14.385 Nonlinear Econometric Analysis

Randomized experiments

Maximilian Kasy

Department of Economics, MIT

Fall 2022

Outline

- Identification
- Testing:
  - Asymptotic inference.
  - Randomization inference.
- Power calculations.

Takeaways for this part of class

- With exogenous assignment, marginal distributions of potential outcomes are identified.
- Standard errors can be calculated as in intro econometrics.
- Alternatively, we can use randomization inference:
  - Condition on the sample, potential outcomes.
  - Consider only randomness coming from treatment assignment.
  - Under the null of no treatment effects on any unit, this randomization distribution is known.
- Basic considerations for experimental design:
  - With equal variances, a 50/50 split of the sample minimizes the estimator variance.
  - The power of tests for zero average treatment effect is a function of sample size and the true treatment effect.
  - We can use this to choose the sample size.

Identification in Randomized Experiments

- Randomization implies:

$$(Y_1, Y_0) \text{ independent of } D, \quad \text{or} \quad (Y_1, Y_0) \perp D.$$

- We have that $E[Y_0|D=1] = E[Y_0|D=0]$ and therefore

$$\alpha_{ATET} = E[Y_1 - Y_0|D=1] = E[Y|D=1] - E[Y|D=0]$$

- Also, we have that

$$\alpha_{ATE} = E[Y_1 - Y_0] = E[Y_1 - Y_0|D=1] = E[Y|D=1] - E[Y|D=0]$$

- As a result,

$$\underbrace{E[Y|D=1] - E[Y|D=0]}_{\text{Difference in Means}} = \alpha_{ATE} = \alpha_{ATET}$$

Identification in Randomized Experiments

- The identification result extends beyond average treatment effects.
- Given random assignment $(Y_1, Y_0) \perp D$:

$$\begin{aligned} F_{Y_0}(y) &= \Pr(Y_0 \leq y) = \Pr(Y_0 \leq y | D = 0) \\ &= \Pr(Y \leq y | D = 0) \end{aligned}$$

- Similarly,

$$F_{Y_1}(y) = \Pr(Y \leq y | D = 1).$$

- So effect of the treatment at any quantile, $Q_\theta(Y_1) - Q_\theta(Y_0)$ is identified.
  - Randomization identifies the entire marginal distributions of Y_0 and Y_1
  - Does not identify the quantiles of the effect: $Q_\theta(Y_1 - Y_0)$ (the difference of quantiles is not the quantile of the difference)

Estimation in Randomized Experiments

- Consider a randomized trial with N individuals. Suppose that the estimand of interest is ATE:

$$\alpha_{ATE} = E[Y_1 - Y_0] = E[Y|D=1] - E[Y|D=0].$$

- Using the analogy principle, we construct an estimator:

$$\widehat{\alpha} = \bar{Y}_1 - \bar{Y}_0,$$

where

$$\bar{Y}_1 = \frac{\sum Y_i \cdot D_i}{\sum D_i} = \frac{1}{N_1} \sum_{D_i=1} Y_i;$$

$$\bar{Y}_0 = \frac{\sum Y_i \cdot (1-D_i)}{\sum (1-D_i)} = \frac{1}{N_0} \sum_{D_i=0} Y_i$$

with $N_1 = \sum_i D_i$ and $N_0 = N - N_1$.

- $\widehat{\alpha}$ is an unbiased and consistent estimator of α_{ATE}.

Identification

Testing

Experimental design

References

Testing in Large Samples: Two Sample t-Test

- Notice that:

$$\frac{\widehat{\alpha} - \alpha_{ATE}}{\sqrt{\frac{\widehat{\sigma}_1^2}{N_1} + \frac{\widehat{\sigma}_0^2}{N_0}}} \xrightarrow{d} N(0,1),$$

  where

$$\widehat{\sigma}_1^2 = \frac{1}{N_1 - 1} \sum_{D_i = 1} (Y_i - \bar{Y}_1)^2,$$

  and $\widehat{\sigma}_0^2$ is analogously defined.

