Day 03: Sampling variability and random assignment procedures

Erin Rossiter

January 25, 2022

Today's plan

1. Tidy up

– Emma, will you:
  » remind me to take a break at 4:30-4:45
  » jot notes on typos & email after class

2. Lecture
3. Lab

Today's plan

1. Tidy up

   – Emma, will you:
     - » remind me to take a break at 4:30-4:45
     - » jot notes on typos & email after class

2. Lecture
3. Lab

Today's plan

1. Tidy up

– Emma, will you:
  - » remind me to take a break at 4:30-4:45
  - » jot notes on typos & email after class

2. Lecture
3. Lab

Recap

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:

  1. Randomization of treatment
     - $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     - $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     - We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     » $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     » $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     » We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
    1. Randomization of treatment
        - » $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
        - » $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
        - » We can estimate left-hand terms using our observed data!
    2. Excludability
    3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     - $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     - $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     - We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     - » $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     - » $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     - » We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     - $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     - $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     - We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     - $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     - $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     - We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     » $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     » $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     » We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Recap

- The average treatment effect (ATE) is often an interesting, important, useful estimand.
- Difference-in-means is an **unbiased** estimator of the ATE assuming:
  1. Randomization of treatment
     » $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$
     » $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$
     » We can estimate left-hand terms using our observed data!
  2. Excludability
  3. Noninterference
- What about **uncertainty**?

Sampling distribution

Sampling distribution

- Our experiment yields one (unbiased) estimate of the ATE under a *single* random assignment
- **Sampling distribution** the frequency distribution of a statistc (e.g., estimated ATEs) that could have been generated when rerunning the procedure (e.g., every possible random assignment)
  - Last time: $E[\hat{ATE}] = ATE$
  - Today: we're concerned about **variability**

Sampling distribution

- Our experiment yields one (unbiased) estimate of the ATE under a *single* random assignment
- **Sampling distribution** the frequency distribution of a statistc (e.g., estimated ATEs) that could have been generated when rerunning the procedure (e.g., every possible random assignment)
    - Last time: $E[\hat{ATE}] = ATE$
    - Today: we're concerned about **variability**

Sampling distribution

- Our experiment yields one (unbiased) estimate of the ATE under a *single* random assignment
- **Sampling distribution** the frequency distribution of a statistc (e.g., estimated ATEs) that could have been generated when rerunning the procedure (e.g., every possible random assignment)
  - Last time: $E[A\hat{T}E] = ATE$
  - Today: we're concerned about **variability**

Sampling distribution

- Our experiment yields one (unbiased) estimate of the ATE under a *single* random assignment
- **Sampling distribution** the frequency distribution of a statistc (e.g., estimated ATEs) that could have been generated when rerunning the procedure (e.g., every possible random assignment)
  - Last time: $E[A\hat{T}E] = ATE$
  - Today: we're concerned about **variability**

In code

```
N = 6
population <- declare_population(N = N,
                                 u = round(rnorm(N)))
potential_outcomes <- declare_potential_outcomes(Y ~ 3*Z + u)
assignment <- declare_assignment(Z = complete_ra(N=N, m=N/2))
design <- population + potential_outcomes
set.seed(123)
df <- draw_data(design)
df
```

```
##   ID  u Y_Z_0 Y_Z_1
## 1  1 -1    -1     2
## 2  2  0     0     3
## 3  3  2     2     5
## 4  4  0     0     3
## 5  5  0     0     3
## 6  6  2     2     5
```

20 ways to allocate treatment with complete RA

```
df$Z <- NA
df
```

```
##   ID  u Y_Z_0 Y_Z_1  Z
## 1  1 -1    -1     2 NA
## 2  2  0     0     3 NA
## 3  3  2     2     5 NA
## 4  4  0     0     3 NA
## 5  5  0     0     3 NA
## 6  6  2     2     5 NA
```

```
(all_ras <- combn(x = N, m = N/2))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
## [1,]    1    1    1    1    1    1    1    1    1     1     2
## [2,]    2    2    2    2    3    3    3    4    4     5     3
## [3,]    3    4    5    6    4    5    6    5    6     6     4
##      [,15] [,16] [,17] [,18] [,19] [,20]
## [1,]     2     2     3     3     3     4
## [2,]     4     5     4     4     5     5
## [3,]     6     6     5     6     6     6
```

