MI628 / ME920 - Inferência Causal

– Lista 1 –

Carlos Trucíos

Instruções

- A resolução da lista será discutida no dia 12/03 em sala de aula (a participação será avaliada). Os **alunos** apresentarão a solução e farão a discussão pertinente de cada um dos exercícios.
- Os exercícios computacionais deverão ser resolvidos com antecedencia de forma que seja possível ver o código, gráficos, tabelas e outros resultados obtidos (sugestão: Github / Colab / Posit Cloud).
- Os exercícios são para ambas as turmas, exceto quando o contrário seja explicitado.

Exercícios

1. Para o caso de tabelas de contigência, já foram apresentadas as definições de RD, RR e OR. Mostre que:

   a. $Z \perp\!\!\!\perp Y$, $RD = 0$, $RR = 1$ e $OR = 1$ são afirmações equivalentes.
   b. Se todos os p_{zy} são positivos, então $RD > 0$ é equivalente a $RR > 1$ e também a $OR > 1$.
   c. $OR \approx RR$ se $Pr(Y = 1|Z = 1)$ e $Pr(Y = 1|Z = 0)$ são pequenos.

2. Seja

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{XY} & 1 & \rho YZ \\ \rho_{XZ} & \rho YZ & 1 \end{pmatrix}\right).$$

   a. [**MI628**] Mostre que $\rho_{YZ|X} = \dfrac{\rho_{YZ} - \rho_{XY}\rho_{ZX}}{\sqrt{1-\rho^2_{XY}} \times \sqrt{1-\rho^2_{ZX}}}$.
   b. De um exemplo numérico (simule um caso) em que $\rho_{YZ} > 0$ mas $\rho_{YZ|X} < 0$ [Paradoxo de Simpson para um vetor Normal trivariado].

3. O *dataset* `lalonde` do pacote `Matching` contém informação do tratamento (`treat`) [1: tratamento, 0: controle], do salario em 1978 (`re78`), bem como de outras 10 covariáveis (ou seja, existem $2^{10} = 1024$ possíveis subconjuntos de covariáveis). Ajuste as 1024 regressões com todos os possíveis subconjuntos de covariáveis e reporte o coeficiente associado ao tratamento.

   a. Quantas vezes o tratamento foi positivo e significativo?
   b. Quantas vezes o tratamento foi negativo e significativo?
   c. Quantas vezes o tratamento não foi significativo?

4. Assuma que os resultados potenciais são gerador por

$$Y(0) \sim N(0,1), \quad \tau = -0.5 + Y(0), \quad eY(1) = Y(0) + \tau.$$

   O mecanismo de atribuição de tratamento é de tal forma que $Z = 1$ se $\tau \geq 0$ e $Z = 0$ se $\tau < 0$. O resultado observado é dado por $Y = ZY(1) + (1 - Z)Y(0)$.

   a. Calcule $\mathbb{E}(Y|Z = 1) - \mathbb{E}(Y|Z = 0)$.
   b. Simule um caso (apenas uma replicação) para verificar o resultado obtido no item anterior.
   c. **Dica:** se $X \sim N(\mu, \sigma)$, $\mathbb{E}(X|a < X < b) = \mu - \sigma \dfrac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}$.

5. De um exemplo numérico (simule um caso) em que $\tau = n^{-1}\sum_{i=1}^{n}(Y_i(1) - Y_i(0)) > 0$ mas a proporção de vezes que $Y_i(1) > Y_i(0)$ é menor do que 0.5. Ou seja, o efeito causal médio é positivo mas o tratamento beneficia menos da metade das unidades em análise.