# 150C Causal Inference

## Randomized Experiments 3

Jonathan Mummolo

# Main Randomization Designs

1. Complete randomization
   - Of $N$ units, $m$ are randomly assigned to treatment and $N - m$ to control
2. Block randomization
   - $N$ units are partitioned into $J$ subgroups (called strata or blocks) and complete randomization occurs **within each block**
3. Cluster randomization
   - Individual units are nested in clusters, and complete randomization occurs **at the level of the clusters**

# Outline

1. Block randomization
2. Cluster Randomization

# Block Randomization

- Imagine you have data on the units that you are about to randomly assign. Why leave it to "pure" chance to balance the observed characteristics?
- In blocking we pre-stratify the sample into subgroups, and then randomize separately within each subgroup
- Two main motivations for blocking:
  1. Blocking reduces sampling uncertainty and improves precision
     - Enforces balance on blocking factors by design and rules out randomizations that may produce outlandish results
  2. Blocking ensures that certain subgroups are available for analysis
     - Makes it clear a priori which subsets will be analyzed

| store | block | All Stores $Y_1$ | All Stores $Y_0$ | Block A $Y_1$ | Block A $Y_0$ | Block B $Y_1$ | Block B $Y_0$ |
|---|---|---|---|---|---|---|---|
| 1 | A | 4 | 1 | 4 | 1 | | |
| 2 | A | 5 | 3 | 5 | 3 | | |
| 3 | A | 6 | 2 | 6 | 2 | | |
| 4 | A | 4 | 1 | 4 | 1 | | |
| 5 | A | 3 | 2 | 3 | 2 | | |
| 6 | A | 2 | 1 | 2 | 1 | | |
| 7 | B | 10 | 8 | | | 10 | 8 |
| 8 | B | 13 | 10 | | | 13 | 10 |
| 9 | B | 11 | 9 | | | 11 | 9 |
| 10 | B | 10 | 8 | | | 10 | 8 |
| 11 | B | 12 | 10 | | | 12 | 10 |
| 12 | B | 11 | 11 | | | 11 | 11 |
| Mean | | 7.6 | 5.5 | 4.0 | 1.7 | 11.2 | 9.3 |
| Variance | | 14.2 | 15.6 | 1.7 | 0.6 | 1.1 | 1.2 |

Under complete randomization with 6 treated stores we get:

$$\begin{aligned} ATE_{CR} &= E[Y_1] - E[Y_0] = 7.6 - 5.5 \approx 2.08 \\ SE_{\widehat{ATE}} &= \sqrt{\frac{1}{N-1}\left(\frac{m}{N-m}Var[Y_0] + \frac{N-m}{m}Var[Y_1] + 2Cov[Y_1, Y_0]\right)} \\ &= \sqrt{\frac{1}{12-1}\left(\frac{6}{12-6}15.6 + \frac{12-6}{6}14.2 + 2 * 14.3\right)} \approx 2.30 \end{aligned}$$

# Effect and Standard Error Under Block Randomization

## ATE under Block Randomization

Given $N$ units that are partitioned into $J$ blocks of size $N_j$, the average treatment effect if given by

$$ATE_{BR} = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$$

## Standard error for ATE under Block Randomization

The true standard error of the estimated ATE is given by

$$SE_{\widehat{ATE_{BR}}} = \sqrt{\sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 SE^2_{\widehat{ATE}_j}}$$

