Uncertainty, consistency and hypothesis testing

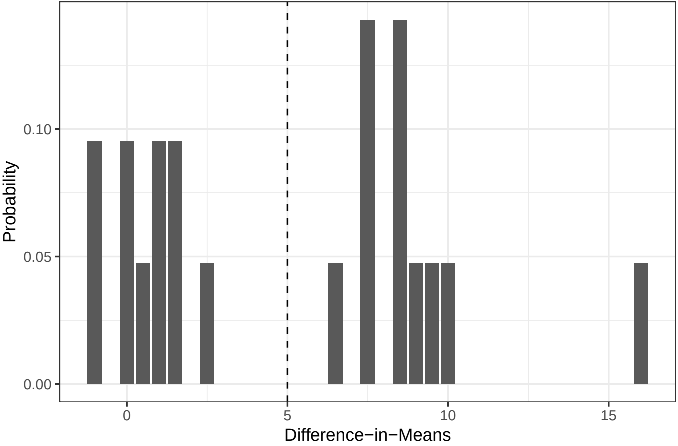
Thomas Leavitt

July 22, 2022

Variance of Difference-in-Means

Variance of Difference-in-Means estimator

- Yesterday we showed that Difference-in-Means estimator in "village heads" example is



- What is the variance of this estimator?
- Why do we care about variance of an estimator?

Variance of Difference-in-Means estimator

- Variance is average squared distance of estimator from its expected value:

$$\underbrace{\mathrm{E}\left[\underbrace{\left(\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\underbrace{\mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}_{\text{Expected value}}\right)^2}_{\text{Squared distance from expected value}}\right]}_{\text{Average (or expected) squared distance}}$$

Variance of Difference-in-Means estimator

- Variance is average squared distance of estimator from its expected value:

$$\underbrace{\mathrm{E}\left[\underbrace{\left(\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\underbrace{\mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}_{\text{Expected value}}\right)^2}_{\text{Squared distance from expected value}}\right]}_{\text{Average (or expected) squared distance}}$$

- Diff-in-Means unbiased for τ, so write variance as

$$\mathrm{E}\left[\left(\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\tau\right)^2\right]$$

Variance of Difference-in-Means estimator

- Variance is average squared distance of estimator from its expected value:

$$\underbrace{\mathrm{E}\left[\underbrace{\left(\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\underbrace{\mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}_{\text{Expected value}}\right)^2}_{\text{Squared distance from expected value}}\right]}_{\text{Average (or expected) squared distance}}$$

- Diff-in-Means unbiased for τ, so write variance as

$$\mathrm{E}\left[\left(\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\tau\right)^2\right]$$

- In "village heads" example with 21 possible assignments

$$\left(\hat{\tau}\left(\boldsymbol{z}_1,\boldsymbol{y}_1\right)-\tau\right)^2 \Pr\left(\boldsymbol{Z}=\boldsymbol{z}_1\right)+\ldots+\left(\hat{\tau}\left(\boldsymbol{z}_{21},\boldsymbol{y}_{21}\right)-\tau\right)^2 \Pr\left(\boldsymbol{Z}=\boldsymbol{z}_{21}\right)$$

Variance is ≈ 21.19

Variance of Difference-in-Means estimator

- Neyman (1923) derived expression for variance of Diff-in-Means under complete random assignment

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

where

$\sigma^2_{y_C} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ci} - \frac{1}{N}\sum_{i=1}^{N} y_{Ci}\right)^2$ is var of control POs

$\sigma^2_{y_T} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ti} - \frac{1}{N}\sum_{i=1}^{N} y_{Ti}\right)^2$ is var of treated POs

$\sigma_{y_C, y_T} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ci} - \frac{1}{N}\sum_{i=1}^{N} y_{Ci}\right)\left(y_{Ti} - \frac{1}{N}\sum_{i=1}^{N} y_{Ti}\right)$ is cov of POs

Variance of Difference-in-Means estimator

- Neyman (1923) derived expression for variance of Diff-in-Means under complete random assignment

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

where

$\sigma^2_{y_C} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ci} - \frac{1}{N}\sum_{i=1}^{N} y_{Ci}\right)^2$ is var of control POs

$\sigma^2_{y_T} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ti} - \frac{1}{N}\sum_{i=1}^{N} y_{Ti}\right)^2$ is var of treated POs

