Type I Errors, Type II Errors, and Power

Lecture 15

Reading: Sections 12.9 – 12.10
("Adaptive Partial Drug Approval: A Health Policy Proposal" Readings page)

1

Type I and Type II Errors

- Type I Error: Reject a true null hyp.
- Type II Error: Fail to reject a false null hyp.
- For example, in trial H_0: innocent; H_1: guilty
  - Type I Error: Convict innocent person (DNA test exonerate)
  - Type II Error: Set guilty person free

	Guilty	Innocent
Convict	No Error	Type I Error
Acquit	Type II Error	No Error

2

Significance Level Recap & Type I Error

- Significance level (α): *Maximum* probability you are willing to tolerate that sampling error caused your observed results: if probability is lower then results are *statistically significant*
  - α is maximum chance of a Type I Error that you would tolerate: i.e. that your sample differs from a true H_0 *only by chance* (sampling error)
    - $\alpha = 0.05$: ready to risk 5% chance of rejecting a true H_0
    - How to reduce the chance of a Type I error?

3

Ex: Lower Sodium

- A gov't agency claims that fewer than 20% of soup eaters notice if sodium is lowered by one-third
- A soup maker wants to prove this wrong
  - H_0:
  - H_1:
- If in fact ___ percent of *all* soup eaters would notice the lower sodium and the P-value for the soup maker's study is ___ then this is an example of ___.

4

β = Probability of a Type II Error

- β = P(fail to reject H_0 when it is false)
  - It's a probability: it must be between 0 and 1
- *Many* factors affect the size of β: one is α
  - Decreasing α (max. tolerable chance of Type I error) increases β (chance of Type II error)
  - If raise burden of proof ($\downarrow \alpha$) so as not to convict the innocent, increase chance guilty go free ($\uparrow \beta$)
  - If lower burden of proof ($\uparrow \alpha$) to "put criminals in jail" ($\downarrow \beta$), increase chance the innocent go to jail

5

Power

- A powerful test is highly likely to lead you to reject a false null hypothesis
  - Power is the complement of Type II error: i.e. the chance you do NOT make a Type II error
  - Power $= 1 - \beta$
    - β = P(Type II Error) = P(fail to reject H_0 when it is false)
  - Power is important: forget costly data collection if the n you are planning will yield insufficient power
    - Increasing the sample size increases power

6

Sex Ratios at Birth in Ontario

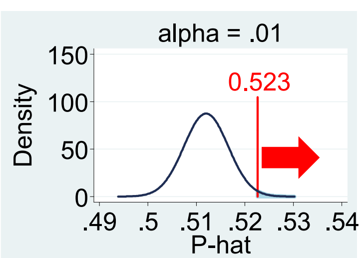
- Recall Ontario sex ratios from Lectures 13, 14
  - Natural proportion of boys born is 51.2%
  - H_0: p = 0.512; H_1: p > 0.512
    - What would a Type I error be?
    - What would a Type II error be?
- How powerful is a statistical test to detect an unnaturally high proportion of males?
  - To calculate power, must also specify α, n, and exactly what we would consider unnaturally high

7

What's Needed to Find Power?

- H_0: p = 0.512; H_1: p > 0.512
- 766,688 births in Ontario from 2002 – 2007
  - But divide it to separately study subgroups
    - i.e. 1st child of Chinese born mom where n = 12,339
  - Consider a "typical" subgroup with n = 12,000
- Choose α = 0.01
- Unnaturally high? Let's say an extra 1 percentage point boys: i.e. p = 0.522