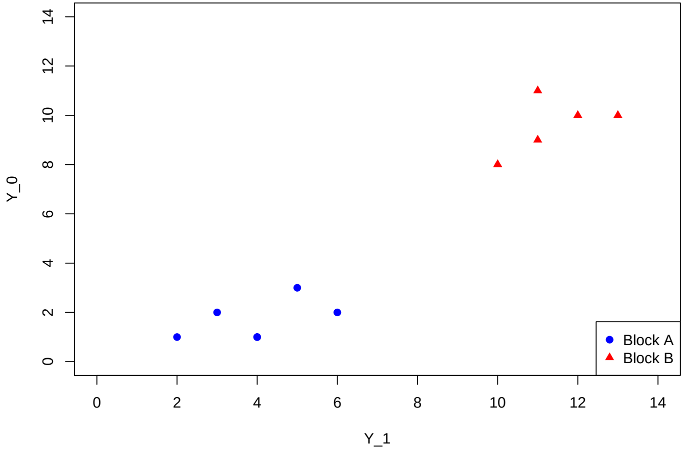
| | | All Stores | | Block A | | Block B | |
|---|---|---|---|---|---|---|---|
| store | block | $Y_1$ | $Y_0$ | $Y_1$ | $Y_0$ | $Y_1$ | $Y_0$ |
| 1 | A | 4 | 1 | 4 | 1 | | |
| 2 | A | 5 | 3 | 5 | 3 | | |
| 3 | A | 6 | 2 | 6 | 2 | | |
| 4 | A | 4 | 1 | 4 | 1 | | |
| 5 | A | 3 | 2 | 3 | 2 | | |
| 6 | A | 2 | 1 | 2 | 1 | | |
| 7 | B | 10 | 8 | | | 10 | 8 |
| 8 | B | 13 | 10 | | | 13 | 10 |
| 9 | B | 11 | 9 | | | 11 | 9 |
| 10 | B | 10 | 8 | | | 10 | 8 |
| 11 | B | 12 | 10 | | | 12 | 10 |
| 12 | B | 11 | 11 | | | 11 | 11 |
| Mean | | 7.6 | 5.5 | 4.0 | 1.7 | 11.2 | 9.3 |
| Variance | | 14.2 | 15.6 | 1.7 | 0.6 | 1.1 | 1.2 |

Under block randomization with 3 treated stores per block we get:

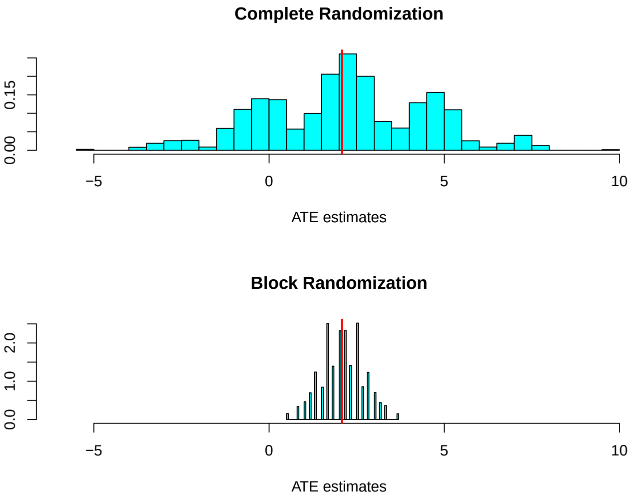
$$ATE_{BR} = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j = \frac{6}{12} 2.33 + \frac{6}{12} 1.83 \approx 2.08$$

$$SE_{\widehat{ATE_{BR}}} = \sqrt{\sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 SE^2_{\widehat{ATE}_j}} = \sqrt{\left(\frac{6}{12}\right)^2 (0.80)^2 + \left(\frac{6}{12}\right)^2 (0.88)^2} \approx 0.84$$

# Block Randomization



# Sampling Distributions: Complete vs. Blocked



# Estimation: "As ye randomize, so shall ye analyze"

## Difference-in-Means Estimator for Block Randomization

Given $J$ blocks of size $N_j$ with $m_j$ treated units the ATE can be estimated using

$$\widehat{ATE}_{BR} = \sum_{j=1}^{J} \frac{N_j}{N} \widehat{ATE}_j = \sum_{j=1}^{J} \frac{N_j}{N} (\bar{Y}_{1j} - \bar{Y}_{0j})$$

with block specific means $\bar{Y}_{dj} = \frac{1}{N_{dj}} \sum_{D_{ij}=d} Y_{ij}$

## Estimator for Standard Error of ATE under Block Randomization

$$\widehat{SE}_{\widehat{ATE}_{BR}} = \sqrt{\sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 \widehat{SE}^2_{\widehat{ATE}_j}} = \sqrt{\sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 \left(\frac{\widehat{\sigma}^2_{Y_{ij}|D_{ij}=1}}{m_j} + \frac{\widehat{\sigma}^2_{Y_{ij}|D_{ij}=0}}{N_j - m_j}\right)}$$

# Estimation in Blocked Design

## Regression Estimator for Block Randomization

Given $J$ blocks of size $N_j$ with $m_j$ treated units, the ATE can be estimated using the following regression

$$Y_i = \beta_1 + \alpha_{BR} D_i + \beta_2 J_2 + \beta_3 J_3 + ... + \beta_J J_J + u_i$$

where $J_2, J_3, ..., J_J$ are dummy variables that indicate each block.

