UNIT 1.6
STATISTICAL POWER OF STUDY DESIGNS

Type I Error and Type II Error of hypothesis testing

- Design issues to consider before collecting data.
  - Type I error: rejecting the null hypothesis when it is true.
  - Type II error: failing to reject the null hypothesis when it is false.

Type I and Type II Errors

		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	**Reject the null hypothesis**	**Type I error** (rejecting a true null hypothesis) α	Correct decision
	Fail to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) β

Elementary Statistics, 10th Edition

Design issues

3

- Typical study process:

  Before collecting data…
  - Set the level of significance α.
  - Set the sample size *n*.

  → Type 1 error and power are determined before collecting data

  After collecting data…
  - Compute p-value.
  - If the p-value is less than α, reject the null hypothesis.

Type I Error and Type II Error of hypothesis testing

4

- Probability of type 1 error (= α) → Usually small because people pay attention
  - P(reject H_0 | H_0 is true)
    - For continuous outcomes, type 1 error is automatically the cutoff for p-value (i.e. the level of significance).
- Probability of type 2 error (= β) → Often big for poorly designed studies
  - P(fail to reject H_0 | H_0 is false)
- Power= P(reject H_0 | H_0 is false)
  - Power=1- type 2 error
- *These probabilities are measured before collecting data: that is, they are qualities of the whole study process (and not just of the observed data.)*

Power= P(reject H_0 | H_0 is false)

5

The following factors affect power: must be considered BEFORE data collection.

- Null hypothesis
- Choice of statistical test
- Type 1 error
- Reasonably conservative effect size (= Alternative hypothesis)
- Sample size
- Nuisance parameter (e.g. standard deviation)

5

Example – serum cholesterol levels

6

- A researcher aims to test whether adult diabetic patients in Bronx have higher mean serum cholesterol levels than US adults. It is known that the mean cholesterol levels for US adults is 211 mg/100ml. He is considering a design with 200 adult sample and level of significance 0.05. What is the statistical power of the study design?
    - Sample size = 200
    - Null hypothesis: Cholesterol level among Bronx adults = 211 mg/100ml
- After statisticians look at the problem, they will ask, …
    - "What is the standard deviation for these patients?"
        - Use US adults cholesterol standard deviation σ =46 mg/100ml.
    - "What is the (reasonably conservative) effect size? "
        - 220.0 mg/100ml
    - "Then, you have 79% power to detect the mean at 220 mg/100ml."

6

Example – serum cholesterol levels

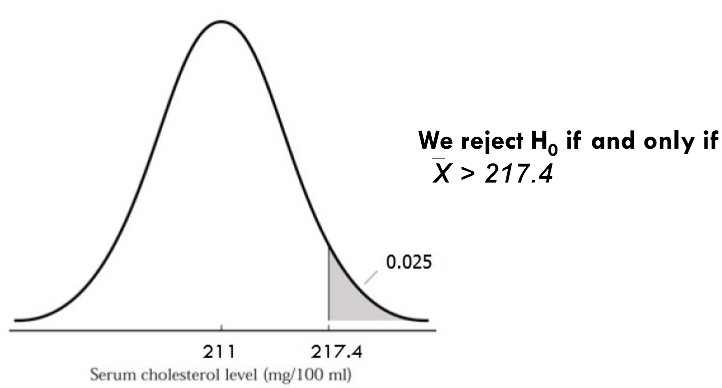
7

- The standard deviation of the cholesterol levels of US adults is σ=46 mg/100ml. We will make an assumption that the variation is the same among Bronx diabetic adults.
- $H_0 : \mu$ = 211 mg/100ml, $H_A : \mu \neq$ 211 mg/100ml
- We reject H_0 if |Z| > 1.96 because α is 0.05.
- $\left|\frac{\bar{x}-211}{46/\sqrt{200}}\right| > 1.96$. So we reject H_0 if and only if $\bar{X}$ > *217.4*
- You will see that we may ignore the situation $\bar{X}$ < *204.4* for the purpose of power calculation because it is very unlikely.