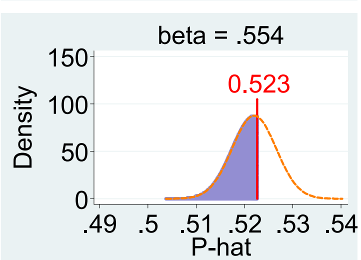
8



- H_0: p = 0.512
- H_1: p > 0.512
- n = 12,000

$$SD[\hat{P}|H_0] = \sqrt{\frac{0.512(0.488)}{12{,}000}} = 0.0046$$

- α = 0.01



- If p = 0.522
- Chance we would fail to reject the false null is very high: P(Type II error) is 0.554
- Power is low: there is only a 44.6% chance our sample would allow us to reject H_0

Everything on this slide determined BEFORE collecting data

9

Size of Type I and II Errors

- Type I Error: Reject a true H_0
  - Set maximum chance of Type I error when pick α
- Type II Error: Fail to reject a false H_0
  - P(Type II error) is β; It depends on many factors:
    - Parameter value in H_0 and direction of H_1
    - Significance level (α)
    - Sample size (n)
    - True parameter value (e.g. p)
    - Which of these 4 factors are observed?

Which type of error is more serious? (See page 388.)

10

Pharmaceutical Ex. (p. 390)

- Huge sunk costs in drug development
  - Pharmaceutical companies do not want to fail to market an effective drug
- Suppose a cancer drug deemed effective if it stops tumor growth in at least 40% of patients
  - H_0: p = 0.40
  - H_1: p > 0.40
  - Where is the burden of proof?

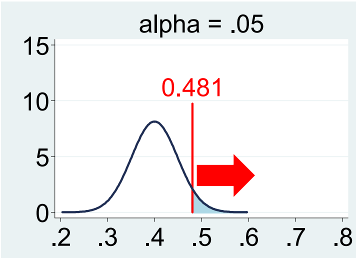
If interested in learning more: Lakdawalla (2018) "Economics of the Pharmaceutical Industry" https://doi.org/10.1257/jel.20161327, which discusses Manski (2009).

11

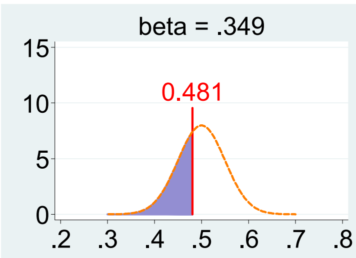
Type II Error: Drug Example

- H_0 value, H_1 direction
  - H_0: p = 0.4
  - H_1: p > 0.4
- Significance level (α)
  - α = 0.05
- Sample size (n)
  - n = 100
- True parameter
  - p = 0.5
- In this case, clearly H_0 is wrong and H_1 is correct
  - Why? Because p = 0.5 (0.5 is greater than 0.4)
- Hence whenever we do not reject H_0 we are making a mistake
  - Which kind of mistake?

12

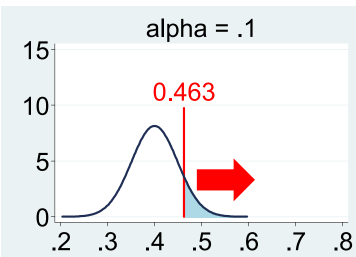


- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.05
- n = 100

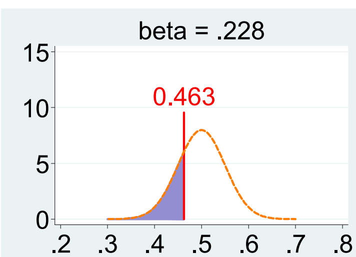


- If p is really 0.5

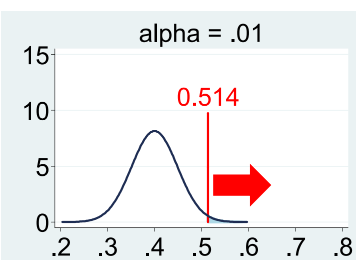
$$P(\hat{P} < 0.481 \mid p = 0.5, n = 100)$$
$$= P\left(Z < \frac{0.481 - 0.5}{\sqrt{\frac{0.5(0.5)}{100}}}\right)$$
$$= P(Z < -0.38) = 0.35$$



