# Advanced Econometrics

Mestrado Profissional em Economia

Insper Instituto de Ensino e Pesquisa

Prof. Hedibert Freitas Lopes

**Lista de Exercícios 3**

**Decomposição em produto, confundimento e paradoxo de Simpson**

Data de entrega via blackboard: 27 de novembro de 2026 (até 19h30)
Cada grupo de trabalho pode conter até 4 estudantes

---

## Contexto

Esta lista é baseada na *Study question 1.5.2* de Pearl, Glymour & Jewell (2016), *Causal Inference in Statistics: A Primer*, Wiley, Seção 1.5.2 (pp. 29–31). Um resumo da seção sobre a *regra da decomposição em produto* está no Apêndice, ao final da lista.

Considere uma população de pacientes em que uma fração $r$ dos indivíduos é portadora de uma síndrome fatal $Z$. A síndrome também torna desconfortável o uso de um medicamento $X$ que prolonga a vida. A notação é:

- $Z = z_1$: portador da síndrome; $Z = z_0$: não portador;
- $X = x_1$: toma o medicamento; $X = x_0$: não toma;
- $Y = y_1$: morte; $Y = y_0$: sobrevivência.

As probabilidades que definem o modelo estão na tabela abaixo. Portadores da síndrome tendem a evitar o medicamento, isto é, $q_2 < q_1$.

| Quantidade | Parâmetro |
|---|---|
| $P(z_1)$ | $r$ |
| $P(x_1 \mid z_0)$ | $q_1$ |
| $P(x_1 \mid z_1)$ | $q_2$ |
| $P(y_1 \mid x_0, z_0)$ | $p_1$ |
| $P(y_1 \mid x_1, z_0)$ | $p_2$ |
| $P(y_1 \mid x_0, z_1)$ | $p_3$ |
| $P(y_1 \mid x_1, z_1)$ | $p_4$ |

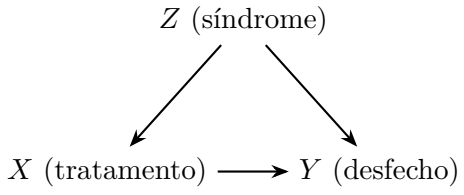



Figura: modelo em que a síndrome $Z$, não observada, afeta tanto o tratamento $X$ quanto o desfecho $Y$ (adaptado da Figura 1.10 de Pearl *et al.*, 2016).

## Exercícios

**1. Lendo o grafo.**

(a) Complete a tabela com os *pais* de cada variável no grafo da figura.

*Dica:* os pais de uma variável são as variáveis de onde saem setas que chegam nela. Uma variável sem setas chegando não tem pais.

| Variável | Pais |
|---|---|
| $Z$ | _____ |
| $X$ | _____ |
| $Y$ | _____ |

(b) Usando a regra da decomposição em produto (eq. (1.29) do Apêndice), escreva $P(x, y, z)$ como um produto de três probabilidades.

*Dica:* um fator para cada variável, sempre na forma $P(\text{variável} \mid \text{seus pais})$.

**2. Um exemplo numérico.** A partir daqui use os valores

$$r = 0{,}5, \quad q_1 = 0{,}9, \quad q_2 = 0{,}1, \quad p_1 = 0{,}10, \quad p_2 = 0{,}15, \quad p_3 = 0{,}60, \quad p_4 = 0{,}70.$$

Ou seja: metade dos pacientes tem a síndrome; 90% dos não portadores tomam o medicamento, mas só 10% dos portadores o tomam.

(a) Complete a tabela da distribuição conjunta $P(x, y, z)$. As duas primeiras linhas já estão preenchidas como exemplo.

