Randomized Experiments

An Adventure in Nonparametric Inference

Randomized Experiments

Opening Comments

But *WHY?!?*

Nonparametrics and Causal Inference

Review

A Population Mean

ATE

$\widehat{\text{ATE}}$

$\text{Var}\left(\widehat{\text{ATE}}\right)$

$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$

Asymptotics

Best Practices

Example

Opening Comments

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments

This week, we continue applied modeling!

1. **engines**: design-based inference!
2. **distribution**: none!
3. **quantities of interest**: the ATE!
4. **evaluating models**: frequentist properties

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

But *WHY?!?*

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

But *WHY?!?*

1. These ideas are foundational (like SE), so it's good to see them again.
2. I think it's helpful for *me* to talk about what inference looks like without parametric models, so that the approach isn't confounded with the instructor.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Nonparametrics and Causal Inference

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Nonparametrics and Causal Inference

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Agnostic inference and causal inference tend to go together.

- I don't know the exact reason for this.
- I suspect it's not principled.
- I suspect it's because randomization (which gives us causal leverage) yeilds a really beautiful and and (largely) complete set of agnostic tools (i.e., tools that can be built only or mostly on the assumption of randomization).

Absent randomization, agnostic inference doesn't feel very strong, important, or interesting, because randomization is the only design-based assumption that feels under the researcher's control. (Random samples, parallel trends.)

But we can use an agnostic approach for descriptive (i.e., non-causal) quantities.

Review

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

- For discrete random variable X, $E(X) = \sum_{\text{all } x} x \cdot \Pr(X = x)$
- $\text{Var}(X) = E\left[(X - E(X))^2\right] = E\left[X^2\right] - E(X)^2$.

It's almost a defining feature of agnostic inference that that's lots of *very tedious* and *uninteresting* algebra. I've put it in these slides, so you can see it.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Our goal is to obtain point estimates, variance estimates, and confidence intervals from randomized experiments **using only the assumptions of the design**.

The notation here follows Imbens and Rubin (2015). An exception is that I use ATE to represent their τ_{fs} and $\widehat{\text{ATE}}$ to represent their $\hat{\tau}_{\text{dif}}$.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

A Population Mean

A Population Mean

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Before the ATE + causal inference, let's look at a *descriptive* problem, just so we can separate the two. Let's consider estimating a population mean **using only the assumption of a random sample**.

We follow Chapter 3 of Aronow and Miller (2019).[1] See the book for more details—they have a whole book about agnotistic inference that isn't motivated by causal quantities like the ATE.

[1] Aronow, Peter M., and Benjamin T. Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge University Press.

A Random Sample

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Suppose a population has mean μ and variance σ^2. We draw n units at random, with replacement, and record their values $Y_1, Y_2, ..., Y_n$.

Because we sample at random with replacement, $Y_1, Y_2, ..., Y_n$ are independent and identically distributed (i.i.d.), so that

$$E(Y_i) = \mu \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2 \quad \text{for all } i \in \{1, ..., n\}.$$

This is our only assumption. We assume nothing about the shape of the population distribution (e.g., normal or Poisson). The only randomness comes from the sampling, which *we* control. *(BUT DO WE?!?)*

We want three things for μ:

1. a point estimate,
2. a variance estimate, and
3. a confidence interval.

A Small Population

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Aronow and Miller (2019, Example 3.1.3) use a population of $N = 4$ units with values 3, 4, 3, and 10. Then

$$\mu = \frac{3+4+3+10}{4} = 5 \quad \text{and}$$
$$\sigma^2 = \frac{(3-5)^2+(4-5)^2+(3-5)^2+(10-5)^2}{4} = \frac{34}{4} = \frac{17}{2}.$$

Suppose we draw $n = 2$ units at random, with replacement. There are $4 \times 4 = 16$ equally likely samples. Here's $\overline{Y} = \frac{1}{2}(Y_1 + Y_2)$ for each.