- This regression estimator is valid if the treatment probability $p_j = m_j/N_j$ is the same in all blocks.
  - Regression weights each block specific $\widehat{ATE}_j$ by $(N_j/N)p_j(1-p_j)$
- If $p_j$ varies across blocks, regression can lead to bias since treatment assignment is correlated with block characteristics
- Need to use weighted regression with unit weights

$$w_{ij} \equiv \left(\frac{1}{p_{ij}}\right) D_i + \left(\frac{1}{1-p_{ij}}\right)(1-D_i)$$

# Blocking

- What to block on?
    - "Block what you can, randomize what you can't"
    - The baseline of the outcome variable and other main predictors
    - Variables desired for subgroup analysis
- How to block?
    - Stratification
    - Pair-matching
    - Check: `blockTools` library

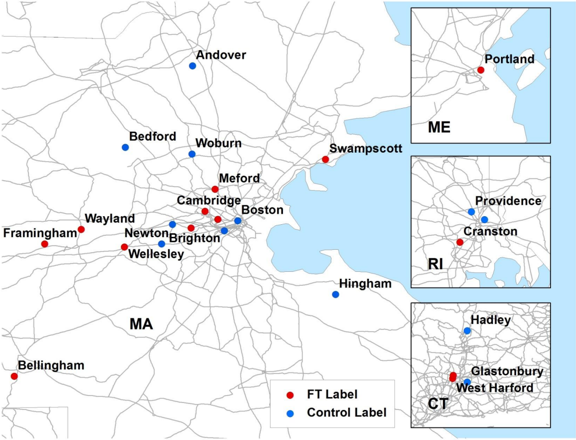
# Label Experiment

Treatment

Control

# Example: Fair Trade Labeling Experiment

# Matched Pairs: Phase 1



# Example: Fair Trade Labeling Experiment

R Code

```
> d <- read.dta("FTdata.dta")
> head(d)
  store pair FTweek lnsalesd
1     1    1      1     3.20
2     4    1      0     2.77
3     6    2      1     4.18
4     9    2      0     4.04
5    21    3      1     4.30
6    24    3      0     3.93
```

# Example: Fair Trade Labeling Experiment

R Code

```
> cr.out <- lm(lnsalesd~FTweek,data=d)
> coeftest(cr.out,vcov = vcovHC(cr.out, type = "HC1"))

t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.35000    0.16079 27.0537   <2e-16 ***
FTweek       0.12385    0.21424  0.5781   0.5686
---
```

