Randomized Experiments with Clusters

Ryan T. Moore

American University

2024-08-21

Table of contents I

Clusters in Experiments

Clusters in Experiments

- So far, *individual units* assigned to treatment

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum
  - Schools (not teachers or students) to new curriculum

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum
  - Schools (not teachers or students) to new curriculum
  - Media markets (not viewers) to ad campaign

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum
  - Schools (not teachers or students) to new curriculum
  - Media markets (not viewers) to ad campaign
  - Clinic catchment areas (not patients) to health infrastructure spending

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum
  - Schools (not teachers or students) to new curriculum
  - Media markets (not viewers) to ad campaign
  - Clinic catchment areas (not patients) to health infrastructure spending
- These groups of individuals are *clusters*

Clusters in Experiments

- So far, *individual units* assigned to treatment
- But, sometimes, *groups* of units assigned to treatment.
- Consider assigning ...
  - Classrooms (not students) to new curriculum
  - Schools (not teachers or students) to new curriculum
  - Media markets (not viewers) to ad campaign
  - Clinic catchment areas (not patients) to health infrastructure spending
- These groups of individuals are *clusters*
- (Let's start with easy case: all clusters same size ...) (Gerber and Green 2012)

Variance

Variance

- CLT SE for diff in means for clustered design:

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\mathrm{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\mathrm{Var}(\overline{Y_j}(1))}{m} + 2\mathrm{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE large?

- Dividing by $k \equiv$ number of *clusters*

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE large?

- Dividing by $k \equiv$ number of *clusters*
- So, small number of clusters $\rightsquigarrow$ larger SE

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE large?

- Dividing by $k \equiv$ number of *clusters*
- So, small number of clusters $\rightsquigarrow$ larger SE
- Small clusters (more variability in means)

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE large?

- Dividing by $k \equiv$ number of *clusters*
- So, small number of clusters $\rightsquigarrow$ larger SE
- Small clusters (more variability in means)
- If clusters differ in potential outcomes $\rightsquigarrow$ larger SE

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE large?

- Dividing by $k \equiv$ number of *clusters*
- So, small number of clusters $\rightsquigarrow$ larger SE
- Small clusters (more variability in means)
- If clusters differ in potential outcomes $\rightsquigarrow$ larger SE
- I.e., if clusters meaningful $\rightsquigarrow$ larger SE

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\mathrm{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\mathrm{Var}(\overline{Y_j}(1))}{m} + 2\mathrm{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE small?

- Many clusters

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE small?

- Many clusters
- If clusters look like each other, $\rightsquigarrow$ smaller SE

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE small?

- Many clusters
- If clusters look like each other, $\rightsquigarrow$ smaller SE
- $\text{Var}(\overline{Y_j}(1))$: variance in cluster averages of Y_1

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE small?

- Many clusters
- If clusters look like each other, $\rightsquigarrow$ smaller SE
- $\text{Var}(\overline{Y_j}(1))$: variance in cluster averages of Y_1
- $\text{Var}(\overline{Y_j}(0))$: variance in cluster averages of Y_0

Variance

- CLT SE for diff in means for clustered design:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left[\frac{m\text{Var}(\overline{Y_j}(0))}{N-m} + \frac{(N-m)\text{Var}(\overline{Y_j}(1))}{m} + 2\text{Cov}(\overline{Y_j}(0), \overline{Y_j}(1))\right]}$$

What might make this SE small?

- Many clusters
- If clusters look like each other, $\rightsquigarrow$ smaller SE
- $\text{Var}(\overline{Y_j}(1))$: variance in cluster averages of Y_1
- $\text{Var}(\overline{Y_j}(0))$: variance in cluster averages of Y_0
- If $\text{Var}(\overline{Y_j}(1)), \text{Var}(\overline{Y_j}(0))$ are small, SE is small

Bias

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0
  - Classrooms differ in number of students, and larger classes have lower test scores

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0
  - Classrooms differ in number of students, and larger classes have lower test scores
  - Media markets vary in number of viewers, and voters in cities less likely to turn out

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0
  - Classrooms differ in number of students, and larger classes have lower test scores
  - Media markets vary in number of viewers, and voters in cities less likely to turn out
  - Health clusters vary in patient population, and larger clusters are healthier

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0
  - Classrooms differ in number of students, and larger classes have lower test scores
  - Media markets vary in number of viewers, and voters in cities less likely to turn out
  - Health clusters vary in patient population, and larger clusters are healthier
- In this case, individual difference in means estimator is *biased* for true ATE

Bias

- Consider case where clusters vary in size, and size correlates with Y_1, Y_0
  - Classrooms differ in number of students, and larger classes have lower test scores
  - Media markets vary in number of viewers, and voters in cities less likely to turn out
  - Health clusters vary in patient population, and larger clusters are healthier
- In this case, individual difference in means estimator is *biased* for true ATE
  - (Different assignments of clusters will produce different counts of treated units!)

