Applied Microeconometrics, Lecture 1

Ciprian Domnisoru
Aalto University

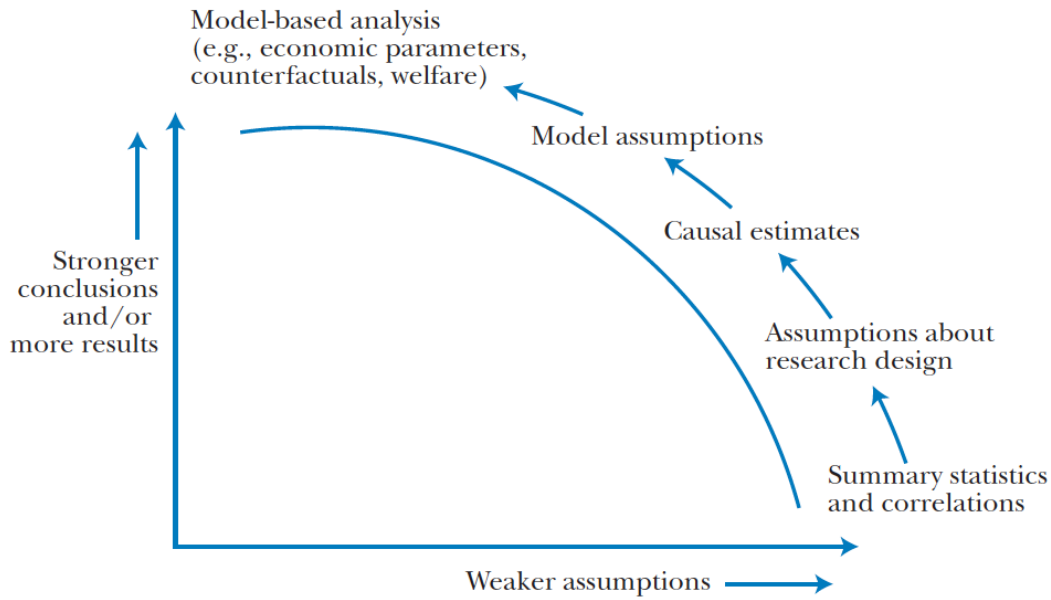
Introductions

- Ciprian Domnisoru, Assistant Professor, www.cipriandomnisoru.net
- Research in Labor and Education Economics with a focus on social mobility
- Reduced-form work: compulsory schooling, school quality, teacher gender, effects of education on mortality.
- Structural work: competition between private and public providers in the higher education system, overeducation(underemployment),job search and mismatch, entrepreneurship.
- Ellen Sahlström, PhD Student, ellen.sahlstrom@aalto.fi

This course

Figure 1

The Frontier Between Strength of the Assumptions and More Economically Relevant Results



Note: Figure depicts the trade-off between the strength of the assumptions and more economically relevant results.

Source: Neale Mahoney, 2022. Principles for Combining Descriptive and Model-Based Analysis in Applied Microeconomics Research, *Journal of Economic Perspectives*—Volume 36, Number 3.

Applied Econometrics II

- This course:
- Learn new methods and robustness checks
- Have a broader picture of methods: descriptive, experimental, quasi-experimental and structural methods

 - **hard skills:**

- some mathematical/ statistics calculations in homework
- Stata work in homework;

4 X Homework=40 points

 - **soft skills:** review the strenghts and weaknesses of research articles.

- 2 paper summaries, 30 points, can turn in anytime until the end of the course. Goal: develop some content knowledge in your field: read papers related to your research interests.
- slides + presentation, 20 points
- participation in the course, 10 points, polls/quizzes answered with a reasonable (not necessarily correct) answer.

Applied Econometrics II

- Email me or Ellen for office hours.
- Plan to meet with me before presentations last week. Submit slides in advance of that meeting. I will send out a doodle.
- You should work in groups. If one team member does not work on the assignment that week, submit names of those who did work. Submit assignments through MyCourses.
- PhD students : one paper summary (15%) + paper proposal (35%). We will likely have a separate session for presentations, depending on final student count in the course (everyone welcome to attend).

Presentations

The presentation should take 15-20 minutes, and include a maximum of 10 slides (in PDF format). The main presentation should cover: main research question, data used in the paper (briefly), main empirical strategy, results and **robustness checks**- what threats to validity they address. Include relevant figures/tables.

