Potential Outcomes and Randomized Experiments

Ryan T. Moore

23 July 2019

Preliminaries

Preliminaries

- PS1 for today
- PS2 for tomorrow
- `.Rmd`

		Treatment Assignment	
		Randomized	Not Randomized
Unit Selection	Randomized	Randomized Experiment (gold standard)	Survey Sampling (allows population inference)
	Not Randomized	Controlled Experiment (allows causal inference)	Observational Study (large potential for bias)

Figure 1: Ramsey & Schafer, *The Statistical Sleuth*

Review Questions

1. When can we observe both $Y_i(1)$ and $Y_i(0)$?
2. What is this fact called?
3. When does the empirical difference-in-means estimator exactly equal the true, underlying ATE?
4. What are the two parts of SUTVA?

Exercise in Potential Outcomes

Statistical Refreshment

Expectation

$$E(Y) = \sum_{i=1}^{n} [y_i \cdot p(y_i)]$$

Expectation

$$E(Y) = \sum_{i=1}^{n} [y_i \cdot p(y_i)]$$

Calculate $E(\text{Ideology})$.

Respondent	Ideology
1	3
2	-2
3	3
4	1
5	1

Conditional Expectation

$$E(Y|X=x) = \sum_{i=1}^{n} [y_i \cdot p(y_i|X=x)]$$

Conditional Expectation

$$E(Y|X = x) = \sum_{i=1}^{n}[y_i \cdot p(y_i|X = x)]$$

Calculate $E(\text{Ideology}|\text{Dem} = 0)$.

Respondent	Ideology	Dem
1	3	0
2	-2	1
3	3	0
4	1	0
5	1	1

Conditional Expectation

$$E(Y|X = x) = \sum_{i=1}^{n}[y_i \cdot p(y_i|X = x)]$$

Calculate $E(\text{Ideology}|\text{Dem} = 0)$ – wide data, as counts.

		Dem?	
		0	1
	3	2	0
Ideology	1	1	1
	-2	0	1

Conditional Expectation

$$E(Y|X=x) = \sum_{i=1}^{n} [y_i \cdot p(y_i|X=x)]$$

Calculate $E(\text{Ideology}|\text{Dem} = 0)$ – wide data, as probs.

		Dem?	
		0	1
Ideology	3	0.4	0.0
	1	0.2	0.2
	-2	0.0	0.2

Unbiasedness

$$E(\hat{\theta}) = \theta$$

Unbiasedness

$$E(\hat{\theta}) = \theta$$

- $E\left[\widehat{\overline{Y_1 - Y_0}}\right] = \overline{Y_1 - Y_0}$

Unbiasedness

$$E(\hat{\theta}) = \theta$$

- $E\left[\widehat{\overline{Y_1 - Y_0}}\right] = \overline{Y_1 - Y_0}$
- $E\left[\widehat{\beta_1}\right] = \beta_1$

Unbiasedness

What estimator used to estimate $\overline{Y_1 - Y_0}$?

Unbiasedness

What estimator used to estimate $\overline{Y_1 - Y_0}$?

The difference-in-means estimator.

Unbiasedness

What estimator used to estimate $\overline{Y_1 - Y_0}$?

The difference-in-means estimator.

$$\widehat{\overline{Y_1 - Y_0}} = \left(\widehat{Y_1} | T_i = 1\right) - \left(\widehat{Y_0} | T_i = 0\right)$$

Probability

The Three Axioms

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. If events *mutually exclusive* (or, sets *disjoint*), then

$$P(A \text{ or } B) = P(A) + P(B)$$

Conditional Probability

For events A and B,

$$\begin{aligned} P(A \text{ and } B) &= P(A)P(B|A) \\ &= P(B)P(A|B) \end{aligned}$$

Divide both sides by marginal probability $P(B)$ yields

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Statistical Independence

A and B are independent iff both

$$P(A|B) = P(A)$$

and

$$P(B|A) = P(B)$$

Conditional Independence

A and B can be independent *conditional on* C.

Conditional Independence

A and B can be independent *conditional on* C.

Many times, *only* independent given C.

Conditional Independence

A and B can be independent *conditional on* C.

Many times, *only* independent given C.

E.g., Y_1, Y_0 indep of T, given X.

Conditional Independence

A and B can be independent *conditional on* C.

Many times, *only* independent given C.

E.g., Y_1, Y_0 indep of T, given X.

$$P(A \text{ and } B|C) = P(A|C)P(B|C)$$

Potential Outcomes Model: Estimands

- Individual TE
$$\tau_i = Y_i(1) - Y_i(0)$$
- Average treatment effect
$$ATE = E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$$
- Average treatment effect for the treated
$$ATT = E(Y_1|T = 1) - E(Y_0|T = 1)$$

Potential Outcomes Model: Estimands

The average treatment effect (ATE):

$$
\begin{aligned}
E(Y_1 - Y_0) &= \frac{1}{2n}\sum_{i=1}^{2n}(Y_{i1} - Y_{i0}) \\
&= \frac{1}{2n}\sum_{i=1}^{2n}(Y_{i1}) - \frac{1}{2n}\sum_{i=1}^{2n}(Y_{i0}) \\
&= E(Y_1) - E(Y_0)
\end{aligned}
$$