Bias

Alternative estimator: Difference in totals (not means)

$$\widehat{ATE} = \frac{k_T + k_C}{N} \left(\frac{\sum Y_i(1) | T_i = 1}{k_T} - \frac{\sum Y_i(0) | T_i = 0}{k_C} \right)$$

Bias

Alternative estimator: Difference in totals (not means)

$$\widehat{ATE} = \frac{k_T + k_C}{N} \left(\frac{\sum Y_i(1)|T_i = 1}{k_T} - \frac{\sum Y_i(0)|T_i = 0}{k_C} \right)$$

- But, doesn't include cluster sizes, so may be high variance

Bias

Alternative estimator: Difference in totals (not means)

$$\widehat{ATE} = \frac{k_T + k_C}{N}\left(\frac{\sum Y_i(1)|T_i = 1}{k_T} - \frac{\sum Y_i(0)|T_i = 0}{k_C}\right)$$

- But, doesn't include cluster sizes, so may be high variance
- I.e., more students per classroom doesn't help precision here (unlike above)

An Example

An Example

- 1000 voters, split across 10 cities

An Example

- 1000 voters, split across 10 cities
- Measure ideology, which varies by city and voter

An Example

- 1000 voters, split across 10 cities
- Measure ideology, which varies by city and voter
- Do 2 hypothetical assignments

An Example

- 1000 voters, split across 10 cities
- Measure ideology, which varies by city and voter
- Do 2 hypothetical assignments
  - Individual-level assignment to news

An Example

- 1000 voters, split across 10 cities
- Measure ideology, which varies by city and voter
- Do 2 hypothetical assignments
    - Individual-level assignment to news
    - City-level cluster assignment to news

An Example

- 1000 voters, split across 10 cities
- Measure ideology, which varies by city and voter
- Do 2 hypothetical assignments
  - Individual-level assignment to news
  - City-level cluster assignment to news
- Use RI to find SE in each case

An Example

```
library(fabricatr)
library(randomizr)
library(tidyverse)

# Make data and visualise

set.seed(95852894)

voters <- fabricate(
  N = 1000,
  city_id = rep(1:10, 100),
  ideology = draw_normal_icc(mean = 0, N = N,
                             clusters = city_id, ICC = 0.7)
  city_id_fac = as.factor(city_id),
  city_id_fac = fct_reorder(city_id_fac, ideology)
  )
```

An Example

```r
table(voters$city_id)
```

```
  1   2   3   4   5   6   7   8   9  10 
100 100 100 100 100 100 100 100 100 100 
```

An Example

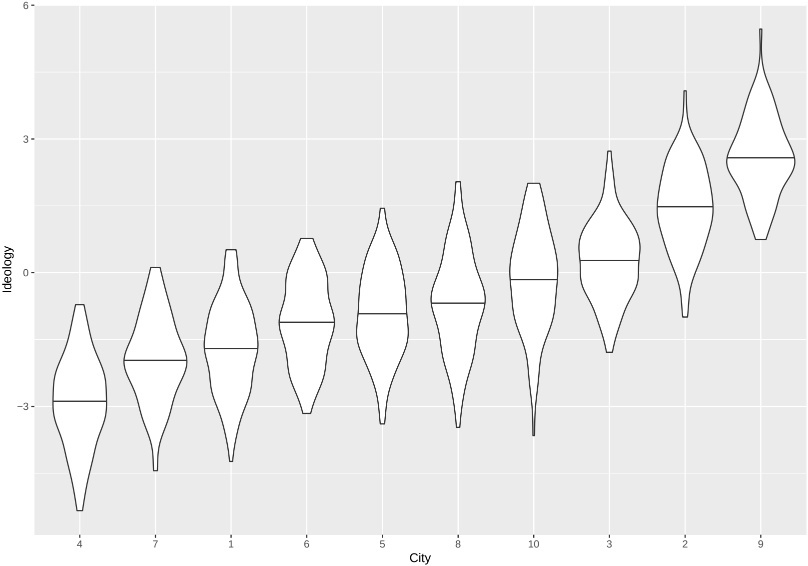
```r
table(voters$city_id)
```

```

  1   2   3   4   5   6   7   8   9  10 
100 100 100 100 100 100 100 100 100 100 
```

```r
voters |> group_by(city_id) |>
  summarise(mean_ideo = mean(ideology)) |>
  arrange(mean_ideo)
```

```
# A tibble: 10 x 2
   city_id mean_ideo
     <int>     <dbl>
 1       4    -2.90 
 2       7    -1.94 
 3       1    -1.68 
 4       6    -1.10 
 5       5    -0.903
 6       8    -0.686
 7      10    -0.161
 8       3     0.267
 9       2     1.46 
10       9     2.60 
```

