

MAT02010 - Tópicos Avançados em Estatística II

Métodos de análise

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

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

Porto Alegre, 2019

Relembrando

- ▶ 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.

Métodos de estimação

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

Métodos de estimação: estratificação

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

Métodos de estimação: estratificação

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

Métodos de estimação: estratificação

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

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

Métodos de estimação: regressão

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

Métodos de estimação: regressão

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

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

- ▶ Assim,

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

Métodos de estimação: regressão

Se

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

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

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

Métodos de estimação: regressão

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

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

Métodos de estimação: regressão

► Assim,

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

Métodos de estimação: regressão

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

Métodos de estimação: o escore de propensão

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

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

Métodos de estimação: o escore de propensão

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

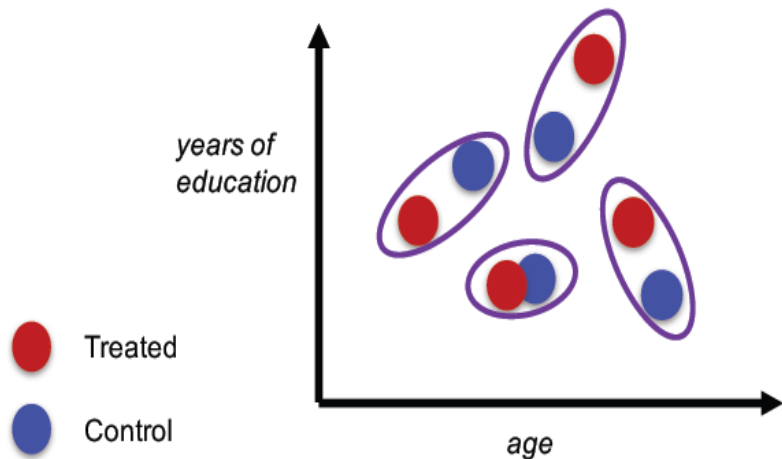
Métodos de estimação: o escore de propensão

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

Métodos de estimação: pareamento

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

Métodos de estimação: pareamento



Comentários finais

- ▶ Neste curso apresentamos os conceitos básicos da abordagem de desfechos potenciais para inferência causal.
 - ▶ Definição dos efeitos causais.
 - ▶ Identificação dos efeitos causais.
 - ▶ Estimação dos efeitos causais.
- ▶ O que vem depois?
 - ▶ Análise de sensibilidade.
 - ▶ Análise de mediação: decomposição do efeito total em efeitos direto e indireto.
 - ▶ Análise de interação: modificação do efeito.
 - ▶ Confundidores intermediários.
 - ▶ Confundidores tempo-dependentes.

Avisos

Avisos

- ▶ **Próxima semana (04/12): SEMINÁRIOS!**
- ▶ **Para casa:** Ler artigos de referência dos seminários que serão apresentados.

Por hoje é só!

Today's Random Medical News

from the New England Journal of Panic-Inducing Gobbledygook

JIM BROWN



CAN CAUSE



IN



ACCORDING TO A REPORT RELEASED TODAY...

NEWS