- In particular, let

$$t = \frac{\widehat{\alpha}}{\sqrt{\frac{\widehat{\sigma}_1^2}{N_1} + \frac{\widehat{\sigma}_0^2}{N_0}}}.$$

- We reject the null hypothesis H_0: $\alpha_{ATE} = 0$ against the alternative H_1: $\alpha_{ATE} \neq 0$ at the 5% significance level if $|t| > 1.96$.

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : E[Y_1] = E[Y_0], \quad H_1 : E[Y_1] \neq E[Y_0]$$

- Fisher's Exact Test with small N:

$$H_0 : Y_1 = Y_0, \quad H_1 : Y_1 \neq Y_0 \qquad \text{(sharp null)}$$

- Let Ω be the set of all possible randomization realizations.
- We only observe the outcomes, Y_i, for one realization of the experiment. We calculate $\hat{\alpha} = \bar{Y}_1 - \bar{Y}_0$.
- Under the sharp null hypothesis we can calculate the value that the difference of means would have taken under any other realization, $\hat{\alpha}(\omega)$, for $\omega \in \Omega$.

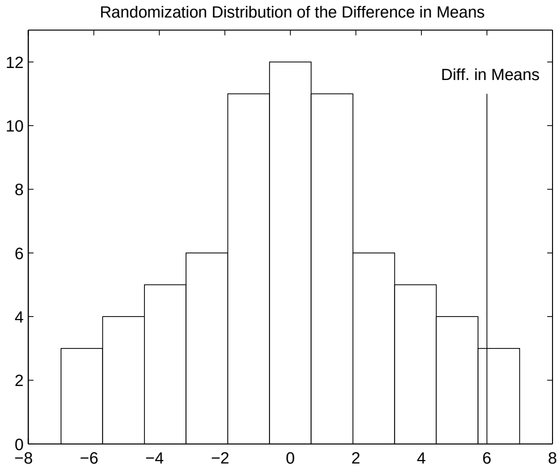
Testing in Small Samples: Fisher's Exact Test

Suppose that we assign 4 individuals out of 8 to the treatment:

Y_i	12	4	6	10	6	0	1	1	
D_i	1	1	1	1	0	0	0	0	$\hat{\alpha} = 6$
									$\hat{\alpha}(\omega)$
$\omega = 1$	1	1	1	1	0	0	0	0	6
$\omega = 2$	1	1	1	0	1	0	0	0	4
$\omega = 3$	1	1	1	0	0	1	0	0	1
$\omega = 4$	1	1	1	0	0	0	1	0	1.5
				$\ldots$					
$\omega = 70$	0	0	0	0	1	1	1	1	-6

- The randomization distribution of $\widehat{\alpha}$ (under the sharp null hypothesis) is $\Pr(\widehat{\alpha} \leq z) = \frac{1}{70} \sum_{\omega \in \Omega} 1\{\widehat{\alpha}(\omega) \leq z\}$
- Now, find $\bar{z} = \inf\{z : P(|\widehat{\alpha}| > z) \leq 0.05\}$
- Reject the null hypothesis, H_0: $Y_{1i} - Y_{0i} = 0$ for all i, against the alternative hypothesis, H_1: $Y_{1i} - Y_{0i} \neq 0$ for some i, at the 5% significance level if $|\widehat{\alpha}| > \bar{z}$

Testing in Small Samples: Fisher's Exact Test



$$\Pr(|\hat{\alpha}(\omega)| \geq 6) = 0.0857$$

Identification

Testing

Experimental design

References

Experimental Design: Relative Sample Sizes for Fixed N

- Suppose that you have N experimental subjects and you have to decide how many will be in the treatment group and how many in the control group.

- We know that:
$$\bar{Y}_1 - \bar{Y}_0 \sim \left(\mu_1 - \mu_0, \frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0} \right).$$

- We want to choose N_1 and N_0, subject to $N_1 + N_0 = N$, to minimize the variance of the estimator of the average treatment effect.

- The variance of $\bar{Y}_1 - \bar{Y}_0$ is:
$$\text{var}(\bar{Y}_1 - \bar{Y}_0) = \frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}$$
where $p = N_1/N$ is the proportion of treated in the sample.

