Statistical Inference and Interpretation

Tara Slough

June 12, 2019

Learning Days XI, African School of Economics

Review: What is an Estimand?

- A quantity of interest that summarizes the data
- Example: the average treatment effect (ATE)
  - The difference between average potential outcomes in treatment and control.

$$\tau \equiv E[Y_i(1) - Y_i(0)] \tag{1}$$

- But there are many others: LATEs, CATEs, ATT, ATC and more(!)

Statistical Inference

- Inference: reasoning about the unobserved
- What is unobserved?
  - Unrevealed potential outcomes: **?**
  - Population that is not in the sample: **?**

Subject	Sample	Z_i	$Y_i(1)$	$Y_i(0)$
1	Yes	1	4	**?**
2	No		**?**	**?**
3	Yes	0	**?**	1
4	Yes	0	**?**	2
5	No		**?**	**?**
6	Yes	1	3	**?**

Causal Inference

- Causal inference $\rightarrow$ identification of causal effects
    1. Could we recover a causal estimand (parameter) in the presence of infinite data?
    2. How do we make inferences with finite data?
- Focus today on $\#2$
    - Estimating the ATE (refresher)
    - Quantifying uncertainty (uncertainty emerges because of unobserved data)
    - Making inferences
    - Interpreting reported estimates

The Logic of Frequentist Statistics

In hypothetical Frequentistland we would:

1. Do experiment
   1.1 Estimate the estimand, for example, the ATE
   1.2 (Calculate relevant test statistic)
2. Repeat #1 many, many times.
3. Construct the sampling distribution of the estimate or test statistic.

Given the (single) experiment we actually:

1. Compare statistic to the sampling distribution under H_0.
2. Compute p-value.
3. Reject or fail to reject H_0

Fisher v. Neyman

	Fischer	Neyman
H_0	*Sharp null*: The treatment effect is 0 for all subjects.	*Null*: The average treatment effect is 0.
Sampling distribution under H_0	ATE calculated under many permutations of treatment assignment.	Central limit theorem provides asymptotic (as $N \to \infty$) distribution of test statistics.

- *p*-values based on **t-tests**, **regression** (unless otherwise stated) use Neyman inference.
    - As N increases, inferences under Neyman and Fischer become very similar.
    - Estimate of ATE is the same!

Signal and Noise

- Statistical inference can be thought of as distinguishing **signal** from **noise**
  - **Signal**: the estimate
  - **Noise**: uncertainty about the estimate
- Focus here on inference on the ATE, but these properties are applicable to other estimands in experiments, other research designs.

Signal: Estimating the ATE

- Difference-in-means estimator

$$\widehat{\tau} = \overline{Y_i}(Z_i = 1) - \overline{Y_i}(Z_i = 0) \tag{2}$$

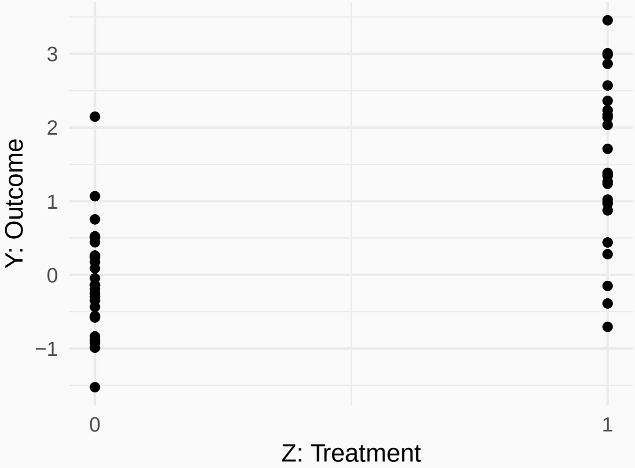
  - This is what we have been doing all week!
- Other ways to estimate the ATE: regression
  - Estimate, via OLS

$$Y_i = \beta_0 + \tau Z_i + \epsilon_i \tag{3}$$

- $\hat{\tau}$ from difference in mean $=$ $\hat{\tau}$ from OLS (univariate setting).
- Nonlinear models do not (directly) estimate the ATE.

