Module 2: Randomization Inference

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

1/ Randomized experiments

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
    - Message: randomization allows for inference under practically no assumptions.

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
    - Message: randomization allows for inference under practically no assumptions.
- No point estimation yet, just inference via tests and intervals.

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
  - Message: randomization allows for inference under practically no assumptions.
- No point estimation yet, just inference via tests and intervals.
- Useful to have notation for vector of all r.v.s

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
    - Message: randomization allows for inference under practically no assumptions.
- No point estimation yet, just inference via tests and intervals.
- Useful to have notation for vector of all r.v.s
    - Treatment: $\mathbf{D} = (D_1, D_2, \dots, D_n)$.

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
    - Message: randomization allows for inference under practically no assumptions.
- No point estimation yet, just inference via tests and intervals.
- Useful to have notation for vector of all r.v.s
    - Treatment: $\mathbf{D} = (D_1, D_2, \dots, D_n)$.
    - Potential outcomes: $\mathbf{Y}(1) = \{Y_1(1), \dots, Y_n(1)\}$.

Motivation

- Last time: defining causal effects as **counterfactual contrasts**.
- What can we learn about these contrasts in randomized experiments?
    - Message: randomization allows for inference under practically no assumptions.
- No point estimation yet, just inference via tests and intervals.
- Useful to have notation for vector of all r.v.s
    - Treatment: $\mathbf{D} = (D_1, D_2, \dots, D_n)$.
    - Potential outcomes: $\mathbf{Y}(1) = \{Y_1(1), \dots, Y_n(1)\}$.
    - Covariates: $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
    - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
    - p_i is controlled and known by researcher in an experiment.

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

1. **Positivity**: assignment is probabilistic: $0 < \mathbb{P}(D_i = 1) < 1$

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

1. **Positivity**: assignment is probabilistic: $0 < \mathbb{P}(D_i = 1) < 1$
   - No deterministic assignment.

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

1. **Positivity**: assignment is probabilistic: $0 < \mathbb{P}(D_i = 1) < 1$
   - No deterministic assignment.
2. **Unconfoundedness**: $\mathbb{P}[D_i = 1 | \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

1. **Positivity**: assignment is probabilistic: $0 < \mathbb{P}(D_i = 1) < 1$
   - No deterministic assignment.
2. **Unconfoundedness**: $\mathbb{P}[D_i = 1 | \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$
   - Treatment assignment does not depend on any potential outcomes.

Randomized Experiments

- **Experiment**: when the researcher controls the treatment assignment.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- **Randomized experiment** is an experiment with two properties:

1. **Positivity**: assignment is probabilistic: $0 < \mathbb{P}(D_i = 1) < 1$
   - No deterministic assignment.
2. **Unconfoundedness**: $\mathbb{P}[D_i = 1 | \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$
   - Treatment assignment does not depend on any potential outcomes.
   - Sometimes written as $D_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$

Effect of political information on accountability

- Does information help citizens hold politicians accountable?

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
    - Difficult with observational studies: having information correlated with lots of stuff!

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
    - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
  - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.
- Setup:

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
    - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.
- Setup:
    - Units: villages i

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
  - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.
- Setup:
  - Units: villages i
  - Treatment: post information about incumbent corruption in village $(D_i = 1)$ or not $(D_i = 0)$

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
  - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.
- Setup:
  - Units: villages i
  - Treatment: post information about incumbent corruption in village $(D_i = 1)$ or not $(D_i = 0)$
  - Outcome: incumbent wins vote in village $(Y_i = 1)$ or not $(Y_i = 0)$

Effect of political information on accountability

- Does information help citizens hold politicians accountable?
  - Difficult with observational studies: having information correlated with lots of stuff!
- Randomized controlled trial can be helpful.
- Setup:
  - Units: villages i
  - Treatment: post information about incumbent corruption in village $(D_i = 1)$ or not $(D_i = 0)$
  - Outcome: incumbent wins vote in village $(Y_i = 1)$ or not $(Y_i = 0)$
- If information $\rightsquigarrow$ accountability, we should see a difference between the treatment and control groups.

Why randomize?

- Randomization makes treated and control groups **comparable**.

Why randomize?

- Randomization makes treated and control groups **comparable**.
    - Both groups are random samples from all units in the study.

Why randomize?

- Randomization makes treated and control groups **comparable**.
    - Both groups are random samples from all units in the study.
    - $\rightsquigarrow$ **balanced** on all variables: roughly $=$ men and women, etc.

Why randomize?

- Randomization makes treated and control groups **comparable**.
    - Both groups are random samples from all units in the study.
    - $\rightsquigarrow$ **balanced** on all variables: roughly $=$ men and women, etc.
    - True for all **pretreatment** observed and unobserved variables.

Why randomize?

- Randomization makes treated and control groups **comparable**.
    - Both groups are random samples from all units in the study.
    - $\rightsquigarrow$ **balanced** on all variables: roughly = men and women, etc.
    - True for all **pretreatment** observed and unobserved variables.
    - Most important: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

Why randomize?

- Randomization makes treated and control groups **comparable**.
  - Both groups are random samples from all units in the study.
  - $\rightsquigarrow$ **balanced** on all variables: roughly $=$ men and women, etc.
  - True for all **pretreatment** observed and unobserved variables.
  - Most importantly: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

- Note: groups aren't comparable on **post-treatment** variables.

Why randomize?

- Randomization makes treated and control groups **comparable**.
  - Both groups are random samples from all units in the study.
  - $\rightsquigarrow$ **balanced** on all variables: roughly = men and women, etc.
  - True for all **pretreatment** observed and unobserved variables.
  - Most importantly: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

- Note: groups aren't comparable on **post-treatment** variables.
  - $Y_i(1) \perp\!\!\!\perp D_i$ but not $Y_i \perp\!\!\!\perp D_i$

Why randomize?

- Randomization makes treated and control groups **comparable**.
  - Both groups are random samples from all units in the study.
  - $\rightsquigarrow$ **balanced** on all variables: roughly = men and women, etc.
  - True for all **pretreatment** observed and unobserved variables.
  - Most importantly: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

- Note: groups aren't comparable on **post-treatment** variables.
  - $Y_i(1) \perp\!\!\!\perp D_i$ but not $Y_i \perp\!\!\!\perp D_i$
- Really talking about **ideal** randomized experiment:

Why randomize?

- Randomization makes treated and control groups **comparable**.
  - Both groups are random samples from all units in the study.
  - $\rightsquigarrow$ **balanced** on all variables: roughly = men and women, etc.
  - True for all **pretreatment** observed and unobserved variables.
  - Most importantly: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

- Note: groups aren't comparable on **post-treatment** variables.
  - $Y_i(1) \perp\!\!\!\perp D_i$ but not $Y_i \perp\!\!\!\perp D_i$
- Really talking about **ideal** randomized experiment:
  - Full compliance, no missing data

Why randomize?

- Randomization makes treated and control groups **comparable**.
  - Both groups are random samples from all units in the study.
  - $\rightsquigarrow$ **balanced** on all variables: roughly = men and women, etc.
  - True for all **pretreatment** observed and unobserved variables.
  - Most important: potential outcomes are comparable by unconfoundedness:

$$\mathbb{P}(Y_i(1) = 1 \mid D_i = 1) = \mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(1) = 1 \mid D_i = 0)$$

- Note: groups aren't comparable on **post-treatment** variables.
  - $Y_i(1) \perp\!\!\!\perp D_i$ but not $Y_i \perp\!\!\!\perp D_i$
- Really talking about **ideal** randomized experiment:
  - Full compliance, no missing data
  - Important to admit limitations: external validity, sample selection, Hawthorne effect

Types of experiments

- Experiments can be classified their **assignment mechanism**.

Types of experiments

- Experiments can be classified their **assignment mechanism**.
    - What (random) function do we use to assign treatment?

Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)

Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
  - Each unit is assigned $D_i = 1$ with prob. p independently.

Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
  - Each unit is assigned $D_i = 1$ with prob. p independently.
  - Downside: "bad" randomizations possible (all treated/control)

Types of experiments

- Experiments can be classified their **assignment mechanism**.
    - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
    - Each unit is assigned $D_i = 1$ with prob. p independently.
    - Downside: "bad" randomizations possible (all treated/control)
- **Completely randomized experiment**:

Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
  - Each unit is assigned $D_i = 1$ with prob. p independently.
  - Downside: “bad” randomizations possible (all treated/control)
- **Completely randomized experiment**:
  - Randomly sample n_1 units from the population to be treated.

Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
  - Each unit is assigned $D_i = 1$ with prob. p independently.
  - Downside: "bad" randomizations possible (all treated/control)
- **Completely randomized experiment**:
  - Randomly sample n_1 units from the population to be treated.
  - Equal probability of any assignment with $\sum_i D_i = n_1$:

$$\mathbb{P}(\mathbf{D} = (d_1, \dots, d_n) \mid \mathbf{Y}(1), \mathbf{Y}(0)) = \begin{cases} \binom{n}{n_1}^{-1} & \text{if } \sum_{i=1}^{n} d_i = n_1 \\ 0 & \text{otherwise} \end{cases}$$

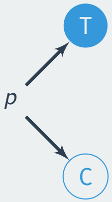
Types of experiments

- Experiments can be classified their **assignment mechanism**.
  - What (random) function do we use to assign treatment?
- **Bernoulli randomization** (coin flips)
  - Each unit is assigned $D_i = 1$ with prob. p independently.
  - Downside: "bad" randomizations possible (all treated/control)
- **Completely randomized experiment**:
  - Randomly sample n_1 units from the population to be treated.
  - Equal probability of any assignment with $\sum_i D_i = n_1$:

$$\mathbb{P}(\mathbf{D} = (d_1, \dots, d_n) \mid \mathbf{Y}(1), \mathbf{Y}(0)) = \begin{cases} \binom{n}{n_1}^{-1} & \text{if } \sum_{i=1}^n d_i = n_1 \\ 0 & \text{otherwise} \end{cases}$$

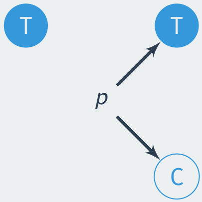
  - For any given i, implies $\mathbb{P}(D_i = 1 \mid \mathbf{Y}(1), \mathbf{Y}(0)) = \frac{n_1}{n}$.

Bernoulli assignment

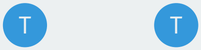


Bernoulli assignment

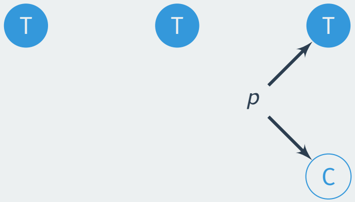
Bernoulli assignment



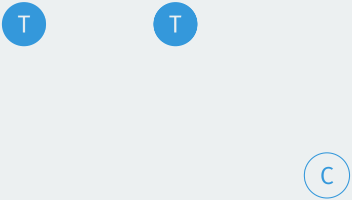
Bernoulli assignment



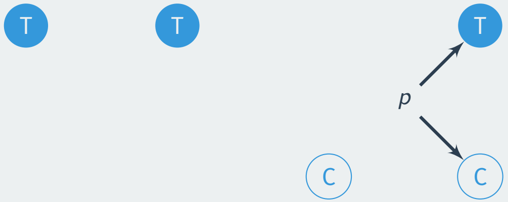
Bernoulli assignment



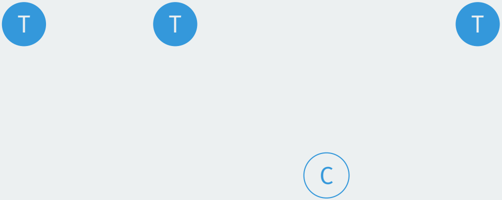
Bernoulli assignment



Bernoulli assignment



Bernoulli assignment



Completely randomized design

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$:

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Fixed number of treated units induces dependence between D_i and D_j

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Fixed number of treated units induces dependence between D_i and D_j
  - Knowing 2 is treated $\rightsquigarrow$ 3 is less likely to be treated.

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Fixed number of treated units induces dependence between D_i and D_j
  - Knowing 2 is treated $\rightsquigarrow$ 3 is less likely to be treated.
  - Makes variance calculations tricky (we'll come back to this)

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Fixed number of treated units induces dependence between D_i and D_j
  - Knowing 2 is treated $\rightsquigarrow$ 3 is less likely to be treated.
  - Makes variance calculations tricky (we'll come back to this)
- We can also randomize within groups (**block/stratified randomization**).

Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Fixed number of treated units induces dependence between D_i and D_j
  - Knowing 2 is treated $\rightsquigarrow$ 3 is less likely to be treated.
  - Makes variance calculations tricky (we'll come back to this)
- We can also randomize within groups (**block/stratified randomization**).
  - When blocks are of size 2, this is a **pair-matched design**.

Example data from information RCT

Village	Information D_i	Incumbent Won? Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	?	0
2	1	0	?	0
3	0	1	1	?
4	1	0	?	0
5	1	1	?	1
6	0	1	1	?
7	0	0	0	?
8	1	1	?	1
9	0	1	1	?
10	0	0	0	?

- Incumbent won 2/5 treated villages vs 3/5 control villages.

Example data from information RCT

Village	Information D_i	Incumbent Won? Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	?	0
2	1	0	?	0
3	0	1	1	?
4	1	0	?	0
5	1	1	?	1
6	0	1	1	?
7	0	0	0	?
8	1	1	?	1
9	0	1	1	?
10	0	0	0	?

- Incumbent won 2/5 treated villages vs 3/5 control villages.
- Very small sample size $\rightsquigarrow$ can we learn anything from this data?

2/ Randomization inference

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
    - Fisher: randomization is the "reasoned basis for inference."

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
  - Fisher: randomization is the "reasoned basis for inference."
  - We can generate exact p-values for tests of a "sharp" null hypothesis.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
  - Fisher: randomization is the "reasoned basis for inference."
  - We can generate exact p-values for tests of a "sharp" null hypothesis.
  - Also called: **design-based inference**.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
  - Fisher: randomization is the "reasoned basis for inference."
  - We can generate exact p-values for tests of a "sharp" null hypothesis.
  - Also called: **design-based inference**.
- Null hypothesis of no effect for any unit $\rightsquigarrow$ very strong.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
    - Fisher: randomization is the "reasoned basis for inference."
    - We can generate exact p-values for tests of a "sharp" null hypothesis.
    - Also called: **design-based inference**.
- Null hypothesis of no effect for any unit $\rightsquigarrow$ very strong.
- Allows us to make **exact, distribution-free** inferences.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
    - Fisher: randomization is the "reasoned basis for inference."
    - We can generate exact p-values for tests of a "sharp" null hypothesis.
    - Also called: **design-based inference**.
- Null hypothesis of no effect for any unit $\rightsquigarrow$ very strong.
- Allows us to make **exact, distribution-free** inferences.
    - No reliance on normality, etc.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
  - Fisher: randomization is the "reasoned basis for inference."
  - We can generate exact p-values for tests of a "sharp" null hypothesis.
  - Also called: **design-based inference**.
- Null hypothesis of no effect for any unit $\rightsquigarrow$ very strong.
- Allows us to make **exact, distribution-free** inferences.
  - No reliance on normality, etc.
  - No reliance on large-sample approximations.

What is randomization inference?

- **Randomization inference**: inference based on different possible randomizations of treatment.
    - Fisher: randomization is the "reasoned basis for inference."
    - We can generate exact p-values for tests of a "sharp" null hypothesis.
    - Also called: **design-based inference**.
- Null hypothesis of no effect for any unit $\rightsquigarrow$ very strong.
- Allows us to make **exact, distribution-free** inferences.
    - No reliance on normality, etc.
    - No reliance on large-sample approximations.
    - $\rightsquigarrow$ truly nonparametric, but less flexible.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.
    - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.
    - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$
3. Determine the distribution of the test statistic under the null.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
   - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
   - No average treatment effect.
   - Claim we would like to reject.
2. Choose a test statistic.
   - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$
3. Determine the distribution of the test statistic under the null.
   - Statistical thought experiment: we know the truth, what data should we expect?

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.
    - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$
3. Determine the distribution of the test statistic under the null.
    - Statistical thought experiment: we know the truth, what data should we expect?
4. Calculate the probability of the test statistics under the null.

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.
    - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$
3. Determine the distribution of the test statistic under the null.
    - Statistical thought experiment: we know the truth, what data should we expect?
4. Calculate the probability of the test statistics under the null.
    - What is this called?

Brief review of hypothesis testing

RI focuses on hypothesis testing, so it's helpful to review.

1. Choose a null hypothesis:
    - $H_0 : \beta_1 = 0$ or $H_0 : \tau = 0$.
    - No average treatment effect.
    - Claim we would like to reject.
2. Choose a test statistic.
    - $Z_i = (X_i - \bar{X})/(s/\sqrt{n})$
3. Determine the distribution of the test statistic under the null.
    - Statistical thought experiment: we know the truth, what data should we expect?
4. Calculate the probability of the test statistics under the null.
    - What is this called? **p-value**

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

- What if treatment affected no one at all?

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

- What if treatment affected no one at all?
- Implies no **average** treatment effect, but no ATE $\not\Rightarrow$ sharp null.

