Stratified Randomized Experiments

Park Seok Hoon

July 19, 2022

Seoul National University

Table of Contents

Table of Contents

The structure of stratified randomized experiments

- Units are grouped together according to some pre-treatment characteristics into strata.
- Within each stratum, a completely randomized experiment is conducted.
- So, the methods discussed in Chapters 5-8 are directly applicable.
- This chapter is about hypotheses and treatment effects across all strata.

Notation

- Suppose that the stratification is "f" and "m"
- N(f) : number of "f" , N(m) : number of "m"
- $N_c(f)$: the number of control unit in "f"
- $N_c(m)$: the number of control unit in "m"
- $W_i = 0$: $unit_i$ control treatment , $W_i = 1$: $unit_i$ active treatment

Notation

- $r_{fs}(f)$: finite-sample average treatment effect in "f"
- $r_{fs}(m)$: finite-sample average treatment effect in "m"

The assignment distribution

- $Pr(\mathcal{W}|Y(0), Y(1), S) = \binom{N(f)}{N_t(f)}^{-1}\binom{N(m)}{N_t(m)}^{-1}$ for $\mathcal{W} \in W^+$
- $W^+ = \{\mathcal{W}|\sum_{i:G_i=f} W_i = N_t(f), \sum_{i:G_i=m} W_i = N_t(m)\}$
- $\mathcal{W} = [W_1, ..., W_n]^t$

Table of Contents

Fisher's Exact P-values

- In stratified randomized experiments, just as in completely randomized experiments, the assignment mechanism is completely known.

- Hence, we can directly apply Fisher's approach in chapter 5

example)

$$\left|\frac{N(f)}{N(f)+N(m)}(\bar{Y}_t^{obs}(f) - \bar{Y}_c^{obs}(f)) + \frac{N(m)}{N(f)+N(m)}(\bar{Y}_t^{obs}(m) - \bar{Y}_c^{obs}(m))\right|$$

Table of Contents

Neyman's repeated sampling perspective

- In this chapter, we are interested in a weighted average of the two within-stratum average effects in the finite sample.

- Suppose the stratification is "f" and "m", then the average treatment effect is

$$r_{fs} = \frac{N(f)}{N(f)+N(m)} r_{fs}(f) + \frac{N(m)}{N(f)+N(m)} r_{fs}(m)$$

Neyman's repeated sampling perspective

- $\hat{r}(f) = \bar{Y}_t^{obs}(f) - \bar{Y}_c^{obs}(f)$
- The estimator is $\hat{r} = \frac{N(f)}{N(f)+N(m)}\hat{r}(f) + \frac{N(m)}{N(f)+N(m)}\hat{r}(m)$
- Similar to chapter 6, we can get a confidence interval

Table of Contents

Regression analysis

- In ordet to interpret regression-based estimators, we take a super-population perspective with a fixed number of strata, and an infinite number of units within each stratum.
- Model : $Y_i^{obs} = rW_i + \sum_{j=1}^{J} \beta_j B_i(j) + \epsilon_i$
- stratification = $\{1, ..., J\}$
- $B_i(j) = I(unit_i \in stratum\ J)$
- $\hat{\tau}^{ols}, \hat{\beta}^{ols} = argmin_{r,\beta} \sum_{i=1}^{N} (Y_i^{obs} - rW_i - \sum_{j=1}^{J} \beta(j) B_i(j))^2$
- $\tau^*, \beta^* = argmin_{r,\beta} E((Y_i^{obs} - rW_i - \sum_{j=1}^{J} \beta(j) B_i(j))^2)$

Regression analysis

- $q(j) = \frac{N(j)}{N}$: proportion of each stratum in the sample from the infinite super-population
- $e(j) = \frac{N_t(j)}{N(j)}$: proportion of treated units in each stratum
- $w(j) = q(j)e(j)(1 - e(j)), \tau_{sp}(j) = E(Y_i(1) - Y_i(0)|B_i(j) = 1)$
- $\tau_w = \sum_j^J w(j)\tau_{sp}(j) / \sum_{j=1}^J w(j)$

Regression analysis

Theorem 9.1 *Suppose we conduct a stratified randomized experiment in a sample drawn at random from an infinite population. Then, for estimands τ^* and τ_w defined in (9.3) and (9.4), the estimator $\hat{\tau}^{\text{ols}}$ satisfies, (i)*

$$\tau^* = \tau_\omega,$$

and (ii),

$$\sqrt{N} \cdot \left(\hat{\tau}^{\text{ols}} - \tau_\omega\right) \xrightarrow{d}$$

$$\mathcal{N}\left(0, \frac{\mathbb{E}\left[\left(W_i - \sum_{j=1}^J q(j) \cdot B_i(j)\right)^2 \cdot \left(Y_i^{\text{obs}} - \tau^* \cdot W_i - \sum_{j=1}^J \beta_j^* \cdot B_i(j)\right)^2\right]}{\left(\sum_{j=1}^J q(j) \cdot e(j) \cdot (1 - e(j))\right)^2}\right).$$

Table of Contents

Model-based analysis

1. As in the analysis of completely randomized experiments, we combine the specifcation of the joint distribution of the potential outcomes with the known distribution of the vector of assignment indicators to derive the posterior distribution of the causal estimand
2. In other words, we can use bayesian model in chapter 8

Model Example

- With few strata and a substantial number of units per stratum, we may wish to use a prior distribution that makes all elements of θ a priori independent.

- $\binom{Y_i(0)}{Y_i(1)} | B_i(j), \theta \sim N(\binom{\mu_c(j)}{\mu_t(j)}, diag(\sigma_c^2(j), \sigma_t^2(j))), j = 1, ..., J$

- $\mu_c(j), \mu_t(j) \sim^{iid} Normal \ , \sigma_c^2, \sigma_t^2 \sim^{iid} inverse - \chi^2$