$\sigma_{y_C, y_T} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{Ci} - \frac{1}{N}\sum_{i=1}^{N} y_{Ci}\right)\left(y_{Ti} - \frac{1}{N}\sum_{i=1}^{N} y_{Ti}\right)$ is cov of POs

$\star$ Note that $\sigma^2_{y_C}$, $\sigma^2_{y_T}$ and σ_{y_C, y_T} depend on unknown potential outcomes

Variance of Difference-in-Means estimator

- Sometimes you might see equivalent expression (Imbens and Rubin 2015)

$$\frac{S_{y_C}}{n_C} + \frac{S_{y_T}}{n_T} - \frac{S_\tau}{N},$$

where

$$S_{y_C} = \tfrac{1}{N-1} \sum_{i=1}^{N} \left(y_{Ci} - \tfrac{1}{N} \sum_{i=1}^{N} y_{Ci} \right)^2$$

$$S_{y_T} = \tfrac{1}{N-1} \sum_{i=1}^{N} \left(y_{Ti} - \tfrac{1}{N} \sum_{i=1}^{N} y_{Ti} \right)^2$$

$$S_\tau = \tfrac{1}{N-1} \sum_{i=1}^{N} \left(\tau_i - \tfrac{1}{N} \sum_{i=1}^{N} \tau_i \right)^2$$

Variance of Difference-in-Means estimator

- Sometimes you might see equivalent expression (Imbens and Rubin 2015)

$$\frac{S_{y_C}}{n_C} + \frac{S_{y_T}}{n_T} - \frac{S_\tau}{N},$$

where

$$S_{y_C} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_{Ci} - \frac{1}{N} \sum_{i=1}^{N} y_{Ci} \right)^2$$

$$S_{y_T} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_{Ti} - \frac{1}{N} \sum_{i=1}^{N} y_{Ti} \right)^2$$

$$S_\tau = \frac{1}{N-1} \sum_{i=1}^{N} \left(\tau_i - \frac{1}{N} \sum_{i=1}^{N} \tau_i \right)^2$$

⋆ Note that S_{y_C}, S_{y_T} and S_τ depend on unknown potential outcomes

Variance of Difference-in-Means estimator

- Example: "Village heads" study (Gerber and Green 2012, Chapter 2):

	Budget share (%)		
Village	y_C	y_T	τ
1	10	15	5
2	15	15	0
3	20	30	10
4	20	15	-5
5	10	20	10
6	15	15	0
7	15	30	15
Average	15	20	5

Variance of Difference-in-Means estimator

- Example: "Village heads" study (Gerber and Green 2012, Chapter 2):

	Budget share (%)		
Village	$\boldsymbol{y_C}$	$\boldsymbol{y_T}$	$\boldsymbol{\tau}$
1	10	15	5
2	15	15	0
3	20	30	10
4	20	15	-5
5	10	20	10
6	15	15	0
7	15	30	15
Average	15	20	5

- With access to true POs, we can directly calculate $\text{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$:

  $\sigma^2_{y_C} \approx 14.29$, $\sigma^2_{y_T} \approx 42.86$, $\sigma_{y_C, y_T} \approx 7.14$

- So, $\text{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right) \approx 21.19$

Variance of Difference-in-Means estimator

- Example: "Village heads" study (Gerber and Green 2012, Chapter 2):

Village	Budget share (%) y_C	y_T	τ
1	10	15	5
2	15	15	0
3	20	30	10
4	20	15	-5
5	10	20	10
6	15	15	0
7	15	30	15
Average	15	20	5

- With access to true POs, we can directly calculate $\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]$:

  $\sigma^2_{y_C} \approx 14.29$, $\sigma^2_{y_T} \approx 42.86$, $\sigma_{y_C,y_T} \approx 7.14$

- So, $\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T\sigma^2_{y_C}}{n_C} + \frac{n_C\sigma^2_{y_T}}{n_T} + 2\sigma_{y_C,y_T}\right) \approx 21.19$
- In practice, $\sigma^2_{y_C}$, $\sigma^2_{y_T}$ and σ_{y_C,y_T} unknown, so we estimate $\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]$

Variance estimation

Variance estimation

We showed that the Diference-in-Means estimator's variance is

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- We have unbiased estimators for $\sigma^2_{y_C}$ and $\sigma^2_{y_T}$, but not σ_{y_C, y_T}
- So what do we do?

