Basic Concepts of Medical Statistics II

Sasivimol Rattanasiri, Ph.D
Section for Clinical Epidemiology and Biostatistics
Ramathibodi Hospital, Mahidol University

1

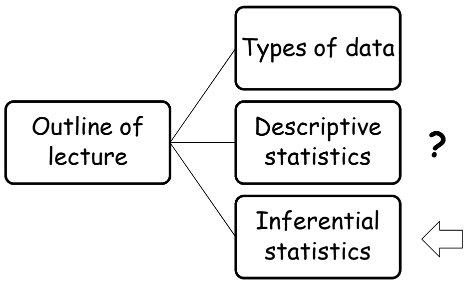

Figure 1 Flow chart for the outline of lecture

2

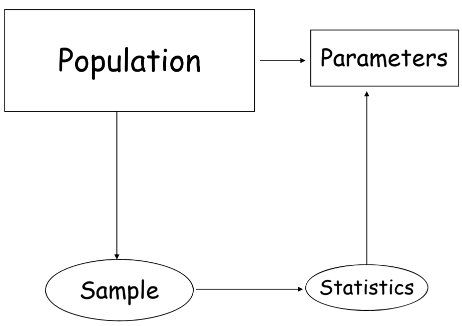

Figure 2 Relationship between population and sample

3

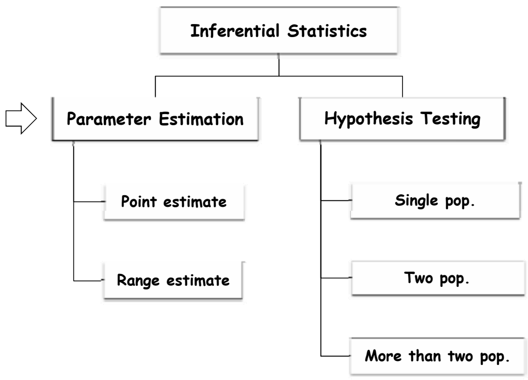

Figure 3 Flow chart for inferential statistics

4

Parameter estimation

For example:

Estimation of the mean age of all patients who have breast cancer in Thailand.

Estimation of the prevalence of patients with diabetes mellitus (DM) in the Thai population.

5

Point Estimate

For example:

- Researchers would like to find the mean age of patients who have DM in Thailand.
- A random sample of 350 patients with DM was taken. The mean age of this sample was 56.12 years.

Therefore, the point estimate is 56.12.

6

Range Estimate

- A range estimate consists of two numerical values which define a range of values that contain the true value of the population parameter.
-A range is always calculated based upon a confidence level.
-The confidence level states how much confidence, there is, that the range contains the true population parameter.

7

Confidence interval

For example:

-The mean age of 350 patients with DM is 56 years with 95% CI of 47 to 62 years.

-This means that the true mean age of patients with DM lies somewhere in range 47 to 62 years.

8

Confidence interval

The main purpose of confidence intervals is to indicate the precision of the point estimates.

9

Confidence interval

Sample size	Mean age	95% CI
30	54	15, 80
500	55	50, 58

10

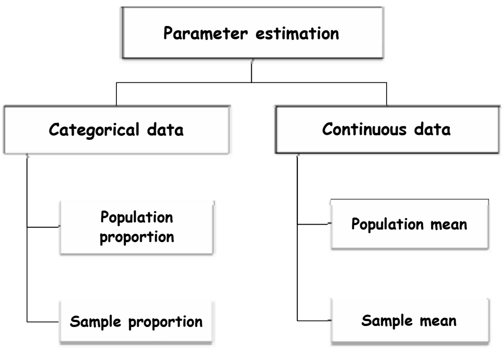

Figure 4 Flow charts for parameter estimation

11

Example I:

- Researcher would like to estimate the prevalence of hypertension in the Thai population.
- A random sample of 749 was taken from the general population. One hundred and twenty-five out of 749 people in the sample had hypertension.

12

STATA command for estimate CI:

- syntax: ci varlist*

***This is binary outcome (code as 1 or 0)**

- immediate command: cii #obs #succ

<u>For example I</u>

- ci hbp

- immediate command

cii 749 125

13

Example II:

-Researcher would like to estimate the mean age of patients with DM in Bangkok.

-A random sample of 750 patients with DM was taken and the mean age of the patients in the sample was estimated.

-The mean age of the patients in the sample was 44.98 with a standard deviation of 14.99.

14

STATA command for estimate CI:

- syntax: ci varlist*

***This is continuous outcome**

- immediate command: cii #obs #mean #sd

<u>For example II</u>

- ci age

- use immediate command

cii 750 44.98 14.99

15

Recommendation

If the primary objective is to estimate the population characteristics, the point estimates should be reported with their confidence intervals to indicate their precision.

