1 + 0.5*x + 2*z) # y ~ normal(1 + 0.5x + 2z, 1)
```

Modelo estatístico sem a variável de confusão Z

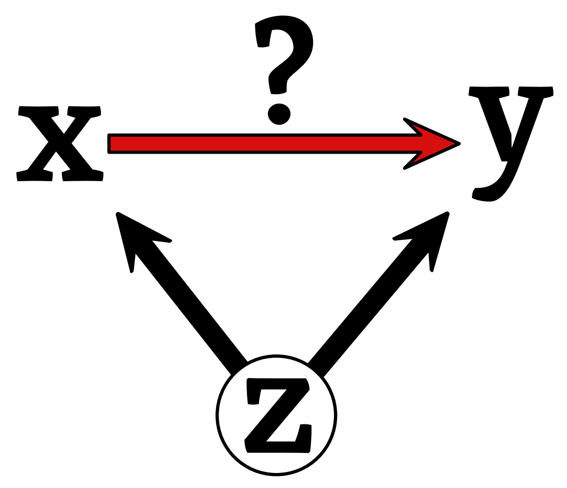


Code

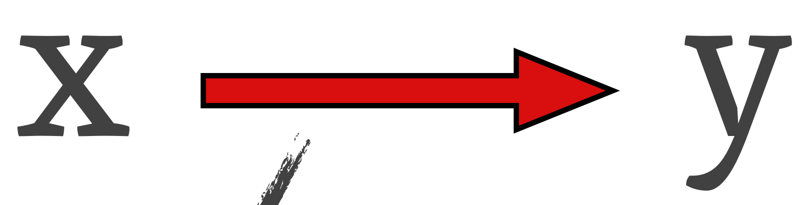



```r
N = 200
z = rbinom(N, 1, 0.5)
x = rnorm(N, 1 + z)
y = rnorm(N, 1 + 0.5*x + 2*z)
m1 = lm(y ~ x)
```

Modelo estatístico sem a variável de confusão Z



Code




```
N = 200
z = rbinom(N, 1, 0.5)
x = rnorm(N, 1 + z)
y = rnorm(N, 1 + 0.5*x + 2*z)
m1 = lm(y ~ x)
```

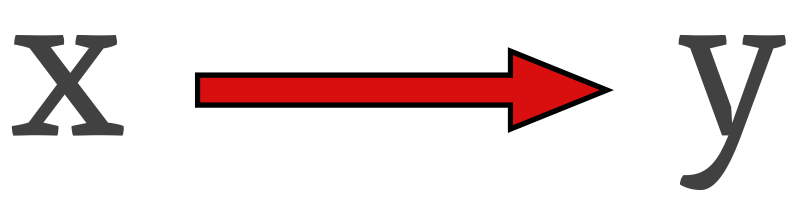
Estimativas sem a Variável de confusão

```
> (pm1 = precis(m1))
            mean   sd 5.5% 94.5%
(Intercept) 1.18 0.16 0.93  1.43
x           0.94 0.08 0.82  1.06
```

$$x \longrightarrow y$$

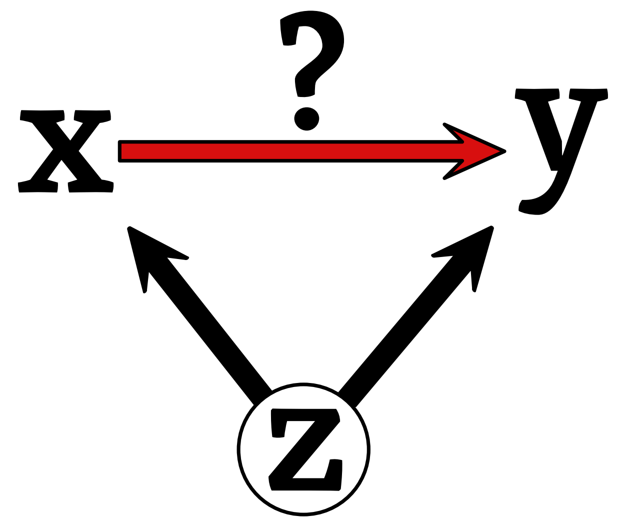
Estimativas sem a Variável de confusão

```
> (pm1 = precis(m1))
             mean   sd 5.5% 94.5%
(Intercept) 1.18 0.16 0.93  1.43
x           0.94 0.08 0.82  1.06
```



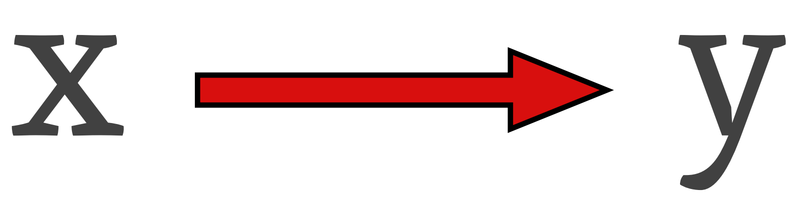
Simulation R code

```
N = 200
z = rbinom(N, 1, 0.5)
x = rnorm(N, 1 + z)
y = rnorm(N, 1 + 0.5*x + 2*z)
```



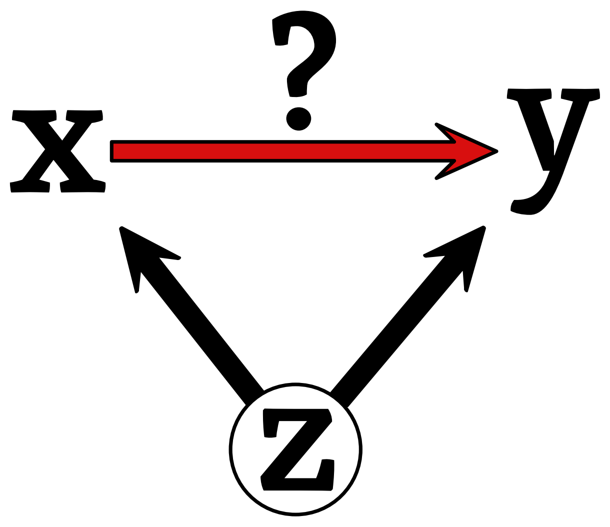
Estimativas sem a Variável de confusão

```
> (pm1 = precis(m1))
             mean   sd 5.5% 94.5%
(Intercept) 1.18 0.16 0.93  1.43
x           0.94 0.08 0.82  1.06
```



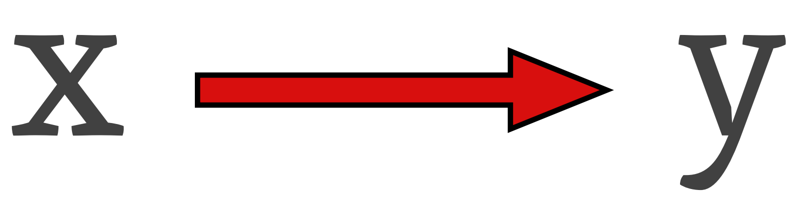
Simulation R code

```
N = 200
z = rbinom(N, 1, 0.5)
x = rnorm(N, 1 + z)
y = rnorm(N, 1 + 0.5*x + 2*z)
```



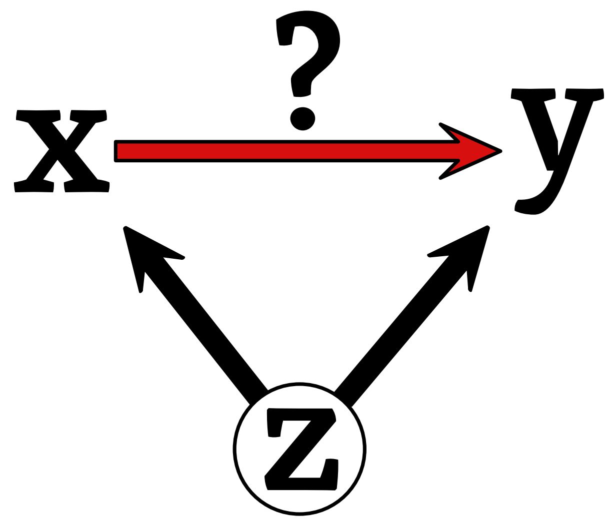
Estimativas sem a Variável de confusão

```
> (pm1 = precis(m1))
            mean   sd 5.5% 94.5%
(Intercept) 1.18 0.16 0.93  1.43
x           0.94 0.08 0.82  1.06
```



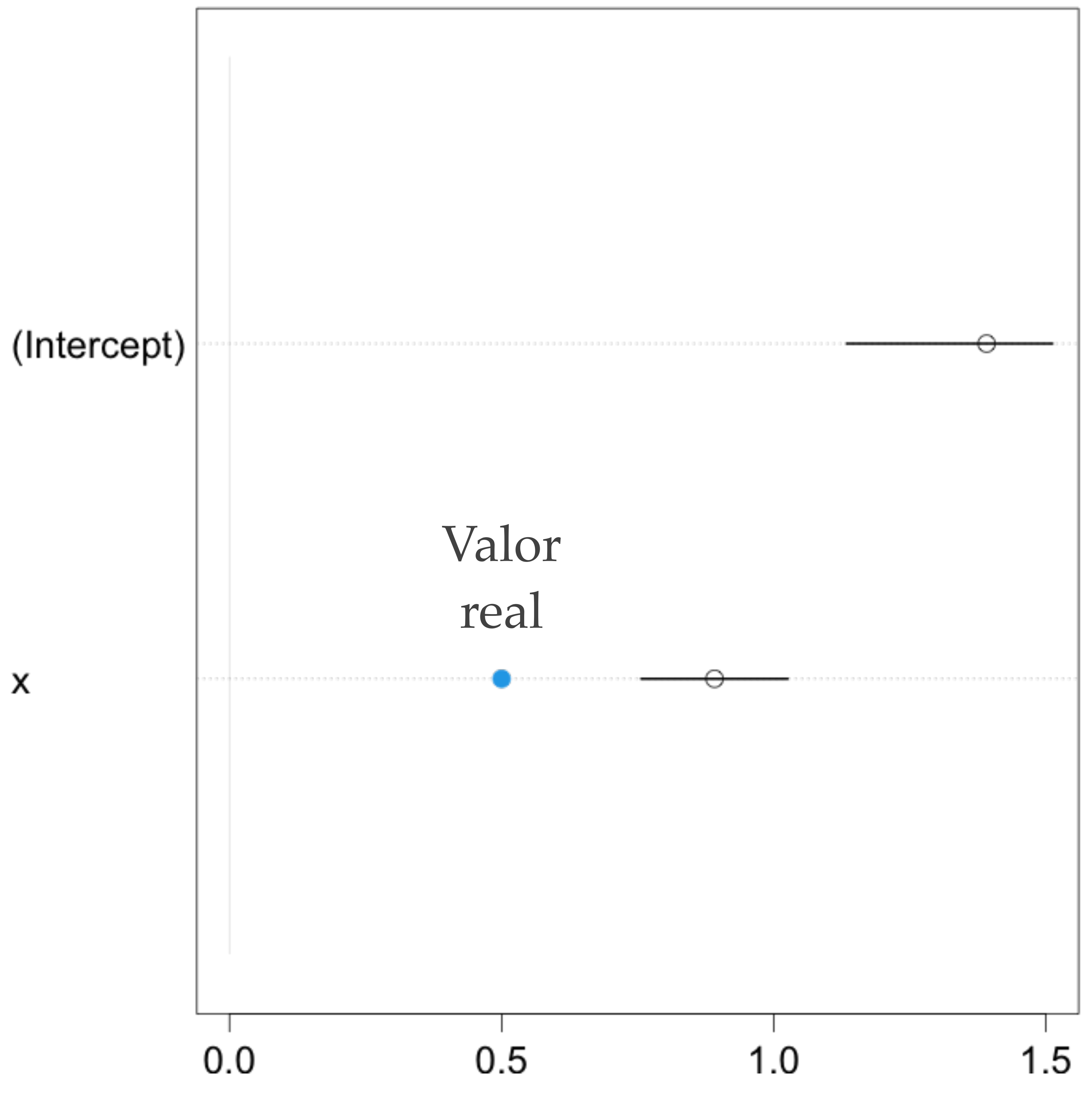
Simulation R code

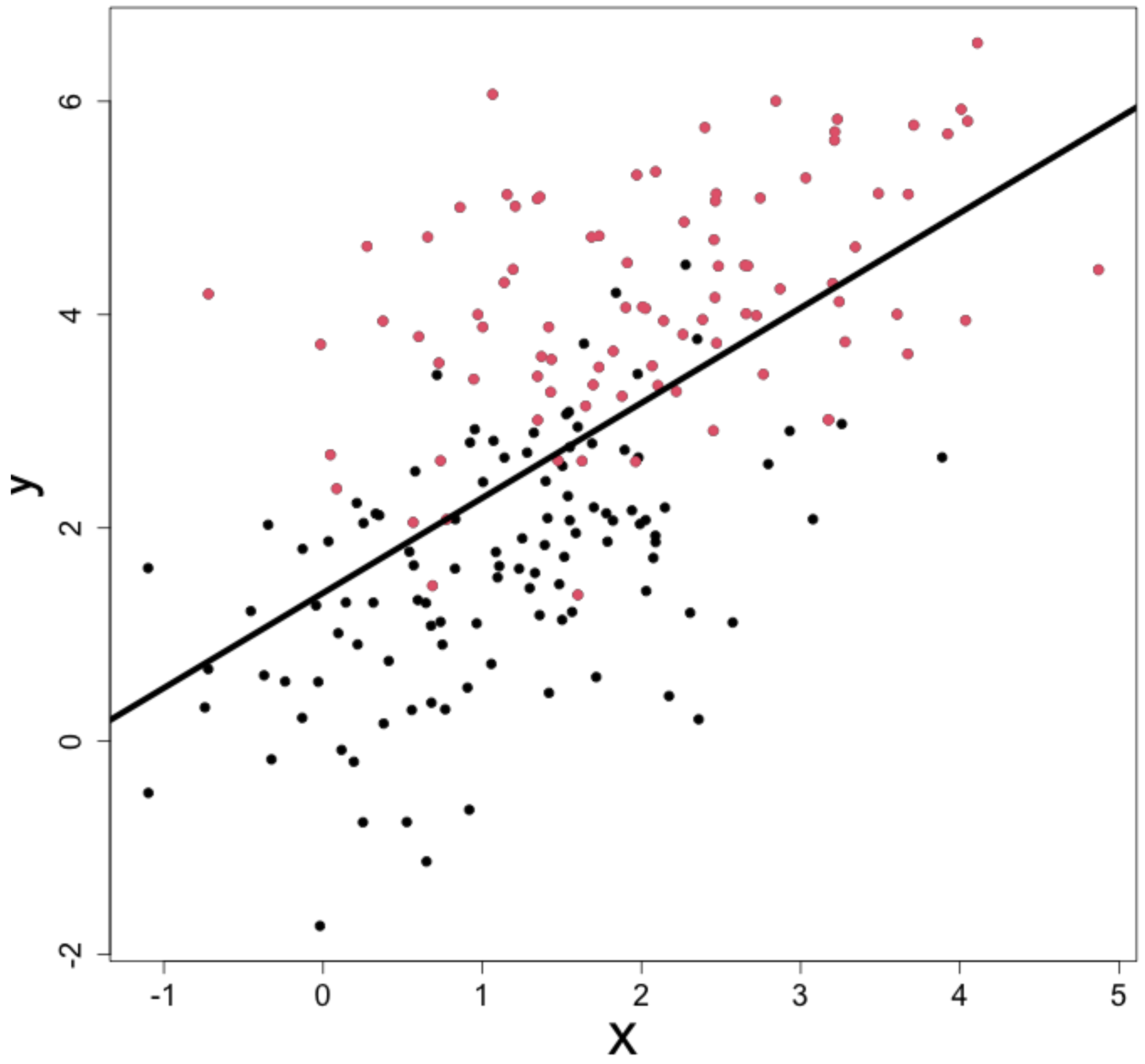
```
N = 200
z = rbinom(N, 1, 0.5)
x = rnorm(N, 1 + z)
y = rnorm(N, 1 + 0.5*x + 2*z)
```



Estimativa do efeito de x em y sem a variável de confusão z

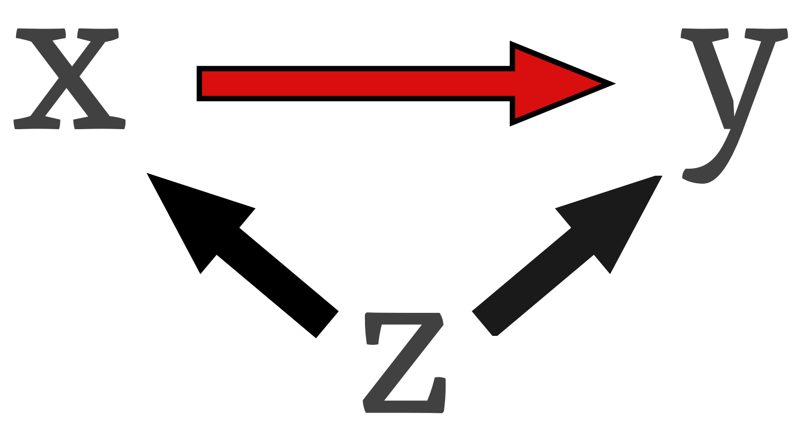
$x \longrightarrow y$





Incluindo a causa compartilhada no modelo

Usando uma regressão múltipla para estratificar por z

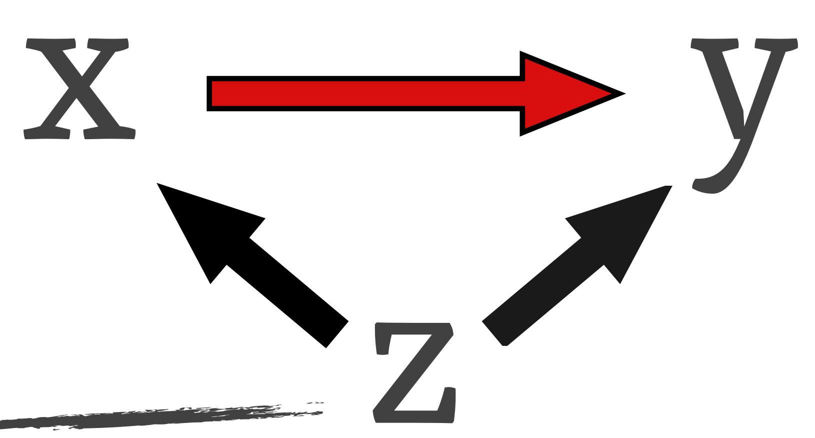