Potential Outcomes Model: Estimands

- If we know (Y_1, Y_0) indep of T
- Then,

$$
\begin{aligned}
E(Y_1) &= E(Y_1|T=1) \\
E(Y_0) &= E(Y_0|T=0)
\end{aligned}
$$

- Then, can substitute

$$
\begin{aligned}
ATE &= E(Y_1) - E(Y_0) \\
&= E(Y_1|T=1) - E(Y_0|T=0)
\end{aligned}
$$

Potential Outcomes Model: Estimands

- If we know (Y_1, Y_0) indep of T
- Then,

$$\begin{aligned} E(Y_1) &= E(Y_1|T=1) \\ E(Y_0) &= E(Y_0|T=0) \end{aligned}$$

- Then, can substitute

$$\begin{aligned} ATE &= E(Y_1) - E(Y_0) \\ &= E(Y_1|T=1) - E(Y_0|T=0) \end{aligned}$$

Observed diff in Tr and Co group means gives ATE!

Potential Outcomes Model: Estimands

Holland (1986): "*prima facie effect*":

$$E(Y_t|S = t) - E(Y_c|S = c)$$

Potential Outcomes Model: Estimands

Holland (1986): "*prima facie effect*":

$$E(Y_t|S = t) - E(Y_c|S = c)$$

"It is important to recognize that $E(Y_t)$ and $E(Y_t|S = t)$ are *not* the same thing ..."

Potential Outcomes Model: Estimands, Interpretation

Gerber & Green:

When $Y(1)$ and $Y(0)$ indep of T,

$$ATE \quad = \quad E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 0)$$

Potential Outcomes Model: Estimands, Interpretation

Gerber & Green:

When $Y(1)$ and $Y(0)$ indep of T,

$$\begin{aligned} ATE &= E(Y_i(1)|T_i=1) - E(Y_i(0)|T_i=0) \\ &= E(Y_i(1)|T_i=1) - E(Y_i(0)|T_i=1) + \\ &\qquad E(Y_i(0)|T_i=1) - E(Y_i(0)|T_i=0) \end{aligned}$$

Potential Outcomes Model: Estimands, Interpretation

Gerber & Green:

When $Y(1)$ and $Y(0)$ indep of T,

$$
\begin{aligned}
ATE &= E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 0) \\
&= E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1) + \\
&\quad\; E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0) \\
&= [E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1)] + \\
&\quad\; [E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0)]
\end{aligned}
$$

Potential Outcomes Model: Estimands, Interpretation

Gerber & Green:

When $Y(1)$ and $Y(0)$ indep of T,

$$
\begin{aligned}
ATE &= E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 0) \\
&= E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1) + \\
&\quad E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0) \\
&= [E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1)] + \\
&\quad [E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0)] \\
&= E(Y_i(1) - Y_i(0)|T_i = 1) + \\
&\quad E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0)
\end{aligned}
$$

Potential Outcomes Model: Estimands, Interpretation

Gerber & Green:

When $Y(1)$ and $Y(0)$ indep of T,

$$
\begin{aligned}
ATE &= E(Y_i(1)|T_i=1) - E(Y_i(0)|T_i=0) \\
&= E(Y_i(1)|T_i=1) - E(Y_i(0)|T_i=1) + \\
&\quad E(Y_i(0)|T_i=1) - E(Y_i(0)|T_i=0) \\
&= [E(Y_i(1)|T_i=1) - E(Y_i(0)|T_i=1)] + \\
&\quad [E(Y_i(0)|T_i=1) - E(Y_i(0)|T_i=0)] \\
&= E(Y_i(1) - Y_i(0)|T_i=1) + \\
&\quad E(Y_i(0)|T_i=1) - E(Y_i(0)|T_i=0)
\end{aligned}
$$

$$
\underbrace{E(Y_i(1) - Y_i(0)|T_i=1)}_{\text{ATT}} + \underbrace{E(Y_i(0)|T_i=1) - E(Y_i(0)|T_i=0)}_{\text{Selection Bias}}
$$

Assignment of Treatment

Ideas

- Causal effects: relative to some other condition

Ideas

- Causal effects: relative to some other condition
- Treatment

Ideas

- Causal effects: relative to some other condition
- Treatment
  - timing defined

Ideas

- Causal effects: relative to some other condition
- Treatment
  - timing defined
  - must be excludable

Ideas

- Causal effects: relative to some other condition
- Treatment
  - timing defined
  - must be excludable
  - cannot be "attribute"

Ideas

- Causal effects: relative to some other condition
- Treatment
  - timing defined
  - must be excludable
  - cannot be "attribute"
- Covariates: causally prior to treatment

Ideas

- Causal effects: relative to some other condition
- Treatment
  - timing defined
  - must be excludable
  - cannot be "attribute"
- Covariates: causally prior to treatment
- Dose-response "biological gradient" evidence