Paper reviews

- Section 1. Describe the research question(s) -1 paragraph
- Section 2. Explain the problem of causal identification: why can't we just run OLS? Is there selection bias, are there unobservables, what are the identification problems?- 1 paragraph
- Section 3. Explain what estimator(s) is/are used. If several estimators are used, explain why. Explain how the estimators solve the identification problem. 2-3 par.
- Section 4. What are the authors' findings? How do you expect these findings differ from what we would find with just a naive OLS estimate? 1 par.
- Section 5. What robustness checks do the authors implement? What validity concerns do the robustness checks address? - 3 par.
- Section 6. Critique the paper. Think about potential confounding factors that may hinder identification. Challenge the assumptions made in the paper. What are the remaining threats to validity? - 2-4 par.

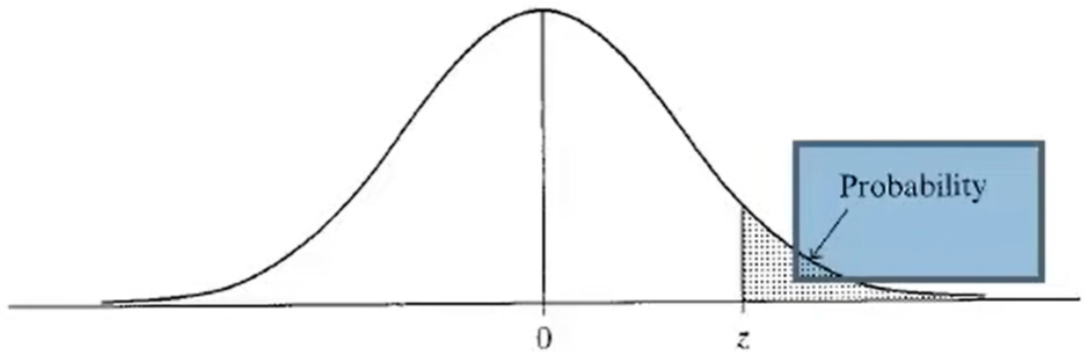
Paper proposal-PhD students- 5-6 pages

- Section 1. Statement of the policy/research question.
- Section 2. Background on the policy/research question. Please make the background relevant to the specific policy question.
- Section 3. Other approaches to solving this problem (briefly) review other policy/research papers dealing with the same (or similar) research question).
- Section 4. Data and sample- describe what data you would use. Provide details about what variables you would use, how they are measured, provide sample summary statistics if they are available. Discuss how you would go about getting the data.
- Section 5. Method- Explain the problem of causal identification you are trying to solve.
- Section 6. Explain what estimator(s) you want to use and how it/they would solve the identification problem.
- Section 7. Discuss the limitations of your proposed study and validity concerns.

Power in experiments

Brief reminder on z scores

TABLE A: Normal curve tail probabilities. Standard normal probability in right-hand tail (for negative values of z, probabilities are found by symmetry).



	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148

Review of basic conceps. Example: Testing vaccines

Treatment: Receives one intramuscular (IM) injection of 100 micrograms (ug) mRNA-1273 .**Control:** Receives one IM injection of saline solution (placebo). Let p_1 be the fraction of the control group population who develop symptoms and are tested positive. Let p_2 be the fraction of the treatment population who develop symptoms and are tested positive. Let's consider a simplified analysis:

- Null Hypothesis H_0 is the claim of "no difference." , $p_1 = p_2$
- Alternative Hypothesis H_a is the claim we are evaluating/ trying to prove. $p_1 > p_2$ or $p_1 \neq p_2$

	Parameter	Statistic	Null Hypothesis in Hypothesis Test	Standard Error for Hypothesis Test	Test Statistic for Hypothesis Test
Proportion of Two Populations	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$p_1 = p_2$	$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

- In a regression context, if you had a "Treatment" indicator variable, what would the null hypothesis be ?