```
> m2 = lm(y ~ x + z)
> (pm2 = precis(m2))
            mean   sd 5.5% 94.5%
(Intercept) 0.92 0.12 0.73  1.12
x           0.48 0.07 0.37  0.60
z           2.03 0.17 1.75  2.31
```

Incluindo a causa compartilhada no modelo

Usando uma regressão múltipla para estratificar por z

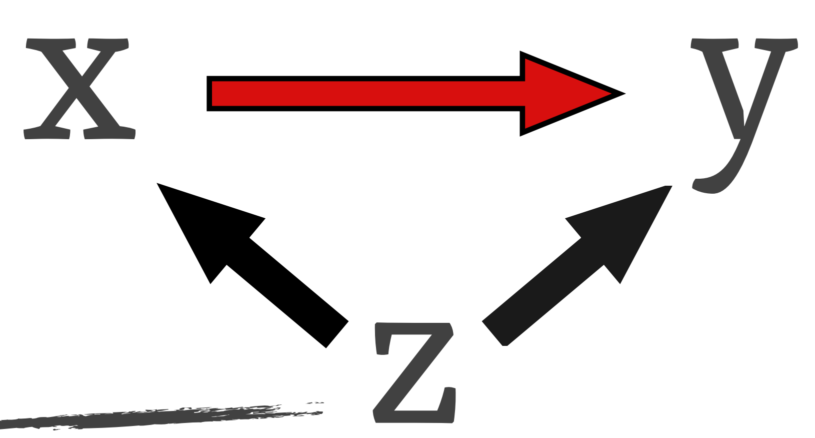
```
> m2 = lm(y ~ x + z)
> (pm2 = precis(m2))
            mean   sd 5.5% 94.5%
(Intercept) 0.92 0.12 0.73  1.12
x           0.48 0.07 0.37  0.60
z           2.03 0.17 1.75  2.31
```



Incluindo a causa compartilhada no modelo

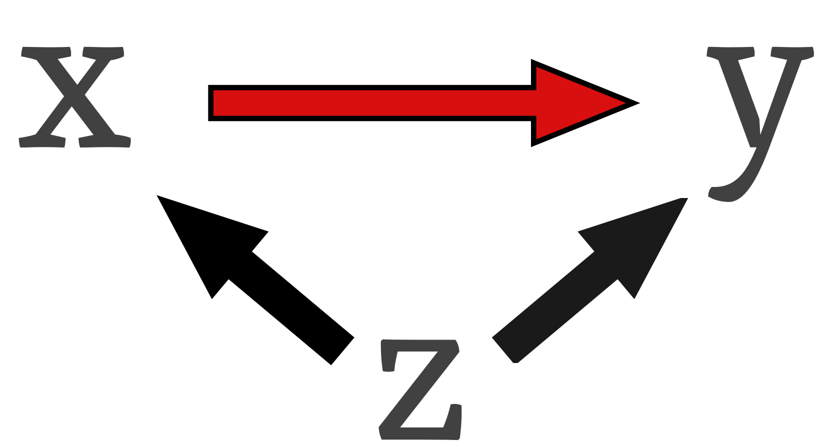
Usando uma regressão múltipla para estratificar por z

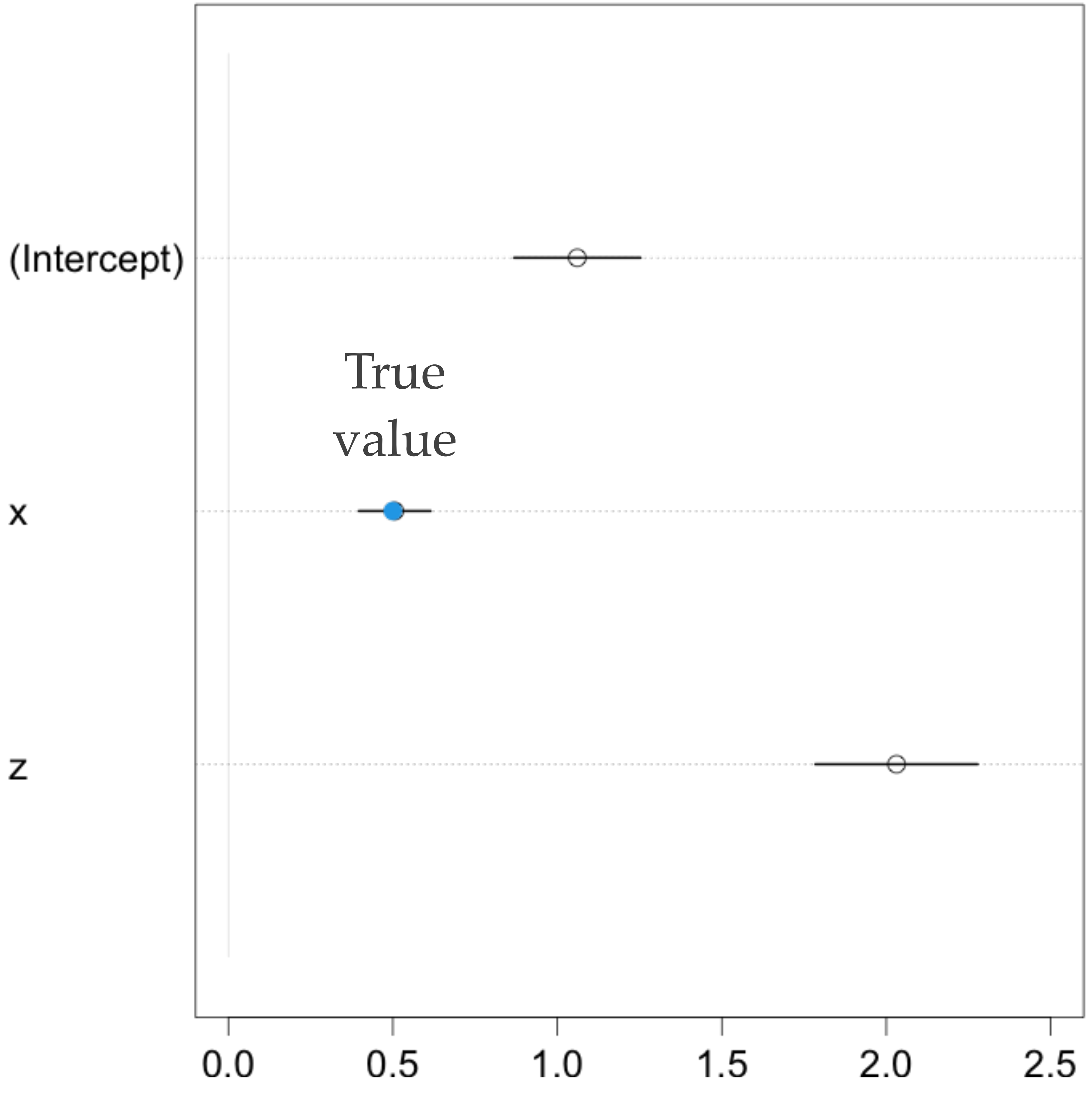
```
> m2 = lm(y ~ x + z)
> (pm2 = precis(m2))
            mean   sd 5.5% 94.5%
(Intercept) 0.92 0.12 0.73  1.12
x           0.48 0.07 0.37  0.60
z           2.03 0.17 1.75  2.31
```

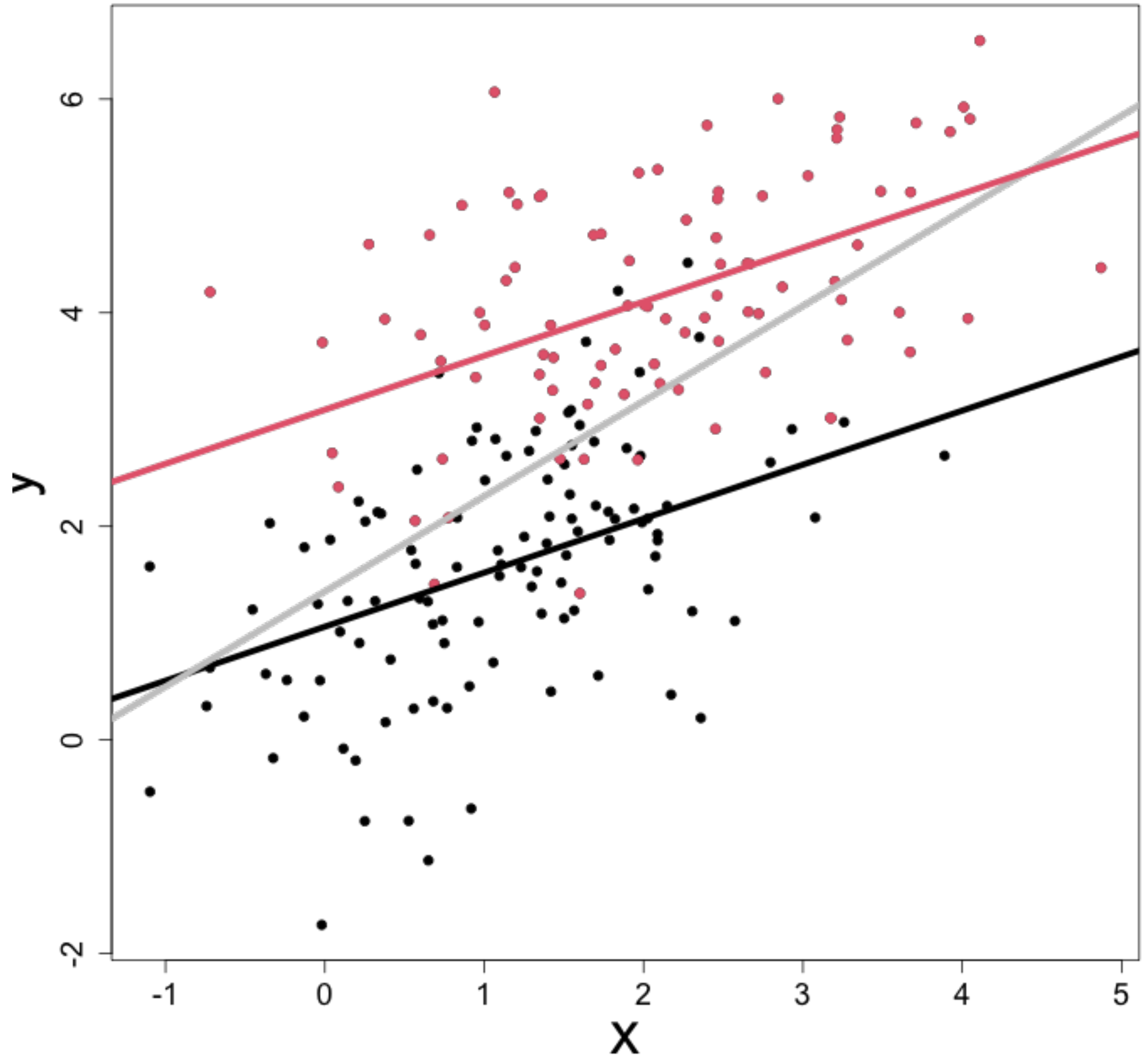


Incluindo a variável de confusão z.

Usando uma regressão múltipla para estratificar por z

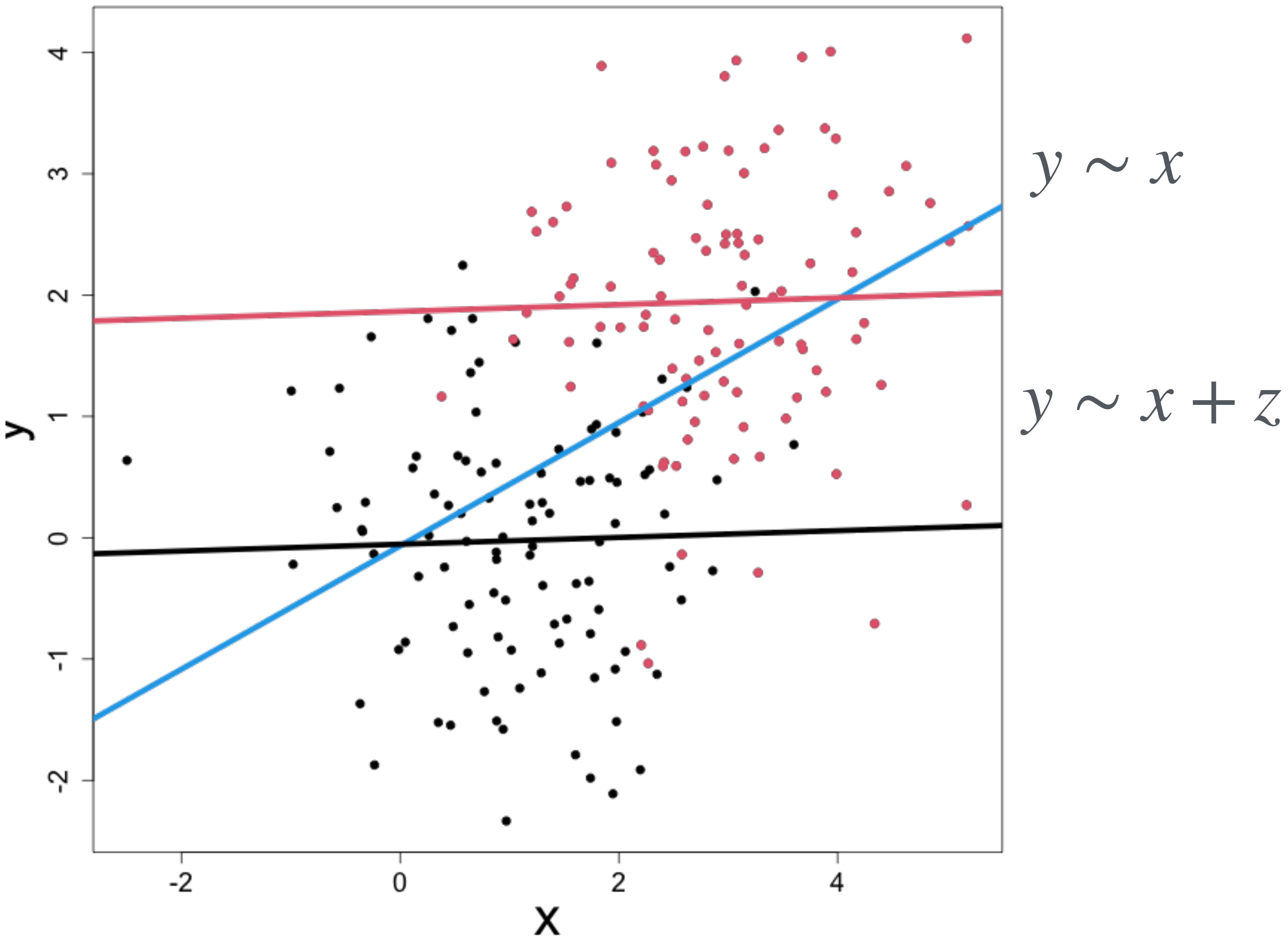




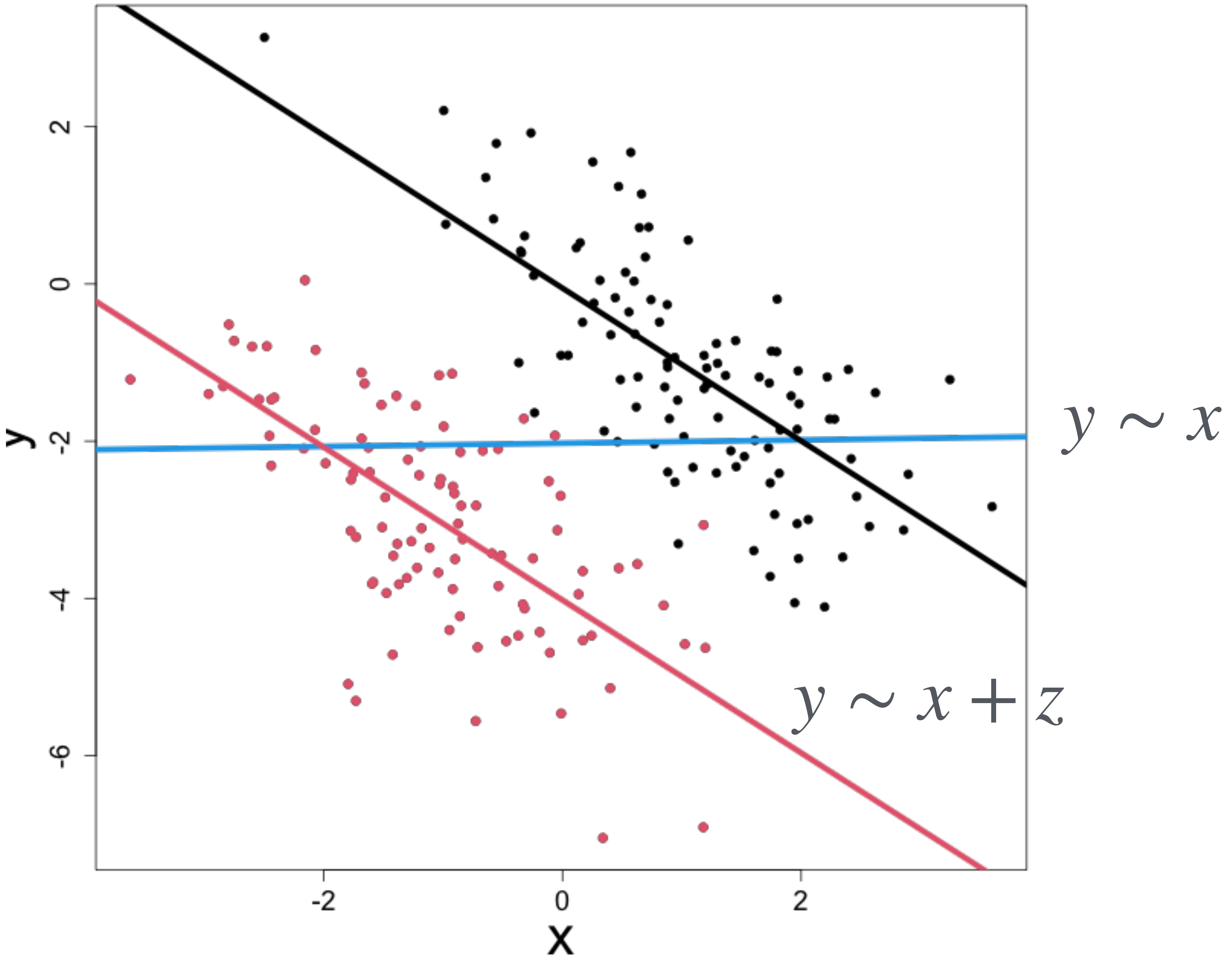


Mudança do efeito com a inclusão da variável de confusão

Efeito de x em y é criado por z

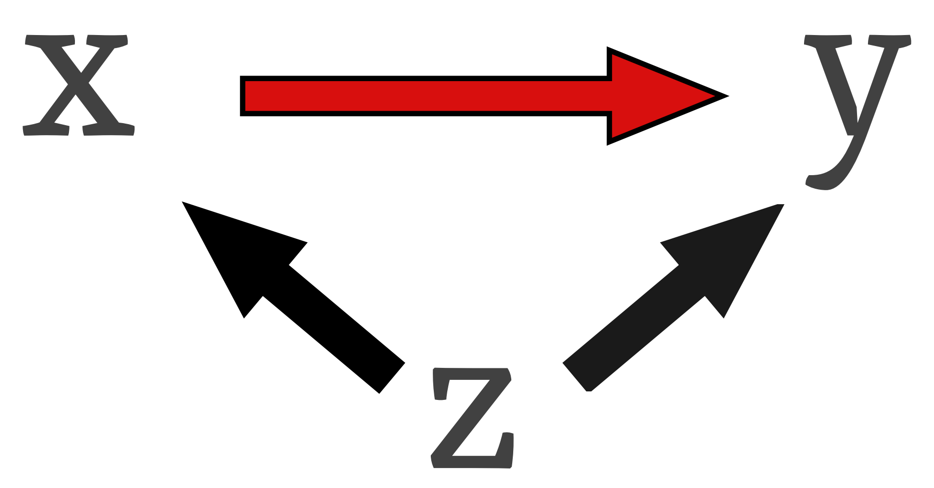


Efeito de x em y é oculto por z



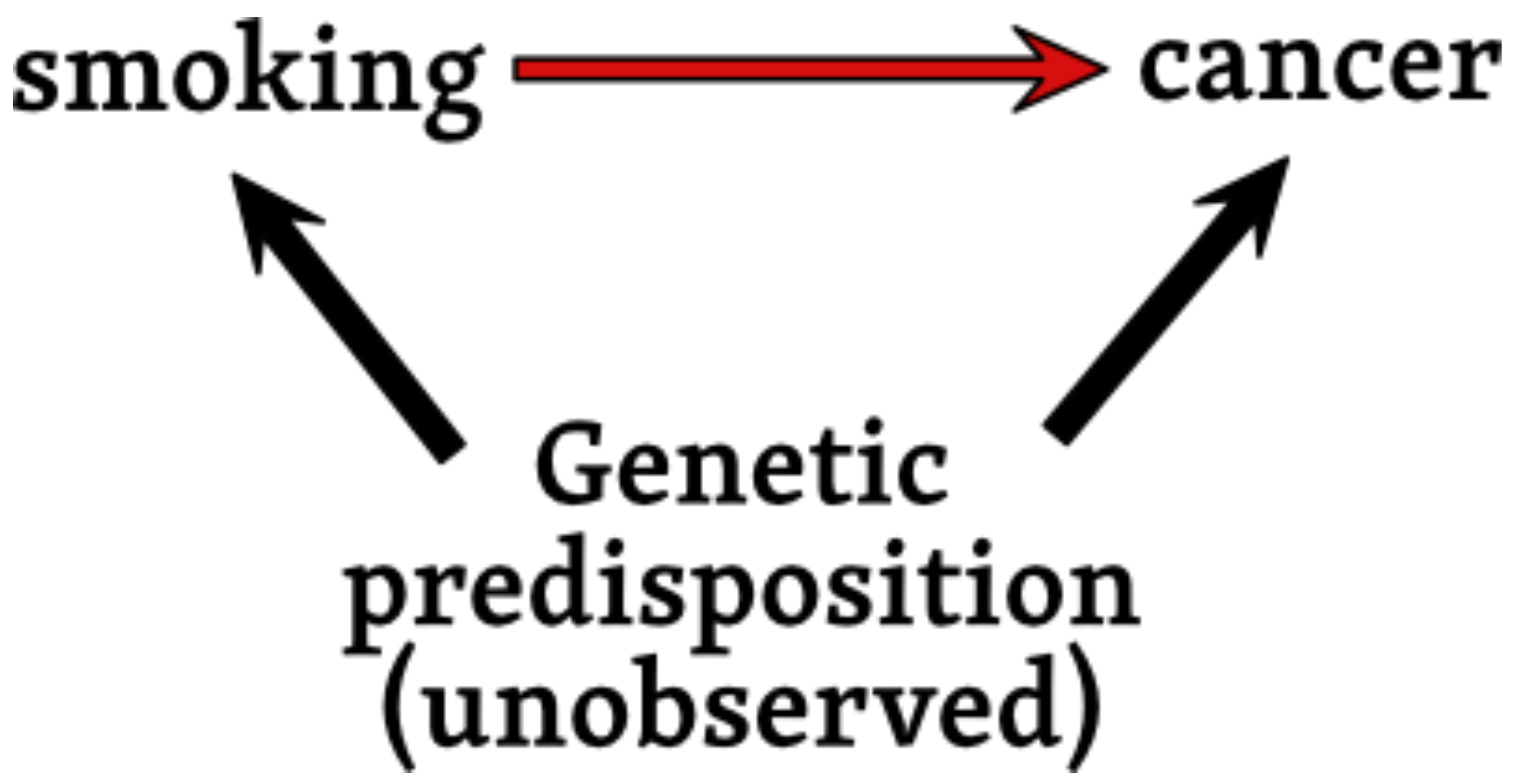
Examples of forks or confounders

- A forquilha é a variável “controle" clássica
- A maioria das variáveis é incluída nos modelo sob a suposição de que são variáveis de confusão e precisam ter seus efeitos levados em consideração
- Condicionando em z, o efeito causal de x em y pode ser estimado de forma não-enviezada
- A variável z cria um **caminho não-causal** entre x e y. Adicionar z ao modelo fecha esse caminho.



Tabaco e Cancer

R. A. Fisher não estava convencido de que fumar causava câncer, e propôs que uma variável de propensão não observada causava tanto o câncer quanto o tabagismo.

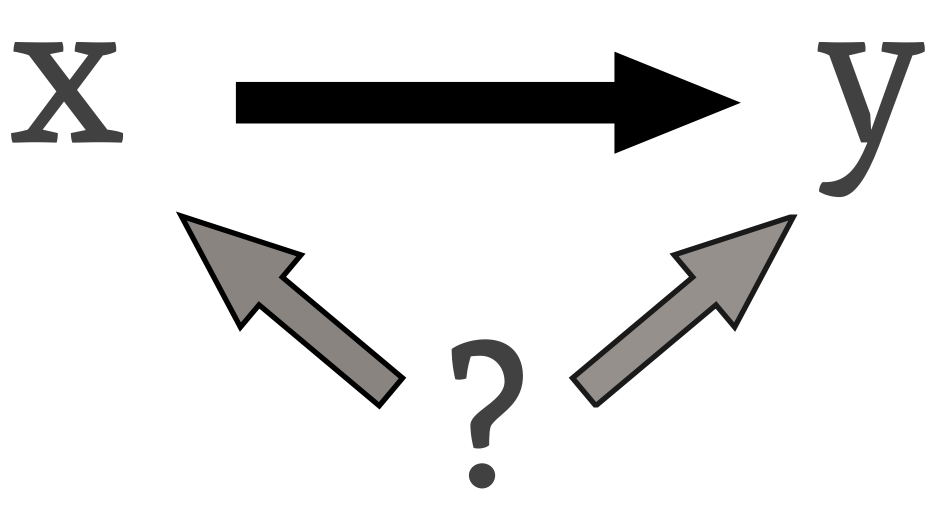


Quantificando Viés por variáveis omitidas

- Mesmo sem conhecer as variáveis omitidas, podemos estimar quão intenso seu efeito precisaria ser para mascarar/causar o efeito observado

