Statistical Significance and Power

Erick Gong (thanks to C. Null, author of original lecture)

UC Berkeley & CEGA

November 16

Schedule

- Problem Set 4 due Tuesday, Dec 1st
- Nov 24 Class (Canceled?)
- Dec 1st Final Class (Review & Group Presentations)

Today

- Power Theory
  Statistical Term
  "What are the chances that you will detect a treatment effect?"
- Power Applications
- Guest Speakers

Super Quick Stat Review

- Mean
  We can use sample averages to estimate the mean.
- Standard Deviation ("Spread")
  We can use the standard error to estimate the standard deviation.

What do we do in this class?

- Estimate the casual effect of a "treatment" (i.e. policy intervention) for a specific population
- We use methods to estimate the following

$$\mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0]$$

The average outcome of those getting the treatment vs. the average outcome of those who don't receive it.

- Average Treatment Effect (ATE)
- If treatment is randomly assigned (or as good as randomly assigned) we estimate the treatment effect

  Examples

  1) The effect of scholarships for Kenyan students on educational outcomes

  2) Clean water devices for rural households in Kenya on diarrhea

  3) HIV education for Kenyan girls attending school on sexual behavior

What do we do in this class?

- Estimate the casual effect of a "treatment" (i.e. policy intervention) for a specific population
- We use methods to estimate the following

$$\mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0]$$

  The average outcome of those getting the treatment vs. the average outcome of those who don't receive it.

- Average Treatment Effect (ATE)
- If treatment is randomly assigned (or as good as randomly assigned) we estimate the treatment effect

  Examples

  1) The effect of scholarships for Kenyan students on educational outcomes

  2) Clean water devices for rural households in Kenya on diarrhea

  3) HIV education for Kenyan girls attending school on sexual behavior

Estimation

- Quality of estimation
  - "Good Estimate" (Unbiased)
  - "Precise Estimate" (Consistency)
- Good Estimate
  1. Unbiased
  2. No Spill Overs
  3. High Quality Data (no measurement error)
- Precise Estimate?
  "How Close is the estimate to the truth?"

Sampling & Population

- Specify specific population of interest
  Students in rural areas
  Mothers with low education
- Sample from this population
  Method of sampling matters
  Size matters
- Intuition: Precision is how close estimate from sample is to actual statistic from population

Example: Estimating Avg Height in Classroom

- Population of Interest: Instructors for DeCal course
- What do we want? The mean (average) height of this population
  The true average height:

$$(G)\ 5'4'' + (E)\ 5'10'' + (T)\ 6'2'' = 5'8''$$

- Sample Size = 1
  Estimate ranges from 5'4" to 6'2"
- Sample Size = 2

$$(G+E) = 5'6''\ (G+T) = 5'7.5''\ (E+T) = 6'0''$$

- Increasing sample size generally gets estimate closer to truth

Estimating Treatment Effects

- As long as we're sampling (not using the whole population), our estimate using a sample will not be the same as using the whole population
- In fact, every sample will give us a different estimate
- Example

  Population of Interest: UC Berkeley Students

  Sample: 100 randomly sampled students

  Treatment: \$10 for each "A" - randomly give 50 students treatment, 50 control

  Outcome: GPA for semester

  $$\mathbb{E}[Y_i | T = 1] - \mathbb{E}[Y_i | T_i = 0]$$

  We estimate this using sample means

  $$\overline{GPA}_{Treat} - \overline{GPA}_{Control}$$

- What happens if we sample another 100 people?
- What happens if the difference in GPA looks like this?

  $$3.92 - 3.90 = .02$$

Statistical Testing

Average Treatment Effects

- Can Express as a difference in two means

$$\mathbb{E}[Y_i | T = 1] - \mathbb{E}[Y_i | T_i = 0] = \beta$$

- Or in a regression

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

- It's the same!
- Null Hypothesis $\beta = 0$
  In other words, let's initially assume that there is no effect.
- Can we reject the null hypothesis?
  Loosely speaking, is the null hypothesis incorrect?

Estimate Effect

$$\mathbb{E}[Y_i|T=1] - \mathbb{E}[Y_i|T_i=0] = \beta$$

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

- Using sample data, we estimate β which is known as $\hat{\beta}$. Everything with a hat over it ˆ is known as an estimate (using sample data).
- What would our estimates of β be under the null hypothesis? Null $\beta = 0$
- What would our estimates of β be under the alternative hypothesis?

Statistical Significance

- Our estimate $\hat{\beta}$ will have a confidence interval
  Intuition: If we took repeated samples, 95% of our estimates would fall in this interval
  AND: In 5% of our samples, our estimate would fall out of this interval.
- Is the null hypothesis in this interval? If not, we can reject the null at the 5% statistical significance level.
  Null hypothesis: $\beta = 0$

Confidence Intervals

- Confidence Intervals (95%)

$$\hat{\beta} \pm 1.96 \times SE$$

  Insert nice formula for standard errors
- Confidence Intervals (99%) Level

$$\hat{\beta} \pm 1.96 \times SE$$

- This Statistically Significance Level is 5%, 1% or the probability that we reject the null when it is actually true. (Type I error)
- The 5% level is generally accepted
  Intuition: If we had multiple samples, 95% of them would fall into our interval, and we would feel fairly confident that this is the actual effect.