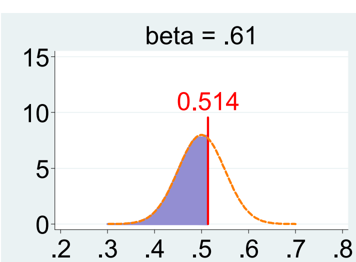
- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.10 ⮜⮜
- n = 100



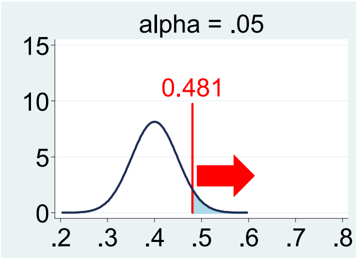
- If p is really 0.5
- Is this test more powerful?



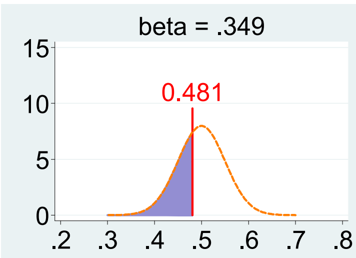
- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.01 ⮜⮜
- n = 100



- If p is really 0.5
- Is this test less powerful?

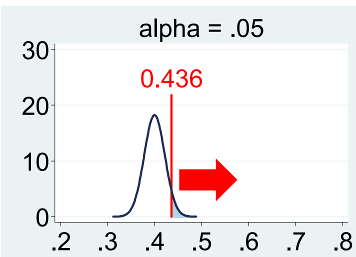


- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.05
- n = 100

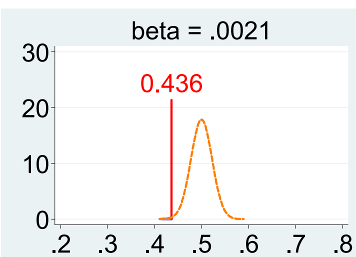


- If p is really 0.5

16

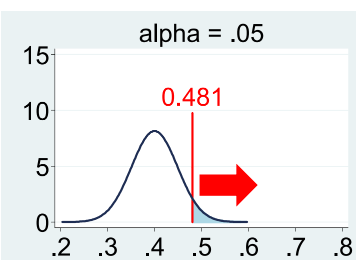


- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.05
- n = 500 ≺≺

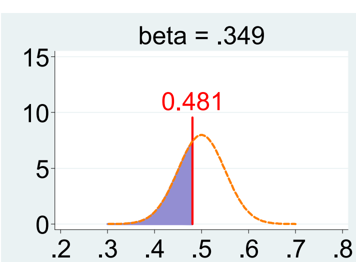


- If p is really 0.5
- Is this test powerful?

17

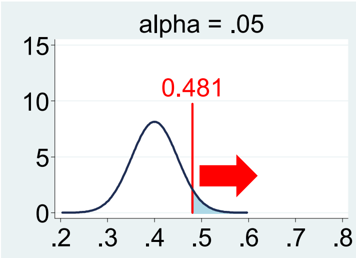


- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.05
- n = 100

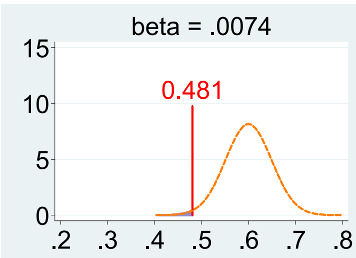


- If p is really 0.5

18



- H_0: p = 0.4
- H_1: p > 0.4
- α = 0.05
- n = 100



- If p is really 0.6

19

Power: Got It?

- Can you compute power before seeing data?
- Should you draw graphs to find power?
- What do you need to specify to find power (or its complement: probability of Type II error)?
  - Review today's notes and chart how changes in each factor affect power and explain why
- What does it mean if your statistical test is not very powerful (i.e. has a high chance of Type II error)?

20