- Podemos comparar esse efeito não conhecido com efeitos conhecidos

- **Exemplo**: uma variável omitida precisaria ter um efeito 2x maior que fumar para causar o efeito observado



Making sense of sensitivity: extending omitted variable bias

Carlos Cinelli and Chad Hazlett

University of California, Los Angeles, USA

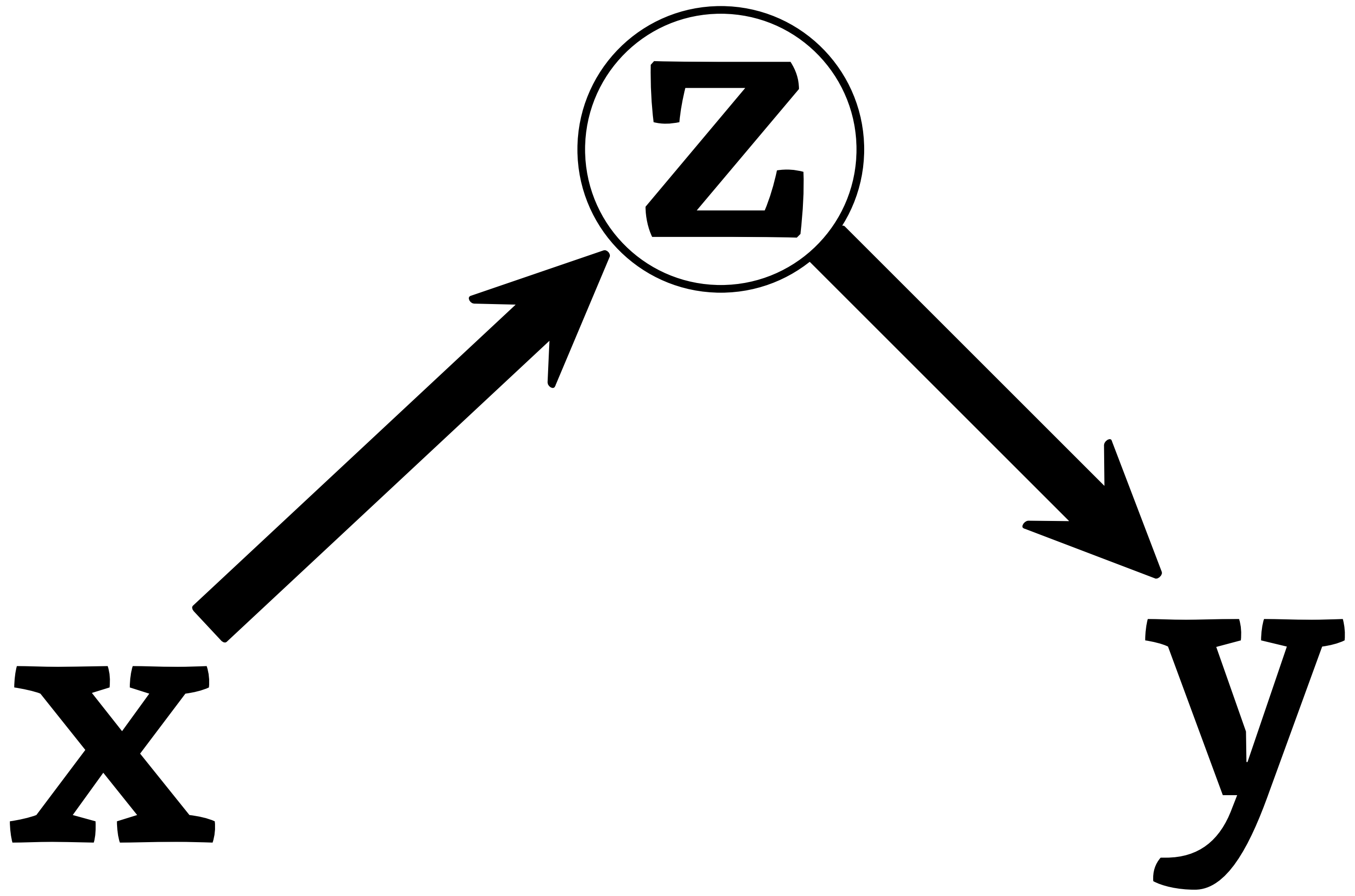
[Received August 2018. Final revision October 2019]

Summary. We extend the omitted variable bias framework with a suite of tools for sensitivity analysis in regression models that does not require assumptions on the functional form of the treatment assignment mechanism nor on the distribution of the unobserved confounders, naturally handles multiple confounders, possibly acting non-linearly, exploits expert knowledge to bound sensitivity parameters and can be easily computed by using only standard regression results. In particular, we introduce two novel sensitivity measures suited for routine reporting. The robustness value describes the minimum strength of association that unobserved confounding would need to have, both with the treatment and with the outcome, to change the research conclusions. The partial R^2 of the treatment with the outcome shows how strongly confounders explaining all the residual outcome variation would have to be associated with the treatment to eliminate the estimated effect. Next, we offer graphical tools for elaborating on problematic confounders, examining the sensitivity of point estimates and t-values, as well as 'extreme scenarios'. Finally, we describe problems with a common 'benchmarking' practice and introduce a novel procedure to bound the strength of confounders formally on the basis of a comparison with observed covariates. We apply these methods to a running example that estimates the effect of exposure to violence on attitudes toward peace.

Cinelli, C. & Hazlett, C. Making sense of sensitivity: extending omitted variable bias. J. R. Stat. Soc. Series B Stat. Methodol. ***82****, 39–67 (2020)*

The pipe

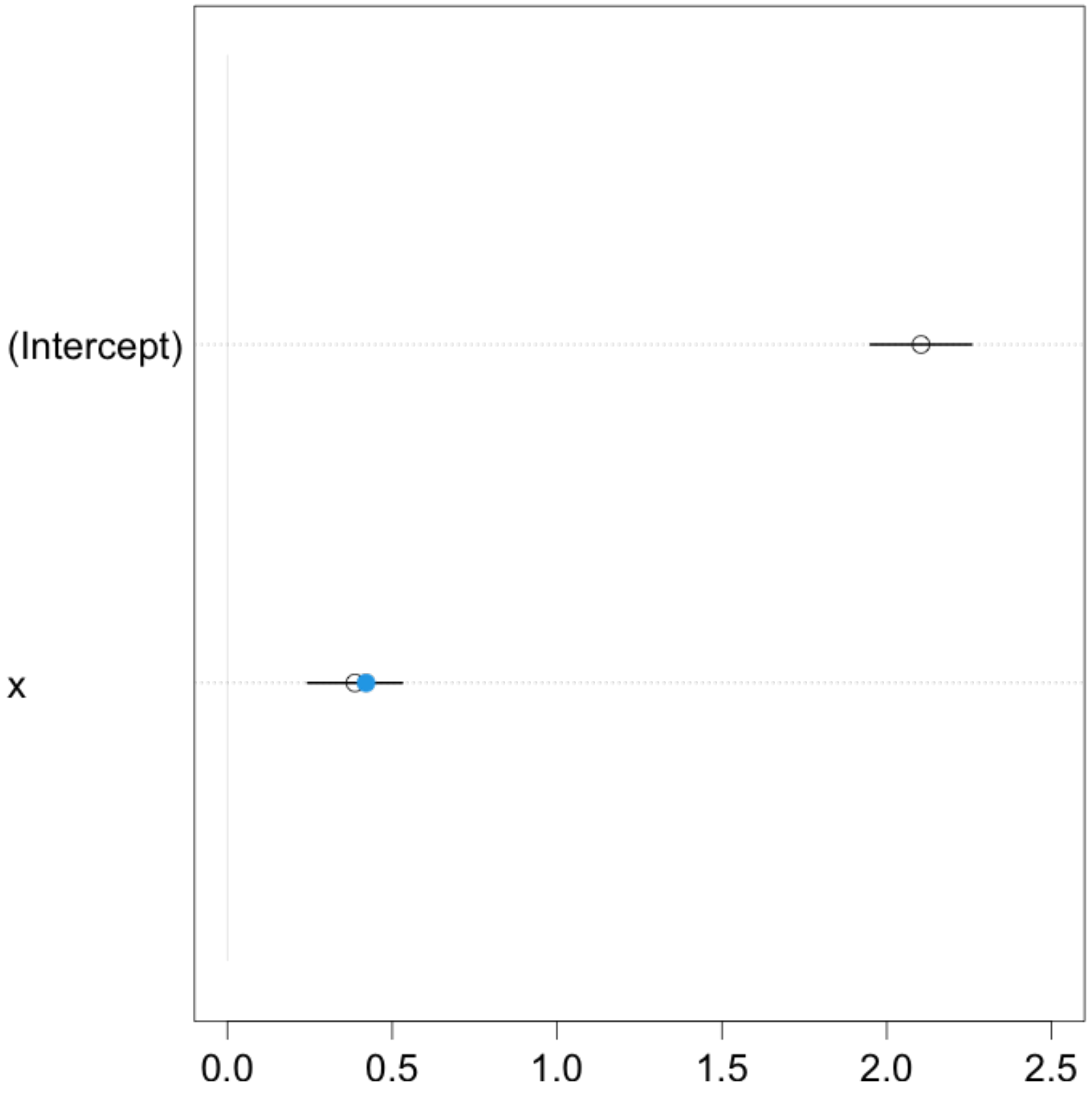
Todo o efeito de x é mediado por z

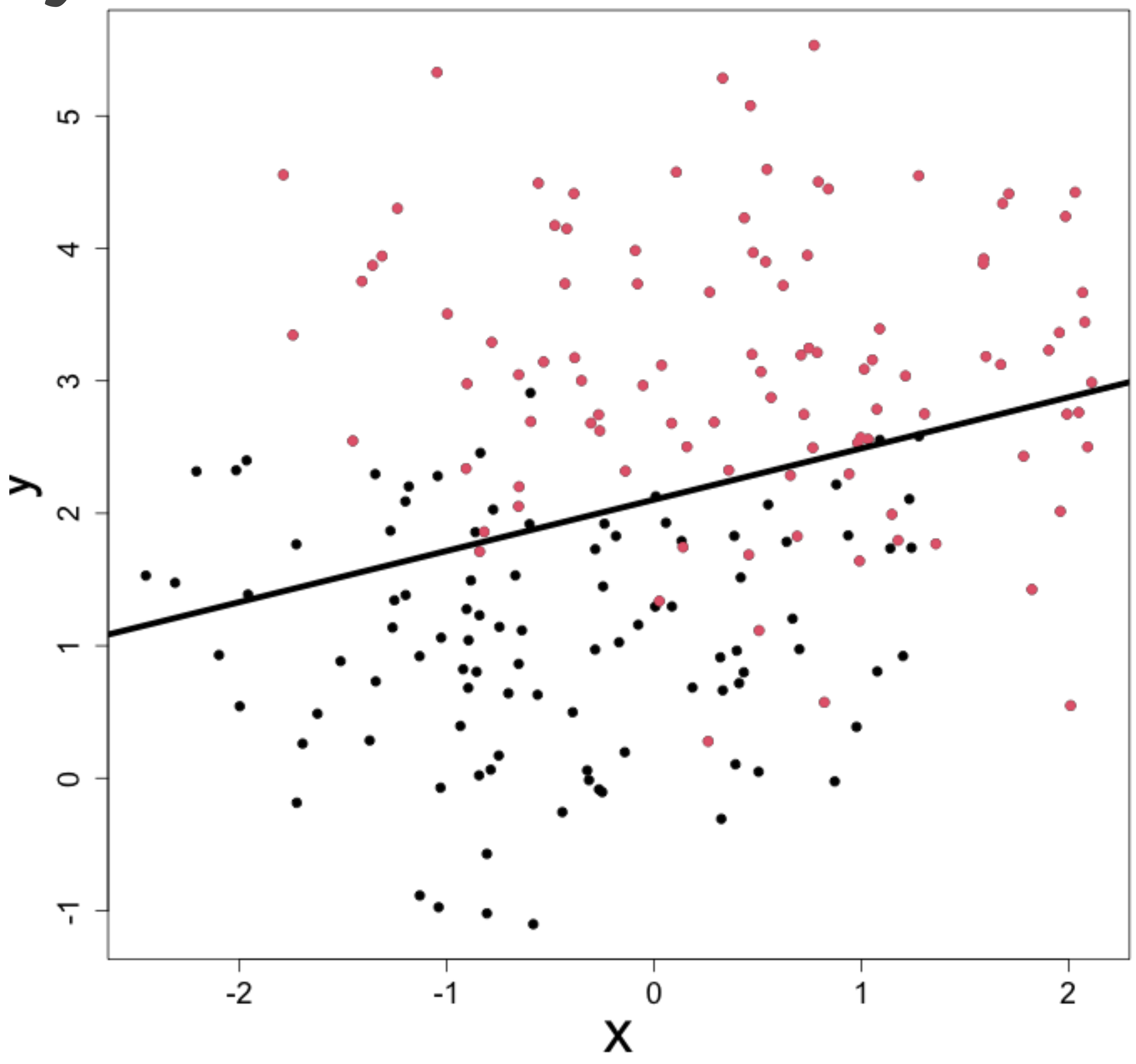



