

MI628 / ME920 - Inferência Causal

– Lista 3 –

Carlos Trucíos

Instruções

- A resolução da lista será discutida no dia 09/04 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 antecedência de forma que seja possível ver o código, gráficos, tabelas e outros resultados obtidos (sugestão: Github / Colab / Posit Cloud).

Exercícios

1. [SRE] Assuma que $\tau_i = \tau \forall i = 1, \dots, n$. Considere a seguinte classe de estimadores para τ :

$$\hat{\tau}_\omega = \sum_{k=1}^K \omega_{[k]} \hat{\tau}_{[k]}.$$

- a. Quais as condições em $\omega_{[k]}$ s para que $\hat{\tau}_\omega$ seja não viesado para τ ?
 - b. Quais as condições em $\omega_{[k]}$ s para que $\hat{\tau}_\omega$ tenha variância mínima?
2. [SRE] Um experimento foi conduzido para estudar a eficiência de um medicamento para tratar a hiperplasia prostática benigna. Em cada centro, pacientes foram atribuídos aleatoriamente em três grupos: controle, finasterida 1mg e finasterida 5mg. O conjunto de dados `multicenter.csv` ([disponível aqui](#)) fornece, para cada centro, algumas estatísticas resumo para a pontuação dos sintomas (Y). Esta pontuação é a soma das respostas a nove questões (0–4) sobre diversos sintomas. Os resultados das colunas são:
 - `center`: ID dos centros,
 - `n0`, `n1`, `n5`: tamanho da amostra em cada um dos grupos (controle, finasterida 1mg e finasterida 5mg).
 - `mean0`, `mean1`, `mean5`: média dos resultados em cada um dos grupos (controle, finasterida 1mg e finasterida 5mg).

- `sd0`, `sd1`, `sd5`: desvio padrão dos resultados em cada um dos grupos (controle, finasterida 1mg e finasterida 5mg).

Estabeleça as hipóteses apropriadas e faça o teste correspondente, separadamente, para `finasterida 1mg vs control` e `finasterida 5mg vs control`. Considere cada centro como um estrato.

3. [SRE] Reanalize o *dataset lalonde* sob a perspectiva de Fisher e de Neyman.
 - Considere `race (black, hisp, other)` como estrato.
 - Considere `marital status (married)` como estrato.
 - Considere o `high school diploma (nodegr)` como estrato.
 - Discuta os resultados e compare com os obtidos sob CRE.
4. [ReM] Assuma que randomizamos $\mathbf{Z} = (Z_1, \dots, Z_n)$ segundo um CRE e aceitamos este se e somente se $\phi(\mathbf{Z}, \mathbf{X}) = 1$ (em que $\phi(\cdot)$ é um critério de balanceamento pre-estabelecido):
 - a. Mostre que se $n_1 = n_0$ e $\phi(\mathbf{Z}, \mathbf{X}) = \phi(1_n - \mathbf{Z}, \mathbf{X})$ então $\hat{\tau}$ é não viesado para τ .
 - b. Verifique que rerandomização utilizando a distância de Mahalanobis satisfaz $\phi(\mathbf{Z}, \mathbf{X}) = \phi(1_n - \mathbf{Z}, \mathbf{X})$ se $n_1 = n_0$.
 - c. Dê um contraexemplo que ilustre que $\hat{\tau}$ é viesado para τ quando as condições acima não são satisfeitas.

Obs: $\phi(\mathbf{Z}, \mathbf{X})$ pode ser um critério geral. No caso visto em aula, $\phi(\mathbf{Z}, \mathbf{X}) = I(M \leq a)$

5. [ReM] Mostre que

$$\hat{\tau}_F = \hat{\tau} - \hat{\gamma}'_F \hat{\tau}_X,$$

em que $\hat{\gamma}_F$ é o coeficiente de \mathbf{X} na regressão por MQO de Y sob $(1, \mathbf{Z}, \mathbf{X})$.

6. [ReM] Reanalize o *dataset Penn46_ascii.txt* ([disponível aqui](#)). O conjunto de dados já foi analisado [na aula 5](#) e consideramos `treatment` como indicador de tratamento (\mathbf{Z}), `log(duration)` como resultado observado e `quartes` como variável estratificadora. Agora, queremos incluir todas as outras covariáveis na análise.
 - a. Realize ajustes por regressão dentro dos estratos do experimento e, em seguida, combine esses estimadores ajustados para estimar o efeito causal médio.
 - b. Reporte o estimador pontual, o erro padrão estimado e o intervalo de confiança de 95%.
 - c. Compare os resultados obtidos com aqueles sem ajuste por regressão.