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

- What if treatment affected no one at all?
- Implies no **average** treatment effect, but no ATE $\not\Rightarrow$ sharp null.
  - Take a simple example with two units: $\tau_1 = 1 \qquad \tau_2 = -1$

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

- What if treatment affected no one at all?
- Implies no **average** treatment effect, but no ATE $\not\Rightarrow$ sharp null.
  - Take a simple example with two units: $\tau_1 = 1 \qquad \tau_2 = -1$
  - Here, $\tau = 0$ but the sharp null is violated.

Sharp null hypothesis of no effect

- **Sharp null hypothesis**:

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i$$

- What if treatment affected no one at all?
- Implies no **average** treatment effect, but no ATE $\not\Rightarrow$ sharp null.
  - Take a simple example with two units: $\tau_1 = 1 \qquad \tau_2 = -1$
  - Here, $\tau = 0$ but the sharp null is violated.
- If the sharp null is true, we know all the potential outcomes:

$$Y_i(1) = Y_i(0) = Y_i$$

Life under the sharp null

We can use the sharp null ($Y_i(1) - Y_i(0) = 0$) to fill in the missing potential outcomes:

	Information	Incumbent Won?		
Village	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	?	0
2	1	0	?	0
3	0	1	1	?
4	1	0	?	0
5	1	1	?	1
6	0	1	1	?
7	0	0	0	?
8	1	1	?	1
9	0	1	1	?
10	0	0	0	?

Life under the sharp null

We can use the sharp null ($Y_i(1) - Y_i(0) = 0$) to fill in the missing potential outcomes:

	Information	Incumbent Won?		
Village	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	0	0
2	1	0	0	0
3	0	1	1	1
4	1	0	0	0
5	1	1	1	1
6	0	1	1	1
7	0	0	0	0
8	1	1	1	1
9	0	1	1	1
10	0	0	0	0

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.
- Want a test statistic with high **statistical power**:

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.
- Want a test statistic with high **statistical power**:
  - Has large values when the null is false

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.
- Want a test statistic with high **statistical power**:
  - Has large values when the null is false
  - These large values are unlikely when then null is true.

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.
- Want a test statistic with high **statistical power**:
    - Has large values when the null is false
    - These large values are unlikely when then null is true.
- These will help us perform a test of the sharp null.

Test statistic

Test statistic

A test statistic is a known, scalar quantity calculated from the treatment assignments, observed outcomes, and possibly covariates: $T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$

- Typically measures the relationship between two variables.
- Test statistics measure how unusual the data is under the null.
- Want a test statistic with high **statistical power**:
  - Has large values when the null is false
  - These large values are unlikely when then null is true.
- These will help us perform a test of the sharp null.
- Many possible tests to choose from!

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?
  - If there was no effect, what test statistics would we expect over different randomizations?

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?
  - If there was no effect, what test statistics would we expect over different randomizations?
- **Key insight of RI**: sharp null $\rightsquigarrow$ treatment assignment doesn't matter.

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?
    - If there was no effect, what test statistics would we expect over different randomizations?
- **Key insight of RI**: sharp null $\rightsquigarrow$ treatment assignment doesn't matter.
    - Shuffling treatment vector won't change outcomes.

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?
  - If there was no effect, what test statistics would we expect over different randomizations?
- **Key insight of RI**: sharp null $\rightsquigarrow$ treatment assignment doesn't matter.
  - Shuffling treatment vector won't change outcomes.
  - $Y_i(1) = Y_i(0) = Y_i$

Null/randomization disitribution

- What is the distribution of the test statistic under the sharp null?
  - If there was no effect, what test statistics would we expect over different randomizations?
- **Key insight of RI**: sharp null $\rightsquigarrow$ treatment assignment doesn't matter.
  - Shuffling treatment vector won't change outcomes.
  - $Y_i(1) = Y_i(0) = Y_i$
- **Randomization distribution**: distribution of T if the sharp null were true.

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.
    - $\Omega =$ set of 2^N assignment vectors (any N-vector of 0s and 1s)

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.
    - $\Omega =$ set of 2^N assignment vectors (any N-vector of 0s and 1s)
    - We can also define the set of feasible assignments under the design:

$$\Omega_0 = \{\mathbf{d} : \mathbb{P}(\mathbf{D} = \mathbf{d}) > 0\}$$

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.
    - $\Omega =$ set of 2^N assignment vectors (any N-vector of 0s and 1s)
    - We can also define the set of feasible assignments under the design:

$$\Omega_0 = \{\mathbf{d} : \mathbb{P}(\mathbf{D} = \mathbf{d}) > 0\}$$

- **Exact p-values**:

$$\Pr(T \geq T^{\text{obs}} \mid \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{X}, H_0) = \frac{1}{|\Omega_0|} \sum_{\mathbf{d} \in \Omega_0} \mathbb{I}(T(\mathbf{d}, \mathbf{Y}, \mathbf{X}) \geq T^{\text{obs}})$$

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.
    - $\Omega =$ set of 2^N assignment vectors (any N-vector of 0s and 1s)
    - We can also define the set of feasible assignments under the design:

$$\Omega_0 = \{\mathbf{d} : \mathbb{P}(\mathbf{D} = \mathbf{d}) > 0\}$$

- **Exact p-values**:

$$\Pr(T \geq T^{\text{obs}} \mid \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{X}, H_0) = \frac{1}{|\Omega_0|} \sum_{\mathbf{d} \in \Omega_0} \mathbb{I}(T(\mathbf{d}, \mathbf{Y}, \mathbf{X}) \geq T^{\text{obs}})$$

    - How often T is larger than the observed divided by total number of randomizations.

Calculate p-values

- How often would we get a test statistic this big or bigger if the sharp null holds?
    - Let $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{Z})$ be the observed value of the test statistic.
    - $\Omega =$ set of 2^N assignment vectors (any N-vector of 0s and 1s)
    - We can also define the set of feasible assignments under the design:

$$\Omega_0 = \{\mathbf{d} : \mathbb{P}(\mathbf{D} = \mathbf{d}) > 0\}$$

- **Exact p-values**:

$$\Pr(T \geq T^{\text{obs}} \mid \mathbf{Y}(1), \mathbf{Y}(0), \mathbf{X}, H_0) = \frac{1}{|\Omega_0|} \sum_{\mathbf{d} \in \Omega_0} \mathbb{I}(T(\mathbf{d}, \mathbf{Y}, \mathbf{X}) \geq T^{\text{obs}})$$

    - How often T is larger than the observed divided by total number of randomizations.
    - p-values will be below α exactly $100\alpha\%$ of the time

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,
2. Calculate observed test statistic: $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$.

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,
2. Calculate observed test statistic: $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$.
3. Randomly select different treatment vector $\tilde{\mathbf{D}}_1$ from Ω_0

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,
2. Calculate observed test statistic: $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$.
3. Randomly select different treatment vector $\tilde{\mathbf{D}}_1$ from Ω_0
4. Calculate $\widetilde{T}_1 = T(\tilde{\mathbf{D}}_1, \mathbf{Y}, \mathbf{X})$.

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,
2. Calculate observed test statistic: $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$.
3. Randomly select different treatment vector $\tilde{\mathbf{D}}_1$ from Ω_0
4. Calculate $\widetilde{T}_1 = T(\tilde{\mathbf{D}}_1, \mathbf{Y}, \mathbf{X})$.
5. Repeat steps 3-4 for all Ω_0 to get $\widetilde{T} = \{\widetilde{T}_1, \dots, \widetilde{T}_K\}$.

Randomization inference step-by-step

1. Choose a sharp null hypothesis and a test statistic,
2. Calculate observed test statistic: $T^{\text{obs}} = T(\mathbf{D}, \mathbf{Y}, \mathbf{X})$.
3. Randomly select different treatment vector $\tilde{\mathbf{D}}_1$ from Ω_0
4. Calculate $\widetilde{T}_1 = T(\tilde{\mathbf{D}}_1, \mathbf{Y}, \mathbf{X})$.
5. Repeat steps 3-4 for all Ω_0 to get $\widetilde{T} = \{\widetilde{T}_1, \ldots, \widetilde{T}_K\}$.
6. Calculate the p-value: $p = \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}(\widetilde{T}_k \geq T)$

Difference in means

- Many different types of test statistics with different strengths.

Difference in means

- Many different types of test statistics with different strengths.
- Natural (if not optimal): absolute difference in means estimator

$$T_{\text{diff}} = \left| \frac{1}{n_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{n_0} \sum_{i=1}^{N} (1 - D_i) Y_i \right|$$

Difference in means

- Many different types of test statistics with different strengths.
- Natural (if not optimal): absolute difference in means estimator

$$T_{\text{diff}} = \left| \frac{1}{n_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{n_0} \sum_{i=1}^{N} (1 - D_i) Y_i \right|$$

- Larger values of T_{diff} are evidence against the sharp null.