Variance estimation

We showed that the Diference-in-Means estimator's variance is

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- We have unbiased estimators for $\sigma^2_{y_C}$ and $\sigma^2_{y_T}$, but not σ_{y_C, y_T}
- So what do we do?
- We use **conservative** "plug-in" estimator (Neyman 1923)
- **Conservative** means that

$$\mathrm{E}\left[\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right] \geq \mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$$

Variance estimation

We showed that the Diference-in-Means estimator's variance is

$$\operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- We have unbiased estimators for $\sigma_{y_C}^2$ and $\sigma_{y_T}^2$, but not σ_{y_C, y_T}
- So what do we do?
- We use **conservative** "plug-in" estimator (Neyman 1923)
- **Conservative** means that

$$\mathrm{E}\left[\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right] \geq \operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$$

- When potential outcomes perfectly positively correlated

$$\mathrm{E}\left[\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right] = \operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$$

- Otherwise, $\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$ is positively biased (conservative)

Variance estimation

We showed that the Diference-in-Means estimator's variance is

$$\operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- We have unbiased estimators for $\sigma_{y_C}^2$ and $\sigma_{y_T}^2$, but not σ_{y_C, y_T}
- So what do we do?
- We use **conservative** "plug-in" estimator (Neyman 1923)
- **Conservative** means that

$$\mathrm{E}\left[\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right] \geq \operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$$

- When potential outcomes perfectly positively correlated

$$\mathrm{E}\left[\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right] = \operatorname{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$$

- Otherwise, $\widehat{\operatorname{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$ is positively biased (conservative)
- Potential outcomes will be perfectly positively correlated if and only if τ_i is the same for all $i = 1, \ldots, N$ units

Conservative variance estimator

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + 2\sigma_{y_C, y_T}\right)$$

Conservative variance estimator

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- The maximum possible value of $2\sigma_{y_C, y_T}$ is $\sigma^2_{y_C} + \sigma^2_{y_T}$
- So substitute $\sigma^2_{y_C} + \sigma^2_{y_T}$ for $2\sigma_{y_C, y_T}$:

Conservative variance estimator

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- The maximum possible value of $2\sigma_{y_C, y_T}$ is $\sigma^2_{y_C} + \sigma^2_{y_T}$
- So substitute $\sigma^2_{y_C} + \sigma^2_{y_T}$ for $2\sigma_{y_C, y_T}$:

$$\begin{aligned}\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] &= \frac{1}{N-1}\left(\frac{n_T \sigma^2_{y_C}}{n_C} + \frac{n_C \sigma^2_{y_T}}{n_T} + \left(\sigma^2_{y_C} + \sigma^2_{y_T}\right)\right) \\ &= \frac{N}{N-1}\left(\frac{\sigma^2_{y_C}}{n_C} + \frac{\sigma^2_{y_T}}{n_T}\right)\end{aligned}$$

Conservative variance estimator

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- The maximum possible value of $2\sigma_{y_C, y_T}$ is $\sigma_{y_C}^2 + \sigma_{y_T}^2$
- So substitute $\sigma_{y_C}^2 + \sigma_{y_T}^2$ for $2\sigma_{y_C, y_T}$:

$$\begin{aligned}\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] &= \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + \left(\sigma_{y_C}^2 + \sigma_{y_T}^2\right)\right) \\ &= \frac{N}{N-1}\left(\frac{\sigma_{y_C}^2}{n_C} + \frac{\sigma_{y_T}^2}{n_T}\right)\end{aligned}$$

- Now all quantities can be estimated!

Conservative variance estimator

$$\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + 2\sigma_{y_C, y_T}\right)$$

- The maximum possible value of $2\sigma_{y_C, y_T}$ is $\sigma_{y_C}^2 + \sigma_{y_T}^2$
- So substitute $\sigma_{y_C}^2 + \sigma_{y_T}^2$ for $2\sigma_{y_C, y_T}$:

$$\begin{aligned}\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] &= \frac{1}{N-1}\left(\frac{n_T \sigma_{y_C}^2}{n_C} + \frac{n_C \sigma_{y_T}^2}{n_T} + \left(\sigma_{y_C}^2 + \sigma_{y_T}^2\right)\right) \\ &= \frac{N}{N-1}\left(\frac{\sigma_{y_C}^2}{n_C} + \frac{\sigma_{y_T}^2}{n_T}\right)\end{aligned}$$