STATA Table

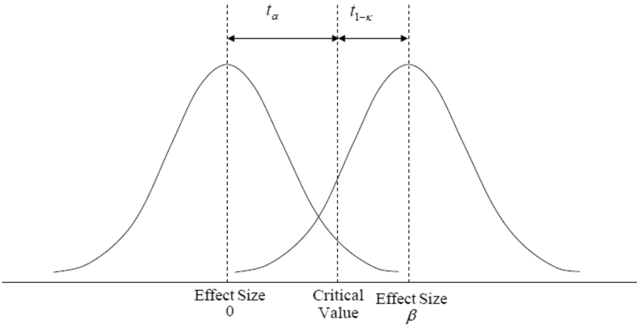
- Using data from PS 4

```
. reg education treat, robust

Linear regression                               Number of obs =      1199
                                                  F(  1,  1197) =     13.98
                                                  Prob > F      =    0.0002
                                                  R-squared     =    0.0099
                                                  Root MSE      =    3.2033

------------------------------------------------------------------------------
             |               Robust
   education |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .7778711    .208044     3.74   0.000     .3696996    1.186043
       _cons |    5.82678   .1072629    54.32   0.000     5.616336    6.037224
------------------------------------------------------------------------------
```

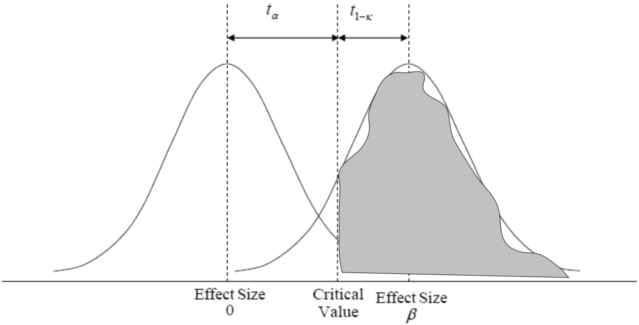
So What is Power?

- Before we collect data
- Definition: The probability that for a given treatment effect we will be able to reject the null hypothesis of zero effect.
- Example: Cash Transfers & GPA
  If we assumed that the true effect of cash for grades was .3 GPA points, if we did a study using a random sample, what is the likelihood of us rejecting the null and concluding that there was a positive effect?
- More Power is good
  90% Power => 90% of the time we would detect a statistically significant effect
  99% Power => 99% of the time we would detect a statistically significant effect

Graphically



Power


t_α
$t_{1-\kappa}$
Effect Size
0
Critical
Value
Effect Size
β

What affects power?

- [Chalkboard]
- Increase Effect Size
- Decrease Standard Deviation

What does this all mean?

- When you do a randomized impact evaluation you need to figure out the sample size (n)
- How many people do you want in your study?
- Assume the following:
  Statistically Significant level = 5%
  Power = 90%
  Effect Size =
  Standard Deviation =

Example 1: GPA

- Treatment Group pay $10 for each "A"

$$\mathbb{E}[Y_i|T=1]-\mathbb{E}[Y_i|T_i=0]=\beta$$

$$Y_i=\alpha+\beta T_i+\varepsilon_i$$

- Key Assumptions
  Treatment Effect: $\beta=.2$
  Standard Deviation: $\sigma=.3$
- This implies

$$\mathbb{E}[Y_i|T=1]-\mathbb{E}[Y_i|T_i=0]=3.5-3.3=.2=\beta$$

STATA: Calculating Sample Size

- [Demonstrate Using STATA]
- Statistics>Power and Sample Size>Test of Means
- Interest in n1+n2 = 96
- What happens when we assume a small effect ($\beta = .05$)

$$\mathbb{E}[Y_i | T = 1] - \mathbb{E}[Y_i | T_i = 0] = 3.35 - 3.3 = .05 = \beta$$

  Look at what happens to n1 + n2 = 1514!
- What if we increase the power?

STATA: What is my power?

- Work with NGO/Government
- Tell you the sample size is 500
  Effect Size

$$\mathbb{E}[Y_i|T=1]-\mathbb{E}[Y_i|T_i=0]=3.5-3.3=.2=\beta$$

Standard Deviation = .30

What affects power?

1. Sample Size
   Larger sample => more power
   More power => Higher probability that I will reject null if false
   In other words, "Good chance that my impact evaluation will find something . . ."
2. Treatment Effect
   Large treatment effect => more power
3. Standard Deviation
   Smaller standard deviation => more power
   Variation in sample

- Before intervention/study, you need to make best guess on treatment effect and standard deviation.
- If you want 90% power => calculation gives you sample size
- If you already know sample size => calculation gives you the power (or likelihood you will find an effect)

Practically Speaking

- What wouldn't I want a huge sample size?
  - Cost of increasing sample size
  - Funding could be used elsewhere
- Assume treatment effect and standard deviation
  - As we've shown, assumptions play huge role in power calculations - need to be realistic

Other Considerations

- You mean there's more?
- Clustering
  Individual observations might be correlated
  (i.e. unobserved factor affects several people in study who live close to each other)
- Proportion in treatment and control arms
- Control Variables
  Observed variables that affect outcome can be used to reduce standard errors => resulting in increased power

How is this useful?

- Avoid starting an evaluation that is doomed from the start – no power to detect impacts (waste of time & money)
- Spend enough, but only that much, on the studies you really need
- Impact Evaluation Proposals

Great Resources

- Basics
  - Introductory Econometrics (Wooldridge)
  - Randomization Toolkit (Duflo et. al. 2009)
  - STATA
- Advanced
  - Mostly Harmless Econometrics (Angrist &)
  - Randomization Toolkit (Duflo et. al. 2009)
  - Minimum Detectable Effects (Bloom 1995)
  - Optimal Design: http://sitemaker.umich.edu/group-based/optimal_design_software