```
N = 200
x = rnorm(N)                          # x ~ normal(0, 1)
z = rbinom(N, 1, inv_logit(x)) # z ~ bernoulli(invlogit(x))
y = rnorm(N,  1 + 2*z)               # y ~ normal(1 + 2z, 1)
```

Modelo sem o mediador

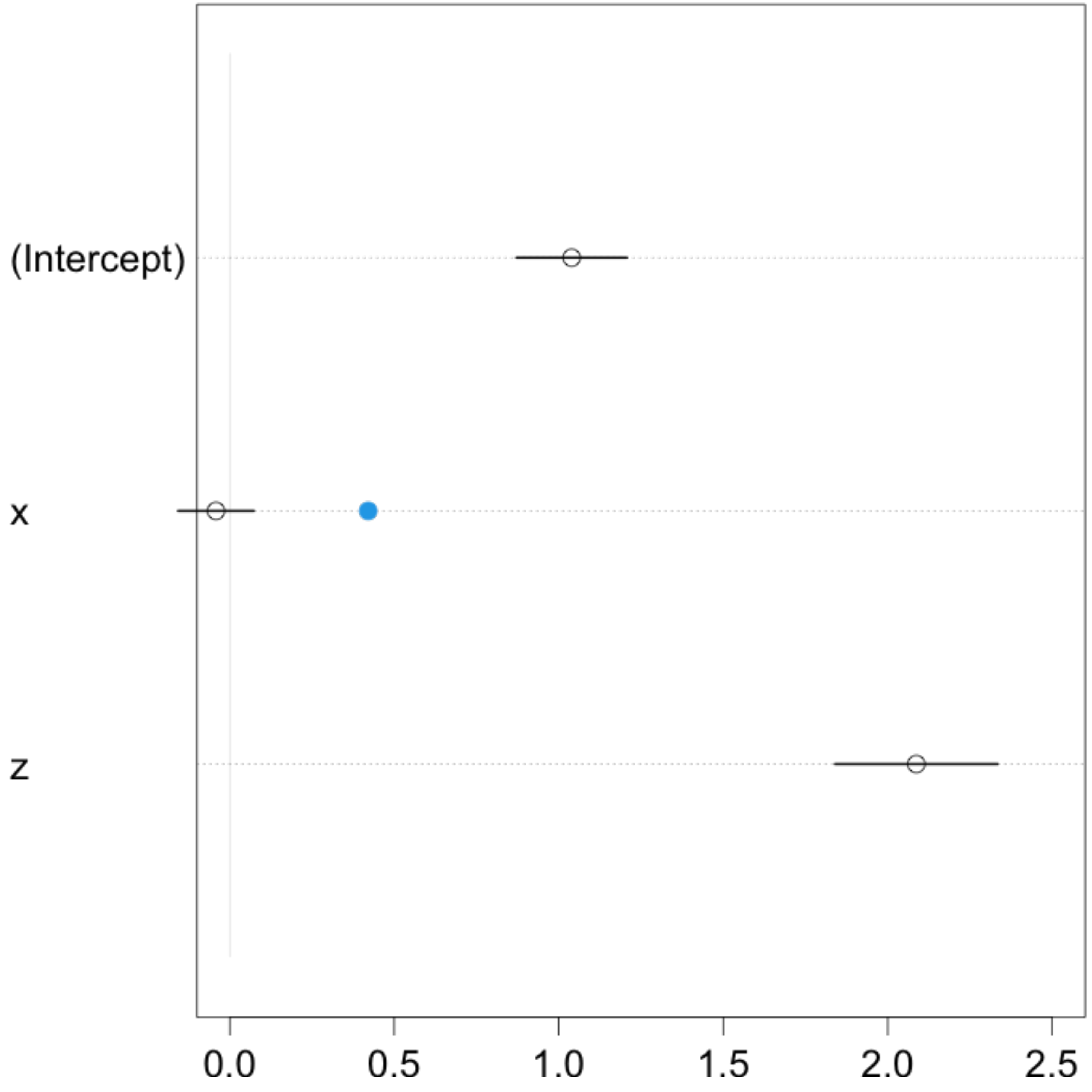
$x \rightarrow y$



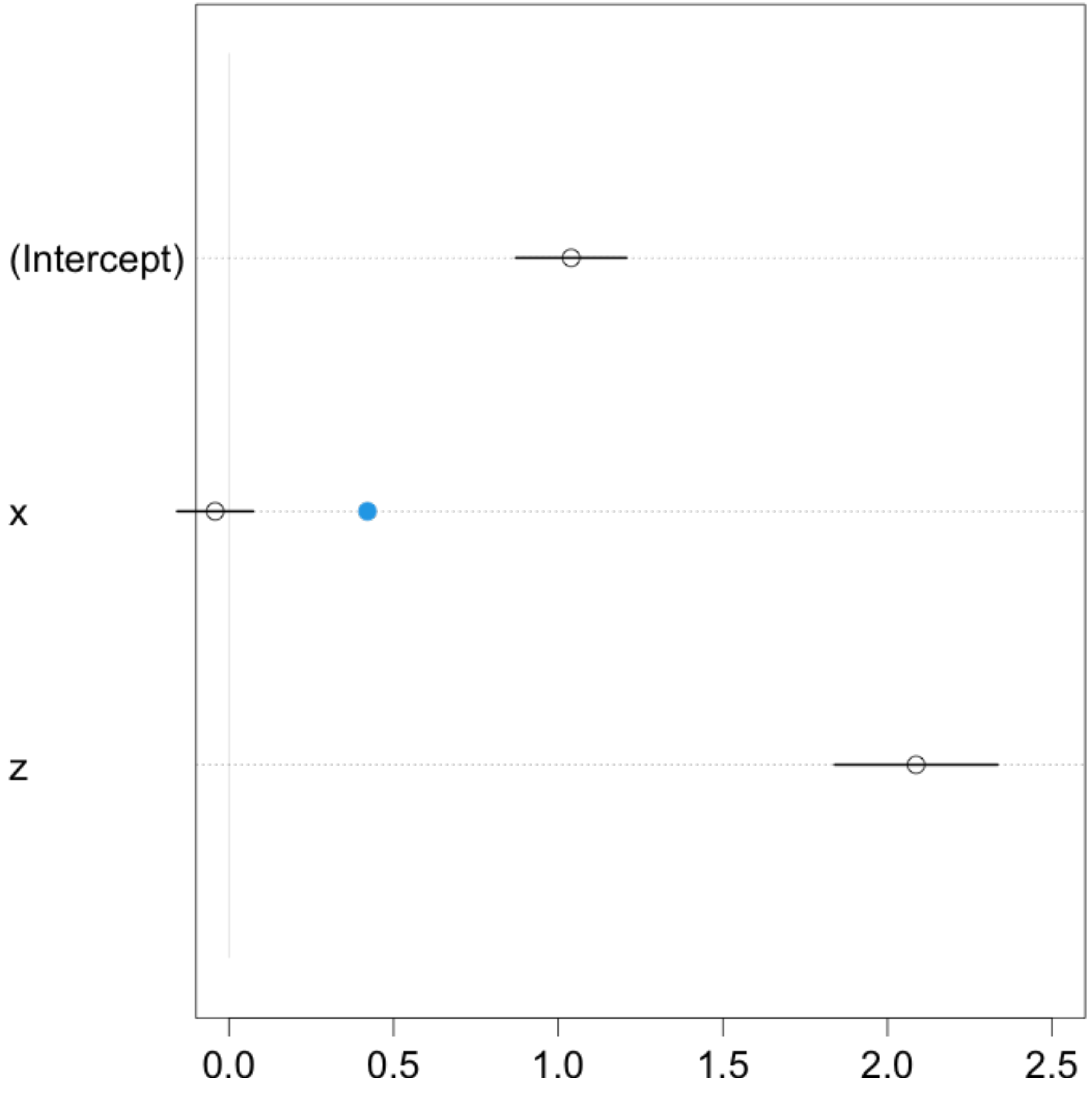


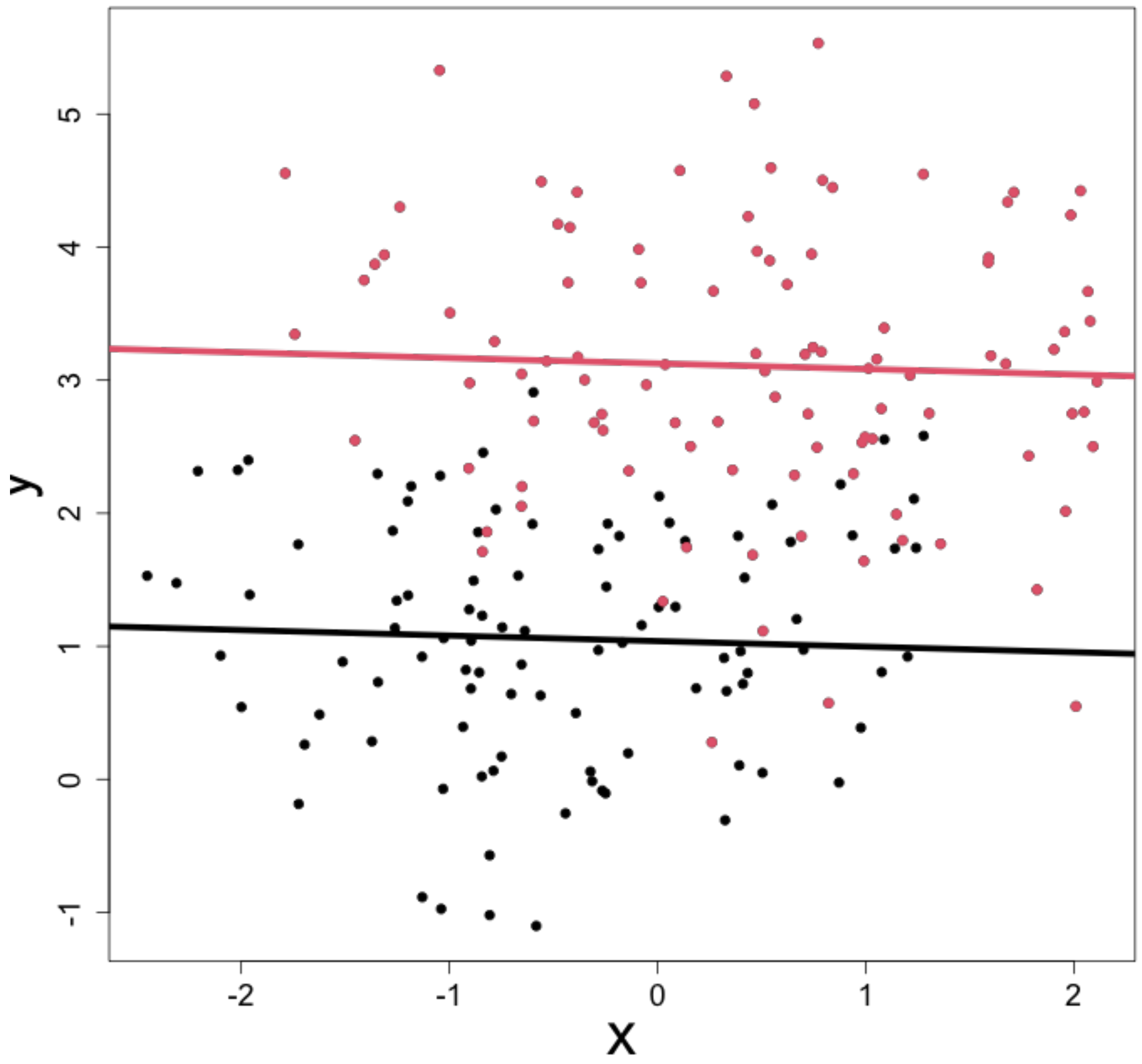
Incluindo o mediador