*Dica:* use o item 1(b) e lembre que $P(z_0) = 1 - r$, $P(x_0 \mid z) = 1 - P(x_1 \mid z)$ e $P(y_0 \mid x, z) = 1 - P(y_1 \mid x, z)$. Pense numa árvore: primeiro $Z$, depois $X$, depois $Y$, e multiplique as probabilidades ao longo de cada caminho.

| $z$ | $x$ | $y$ | $P(z)$ | $P(x \mid z)$ | $P(y \mid x, z)$ | $P(x, y, z)$ |
|---|---|---|---|---|---|---|
| $z_0$ | $x_0$ | $y_1$ | 0,5 | 0,1 | 0,10 | 0,0050 |
| $z_0$ | $x_0$ | $y_0$ | 0,5 | 0,1 | 0,90 | 0,0450 |
| $z_0$ | $x_1$ | $y_1$ | 0,5 | _____ | _____ | _____ |
| $z_0$ | $x_1$ | $y_0$ | 0,5 | _____ | _____ | _____ |
| $z_1$ | $x_0$ | $y_1$ | 0,5 | _____ | _____ | _____ |
| $z_1$ | $x_0$ | $y_0$ | 0,5 | _____ | _____ | _____ |
| $z_1$ | $x_1$ | $y_1$ | 0,5 | _____ | _____ | _____ |
| $z_1$ | $x_1$ | $y_0$ | 0,5 | _____ | _____ | _____ |
| | | | | | Soma | _____ |

Confira que a última coluna soma 1.

(b) Usando a tabela, calcule a proporção de mortes entre os que tomam e entre os que não tomam o medicamento, $P(y_1 \mid x_1)$ e $P(y_1 \mid x_0)$.

*Dica:* $P(y_1 \mid x_1) = P(x_1, y_1)/P(x_1)$. Para obter $P(x_1, y_1)$, some as linhas com $x_1$ e $y_1$ (uma com $z_0$ e outra com $z_1$). Para obter $P(x_1)$, some todas as linhas com $x_1$.

(c) Calcule a diferença $P(y_1 \mid x_1) - P(y_1 \mid x_0)$ em três populações: (1) só os portadores da síndrome; (2) só os não portadores; (3) a população toda.

*Dica:* para (1) e (2) não é preciso fazer contas: as respostas estão na tabela de parâmetros $(p_1, \ldots, p_4)$. Para (3), use o item (b).

(d) O medicamento parece bom ou ruim em cada uma das três populações? Esse fenômeno é conhecido como *paradoxo de Simpson*. Explique com suas palavras o que está acontecendo.

*Dica:* calcule a proporção de portadores da síndrome entre os que tomam o medicamento, $P(z_1 \mid x_1) = P(x_1, z_1)/P(x_1)$, e entre os que não tomam, $P(z_1 \mid x_0)$. Os dois grupos são comparáveis?

**3. Afinal, o medicamento deve ser receitado?**

(a) O *efeito causal médio* do medicamento é obtido pela *fórmula de ajuste*:

$$P(y_1 \mid do(x)) = P(y_1 \mid x, z_0)\, P(z_0) + P(y_1 \mid x, z_1)\, P(z_1).$$

Calcule $P(y_1 \mid do(x_1))$, $P(y_1 \mid do(x_0))$ e a diferença entre elas.

*Dica:* é uma média das mortalidades de cada grupo, com pesos iguais à proporção de cada grupo *na população* ($P(z_0) = P(z_1) = 0{,}5$), e não no grupo de tratados. Basta usar os $p$'s.

(b) Compare o resultado de (a) com a diferença do item 2(c)(3). Qual delas responde à pergunta "o medicamento deve ser receitado?". Justifique.

(c) Na figura, a síndrome $Z$ é *não observada*. Se você tivesse apenas os dados de $X$ e $Y$, conseguiria calcular o efeito de (a)? Cite uma forma de contornar o problema.

*Dica:* pense no que um experimento aleatorizado faz com a seta $Z \to X$.