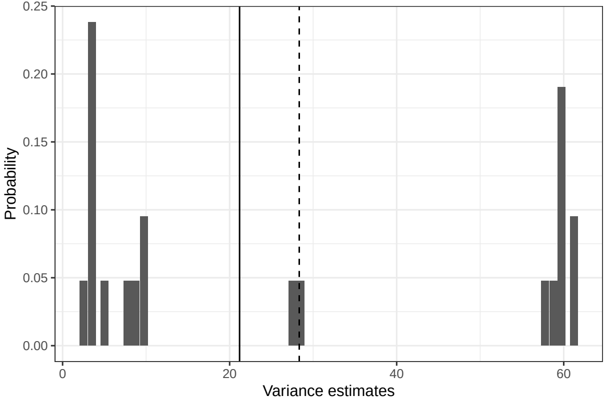
- Now all quantities can be estimated!
- So, just "plug-in" estimators $\hat{\sigma}_{y_C}^2$ and $\hat{\sigma}_{y_T}^2$ for $\sigma_{y_C}^2$ and $\sigma_{y_T}^2$

$$\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{N}{N-1}\left(\frac{\hat{\sigma}_{y_C}^2}{n_C} + \frac{\hat{\sigma}_{y_T}^2}{n_T}\right) \tag{1}$$

- For exact expressions of $\hat{\sigma}_{y_C}^2$ and $\hat{\sigma}_{y_T}^2$, see Variance estimators

Variance estimation

- Here is the conservative variance estimator in “village heads” example:



Solid line is true variance of Difference-in-Means estimator, $\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]$

Dashed line is expected value of conservative estimator, $\mathrm{E}\left[\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]\right]$

Asymptotic properties

Asymptotics

- So far, we have derived

    1. unbiased Difference-in-Means estimator of ATE
    2. variance of Difference-in-Means estimator
    3. conservative estimator of Difference-in-Means estimator's variance

Asymptotics

- So far, we have derived

  1. unbiased Difference-in-Means estimator of ATE
  2. variance of Difference-in-Means estimator
  3. conservative estimator of Difference-in-Means estimator's variance

- All of these derivations were for a fixed N, either small or large
- Now let's see what happens to our estimator when N grows large, $N \to \infty$
- But first, why do we care?

Asymptotics

- So far, we have derived

    1. unbiased Difference-in-Means estimator of ATE
    2. variance of Difference-in-Means estimator
    3. conservative estimator of Difference-in-Means estimator's variance

- All of these derivations were for a fixed N, either small or large
- Now let's see what happens to our estimator when N grows large, $N \to \infty$
- But first, why do we care?

  N never goes to ∞ in an actual experiment

  But properties as $N \to \infty$ **approximate** properties with fixed, but large N

Consistent estimation

Consistent estimation

- Difference-in-Means estimator is consistent:

Consistent estimation

- Difference-in-Means estimator is consistent:

$$\lim_{N \to \infty} \Pr\left(\left|\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau\right| < \epsilon\right) = 1 \text{ for all } \epsilon > 0$$

Consistent estimation

- Difference-in-Means estimator is consistent:

$$\lim_{N\to\infty} \Pr\left(\left|\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau\right| < \epsilon\right) = 1 \text{ for all } \epsilon > 0$$

- In words:

Pick any ϵ you want, no matter how small. There will be some value N^* such that, if size of experiment is greater than N^*, the probability of an estimate within a distance of ϵ from the truth is equal to 1.

Consistent estimation

- Difference-in-Means estimator is consistent:

$$\lim_{N\to\infty} \Pr\left(\left|\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\tau\right|<\epsilon\right)=1 \text{ for all } \epsilon>0$$

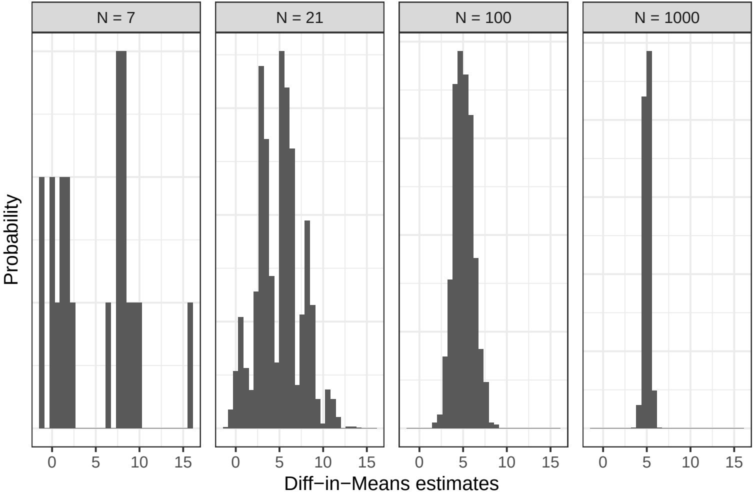
- In words:

  Pick any ϵ you want, no matter how small. There will be some value N^* such that, if size of experiment is greater than N^*, the probability of an estimate within a distance of ϵ from the truth is equal to 1.