	Second draw			
First draw	Unit 1 (3)	Unit 2 (4)	Unit 3 (3)	Unit 4 (10)
Unit 1 (3)	3	3.5	3	6.5
Unit 2 (4)	3.5	4	3.5	7
Unit 3 (3)	3	3.5	3	6.5
Unit 4 (10)	6.5	7	6.5	10

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Counting the cells, the sampling distribution of $\overline{Y}$ is

$\overline{y}$	3	3.5	4	6.5	7	10
$\Pr(\overline{Y} = \overline{y})$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Using the formulas from the review,

$$E(\overline{Y}) = 3 \cdot \frac{4}{16} + 3.5 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} + 6.5 \cdot \frac{4}{16} + 7 \cdot \frac{2}{16} + 10 \cdot \frac{1}{16} = \frac{80}{16} = 5$$

and

$$\text{Var}(\overline{Y}) = E(\overline{Y}^2) - E(\overline{Y})^2 = \frac{468}{16} - 5^2 = 4.25.$$

Notice that $E(\overline{Y}) = 5 = \mu$ and $\text{Var}(\overline{Y}) = 4.25 = \frac{17/2}{2} = \frac{\sigma^2}{n}$. The next two theorems show that this happens for *any* population.

Is $\overline{Y}$ a good estimator?

Theorem: $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is an unbiased estimator of μ. (Aronow and Miller 2019, Theorem 3.2.3)

Proof.

$$E(\overline{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\mu = \mu.$$

This uses only the linearity of expectation and $E(Y_i) = \mu$.

Theorem: The variance of $\overline{Y}$ is $\text{Var}(\overline{Y}) = \frac{\sigma^2}{n}$. (Aronow and Miller 2019, Theorem 3.2.4)

Proof.

$$\begin{aligned}
\text{Var}(\overline{Y}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(Y_i) \qquad \text{(because the } Y_i \text{ are independent)} \\
&= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Theorem (Weak Law of Large Numbers): $\overline{Y}$ is a consistent estimator of μ. That is, for any $\epsilon > 0$, $\Pr\left(|\overline{Y} - \mu| \geq \epsilon\right) \to 0$ as $n \to \infty$. (Aronow and Miller 2019, Theorem 3.2.8)

Proof. Chebyshev's inequality says that $\Pr\left(|Z - E(Z)| \geq \epsilon\right) \leq \frac{\text{Var}(Z)}{\epsilon^2}$ for any random variable Z with finite variance. Applying it to $\overline{Y}$,

$$\Pr\left(|\overline{Y} - \mu| \geq \epsilon\right) \leq \frac{\text{Var}(\overline{Y})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0.$$

Estimating $\text{Var}(\overline{Y})$

$\text{Var}(\overline{Y}) = \frac{\sigma^2}{n}$ depends on σ^2, which we do not know. But we can estimate σ^2 with the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

which gives the variance estimate

$$\widehat{\text{Var}}(\overline{Y}) = \frac{s^2}{n}.$$

Theorem: s^2 is an unbiased estimator of σ^2, so $\widehat{\text{Var}}(\overline{Y})$ is an unbiased estimator of $\text{Var}(\overline{Y})$. (Aronow and Miller 2019, Theorems 3.2.21 and 3.2.33)

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Proof. First, expand the sum of squares.

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + n\overline{Y}^2 = \sum_{i=1}^{n} Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2.$$

Second, rearrange $\text{Var}(Z) = E(Z^2) - E(Z)^2$ to get $E(Z^2) = \text{Var}(Z) + E(Z)^2$. Then $E(Y_i^2) = \sigma^2 + \mu^2$ and, using the two theorems above, $E(\overline{Y}^2) = \frac{\sigma^2}{n} + \mu^2$.

Lastly, take the expectation of the sum of squares.

$$E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2.$$

Dividing by $n-1$ shows that $E(s^2) = \sigma^2$. Dividing again by n shows that $E\left[\widehat{\text{Var}}(\overline{Y})\right] = \frac{\sigma^2}{n} = \text{Var}(\overline{Y})$.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

In the small population, $s^2 = \frac{1}{2-1}\left[(Y_1 - \overline{Y})^2 + (Y_2 - \overline{Y})^2\right]$ takes four values across the 16 samples.

Values drawn	s^2	Number of samples (of 16)
$\{3,3\}$, $\{4,4\}$, or $\{10,10\}$	0	6
$\{3,4\}$	0.5	4
$\{4,10\}$	18	2
$\{3,10\}$	24.5	4

So $E(s^2) = \frac{6\cdot 0 + 4\cdot 0.5 + 2\cdot 18 + 4\cdot 24.5}{16} = \frac{136}{16} = 8.5 = \frac{17}{2} = \sigma^2$, as the theorem says.