Difference in means

- Many different types of test statistics with different strengths.
- Natural (if not optimal): absolute difference in means estimator

$$T_{\text{diff}} = \left| \frac{1}{n_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{n_0} \sum_{i=1}^{N} (1 - D_i) Y_i \right|$$

- Larger values of T_{diff} are evidence against the sharp null.
- Good estimator for constant, additive treatment effects and relatively few outliers in the the potential outcomes.

Example

- Suppose we are targeting 6 people for donations to Harvard.

Example

- Suppose we are targeting 6 people for donations to Harvard.
- As an encouragement, we send 3 of them a mailer with inspirational stories of learning from our graduate students.

Example

- Suppose we are targeting 6 people for donations to Harvard.
- As an encouragement, we send 3 of them a mailer with inspirational stories of learning from our graduate students.
- Afterwards, we observe them giving between \$0 and \$5.

Example

- Suppose we are targeting 6 people for donations to Harvard.
- As an encouragement, we send 3 of them a mailer with inspirational stories of learning from our graduate students.
- Afterwards, we observe them giving between $0 and $5.
- Simple example to show the steps of RI in a concrete case.

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(3)	3
Sansa	1	5	(5)	5
Arya	1	0	(0)	0
Robb	0	4	4	(4)
Bran	0	0	0	(0)
Rickon	0	1	1	(1)

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(3)	3
Sansa	1	5	(5)	5
Arya	1	0	(0)	0
Robb	0	4	4	(4)
Bran	0	0	0	(0)
Rickon	0	1	1	(1)

$$T_{\text{diff}} = |8/3 - 5/3| = 1$$

Randomization distribution

Unit	Mailer $\widetilde{D}_i$	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(3)	3
Sansa	1	5	(5)	5
Arya	0	0	(0)	0
Robb	1	4	4	(4)
Bran	0	0	0	(0)
Rickon	0	1	1	(1)

$$\widetilde{T}_{\text{diff}} = |12/3 - 1/3| = 3.67$$

Randomization distribution

Unit	Mailer $\widetilde{D}_i$	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(3)	3
Sansa	1	5	(5)	5
Arya	0	0	(0)	0
Robb	0	4	4	(4)
Bran	1	0	0	(0)
Rickon	0	1	1	(1)

$$\widetilde{T}_{\text{diff}} = |8/3 - 5/3| = 1$$

Randomization distribution

Unit	Mailer $\widetilde{D}_i$	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(3)	3
Sansa	1	5	(5)	5
Arya	0	0	(0)	0
Robb	0	4	4	(4)
Bran	0	0	0	(0)
Rickon	1	1	1	(1)

$$\widetilde{T}_{\text{diff}} = |9/3 - 4/3| = 1.67$$

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|
1	1	1	0	0	0	1.00

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|
1	1	1	0	0	0	1.00
1	1	0	1	0	0	3.67

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|
1	1	1	0	0	0	1.00
1	1	0	1	0	0	3.67
1	1	0	0	1	0	1.00

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|
1	1	1	0	0	0	1.00
1	1	0	1	0	0	3.67
1	1	0	0	1	0	1.00
1	1	0	0	0	1	1.67

Randomization distribution

D_1	D_2	D_3	D_4	D_5	D_6	|Diff in means|
1	1	1	0	0	0	1.00
1	1	0	1	0	0	3.67
1	1	0	0	1	0	1.00
1	1	0	0	0	1	1.67
1	0	1	1	0	0	0.33
1	0	1	0	1	0	2.33
1	0	1	0	0	1	1.67
1	0	0	1	1	0	0.33
1	0	0	1	0	1	1.00
1	0	0	0	1	1	1.67
0	1	1	1	0	0	1.67
0	1	1	0	1	0	1.00
0	1	1	0	0	1	0.33
0	1	0	1	1	0	1.67
0	1	0	1	0	1	2.33
0	1	0	0	1	1	0.33
0	0	1	1	1	0	1.67
0	0	1	1	0	1	1.00
0	0	1	0	1	1	3.67
0	0	0	1	1	1	3.67

In R

```
library(ri)
y <- c(3, 5, 0, 4, 0, 1)
D <- c(1, 1, 1, 0, 0, 0)
T_obs <- abs(mean(y[D == 1]) - mean(y[D == 0]))
D_bold <- ri::genperms(D)
D_bold[, 1:7]
```

In R

```r
library(ri)
y <- c(3, 5, 0, 4, 0, 1)
D <- c(1, 1, 1, 0, 0, 0)
T_obs <- abs(mean(y[D == 1]) - mean(y[D == 0]))
D_bold <- ri::genperms(D)
D_bold[, 1:7]
```

```
##   [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## 1    1    1    1    1    1    1    1
## 2    1    1    1    1    0    0    0
## 3    1    0    0    0    1    1    1
## 4    0    1    0    0    1    0    0
## 5    0    0    1    0    0    1    0
## 6    0    0    0    1    0    0    1
```

Calculate means

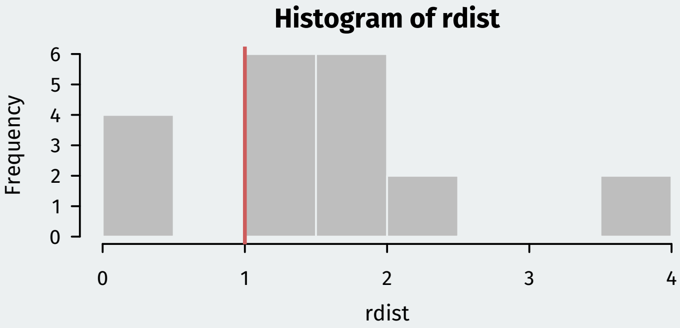
```r
rdist <- rep(NA, times = ncol(D_bold))
for (i in seq_len(ncol(D_bold))) {
  D_tilde <- D_bold[, i]
  rdist[i] <- abs(mean(y[D_tilde == 1]) - mean(y[D_tilde == 0]))
}
rdist
```

Calculate means

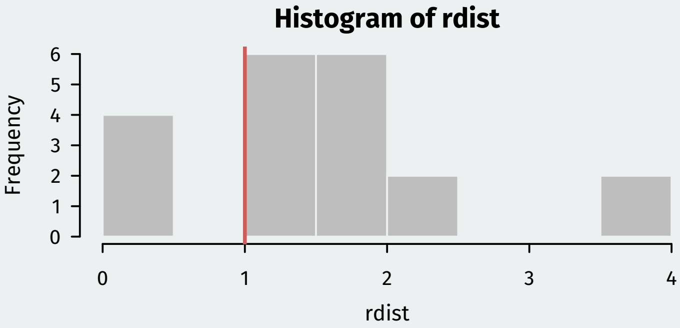
```
rdist <- rep(NA, times = ncol(D_bold))
for (i in seq_len(ncol(D_bold))) {
  D_tilde <- D_bold[, i]
  rdist[i] <- abs(mean(y[D_tilde == 1]) - mean(y[D_tilde == 0]))
}
rdist
```

```
##  [1] 1.000 3.667 1.000 1.667 0.333 2.333 1.667 0.333
##  [9] 1.000 1.667 1.667 1.000 0.333 1.667 2.333 0.333
## [17] 1.667 1.000 3.667 1.000
```

p-value



p-value




```
# p-value
mean(rdist >= T_obs)
```

```
## [1] 0.8
```

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow$ 20 assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow$ 252 vectors.

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.
- Workaround: simulation!

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.
- Workaround: simulation!
    - take K samples from the treatment assignment space, Ω_0.

Computation

Computing the exact randomization distribution not always feasible:

- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.
- Workaround: simulation!
  - take K samples from the treatment assignment space, Ω_0.
  - calculate the randomization distribution in the K samples.

Computation

Computing the exact randomization distribution not always feasible:

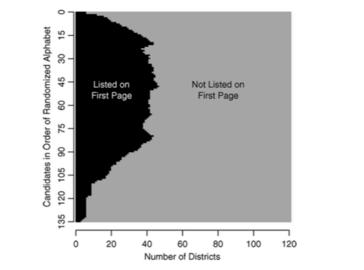
- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.
- Workaround: simulation!
    - take K samples from the treatment assignment space, Ω_0.
    - calculate the randomization distribution in the K samples.
    - tests no longer exact, but bias is under your control!