# Example: Fair Trade Labeling Experiment

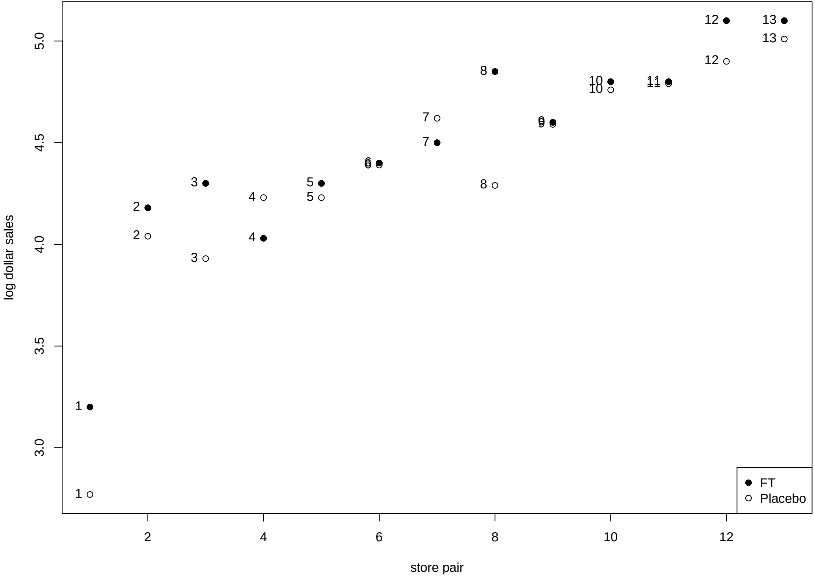
R Code

```
> br.out <- lm(lnsalesd~FTweek+as.factor(pair),data=d)
> coeftest(br.out,vcov = vcovHC(br.out, type = "HC1"))

t test of coefficients:

                  Estimate Std. Error t value  Pr(>|t|)
(Intercept)       2.923077   0.162144 18.0277 4.671e-10 ***
FTweek            0.123846   0.060176  2.0581 0.0619840 .
as.factor(pair)2  1.125000   0.159549  7.0511 1.335e-05 ***
as.factor(pair)3  1.130000   0.204440  5.5273 0.0001304 ***
as.factor(pair)4  1.145000   0.231925  4.9369 0.0003439 ***
as.factor(pair)5  1.280000   0.161773  7.9123 4.208e-06 ***
as.factor(pair)6  1.410000   0.169987  8.2948 2.591e-06 ***
as.factor(pair)7  1.575000   0.203689  7.7324 5.317e-06 ***
as.factor(pair)8  1.585000   0.277319  5.7154 9.675e-05 ***
as.factor(pair)9  1.610000   0.169987  9.4713 6.420e-07 ***
as.factor(pair)10 1.795000   0.165195 10.8660 1.450e-07 ***
as.factor(pair)11 1.810000   0.169987 10.6479 1.810e-07 ***
as.factor(pair)12 2.015000   0.164183 12.2729 3.763e-08 ***
as.factor(pair)13 2.070000   0.160298 12.9134 2.127e-08 ***
---
```

# Example: Fair Trade Labeling Experiment



# Example: Fair Trade Labeling Experiment

R Code

```
> summary(lm(lnsalesd~as.factor(pair),data=d))
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.9850     0.1212  24.621 2.72e-12 ***
as.factor(pair)2    1.1250     0.1715   6.562 1.82e-05 ***
as.factor(pair)3    1.1300     0.1715   6.591 1.74e-05 ***
as.factor(pair)4    1.1450     0.1715   6.678 1.52e-05 ***
as.factor(pair)5    1.2800     0.1715   7.466 4.73e-06 ***
as.factor(pair)6    1.4100     0.1715   8.224 1.65e-06 ***
as.factor(pair)7    1.5750     0.1715   9.186 4.77e-07 ***
as.factor(pair)8    1.5850     0.1715   9.245 4.44e-07 ***
as.factor(pair)9    1.6100     0.1715   9.390 3.71e-07 ***
as.factor(pair)10   1.7950     0.1715  10.469 1.05e-07 ***
as.factor(pair)11   1.8100     0.1715  10.557 9.56e-08 ***
as.factor(pair)12   2.0150     0.1715  11.752 2.68e-08 ***
as.factor(pair)13   2.0700     0.1715  12.073 1.94e-08 ***
---
Residual standard error: 0.1715 on 13 degrees of freedom
Multiple R-squared:  0.9474, Adjusted R-squared:  0.8988
F-statistic:  19.5 on 12 and 13 DF,  p-value: 2.356e-06
```

# Outline

1. Block randomization
2. Cluster Randomization

# Unit of Analysis and Randomization

- Imagine we consider the effect of school vouchers on academic performance. Is it better to randomize vouchers to students or to schools?

# Unit of Analysis and Randomization

- Imagine we consider the effect of school vouchers on academic performance. Is it better to randomize vouchers to students or to schools?
- It depends on the question: Do we want to know how students respond to new environment or how schools respond to competition?
- Choice of analytic level determines what the study has the capacity to demonstrate.
- Analytical level also determines the effective number of observations
- Can also help with interference among units (e.g. interactions within and between schools)

# Variance with Clustered Random Assignment

- Assume you randomly assign clusters of units, each cluster has several individual units, and outcomes are measured at the individual level (e.g. assign schools and measure students' test scores)
- Since random assignment occurred at the cluster level and units within each cluster may not be independent, we need to correct our variance estimation for the clustering
- Ignoring the clustering can severely distort/downward bias the true variance
- Effectively the number of observations is the number of clusters not the number of units
- A safe bet is always to simply compute the average outcome for each cluster and then analyze the results using the cluster level averages.