```
> m2 = lm(y ~ x + z)
> (pm2 = precis(m2))
             mean   sd  5.5% 94.5%
(Intercept)  1.04 0.11  0.87  1.21
x           -0.04 0.07 -0.16  0.07
z            2.09 0.15  1.84  2.33
```



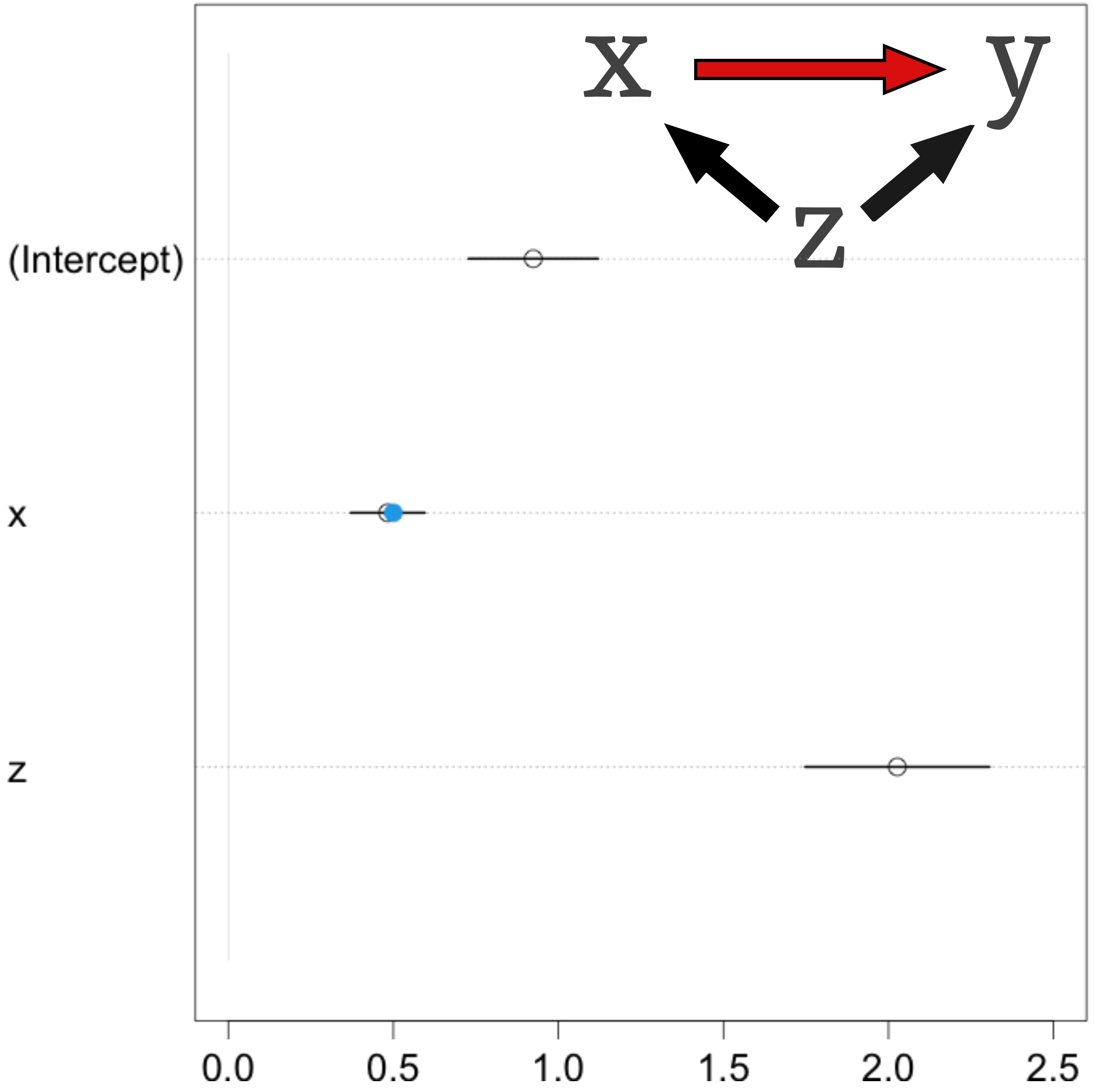
Incluindo o mediador



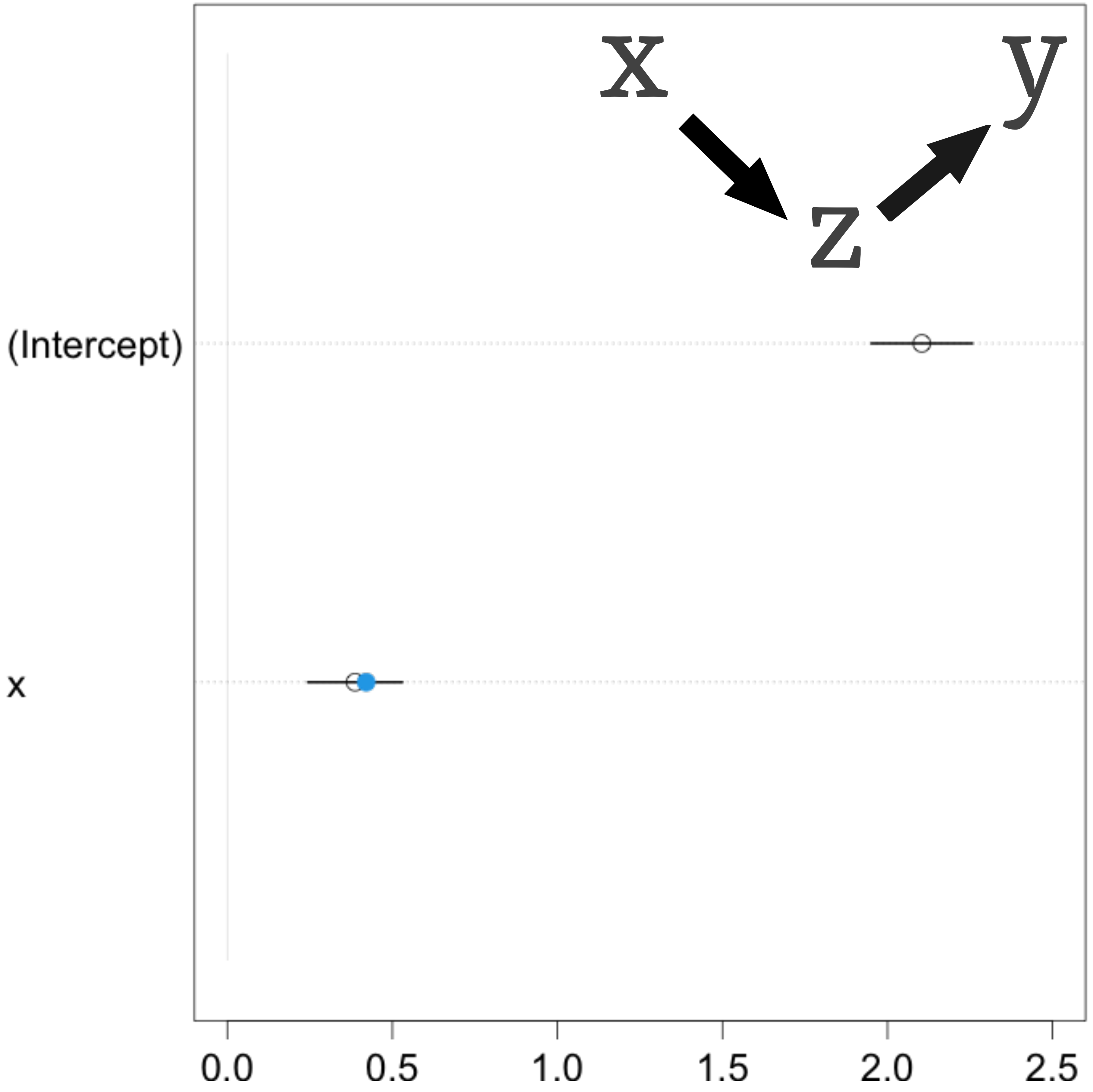


Fork vs. Pipe

Estimador correto **com** o Z



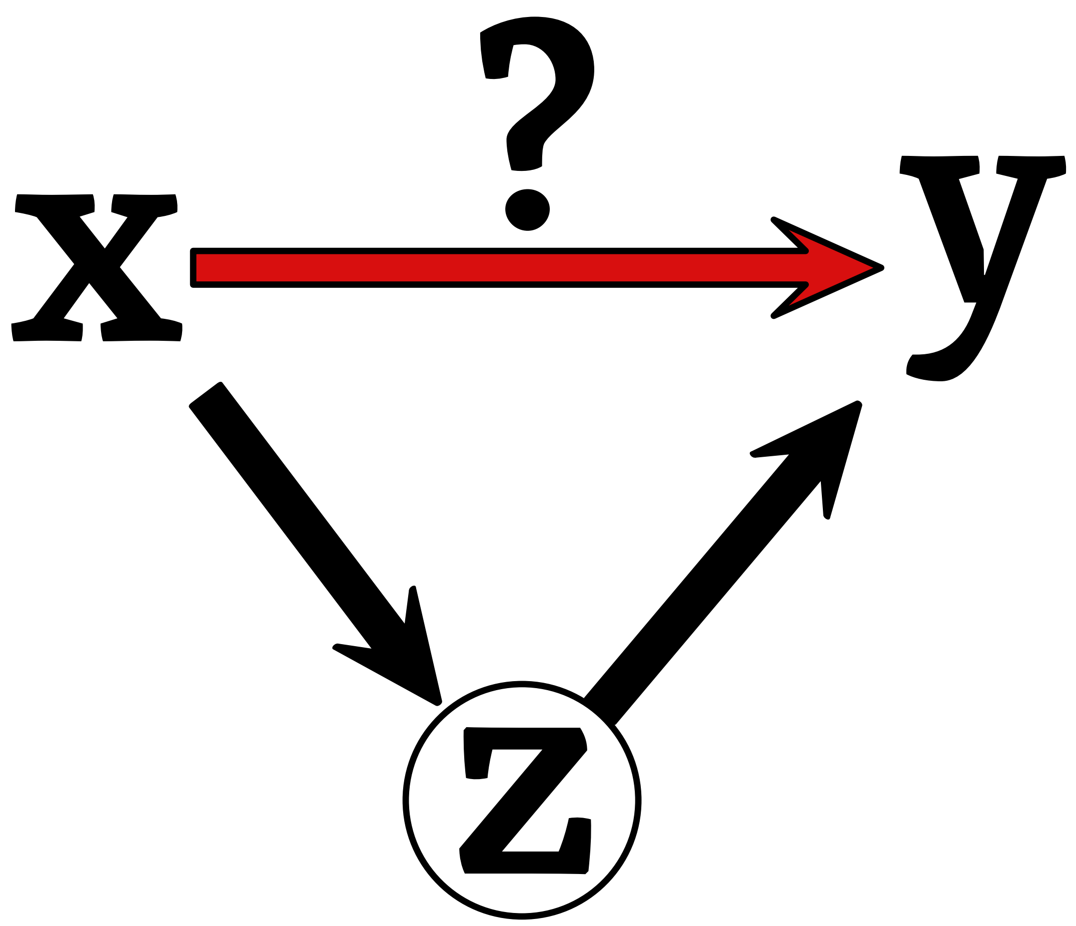
Estimador correto **sem** o Z



Examples of pipes

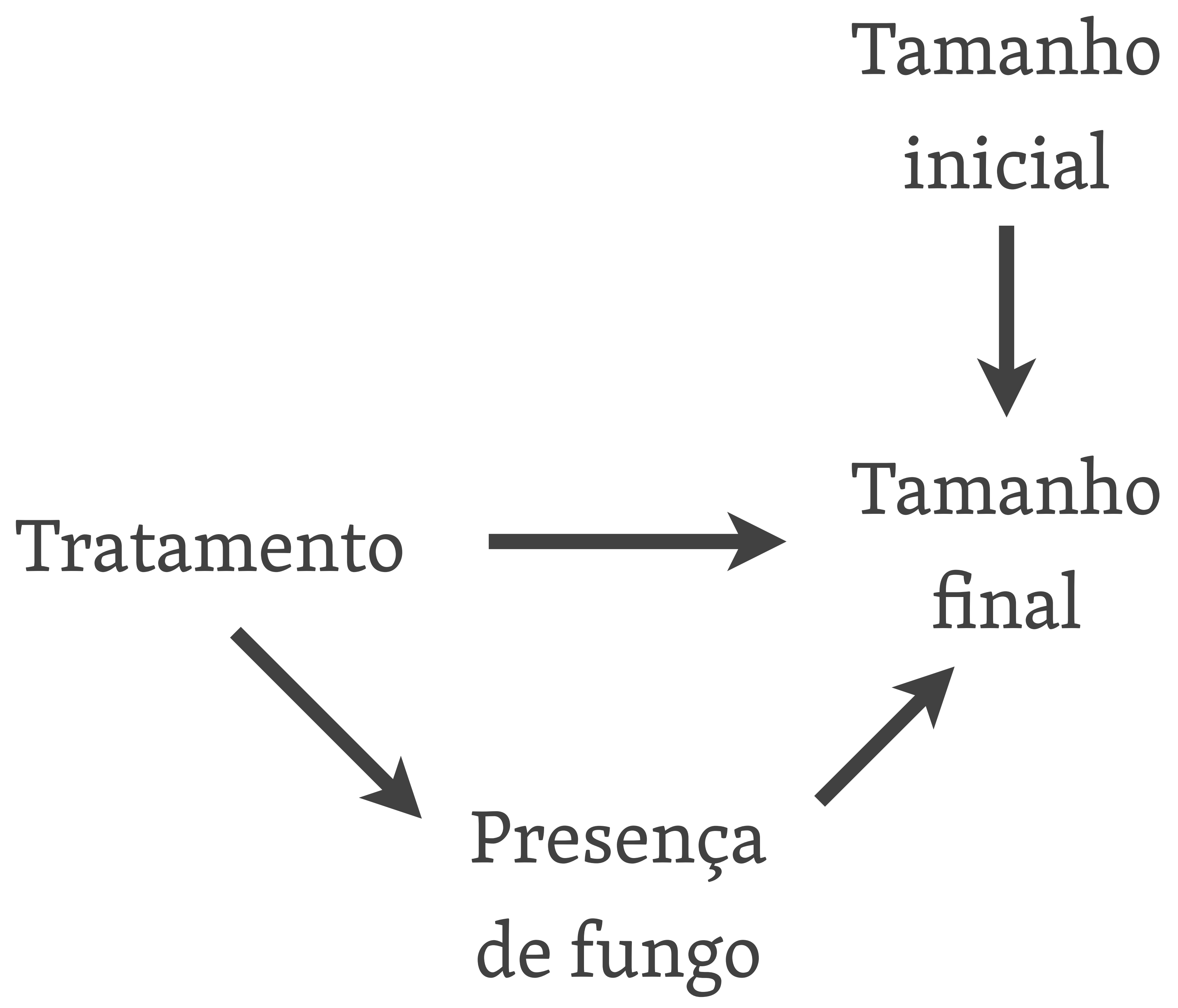
Mediators or descendants

- Nesse caso, a variável z forma um **caminho causal** entre x e y. Adicionar z ao modelo fecha esse caminho.
- Incluir um mediador em nossos modelos pode ter efeitos catastróficos.
- Um erro comum é incluir **variáveis pós-tratamento** no modelo.



Variáveis pós tratamento

Por exemplo, num experimento que procura entender o efeito de um tratamento antifúngico no crescimento de uma planta



Post-treatment variables

AMERICAN JOURNAL of POLITICAL SCIENCE

How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It

Jacob M. Montgomery Washington University in St. Louis
Brendan Nyhan Dartmouth College
Michelle Torres Washington University in St. Louis

Abstract: *In principle, experiments offer a straightforward method for social scientists to accurately estimate causal effects. However, scholars often unwittingly distort treatment effect estimates by conditioning on variables that could be affected by their experimental manipulation. Typical examples include controlling for posttreatment variables in statistical models, eliminating observations based on posttreatment criteria, or subsetting the data based on posttreatment variables. Though these modeling choices are intended to address common problems encountered when conducting experiments, they can bias estimates of causal effects. Moreover, problems associated with conditioning on posttreatment variables remain largely unrecognized in the field, which we show frequently publishes experimental studies using these practices in our discipline's most prestigious journals. We demonstrate the severity of experimental posttreatment bias analytically and document the magnitude of the potential distortions it induces using visualizations and reanalyses of real-world data. We conclude by providing applied researchers with recommendations for best practice.*

Montgomery, J. M., Nyhan, B. & Torres, M. How conditioning on posttreatment variables can ruin your experiment and what to do about it. Am. J. Pol. Sci. ***62****, 760–775 (2018)*

Post-treatment variables

Table 1. Posttreatment Conditioning in Experimental Studies

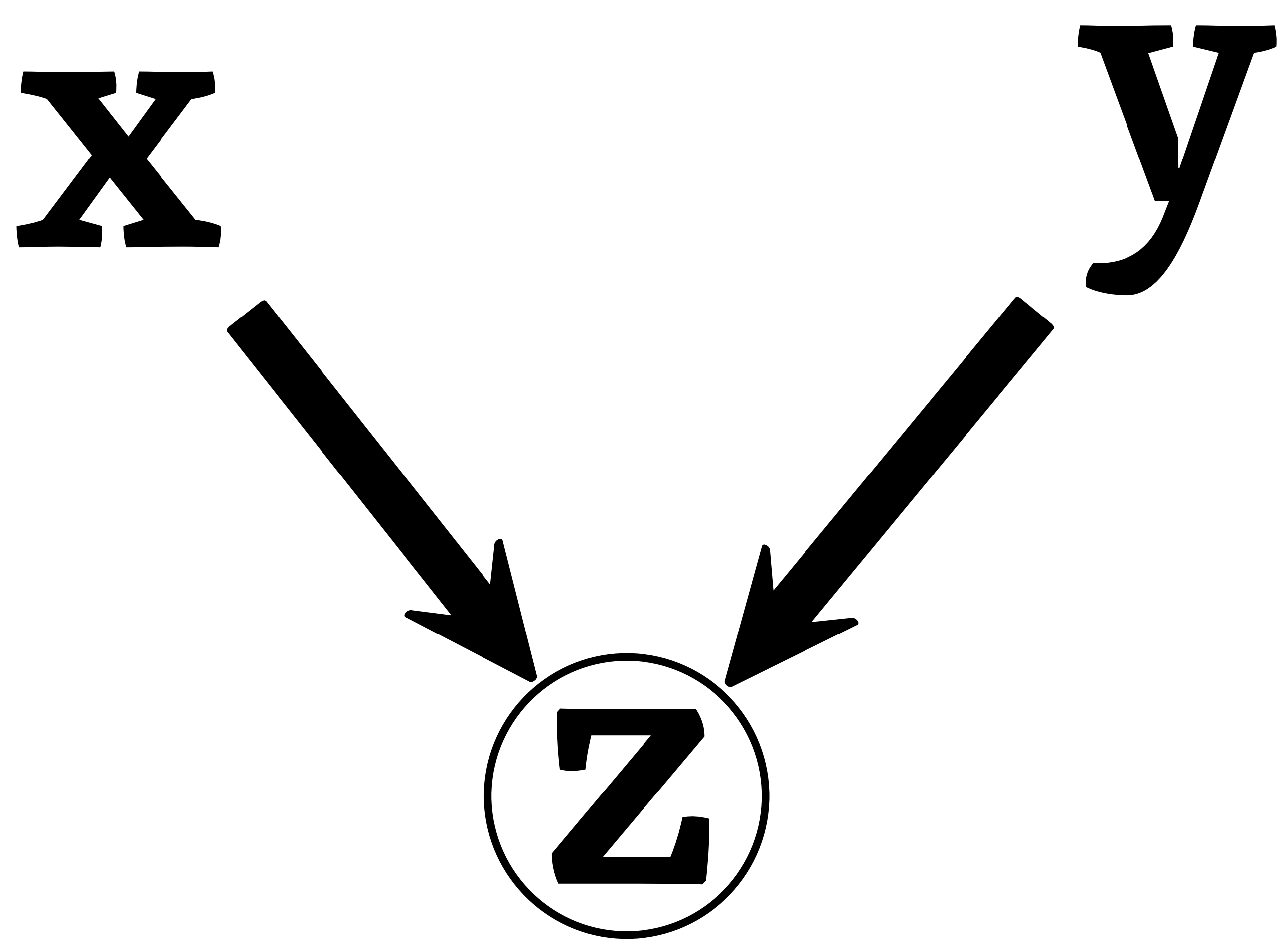
Category	Prevalence
Engages in posttreatment conditioning	46.7%
Controls for/interacts with a posttreatment variable	21.3%
Drops cases based on posttreatment criteria	14.7%
Both types of posttreatment conditioning present	10.7%
No conditioning on posttreatment variables	52.0%
Insufficient information to code	1.3%

Note: The sample consists of 2012–14 articles in the *American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics* including a survey, field, laboratory, or lab-in-the-field experiment (n = 75).

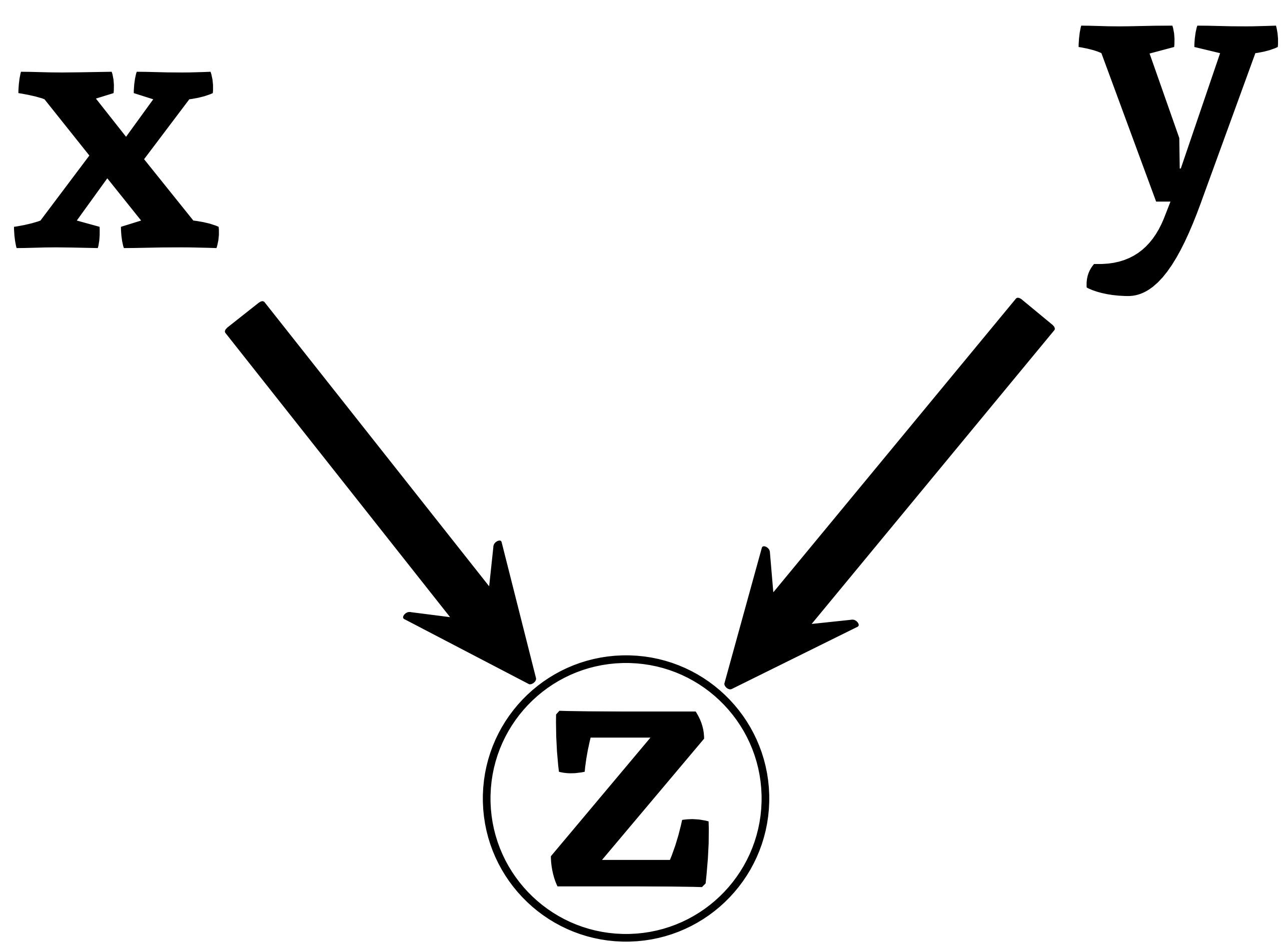
Montgomery, J. M., Nyhan, B. & Torres, M. How conditioning on posttreatment variables can ruin your experiment and what to do about it. Am. J. Pol. Sci. ***62****, 760–775 (2018)*

Collider

Nenhum efeito de x em y, mas ambos afetam Z



Nenhum efeito de x em y, mas ambos afetam Z



Math

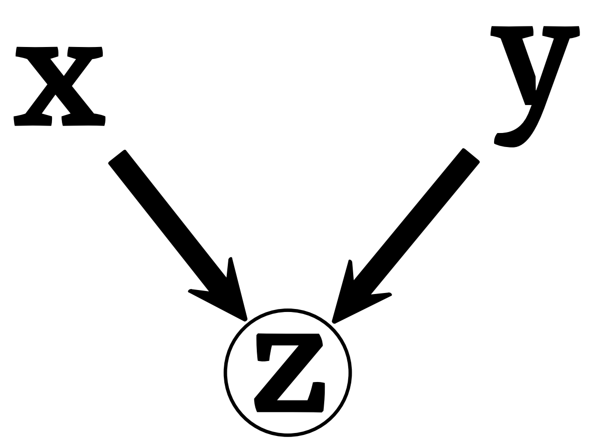
$$y \sim Normal(0,1)$$

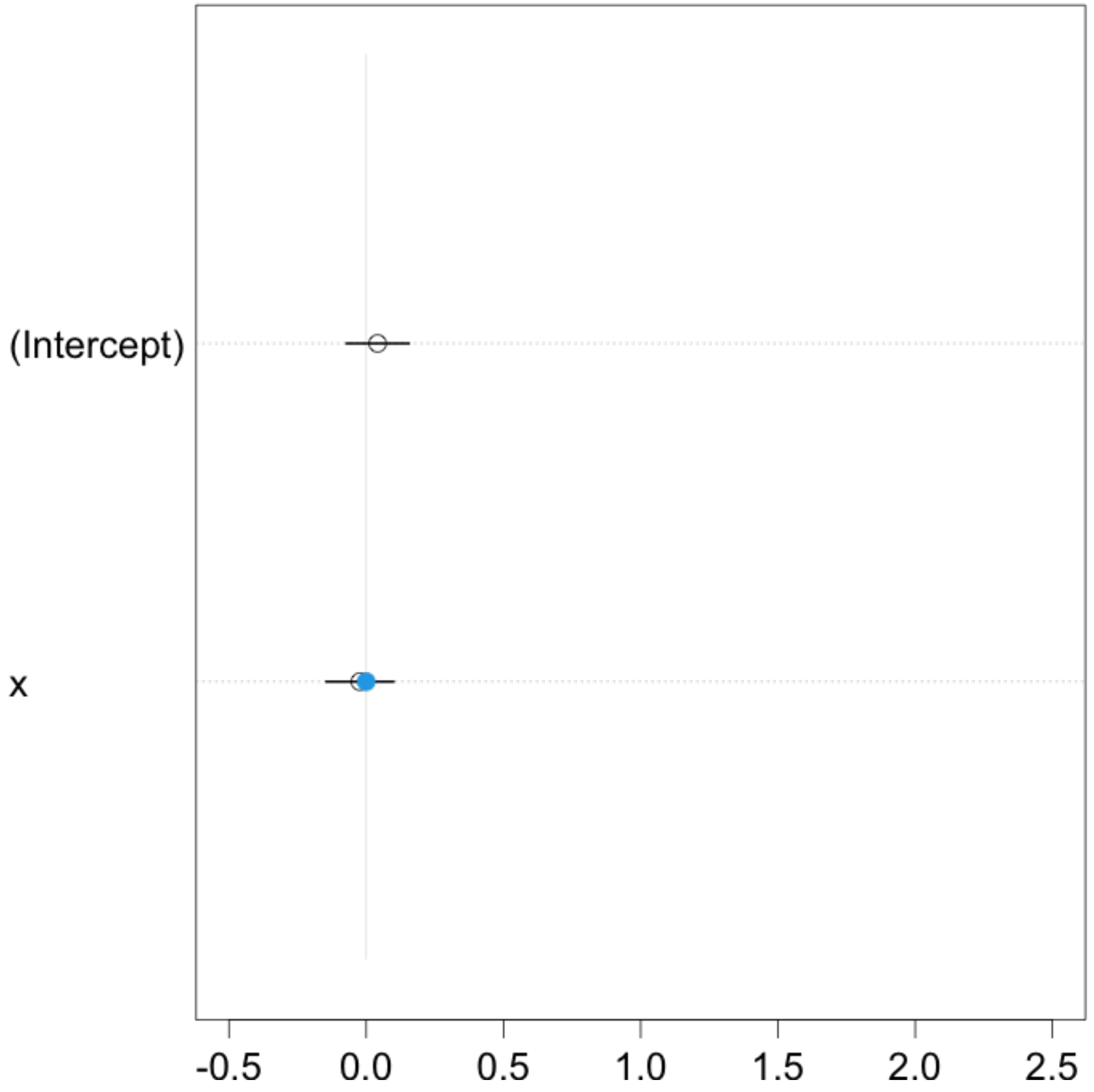
$$x \sim Normal(0,1)$$

$$z \sim Bernoulli(logit^{-1}(2x + 2y - 2))$$

Nenhum efeito de x em y, mas ambos afetam Z

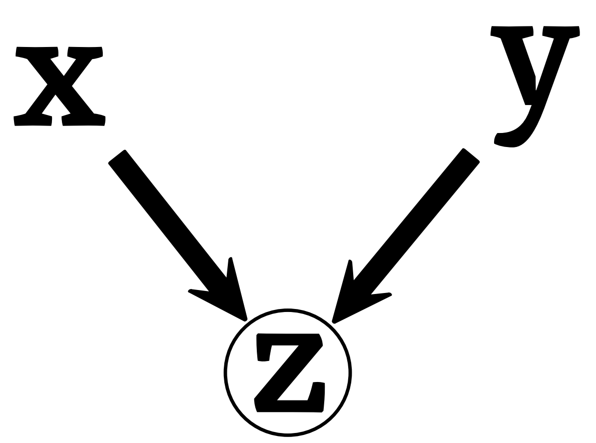
```
set.seed(1)
N = 200
x = rnorm(N)
y = rnorm(N)
z = rbinom(N, 1, inv_logit(2*x + 2*y - 2))