Computation

Computing the exact randomization distribution not always feasible:

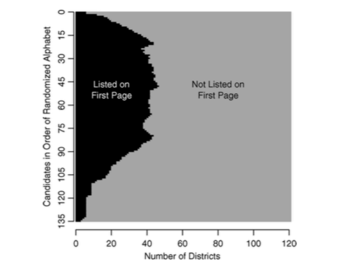
- $n = 6$ and $n_1 = 3 \rightsquigarrow 20$ assignment vectors.
- $n = 10$ and $n_1 = 5 \rightsquigarrow 252$ vectors.
- $n = 100$ and $n_1 = 50 \rightsquigarrow 1.009 \times 10^{29}$ vectors.
- Workaround: simulation!
    - take K samples from the treatment assignment space, Ω_0.
    - calculate the randomization distribution in the K samples.
    - tests no longer exact, but bias is under your control! (increase K)

CA recall election



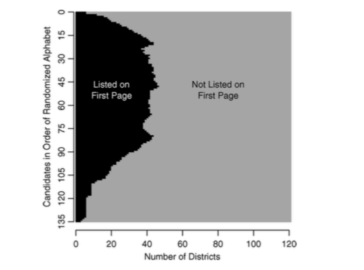
- Ho & Imai (2006): 135 candidates in 2003 CA Gov. recall election.

CA recall election



- Ho & Imai (2006): 135 candidates in 2003 CA Gov. recall election.
- Ballot order randomly assigned so not all candidates were on 1st page

CA recall election



- Ho & Imai (2006): 135 candidates in 2003 CA Gov. recall election.
- Ballot order randomly assigned so not all candidates were on 1st page
- Effect of being on the first page on the vote share for a candidate?

CA recall election

- Randomization process:

CA recall election

- Randomization process:

  1. Choose a random ordering of all 26 letters:

     R W Q O J M V A H B S G Z X N T C I E K U P D Y F L

CA recall election

- Randomization process:
    1. Choose a random ordering of all 26 letters:

       R W Q O J M V A H B S G Z X N T C I E K U P D Y F L

    2. Order candidates on ballot by this in the 1st assembly district.

CA recall election

- Randomization process:

    1. Choose a random ordering of all 26 letters:

       R W Q O J M V A H B S G Z X N T C I E K U P D Y F L

    2. Order candidates on ballot by this in the 1st assembly district.

    3. In the next district, rotate ordering by 1 letter and order names by this.

       W Q O J M V A H B S G Z X N T C I E K U P D Y F L R

CA recall election

- Randomization process:
  1. Choose a random ordering of all 26 letters:

     R W Q O J M V A H B S G Z X N T C I E K U P D Y F L

  2. Order candidates on ballot by this in the 1st assembly district.
  3. In the next district, rotate ordering by 1 letter and order names by this.

     W Q O J M V A H B S G Z X N T C I E K U P D Y F L R

  4. Continue rotating for each district.

CA recall election with RI

1. Pick another possible letter ordering.

CA recall election with RI

1. Pick another possible letter ordering.
2. Assign 1st page/not first page based on this new ordering as was done in the election.

CA recall election with RI

1. Pick another possible letter ordering.
2. Assign 1st page/not first page based on this new ordering as was done in the election.
3. Calculate diff-in-means for this new treatment.

CA recall election with RI

1. Pick another possible letter ordering.
2. Assign 1st page/not first page based on this new ordering as was done in the election.
3. Calculate diff-in-means for this new treatment.
4. Lather, rinse, repeat.

Other test statistics

- The difference in means is great for when effects are:

Other test statistics

- The difference in means is great for when effects are:
    - constant and additive

Other test statistics

- The difference in means is great for when effects are:
    - constant and additive
    - few outliers in the data

Other test statistics

- The difference in means is great for when effects are:
  - constant and additive
  - few outliers in the data
- Outliers $\rightsquigarrow$ more variation in the randomization distribution

Other test statistics

- The difference in means is great for when effects are:
    - constant and additive
    - few outliers in the data
- Outliers $\rightsquigarrow$ more variation in the randomization distribution
- What about alternative test statistics?

Transformations

- What if there was a constant multiplicative effect: $Y_i(1)/Y_i(0) = C$?

Transformations

- What if there was a constant multiplicative effect: $Y_i(1)/Y_i(0) = C$?
- T_{diff} will have low power in this case.

Transformations

- What if there was a constant multiplicative effect: $Y_i(1)/Y_i(0) = C$?
- T_{diff} will have low power in this case.
- $\rightsquigarrow$ transform the observed outcome using the natural logarithm:

$$T_{\log} = \left| \frac{1}{n_1} \sum_{i=1}^{n} D_i \log(Y_i) - \frac{1}{n_0} \sum_{i=1}^{n} (1 - D_i) \log(Y_i) \right|$$

Transformations

- What if there was a constant multiplicative effect: $Y_i(1)/Y_i(0) = C$?
- T_{diff} will have low power in this case.
- $\rightsquigarrow$ transform the observed outcome using the natural logarithm:

$$T_{\log} = \left| \frac{1}{n_1} \sum_{i=1}^{n} D_i \log(Y_i) - \frac{1}{n_0} \sum_{i=1}^{n} (1 - D_i) \log(Y_i) \right|$$

- Useful for skewed distributions of outcomes.

Difference in median/quantiles

- To further protect against outliers: quantiles .

Difference in median/quantiles

- To further protect against outliers: quantiles .
- Let use $\mathbf{Y}_t = \{Y_i; i : D_i = 1\}$ and $\mathbf{Y}_c = \{Y_i; i : D_i = 0\}$.

Difference in median/quantiles

- To further protect against outliers: quantiles .
- Let use $\mathbf{Y}_t = \{Y_i; i : D_i = 1\}$ and $\mathbf{Y}_c = \{Y_i; i : D_i = 0\}$.
- Differences in medians:

$$T_{\text{med}} = |\text{med}(\mathbf{Y}_t) - \text{med}(\mathbf{Y}_c)|$$

Difference in median/quantiles

- To further protect against outliers: quantiles .
- Let use $\mathbf{Y}_t = \{Y_i; i : D_i = 1\}$ and $\mathbf{Y}_c = \{Y_i; i : D_i = 0\}$.
- Differences in medians:

$$T_{\text{med}} = |\text{med}(\mathbf{Y}_t) - \text{med}(\mathbf{Y}_c)|$$

- Remember that the median is the 0.5 quantile.

Difference in median/quantiles

- To further protect against outliers: quantiles .
- Let use $\mathbf{Y}_t = \{Y_i; i : D_i = 1\}$ and $\mathbf{Y}_c = \{Y_i; i : D_i = 0\}$.
- Differences in medians:

$$T_{\text{med}} = |\text{med}(\mathbf{Y}_t) - \text{med}(\mathbf{Y}_c)|$$

- Remember that the median is the 0.5 quantile.
- Could use other quantiles (the 0.25 quantile or the 0.75 quantile).

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
    - with continuous outcomes,

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
  - with continuous outcomes,
  - small datasets, and/or

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
  - with continuous outcomes,
  - small datasets, and/or
  - many outliers

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
    - with continuous outcomes,
    - small datasets, and/or
    - many outliers
- Basic idea:

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
    - with continuous outcomes,
    - small datasets, and/or
    - many outliers
- Basic idea:
    - rank the outcomes (higher values of Y_i are assigned higher ranks)

Rank statistics

- Rank statistics transform outcomes to ranks and then analyze those.
- Useful for situations
  - with continuous outcomes,
  - small datasets, and/or
  - many outliers
- Basic idea:
  - rank the outcomes (higher values of Y_i are assigned higher ranks)
  - compare the average rank of the treated and control groups

Rank statistics formally

- Calculate ranks of the outcomes:

$$\tilde{R}_i = \tilde{R}_i(Y_1, \dots, Y_n) = \sum_{j=1}^{N} \mathbb{I}(Y_j \leq Y_i)$$

Rank statistics formally

- Calculate ranks of the outcomes:

$$\tilde{R}_i = \tilde{R}_i(Y_1, \ldots, Y_n) = \sum_{j=1}^{N} \mathbb{I}(Y_j \leq Y_i)$$

- Normalize the ranks to have mean 0:

$$\dot{R}_i = \tilde{R}_i(Y_1, \ldots, Y_n) - \frac{n+1}{2}$$

Rank statistics formally

- Calculate ranks of the outcomes:

$$\tilde{R}_i = \tilde{R}_i(Y_1, \dots, Y_n) = \sum_{j=1}^{N} \mathbb{I}(Y_j \leq Y_i)$$

- Normalize the ranks to have mean 0:

$$\dot{R}_i = \tilde{R}_i(Y_1, \dots, Y_n) - \frac{n+1}{2}$$

- Minor adjustment for ties yields R_i.