Sampling distribution of ATE

```
poss_ates <- rep(NA, ncol(all_ras))
for(i in 1:ncol(all_ras)){
  treated <- all_ras[,i] #indices of treated units
  df$Z <- NA
  df$Z[treated] <- 1 #assign those indices 1
  df$Z[-treated] <- 0 #assign all others 0
  # Calculate DIM estimate
  poss_ates[i] <- mean(df$Y_Z_1[df$Z == 1]) -
                  mean(df$Y_Z_0[df$Z == 0])
}
#unbiased, but... precise?
mean(poss_ates)
```

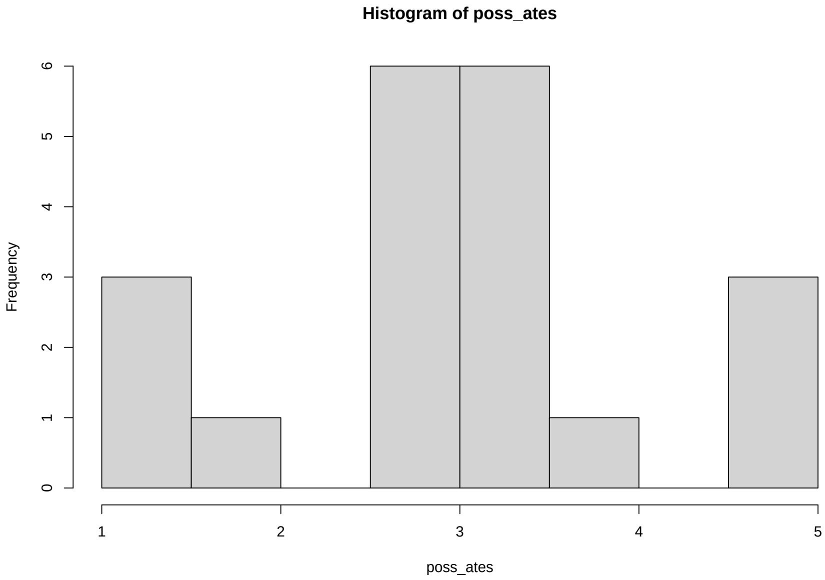
```
## [1] 3
```

```
poss_ates
```

```
##  [1] 2.666667 1.333333 1.333333 2.666667 2.666667 2.666667 4.
##  [9] 2.666667 2.666667 3.333333 3.333333 4.666667 2.000000 3.
## [17] 3.333333 4.666667 4.666667 3.333333
```

Sampling distribution of ATE

```r
hist(poss_ates)
```



Standard error

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, A\hat{T}E, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, A\hat{T}E, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Standard error

- To be clear:
  - a sampling distribution is a *distribution* of a statistic (e.g., $\hat{\beta}, \hat{ATE}, \hat{\mu}$)
  - a standard deviation is a single value quantifying variability for any variable X:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- **Standard error is the standard deviation of the sampling distribution**
  - How we summarize sampling variability
  - In experiments, how much might our estimate vary across all hypothetical D_i's?
  - Bigger SE, more uncertainty about our single estimate

Our example

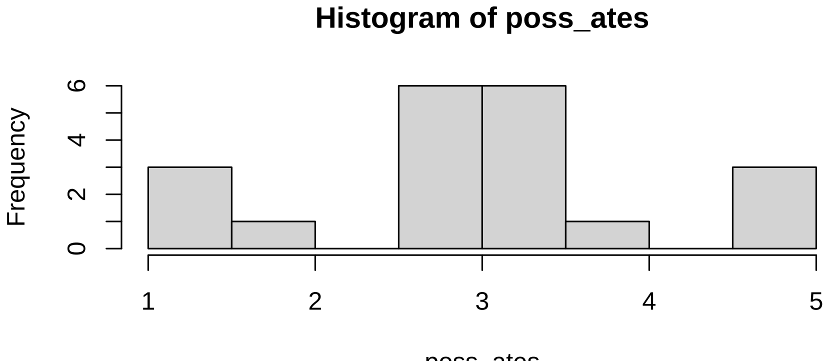
Note: we're talking about **the true standard error** of $\hat{ATE}$

- Let's calculate it with our simulation!
- Later... how we estimate it in practice

```
sd(poss_ates)
```

```
## [1] 1.025978
```

```
hist(poss_ates)
```