# Estimation with Clustered Random Assignment

## Estimator of Standard Error of ATE with Cluster Randomization

Consider $K$ equal-sized clusters, $m$ of which are assigned to treatment. Now imagine we take cluster-level averages of the outcome variable $Y$. A conservative estimator for the standard error of the ATE is given by

$$\widehat{SE}_{\widehat{ATE}} = \sqrt{\frac{N}{K(N-m)}\widehat{Var}[\bar{Y}_{k|D_k=0}] + \frac{N}{Km}\widehat{Var}[\bar{Y}_{k|D_k=1}]}$$

where $\bar{Y}_k$ is the average outcome in the k-th cluster.

- With regression we can cluster the standard errors by the cluster unit, which works if numbers of clusters is large enough
- Difference in means estimator for ATE not unbiased if cluster sizes vary with potential outcomes, but bias vanishes as number of clusters increases
- Pick equal sized clusters or block on cluster size if possible

# Example: Kansas City Mobilization Experiment

- Kansas City experiment studied a local activist group canvassing on behalf of a public transportation ballot initiative in the weeks before the election
- 28 predominantly African-American precincts (clusters), were randomly assigned to treatment and control (equal allocation)
- Overall there are $N = 9,712$ voters in the 28 precincts. Average number of voters per precincts was 347
- Since randomization occurred at the precinct level we need to cluster the standard errors by precinct or analyze the precinct averages to adjust for the non-independence.

# Kansas City Mobilization Experiment

R Code

```
> library(foreign)
> library(lmtest)
> d <- read.dta("Arceneaux_AAAPSSsubset_2005.dta")
>
> head(d[,c("precint","vote03","treatmen")])
  precint vote03 treatmen
1      27      1        1
2       9      0        0
3      16      0        1
4       8      1        0
5       8      1        0
6       8      0        0
> table(d$precint)

  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17 
655 386 245 364 272 124 288 386 417 336 309 237 487 273 268 434 211 
 18  19  20  21  22  23  24  25  26  27  28 
568 211 496  80 300 416 344  31 491 611 472 
```

# Kansas City Mobilization Experiment

R Code

```
> out <- lm(vote03~treatmen,data=d)
> coeftest(out)

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 0.2912743  0.0067046 43.4436 < 2.2e-16 ***
treatmen    0.0440186  0.0094075  4.6791  2.92e-06 ***
---
```

# Kansas City Mobilization Experiment

R Code

```
> d.ag <- aggregate(d[,c("vote03","treatmen")],by=list(d$precint),mean)
> head(d.ag)
  Group.1   vote03 treatmen
1       1 0.3832061        0
2       2 0.1865285        0
3       3 0.3306122        0
4       4 0.3379121        0
5       5 0.3382353        0
6       6 0.2983871        0
> out2 <- lm(vote03~treatmen,data=d.ag)
> coeftest(out2)

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 0.288516   0.017244 16.7318 1.947e-15 ***
treatmen    0.036294   0.024386  1.4883    0.1487
---
```

# Kansas City Mobilization Experiment

R Code

```
> source("vcovCluster.r")
> coeftest(out,vcov = vcovCluster(out, cluster = d$precint))

t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.291274   0.020785 14.0140  < 2e-16 ***
treatmen    0.044019   0.025163  1.7494  0.08026 .
---
```

# Conclusion: Experiments

- Random assignment solves the identification problem for causal inference based on minimal assumptions that we can control as researchers
- Random assignment balances observed and unobserved confounders, which is why it is considered the gold standard for causal inference
- Statistical analysis is simple, transparent, and results are typically not model dependent, since confounders are controlled for "by design"
- Design features can help to improve inferences
- Always important to think about theory and external validity prior to experimentation