Rank statistics formally

- Calculate ranks of the outcomes:

$$\tilde{R}_i = \tilde{R}_i(Y_1, \dots, Y_n) = \sum_{j=1}^{N} \mathbb{I}(Y_j \leq Y_i)$$

- Normalize the ranks to have mean 0:

$$\dot{R}_i = \tilde{R}_i(Y_1, \dots, Y_n) - \frac{n+1}{2}$$

- Minor adjustment for ties yields R_i.
- Calculate the absolute difference in average ranks:

$$T_{\text{rank}} = |\bar{R}_t - \bar{R}_c| = \left| \frac{\sum_{i:D_i=1} R_i}{n_1} - \frac{\sum_{i:D_i=0} R_i}{n_0} \right|$$

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3		
Sansa	1	5	(5)	5		
Arya	1	0	(0)	0		
Robb	0	4	4	(4)		
Bran	0	0	0	(0)		
Rickon	0	1	1	(1)		

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3		
Sansa	1	5	(5)	5	6	
Arya	1	0	(0)	0		
Robb	0	4	4	(4)		
Bran	0	0	0	(0)		
Rickon	0	1	1	(1)		

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3		
Sansa	1	5	(5)	5	6	
Arya	1	0	(0)	0		
Robb	0	4	4	(4)	5	
Bran	0	0	0	(0)		
Rickon	0	1	1	(1)		

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3	4	
Sansa	1	5	(5)	5	6	
Arya	1	0	(0)	0		
Robb	0	4	4	(4)	5	
Bran	0	0	0	(0)		
Rickon	0	1	1	(1)		

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3	4	
Sansa	1	5	(5)	5	6	
Arya	1	0	(0)	0		
Robb	0	4	4	(4)	5	
Bran	0	0	0	(0)		
Rickon	0	1	1	(1)	3	

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3	4	
Sansa	1	5	(5)	5	6	
Arya	1	0	(0)	0	1.5	
Robb	0	4	4	(4)	5	
Bran	0	0	0	(0)	1.5	
Rickon	0	1	1	(1)	3	

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3	4	0.5
Sansa	1	5	(5)	5	6	2.5
Arya	1	0	(0)	0	1.5	-2
Robb	0	4	4	(4)	5	1.5
Bran	0	0	0	(0)	1.5	-2
Rickon	0	1	1	(1)	3	-0.5

Randomization distribution

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Rank	R_i
Jon	1	3	(3)	3	4	0.5
Sansa	1	5	(5)	5	6	2.5
Arya	1	0	(0)	0	1.5	-2
Robb	0	4	4	(4)	5	1.5
Bran	0	0	0	(0)	1.5	-2
Rickon	0	1	1	(1)	3	-0.5

$$T_{\text{rank}} = |1/3 - -1/3| = 0.67$$

Effects on outcome distributions

- Focused so far on “average” differences between groups.

Effects on outcome distributions

- Focused so far on "average" differences between groups.
- What about differences in the distribution of outcomes? $\rightsquigarrow$ Kolmogorov-Smirnov test

Effects on outcome distributions

- Focused so far on "average" differences between groups.
- What about differences in the distribution of outcomes? $\rightsquigarrow$ Kolmogorov-Smirnov test
- Define the empirical cumulative distribution function:

$$\widehat{F}_0(y) = \frac{1}{n_0} \sum_{i:D_i=0} \mathbb{1}(Y_i \leq y) \qquad \widehat{F}_1(y) = \frac{1}{n_1} \sum_{i:D_i=1} \mathbb{1}(Y_i \leq y)$$

Effects on outcome distributions

- Focused so far on “average” differences between groups.
- What about differences in the distribution of outcomes? $\rightsquigarrow$ Kolmogorov-Smirnov test
- Define the empirical cumulative distribution function:

$$\widehat{F}_0(y) = \frac{1}{n_0} \sum_{i:D_i=0} \mathbb{1}(Y_i \leq y) \qquad \widehat{F}_1(y) = \frac{1}{n_1} \sum_{i:D_i=1} \mathbb{1}(Y_i \leq y)$$

- Proportion of observed ouctomes below a chosen value for treated and control separately.

Effects on outcome distributions

- Focused so far on "average" differences between groups.
- What about differences in the distribution of outcomes? $\rightsquigarrow$ Kolmogorov-Smirnov test
- Define the empirical cumulative distribution function:

$$\widehat{F}_0(y) = \frac{1}{n_0} \sum_{i:D_i=0} \mathbb{1}(Y_i \leq y) \qquad \widehat{F}_1(y) = \frac{1}{n_1} \sum_{i:D_i=1} \mathbb{1}(Y_i \leq y)$$

- Proportion of observed ouctomes below a chosen value for treated and control separately.
- If two distributions are the same, then $\widehat{F}_0(y) = \widehat{F}_1(y)$

Kolmogorov-Smirnov statistic

- eCDFs are functions, but we need a scalar test statistic.

Kolmogorov-Smirnov statistic

- eCDFs are functions, but we need a scalar test statistic.
- Use the maximum discrepancy between the two eCDFs:

$$T_{\text{KS}} = \max_i |\widehat{F}_1(Y_i) - \widehat{F}_0(Y_i)|$$

Kolmogorov-Smirnov statistic

- eCDFs are functions, but we need a scalar test statistic.
- Use the maximum discrepancy between the two eCDFs:

$$T_{\text{KS}} = \max_i |\widehat{F}_1(Y_i) - \widehat{F}_0(Y_i)|$$

- Summary of how different the two distributions are.

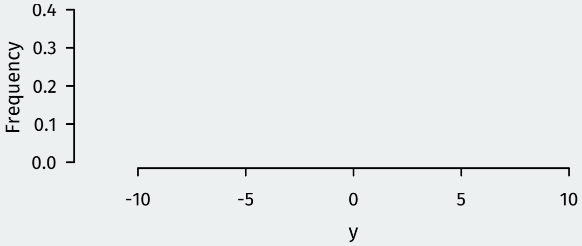
Kolmogorov-Smirnov statistic

- eCDFs are functions, but we need a scalar test statistic.
- Use the maximum discrepancy between the two eCDFs:

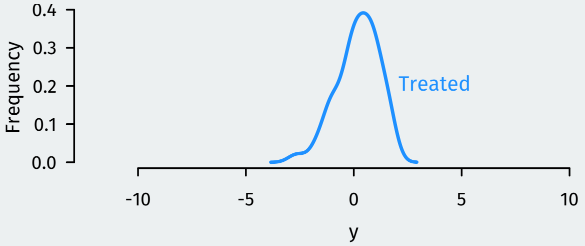
$$T_{\text{KS}} = \max_i |\widehat{F}_1(Y_i) - \widehat{F}_0(Y_i)|$$

- Summary of how different the two distributions are.
- Useful in many contexts!

