Applied Microeconometrics II

Review Session 1

Ellen Sahlström[1]

Aalto University

October 27, 2023

[1]Slides heavily inspired by Kimmo Palanne, thank you!

About me

- Second-year economics PhD student at Aalto University
- Main research interest in education and inequality: *Why are immigrants performing worse in school?*
- Current projects:
  - Differences in primary school social networks (quasi-random variation)
  - Teacher anti-immigrant bias and its consequences for immigrant students (RDD)
  - Preventing social exclusion through workshop on integration? (field experiment)
- Feel free to reach out to discuss research ideas!

Practicalities

- Review sessions main forum for questions: Please be active!
  - Alternatively, email me at ellen.sahlstrom@aalto.fi
  - You can also let me know of a question by email, and I can prepare an answer for the review session
- Submit your solutions to the assignments on MyCourses by the dates specified in the syllabus
  - Work in groups of 2 or 3
  - State the names of your group members in the submissions
  - Keep answers tidy and include log files when required
- Make sure to fill in your research interests and group members in the spreadsheet

Hypothesis testing

Hypothesis Testing: An Example

Example from Bertrand, M and Mullainathan, S. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." American Economic Review 94(4), 991–1013.

- Resumes with traditionally African American or traditionally white names were randomly submitted to employers
  - Out of 2,435 resumes with white-sounding names, the proportion of callbacks was 0.0965
  - Out of 2,435 resumes with African American-sounding names, the proportion of callbacks was 0.0645

Hypothesis Testing: An Example

	White name	African American name
Number of resumes sent	2,435	2,435
Callback rate	0.0965	0.0645

1. What is the null hypothesis?
2. How do we test the null hypothesis?
3. Do we reject the null hypothesis?
4. What do we conclude about discrimination in the labor market?

1. What is the null hypothesis?

- Null hypothesis $H_0 : \mu_W = \mu_{AA}$
  - The callback rates for African American (AA) and White (W) names are equal in the population
- Alternative hypothesis $H_1 : \mu_W \neq \mu_{AA}$
  - The callback rates for African American (AA) and White (W) names are not equal in the population

2. How do we test the null hypothesis?

1. Form the test statistic
2. Calculate the size of the test statistic
3. Calculate (look up) the p-value

2.1 Form the test statistic

- The test statistic under the null
  - $z = \dfrac{\text{difference in callback rates} - 0}{\text{standard error of the difference}}$
- Difference in callback rates
  - $p_W - p_{AA} = 0.0965 - 0.0645 = 0.0320$
- The callback rates can also be written as $p_i = \dfrac{x_i}{n_i}$ for $i = \{W, AA\}$
  - x_i is the number of callbacks
  - n_i is the number of resumes sent

2.1 Form the test statistic

- To get the **standard error** of the difference, first calculate its variance

$$
\begin{aligned}
Var(p_W - p_{AA}) &= Var(p_W) + Var(p_{AA}) \\
&= Var\left(\frac{x_W}{n_W}\right) + Var\left(\frac{x_{AA}}{n_{AA}}\right) \\
&= \frac{1}{n_W^2} Var(x_W) + \frac{1}{n_{AA}^2} Var(x_{AA}) \\
&= \frac{1}{n_W^2} n_W \mu_W (1 - \mu_W) + \frac{1}{n_{AA}^2} n_{AA} \mu_{AA} (1 - \mu_{AA}) \\
&= \frac{\mu_W (1 - \mu_W)}{n_W} + \frac{\mu_{AA} (1 - \mu_{AA})}{n_{AA}}
\end{aligned}
$$

2.1 Form the test statistic

- Under the null we have $\mu_W = \mu_{AA} =: \mu$ so we further get

$$\begin{aligned} Var(p_W - p_{AA}) &= \frac{\mu_W(1-\mu_W)}{n_W} + \frac{\mu_{AA}(1-\mu_{AA})}{n_{AA}} \\ &\overset{H_0}{=} \mu(1-\mu)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right) \end{aligned}$$

- We can estimate μ using the average callback rate $p := \frac{n_W p_W + n_{AA} p_{AA}}{n_W + n_{AA}}$
- An estimator for the standard error of the difference under the null is thus

$$\widehat{SE} = \sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)}$$

2.1 Form the test statistic

- The test statistic under the null
  - $z = \dfrac{\text{difference in callback rates} - 0}{\text{standard error of the difference}}$
- With $H_0 : \mu_W = \mu_{AA}$, we have now defined the test statistic as:

$$z = \frac{p_W - p_{AA}}{\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)}} \xrightarrow{d} \mathcal{N}(0,1)$$