Lecture 1- Power

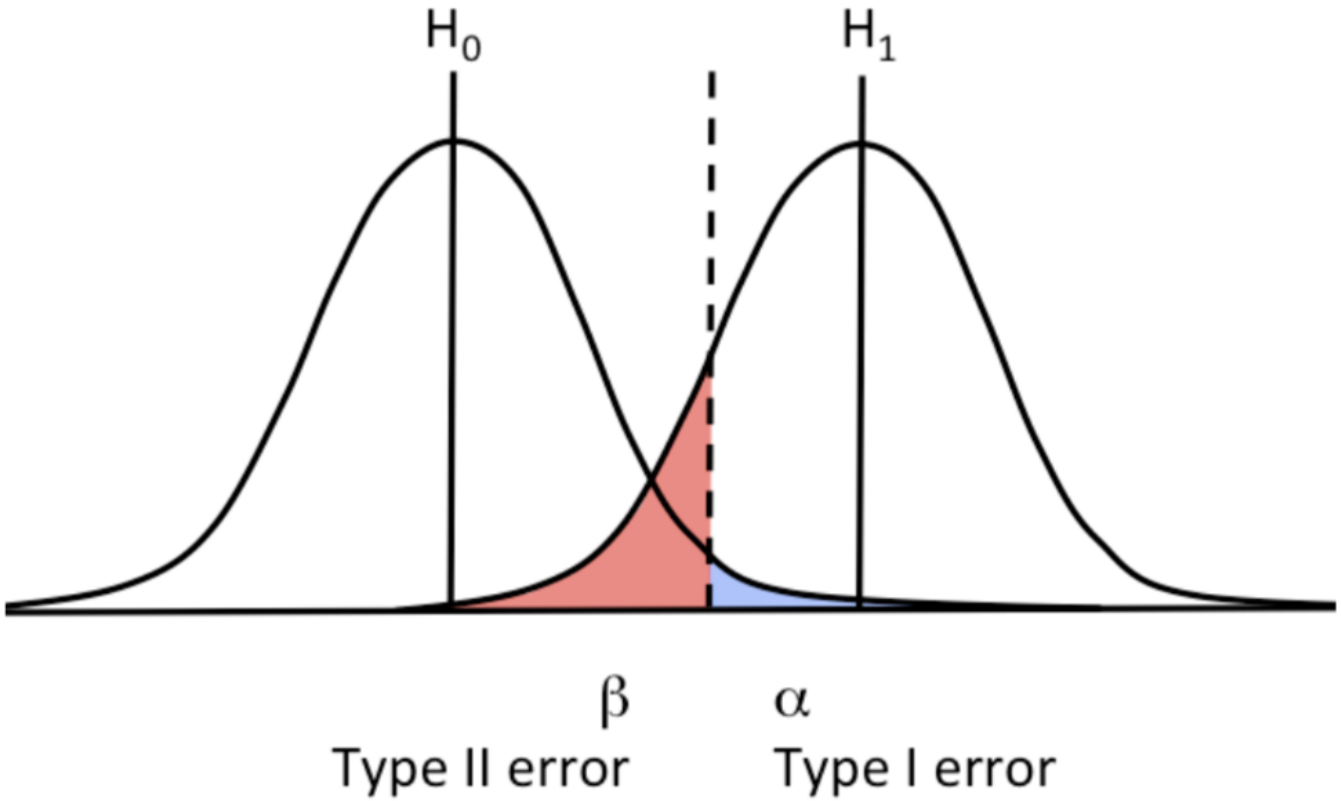
- Type I error- rejecting the null when the null is correct
- Standard notation: α is the probability of a Type I error (e.g., we usually reject the null hypothesis when $\mid t \mid > 2$, α=0.05

	Your decision	
	Reject H_0	Do not reject H_0
H_0 is true	Type I error	Correct decision
H_0 is false	Correct decision	Type II error

- Type II error= failing to reject the null when alternative is correct
- Often β is used to denote the probability of a Type II error.
- 1 - β is said to be the power of a test.
- A test is said to be underpowered if it has low power.
- Increasing sample size helps increase power, but usually more expensive.

