Sharp null hypothesis testing

Thomas Leavitt

July 20, 2022

Recap

Randomization and potential outcomes schedules

- Yesterday we introduced two important concepts:

    1. Random assignment and
    2. potential outcomes schedule

- Today we will focus on how we can use randomization as the basis for testing hypotheses about potential outcomes schedules

Hypothesis testing

Hypothesis testing: Introduction

- Steps of hypothesis testing:

  1. From randomized experiment, we observe data, $(\boldsymbol{z}, \boldsymbol{y})$, and summarize data by test-stat, $t(\boldsymbol{z}, \boldsymbol{y})$

  2. For purposes of argumentation, we postulate a sharp null hypothesis

     A sharp null hypothesis implies complete specification of unit-level responses to experiment under every possible assignment

  3. Under sharp null hypothesis, calculate test-stat over all assignments

  4. Compare observed test-stat in (1) to distribution of test-stats under null in (3)

  5. If observed test-stat inconsistent with distribution of test-stat implied by sharp null, then reject sharp null; otherwise, don't

Errors and goals of hypothesis tests

- Type I Error: Rejecting null when null is true
- Type II: Failing to reject null when null is false
- Hence, two goals of our tests:

  1. Control the Type I Error: $\Pr(\text{Type I Error}) \leq \alpha$, where α is size of test
  2. Control Type II Error: Make power as large as possible, where power is $1 - \Pr(\text{Type II Error})$

- Definitions:

  Unbiased Test: $\text{Power} \geq \Pr(\text{Type I Error})$

  Consistent Test: $\text{Power} \to 1$ as size of experiment $\to \infty$

Test of Sharp Null of No Effects

"Lady Tasting Tea" Example

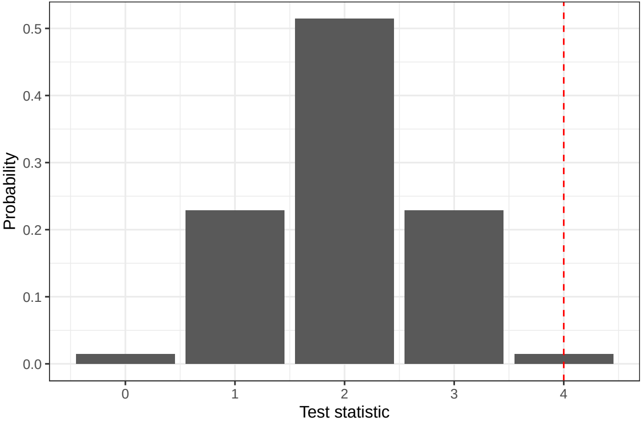

Figure 1: Distribution of test-stat over assignments under sharp null of no effects

- Upper p-value is $1/70 \approx 0.0143$, where upper p-value of sharp null given by

$$\Pr\left(t\left(\boldsymbol{Z}, \boldsymbol{y}\right) \geq t^{\text{obs}}\right) = \sum_{\boldsymbol{z} \in \Omega} \mathbb{1}\left\{t\left(\boldsymbol{Z}, \boldsymbol{y}\right) \geq t^{\text{obs}}\right\} \Pr\left(\boldsymbol{Z} = \boldsymbol{z}\right)$$

“Lady Tasting Tea” Example

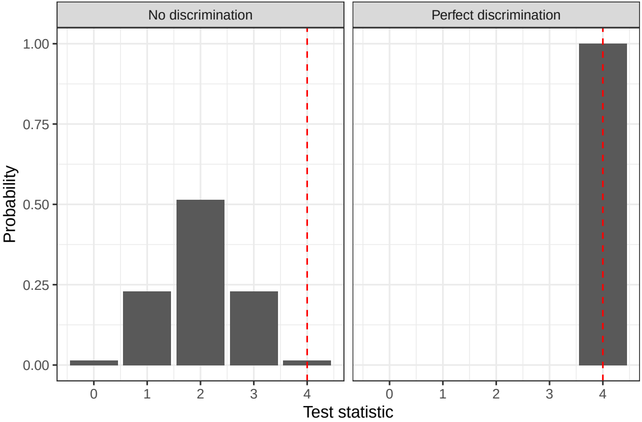

Figure 2: Distribution of test-stat over assignments under two

- Imagine that we test sharp null of no effects when (unbeknownst to us) true effect is that of perfect discrimination. What is power of the test?