An Example



An Example

Assignments

```
voters <- voters |>
  mutate(
    tr_ind = sample(rep(0:1, nrow(voters) / 2)),
    tr_cl = cluster_ra(clusters = voters$city_id, m = 5))
```

An Example

Assignments

```
voters <- voters |>
  mutate(
    tr_ind = sample(rep(0:1, nrow(voters) / 2)),
    tr_cl = cluster_ra(clusters = voters$city_id, m = 5))
```

```
voters |> count(tr_ind)
```

```
  tr_ind   n
1      0 500
2      1 500
```

```
voters |> count(tr_cl)
```

```
  tr_cl   n
1     0 500
2     1 500
```

An Example

Conditions by city, individual-level assignment:

```r
table(voters$city_id, voters$tr_ind)
```

```
     0  1
1   48 52
2   47 53
3   52 48
4   47 53
5   51 49
6   49 51
7   49 51
8   53 47
9   48 52
10  56 44
```

An Example

Conditions by city, cluster-level assignment:

```
table(voters$city_id, voters$tr_cl)
```

```
       0   1
  1    0 100
  2  100   0
  3  100   0
  4    0 100
  5  100   0
  6    0 100
  7    0 100
  8    0 100
  9  100   0
  10 100   0
```

An Example

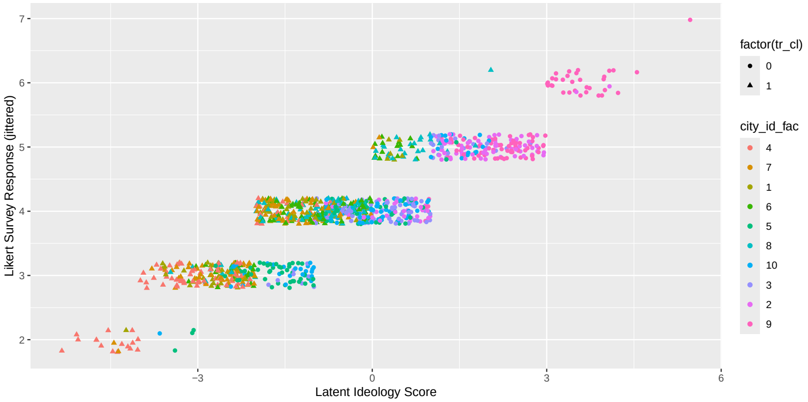
Draw responses under each assignment:

```
voters <- voters |>
  mutate(
    response_ind = draw_likert(x = ideology + tr_ind, m
    response_cl = draw_likert(x = ideology + tr_cl, mir
```

An Example

Draw responses under each assignment:

```
voters <- voters |>
  mutate(
    response_ind = draw_likert(x = ideology + tr_ind, m
    response_cl = draw_likert(x = ideology + tr_cl, mi
```



An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments

An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments
- Do 1000 cluster-level hypothetical assignments

An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments
- Do 1000 cluster-level hypothetical assignments
- Use sharp null $H_0 : \tau_i = 0$ to impute unobserved potential outcomes

An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments
- Do 1000 cluster-level hypothetical assignments
- Use sharp null $H_0 : \tau_i = 0$ to impute unobserved potential outcomes
- Calculate $\widehat{ATE}$ in each case