Experimental Design: Relative Sample Sizes for Fixed N

Practice problem

Derive the value of p which minimizes the variance of the estimator of the average treatment effect.

Experimental Design: Relative Sample Sizes for Fixed N

- Find the value p^* that minimizes $\text{var}(\bar{Y}_1 - \bar{Y}_0)$:

$$-\frac{\sigma_1^2}{p^{*2}N} + \frac{\sigma_0^2}{(1-p^*)^2 N} = 0.$$

- Therefore:

$$\frac{1-p^*}{p^*} = \frac{\sigma_0}{\sigma_1},$$

  and

$$p^* = \frac{\sigma_1}{\sigma_1 + \sigma_0} = \frac{1}{1 + \sigma_0/\sigma_1}.$$

- A "rule of thumb" for the case $\sigma_1 \approx \sigma_0$ is $p* = 0.5$
- For practical reasons it is sometimes better to choose unequal sample sizes (even if $\sigma_1 \approx \sigma_0$)

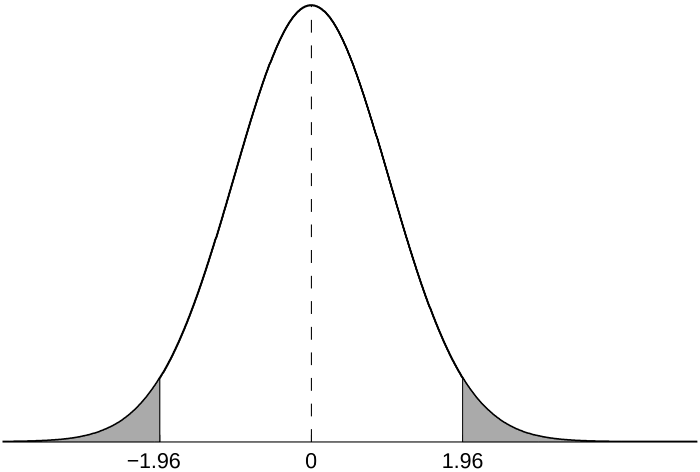
Experimental Design: Power Calculations to Choose N

- Recall that for a statistical test:
  - Type I error: Rejecting the null if the null is true.
  - Type II error: Not rejecting the null if the null is false.
- Size of a test is the probability of type I error, usually 0.05.
- Power of a test is one minus the probability of type II error, i.e. the probability of rejecting the null if the null is false.
- Statistical power increases with the sample size.
- But when is a sample "large enough"?
- We want to find N such that we will be able to detect an average treatment effect of size α or larger with high probability.

Experimental Design: Power Calculations to Choose N

- Assume a particular value, α, for $\mu_1 - \mu_0$.
- Let $\widehat{\alpha} = \bar{Y}_1 - \bar{Y}_0$ and
$$\text{s.e.}(\widehat{\alpha}) = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}.$$
- For a large enough sample, we can approximate:
$$\frac{\widehat{\alpha} - \alpha}{\text{s.e.}(\widehat{\alpha})} \sim N(0,1).$$
- Therefore, the t-statistic for a test of significance is:
$$t = \frac{\widehat{\alpha}}{\text{s.e.}(\widehat{\alpha})} \sim N\left(\frac{\alpha}{\text{s.e.}(\widehat{\alpha})}, 1\right).$$

Probability of Rejection if $\mu_1 - \mu_0 = 0$

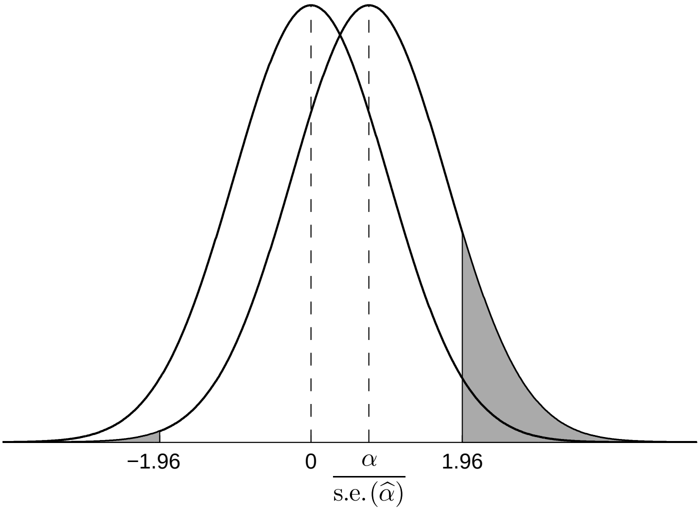


Power

Practice problem

Derive the probability of rejection as a function of α, p, and N.