Tests of general sharp null hypotheses

A simple model of effects for the Acorn GOTV experiment

- Consider ACORN GOTV experiment by Arceneaux (2005)
- Fisher's sharp null hypothesis of no effect states that individual effect is $p = 0$ percentage points for all precincts

	GOTV?	vote03(%)	y_C	y_T	τ
1	0	38	38	38	0
$\vdots$					
13	0	19	19	19	0
14	0	34	34	34	0
15	1	49	49	49	0
16	1	38	38	38	0
$\vdots$					
28	1	29	29	29	0

Distribution of test-stat under sharp null of no effect

- To get a p-value, we could exactly enumerate all assignments, Ω
- But with $\binom{28}{14} = 40,116,600$, this is too computationally intensive
- Instead, we randomly sample from set of $\binom{28}{14}$ possible assignments
- Then calculate test-stat under each assignment holding outcomes fixed

  E.g., Diff-in-Means $t(\boldsymbol{z}, \boldsymbol{y}) = n_T^{-1} \boldsymbol{z}^\top \boldsymbol{y} - n_C^{-1} (1 - \boldsymbol{z})^\top \boldsymbol{y}$

  Note this test-stat is not same as one in Fisher's "Lady Tasting Tea"

- Finally, calculate p-value

$$\Pr\left(t(\boldsymbol{Z}, \boldsymbol{y}) \geq t^{\text{obs}}\right) = \sum_{\boldsymbol{z} \in \Omega} \mathbb{1}\left\{t(\boldsymbol{z}, \boldsymbol{y}) \geq t^{\text{obs}}\right\} \Pr(\boldsymbol{Z} = \boldsymbol{z})$$

Distribution of test-stat under sharp null of no effect

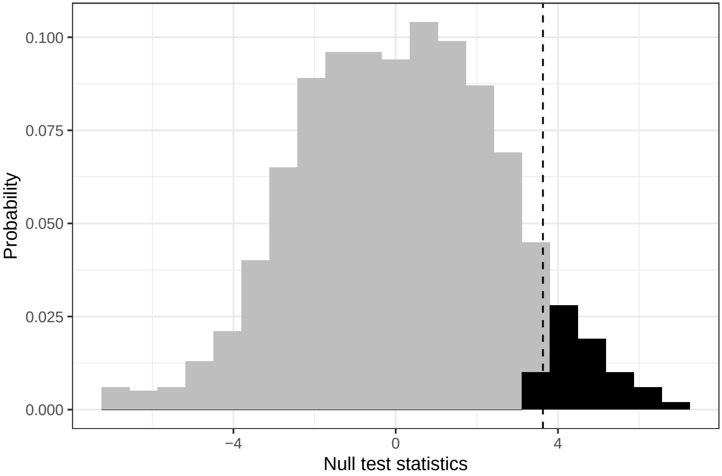

Figure 3: Distribution of Difference-in-Means under sharp null of no effects

Hypothesis tests in adjusted outcomes

- How do we test hypotheses other than no effects for all units?
- Rosenbaum (2002, 2010, 2017): Write units' true adjusted outcomes as $\tilde{y}_i = y_i - \tau_i z_i$, where $\tau_i = y_{Ti} - y_{Ci}$ for $i = 1, \ldots, n$,
- $\tilde{y}_i$ is fixed for every unit regardless if assigned to treatment or control (i.e., $\tilde{\boldsymbol{y}}$ satisfies sharp null of no effects)
- So to conduct a test about $\boldsymbol{\tau}$, we can compare $t(\boldsymbol{z}, \tilde{\boldsymbol{y}}_h)$ to the randomization distribution of sharp null of no effects on adjusted outcomes, $t(\boldsymbol{Z}, \tilde{\boldsymbol{y}}_h)$, where $\tilde{y}_{hi} = y_i - \tau_{hi} z_i$ for all $i = 1, \ldots, n$
- Intuition: Can we make outcomes appear as if there is no effect by removing hypothetical effect from treated units? If so, then this is evidence in favor of that hypothetical effect

