Blocking for Improved Efficiency

- How can we improve the **efficiency** of causal effect estimation while maintaining the unbiasedness?
- Minimize variances of potential outcomes
  \[ \Rightarrow \text{conduct a randomized experiment in a group of similar units} \]

"**Block what you can and randomize what you cannot**"


**Basic procedure:**
1. **Blocking (Stratification):** create groups of similar units based on pre-treatment covariates
2. **Block (Stratified) randomization:** completely randomize treatment assignment within each group
Stratified Design

- **Setup:**
  - Number of units, \( n \)
  - Number of blocks, \( J \)
  - Block size, \( n_j > 2 \)
  - Number of treated in each block, \( n_{1j} > 1 \)
  - Complete randomization within each block, \( \Pr(T_{ij} = 1) = n_{1j}/n_j \)

- **Analysis:**
  1. Apply Neyman's analysis to each block
     
     \[
     \hat{\tau}_j = \frac{1}{n_{1j}} \sum_{i=1}^{n_j} T_{ij} Y_{ij} - \frac{1}{n_{0j}} \sum_{i=1}^{n_j} (1 - T_{ij}) Y_{ij}, \quad \hat{\sigma}^2_j = \frac{\hat{\sigma}^2_{1j}}{n_{1j}} + \frac{\hat{\sigma}^2_{0j}}{n_{0j}}
     \]

  2. Aggregate block-specific estimates and variances
     
     \[
     \hat{\tau}_{\text{block}} = \sum_{j=1}^{J} w_j \cdot \hat{\tau}_j, \quad \text{and} \quad \hat{\sigma}^2_{\text{block}}(\hat{\tau}_{\text{block}}) = \sum_{j=1}^{J} w_j^2 \cdot \hat{\sigma}^2_j
     \]

     where \( w_j \) is the weight for the \( j \)th block, e.g., \( w_j = n_j/n \)
Efficiency Gain due to Blocking

- Simple analytic framework:
  - PATE as the estimand
  - $J$ pre-defined blocks within an infinite population
  - Stratified random sampling of $w_j \cdot n$ units within each block
  - Complete randomization of treatment assignment within each block
  - Identical treatment assignment probability across blocks: $k = n_1/n = n_{1j}/n_j$ for all $j$

- Key equality:

\[
\mathbb{V}(\hat{X}) = \mathbb{E}\{\mathbb{V}(X | Y)\} + \mathbb{V}\{\mathbb{E}(X | Y)\}
\]

- Difference in variance:

\[
\mathbb{V}(\hat{\tau}) - \mathbb{V}_{\text{block}}(\hat{\tau}_{\text{block}}) = \frac{1}{n} \left\{ \frac{\sigma_1^2}{k} + \frac{\sigma_0^2}{1 - k} - \sum_{j=1}^{J} w_j \left( \frac{\sigma_{1j}^2}{k} + \frac{\sigma_{0j}^2}{1 - k} \right) \right\} \geq 0
\]
Randomization was done within each school \(\leadsto\) stratification!

Effect of kindergarten class size on high school graduation:

1. **Permutation tests**
   - Fisher’s exact test: \( p\)-value = 0.51
   - Mantel-Haenszel test: \( p\)-value = 0.37

2. **Average treatment effect estimation**
   - \( \text{est.} = 0.018, \text{se.} = 0.017 \)
   - s.e. without stratification \(\approx\) 7% greater

Effect of kindergarten class size on 8th grade reading score

1. **Permutation tests**
   - Wilcoxon’s test: \( p\)-value = 0.121
   - Aligned rank sum test: \( p\)-value = 0.067

2. **Average treatment effect estimation**
   - \( \text{est.} = 2.76, \text{se.} = 1.73 \)
   - s.e. without stratification \(\approx\) 9% greater
Matched-Pairs Design

- Should we keep blocking until we cannot block any further?

**Procedure:**
1. Create $J = n/2$ pairs of similar units
2. Randomize treatment assignment within each pair
   - $W_j = 1$ if the first unit receives the treatment
   - $W_j = -1$ if the second unit receives the treatment

**Analysis:**

$$\hat{\tau}_{\text{pair}} = \frac{1}{J} \sum_{j=1}^{J} W_j (Y_{1j} - Y_{2j}),$$

$$\text{Var}(\hat{\tau}_{\text{pair}}) = \frac{1}{J(J-1)} \sum_{j=1}^{J} \{W_j (Y_{1j} - Y_{2j}) - \hat{\tau}_{\text{pair}}\}^2$$
Efficiency Analysis

- Neyman’s stratified variance estimator is not applicable
- For SATE, $\hat{\text{V}}(\hat{\tau}_{\text{pair}})$ is conservative unless the average treatment effect is constant across pairs (Imai. 2008. Stat. Med.)

