4(b). Extensions of Completely Randomized Experiments

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Regression in Completely Randomized Experiment

- Regression with no covariates = standard Neyman analysis.
- Regression with (uninteracted) covariates:
  - Consistent for SATE/PATE.
  - Usually will help with precision, but can hurt.
- Regression with interacted covariates:
  - Consistent for SATE/PATE.
  - Asymptotically will never hurt precision.
- Always use robust/HC2 variance estimators unless you have good reasons.

Regression for Stratified Experiments

- Setup: block randomized experiment with block indicators W_{ij}.
  - Block "fixed effects" $W_{ij} = 1$ if i is in block j, 0 otherwise.
  - Blocks $j \in \{1, \ldots, J\}$ with sizes $w_j = n_j/n$ and propensity scores $p_j = n_{1,j}/n_j$.
- Recall STAR Project: within each school (block), classes were randomized.
- Naive approach: just include the block FEs in OLS

$$(\widehat{\tau}_{b,\text{fe}}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_J) = \underset{(\tau,\alpha_1,\ldots,\alpha_J)}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \tau D_i - \sum_{j=1}^{J} \alpha_j W_{ij} \right)^2$$

- $\widehat{\tau}_{b,\text{fe}}$ **not consistent** for the PATE unless:

$$\widehat{\tau}_{b,\text{fe}} \xrightarrow{p} \frac{\sum_{j=1}^{J} \omega_j \tau_j}{\sum_{j=1}^{J} \omega_j}, \quad \text{where} \quad \omega_j = w_j p_j (1 - p_j)$$

  - Propensity scores are equal across blocks: $p_j = p$ for all j.
  - ATEs are equal across strata $\tau_j = \tau$ for all j.

Block Randomized Trials: Correct Analysis

1. Just use original Neyman analysis aggregating within-strata analyses.
2. Weight OLS by inverse of the propensity score: $1/p_j$.
3. Fully interact block FEs with treatment.

- Latter two allow for additional covariates to be added.
- Check Imbens and Rubin (2015) Ch.9.6.1., second model
- See this simulation study using `DeclareDesign`: https://declaredesign.org/blog/posts/biased-fixed-effects.html

Block Randomized Trials: Correct Analysis

2. Weight OLS by inverse of the propensity score.

$$(\widehat{\tau}_{b,\text{fe}}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_J) = \underset{(\tau, \alpha_1, \ldots, \alpha_J)}{\arg\min} \sum_{i=1}^{n} s_i \left(Y_i - \tau D_i - \sum_{j=1}^{J} \alpha_j W_{ij} \right)^2$$

$$\text{where} \quad s_i = \sum_{j=1}^{J} \left\{ \left(\frac{1}{p_j} \right) D_i + \left(\frac{1}{1 - p_j} \right) (1 - D_i) \right\} W_{ij} \quad \text{and} \quad p_j = \frac{n_{1,j}}{n_j}.$$

In R

```
your_formula <- as.formula("outcome ~ treat + x_tilde1 + x_tilde2")

your_data <- data.frame(outcome, treat,
                        x_tilde1, x_tilde2,
                        weights, block)

your_fitted_model <- estimatr::lm_robust(your_formula, data = your_data,
                                         weights = weights, # s
                                         se_type = "HC2",
                                         fixed_effects = block)
```

Cluster Randomized Trials

- Treatment often allocated at a higher level than the data.
  - Suppose schools are randomized and all the classes in the same school receive the same treatment.
  - Now school is not a block, but a cluster!
  - More examples:
    - States are treated, but we have firm-level data.
    - Platforms are treated, but we have user-level data.
- Setup:
  - Clusters: $k \in \{1, \ldots, K\}$
  - Randomly choose K_1 treatment clusters, K_0 control.
  - Each cluster has units $i \in \{1, \ldots, m_k\}$ with $\sum_{k=1}^{K} m_k = n$
  - Treatment assignment at cluster level: $D_{ik} = D_k$
  - Potential outcomes $Y_{ik}(d)$
- Cost of clustering
  - More similarity $\rightsquigarrow$ each unit provides redundant information $\rightsquigarrow$ less efficiency under clustering

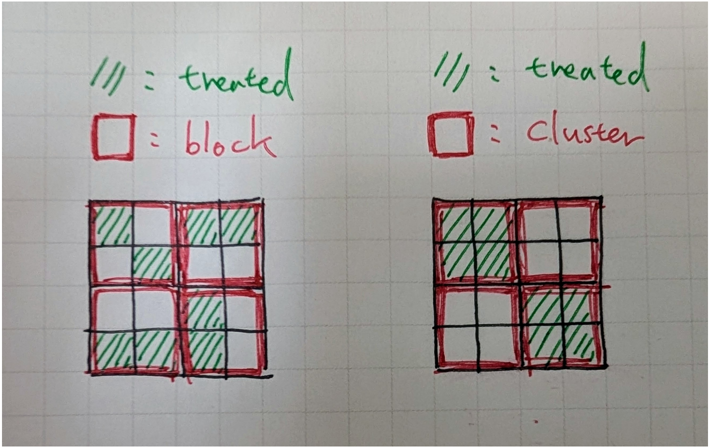
Cluster Randomized Trials: Analysis

Use **cluster-robust variance estimator**:

In R

```
your_formula <- as.formula("outcome ~ treat + x_tilde1 + x_tilde2")

your_data <- data.frame(outcome, treat,
                        x_tilde1, x_tilde2,
                        cluster)

your_fitted_model <- estimatr::lm_robust(your_formula, data = your_data,
                                         clusters = cluster,
                                         se_type = "CR2")

??estimatr::lm_robust # Check more options for se_type

# Or
your_model <- lm(your_formula, data = your_data)
your_vcov <- clubSandwich::vcovCR(your_model, cluster = your_data$cluster,
                                  type = "CR2")
```

- **Cluster at the treatment assignment level** (no higher or lower)!
- Vanilla CR variance estimator is biased, Bell & McCaffrey proposed CR2 adjustment similar to HC2 (usually preferable).
- You may have block and cluster design at the same time.


/// : treated
□ : block
/// : treated
□ : cluster

Have a great weekend!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/