16

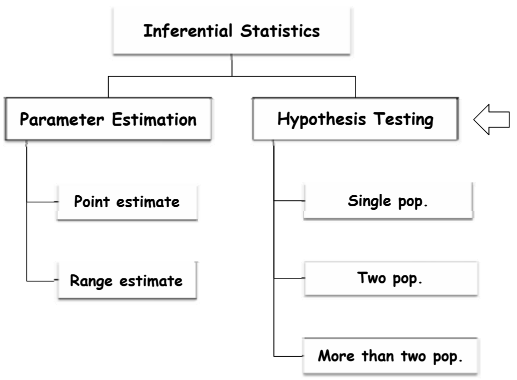

Figure 5 Flow chart for inferential statistics

17

Hypothesis for single population

According to a survey study in Thailand in 2000, the proportion of people with chronic kidney disease (CKD) is 0.091 or 9.1%. Researchers want to test if the proportion of people with CKD in this year is the same as the proportion of people with CKD in 2000 which is 0.091.

18

Hypothesis for two populations

It is known that calcium supplement is associated with increased bone mineral density (BMD) in the adult population. Researchers wanted to test if the means of BMD between postmenopausal women who received and who did not receive calcium supplement are different.

19

Hypothesis testing

The purpose of hypothesis testing is to make a decision about a population by examining a sample from that population.

20

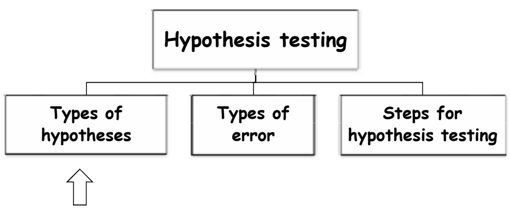

Figure 6 Flow chart for hypothesis testing

21

Types of hypotheses

- Null hypothesis (H_0) :

It is a statement that a population parameter is assumed to be true or there is no difference between groups.

- Alternative hypothesis (H_a):

It is a statement that is opposed to the null hypothesis.

22

For example:

Null hypothesis

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$

Alternative hypothesis

Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

23

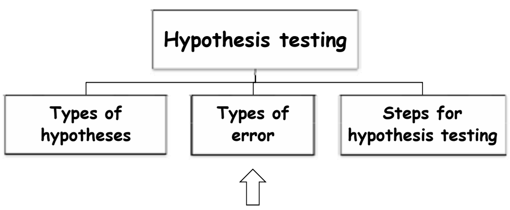

Figure 7 Flow chart for hypothesis testing

24

Types of Errors

There are two possible types of errors that may occur in hypothesis testing:

- Type I error
- Type II error

25

Null hypothesis

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$

Alternative hypothesis

Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

Type I error (α) is the probability of deciding that the means of BMD between women who receive and who did not receive calcium supplement are different, when in fact in the population, the means of BMD between these two groups are not different.

26

Null hypothesis

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$

Alternative hypothesis

Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

Type II error (β) is the probability of deciding that the means of BMD between women who receive and who did not receive calcium supplement are not different, when in fact in the population, the means of BMD between these two groups are different.

27

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$
Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

Statistical Decision Based on Sample	In Population: Ho True	In Population: Ho False
Reject H_0	α **(Type I error)**	$1-\beta$ **(Power of test)**
Do not reject H_0	$1-\alpha$ **(Confidence)**	β **(Type II error)**

28

Diagnostic test

	Actual condition of population	
Test result	Without disease	With disease
Positive	α **(False positive)**	$1-\beta$ **(True positive/ sensitivity)**
Negative	$1-\alpha$ **(True negative/ specificity)**	β **(False negative)**

29

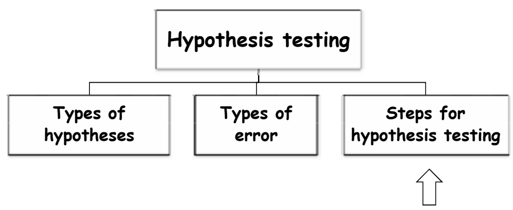

Figure 8 Flow chart for hypothesis testing

30

Steps for Hypothesis testing

Step 1

Generate the null and alternative hypotheses

Null hypothesis

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$

Alternative hypothesis

Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

31

Steps for Hypothesis testing

Step 2

Determine the significance level

The commonly used value of the significance level is 0.05. Usually this value does not exceed 0.10.

32

Steps for Hypothesis testing

Step 3

Select an appropriate test statistic

The objective of the analysis

The type of data

The number of groups of samples

Independent or dependent groups of samples

33

Steps for Hypothesis testing

Step 4

Calculate the test statistic

The general form of the test statistic can be expressed as:

$$\text{test statistic} = \frac{\text{observed value} - \text{hypothesized value}}{\text{standard error of observed value}}$$

34

Steps for Hypothesis testing

Step 5

Convert the test statistic to *p* value

The *p* value is the probability of obtaining our observed data (or more extreme data) when the null hypothesis is true.

35

Steps for Hypothesis testing

Step 6

Draw a conclusion

If the *p* value is less than or equal to the significance level, we reject the null hypothesis.

If the *p* value is greater than the significance level, we do not reject the null hypothesis.

36

For example:

Null hypothesis

Ho: $\mu_{BMD(calcium+)} = \mu_{BMD(calcium-)}$

Alternative hypothesis

Ha: $\mu_{BMD(calcium+)} \neq \mu_{BMD(calcium-)}$

Significance level	**P value**	**Decision**
0.05	0.30	
0.05	0.01	

37