Attributes

"Causal effect of race"?

```r
resume <- read_csv("http://j.mp/2sDjsHI")
dim(resume)
```

```
## [1] 4870    4
```

Attributes

"Causal effect of race"?

```
resume <- read_csv("http://j.mp/2sDjsHI")
dim(resume)
```

```
## [1] 4870    4
```

```
kable(table(resume$race, resume$call))
```

	0	1
black	2278	157
white	2200	235

Attributes

"Causal effect of race"?

```
resume <- read_csv("http://j.mp/2sDjsHI")
dim(resume)
```

```
## [1] 4870    4
```

```
kable(table(resume$race, resume$call))
```

	0	1
black	2278	157
white	2200	235

Sen and Wasow (2016) : "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics" (*elements* of attributes varyingly manipulable)

The Potential Outcomes Model: Assignment

Observed outcome: $Y_i = Y_i(1) \cdot T_i + Y_i(0) \cdot (1 - T_i)$

The assignment mechanism *selects* which potential outcome we observe.

The Potential Outcomes Model: Assignment

Observed outcome: $Y_i = Y_i(1) \cdot T_i + Y_i(0) \cdot (1 - T_i)$

The assignment mechanism *selects* which potential outcome we observe.

(Gerber & Green use $Y_i = Y_i(1) \cdot d_i + Y_i(0) \cdot (1 - d_i)$ to highlight that we observe pot outcome from treatment actually taken, not hypothetical or assigned treatment.)

The Potential Outcomes Model: Assignment

"Assignment mechansims" are really missing-data-generating procedures.

Ignorability

Assignment mechanism is *ignorable* if Y_{obs} conditnly indep of T

$$P(T|X, Y_{obs}, Y_{mis}) = P(T|X, Y_{obs})$$

Ignorability

Assignment mechanism is *ignorable* if Y_{obs} conditnly indep of T

$$P(T|X, Y_{obs}, Y_{mis}) = P(T|X, Y_{obs})$$

Nothing in unobserved Y_{mis} informs relationship between Y_{obs}, T.

Unconfoundedness

Some ignorable mechanisms are *unconfounded*, too.

$$P(T|X, Y_{obs}, Y_{mis}) = P(T|X)$$

Unconfoundedness

Some ignorable mechanisms are *unconfounded*, too.

$$P(T|X, Y_{obs}, Y_{mis}) = P(T|X)$$

Nothing in Y informs T.

Unconfoundedness

Some ignorable mechanisms are *unconfounded*, too.

$$P(T|X, Y_{obs}, Y_{mis}) = P(T|X)$$

Nothing in Y informs T.

These are special cases of conditional independence.

An Assignment Mechanism

Little & Rubin (2000):

Patient	Y	T
1	6	1
2	12	1
3	9	0
4	11	0

An Assignment Mechanism

Little & Rubin (2000):

Patient	Y	T
1	6	1
2	12	1
3	9	0
4	11	0

Clearly, treatment is harmful. $\overline{Y(1)} - \overline{Y(0)} = 9 - 10 = -1$.

An Assignment Mechanism

Little & Rubin (2000):

Patient	Y(0)	Y(1)	τ	T
1		6		1
2		12		1
3	9			0
4	11			0
Mean	10	9		

Clearly, treatment is harmful.

$$\overline{Y(1)|T=1} - \overline{Y(0)|T=0} = 9 - 10 = -1$$

An Assignment Mechanism: Perfect Doctor

Little & Rubin (2000):

Patient	Y(0)	Y(1)	τ	T
1	(1)	6		1
2	(3)	12		1
3	9	(8)		0
4	11	(10)		0
Mean	10	9		

An Assignment Mechanism: Perfect Doctor

Little & Rubin (2000):

Patient	Y(0)	Y(1)	τ	T
1	(1)	6	(5)	1
2	(3)	12	(9)	1
3	9	(8)	(-1)	0
4	11	(10)	(-1)	0
Mean	10	9	(3)	

An Assignment Mechanism: Perfect Doctor

Little & Rubin (2000):

Patient	Y(0)	Y(1)	τ	T
1	(1)	6	(5)	1
2	(3)	12	(9)	1
3	9	(8)	(-1)	0
4	11	(10)	(-1)	0
Mean	10	9	(3)	

Clearly, treatment is beneficial:

$$\overline{Y(1)} - \overline{Y(0)} = 9 - 6 = 3$$

An Assignment Mechanism: Perfect Doctor

Little & Rubin (2000):

Patient	Y(0)	Y(1)	τ	T
1	(1)	6	(5)	1
2	(3)	12	(9)	1
3	9	(8)	(-1)	0
4	11	(10)	(-1)	0
Mean	10	9	(3)	

Clearly, treatment is beneficial:

$$\overline{Y(1)} - \overline{Y(0)} = 9 - 6 = 3$$

This assg mechanism is non-ignorable, confounded.