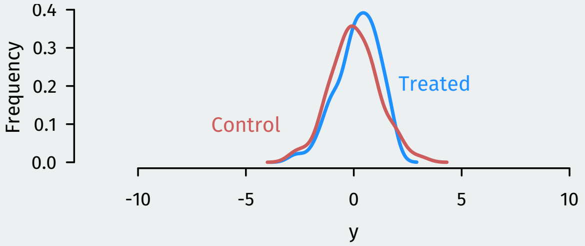
KS statistic, similar



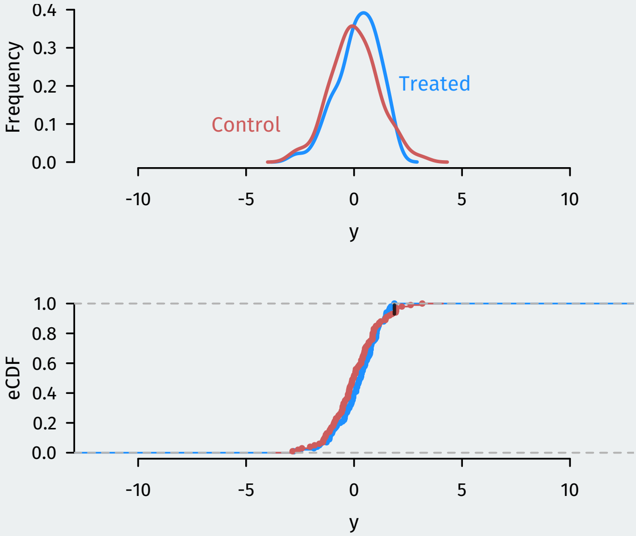
KS statistic, similar



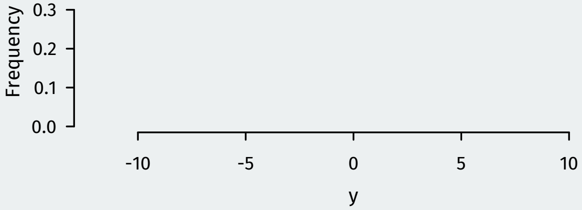
KS statistic, similar



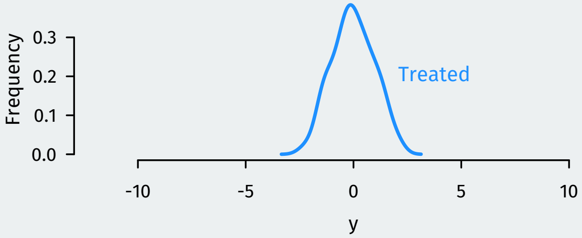
KS statistic, similar



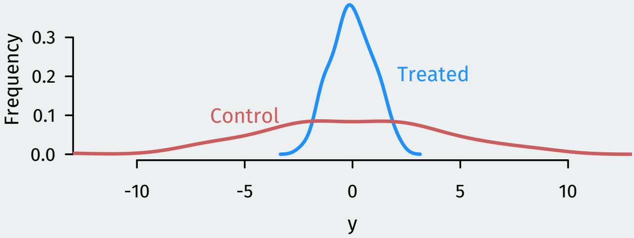
KS statistic, different



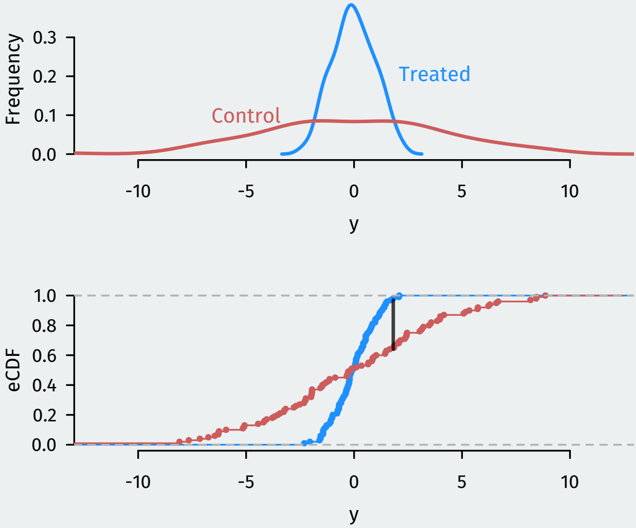
KS statistic, different



KS statistic, different



KS statistic, different



Two-sided or one-sided?

- So far, we have defined all test statistics as absolute values.

Two-sided or one-sided?

- So far, we have defined all test statistics as absolute values.
- $\rightsquigarrow$ testing against a two-sided alternative hypothesis:

$$H_0 : \tau_i = 0 \; \forall i \qquad H_1 : \tau_i \neq 0 \text{ for some } i$$

Two-sided or one-sided?

- So far, we have defined all test statistics as absolute values.
- $\rightsquigarrow$ testing against a two-sided alternative hypothesis:

$$H_0 : \tau_i = 0 \ \forall i \qquad H_1 : \tau_i \neq 0 \text{ for some } i$$

- What about a one-sided alternative?

$$H_0 : \tau_i = 0 \ \forall i \qquad H_1 : \tau_i > 0 \text{ for some } i$$

Two-sided or one-sided?

- So far, we have defined all test statistics as absolute values.
- $\rightsquigarrow$ testing against a two-sided alternative hypothesis:

$$H_0 : \tau_i = 0 \; \forall i \qquad H_1 : \tau_i \neq 0 \text{ for some } i$$

- What about a one-sided alternative?

$$H_0 : \tau_i = 0 \; \forall i \qquad H_1 : \tau_i > 0 \text{ for some } i$$

- For these, use a test statistic that is bigger under the alternative:

$$T^*_{\text{diff}} = \bar{Y}_t - \bar{Y}_c$$

3/ Confidence intervals in randomization inference

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.
- Sharp null in general is one of a constant additive effect: $H_0 : \tau_i = 0.2$.

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.
- Sharp null in general is one of a constant additive effect: $H_0 : \tau_i = 0.2$.
  - Implies that $Y_i(1) = Y_i(0) + 0.2$.

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.
- Sharp null in general is one of a constant additive effect: $H_0 : \tau_i = 0.2$.
    - Implies that $Y_i(1) = Y_i(0) + 0.2$.
    - Can still calculate all the potential outcomes!

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.
- Sharp null in general is one of a constant additive effect: $H_0 : \tau_i = 0.2$.
    - Implies that $Y_i(1) = Y_i(0) + 0.2$.
    - Can still calculate all the potential outcomes!
- More generally, we could have $H_0 : \tau_i = \tau_0$ for a fixed τ_0

Other sharp nulls

- Sharp null of no effect is not the only sharp null of no effect.
- Sharp null in general is one of a constant additive effect: $H_0 : \tau_i = 0.2$.
    - Implies that $Y_i(1) = Y_i(0) + 0.2$.
    - Can still calculate all the potential outcomes!
- More generally, we could have $H_0 : \tau_i = \tau_0$ for a fixed τ_0
- Complications: why constant and additive?

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:
    - A $100(1-\alpha)\%$ confidence interval is equivalent to the set of null hypotheses that **would not be rejected** at the α significance level.

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:
  - A $100(1-\alpha)\%$ confidence interval is equivalent to the set of null hypotheses that **would not be rejected** at the α significance level.
- 95% CI: find all values τ_0 such that $H_0 : \tau = \tau_0$ is not rejected at the 0.05 level.

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:
    - A $100(1-\alpha)\%$ confidence interval is equivalent to the set of null hypotheses that **would not be rejected** at the α significance level.
- 95% CI: find all values τ_0 such that $H_0 : \tau = \tau_0$ is not rejected at the 0.05 level.
    - Choose grid across space of τ: $-0.9, -0.8, -0.7, \dots, 0.7, 0.8, 0.9$.

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:
  - A $100(1-\alpha)\%$ confidence interval is equivalent to the set of null hypotheses that **would not be rejected** at the α significance level.
- 95% CI: find all values τ_0 such that $H_0 : \tau = \tau_0$ is not rejected at the 0.05 level.
  - Choose grid across space of τ: $-0.9, -0.8, -0.7, \dots, 0.7, 0.8, 0.9$.
  - For each value, use RI to test sharp null of $H_0 : \tau_i = \tau_m$ at 0.05 level.

Confidence intervals via test inversion

- CIs usually justified using Normal distributions and approximations.
- Can calculate CIs here using the duality of tests and Cis:
  - A $100(1-\alpha)\%$ confidence interval is equivalent to the set of null hypotheses that **would not be rejected** at the α significance level.
- 95% CI: find all values τ_0 such that $H_0 : \tau = \tau_0$ is not rejected at the 0.05 level.
  - Choose grid across space of τ: $-0.9, -0.8, -0.7, \ldots, 0.7, 0.8, 0.9$.
  - For each value, use RI to test sharp null of $H_0 : \tau_i = \tau_m$ at 0.05 level.
  - Collect all values that you cannot reject as the 95% CI.

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

	Mailer	Contr.		
Unit	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	?	3
Sansa	1	5	?	5
Arya	1	0	?	0
Robb	0	4	4	?
Bran	0	0	0	?
Rickon	0	1	1	?

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

	Mailer	Contr.		
Unit	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(2)	3
Sansa	1	5	(4)	5
Arya	1	0	(-1)	0
Robb	0	4	4	(5)
Bran	0	0	0	(1)
Rickon	0	1	1	(2)

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$
Jon	1	3	(2)	3
Sansa	1	5	(4)	5
Arya	1	0	(-1)	0
Robb	0	4	4	(5)
Bran	0	0	0	(1)
Rickon	0	1	1	(2)

- Assignments will now affect Y_i.

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Adjusted $Y_i - D_i\tau_0$
Jon	1	3	(2)	3	2
Sansa	1	5	(4)	5	4
Arya	1	0	(-1)	0	-1
Robb	0	4	4	(5)	4
Bran	0	0	0	(1)	0
Rickon	0	1	1	(2)	1

- Assignments will now affect Y_i.
- Solution: use **adjusted outcomes**, $Y_i^* = Y_i - D_i\tau_0$.

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Adjusted $Y_i - D_i\tau_0$
Jon	1	3	(2)	3	2
Sansa	1	5	(4)	5	4
Arya	1	0	(-1)	0	-1
Robb	0	4	4	(5)	4
Bran	0	0	0	(1)	0
Rickon	0	1	1	(2)	1

- Assignments will now affect Y_i.
- Solution: use **adjusted outcomes**, $Y_i^* = Y_i - D_i\tau_0$.
- Now, just test sharp null of no effect for Y_i^*.