Lecture 1- Power

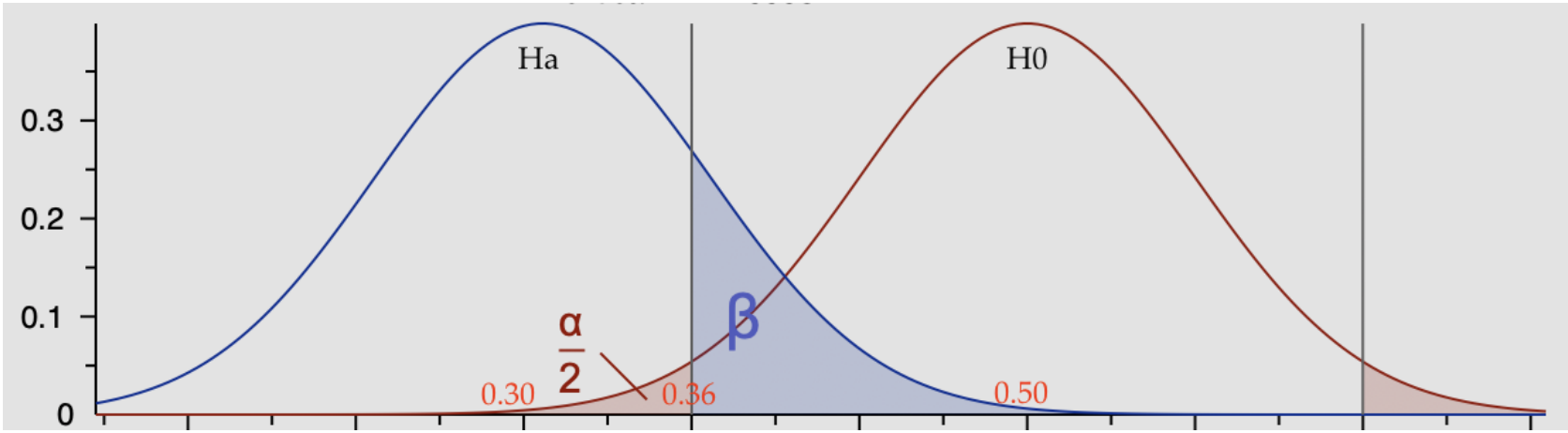
- Consider a one-sided test where the alternative sample mean would be larger than the null hypothesis mean.



- How are α and β related?

Lecture 1- Power

- Suppose now the alternative hypothesis mean is smaller than the null hypothesis mean, and we conduct a double-sided test.



Lecture 1- Power- Example

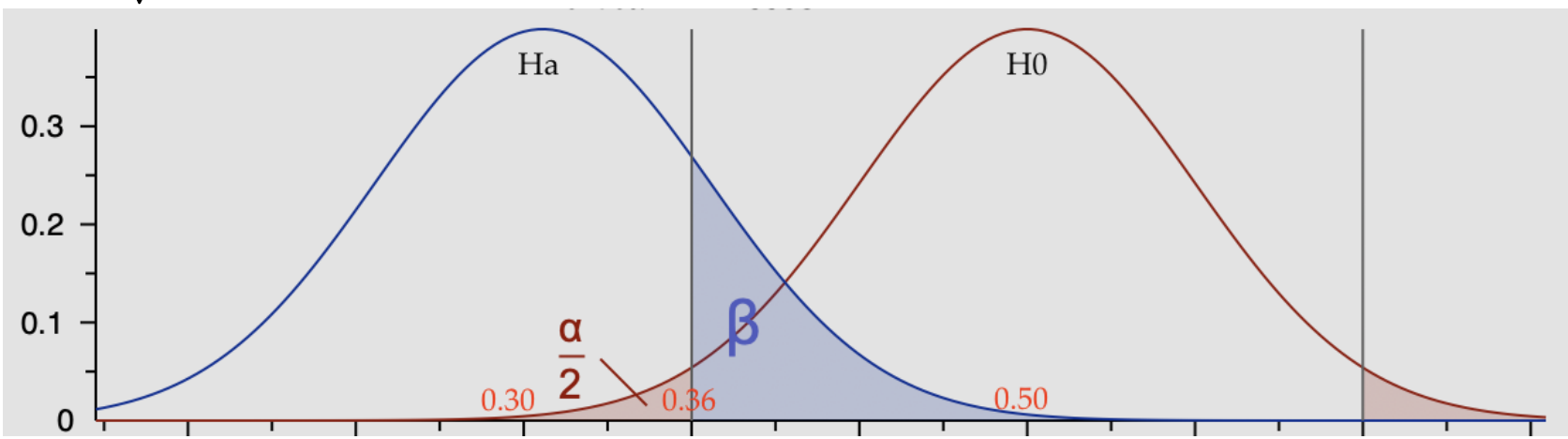
- Example: You know that in Sweden 0.50 of job applicants with Swedish names get callbacks on job applications, and want to know if this rate is the same for those with Middle-Eastern (ME) names.
- You plan to conduct a study with 100 applications in each group.
- Under the null hypothesis, your z-statistic will be:
$$z = \frac{diff}{s.e.} = \frac{diff}{\sqrt{(p(1-p)}\sqrt{1/n_1+1/n_2}} = \frac{diff}{\sqrt{(0.5(1-0.5)}\sqrt{1/100+1/100}}$$
- Reject the null if $\mid z \mid > 2$, i.e. if $\mid diff \mid > 0.14$

	Parameter	Statistic	Null Hypothesis in Hypothesis Test	Standard Error for Hypothesis Test	Test Statistic for Hypothesis Test
Proportion of Two Populations	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$p_1 = p_2$	$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Lecture 1- Power- Example