**4. Confirmando por simulação.** Rode o código abaixo em `R`. Ele simula 100.000 pacientes do modelo do item 2.

```
set.seed(2026)
r <- 0.5; q1 <- 0.9; q2 <- 0.1
p1 <- 0.10; p2 <- 0.15; p3 <- 0.60; p4 <- 0.70
N <- 100000
z <- rbinom(N, 1, r) # 1 = tem a sindrome
x <- rbinom(N, 1, ifelse(z == 1, q2, q1)) # 1 = toma o medicamento
y <- rbinom(N, 1, ifelse(z == 0, ifelse(x == 1, p2, p1),
                                 ifelse(x == 1, p4, p3))) # 1 = morre
table(z, x, y) / N # compare com o item 2(a)

mean(y[x == 1]) - mean(y[x == 0]) # diferenca bruta
mean(y[x == 1 & z == 1]) - mean(y[x == 0 & z == 1]) # so portadores
mean(y[x == 1 & z == 0]) - mean(y[x == 0 & z == 0]) # so nao portadores

summary(lm(y ~ x)) # regressao sem controlar pela sindrome
summary(lm(y ~ x + z)) # regressao controlando pela sindrome
```

(a) A tabela simulada está próxima da tabela do item 2(a)? E as três diferenças estão próximas das do item 2(c)?

(b) Compare o coeficiente de `x` nas duas regressões. Qual delas se aproxima da diferença bruta e qual se aproxima do efeito causal do item 3(a)? O que isso ensina sobre a importância de incluir (ou omitir) uma variável numa regressão?

*Dica:* lembre do *viés de variável omitida* visto em aula.

**5. Para que serve a decomposição em produto? (exercício curto)** Volte ao exemplo do Apêndice: a cadeia $X \to Y \to Z \to W$, com $X$ = nuvens, $Y$ = chuva, $Z$ = pavimento molhado e $W$ = pavimento escorregadio, todas binárias.

(a) Usando os valores "de bolso" do Apêndice (0,5; 0,75; 0,9; 0,05), calcule

$$P(\text{nuvens}, \text{sem chuva}, \text{pavimento seco}, \text{escorregadio}).$$

O valor 0,23 é plausível?

(b) Sem nenhum grafo, quantas probabilidades precisamos especificar para a conjunta de quatro variáveis binárias? E usando a cadeia?

*Dica:* sem grafo, são $2^4$ combinações, mas a última é determinada pelas demais (a soma tem de ser 1). Com a cadeia, conte: $P(x)$ precisa de 1 número; $P(y \mid x)$ precisa de 2 (um para cada valor de $x$); e assim por diante.

(c) Com apenas 45 observações, por que é mais fácil estimar bem os fatores da cadeia do que as 15 probabilidades da tabela completa?

*Dica:* quantas observações, em média, cairiam em cada cela da tabela completa? E em cada cela de $P(x)$?

---

**Observações.** Entregue um relatório com as contas, tabelas e comentários, acompanhado da saída do código `R`. As contas podem ser feitas à mão, em planilha ou no `R`.

**Referência.** Pearl, J., Glymour, M. & Jewell, N. P. (2016), *Causal Inference in Statistics: A Primer*, Wiley, Seção 1.5.2.

# Apêndice: a regra da decomposição em produto

**(resumo da Seção 1.5.2 de Pearl, Glymour & Jewell, 2016, pp. 29–31)**

**Motivação.** Uma distribuição conjunta pode ser apresentada de duas formas. A primeira é uma tabela com a probabilidade de cada combinação de valores: é fácil de ler, mas cresce exponencialmente (10 variáveis binárias já exigem $2^{10} = 1024$ linhas). A segunda é um modelo causal estrutural (SCM) totalmente especificado: basta conhecer as $n$ funções que ligam as variáveis e a distribuição dos termos de erro. Na prática, porém, muitas vezes sabemos *que* uma variável causa outra sem conhecer a forma funcional ou a distribuição dos erros. Mesmo quando as conhecemos, escrevê-las pode ser difícil, sobretudo com variáveis discretas. Os modelos gráficos contornam as duas dificuldades.