m1 = lm(y ~ x)

> (pm1 = precis(m1))
             mean   sd  5.5% 94.5%
(Intercept)  0.04 0.07 -0.07  0.16
x           -0.02 0.08 -0.15  0.10
```

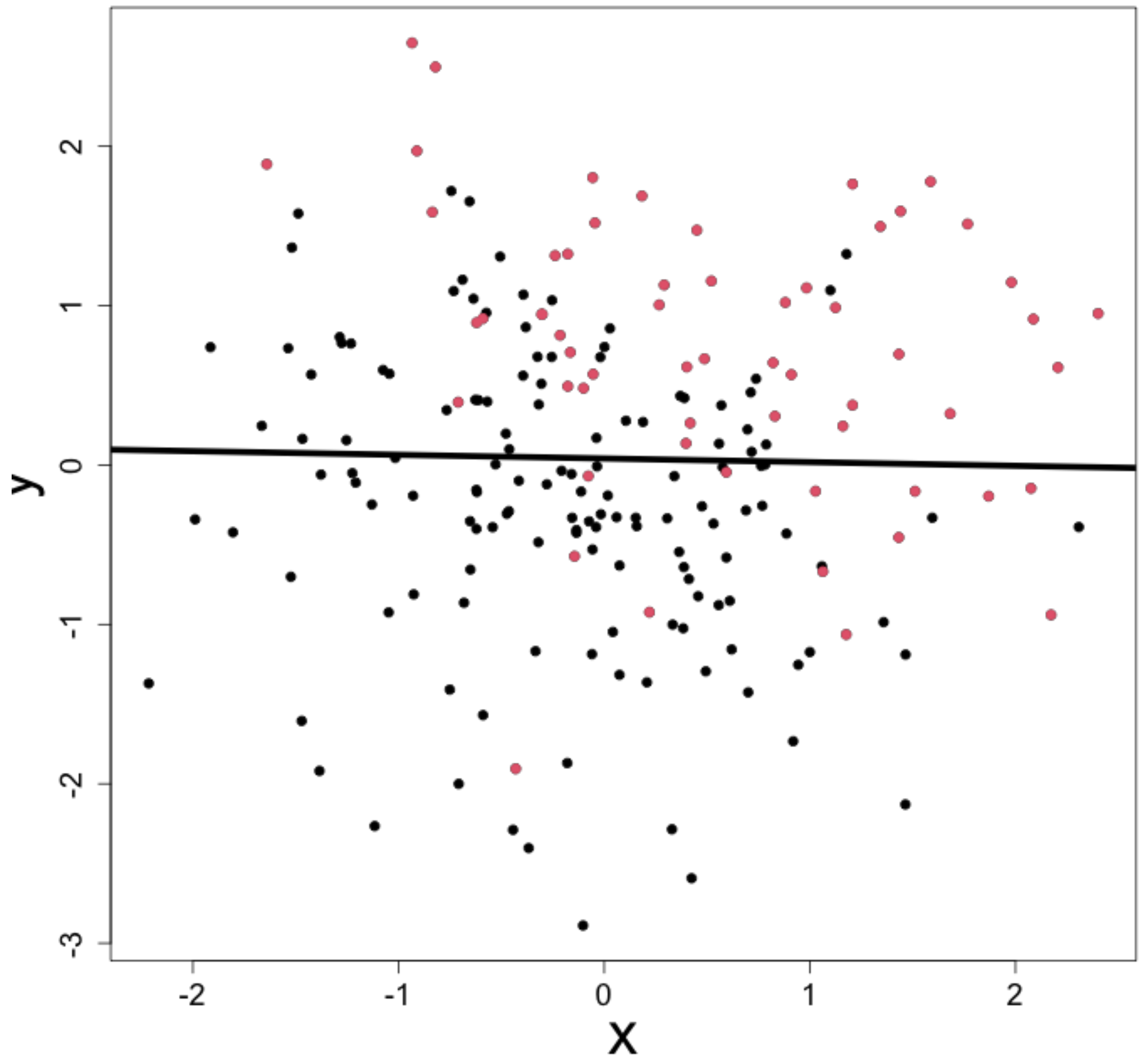




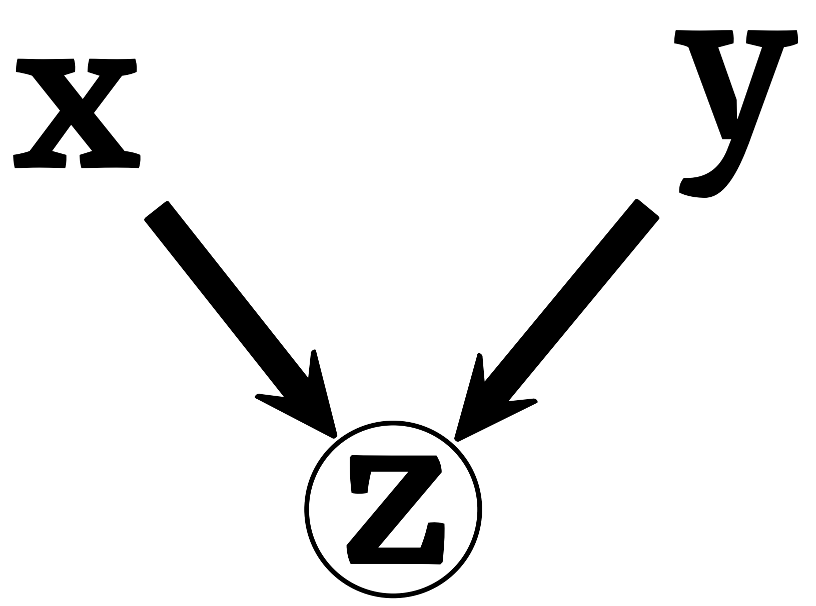
Nenhum efeito de x em y, mas ambos afetam Z

```
set.seed(1)
N = 200
x = rnorm(N)
y = rnorm(N)
z = rbinom(N, 1, inv_logit(2*x + 2*y - 2))

m1 = lm(y ~ x)

> (pm1 = precis(m1))
             mean   sd  5.5% 94.5%
(Intercept)  0.04 0.07 -0.07  0.16
x           -0.02 0.08 -0.15  0.10
```

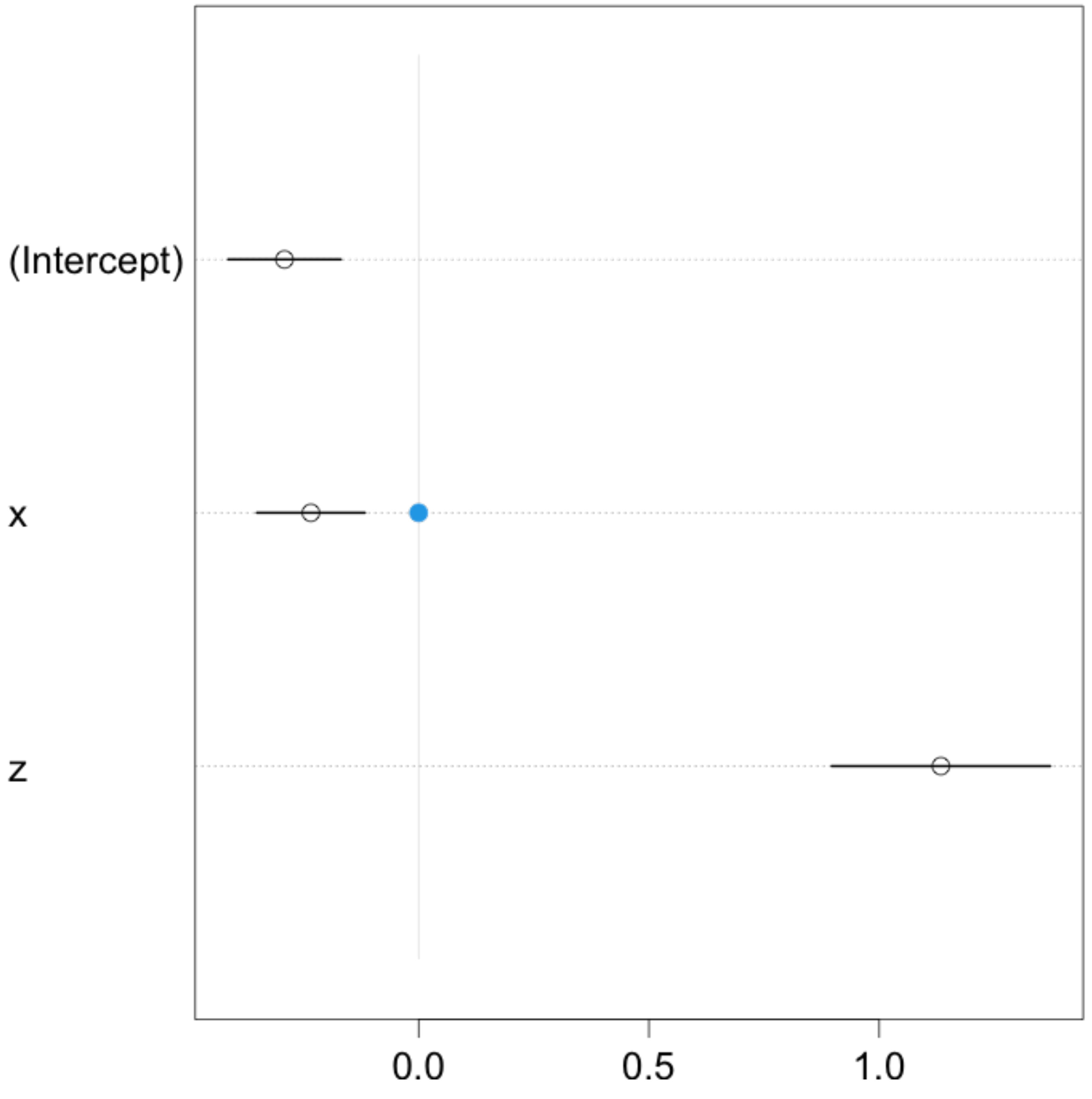




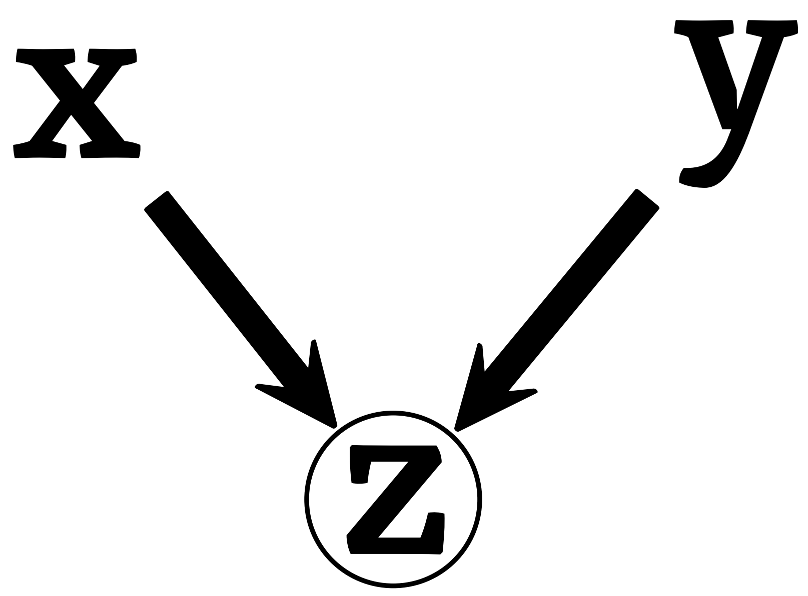
Agora, o modelo com o collider z incluído



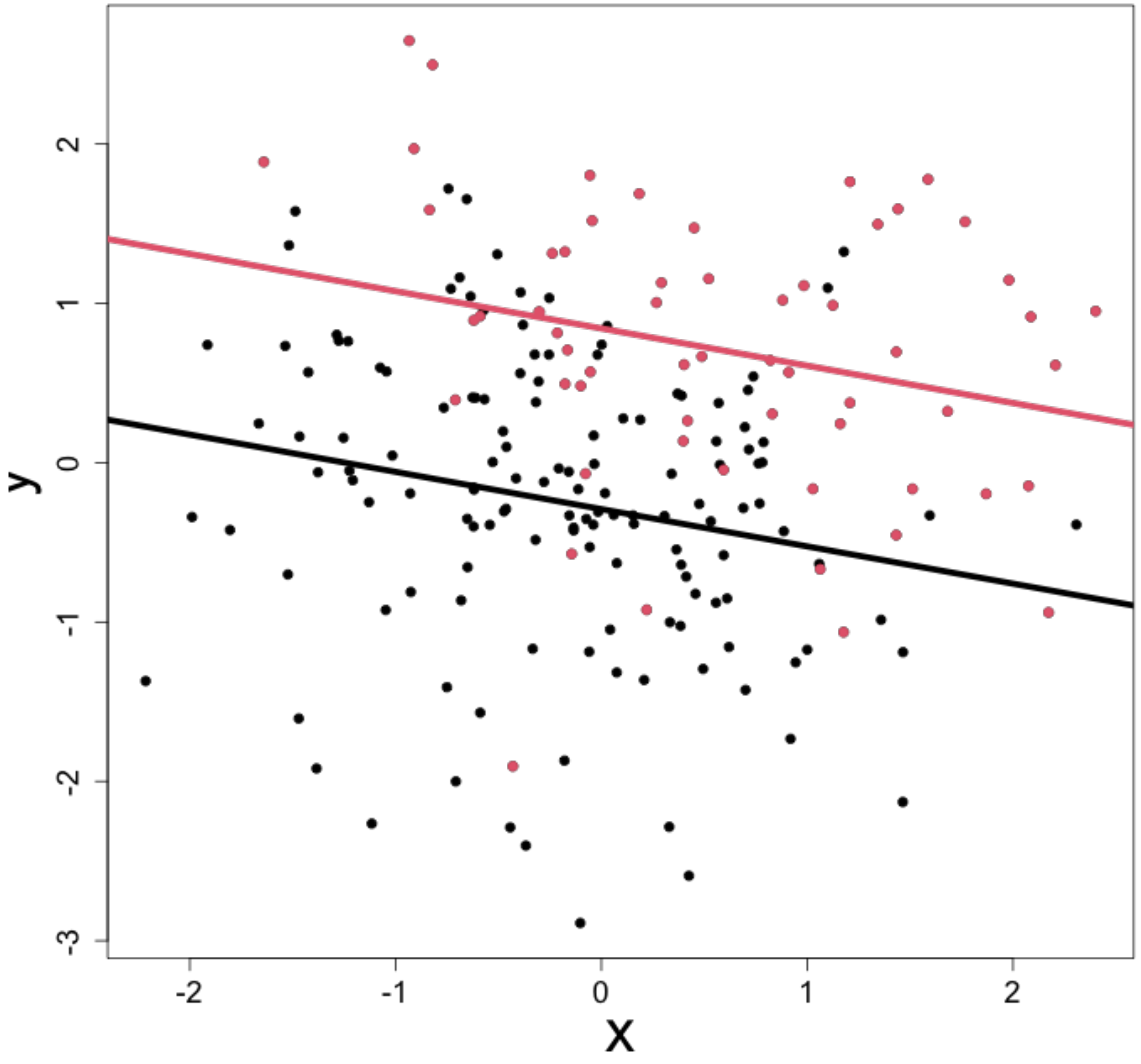

