Causal Statistics — Formulas

Completely Randomized Experiments (CRE)

Treatment Effect Estimator

$$\hat{\tau} = \frac{1}{N_1}\sum_{i=1}^{N} T_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)Y_i$$

Average Potential Outcomes

$$\overline{Z_N} = \frac{1}{N}\sum_{i=1}^{N} Z_i = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i^{(1)} - Y_i^{(0)}\right)$$

Variance Components

$$S_2^2 = S_1^2 + S_0^2 - 2S_{10}$$

$$S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i^{(1)} - \overline{Y^{(1)}}\right)^2$$

$$S_0^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i^{(0)} - \overline{Y^{(0)}}\right)^2$$

$$S_{10} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i^{(1)} - \overline{Y^{(1)}}\right)\left(Y_i^{(0)} - \overline{Y^{(0)}}\right)$$

Variance of Estimator

$$V(\hat{\tau}|T_i) = \frac{1}{N}\left\{\frac{N_0}{N_1}S_1^2 + \frac{N_1}{N_0}S_0^2 + 2S_{10}\right\}$$

$$\hat{V}(\hat{\tau}) = \frac{1}{N_1}\hat{S}_1^2 + \frac{1}{N_0}\hat{S}_0^2$$

Stratified Randomized Experiments (SRE)

Treatment Effect Estimator

$$\hat{\tau}_{pre} = \sum_{h=1}^{H} W_h \cdot \hat{\tau}_h, \quad W_h = \frac{N_h}{N}$$

$$\hat{\tau}_h = \frac{1}{N_{h1}}\sum_{i=1}^{N} I(X_i = h)T_i Y_i - \frac{1}{N_{h0}}\sum_{i=1}^{N} I(X_i = h)(1-T_i)Y_i$$

Variance

$$V(\hat{\tau}_{pre}) = \sum_{h=1}^{H} W_h^2 V(\hat{\tau}_h)$$

$$V(\hat{\tau}_h) = \frac{1}{N_{h1}}S_{h1}^2 + \frac{1}{N_{h0}}S_{h0}^2 - \frac{1}{N_h}S_{h,\tau}^2$$

Post-Stratification in CRE

Estimator

$$\hat{\tau}_{post} = \sum_{h=1}^{H} W_h \hat{\tau}_h$$

$$\hat{\tau}_h = \frac{1}{N_{h1}}\sum_{i=1}^{N} T_i Y_i - \frac{1}{N_{h0}}\sum_{i=1}^{N}(1-T_i)Y_i$$

Variance

$$V_{cre}(\hat{\tau}_{post}) \approx \frac{1}{N}\left\{\frac{N_0}{N_1}\sum_{h=1}^{H} W_h S_{h1}^2 + \frac{N_1}{N_0}\sum_{h=1}^{H} W_h S_{h0}^2 + 2\sum_{h=1}^{H} W_h S_{h,10}\right\}$$

Regression under CRE

Estimator

$$\hat{\tau}_{reg} = \frac{1}{N}\sum_{i=1}^{N} X_i'\hat{\beta}_1 - \frac{1}{N}\sum_{i=1}^{N} X_i'\hat{\beta}_0$$

$$\hat{\beta}_1 = \left(\sum_{i=1}^{N} T_i X_i X_i'\right)^{-1}\sum_{i=1}^{N} T_i X_i Y_i$$

$$\hat{\beta}_0 = \left(\sum_{i=1}^{N}(1-T_i)X_i X_i'\right)^{-1}\sum_{i=1}^{N}(1-T_i)X_i Y_i$$

Variance

$$V(\hat{\tau}_{reg,t}) = V(\hat{\tau}) + E\left[\left(\frac{1}{N_1} - \frac{1}{N}\right)^2 S_{e1}^2 + \left(\frac{1}{N_0} - \frac{1}{N}\right)^2 S_{e0}^2 + \frac{2}{N}S_{e0,e1}\right]$$

Inverse Probability Weighting (IPW)

Propensity Score

$$\pi(X_i) = P(T_i = 1|X_i)$$

IPW Estimators

- **Horvitz-Thompson (HT):**

$$\hat{\tau}_{\mathrm{HT}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{T_i Y_i}{\pi(X_i)} - \frac{(1-T_i)Y_i}{1-\pi(X_i)}\right)$$

- **Hájek**

$$\hat{\tau}_{\mathrm{Hajek}} = \frac{\sum_{i=1}^{N}\frac{T_i Y_i}{\hat{\pi}(X_i)}}{\sum_{i=1}^{N}\frac{T_i}{\hat{\pi}(X_i)}} - \frac{\sum_{i=1}^{N}\frac{(1-T_i)Y_i}{1-\hat{\pi}(X_i)}}{\sum_{i=1}^{N}\frac{1-T_i}{1-\hat{\pi}(X_i)}}$$

Augmented IPW (AIPW)

Oracle AIPW

If $\pi(X)$ and $Q(X,t) = E(Y|X,T=t)$ are known:

$$\hat{\tau}^*_{\text{AIPW}} = \frac{1}{N}\sum_{i=1}^{N}\left[Q(X_i,1) - Q(X_i,0) + \frac{T_i(Y_i - Q(X_i,1))}{\pi(X_i)} - \frac{(1-T_i)(Y_i - Q(X_i,0))}{1-\pi(X_i)}\right]$$

Estimated AIPW

Using estimated $\hat{\pi}(X)$ and $\hat{Q}(X,t)$:

$$\hat{\tau}_{\text{AIPW}} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{Q}(X_i,1) - \hat{Q}(X_i,0) + \frac{T_i(Y_i - \hat{Q}(X_i,1))}{\hat{\pi}(X_i)} - \frac{(1-T_i)(Y_i - \hat{Q}(X_i,0))}{1-\hat{\pi}(X_i)}\right]$$