The standard error. We usually report $\sqrt{\widehat{\text{Var}}(\overline{Y})} = \frac{s}{\sqrt{n}}$, the estimated standard error.

- It is *not* unbiased. Because the square root is concave, Jensen's inequality gives $E\left(\sqrt{s^2/n}\right) \leq \sqrt{E(s^2/n)} = \sqrt{\text{Var}(\overline{Y})}$.
- But it is consistent. *Sketch.* The WLLN applies to the averages of Y_i^2 and Y_i, so $s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2\right) \to E(Y_i^2) - \mu^2 = \sigma^2$ in probability. The square root is continuous, so $s \to \sigma$ in probability as well. (Aronow and Miller 2019, Theorems 3.2.21 and 3.2.35)

A Confidence Interval

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Theorem (Central Limit Theorem): If $0 < \sigma^2 < \infty$, then as $n \to \infty$,

$$\frac{\overline{Y} - \mu}{\sqrt{\sigma^2/n}} \to N(0,1) \quad \text{in distribution.}$$

Proof. Omitted; Aronow and Miller (2019, Theorem 3.2.24) omit it as well.

Because $s^2 \to \sigma^2$ in probability, we can replace σ^2 with s^2 and keep the same limit (Slutsky's theorem; Aronow and Miller 2019, Theorem 3.2.25). Thus,

$$\Pr\left(-1.96 \leq \frac{\overline{Y} - \mu}{\sqrt{s^2/n}} \leq 1.96\right) \to 0.95.$$

Re-arranging the inequalities inside the probability, we have

$$\Pr\left(\overline{Y} - 1.96 \cdot \sqrt{s^2/n} \leq \mu \leq \overline{Y} + 1.96 \cdot \sqrt{s^2/n}\right) \to 0.95.$$

So a $(1-\alpha) \times 100\%$ CI is $\overline{Y} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\widehat{\text{Var}}(\overline{Y})}$. For a 95% CI, this would be $\overline{Y} \pm 1.96 \cdot \sqrt{\widehat{\text{Var}}(\overline{Y})}$. (Aronow and Miller 2019, Theorem 3.4.2 and Example 3.4.3)

How large is "large"?

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

The CI has coverage of $(1-\alpha)\times 100\%$ only *asymptotically*. In small samples, the coverage can be much lower.

In the small population with $n=2$, the 95% CI covers $\mu=5$ in only 6 of the 16 samples (37.5%):

- The 6 samples with $s^2=0$ give a CI of width zero, at 3, 4, or 10.
- The 4 samples of $\{3,4\}$ give $[2.52, 4.48]$.
- The 2 samples of $\{4,10\}$ give $[1.12, 12.88]$, and the 4 samples of $\{3,10\}$ give $[-0.36, 13.36]$. Only these 6 cover $\mu=5$.

Aronow and Miller (2019, Tables 3.4.1 and 3.4.2) simulate the coverage of the 95% CI (1 million simulations each).

n	Uniform(0, 1) population	Bernoulli(0.5) population
10	91.6%	89.0%
100	94.7%	94.3%
1,000	95.0%	94.7%

ATE

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

ATE

We have N units in our experiment and we assign N_t units to treatment and $N_c = N - N_t$ units to control. We require that $1 < N_t < N$, so that at least one unit is assigned to treatment and control.

We use the random variable W_i to assign unit i to treatment ($W_i = 1$) or control ($W_i = 0$).

For each unit i, we have two potential outcomes: $Y_i(1)$ if assigned to treatment and $Y_i(0)$ if assigned to control.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

For unit $i \in \{1, ..., N\}$, we have one observed potential outcome, denoted by Y_i^{obs}.

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases}$$

For each unit, we also have one missing potential outcome, denoted by Y_i^{mis}:

$$Y_i^{\text{mis}} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{if } W_i = 0, \\ Y_i(0) & \text{if } W_i = 1. \end{cases}$$

So what is this $Y_i(\cdot)$ thing?