Random Assignment Mechanisms

- Simple/complete randomization
  (Bernoulli trial, prob π)

Random Assignment Mechanisms

- Simple/complete randomization
  (Bernoulli trial, prob π)
- Complete randomization / random allocation
  (fixed proportion to tr)

Random Assignment Mechanisms

- Simple/complete randomization
  (Bernoulli trial, prob π)
- Complete randomization / random allocation
  (fixed proportion to tr)
- Blocked randomizations
  (fixed proportion to tr, w/in group)

Random Assignment Mechanisms

- Simple/complete randomization
  (Bernoulli trial, prob π)
- Complete randomization / random allocation
  (fixed proportion to tr)
- Blocked randomizations
  (fixed proportion to tr, w/in group)
- Cluster randomizations
  (assignment at higher level)

Randomization (Design-based) Inference

A volunteer?

A volunteer?

The task: select the 2 folders with messages

A volunteer?

The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?

A volunteer?

The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?
  - "No x-ray vision. No ESP. Effect of messages on choice = 0."

A volunteer?

The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?
  - "No x-ray vision. No ESP. Effect of messages on choice = 0."
- What is an alternative?

A volunteer?

The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?
  - "No x-ray vision. No ESP. Effect of messages on choice $= 0$."
- What is an alternative?
  - "Some way to detect messages. Message location $\rightarrow$ choice."

A volunteer?

The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?
  - "No x-ray vision. No ESP. Effect of messages on choice $= 0$."
- What is an alternative?
  - "Some way to detect messages. Message location $\rightarrow$ choice."

A volunteer?

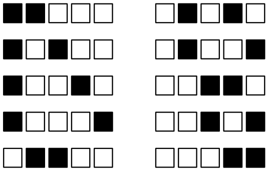
The task: select the 2 folders with messages

- What is our baseline expectation/model for this process?
  - "No x-ray vision. No ESP. Effect of messages on choice $= 0$."
- What is an alternative?
  - "Some way to detect messages. Message location $\rightarrow$ choice."

Select!

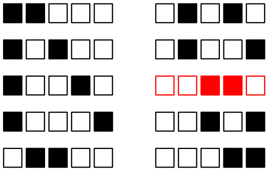
Randomization Inference

The possible choices:



Randomization Inference

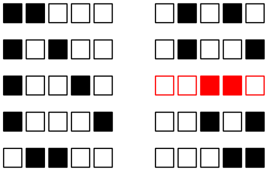
The possible choices:



- You chose _____ and _____. Let X = number found.

Randomization Inference

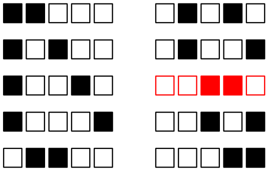
The possible choices:



- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2 | \text{no ESP})$?

Randomization Inference

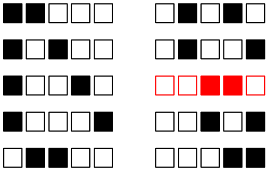
The possible choices:



- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2|\text{no ESP})$? $\frac{1}{10} = 0.1$

Randomization Inference

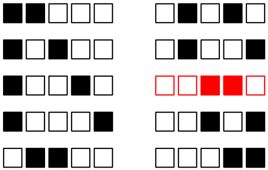
The possible choices:



- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2|\text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1|\text{no ESP})$?

Randomization Inference

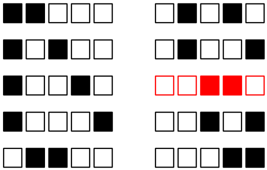
The possible choices:



- You chose ____ and ____. Let X = number found.
- What was $P(X \geq 2|\text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1|\text{no ESP})$? $\frac{7}{10} = 0.7$

Randomization Inference

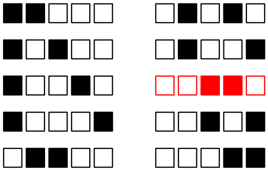
The possible choices:



- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2|\text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1|\text{no ESP})$? $\frac{7}{10} = 0.7$
- What is "prob result at least this extreme, given model of no effect"?

Randomization Inference

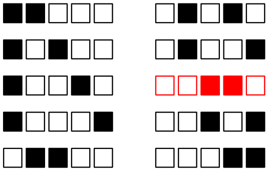
The possible choices:



- You chose ____ and ____. Let X = number found.
- What was $P(X \geq 2|\text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1|\text{no ESP})$? $\frac{7}{10} = 0.7$
- What is "prob result at least this extreme, given model of no effect"?
- Definition of p-value!

Randomization Inference

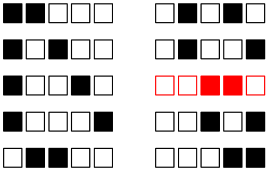
The possible choices:



- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2 | \text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1 | \text{no ESP})$? $\frac{7}{10} = 0.7$
- What is "prob result at least this extreme, given model of no effect"?
- Definition of p-value!
- Valid, exact, with no distributional assumption, no large n.