- Suppose the callback rate for Middle Eastern applicants is 0.30. You can calculate power under this scenario.

$$z = \frac{0.36 - 0.30}{\sqrt{(0.4(1-0.4)}\sqrt{1/100 + 1/100}} = \frac{0.06}{0.069} = 0.87$$



Lecture 1- Power- Example

- Suppose the callback rate for Middle Eastern applicants is 0.30. You can calculate power under this scenario.
  $z = \frac{0.36-0.30}{\sqrt{(0.4(1-0.4)}\sqrt{1/100+1/100}} = \frac{0.06}{0.069} = 0.87$
- The probability that $z > 0.87$ is 0.19, so β=0.19 and power=1-0.19=0.81.
- Suppose you had funds for a larger sample...
- Step 1. Calculate diff
  $z = \frac{diff}{s.e.} = \frac{diff}{\sqrt{(0.5(1-0.5)}\sqrt{1/200+1/200}} = \frac{diff}{0.05} \implies \mid diff \mid > 0.1$
- Step 2. t-stat under alternative
  $z = \frac{0.40-0.30}{\sqrt{(0.4(1-0.4)}\sqrt{1/200+1/200}} = \frac{0.10}{0.049} = 2.04$
- Step β calculation
  The probability that $t > 2.04$ is 0.022, so β=0.022
- Power calculation: 1-0.022=0.98.
- Example continued in Review session.

Lecture 1- Power

- Use power prospectively for planning future studies.
  1. determine an appropriate sample size- for a desired level of power and minimum detectable effect size.
  How would you choose the effect size? What have previous studies found? What would be the smallest effect size that would be interesting to be able to reject?
- Evaluate a planned study to see if it is likely to yield useful information.
  2. calculate power, given sample size and minimum effect size
  3. minimum effect size, given power and sample size.
- Sample size, power, and the minimum detectable effect size should be calculated at the impact evaluation stage.
- Typically, researchers set Type I significance level α to 0.05 and Type II significance level, or power, to 0.80 .

Lecture 1- Power

- Use power prospectively for planning future studies.
  1. determine an appropriate sample size- for a desired level of power and minimum detectable effect size.

```
. power twoproportions 0.5 0.3

Performing iteration ...

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =   -0.2000  (difference)
           p1 =    0.5000
           p2 =    0.3000

Estimated sample sizes:

            N =       186
  N per group =        93
```