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

A Data Set of Fully Observed Potential Outcomes

Unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	0	1	1
2	1	2	1
3	0	0	0
4	1	3	2
5	0	1	1
6	1	2	1
7	0	0	0
8	1	3	2
9	0	1	1
10	1	2	1

Table 1: Mock Dataset with Potential Outcomes and Their Differences

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Average Treatment Effect: Let's use these potential outcomes to define the average treatment effect (ATE) as
$\text{ATE} = \frac{1}{N}\sum_{i=1}^{N}\left[Y_i(1) - Y_i(0)\right] = \overline{Y}(1) - \overline{Y}(0)$, where $\overline{Y}(0)$ and $\overline{Y}(1)$ are the averages of the potential control and treated outcomes respectively, so that $\overline{Y}(0) = \frac{1}{N}\sum_{i=1}^{N} Y_i(0)$ and $\overline{Y}(1) = \frac{1}{N}\sum_{i=1}^{N} Y_i(1)$.

In our case, we are explicitly interested in what we call today the **sample average treatment effect** (SATE). That is, we are interested in the ATE among the units *in our sample* not an existing or imagined larger population.

Two Potential Realizations for the Experiment

Unit	$Y_i(0)$	$Y_i(1)$	W_i
1	?	1	1
2	1	?	0
3	?	0	1
4	1	?	0
5	0	?	0
6	?	2	1
7	0	?	0
8	?	3	1
9	0	?	0
10	?	2	1

Unit	$Y_i(0)$	$Y_i(1)$	W_i
1	0	?	0
2	?	2	1
3	0	?	0
4	?	3	1
5	?	1	1
6	1	?	0
7	?	0	1
8	1	?	0
9	?	1	1
10	1	?	0

Notice that we cannot compute the ATE from these data because we can never obtain $Y_i(1) - Y_i(0)$. (This is the fundamental problem of causal inference).

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

$\widehat{\text{ATE}}$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

An Estimator for the Average Treatment Effect

Suppose that we observe data from a completely randomized experiment in which $N_t = \sum_{i=1}^{N} W_i$ units are randomly selected to be assigned to treatment and the remaining $N_c = \sum_{i=1}^{N} (1 - W_i)$ are assigned to control. Because of the randomization, a natural[2] estimator for the average treatment effect is the difference in the average outcomes between those assigned to treatment and those assigned to control, so that

$$\widehat{\mathsf{ATE}} = \overline{Y}_t^{\mathsf{obs}} - \overline{Y}_c^{\mathsf{obs}},$$

where

$$\overline{Y}_c^{\mathsf{obs}} = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{\mathsf{obs}} \quad \text{and} \quad \overline{Y}_t^{\mathsf{obs}} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{\mathsf{obs}}.$$

Recall that $Y_i^{\mathsf{obs}} = Y_i(1)$ if $W_i = 1$ and $Y_i^{\mathsf{obs}} = Y_i(0)$ if $W_i = 0$. Thus we can write $\widehat{\mathsf{ATE}}$ as

$$\widehat{\mathsf{ATE}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1 - W_i) \cdot Y_i(0)}{N_c/N} \right).$$

[2] See next slide

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\mathsf{ATE}}$
$\mathsf{Var}\left(\widehat{\mathsf{ATE}}\right)$
$\widehat{\mathsf{Var}}\left(\widehat{\mathsf{ATE}}\right)$
Asymptotics
Best Practices
Example

Why $\widehat{\text{ATE}}$?

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Why is $\widehat{\text{ATE}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$ a "natural" estimator of the ATE?

For both intuitive and rigorous reasons, simply replacing population moments (e.g., the population mean) with sample moments (e.g., the sample mean) produces a good estimator.

Here's a sketch of the theory. For distributions F that are "not too weird" (i.e., meet fairly weak regularity conditions), then the empirical cdf $\hat{F}$ is a "good substitute" for F. It's a good substitute in the sense that $\hat{F}_N \to F$ as $N \to \infty$. This is an important "brute force" approach to nonparametric estimation.

In our case, we don't use it to defend the estimator, but to obtain a starting point.

What makes $\widehat{\text{ATE}}$ uncertain?

From our usual parametric modeling perspective, we *imagine* that the data are drawn from a distribution $y_i \sim f(\theta_i)$.