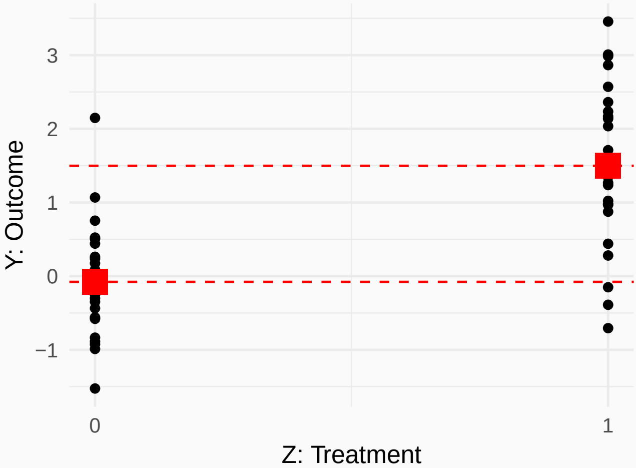
Analogue to Regression

- Visualization of data from two-arm experiment



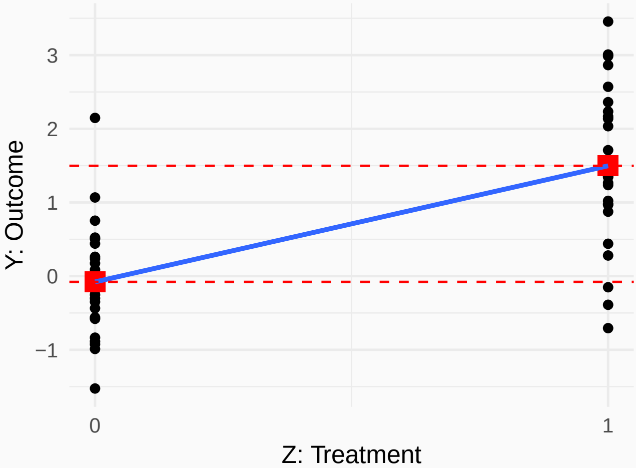
Analogue to Regression

- Visualization of a difference-in-means estimate



Analogue to Regression

- Difference-in-means estimate = univariate OLS estimate



Quantifying the noise: the Standard Error

- A statistic measuring sampling variability

Subject	Sample	Z_i	$Y_i(1)$	$Y_i(0)$
1	Yes	1	4	**?**
2	No		**?**	**?**
3	Yes	0	**?**	1
4	Yes	0	**?**	2
5	No		**?**	**?**
6	Yes	1	3	**?**

Quantifying the noise: the Standard Error

- A statistic measuring sampling variability
- Standard deviation of the sampling distribution about the estimate
- Conservative formula for the standard error of $\widehat{ATE}$:
  - m subjects in treatment, $N - m$ subjects in control:

$$\widehat{SE}_{\tau} = \sqrt{\frac{\widehat{Var}(Y_i(0))}{N - m} + \frac{\widehat{Var}(Y_i(1))}{m}} \tag{4}$$

  - *This formula is for experiments with simple or complete random assignment!*

Refresher: Variance

- The variance of a random sample of a variable, X, of size N is:

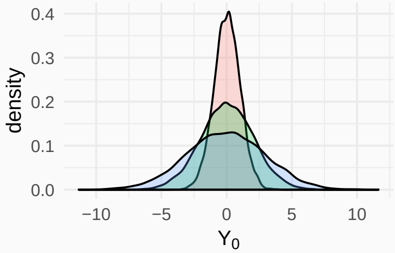
$$Var(X) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - E[X_i])^2 \tag{5}$$

- Consider 5 realized values of $Y_i(0) = \{1, 2, 3, 4, 5\}$
  - $E[Y_i(0)] = 3$
  - The variance is $\frac{1}{4}(4 + 1 + 1 + 4) = 2.5$
- As dispersion grows, so does variance:
  - If $Y_i(1) = \{-1, 1, 3, 5, 7\}$, $E[Y_i(1)] = 3$ and $Var(Y_i(1)) = 10$.
  - Informally, we say this is "noisier"

Standard Error, ctd.

$$\widehat{SE}_{\tau} = \sqrt{\frac{\widehat{Var}(Y_i(0))}{N-m} + \frac{\widehat{Var}(Y_i(1))}{m}} \tag{6}$$

- As sample size in each group $\uparrow$, $N - m$ and m, $\widehat{SE}_{\tau} \downarrow$



- As variance $\uparrow$, $\widehat{SE}_{\widehat{ATE}} \uparrow$
- Lowest variance in pink
- Lowest variance in blue

Inference and Standard Errors

- Form a test statistic:
  - The ratio of signal to noise is referred to as a Z-statistic.
  - Under H_0 of no effect on average:

$$\frac{\tau}{SE_\tau} \text{ very closely approximates standard normal}$$

Inference and Standard Errors

- Form a test statistic:
  - The ratio of signal to noise is referred to as a Z-statistic.
  - Under H_0 of no effect on average:

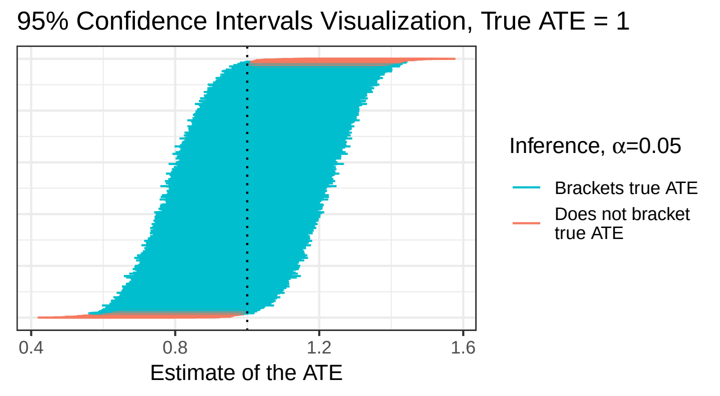
$$\frac{\tau}{SE_\tau} \text{ very closely approximates standard normal}$$

Inference and Standard Errors, ctd.