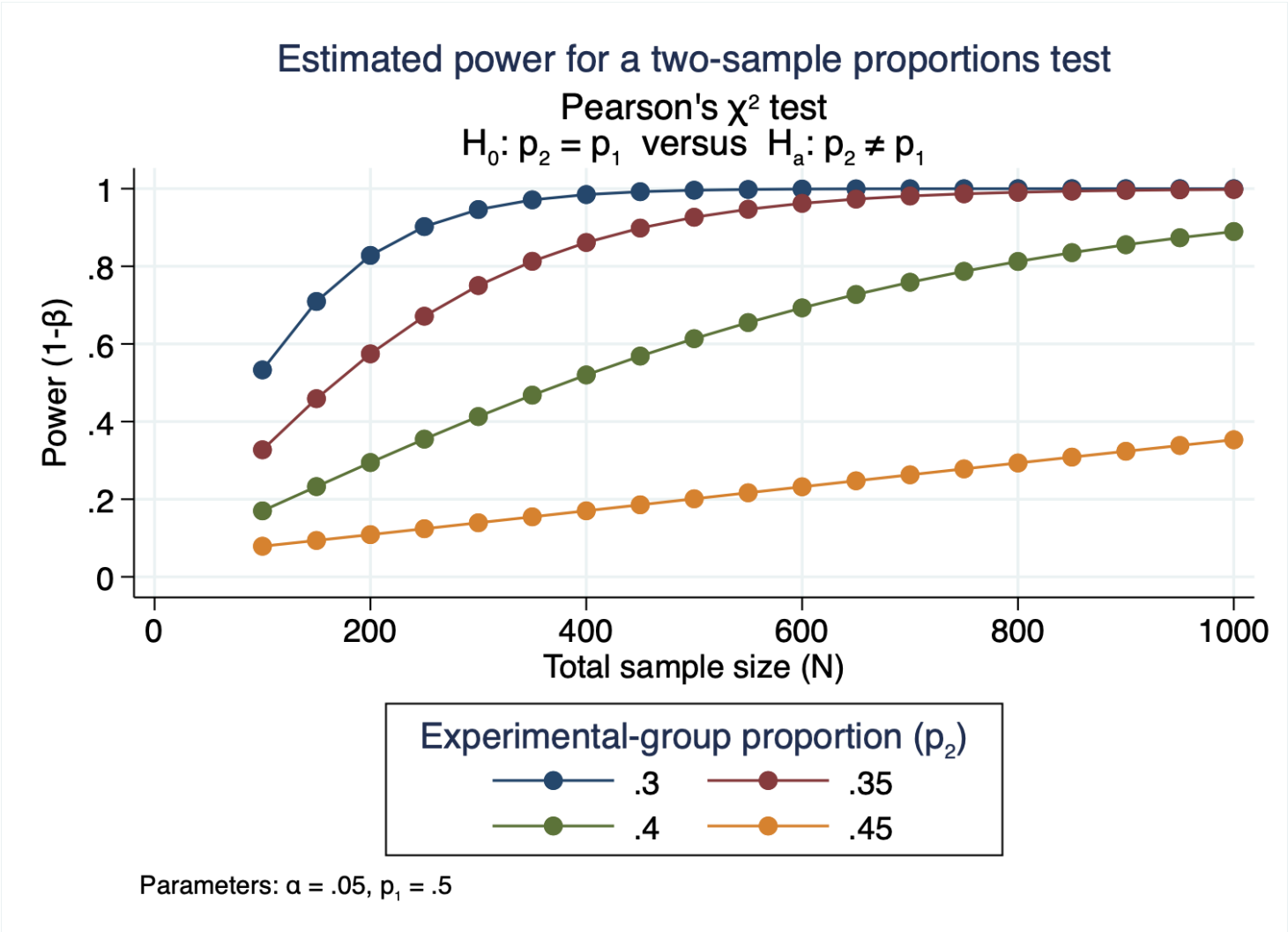
Lecture 1- Power

- 2. calculate power, given sample size and effect size

```
. power twoproportions 0.5 0.3, n1(100) n2(100)

Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
            N =       200
           N1 =       100
           N2 =       100
        delta =   -0.2000  (difference)
           p1 =    0.5000
           p2 =    0.3000

Estimated power:

        power =    0.8281
```

Lecture 1- Power

- power twoproportions 0.5 (0.3 (0.05) 0.45), n(100 (50) 1000) graph



Use power prospectively for planning future studies.

- 3. minimum detectable effect size, given power and sample size.

```
. power twoproportions 0.5, n1(100) n2(100) power(0.8)

Performing iteration ...

Estimated experimental-group proportion for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1; p2 > p1

Study parameters:

        alpha =    0.0500
        power =    0.8000
            N =       200
           N1 =       100
           N2 =       100
           p1 =    0.5000

Estimated effect size and experimental-group proportion:

        delta =    0.1932  (difference)
           p2 =    0.6932
```

Use power prospectively for planning future studies.

- 3. minimum detectable effect size, given power and sample size.