But this approach to randomized experiments is not like that; we're not imagining the $Y_i(1)$ and $Y_i(0)$ are random. Instead, they are fixed values.

The only thing that's random in the equation below is W_i, which is the assignment to treatment and control. Because we designed that mechanism, we know its distribution. Thus we have *design-based inference* rather than model-based inference.

$$\widehat{\text{ATE}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1 - W_i) \cdot Y_i(0)}{N_c/N} \right).$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Is $\widehat{\text{ATE}}$ a good estimator?

Theorem: $\widehat{\text{ATE}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$ is an unbiased estimator of ATE.

Proof.

We're trying to show that $E_W\left[\widehat{\text{ATE}}\right] = \text{ATE}$. *Walk through* $(2) \to (3)$ *together.*

$$E_W\left[\widehat{\text{ATE}}\right] = E_W\left[\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right] \tag{1}$$

$$= E_W\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1-W_i)\cdot Y_i(0)}{N_c/N}\right)\right] \tag{2}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(E_W[W_i]\cdot Y_i(1)\frac{1}{N_t/N} - E_W[1-W_i]\cdot Y_i(0)\frac{1}{N_c/N}\right) \tag{3}$$

$$= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \tag{4}$$

$$= \text{ATE} \tag{5}$$

Done! $\widehat{\text{ATE}}$ is an unbiased estimator of ATE under the assumptions of the design. No parametric model of the data needed, just a (known) model of the design.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

$$\text{Var}\left(\widehat{\text{ATE}}\right)$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Some Initial Results

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

To calculate the variance of $\widehat{\text{ATE}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$, we need $E_W[W_i^2]$ and $\text{Var}_W(W_i)$.

We have

$$E_W[W_i^2] = E_W[W_i] = \frac{N_t}{N} \quad \text{and} \quad \text{Var}_W(W_i) = \frac{N_t}{N}\left(1 - \frac{N_t}{N}\right).$$

Proof. Homework Exercise.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

We also need $E_W[W_i \cdot W_{i'}]$.

$$E_W[W_i \cdot W_{i'}] = P_W(W_i = 1) \cdot P_W(W_{i'} = 1 | W_i = 1) = \frac{N_t}{N} \cdot \frac{N_t - 1}{N - 1}, \quad \text{for } i \neq i',$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Theorem The variance of $\widehat{\text{ATE}}$ is

$$\text{Var}_W(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N},$$

where S_c^2 and S_t^2 are the variances of $Y_i(0)$ and $Y_i(1)$:

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(0) - \overline{Y}(0))^2, \quad \text{and} \quad S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(1) - \overline{Y}(1))^2,$$

and S_{tc}^2 is the variance of the unit-level treatment effects:

$$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - (\overline{Y}(1) - \overline{Y}(0))\right)^2.$$

Proof. Long and tedious! So here we go...

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Remember,

$$\widehat{\text{ATE}} = \frac{1}{N_t}\sum_{i=1}^{N} W_i \cdot Y_i^{\text{obs}} - \frac{1}{N_c}\sum_{i=1}^{N}(1-W_i)\cdot Y_i^{\text{obs}}.$$

Re-arranging, we have

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N}{N_t}\cdot W_i \cdot Y_i(1) - \frac{N}{N_c}\cdot(1-W_i)\cdot Y_i(0)\right).$$

It happens to be helpful to replace W_i with D_i.

$$D_i = W_i - \frac{N_t}{N} = \begin{cases} \frac{N_c}{N} & \text{if } W_i = 1, \\ -\frac{N_t}{N} & \text{if } W_i = 0. \end{cases}$$

Facts:

- $E(D_i) = 0$.
- $\text{Var}_W(D_i) = E[D_i^2] = \frac{N_t N_c}{N^2}$.