- For PATE, simple random sampling of pairs instead of stratified random sampling within pre-defined strata

$$\mathbb{E}(\hat{\text{V}}(\hat{\tau}_{\text{pair}})) = \frac{\sigma^2_1}{J} + \frac{\sigma^2_0}{J} - 2 \times \text{Cov}(Y_{1j}(1), Y_{2j}(0))$$

- Improved inference under stratified random sampling:
  - group similar pairs (Imbens and Rubin. Chapter 10)

- 50 million uninsured Mexicans $\rightsquigarrow$ catastrophic medical expenditure among poor households
- Seguro popular: delivery of health insurance, regular and preventive healthcare, medicines and health facilities
- Units: health clusters = predefined health facility catchment areas
- Randomization within 74 matched pairs of “similar” health clusters
- 10 months followup survey for 50 pairs
- Outcome: proportion of households within each health cluster who experienced catastrophic medical expenditure
  - est. $= -0.013$, s.e. $= 0.007$
  - within-pair correlation: $\text{corr}(Y_{1j}(1), Y_{2j}(0)) = 0.482$
  - estimated s.e. under complete randomization $= 0.010$
- Wilcoxon’s signed rank test:
  - $p$-value $= 0.10$, 95% conf. int. $= [-0.026, 0.002]$
Blocking in Practice

- Univariate blocking: discrete or discretized variable
- Multivariate blocking: Mahalanobis distance

\[ D(X_i, X_j) = \sqrt{(X_i - X_j)^\top \hat{V}(X)^{-1} (X_i - X_j)} \]

- Greedy algorithms
  1. Matching: pair two units with the shortest distance, set them aside, and repeat
  2. Blocking: randomly choose one unit and choose \( n_j \) units with the shortest distances, set them aside, and repeat

- But the resulting matches may not be optimal
Optimal Matching

- **D**: $n \times n$ matrix of pairwise distance or a cost matrix

Select $n$ elements of $D$ such that there is only one element in each row and one element in each column and the sum of pairwise distances is minimized

- **Linear Sum Assignment Problem (LSAP)**
  - Binary $n \times n$ matching matrix: $M$ with $M_{ij} \in \{0, 1\}$
  - Optimization problem

$$\minimize_{M} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} D_{ij} \quad \text{subject to} \quad \sum_{i=1}^{n} M_{ij} = 1, \sum_{j=1}^{n} M_{ij} = 1$$

where we set $D_{ii} = \infty$ for all $i$

- can apply the Hungarian algorithm

- To be truly optimal, we must consider variable strata size
Optimal Matching for Seguro Popular

- 4 pre-treatment cluster-average covariates: age, education, household size, household assets
- 100 clusters, 50 pairs
- Minimize the sum of pairwise Mahalanobis distance
Other Strategies to Increase Efficiency

  - Blocking after the experiment is conducted
  - Number of treated units is a random variable
  - Variance is conditional on “valid” randomization draws
  - Post-stratification is nearly as efficient as pre-randomization blocking except with a large number of small strata
  - Risk of p-hacking $\Rightarrow$ need for pre-registration

  - Procedure:
    1. Specify the acceptable level of covariate balance
    2. Randomize the treatment and check covariate balance
    3. Repeat until the covariate balance criterion is met
  - permutation inference
  - asymptotic inference (Li *et al.* 2018. *PNAS*)
Adaptive Designs

What should we do if units arrive sequentially?

Biased coin design (Efron. 1971. *Biometrika*)

1. For the first $2m$ units, we use the Bernoulli design
2. For a new unit, assign it to the treatment group with

\[
\begin{align*}
\text{probability } p & \quad \text{if more units are in treatment group} \\
\text{probability } 1/2 & \quad \text{if treatment and control groups have same size} \\
\text{probability } q & \quad \text{if more units are in control group}
\end{align*}
\]

- Efron suggests $p = 1/3$ and $q = 2/3$
- Used for small trials with unknown sample size
- Possible to make it adaptive using covariates and/or outcomes

Thompson sampling (Thompson. 1933. *Biometrika*)

- Consider $K$ treatments $\sim$ multi-armed bandit
- Binary outcome here but can be extended to other settings

1. Sample $\theta_k$ from Beta($\alpha_k, \beta_k$) for $k = 1, 2, \ldots, K$
2. Choose $T_t \leftarrow \arg\max_k \theta_k$ and observe a binary outcome $Y_t$
3. Bayesian update $(\alpha_k, \beta_k) \leftarrow (\alpha_{T_t} + Y_t, \beta_{T_t} + 1 - Y_t)$
Summary

- Blocking improves efficiency of inference with randomized experiments while preserving the advantages of randomization.

- Neyman’s randomization inference allows for the efficiency analysis:
  - Stratified designs
  - Matched-pair designs

- Optimal matching algorithm → stratification of variable size?

- Other designs:
  - Post-stratification and rerandomization
  - Adaptive designs

- Reading: IMBENS AND RUBIN, Chapters 9 and 10