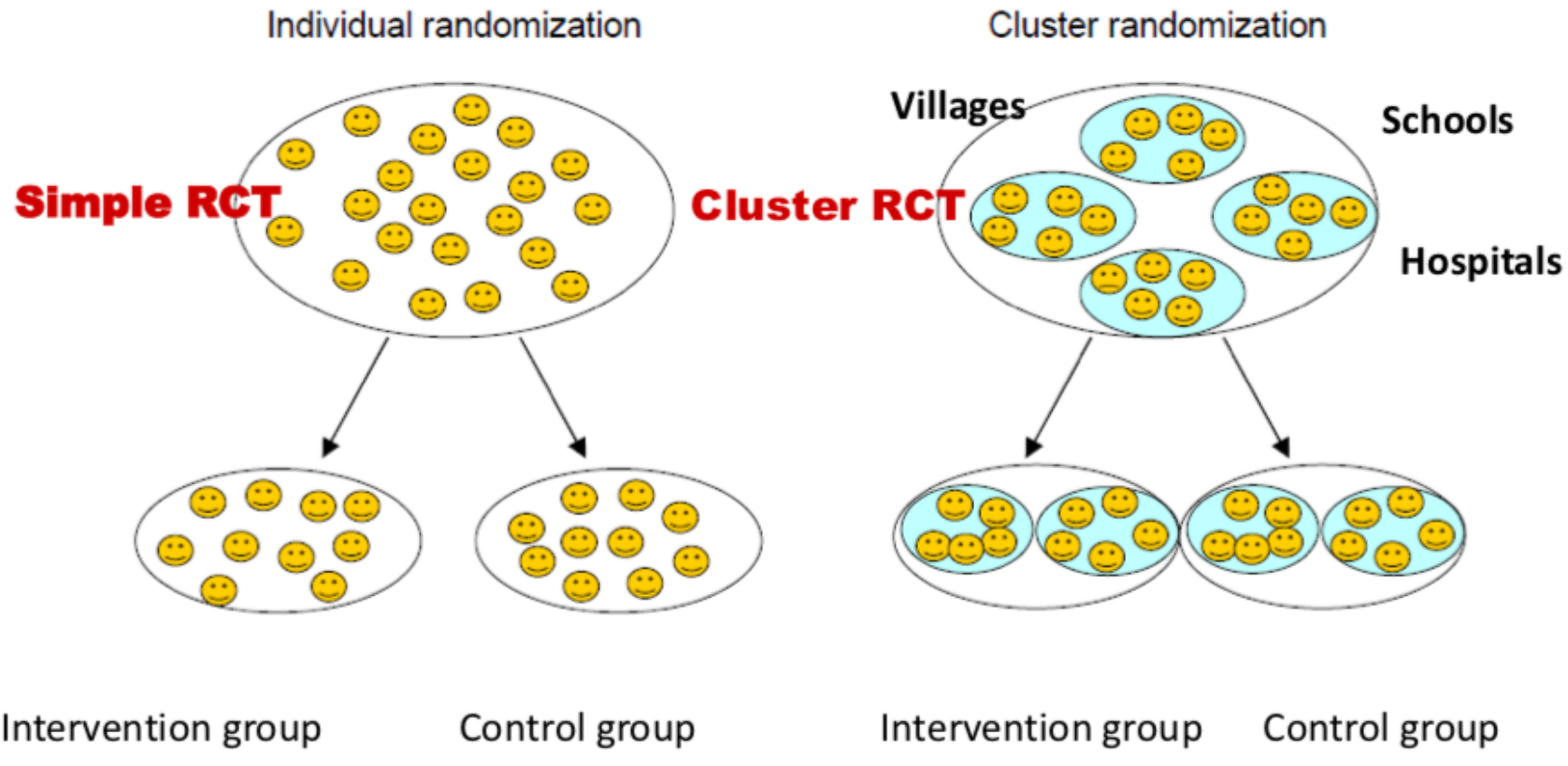
```
. power twoproportions 0.5, n1(100) n2(100) power(0.8281)

Performing iteration ...

Estimated experimental-group proportion for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1; p2 > p1

Study parameters:

        alpha =    0.0500
        power =    0.8281
            N =       200
           N1 =       100
           N2 =       100
           p1 =    0.5000

Estimated effect size and experimental-group proportion:

        delta =    0.2000  (difference)
           p2 =    0.7000
```

Clustering

Why account for clustering?

- Sometimes interventions in randomized experiments are not allocated to individuals, but rather to groups. Similarly, in some observational studies, observations are found in naturally occurring groups, such as neighborhoods.
- observations within a cluster tend to be more similar than observations selected entirely at random.
- This violates the assumption of independence that is at the heart of common methods of statistical estimation and hypothesis testing.
- Typical consequences of failing to account for clustering: p-values are too small, confidence intervals too narrow, and sample size estimates too small.

Related outcomes of experimental subjects



Related outcomes of experimental subjects

- Suppose student i in classroom j has a test score
$$S_{ij} = \mu + \gamma_j + \epsilon_{ij}$$
, where $E[\gamma_j = 0]$ and E[ϵ_{ij}]=0
- The variance of γ_j across classrooms , is σ^2_γ, and the individual-level term ϵ_{ij} has variance σ^2_ϵ.
- The intraclass correlation (ICC) is defined as:
$$\rho = \frac{\sigma^2_\gamma}{\sigma^2_S} = \frac{\sigma^2_\gamma}{\sigma^2_\gamma + \sigma^2_\epsilon}$$
- Suppose you have n_j students in each of j=1...m classrooms
- Then the variance of the mean will be :
$$\sigma^2_{\bar{S}} = \frac{\sigma^2_S}{mn_j}[1 + ((n_j - 1)\rho)]$$
- The unclustered variance of the mean would be just
$$\sigma^2_{\bar{S}} = \frac{\sigma^2_S}{mn_j}$$
- This variance is the one used in computing hypothesis tests or confidence intervals when describing this sample or comparing it with some other samples.