Probability of Rejection if $\mu_1 - \mu_0 = \alpha$



Experimental Design: Power Calculations to Choose N

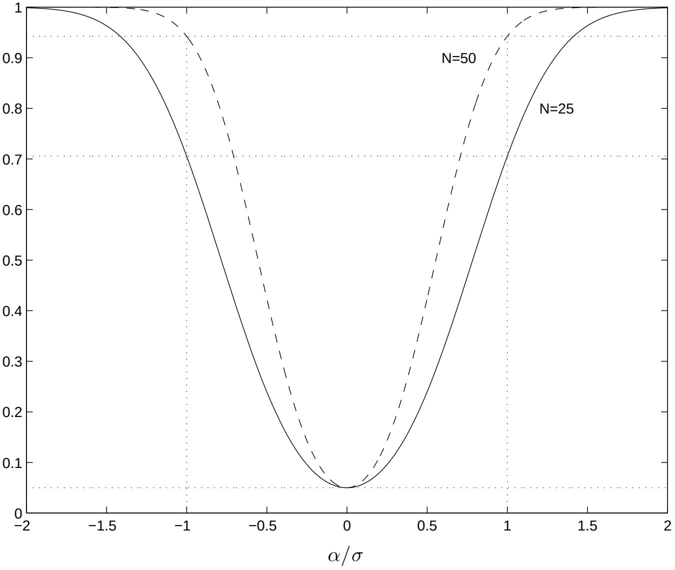
- The probability of rejecting the null $\mu_1 - \mu_0 = 0$ is:

$$\begin{aligned}
\Pr(|t| > 1.96) &= \Pr(t < -1.96) + \Pr(t > 1.96) \\
&= \Pr\left(t - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})} < -1.96 - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})}\right) \\
&+ \Pr\left(t - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})} > 1.96 - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})}\right) \\
&= \Phi\left(-1.96 - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})}\right) + \left(1 - \Phi\left(1.96 - \frac{\alpha}{\text{s.e.}(\widehat{\alpha})}\right)\right)
\end{aligned}$$

- Suppose that $p = 1/2$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$. Then,

$$\text{s.e.}(\widehat{\alpha}) = \sqrt{\frac{\sigma^2}{N/2} + \frac{\sigma^2}{N/2}} = \frac{2\sigma}{\sqrt{N}}.$$

Power Functions with $p = 1/2$ and $\sigma_1^2 = \sigma_0^2$



General formula for the power function ($p \neq 1/2$, $\sigma_0^2 \neq \sigma_1^2$)

$$
\begin{aligned}
\Pr(\text{reject } \mu_1 - \mu_0 = 0 | \mu_1 - \mu_0 = \alpha) \\
&= \Phi\left(-1.96 - \alpha \bigg/ \sqrt{\frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}}\right) \\
&\quad + \left(1 - \Phi\left(1.96 - \alpha \bigg/ \sqrt{\frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}}\right)\right).
\end{aligned}
$$

To choose N we need to specify:

1. α: minimum detectable magnitude of treatment effect
2. Power value (usually 0.80 or higher)
3. σ_1^2 and σ_0^2 (usually $\sigma_1^2 = \sigma_0^2$) (e.g., using previous measures)
4. p: proportion of observations in the treatment group If $\sigma_1 = \sigma_0$, then the power is maximized by $p = 0.5$

References

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In Handbook of Economic Field Experiments, *volume 1, pages 73–140. Elsevier*

Kasy, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. Political Analysis, *24(3):324–338*

These slides are based on the slides by **Alberto Abadie** for previous iterations of 14.385.