Our example

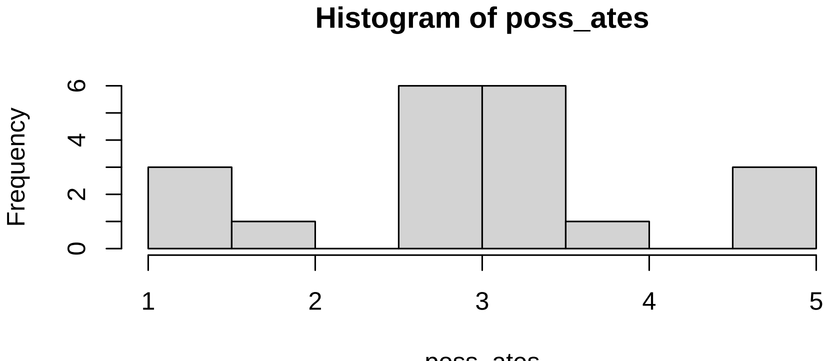
Note: we're talking about **the true standard error** of $\hat{ATE}$

- Let's calculate it with our simulation!
- Later… how we estimate it in practice

```
sd(poss_ates)
```

```
## [1] 1.025978
```

```
hist(poss_ates)
```



Our example

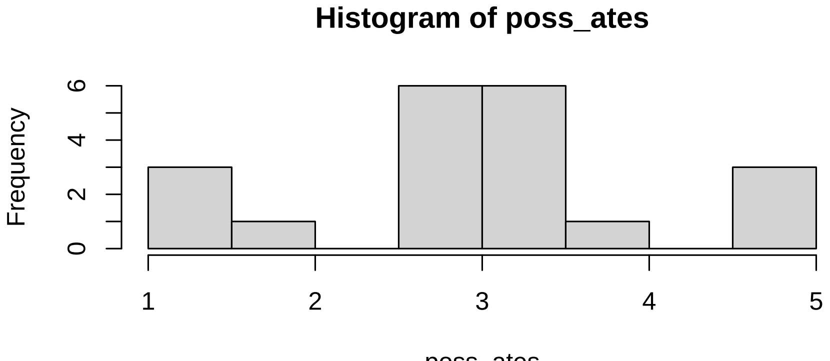
Note: we're talking about **the true standard error** of $\hat{ATE}$

- Let's calculate it with our simulation!
- Later... how we estimate it in practice

```
sd(poss_ates)
```

```
## [1] 1.025978
```

```
hist(poss_ates)
```



Our example

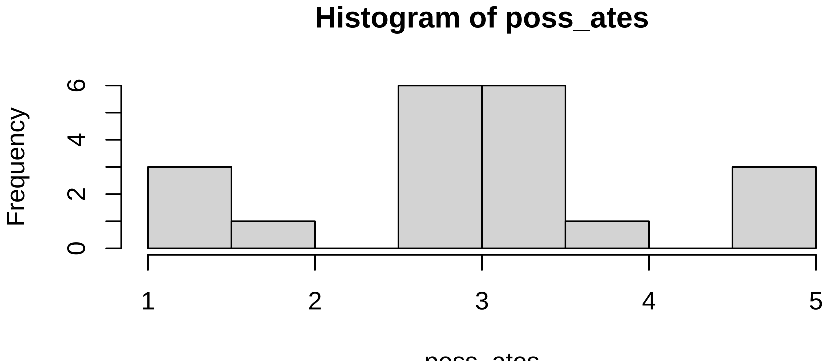
Note: we're talking about **the true standard error** of $\hat{ATE}$

- Let's calculate it with our simulation!
- Later... how we estimate it in practice

```
sd(poss_ates)
```

```
## [1] 1.025978
```

```
hist(poss_ates)
```



Today

How do we reduce the standard error?

In other words, how do we **design our experiment** so that it produces more precise estimates of the ATE?

A thought experiment

How do we reduce the standard error? This isn't p-hacking! We can do this by being thoughtful about our design...

$$SE(A\hat{T}E) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Term by term: N

All else being equal, how does adding/subtracting subjects affect sampling variability?

$$SE(A\hat{T}E) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Term by term: Variances

All else being equal, how does increasing/decreasing variances of the potential outcomes affect sampling variability?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Term by term: Covariance

All else being equal, how does increasing/decreasing the covariance of the potential outcomes affect sampling variability?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(A\hat{T}E)$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{S}E(A\hat{T}E)$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(A\hat{T}E) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(A\hat{T}E)$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(A\hat{T}E)$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(A\hat{T}E) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(\hat{ATE})$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(\hat{ATE})$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(\hat{ATE})$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(\hat{ATE})$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(\hat{ATE})$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(\hat{ATE})$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(\hat{ATE})$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(\hat{ATE})$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Estimation