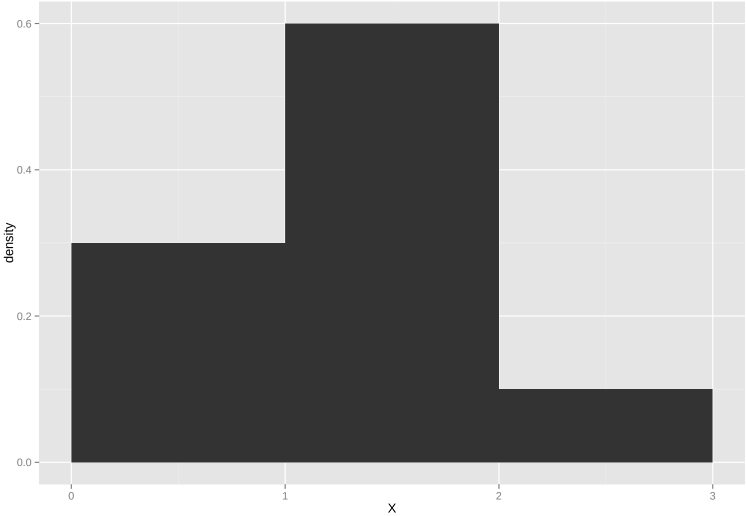
Randomization Inference

The possible choices:

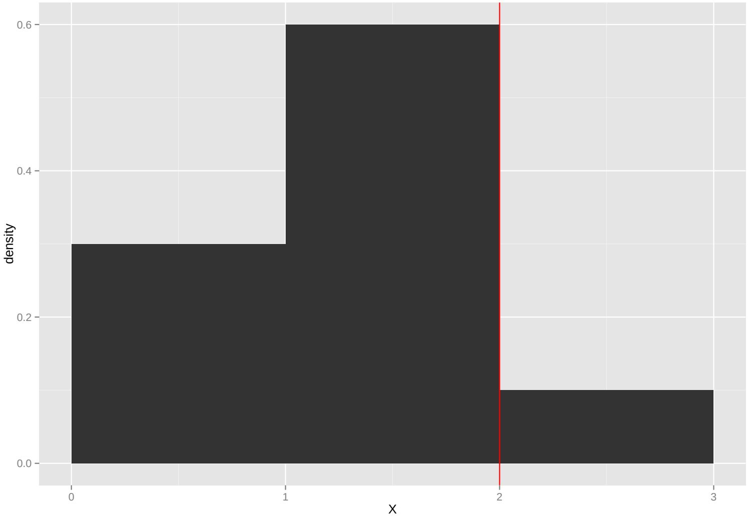


- You chose _____ and _____. Let X = number found.
- What was $P(X \geq 2 | \text{no ESP})$? $\frac{1}{10} = 0.1$
- What was $P(X \geq 1 | \text{no ESP})$? $\frac{7}{10} = 0.7$
- What is "prob result at least this extreme, given model of no effect"?
- Definition of p-value!
- Valid, exact, with no distributional assumption, no large n.
- *Randomization* creates dist'n of possible numbers correct

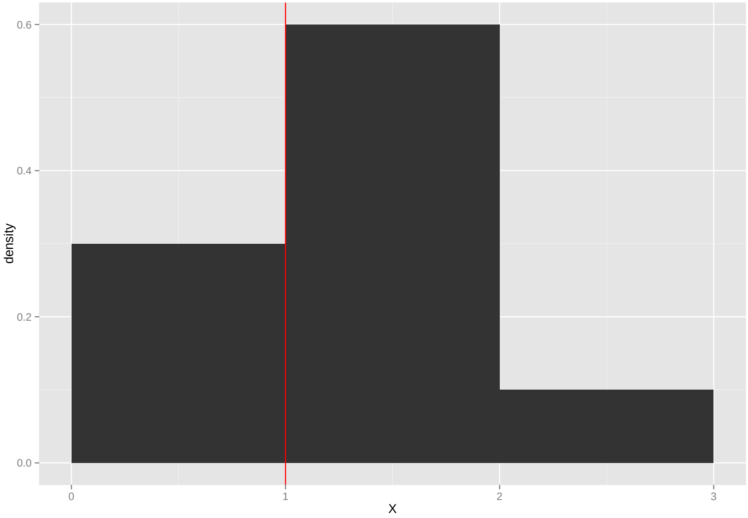
The Randomization Distribution of X



The Randomization Distribution of X



The Randomization Distribution of X



Parametric Null Hypothesis Significance Testing

- Specify and assume H_0
- Define H_A
- Examine reference dist'n (t, χ^2, ...) under H_0
- Calculate p-value
- Compare to some α; reject H_0 if $p < \alpha$

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)
- Define H_A

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)
- Define H_A
- Create reference dist'n from all possible values of X under H_0
  (or at least a big sample of them)

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)
- Define H_A
- Create reference dist'n from all possible values of X under H_0
  (or at least a big sample of them)
- What prop. of possible "at least as extreme as" observed?
  $\rightsquigarrow$ p-value!