- In a two-tailed test, we reject H_0 at the $\alpha = 0.05$ level if:
    - $\frac{\widehat{\tau}}{\widehat{SE}_\tau} > 1.96$
    - $\frac{\widehat{\tau}}{\widehat{SE}_\tau} < -1.96$
- Intuition:
    - If signal is strong enough (either positive or negative) relative to the noise, we reject the null hypothesis of zero average effect.

Confidence Intervals

- Form confidence intervals
  - Confidence intervals: by convention we estimate 95% CIs
  - Interval that has a 95% (1-α) probability of bracketing the true (unknown) ATE.



Confidence Intervals, ctd.

- Confidence interval for $\widehat{\tau}$:

$$[\widehat{\tau} - 1.96 \times \widehat{SE}_\tau, \widehat{\tau} + 1.96 \times \widehat{SE}_\tau]$$

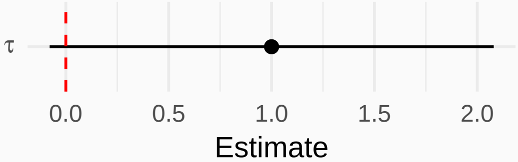
- So if $\widehat{\tau} = 1$ and $\widehat{SE_\tau} = .55$ ...

Confidence Intervals, ctd.

- Confidence interval for $\widehat{\tau}$:

$$[\widehat{\tau} - 1.96 \times \widehat{SE}_{\tau}, \widehat{\tau} + 1.96 \times \widehat{SE}_{\tau}]$$

- So if $\widehat{\tau} = 1$ and $\widehat{SE_{\tau}} = .55$ ...
- $CI_{\tau} = [1 - 1.96 \times .55, 1 + 1.96 \times .55] = [-0.078, 2.078]$



- What does it mean if a confidence interval bounds 0?

Caveats

- The standard error fomula here is for an experiment with simple or complete random assignment.
- If we use blocked or clustered assignment, the standard error estimator is different.
- In practice, most people estimate standard errors by regression:
    - For individually-randomized experiments `,robust` in Stata (heteroskedasticity robust SEs)
    - For cluster-randomized experiments: cluster robust SEs

Implications

- Do you want to be able to detect treatment effects?

$$\frac{\widehat{\tau}}{\widehat{SE_\tau}}$$

- Increase **signal**
    - Make treatments stronger
- Reduce **noise**
    - Get a bigger sample (or more clusters!)
    - Reduce variance by blocking or covariate adjustment

Relation to Randomization Inference

- What is the *same*:
  - Our estimator, estimate of the ATE ($\widehat{\tau}$)
  - We reject/fail to reject a null hypothesis by comparing the $\widehat{\tau}$ to a probability distribution under the null hypothesis.
- What is definitely *different*:
  - The null hypothesis: sharp null for RI.
  - The construction of the null distribution: in RI, the null distribution is constructed by permutation tests.
  - Construction of CIs.

Interpreting Reported Evidence

- Green et al. (2019) examine a media campaign in 112 villages in rural Uganda:
  - Treatment conditions:
    - Treatment: 6 Hollywood movie screenings with anti-Violence Against Women (VAW) ads
    - Control: 6 Hollywood movie screenings without ads (placebo)
  - Outcomes: Women's survey reports of domestic violence
    - DV #1: Number of incidents
    - DV #2: Any incidents

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- What is the difference between Columns 1 and 2?

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	-0.177	-0.146	-0.346	-0.069^{***}	-0.048^{**}	-0.132^{***}
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	-0.033	0.002	-0.006	0.057	0.014	0.026

- How does the "Neyman" p-value in Column 1 compare to the RI p-value?

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- What should the authors conclude about the “Number of Incidents” measure?

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- How do we interpret the -0.069*** in Column 4?

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- What does the "Control Mean" indicate in Column 4?

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- Construct and interpret 95% CIs on the estimate reported in Column 6.

Results

	Number of Incidents			Any Incidents		
	(1)	(2)	(3)	(4)	(5)	(6)
Anti-VAW Media	−0.177	−0.146	−0.346	−0.069***	−0.048**	−0.132***
	(0.113)	(0.091)	(0.226)	(0.026)	(0.022)	(0.049)
Control Mean	0.56	0.59	0.59	0.19	0.2	0.2
RI p-values: IPV	0.128	0.159	0.138	0.009	0.038	0.007
Hypothesis	Two	Two	Two	Two	Two	Two
Sample	All W	All W	W compl.	All W	All W	W compl.
Analysis Level	Clus.	Indiv.	Indiv.	Clus.	Indiv.	Indiv.
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Observations	110	1,036	356	110	1,036	356
Adjusted R^2	−0.033	0.002	−0.006	0.057	0.014	0.026

- What should we conclude about the effect of anti-VAW messaging on the incidence of violence against women?