- Remember, we've been discussing the *true* standard error of our estimate $SE(\hat{ATE})$
  - Useful to understand theoretical sampling variability inherent in randomization
- What do we do in practice to describe uncertainty of our estimate ATE?
  - We need to generate an estimate of the standard error $\hat{SE}(\hat{ATE})$
- Spoiler: there's a problem estimating this.
  - What term is **not** estimable from our sample?

$$SE(\hat{ATE}) = \sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\right)}$$

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE
   - Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1
2. Note that variances are estimated themselves. See GG pg 61

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE
   - Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1
2. Note that variances are estimated themselves. See GG pg 61

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE
   - Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1
2. Note that variances are estimated themselves. See GG pg 61

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE

   – Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1

2. Note that variances are estimated themselves. See GG pg 61

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE
   - Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1
2. Note that variances are estimated themselves. See GG pg 61

Conservative estimation

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N-m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

- We won't walk through this in detail- Big picture:

1. We make a simplifying assumption so $\hat{SE}$ is at least as large as true SE

   – Assume correlation between $Y_i(0)$ and $Y_i(1)$ is 1

2. Note that variances are estimated themselves. See GG pg 61

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

So far

- Standard error is the standard deviation of the sampling distribution
  - There's some true standard error associated with our design
  - Depends on:
    - Variability in potential outcomes
    - Sample size
    - etc.
  - We estimate the standard error to learn uncertainty of our one estimate
- Next: randomization schemes also affect sampling variability.
  - Block randomization
  - Cluster randomization

Block randomization

Blocking

A randomization procedure where:

1. subjects are partitioned into subgroups (or blocks, or strata)
2. complete random assignment occurs within each block

Blocking

A randomization procedure where:

1. subjects are partitioned into subgroups (or blocks, or strata)
2. complete random assignment occurs within each block

Blocking

A randomization procedure where:

1. subjects are partitioned into subgroups (or blocks, or strata)
2. complete random assignment occurs within each block

Blocking

A randomization procedure where:

1. subjects are partitioned into subgroups (or blocks, or strata)
2. complete random assignment occurs within each block

Example

- Imagine a context where gender relates to our outcome
- $N = 20$, 10 men, 10 women
- We could use complete randomization
  - What might happen?

Example

- Imagine a context where gender relates to our outcome
- $N = 20$, 10 men, 10 women
- We could use complete randomization
  - What might happen?

Example

- Imagine a context where gender relates to our outcome
- $N = 20$, 10 men, 10 women
- We could use complete randomization
  - What might happen?

Example

- Imagine a context where gender relates to our outcome
- $N = 20$, 10 men, 10 women
- We could use complete randomization
  - What might happen?

Example

- Imagine a context where gender relates to our outcome
- $N = 20$, 10 men, 10 women
- We could use complete randomization
  - What might happen?

Example cont.

- If we want gender to be "balanced" across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Example cont.

- If we want gender to be "balanced" across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Example cont.

- If we want gender to be "balanced" across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Example cont.

- If we want gender to be “balanced” across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Example cont.

- If we want gender to be “balanced” across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Example cont.

- If we want gender to be "balanced" across treatment and control groups, we can:
  - Split group on gender
  - Conduct complete randomization *within gender blocks*:
    - Treatment group will have 5 men, 5 women
    - Control group will have 5 men, 5 women
  - Think of it as several mini-experiments we then aggregate

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(\hat{ATE}) = \sqrt{(}\sum_{j=1}^{J} SE^2(\hat{ATE_j}))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(\hat{ATE}) = \sqrt(\sum_{j=1}^{J} SE^2(\hat{ATE_j}))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(A\hat{T}E) = \sqrt{(}\sum_{j=1}^{J} SE^2(A\hat{T}E_j))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(\hat{ATE}) = \sqrt(\sum_{j=1}^{J} SE^2(\hat{ATE_j}))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(\hat{ATE}) = \sqrt{(} \sum_{j=1}^{J} SE^2(\hat{ATE_j}))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - » $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - » $SE(A\hat{T}E) = \sqrt{(}\sum_{j=1}^{J} SE^2(A\hat{T}E_j))$
  - » Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - » Why does this reduce sampling variability?

Why block?

Blocking allows us to estimate the ATE while reducing sampling variability!

- Relationship between overall ATE and ATE within each block j is:
  - $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
- Relationship between overall uncertainty and uncertainty within each block is:
  - $SE(\hat{ATE}) = \sqrt{(}\sum_{j=1}^{J} SE^2(\hat{ATE}_j))$
  - Recall idea behind blocking is to create blocks where units have similar potential outcomes
  - Why does this reduce sampling variability?