Example – serum cholesterol levels

8

Following is the distribution of $\bar{X}$ assuming H_0 is true. It is Normal distribution with mean 211 and standard deviation $46/\sqrt{200}$.



Distribution of means of samples of size 200 for the serum cholesterol levels of Bronx adult diabetic patients (under null hypothesis)

Type II Error

9

- Type II error depends on what the true state of nature (or point alternative hypothesis) is.

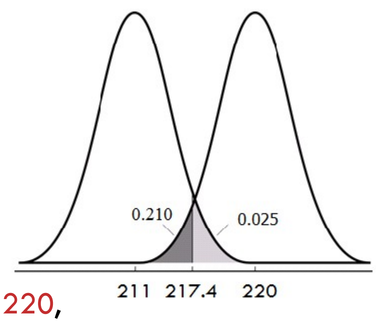
  $H_0 : \mu = 211$ mg/100ml vs. $H_A : \mu = 220\ (=\mu_1)$ mg/100ml
- Probability of type II error $(= \beta)$
  - β = P(fail to reject $H_0 | \mu = \mu_1$)

    $= P(\bar{X} < 217.4 \mid \mu = 220)$

    $= P(Z < [217.4-220]/[46/\sqrt{200}])$

    $= P(Z < -0.807)$

    $= 0.210$



- With type I error 0.05, sample size 200, $\mu_1 = 220$,

the probability of type II error is 0.210.

Power of a hypothesis test

10

- Power of a hypothesis test
  - P(reject $H_0 |$ H_0 is false)
  - probability of rejecting a false null hypothesis

  Power = 1-probability of type 2 error

- Example (cont.)
  - $1-\beta = 1-0.210 = 0.790$
  - "We have 79.0% power to detect the mean at 220 mg/100ml."

Sample size and μ_1 affects power

11

- Increase sample size to reduce both α and β
- Power increases as μ_1 is further from μ_0.
- Increasing α will decrease β (and increase power)
- Reducing α will increase β (and decrease power)

Calculating sample size when σ is unknown

12

1. Use the range rule of thumb to estimate the standard deviation as follows:
   $\sigma \approx$ range/4.
2. Conduct a pilot study. Start the sample collection process and, using the first several values, calculate the sample standard deviation s and use it in place of σ.
3. Estimate the value of σ from literature review.

Valadez's Erroneous Lot Quality Assurance Sampling Method

13

- According to Asher et al. *Statistical Methods for Human Rights* (2008) , CORE group used Valadez's method do determine if vaccination rate of a community is adequate (>20%).
  - It uses the following null hypothesis and controlled only for type 1 error at 5%.
    - H_0: vaccination rate is adequate
  - What is the problem ?
    - The team in a community will survey how many are vaccinated among a sample and if the number is too little, it will reject the null hypothesis (and conclude that the rate is inadequate) .

Erroneous Method (cont.)

14

- The table shows p-values for all possible events, when the sample size is n=28.
- The decision rule will make the team to conclude adequacy when $x \geq 2$ (two or more vaccinated).
- However, when the true vaccination rate is 15% (and hence inadequate), the probability of $x \geq 2$ is 94%. In other words, the team will *falsely* conclude that community children are protected 94 percent of the times.
- Lessons:

  H_0 should be the conservative status quo.

  Always check for statistical power.

Number of vaccinated	P-values
0	0.002
1	0.015
2	0.061
3	0.160
4	0.315
5	0.501
6	0.678
7	0.818
8	0.910
9	0.961
10	0.985
11	0.995
12	0.999
13~28	1.000

- Reference
  - Principles of Biostatistics (Pagano and Gauvreau)
  - *Elementary Statistics* by Triola, 10th edition.
  - *Applied Statistics for Engineers and Scientists* by Petruccelli et al.

- Acknowledgements
  - Prof. Jayson Wilbur, WPI
  - Prof. Balgobin Nandram, WPI
  - Prof. Lee Jaeyong, Seoul National University
  - Some slides provided by Pearson Education, Inc Publishing as Pearson Addison-Weslely