```
> m2 = lm(y ~ x + z)

> (pm2 = precis(m2))
             mean   sd  5.5% 94.5%
(Intercept) -0.29 0.08 -0.41 -0.17
x           -0.23 0.07 -0.35 -0.12
z            1.13 0.15  0.90  1.37
```



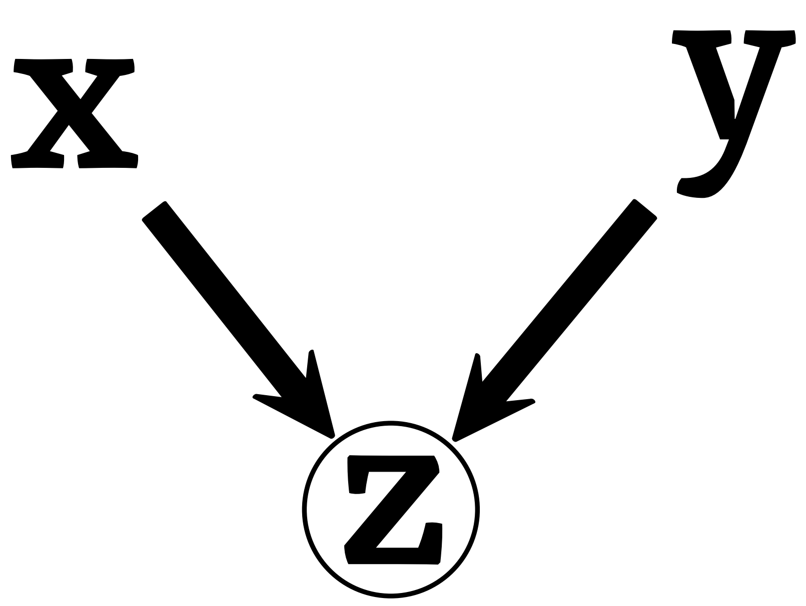
Agora, o modelo com o collider z incluído




```
> m2 = lm(y ~ x + z)

> (pm2 = precis(m2))
             mean   sd  5.5% 94.5%
(Intercept) -0.29 0.08 -0.41 -0.17
x           -0.23 0.07 -0.35 -0.12
z            1.13 0.15  0.90  1.37
```

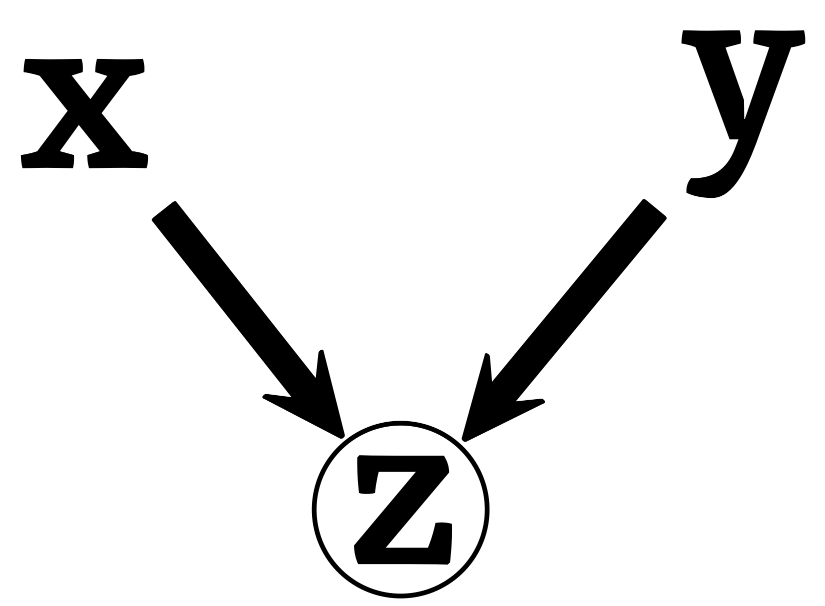


E o p-valor?!?




```
> summary(m1)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.93275 -0.54273 -0.02523  0.66833  2.58615 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04146    0.07168   0.578    0.564
x           -0.02308    0.07729  -0.299    0.766
```

E o p-valor?




```
> summary(m2)

Call:
lm(formula = y ~ x + z)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.84828 -0.48711  0.02999  0.59867  1.89105 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.29123    0.07672  -3.796 0.000196 ***
x           -0.23391    0.07347  -3.184 0.001689 ** 
z            1.13392    0.14858   7.632 9.72e-13 ***
```

No modelo enviezado pelo collider, dois preditores são significativos!

Talvez com comparação de modelos?

Nosso salvador?!!?