- Intuitively, with large experiment, our estimate will be close to true ATE!

Consistent estimation

- "Village heads" example:



Hypothesis testing

Hypothesis tests of the weak null

- The finite population CLT tells us that

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

Hypothesis tests of the weak null

- The finite population CLT tells us that

$$\frac{\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y}) - \mathrm{E}[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})]}{\sqrt{\mathrm{Var}[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})]}} \xrightarrow{d} \mathcal{N}(0, 1)$$

- Diff-in-Means is unbiased, so write

$$\frac{\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y}) - \tau}{\sqrt{\mathrm{Var}[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})]}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Hypothesis tests of the weak null

- The finite population CLT tells us that

$$\frac{\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}}\xrightarrow{d}\mathcal{N}\left(0,1\right)$$

- Diff-in-Means is unbiased, so write

$$\frac{\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)-\tau}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z},\boldsymbol{Y}\right)\right]}}\xrightarrow{d}\mathcal{N}\left(0,1\right)$$

- The CLT is an asymptotic results as $N \to \infty$
- But we can bound error of Normal approximation for fixed N

Hypothesis tests of the weak null

- The finite population CLT tells us that

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

- Diff-in-Means is unbiased, so write

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

- The CLT is an asymptotic results as $N \to \infty$
- But we can bound error of Normal approximation for fixed N
- Thus, with experiments of at least moderate size and outcomes that aren't too skewed or have extreme outliers,

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \overset{\text{approx.}}{\sim} \mathcal{N}\left(0, 1\right)$$

Hypothesis tests of the weak null

- The finite population CLT tells us that

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \mathrm{E}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

- Diff-in-Means is unbiased, so write

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

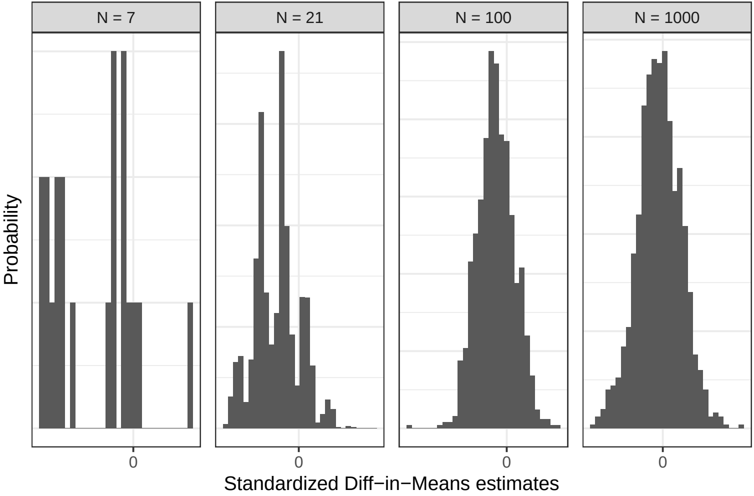
- The CLT is an asymptotic results as $N \to \infty$
- But we can bound error of Normal approximation for fixed N
- Thus, with experiments of at least moderate size and outcomes that aren't too skewed or have extreme outliers,

$$\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau}{\sqrt{\mathrm{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \overset{\text{approx.}}{\sim} \mathcal{N}\left(0, 1\right)$$

- This justifies use of standard Normal distribution for hypothesis tests

Hypothesis tests of the weak null

- "Village heads" example:



Hypothesis tests of the weak null

- To test null hypothesis relative to alternative

$$H_0 : \tau = \tau_0 \text{ versus either}$$
$$H_A : \tau > \tau_0,\ H_A : \tau < \tau_0 \text{ or } H_A : |\tau| > |\tau_0|$$

Hypothesis tests of the weak null

- To test null hypothesis relative to alternative

$$H_0 : \tau = \tau_0 \text{ versus either}$$

$$H_A : \tau > \tau_0,\ H_A : \tau < \tau_0 \text{ or } H_A : |\tau| > |\tau_0|$$

