

Rerandomização

Felipe Marques

Introdução

Quando tratamos de dados desbalanceados (o que pode causar problemas nas análises), podemos tentar fazer pós-estratificação (covariáveis categóricas) ou rerandomização ou ajuste por covariável

Rerandomização

Distância de Mahalanobis

Usamos a distância para verificar desbalanceamento e como “filtro” no FRT.

$$M = \hat{\tau}'_X \left(\frac{n}{n_1 n_0} S_X^2 \right)^{-1} \hat{\tau}_X$$

Quanto maior for o M , mais indícios temos sob o desbalanceamento das covariáveis.

Resultado: sob CRE ou para $n \rightarrow \infty$, $M \sim \chi_K^2$.

Na rerandomização, escolhemos Z caso $M \leq a$.

Podemos escolher a , fixando $p_a = P(M \leq a)$. A escolha de p_a é feita considerando o *PRIV*.

$$PRIV = 100 \times \left(1 - \frac{P(\chi_{K+2}^2 \leq a)}{P(\chi_K^2 \leq a)} \right)$$

Análise sob rerandomização

Podemos utilizar FRT desde que simulemos Z sob a restrição que $M \leq a$

Ajuste de Regressão

FRT

Nesse caso X é fixo e sob H_0 , os resultados potenciais também são. Assim, construímos $T(Y, Z, X)$ utilizando uma das duas técnicas.

Pseudo-resultado

Construir $T(Y, Z, X)$ baseado nos resíduos $\hat{\epsilon}$ da regressão de Y em X . Utilizar os resíduos como pseudo-resultado para construir o teste. Rodamos a regressão uma vez.

Model-output

Fazer a regressão de Y em X, Z e utilizar $\hat{\beta}_Z$ como estatística do teste. Rodamos a regressão várias vezes.

Neyman

Rodamos a regressão de Y em $(1, Z, X, Z \times X)$ e pegar o coeficiente atrelado a Z como estimador para $\hat{\tau}_L$. O estimador de EHW é conservador para

Lin e SRE

Caso, além de covariáveis contínuas, existir uma variável estratificadora. Podemos obter um estimador de Lin e sua respectiva variância conservadora da seguinte forma:

$$\hat{\tau}_{L,S} = \sum_{k=1}^K \pi_{[k]} \hat{\tau}_{L,[k]}$$

$$Var(\hat{\tau}_{L,S}) = \hat{V}_{L,S} = \sum_{k=1}^K \pi_{[k]}^2 \hat{V}_{EHW,[k]}$$