```
> AIC(m1, m2)
   df      AIC
m1  3 576.7421
m2  4 526.9382
```

$$m1 : y \sim x$$

$$m2 : y \sim x + z$$

Modelo com enviezado pelo collider tem AIC menor!

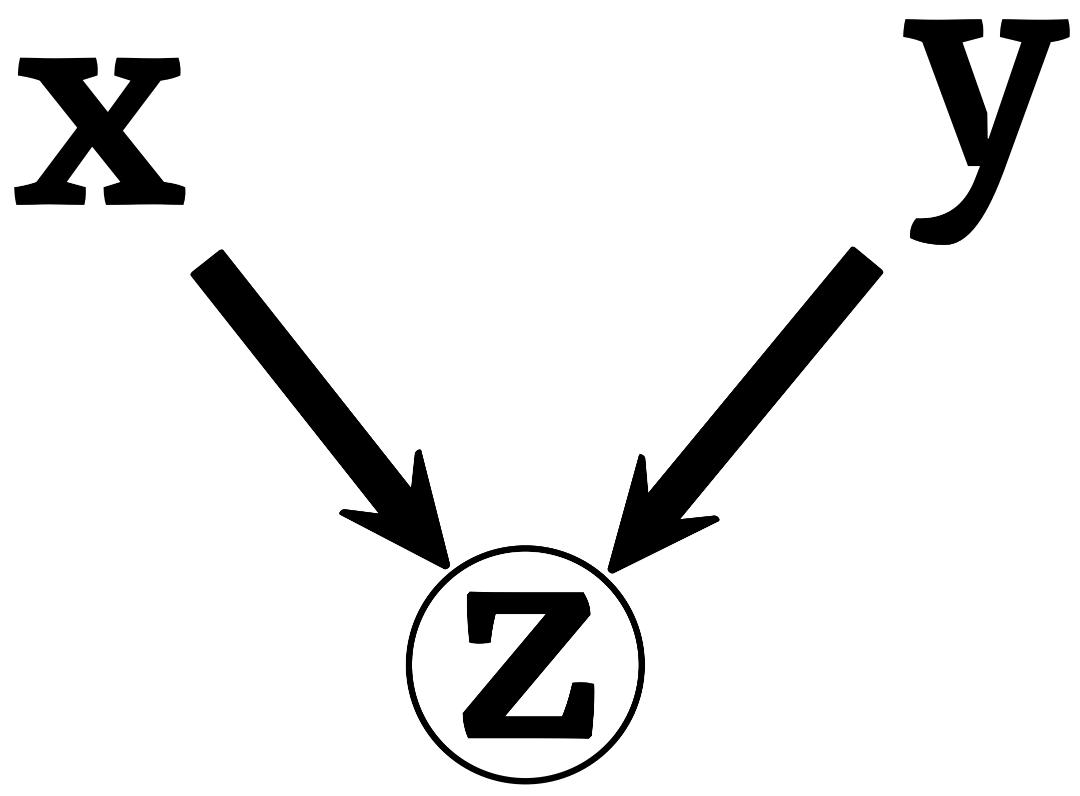
Caminhos e colliders

Incluir colliders abre caminhos!

- No caso do collider, a variável z forma um **caminho não-causal fechado** entre x e y. Adicionar z ao modelo **abre** esse caminho.

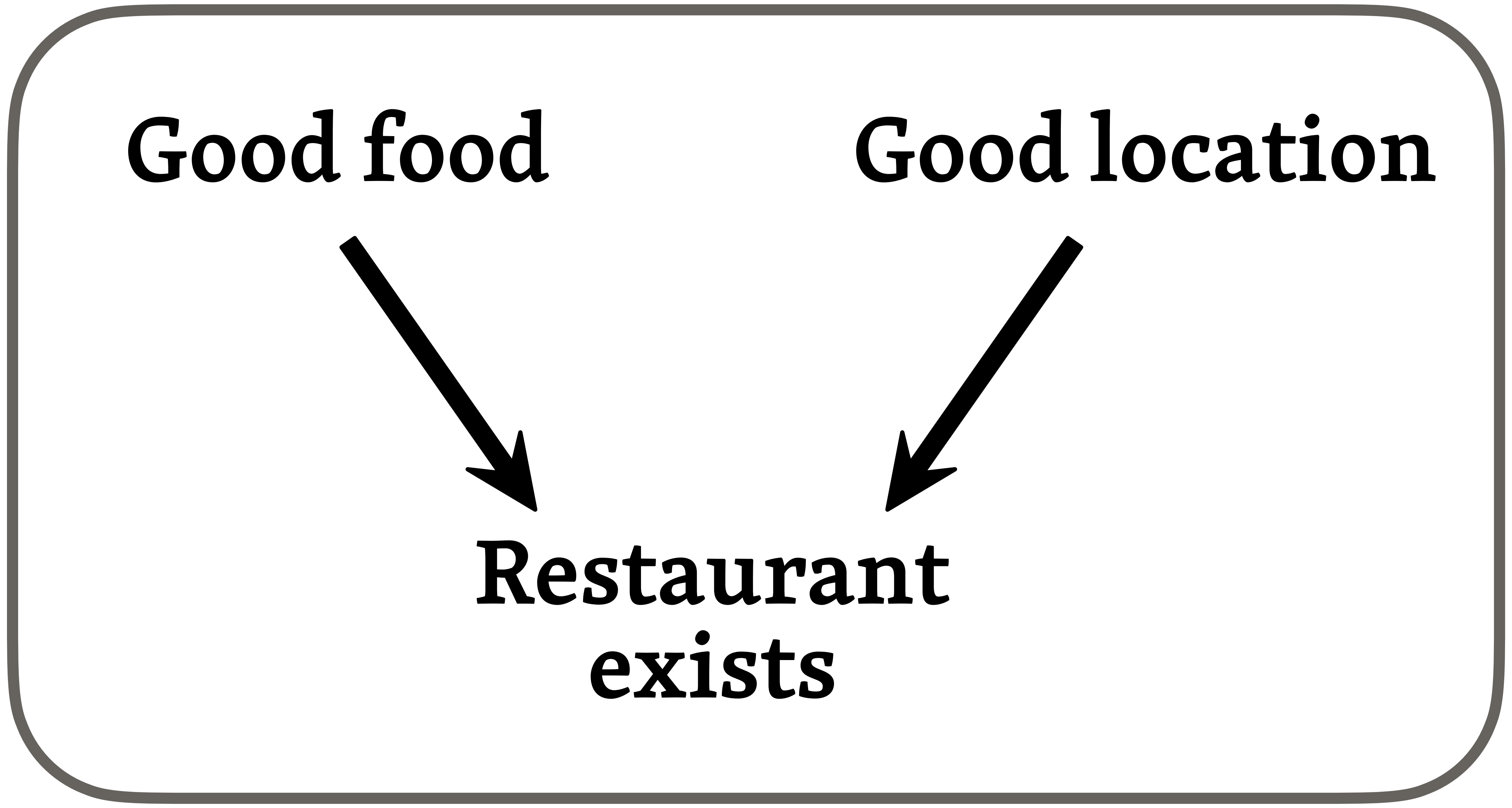
- Incluir um collider em modelos pode abrir caminhos não-causais e prejudicar seriamente as estimativas de efeitos causais

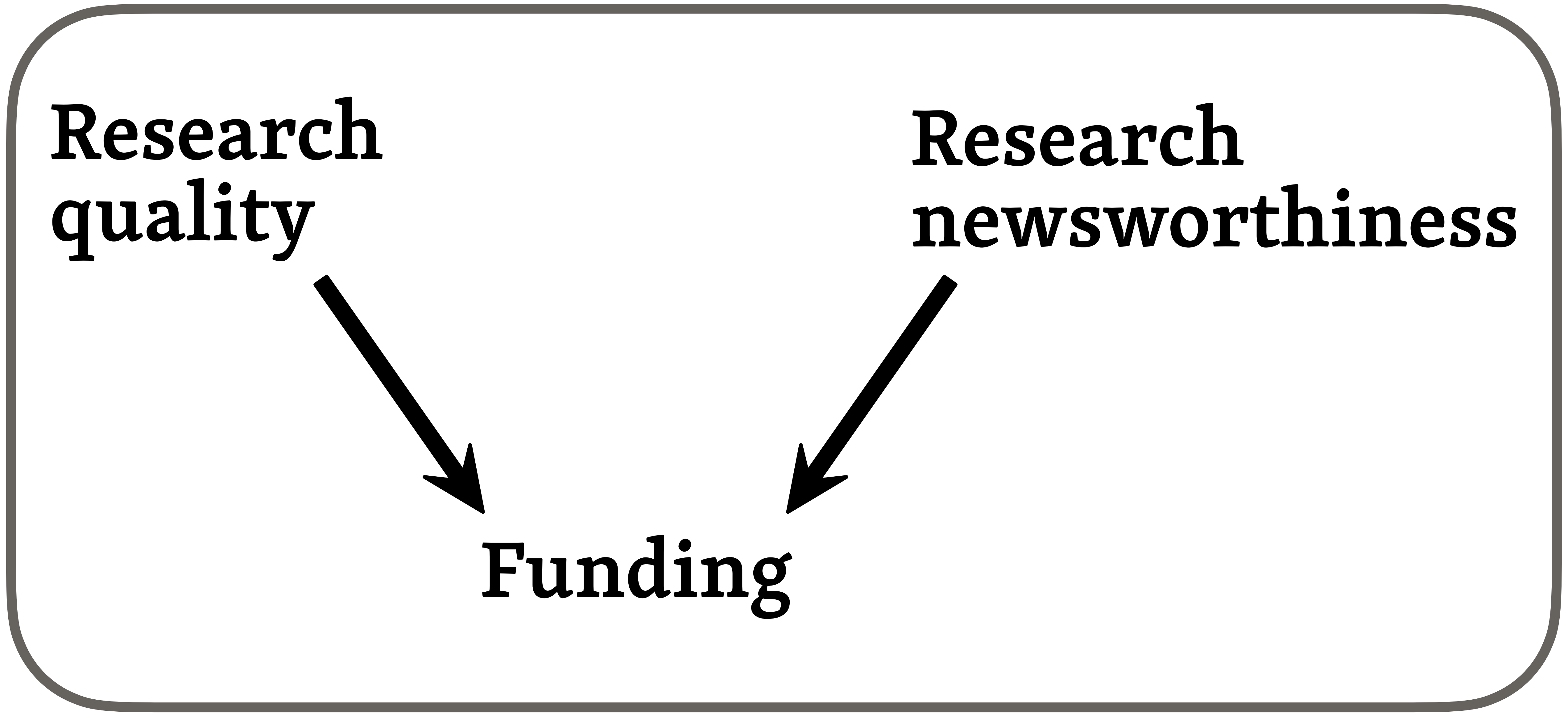
- Apesar disso, colliders trazem informação preditiva, e portanto levam a modelos com maior poder preditivo (menor AIC) e coeficientes significativos!



Colliders naturais podem afetar nossas amostras!

Dados coletáveis são frequentemente poluídos por colliders

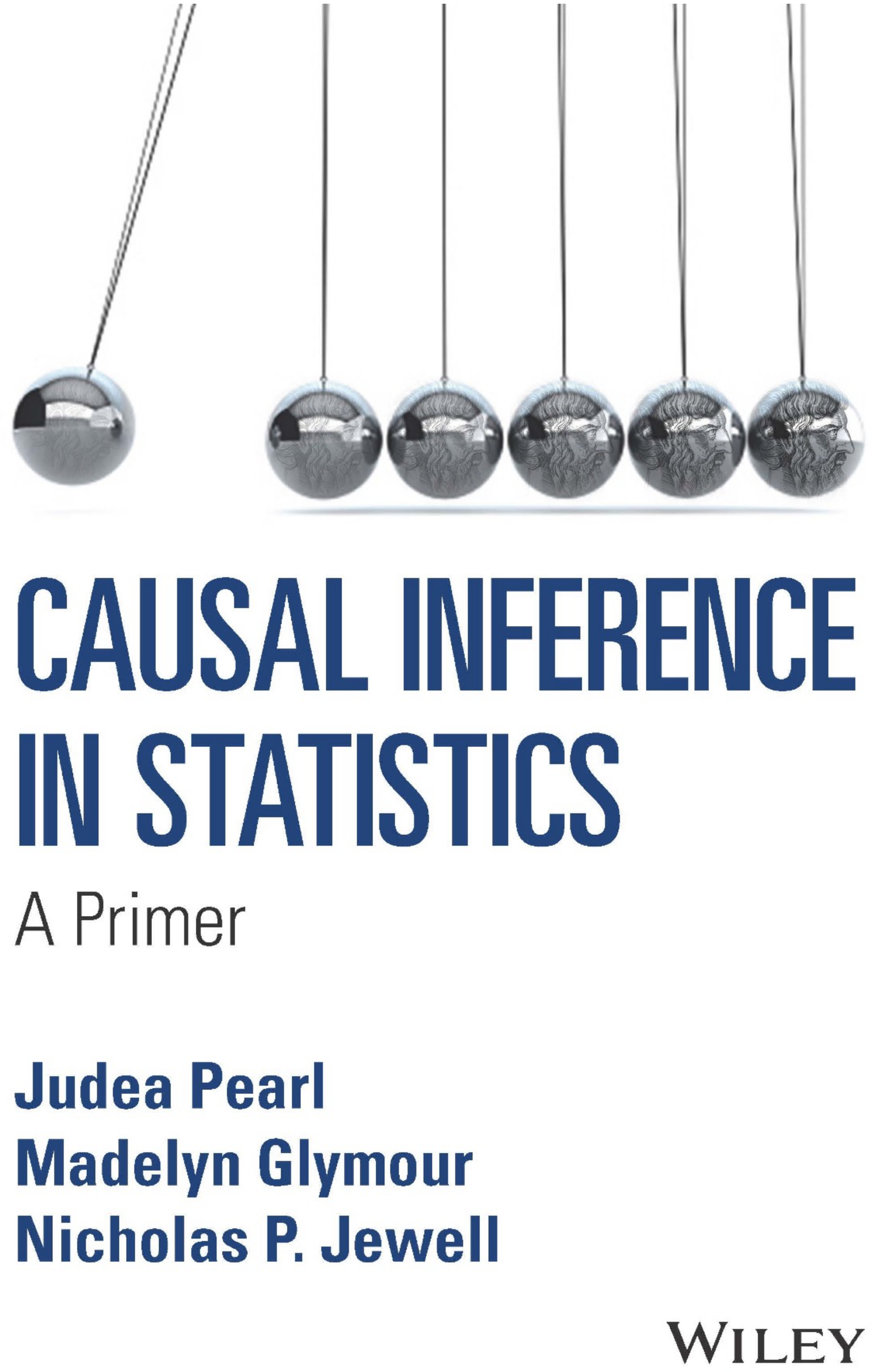




E AGORA?!

Usando DAGs para construir nossos modelos

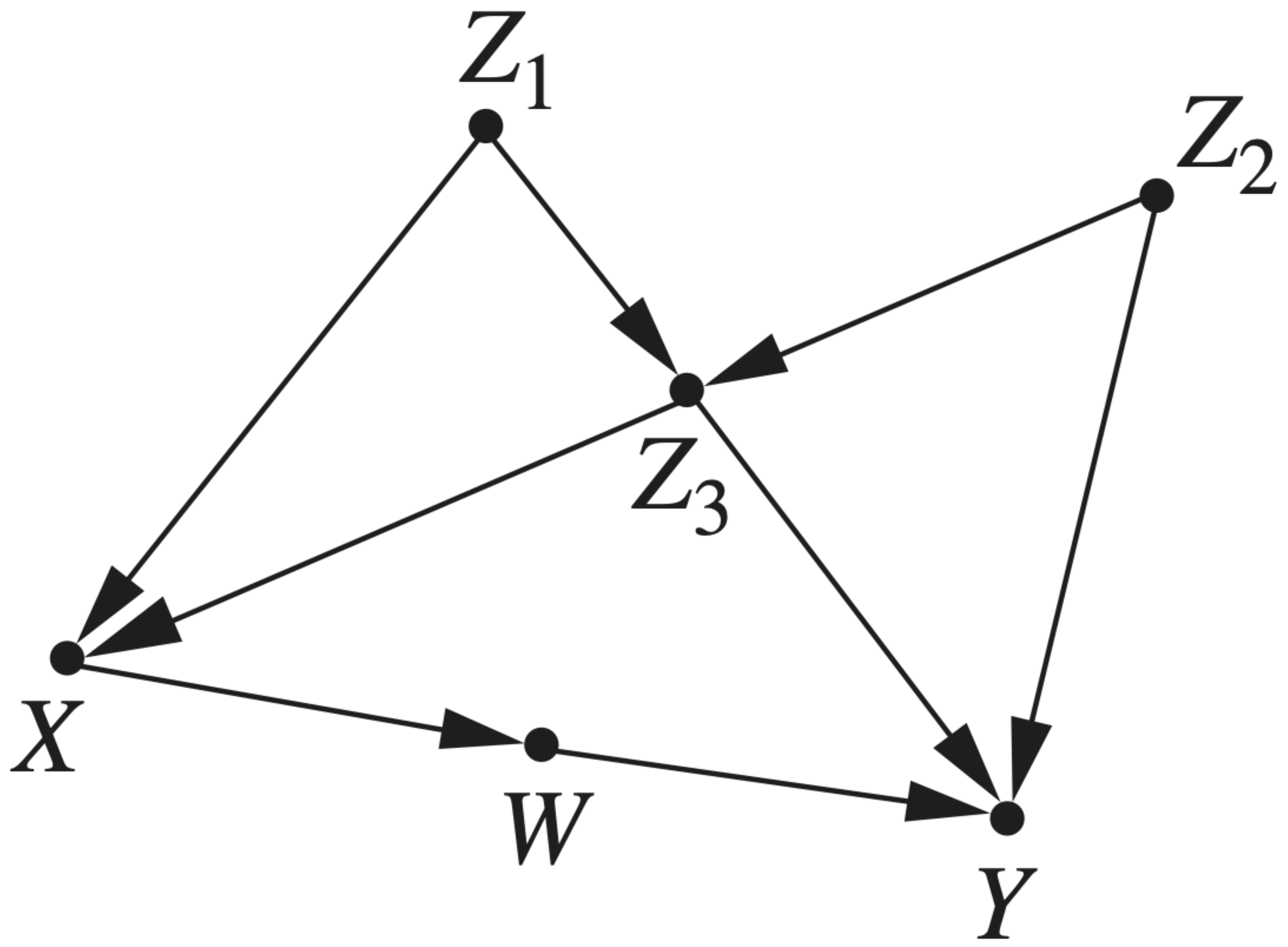
- Se representarmos nossas relações causais usando DAGs, temos um conjunto de regras que nos diz quais variáveis precisamos incluir no modelo para calcular um determinado efeito, ou determinar se o efeito pode ser estimado
  - Esse é um formalismo chamado **do-calculus**
- Existem MODELOS mais complicados, como Structural Equation Modeling ou **Full-Luxury Bayesian Inference**:
  - Regression, Fire, and Dangerous Things (1/3)



Caminhos abertos e Fechados

- Caminhos com forks ou pipes **não controlados** estão abertos
- Caminhos com colliders **estão bloqueados**, mas ficam abertos se **incluirmos o collider** no modelo.
- Para estimar o efeito causal de x em y, nós precisamos que todos os caminhos **não causais** estejam bloqueados

Identifique todos os caminhos entre x e y



Back door criterion

Para estimar o efeito causal de X sobre Y, identifique um conjunto de variáveis de **controle** de modo que nenhum descendente de X esteja no conjunto de controle, e todos os caminhos entre X e Y que contenham uma seta para X estejam bloqueados.

Nem Toda estimativa é equivalente

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients FREE

Daniel Westreich, Sander Greenland Author Notes

American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298,
https://doi.org/10.1093/aje/kws412
Published: 30 January 2013 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

Westreich, D. & Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am. J. Epidemiol. ***177****, 292–298 (2013)*

Nem Toda estimativa é equivalente

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients FREE

Daniel Westreich ✉, Sander Greenland   Author Notes

American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298,
https://doi.org/10.1093/aje/kws412
Published: 30 January 2013   **Article history** ▾

PDF   Split View   Cite   Permissions   Share ▾

Abstract

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

Table 1. Characteristics of the Patients at Baseline.*

Characteristic	Ivermectin (N=679)	Placebo (N=679)	Total (N=1358)
Age			
Median (IQR) — yr	49 (39–57)	49 (37–56)	49 (38–57)
Distribution — no. (%)			
≤50 yr	359 (52.9)	372 (54.8)	731 (53.8)
>50 yr	320 (47.1)	307 (45.2)	627 (46.2)
Female sex — no. (%)	383 (56.4)	408 (60.1)	791 (58.2)
Race — no. (%)†			
Mixed race	648 (95.4)	645 (95.0)	1293 (95.2)
White	6 (0.9)	6 (0.9)	12 (0.9)
Black	7 (1.0)	5 (0.7)	12 (0.9)
Other	1 (0.1)	0	1 (0.1)
Unknown	17 (2.5)	23 (3.4)	40 (2.9)
Body-mass index — no. (%)			
<30	347 (51.1)	336 (49.5)	683 (50.3)
≥30	332 (48.9)	343 (50.5)	675 (49.7)
Time since onset of symptoms — no. (%)			
0–3 days	302 (44.5)	295 (43.4)	597 (44.0)
4–7 days	377 (55.5)	384 (56.6)	761 (56.0)
Risk factors — no. (%)			
Chronic cardiac disease	14 (2.1)	10 (1.5)	24 (1.8)
Uncontrolled hypertension	55 (8.1)	59 (8.7)	114 (8.4)
Chronic pulmonary disease	18 (2.7)	23 (3.4)	41 (3.0)
Asthma	54 (8.0)	60 (8.8)	114 (8.4)
Chronic kidney disease	2 (0.3)	5 (0.7)	7 (0.5)
Type 1 diabetes mellitus	3 (0.4)	9 (1.3)	12 (0.9)
Type 2 diabetes mellitus	79 (12)	89 (13)	168 (12)
Autoimmune disease	2 (0.3)	2 (0.3)	4 (0.3)
Any other risk factor or coexisting condition	22 (3.2)	19 (2.8)	41 (3.0)

* Missingness in covariate data was handled with multiple imputation by chained equations.[16] IQR denotes interquartile range.
† Race was reported by the patient.

Westreich, D. & Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am. J. Epidemiol. **177**, *292–298 (2013)*

Nem Toda estimativa é equivalente

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients FREE

Daniel Westreich ✉, Sander Greenland   Author Notes

American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298,
https://doi.org/10.1093/aje/kws412
Published: 30 January 2013   **Article history** ▾

PDF   Split View   Cite   Permissions   Share ▾

Abstract

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

	General Linear Model				Ordinary Least Squares	
Independent variable	*df*	*MS*	*F*	*p*	β[a]	*SE*
Age	1	44.7	1.8	.175	−.04	.024
Gender (male)	1	294.7	12.1	.001	.10	.391
Education	1	35.2	1.4	.229	.04	.052
Financial strain	1	687.9	28.3	.000	.14	.206
Volunteer work	1	95.9	3.9	.047	.05	.409
Social support	1	95.6	3.9	.048	.05	.021
Religious participation	1	264.4	10.9	.001	−.09	.168
Cognitive deficit	1	202.1	8.3	.004	.08	.074
Stressful life events	1	591.3	24.3	.000	−.13	.082
Health status	1	1145.1	47.1	.000	−.21	.103
Daily activity limitations	1	1508.2	62.1	.000	−.24	.045
Vision	3	66.5	2.74	.021	−.11	.175
Hearing	3	2.2	1.0	.965	−.04	.169
Vision × Hearing	9	12.1	0.5	.876	.01	.160
Corrected model	26	577.9	23.8	.000		
R^2 (adjusted)					.376	

[a]Standardized regression coefficients.

Westreich, D. & Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am. J. Epidemiol. ***177****, 292–298 (2013)*

Nem Toda estimativa é equivalente

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients FREE

Daniel Westreich, Sander Greenland Author Notes

American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298,
https://doi.org/10.1093/aje/kws412
Published: 30 January 2013 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

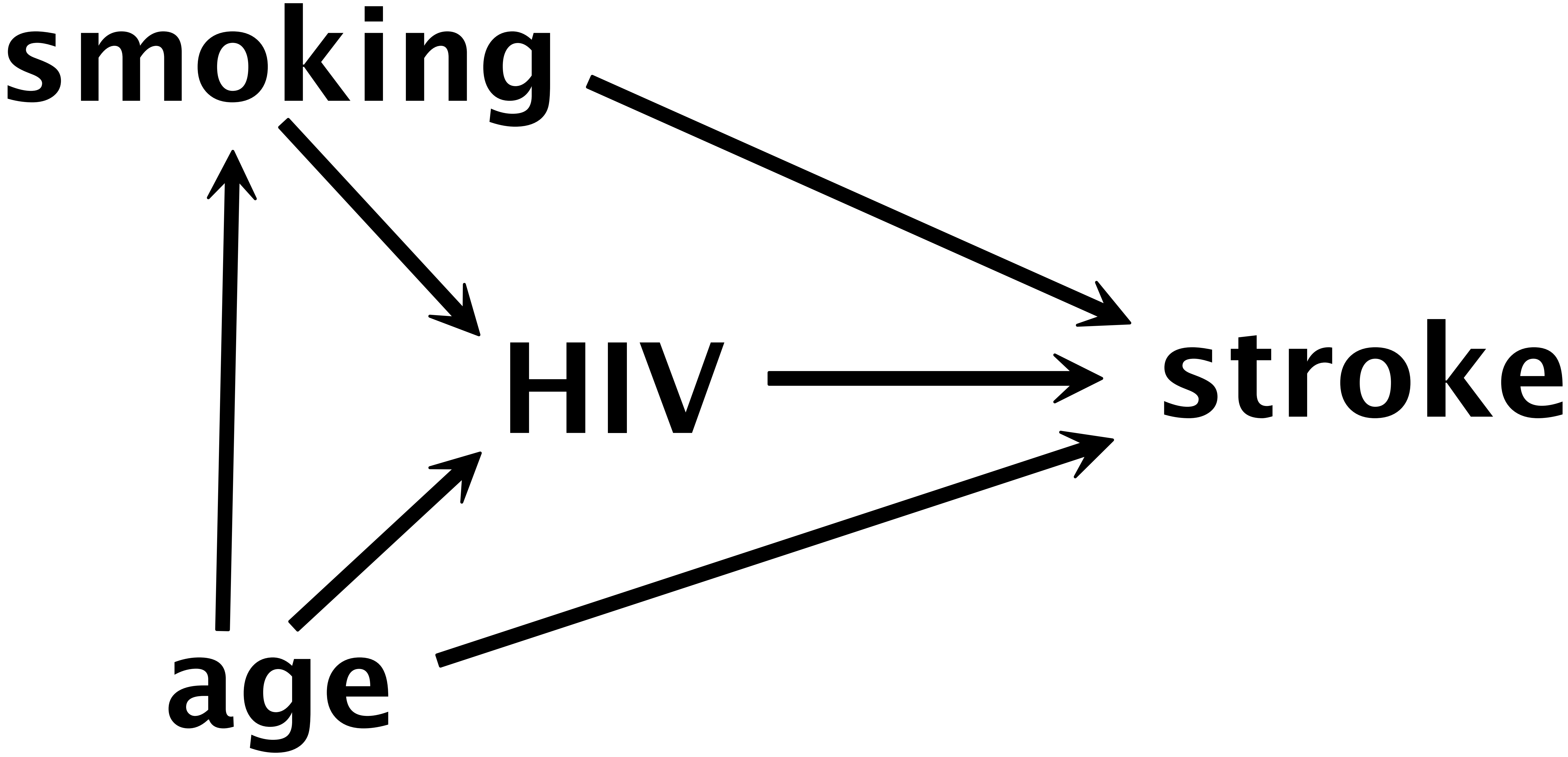
It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

	General Linear Model				Ordinary Least Squares	
Independent variable	*df*	*MS*	*F*	*p*	β[a]	*SE*
Age	1	44.7	1.8	.175	−.04	.024
Gender (male)	1	294.7	12.1	.001	.10	.391
Education	1	35.2	1.4	.229	.04	.052
Financial strain	1	687.9	28.3	.000	.14	.206
Volunteer work	1	95.9	3.9	.047	.05	.409
Social support	1	95.6	3.9	.048	.05	.021
Religious participation	1	264.4	10.9	.001	−.09	.168
Cognitive deficit	1	202.1	8.3	.004	.08	.074
Stressful life events	1	591.3	24.3	.000	−.13	.082
Health status	1	1145.1	47.1	.000	−.21	.103
Daily activity limitations	1	1508.2	62.1	.000	−.24	.045
Vision	3	66.5	2.74	.021	−.11	.175
Hearing	3	2.2	1.0	.965	−.04	.169
Vision × Hearing	9	12.1	0.5	.876	.01	.160
Corrected model	26	577.9	23.8	.000		
R^2 (adjusted)					.376	

[a]Standardized regression coefficients.

Westreich, D. & Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am. J. Epidemiol. ***177****, 292–298 (2013)*

HIV e Aneurismas



Good and bad controls

Bons controles

- Bloqueiam caminhos não causais.
- Aumentam a precisão das estimativas.
- Permitem a inferência de efeitos causais.

Controles ruins

- Bloqueiam caminhos causais.
- Abrem caminhos não causais (colliders)
- Reduzem a precisão das estimativas.
- Levam a estimativas enviesadas de efeitos causais.

Cinelli, C., Forney, A. & Pearl, J. A crash course in good and bad controls. (2020) doi:10.2139/ssrn.3689437

Sumário

- A única maneira de responder perguntas causais sobre os nossos sistemas é pensar sobre eles cientificamente.
- Métodos estatísticos não conseguem diferenciar entre associações causais e não causais (só estamos medindo diferenças!)
- Comparação de modelos, significância e outras ferramentas estatísticas são úteis, mas não podem substituir o pensamento científico e o conhecimento de area.

- Use diagramas causais sempre para pensar nas relações que você está tentando medir.
- Use o **backdoor criteria** para decidir quais variáveis você deveria incluir no modelo, dependendo de qual pergunta você está tentando responder.
  - Qual efeito você quer estimar?
  - **Estimativas diferentes** podem precisar de **modelos diferentes**.

Quero saber mais!