Proof. Homework exercise.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Another fact: For $i \neq j$ the distribution of $D_i \cdot D_j$ is

$$P_W(D_i \cdot D_j = d) = \begin{cases} \frac{N_t \cdot (N_t - 1)}{N \cdot (N-1)} & \text{if } d = \frac{N_c^2}{N^2}, \\ \frac{2 \cdot N_t \cdot N_c}{N \cdot (N-1)} & \text{if } d = -\frac{N_t N_c}{N^2}, \\ \frac{N_c \cdot (N_c - 1)}{N \cdot (N-1)} & \text{if } d = \frac{N_t^2}{N^2}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$E_W[D_i \cdot D_j] = \begin{cases} \frac{N_c \cdot N_t}{N^2} & \text{if } i = j, \\ -\frac{N_t \cdot N_c}{N^2 \cdot (N-1)} & \text{if } i \neq j. \end{cases}$$

Proof. Homework exercise.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Written using D_i rather than W_i, the estimate is:

$$\widehat{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{N}{N_t}\left(D_i + \frac{N_t}{N}\right)\cdot Y_i(1) - \frac{N}{N_c}\left(\frac{N_c}{N} - D_i\right)\cdot Y_i(0)\right].$$

Re-arranging, this becomes

$$= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) + \frac{1}{N}\sum_{i=1}^{N} D_i \cdot \left(\frac{N}{N_t}\cdot Y_i(1) + \frac{N}{N_c}\cdot Y_i(0)\right).$$

Exercise: Using the equation above, show that $\widehat{\text{ATE}}$ is an unbiased estimator of ATE.

$$\widehat{\mathsf{ATE}} = \overbrace{\frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))}^{\text{fixed}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N} D_i \cdot \overbrace{\left(\frac{N}{N_t}\cdot Y_i(1) + \frac{N}{N_c}\cdot Y_i(0)\right)}^{Y_i^+}}_{\text{random variable}}$$

Let $Y_i^+ = (N/N_t)Y_i(1) + (N/N_c)Y_i(0)$. Then

$$\mathsf{Var}_W(\widehat{\mathsf{ATE}}) = \mathsf{Var}_W\left(\frac{1}{N}\sum_{i=1}^{N} D_i \cdot Y_i^+\right) = \frac{1}{N^2} E_W\left[\left(\sum_{i=1}^{N} D_i \cdot Y_i^+\right)^2\right].$$

Expanding, we get:

$$\mathsf{Var}_W(\widehat{\mathsf{ATE}}) = E_W\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_i D_j Y_i^+ Y_j^+\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}(Y_i^+)^2 E_W[D_i^2] + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} E_W[D_i \cdot D_j]\cdot Y_i^+ \cdot Y_j^+$$

$$= \frac{N_c \cdot N_t}{N^4}\sum_{i=1}^{N}(Y_i^+)^2 - \frac{N_c \cdot N_t}{N^4\cdot(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} Y_i^+ \cdot Y_j^+$$

$$= \frac{N_c \cdot N_t}{N^4\cdot(N-1)}\left[(N-1)\sum_{i=1}^{N}(Y_i^+)^2 - \sum_{i=1}^{N}\sum_{j\neq i} Y_i^+ \cdot Y_j^+\right]$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\mathsf{ATE}}$
$\mathsf{Var}\left(\widehat{\mathsf{ATE}}\right)$
$\widehat{\mathsf{Var}}\left(\widehat{\mathsf{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

$$= \frac{N_c \cdot N_t}{N^3 \cdot (N-1)} \sum_{i=1}^{N} \left(Y_i^+ - \overline{Y}^+ \right)^2$$

$$= \frac{N_c \cdot N_t}{N^3 \cdot (N-1)} \sum_{i=1}^{N} \left(\frac{N}{N_t} \cdot Y_i(1) + \frac{N}{N_c} \cdot Y_i(0) - \left(\frac{N}{N_t} \cdot \overline{Y}(1) + \frac{N}{N_c} \cdot \overline{Y}(0) \right) \right)^2$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