Hypothesis tests in adjusted outcomes

- H_0: GOTV campaign increases voter turnout by p percentage points per precinct

	GOTV?	vote03(%)	$\tilde{\boldsymbol{y}}_h$	$\boldsymbol{y_C}$	$\boldsymbol{y_T}$	$\boldsymbol{\tau}$
1	0	38	38	38	?	?
$\vdots$						
13	0	19	19	19	?	?
14	0	34	34	34	?	?
15	1	49	49 - p	?	49	?
16	1	38	38 - p	?	38	?
$\vdots$						
28	1	29	29 - p	?	29	?

Hypothesis tests in adjusted outcomes

- For example, H_0: $p = 2.5$

	GOTV?	vote03(%)	$\tilde{\boldsymbol{y}}_h$	$\boldsymbol{y_C}$	$\boldsymbol{y_T}$	$\boldsymbol{\tau}$
1	0	38	38	38	?	?
⋮						
13	0	19	19	19	?	?
14	0	34	34	34	?	?
15	1	49	46.5	?	49	?
16	1	38	35.5	?	38	?
⋮						
28	1	29	26.5	?	29	?

- The observed test-stat calculated on adjusted outcomes is $t(\boldsymbol{z}, \tilde{\boldsymbol{y}}_h) \approx 1.13$
- How does it compare to $t\left(\boldsymbol{Z}, \tilde{\boldsymbol{y}}_h\right)$?

Hypothesis tests in adjusted outcomes

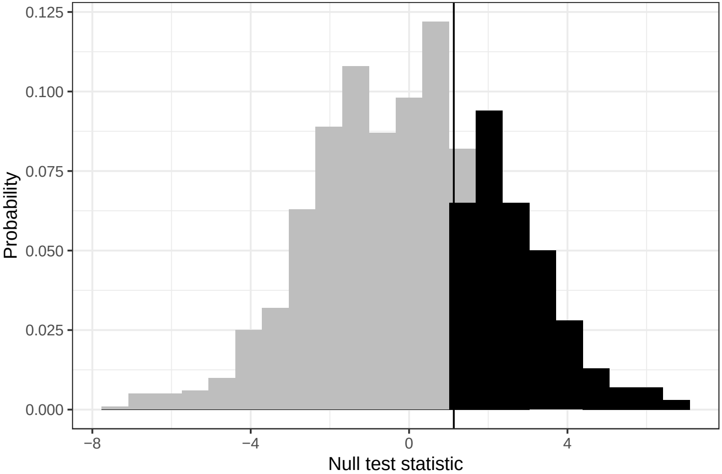

Figure 4: Distribution of Difference-in-Means under sharp null of no effects with outcomes adjusted by $p = 2.5$

Confidence sets

- To get a confidence set we do what we just did with $p = 2.5$ over an entire grid of values of p
- A $1 - \alpha$ confidence set for $p = \{p$: H_p not rejected at level $\alpha\}$
- We test over all values of p and retain those we fail to reject with adjusted outcomes
- Two sided confidence set for ACORN example: $\{-0.4, 7.5\}$
- We fail to reject sharp null of no effect

References

Arceneaux, K. (2005). Using cluster randomized field experiments to study voting behavior. *The Annals of the American Academy of Political and Social Science 601*(1), 169–179.

Rosenbaum, P. R. (2002). *Observational Studies* (2nd ed.). New York, NY: Springer.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York, NY: Springer.

Rosenbaum, P. R. (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press.