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)
- Define H_A
- Create reference dist'n from all possible values of X under H_0
  (or at least a big sample of them)
- What prop. of possible "at least as extreme as" observed?
  $\rightsquigarrow$ p-value!
- Compare to some α; reject H_0 if $p < \alpha$

Randomization Inference

- Specify and assume H_0
  (sharp null of no treatment effect)
- Define H_A
- Create reference dist'n from all possible values of X under H_0
  (or at least a big sample of them)
- What prop. of possible "at least as extreme as" observed?
  $\rightsquigarrow$ p-value!
- Compare to some α; reject H_0 if $p < \alpha$
- CA ballot ordering effects (JASA 2006)

Randomization Inference

The RI p-value is

$$p = \frac{\#\text{ outcomes } \geq \text{as extreme as obs}}{\text{total }\#\text{ outcomes}}$$

Randomization Inference

The RI p-value is

$$p = \frac{\#\text{ outcomes } \geq \text{as extreme as obs}}{\text{total }\#\text{ outcomes}}$$

or

$$p = \frac{\#\text{ randomizations producing extreme } \widehat{ATE}}{\text{total }\#\text{ randomizations}}$$

Randomization Inference

The RI p-value is

$$p = \frac{\#\text{ outcomes } \geq \text{as extreme as obs}}{\text{total }\#\text{ outcomes}}$$

or

$$p = \frac{\#\text{ randomizations producing extreme } \widehat{ATE}}{\text{total }\#\text{ randomizations}}$$

How many randomizations are there?

Combinations: Counting selected sets

How many ways to **select** k things from a set of n things?

$$ {}_nC_k = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} $$

Combinations: Counting selected sets

How many ways to **select** k things from a set of n things?

$$ {}_nC_k = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} $$

How many ways to choose 5 villages of 10 for treatment?

Combinations: Counting selected sets

How many ways to **select** k things from a set of n things?

$$ {}_nC_k = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} $$

How many ways to choose 5 villages of 10 for treatment?

$$ {}_{10}C_5 = \binom{10}{5} = \frac{10!}{5!(10-5)!} $$

Combinations: Counting selected sets

How many ways to **select** k things from a set of n things?

$$_nC_k = \binom{n}{k} = \frac{_nP_k}{k!} = \frac{n!}{k!(n-k)!}$$

How many ways to choose 5 villages of 10 for treatment?

$$_{10}C_5 = \binom{10}{5} = \frac{10!}{5!(10-5)!}$$

$$\frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 252$$

Common Assumptions, Null Hypotheses

- Constant effect:

$$\tau_i = Y_{i1} - Y_{i0} = \tau \quad \forall i$$

- Null hypothesis of no average effect:

$$ATE = \overline{\tau} = 0$$

- Sharp null hypothesis of no effect:

$$\tau_i = 0$$

An Assignment Mechanism: Perfect Doctor

Calculate RI p-value for Perfect Doctor, under sharp null.

Patient	Y(0)	Y(1)	τ	T
1	(1)	6	(5)	1
2	(3)	12	(9)	1
3	9	(8)	(-1)	0
4	11	(10)	(-1)	0
Mean	10	9	(3)	

An Assignment Mechanism: Perfect Doctor

Calculate RI p-value for Perfect Doctor, under sharp null.

Patient	Y(0)	Y(1)	τ	T
1	(1)	6	(5)	1
2	(3)	12	(9)	1
3	9	(8)	(-1)	0
4	11	(10)	(-1)	0
Mean	10	9	(3)	

(See `02-ri-perfect-dr.R`)

RI versus the t-test

Perfect Doctor:

- RI: $p = 1$
- `t.test()`: $p \approx 0.8$
- "If no tr effect, then this result typical"

RI versus the t-test

Perfect Doctor:

- RI: $p = 1$
- `t.test()`: $p \approx 0.8$
- "If no tr effect, then this result typical"

(Odd logic of NHST: "assume false thing, how strange is data?"")

Randomization Inference

- Resume audit study, Bertrand and Mullainathan (2004)

	0	1
black	2278	157
white	2200	235

Randomization Inference

- Resume audit study, Bertrand and Mullainathan (2004)

	0	1
black	2278	157
white	2200	235

- Only possible values: $\tau_i \in \{-1, 0, 1\}$

Randomization Inference

- Resume audit study, Bertrand and Mullainathan (2004)

	0	1
black	2278	157
white	2200	235

- Only possible values: $\tau_i \in \{-1, 0, 1\}$

```
resume %>% group_by(race) %>% summarise(call_rate = me
```

```
## # A tibble: 2 x 2
##   race  call_rate
##   <fct>     <dbl>
## 1 black    0.0645
## 2 white    0.0965
```

- Assume the sharp null $\tau_i = 0$ for every employer.

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

$$_{4870}C_{2435} = \binom{4870}{2435} = \frac{4870 \cdot 4869 \cdot \ldots \cdot 2436}{2435!}$$

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

$$ {}_{4870}C_{2435} = \binom{4870}{2435} = \frac{4870 \cdot 4869 \cdot \ldots \cdot 2436}{2435!} $$

$\approx 1.1 \times 10^{1464}$

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

$$ {}_{4870}C_{2435} = \binom{4870}{2435} = \frac{4870 \cdot 4869 \cdot \ldots \cdot 2436}{2435!} $$

$\approx 1.1 \times 10^{1464}$

(There are $\approx 10^{86}$ fundamental particles in the universe.)

- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

$$_{4870}C_{2435} = \binom{4870}{2435} = \frac{4870 \cdot 4869 \cdot \ldots \cdot 2436}{2435!}$$

$\approx 1.1 \times 10^{1464}$

(There are $\approx 10^{86}$ fundamental particles in the universe.)

- Let's do 1000, or 100,000 – something reasonable

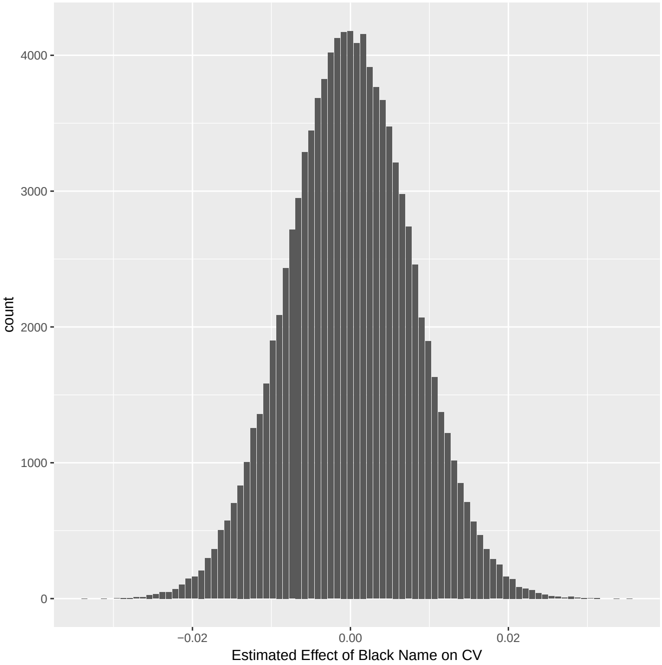
- Assume the sharp null $\tau_i = 0$ for every employer.
- $H_0 : \mu_{\text{black name}} = \mu_{\text{white name}}$
- $H_A : \mu_{\text{black name}} \neq \mu_{\text{white name}}$
- Create reference dist'n of all possible assignments

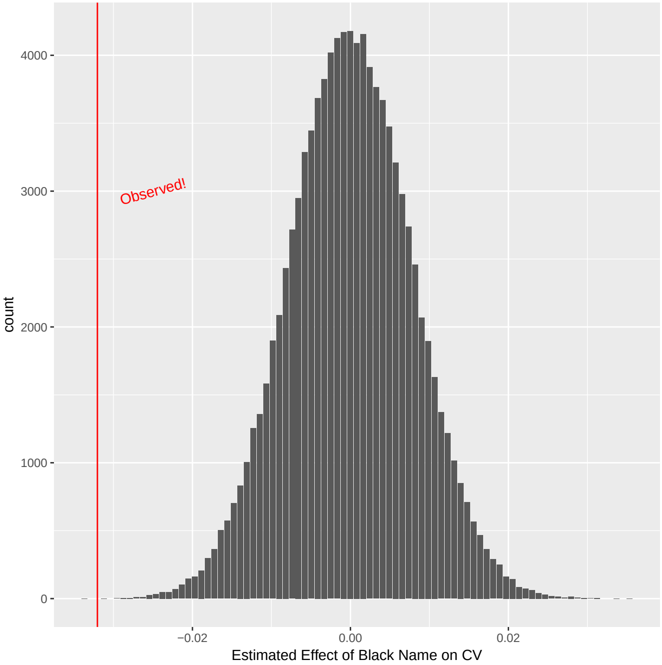
$$
{}_{4870}C_{2435} = \binom{4870}{2435} = \frac{4870 \cdot 4869 \cdot \ldots \cdot 2436}{2435!}
$$

$\approx 1.1 \times 10^{1464}$

(There are $\approx 10^{86}$ fundamental particles in the universe.)