$$
\begin{aligned}
&= \frac{N_c \cdot N_t}{N^3 \cdot (N-1)} \sum_{i=1}^{N} \left(\frac{N}{N_t} \cdot Y_i(1) - \frac{N}{N_t} \cdot \overline{Y}(1) \right)^2 \\
&\quad + \frac{N_c \cdot N_t}{N^3 \cdot (N-1)} \sum_{i=1}^{N} \left(\frac{N}{N_c} \cdot Y_i(0) - \frac{N}{N_c} \cdot \overline{Y}(0) \right)^2 \\
&\quad + \frac{2 \cdot N_c \cdot N_t}{N^3 \cdot (N-1)} \sum_{i=1}^{N} \left(\frac{N}{N_t} \cdot Y_i(1) - \frac{N}{N_t} \cdot \overline{Y}(1) \right) \cdot \left(\frac{N}{N_c} \cdot Y_i(0) - \frac{N}{N_c} \cdot \overline{Y}(0) \right) \\
&= \frac{N_c}{N \cdot N_t \cdot (N-1)} \sum_{i=1}^{N} (Y_i(1) - \overline{Y}(1))^2 + \frac{N_t}{N \cdot N_c \cdot (N-1)} \sum_{i=1}^{N} (Y_i(0) - \overline{Y}(0))^2 \\
&\quad + \frac{2}{N \cdot (N-1)} \sum_{i=1}^{N} (Y_i(1) - \overline{Y}(1)) \cdot (Y_i(0) - \overline{Y}(0)).
\end{aligned}
$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Recall the definition of S_{tc}^2, which implies that

$$\begin{aligned}
S_{tc}^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)-(Y_i(0)-\overline{Y}(0))\right)^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)^2 + \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0)-\overline{Y}(0)\right)^2 \\
&\quad - \frac{2}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)\cdot\left(Y_i(0)-\overline{Y}(0)\right) \\
&= S_t^2 + S_c^2 - \frac{2}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)\cdot\left(Y_i(0)-\overline{Y}(0)\right).
\end{aligned}$$

Hence, the expression in (Result 1) is equal to

$$\begin{aligned}
\text{Var}_W(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}) &= \frac{N_c}{N\cdot N_t}\cdot S_t^2 + \frac{N_t}{N\cdot N_c}\cdot S_c^2 \\
&\quad + \frac{1}{N}\left(S_t^2 + S_c^2 - S_{tc}^2\right) \\
&= \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{S_{tc}^2}{N}.
\end{aligned}$$

$$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$$

$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

For a given set of potential outcomes, we now know the variance. But to make inferences, we need to be able to *estimate* this variance using the observed data Y_i^{obs}.

An unbiased estimator for S_c^2 is

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i(0) - \overline{Y}_c^{\text{obs}}\right)^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2.$$

Similarly, an unbiased estimator for S_t^2 is

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i(1) - \overline{Y}_t^{\text{obs}}\right)^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2.$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

The third term, S_{tc}^2, is the tricky one. This is the population variance of the *unobservable* unit-level treatment effects—we cannot estimate this quantity.

We can only make assumptions about this quantity.

Assumption 1 (Sharp Null): If the treatment effect $Y_i(1) - Y_i(0)$ equals zero, then an unbiased estimator of the variance of $\widehat{\text{ATE}}$ is

$$\widehat{\text{Var}}(\widehat{\text{ATE}}) = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}. \qquad \text{(Neyman Variance Estimator)}$$

Assumption 2 (Constant Effects): If the treatment effect $Y_i(1) - Y_i(0)$ is constant, then an unbiased estimator of the variance of $\widehat{\text{ATE}}$ is also Neyman's variance estimator.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Without an Assumption: For any $Y_i(1)$ and $Y_i(0)$, the expectation of Neyman's variance estimator is *at least* as large as the actual variance.

$$\text{Var}_W(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}) = \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{S_{tc}^2}{N}.$$

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Asymptotics

A CI with Conservative Coverage Asymptotically

In Section 3 (pp. 1763-1763), Li and Ding (2017) provide a finite version of the CLT, so long as the potential outcomes are "not too weird."

A *conservative* $(1-\alpha)\times 100\%$ CI is $\widehat{\mathsf{ATE}} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\widehat{\mathsf{Var}}\left(\widehat{\mathsf{ATE}}\right)}$.

For a 95% confidence interval, this would be $\widehat{\mathsf{ATE}} \pm 1.96\cdot\sqrt{\widehat{\mathsf{Var}}\left(\widehat{\mathsf{ATE}}\right)}$.

Asymptotically, this interval has coverage of *at least* $(1-\alpha)\times 100\%$.