Related outcomes of experimental subjects

- Suppose you implement a new curriculum, which you believe will have a positive treatment effect on μ. There are 40 students in each classroom, there are 10 classrooms, and, from the literature, you believe $\rho = 0.1$.
- If you treat scores as independent, the standard error of the mean is $\frac{\sigma_S}{\sqrt{mn_j}} = 0.05\sigma_S$
- If you account for intraclass correlation,

  The standard error of the mean is $\sigma_S \sqrt{\frac{1+(n_j-1)\rho}{mn_j}} = 0.11\sigma_S$
- Not accounting for intraclass correlation will mean your standard errors are incorrectly too small. If you are conducting power/ sample size calculations, it will mean your sample size is too small.

Related outcomes of experimental subjects

- Suppose job applicants have a common case worker, and the intraclass correlation is 0.1. k clusters, m applicants.

```
. power twoproportions 0.5 0.3, k1(10) k2(10) m1(10) m2(10) rho(0.1)

Estimated power for a two-sample proportions test
Cluster randomized design, Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        delta =   -0.2000  (difference)
           p1 =    0.5000
           p2 =    0.3000

Cluster design:

           K1 =        10
           K2 =        10
           M1 =        10
           M2 =        10
           N1 =       100
           N2 =       100
          rho =    0.1000

Estimated power:

        power =    0.5546
```

You will need a larger sample size whenever:

- The effects you want to detect are small
- The statistical precision (α) is smaller
- There is more clustering in samples (High ICC also implies a need for more clusters)
- You have more strata (subgroups for which you want to evaluate the treatment. Example: Moderna vaccine: $\geq$ 65 years; $<$ 65 years and at increased risk for COVID-19 complications ("at risk"); $<$ 65 years "not at risk" for COVID-19 complications.

Duflo, Hanna, Ryan."Incentives Work: Getting Teachers to Come to School"

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

- Regular teachers in India have few incentives, often don't show up to work.
- Can an incentive program for para-teachers increase teacher presence? Student outcomes?
- Increase in student outcomes not obvious: Multitasking (teachers reduce efforts in non-tested dimensions); Loss in intrinsic motivation; Incompetence; Target earnings; Level of incentives?
- Paper combines:
  1. A randomized experiment in teacher incentives
  2. A regression discontinuity design that tests how teachers respond to financial incentives : change in teacher behavior just before and after the end of he month
  3. Structural model estimated using the treatment group: simple dynamic labor supply model, teachers choose each day whether to go and teach or not.

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

- Need for structural estimation: "While the reduced form results inform us that this program was effective in reducing absenteeism, they do not tell us what the effect of another scheme with a different payment structure would be."

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

- Seva Mandir, an NGO in rural Rajasthan, who runs 150 "non-formal education center" (NFE): single teacher school for students who do not attend regular school.
- Students are 7-14 year old, illiterate when they join.
- Teacher absence rate 35%
- Schools teach basic hindi and math skills and prepare students to "graduate" to primary school.
- In 1997, 20 million children were served by such NFEs

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

- Teachers in intervention schools received a camera with non-temperable time and date stamp.
- Instructed to take two pictures of themselves and the children every day (pictures separated by at least 5 hours, at least 8 children per picture).
- Payment is calculated each month and is a non-linear function of attendance: • Up to 10 days: Rs 500. • Each day above 10 days: Rs 50. • In non-intervention schools, teachers receive Rs 1000, and are reminded by attending at least 20 days is compulsory.

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

- "We originally picked 120 schools, out of which 7 closed immediately after they were picked to be in the study (unrelated to the study)".
- 57 treatment schools, the rest control. Data collection: • Teacher and student attendance: Monthly random checks. • In treatment schools: Camera data • Students learning: tests in September 03-April 04-Oct 04 • Long term impact: a new sets of random checks was done in 2006-2007, and a new set of test scores were done in 2007
- Findings: "Over the 30 months in which attendance was tracked, teachers at program schools had an absence rate of 21 percent, compared to 44 percent at baseline and the 42 percent in the comparison schools."

Duflo, Hanna, Ryan: Incentives Work: Getting Teachers to Come to School

Duflo discusses a "randomization checklist":

- 1. What was the power of the Experiment? At what level was the experiment randomized? "Need to take into account clustering at that level in computing our standard error"
- 2. What the randomization successful (was there balance between treatment and control group in covariates)
- 3. Did we have attrition (lost observations)?
- 4. Did we have non-compliance?
- 5. Did we have contagion (externalities) between treatment and control group?