- Let's do 1000, or 100,000 – something reasonable
- See `02-ri-resume-donate.R`

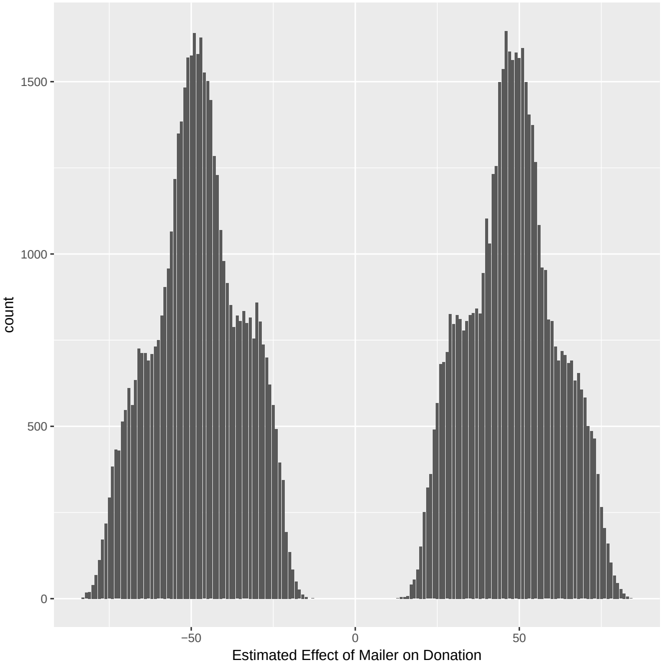

4000
3000
2000
1000
0
count
−0.02
0.00
0.02
Estimated Effect of Black Name on CV



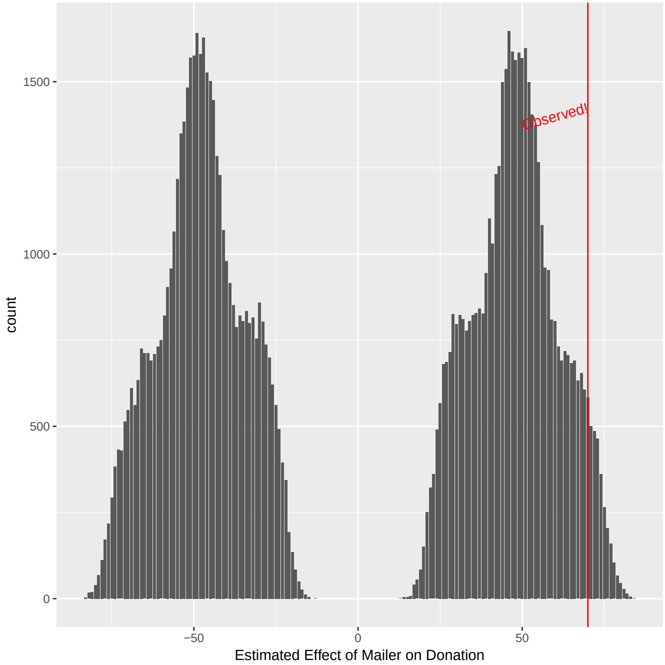
Observed!
count
4000
3000
2000
1000
0
−0.02
0.00
0.02
Estimated Effect of Black Name on CV

Randomization Inference

- Gerber & Green donations example, p. 65
- Possible values $\tau_i \in (-\infty, \infty)$
- Y_1, Y_0, τ likely very skewed
- See `02-ri-resume-donate.R`



count
1500
1000
500
0
−50
0
50
Estimated Effect of Mailer on Donation


Observed!
1500
1000
500
0
count
−50
0
50
Estimated Effect of Mailer on Donation

The RI Confidence Interval

Recall that

$$\text{“reject } H_0 \text{ at } \alpha = 0.05\text{”} \equiv \text{“}H_0 \text{ falls outside 95\% CI”}$$

The RI Confidence Interval

Recall that

"reject H_0 at $\alpha = 0.05$" $\equiv$ "H_0 falls outside 95% CI"

Create RI confidence intervals

- Posit $H_0 : \tau = \tau^* \in \{\ldots, -2, 1, 0, 1, 2, \ldots\}$

The RI Confidence Interval

Recall that

$$\text{``reject } H_0 \text{ at } \alpha = 0.05\text{''} \equiv \text{``}H_0 \text{ falls outside 95\% CI''}$$

Create RI confidence intervals

- Posit $H_0 : \tau = \tau^* \in \{\ldots, -2, 1, 0, 1, 2, \ldots\}$
- RI test whether to reject H_0

The RI Confidence Interval

Recall that

$$\text{``reject } H_0 \text{ at } \alpha = 0.05\text{''} \equiv \text{``}H_0 \text{ falls outside 95\% CI''}$$

Create RI confidence intervals

- Posit $H_0 : \tau = \tau^* \in \{\ldots, -2, 1, 0, 1, 2, \ldots\}$
- RI test whether to reject H_0
- If not, then τ^* is in CI

The RI Confidence Interval

Recall that

$$\text{"reject } H_0 \text{ at } \alpha = 0.05\text{"} \equiv \text{"}H_0 \text{ falls outside 95\% CI"}$$

Create RI confidence intervals

- Posit $H_0 : \tau = \tau^* \in \{\dots, -2, 1, 0, 1, 2, \dots\}$
- RI test whether to reject H_0
- If not, then τ^* is in CI
- CI consists of set of τ^* not unusual, given data

The RI Confidence Interval

Recall that

$$\text{“reject } H_0 \text{ at } \alpha = 0.05\text{''} \equiv \text{“}H_0 \text{ falls outside 95\% CI''}$$

Create RI confidence intervals

- Posit $H_0 : \tau = \tau^* \in \{\ldots, -2, 1, 0, 1, 2, \ldots\}$
- RI test whether to reject H_0
- If not, then τ^* is in CI
- CI consists of set of τ^* not unusual, given data
- See `02-ri-resume-donate.R`

Next:
Covariates in Experiments
PS1 due

Bertrand, Marianne, and Sendhil Mullainathan. 2004. “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94 (4): 991–1013.

Sen, Maya, and Omar Wasow. 2016. “Race as a Bundle of Sticks: Designs That Estimate Effects of Seemingly Immutable Characteristics.” *Annual Reviews of Political Science* 19: 499–522.