An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments
- Do 1000 cluster-level hypothetical assignments
- Use sharp null $H_0 : \tau_i = 0$ to impute unobserved potential outcomes
- Calculate $\widehat{ATE}$ in each case
- Compare distribution of $\widehat{ATE}_{\text{Individual}}$ to distribution of $\widehat{ATE}_{\text{Cluster}}$

An Example

Analyse precision with RI:

- Do 1000 individual-level hypothetical assignments
- Do 1000 cluster-level hypothetical assignments
- Use sharp null $H_0 : \tau_i = 0$ to impute unobserved potential outcomes
- Calculate $\widehat{ATE}$ in each case
- Compare distribution of $\widehat{ATE}_{\text{Individual}}$ to distribution of $\widehat{ATE}_{\text{Cluster}}$
- With partner, sketch the two randomization distributions!

An Example

Analyse precision with RI:

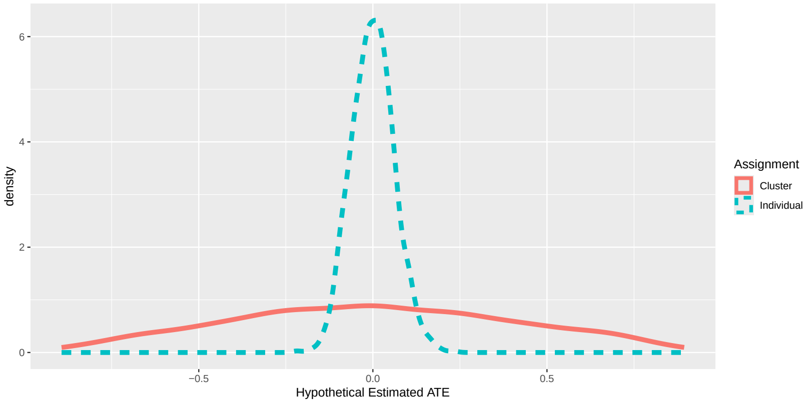
```
n_sims <- 1000
df_ates <- tibble(ate_ind = NA,
                  ate_cl = NA)

for(idx in 1:n_sims){

  voters <- voters |>
    mutate(
      hyp_tr_individ = sample(rep(0:1, nrow(voters) / 2)),
      hyp_tr_cl = cluster_ra(clusters = voters$city_id, m = 5))

  ate_ind <- mean(voters$response_ind[voters$hyp_tr_ind == 1]) -
    mean(voters$response_ind[voters$hyp_tr_ind == 0])

  ate_cl <- mean(voters$response_cl[voters$hyp_tr_cl == 1]) -
    mean(voters$response_cl[voters$hyp_tr_cl == 0])

  df_ates[idx, "ate_ind"] <- ate_ind
  df_ates[idx, "ate_cl"] <- ate_cl
}
```

An Example

Analyse precision with RI:

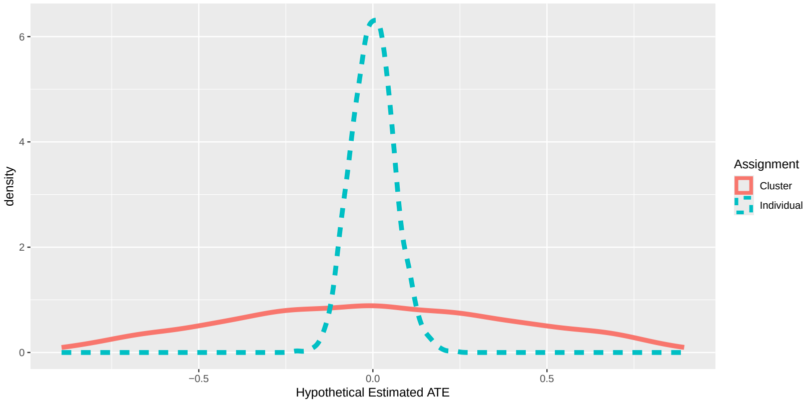
An Example

Analyse precision with RI:



An Example

Analyse precision with RI:




```
# A tibble: 1 x 2
  se_ind se_cl
   <dbl> <dbl>
1 0.0614 0.408
```

Next:
Regression and Experiments

`rtm@american.edu`
`www.ryantmoore.org`

References I

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: WW Norton.