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\mathsf{ATE}}$
$\mathsf{Var}\left(\widehat{\mathsf{ATE}}\right)$
$\widehat{\mathsf{Var}}\left(\widehat{\mathsf{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Best Practices

Best Practices

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Estimation

- It turns out that OLS with robust standard errors (HC2) is equivalent to the difference-in-means and Neyman variance estimator we proposed above. See Samii and Aronow (2012).
- In practice, we obtain estimates with something like the following:

```
fit <- lm(y ~ treatment_indicator, data = df)
V_hat <- sandwich::vcovHC(fit, type = "HC2")
```

Adjustment

- Some authors are critical of regression adjustment for experimental data (Freedman 2008).
- However, in my view the dangers are mostly theoretical. In real world problems, regression adjustment will almost always give you better estimates, sometimes *much* better estimates
  - Simply include controls.
  - Include mean-centered controls fully interacted with treatment.
  - Select controls with LASSO.

Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
$\text{Var}\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Example

Here's the packages we're using.

```
# generally useful packages
library(tidyverse)
library(modelsummary)
```

The Coversight Experiment

```
# load packages
library(tidyverse)

# read cleaned data from dropbox
coversight <- read_rds("https://www.dropbox.com/s/9grn8kkb5yzwagx/data.rds?raw=1")%>%
  select(pres_overall, amplify, failure, pid_strength, passed_mvc1) %>%
  filter(passed_mvc1) %>%  # remove respondents who failed MVC 1
  mutate(pid_strength_rs = arm::rescale(pid_strength)) # center partisan strength

# quick look
glimpse(coversight)
```

```
Rows: 1,462
Columns: 6
$ pres_overall    <dbl> 0, 1, -3, 0, 2, 3, 2, -3, -3, 0, -~
$ amplify         <fct> Amplify, Ignore, Amplify, Ignore, ~
$ failure         <fct> Failure, Failure, Failure, Failure~
$ pid_strength    <dbl> 3, 3, 3, 0, 2, 2, 3, 3, 2, 3, 3, 3~
$ passed_mvc1     <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE~
$ pid_strength_rs <dbl> 0.5027, 0.5027, 0.5027, -0.9731, 0~
```


Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

```r
# use only data from amplify condition (failure/success varies)
cs_failure <- coversight %>%
  filter(amplify == "Amplify")

table(cs_failure$failure)
```

```
Success Failure
    366     364
```

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

```r
# make y for convenience
y <- cs_failure$pres_overall
treat <- as.numeric(cs_failure$failure == "Failure")

# compute point estimate
mean(y[treat == 1]) - mean(y[treat == 0])
```

```
[1] -1.23
```

```r
# compute variance estimate
var(y[treat == 1])/sum(treat) + var(y[treat == 0])/sum(1 - treat)
```

```
[1] 0.0147
```

```r
# ols with robust ses
fit <- lm(y ~ treat)
coef(fit)["treat"]
```

```
treat
-1.23
```

```r
diag(sandwich::vcovHC(fit, type = "HC2"))["treat"]
```

```
 treat
0.0147
```

Opening Comments
But *WHY?!?*
Nonparametrics and Causal Inference
Review
A Population Mean
ATE
$\widehat{\text{ATE}}$
Var $\left(\widehat{\text{ATE}}\right)$
$\widehat{\text{Var}}\left(\widehat{\text{ATE}}\right)$
Asymptotics
Best Practices
Example

```r
fit1 <- lm(pres_overall ~ failure, data = cs_failure)
fit2 <- lm(pres_overall ~ failure + I(pid_strength_rs + 100), data = cs_failure)
fit3 <- lm(pres_overall ~ failure*pid_strength_rs, data = cs_failure)

modelsummary(list("No Controls" = fit1,
                  "One Control" = fit2,
                  "One Control, Interacted" = fit3),
             vcov = "HC2",
             gof_map = NA,
             escape = TRUE,
             output = "latex")
```

	No Controls	One Control	One Control, Interacted
(Intercept)	−0.650	17.973	−0.657
	(0.099)	(12.081)	(0.099)
failureFailure	−1.234	−1.224	−1.225
	(0.121)	(0.121)	(0.121)
I(pid_strength_rs + 100)		−0.186	
		(0.121)	
pid_strength_rs			−0.156
			(0.199)
failureFailure × pid_strength_rs			−0.061
			(0.241)