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Adjusted $Y_i - D_i\tau_0$
Jon	1	3	(2)	3	2
Sansa	1	5	(4)	5	4
Arya	1	0	(-1)	0	-1
Robb	0	4	4	(5)	4
Bran	0	0	0	(1)	0
Rickon	0	1	1	(2)	1

- Assignments will now affect Y_i.
- Solution: use **adjusted outcomes**, $Y_i^* = Y_i - D_i\tau_0$.
- Now, just test sharp null of no effect for Y_i^*.
    - $Y_i^*(1) = Y_i(1) - 1 \times 1 = Y_i(0)$

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Adjusted $Y_i - D_i\tau_0$
Jon	1	3	(2)	3	2
Sansa	1	5	(4)	5	4
Arya	1	0	(-1)	0	-1
Robb	0	4	4	(5)	4
Bran	0	0	0	(1)	0
Rickon	0	1	1	(2)	1

- Assignments will now affect Y_i.
- Solution: use **adjusted outcomes**, $Y_i^* = Y_i - D_i\tau_0$.
- Now, just test sharp null of no effect for Y_i^*.
    - $Y_i^*(1) = Y_i(1) - 1 \times 1 = Y_i(0)$
    - $Y_i^*(0) = Y_i(0) - 0 \times 1 = Y_i(0)$

Testing non-zero sharp nulls

- Suppose that we had: $H_0 : \tau_i = Y_i(1) - Y_i(0) = 1$

Unit	Mailer D_i	Contr. Y_i	$Y_i(0)$	$Y_i(1)$	Adjusted $Y_i - D_i\tau_0$
Jon	1	3	(2)	3	2
Sansa	1	5	(4)	5	4
Arya	1	0	(-1)	0	-1
Robb	0	4	4	(5)	4
Bran	0	0	0	(1)	0
Rickon	0	1	1	(2)	1

- Assignments will now affect Y_i.
- Solution: use **adjusted outcomes**, $Y_i^* = Y_i - D_i\tau_0$.
- Now, just test sharp null of no effect for Y_i^*.
    - $Y_i^*(1) = Y_i(1) - 1 \times 1 = Y_i(0)$
    - $Y_i^*(0) = Y_i(0) - 0 \times 1 = Y_i(0)$
    - $\tau_i^* = Y_i^*(1) - Y_i^*(0) = 0$

Notes on RI CIs

- CIs are correct, but might have **overcoverage**.

Notes on RI CIs

- CIs are correct, but might have **overcoverage**.
- With RI, p-values are discrete and depend on n and n_1.

Notes on RI CIs

- CIs are correct, but might have **overcoverage**.
- With RI, p-values are discrete and depend on n and n_1.
    - With n and n_1, the lowest p-value is $1/20$.

Notes on RI CIs

- CIs are correct, but might have **overcoverage**.
- With RI, p-values are discrete and depend on n and n_1.
    - With n and n_1, the lowest p-value is $1/20$.
    - Next lowest p-value is $2/20 = 0.10$.

Notes on RI CIs

- CIs are correct, but might have **overcoverage**.
- With RI, p-values are discrete and depend on n and n_1.
    - With n and n_1, the lowest p-value is $1/20$.
    - Next lowest p-value is $2/20 = 0.10$.
- If the p-value of 0.05 falls "between" two of these discrete points, a 95% CI will cover the true value more than 95% of the time.

Point estimates

- Is it possible to get point estimates?

Point estimates

- Is it possible to get point estimates?
- Not really the point of RI, but still possible:

Point estimates

- Is it possible to get point estimates?
- Not really the point of RI, but still possible:
    1. Create a grid of possible sharp null hypotheses.

Point estimates

- Is it possible to get point estimates?
- Not really the point of RI, but still possible:
    1. Create a grid of possible sharp null hypotheses.
    2. Calculate p-values for each sharp null.

Point estimates

- Is it possible to get point estimates?
- Not really the point of RI, but still possible:
    1. Create a grid of possible sharp null hypotheses.
    2. Calculate p-values for each sharp null.
    3. Pick the value that is “least surprising” under the null.

Point estimates

- Is it possible to get point estimates?
- Not really the point of RI, but still possible:
    1. Create a grid of possible sharp null hypotheses.
    2. Calculate p-values for each sharp null.
    3. Pick the value that is “least surprising” under the null.
- Usually this means selecting the value with the highest p-value.

Including covariate information

- Let X_i be a pretreatment measure of the outcome.

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.
- Causal effects are the same: $Y_i'(1) - Y_i'(0) = Y_i(1) - Y_i(0)$.

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.
- Causal effects are the same: $Y_i'(1) - Y_i'(0) = Y_i(1) - Y_i(0)$.
- But the test statistic is different:

$$T_{\text{gain}} = \left|(\bar{Y}_t - \bar{Y}_c) - (\bar{X}_t - \bar{X}_c)\right|$$

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.
- Causal effects are the same: $Y_i'(1) - Y_i'(0) = Y_i(1) - Y_i(0)$.
- But the test statistic is different:

$$T_{\text{gain}} = \left|(\bar{Y}_t - \bar{Y}_c) - (\bar{X}_t - \bar{X}_c)\right|$$

- If X_i is strongly predictive of $Y_i(0)$, then this could have higher power:

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.
- Causal effects are the same: $Y_i'(1) - Y_i'(0) = Y_i(1) - Y_i(0)$.
- But the test statistic is different:

$$T_{\text{gain}} = \left|(\bar{Y}_t - \bar{Y}_c) - (\bar{X}_t - \bar{X}_c)\right|$$

- If X_i is strongly predictive of $Y_i(0)$, then this could have higher power:
    - T_{gain} will have lower variance under the null.

Including covariate information

- Let X_i be a pretreatment measure of the outcome.
- One way is to use this is as a **gain score**: $Y_i'(d) = Y_i(d) - X_i$.
- Causal effects are the same: $Y_i'(1) - Y_i'(0) = Y_i(1) - Y_i(0)$.
- But the test statistic is different:

$$T_{\text{gain}} = \left|(\bar{Y}_t - \bar{Y}_c) - (\bar{X}_t - \bar{X}_c)\right|$$

- If X_i is strongly predictive of $Y_i(0)$, then this could have higher power:
    - T_{gain} will have lower variance under the null.
    - $\rightsquigarrow$ easier to detect smaller effects.

Using regression in RI

- We can extend this to use covariates in more complicated ways.

Using regression in RI

- We can extend this to use covariates in more complicated ways.
- For instance, we can use an OLS regression:

$$(\hat{\beta}_0, \hat{\beta}_D, \hat{\beta}_X) = \underset{\beta_0, \beta_D, \beta_X}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_D \cdot D_i - \beta_X \cdot X_i\right)^2.$$

Using regression in RI

- We can extend this to use covariates in more complicated ways.
- For instance, we can use an OLS regression:

$$(\hat{\beta}_0, \hat{\beta}_D, \hat{\beta}_X) = \underset{\beta_0, \beta_D, \beta_X}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_D \cdot D_i - \beta_X \cdot X_i\right)^2.$$

- Then, our test statistic could be $T_{\text{ols}} = \hat{\beta}_D$.

Using regression in RI

- We can extend this to use covariates in more complicated ways.
- For instance, we can use an OLS regression:

$$(\hat{\beta}_0, \hat{\beta}_D, \hat{\beta}_X) = \underset{\beta_0, \beta_D, \beta_X}{\arg\min} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_D \cdot D_i - \beta_X \cdot X_i)^2.$$

- Then, our test statistic could be $T_{\text{ols}} = \hat{\beta}_D$.
- RI is justified **even if the model is wrong!**

Using regression in RI

- We can extend this to use covariates in more complicated ways.
- For instance, we can use an OLS regression:

$$(\hat{\beta}_0, \hat{\beta}_D, \hat{\beta}_X) = \underset{\beta_0, \beta_D, \beta_X}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_D \cdot D_i - \beta_X \cdot X_i\right)^2.$$

- Then, our test statistic could be $T_{\text{ols}} = \hat{\beta}_D$.
- RI is justified **even if the model is wrong!**
    - OLS is just another way to generate a test statistic.

Using regression in RI

- We can extend this to use covariates in more complicated ways.
- For instance, we can use an OLS regression:

$$(\hat{\beta}_0, \hat{\beta}_D, \hat{\beta}_X) = \underset{\beta_0, \beta_D, \beta_X}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_D \cdot D_i - \beta_X \cdot X_i\right)^2.$$

- Then, our test statistic could be $T_{\text{ols}} = \hat{\beta}_D$.
- RI is justified **even if the model is wrong!**
    - OLS is just another way to generate a test statistic.
    - If the model is "right" (read: predictive of $Y_i(0)$), then T_{ols} will have higher power.