- Calculate upper(u), lower(l) or two-sided(t) p-value as

$$p_u = 1 - \Phi\left(\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}}\right)$$

$$p_l = \Phi\left(\frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}}\right)$$

$$p_t = 2\left(1 - \Phi\left(\frac{\left|\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0\right|}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}}\right)\right)$$

Hypothesis tests of the weak null

- To test null hypothesis relative to alternative

$$H_0 : \tau = \tau_0 \text{ versus either}$$
$$H_A : \tau > \tau_0,\ H_A : \tau < \tau_0 \text{ or } H_A : |\tau| > |\tau_0|$$

- Calculate upper(u), lower(l) or two-sided(t) p-value as

$$p_u = 1 - \Phi\left(\frac{\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y}) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})\right]}}\right)$$
$$p_l = \Phi\left(\frac{\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y}) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})\right]}}\right)$$
$$p_t = 2\left(1 - \Phi\left(\frac{|\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y}) - \tau_0|}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})\right]}}\right)\right)$$

- If p-value is less than size α-level of test, reject. Otherwise, don't
- Note that, since we don't know $\mathrm{Var}\left[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})\right]$, we have used its conservative estimator instead, $\widehat{\mathrm{Var}}\left[\hat{\tau}(\boldsymbol{Z}, \boldsymbol{Y})\right]$

Hypothesis tests of the weak null

Hypothesis tests susceptible to two errors:

- Type I error: Rejecting null hypothesis when it is true
- Type II error: *Failing* to reject null hypothesis when it is false

Hypothesis tests of the weak null

Hypothesis tests susceptible to two errors:

- Type I error: Rejecting null hypothesis when it is true
- Type II error: *Failing* to reject null hypothesis when it is false

A good test controls these errors:

1. Type I error probability is less than or equal to size (α-level) of test
2. Power (1 - type II error probability) is at least as great as α-level
3. Power tends to 1 as $N \to \infty$

Hypothesis tests of the weak null

- We can prove that tests of weak null satisfy (1) – (3) as $N \to \infty$
- Thus, when experiments are large, we can often safely use such tests
- But (1) – (3) may not always be satisfied when experiments are small, have skewed outcome distributions or extreme outliers

Confidence intervals

- Equivalence between hypothesis testing and confidence intervals
- Confidence interval is set of null hypotheses we fail to reject

Consider two-sided confidence interval, $\mathcal{C}_t$:

$$\begin{aligned}
\mathcal{C}_t &= \left\{ \tau_0 : \left| \frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \right| \leq z_{1-\alpha/2} \right\} \\
&= \left\{ \tau_0 : -z_{1-\alpha/2} \leq \frac{\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]}} \leq z_{1-\alpha/2} \right\} \\
&= \left\{ \tau_0 : -z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \leq \hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - \tau_0 \leq z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \right\} \\
&= \left\{ \tau_0 : -\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \leq -\tau_0 \leq -\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \right\} \\
&= \left\{ \tau_0 : \hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \geq \tau_0 \geq \hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \right\} \\
&= \left\{ \tau_0 : \hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \leq \tau_0 \leq \hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right) + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right]} \right\}
\end{aligned}$$

Appendix: Estimator of Difference-in-Means estimator's variance

Neyman's conservative estimator of the Difference-in-Means estimator's variance is

$$\widehat{\text{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{Y}\right)\right] = \frac{N}{N-1}\left(\frac{\hat{\sigma}^2_{y_C}}{n_C} + \frac{\hat{\sigma}^2_{y_T}}{n_T}\right),$$

where

$$\hat{\sigma}^2_{y_C} = \left(\frac{N-1}{n\left(n_C - 1\right)}\right) \sum_{i:Z_i=0}^{N} \left(y_{Ci} - \hat{\mu}_{y_C}\right)^2$$

$$\hat{\sigma}^2_{y_T} = \left(\frac{N-1}{n\left(n_T - 1\right)}\right) \sum_{i:Z_i=1}^{n} \left(y_{Ti} - \hat{\mu}_{y_T}\right)^2$$

$$\hat{\mu}_{y_C} = \left(\frac{1}{n_C}\right) \sum_{i=1}^{N} \left(1 - Z_i\right) y_{Ci}$$

$$\hat{\mu}_{y_T} = \left(\frac{1}{n_T}\right) Z_i y_{Ti}$$