**Regra da decomposição em produto.** Se o grafo do modelo é acíclico, a distribuição conjunta das variáveis é o produto das distribuições condicionais de cada variável dados os seus pais, isto é, um fator $P(\text{filho} \mid \text{pais})$ para cada "família" do grafo:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid pa_i), \tag{1.29}$$

em que $pa_i$ denota os valores dos pais de $X_i$. A regra decorre de certas independências que valem em qualquer modelo com esse grafo, discutidas no Capítulo 2 do livro. Por exemplo, na cadeia $X \to Y \to Z$,

$$P(X = x, Y = y, Z = z) = P(X = x)\, P(Y = y \mid X = x)\, P(Z = z \mid Y = y).$$

**Economia de espaço.** Em vez de uma tabela com todas as triplas $(x, y, z)$, bastam três tabelas pequenas, para $X$, $Y \mid X$ e $Z \mid Y$, multiplicadas quando necessário.

**Economia na estimação.** Para estimar a conjunta a partir de dados gerados por esse modelo, não é preciso contar a frequência de cada tripla: basta contar as frequências de $x$, de $(y \mid x)$ e de $(z \mid y)$ e multiplicar. Isso reduz o tempo de processamento em modelos grandes e torna as estimativas mais precisas. As hipóteses contidas no grafo transformam um problema de estimação de alta dimensão em vários problemas de baixa dimensão. Sem conhecer a estrutura gráfica, a estimação com muitas variáveis e amostras pequenas ou moderadas torna-se inviável (a "maldição da dimensionalidade"). E tudo isso pode ser feito sem conhecer as funções, os parâmetros ou as distribuições dos erros do SCM.

**Exemplo de julgamento.** Na cadeia $X \to Y \to Z \to W$, com $X$ = nuvens/sem nuvens, $Y$ = chuva/sem chuva, $Z$ = pavimento molhado/seco e $W$ = pavimento escorregadio/não escorregadio, é difícil dizer diretamente se é plausível que $P(\text{nuvens}, \text{sem chuva}, \text{seco}, \text{escorregadio}) = 0{,}23$. Pela regra do produto, essa probabilidade é

$$P(\text{nuvens})\, P(\text{sem chuva} \mid \text{nuvens})\, P(\text{seco} \mid \text{sem chuva})\, P(\text{escorregadio} \mid \text{seco}),$$

e cada fator é fácil de avaliar: algo como 0,5, 0,75, 0,9 e 0,05, respectivamente. O resultado, $0{,}5 \times 0{,}75 \times 0{,}9 \times 0{,}05 \approx 0{,}0169$, mostra que 0,23 é implausível. O livro usa essa regra sempre que precisa raciocinar com probabilidades numéricas sem escrever tabelas grandes.

**Exemplo de estimação.** Ainda na cadeia de quatro variáveis binárias, a conjunta tem $16 - 1 = 15$ probabilidades a determinar. Com 45 observações $(x, y, z, w)$, cada cela recebe em média cerca de três; algumas recebem uma ou duas e outras ficam vazias, de modo que as frequências por cela estimam mal as proporções populacionais. Com a decomposição em produto, as 45 observações são divididas em categorias muito maiores: para $P(x)$, só duas celas ($X = 1$ e $X = 0$); para $P(y \mid x)$, $P(z \mid y)$ e $P(w \mid z)$, quatro celas cada (as combinações de valores do filho e do pai). A chance de uma cela ficar vazia é muito menor e as frequências estimadas são muito mais precisas do que na divisão original em 15 celas.

**Conclusão da seção.** O conhecimento qualitativo contido no grafo simplifica a estimação e aumenta sua precisão. Além disso, como mostram as seções seguintes do livro, o grafo da história causal permite inferir muito mais a partir de um conjunto de dados.