2.2 Calculate the size of the test statistic

- The size of the test statistic is now

$$z = \frac{p_W - p_{AA}}{\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)}} = \frac{0.0965 - 0.0645}{\sqrt{0.0805 \cdot (1 - 0.0805)\left(\frac{1}{2435} + \frac{1}{2435}\right)}} \approx 4.10$$

2.3 Calculate the p-value

- The two-tailed p-value is then

$$p = P(|z| > 4.10) = 2 \cdot (1 - P(z \leq 4.10)) = 2 \cdot (1 - \Phi(4.10)) < 0.0001$$

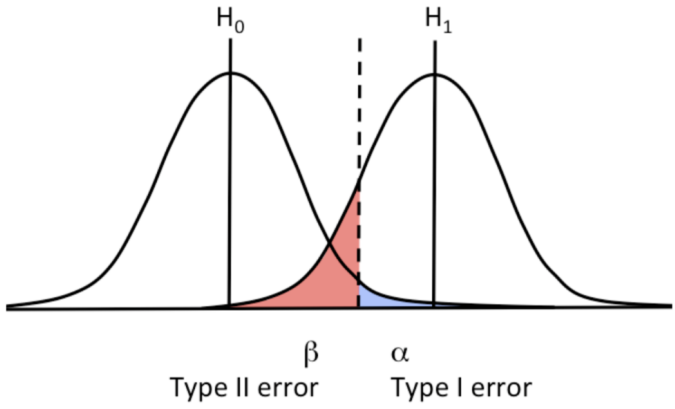
- We can reject the null hypothesis at the 0.01% level

2. What do we conclude about discrimination in the labor market?

- Individuals with African American-sounding names are discriminated in the callback stage of the hiring process solely based on their name

Power Analysis

Power Analysis



- The interpretation of power is $P(\text{Reject } H_0 | H_1 \text{ is true})$
- This is equivalent to calculating the area $1 - \beta$

Power Analysis

- Choose $\alpha = 0.05$
- Now let the alternative hypothesis be $H_1 : \mu_W - \mu_{AA} = \theta > 0$
- In a one-sided test we reject the null of $H_0 : \mu_W = \mu_{AA}$ if

$$\frac{p_W - p_{AA}}{\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)}} > \Phi^{-1}(0.95)$$

$$\iff p_W - p_{AA} > \Phi^{-1}(0.95)\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)}$$

Power Analysis

- Under H_1 we assume that μ_W and μ_{AA} are not equal so we also use a slightly different estimator for the standard error
- We know that

$$Var(p_W - p_{AA}) = \frac{\mu_W(1-\mu_W)}{n_W} + \frac{\mu_{AA}(1-\mu_{AA})}{n_{AA}}$$

- Now we estimate μ_W and μ_{AA} using p_W and p_{AA}, respectively, and get

$$\widehat{SE} = \frac{p_W(1-p_W)}{n_W} + \frac{p_{AA}(1-p_{AA})}{n_{AA}}$$

Power Analysis

- Finally, the type II error is

$$\beta = P\left(\frac{p_W - p_{AA} - \theta}{\sqrt{\frac{p_W(1-p_W)}{n_W} + \frac{p_{AA}(1-p_{AA})}{n_{AA}}}} < \frac{\Phi^{-1}(0.95)\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)} - \theta}{\sqrt{\frac{p_W(1-p_W)}{n_W} + \frac{p_{AA}(1-p_{AA})}{n_{AA}}}}\right)$$

$$= \Phi\left(\frac{\Phi^{-1}(0.95)\sqrt{p(1-p)\left(\frac{1}{n_W} + \frac{1}{n_{AA}}\right)} - \theta}{\sqrt{\frac{p_W(1-p_W)}{n_W} + \frac{p_{AA}(1-p_{AA})}{n_{AA}}}}\right)$$

and power is then $1 - \beta$

What is next?

Questions?

- Find a group and submit name of group members
- Submit also research interests (fields)
- Start working on Assignment 1

Next week:

- Diff-in-diff
- Questions on PS1
- Help with PS2