Note on estimation

- Estimating ATE is still unbiased
  - ATE calculated *within block* first because that's how we randomized!
  - Overall ATE is then a weighted average

$$\hat{ATE} = \sum_{j=1}^{J} \frac{N_j}{N} \hat{ATE}_j$$

Implementation discussion

- How would you implement blocking? Is it feasible?
  - Real time?
  - Pre-test?

Cluster randomization

Cluster randomization

A randomization procedure where:

1. units are clustered in an observable, natural way
2. level of randomization is *cluster*

- all units in a cluster are placed into treatment or control
- rules out D_i where units in a cluster would be assigned to different conditions

Cluster randomization

A randomization procedure where:

1. units are clustered in an observable, natural way
2. level of randomization is *cluster*

- all units in a cluster are placed into treatment or control
- rules out D_i where units in a cluster would be assigned to different conditions

Cluster randomization

A randomization procedure where:

1. units are clustered in an observable, natural way
2. level of randomization is *cluster*

- all units in a cluster are placed into treatment or control
- rules out D_i where units in a cluster would be assigned to different conditions

Cluster randomization

A randomization procedure where:

1. units are clustered in an observable, natural way
2. level of randomization is *cluster*

– all units in a cluster are placed into treatment or control
– rules out D_i where units in a cluster would be assigned to different conditions

Cluster randomization

A randomization procedure where:

1. units are clustered in an observable, natural way
2. level of randomization is *cluster*

- all units in a cluster are placed into treatment or control
- rules out D_i where units in a cluster would be assigned to different conditions

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

In practice

Recall our assumption of no interference

- If clusters are unavoidable, do cluster randomization!
- But, it's a sacrifice (more on that)

Examples

- TV ads broadcast to media markets
- Classrooms getting different curriculum
- Any policy that is administered at a local level

Implications

1. Within clusters, potential outcomes are likely correlated

   - People are similar within towns, networks, classrooms, etc.
   - Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

   - Equal cluster sizes, DIM unbiased :)
   - Unequal cluster sizes, DIM *maybe* biased
     - » Do cluster sizes correlate with potential outcomes?
       - Ex: effect of TV ad on vote choice, randomized at media market level
       - Bigger markets more likely to vote Dem?
     - » Bias goes away as number of clusters increases

Implications

1. Within clusters, potential outcomes are likely correlated

   - People are similar within towns, networks, classrooms, etc.
   - Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

   - Equal cluster sizes, DIM unbiased :)
   - Unequal cluster sizes, DIM *maybe* biased
     - Do cluster sizes correlate with potential outcomes?
       - Ex: effect of TV ad on vote choice, randomized at media market level
       - Bigger markets more likely to vote Dem?
     - Bias goes away as number of clusters increases

Implications

1. Within clusters, potential outcomes are likely correlated

- People are similar within towns, networks, classrooms, etc.
- Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

- Equal cluster sizes, DIM unbiased :)
- Unequal cluster sizes, DIM *maybe* biased
  - Do cluster sizes correlate with potential outcomes?
    - Ex: effect of TV ad on vote choice, randomized at media market level
    - Bigger markets more likely to vote Dem?
  - Bias goes away as number of clusters increases

Implications

1. Within clusters, potential outcomes are likely correlated

– People are similar within towns, networks, classrooms, etc.
– Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

– Equal cluster sizes, DIM unbiased :)
– Unequal cluster sizes, DIM *maybe* biased
  » Do cluster sizes correlate with potential outcomes?
    • Ex: effect of TV ad on vote choice, randomized at media market level
    • Bigger markets more likely to vote Dem?
  » Bias goes away as number of clusters increases

Implications

1. Within clusters, potential outcomes are likely correlated

- People are similar within towns, networks, classrooms, etc.
- Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

- Equal cluster sizes, DIM unbiased :)
- Unequal cluster sizes, DIM *maybe* biased
  - Do cluster sizes correlate with potential outcomes?
    - Ex: effect of TV ad on vote choice, randomized at media market level
    - Bigger markets more likely to vote Dem?
  - Bias goes away as number of clusters increases

Implications

1. Within clusters, potential outcomes are likely correlated

- People are similar within towns, networks, classrooms, etc.
- Less information (smaller effective N) = more sampling variability!

2. Unequal cluster sizes

- Equal cluster sizes, DIM unbiased :)
- Unequal cluster sizes, DIM *maybe* biased
  - » Do cluster sizes correlate with potential outcomes?
    - Ex: effect of TV ad on vote choice, randomized at media market level
    - Bigger markets more likely to vote Dem?
  - » Bias goes away as number of clusters increases