Estimation & Hypothesis Testing (Postgraduate)

Dr Wan Nor Arifin
MBBS (IIUM), MSc (Medical Statistics) (USM).
Lecturer, Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.
E-mail: wnarifin@usm.my

Outlines

Introduction

Statistics

- It "is a field of study concerned with the collection, organization, summarization and analysis of data, and the drawing of inferences about a body of data when only part of the data is observed" (Daniel, 1995).

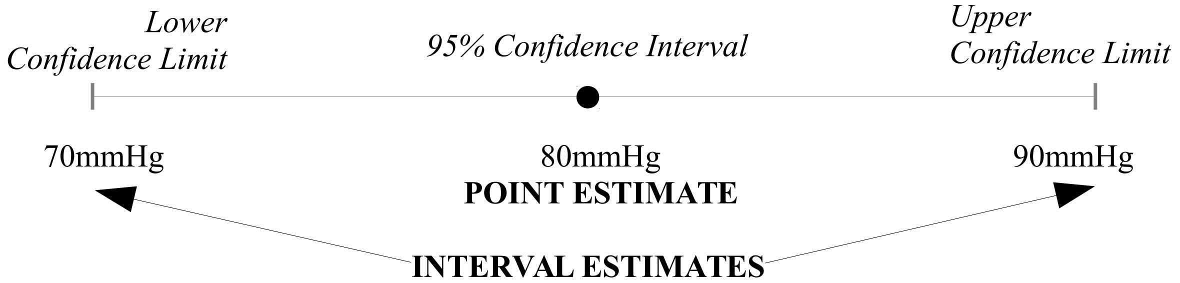
Statistical inference

- It "is the procedure by which we reach a conclusion about a population on the basis of information contained in a sample drawn from that population" (Daniel, 1995).
- It reflects the definition of statistics i.e. making inference about a body of data (population) from part of the data (sample).
- Numerical value calculated from:
  - *Population → parameter*.
  - *Sample → statistic*.
- Two approaches of statistical inference: *Estimation* and *Hypothesis testing*.

Estimation

Introduction

- It is the process of calculating *statistic* from sample data as approximation of *parameter* of the population from which the sample was drawn.
- An *estimate* is used as an approximation of a *parameter*.
- For each parameter, two types of estimate are possible:
  - *Point estimate* – a single numerical value used as an estimation of respective population parameter.
  - *Interval estimates* – two numerical values presented in form of range/interval inside which we believe the parameter is included, given with specified level of confidence.
- By taking a representative random sample → better approximation to population.
- For example in journal "the mean value of diastolic blood pressure is 80mmHg (95% CI: 70, 90).



- Presenting **point estimate** alone not enough → comes from sample, should be accompanied by its **interval estimates**.
- *Mean of sample* is not necessarily → *mean of population*.
- When we estimate, we make an educated guess of the real population parameter.
- It is like saying "I think that based on my sample, the mean DBP is 80 mmHg and I am 95% sure that the real mean DBP in population is between 70 to 90mmHg."
- Present estimate in form of point estimate followed by interval estimates within a certain confidence level:

point estimate (% of confidence: lower confidence limit, upper confidence limit)

- Usually 95% for confidence interval.
- Generally to obtain confidence interval:

$$\text{point estimate} \pm (\text{reliability coefficient}) \times (\text{standard error})$$

or in short

$$\text{point estimate} \pm \text{precision}$$

- **Reliability coefficient** depends on sampling distributions used and confidence level.
- **Standard error** is the standard deviation of sampling distribution.
- Remember we are estimating based on sample, thus in place of commonly use standard deviation to deal with population (i.e in descriptive statistics lecture), we use standard error.
- It shows how "erratic" is our sample as compared to population. Smaller standard error

gives us narrower confidence interval (better as it narrows down possible values of population parameter) and also better precision.

- Think of an analog watch which can give you the time precise to the closest 1 second as compared to a digital atomic clock used in advanced research center which can give you reading as precise as a fraction of a second.

One sample mean

- Recall that to find average/mean $\bar{x}$ of numerical data:

$$\bar{x}=\frac{\sum x_i}{n}$$

- To find confidence interval for one sample mean, can calculate using *standard normal distribution, z.*
- Prerequisites:
  - Normally distributed data. If data not normally distributed, given large sample size *n* (30 or more), the data can be considered as normally distributed.
  - Population standard deviation σ, is known. If σ is not known, again given large sample size (30 or more), sample standard deviation *s*, can be used in place of σ.
- Confidence interval is given by,

$$\text{point estimate} \pm (\text{reliability coefficient}) \times (\text{standard error})$$

$$\bar{x} \pm z_{(1-\alpha/2)} \times \sigma_{\bar{x}}$$

$$\bar{x} \pm z_{(1-\alpha/2)} \times \frac{\sigma}{\sqrt{(n)}}$$

- Commonly used reliability coefficient using z distribution $z_{(1-\alpha/2)}$ by confidence level are:

90% → 1.645
95% → 1.96
99% → 2.58

<u>Example 1</u>:

From his data on systolic blood pressure (SBP) collected from 30 patients, he found that the mean SBP was 120mmHg with standard deviation of 15mmHg. Estimate with 95% confidence the population parameter.

$$\bar{x}=120$$
$$s=15 \approx \sigma$$
$$\bar{x} \pm z_{(1-\alpha/2)} \times \sigma/\sqrt{(n)}$$
$$120 \pm 1.96 \times 15/\sqrt{(30)}$$
$$120 \pm 1.96 \times 2.739$$
$$95\ CI: 114.6, 125.4$$

Stated in sentence:

"We are 95% confident that the population mean is between 114.6 and 125.4."

or in journal presentation form:

120mmHg (95% CI: 114.6, 125.4)

One proportion

- Commonly used in health sciences. For example percentage of HIV positive among drug addicts, proportion of smokers died of lung cancer etc.
- Sample proportion is written as $\hat{p}$.
- Prerequisite:
    - Sampling distribution of $\hat{p}$ is quite close to normal distribution when both *np* and *n(1-p)* greater than 5 (i.e when sample size *n* is large and the proportion in population *p* not too small), for example:

      n = 1000, p = 0.05, 1 – p = 0.95; np = 50, n(1 – p) = 950 → OK.

      n = 50, p = 0.05, 1-p = 0.95; np = 2.5, n(1 – p) = 47.5 → cannot calculate confidence interval based on $z_{(1-\alpha/2)}$.

- Confidence interval is given by,

$$\text{point estimate} \pm (\text{reliability coefficient}) \times \text{standard error}$$

$$\hat{p} \pm z_{(1-\alpha/2)} \times \sigma_{\hat{p}}$$

$$\hat{p} \pm z_{(1-\alpha/2)} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Similarly, commonly used reliability coefficient using z distribution $z_{(1-\alpha/2)}$ by confidence level are:

  90% → 1.645
  95% → 1.96
  99% → 2.58

<u>Example 2</u>:

It was found that in a study on drug addicts in Kelantan, 130 out of 200 are HIV positive. Construct 99% confidence interval for the proportion of HIV positive among the addicts.

$$\hat{p} = 130/200 = .65$$
$$n = 200$$
$$\hat{p} \pm z_{(1-\alpha/2)} \times \sqrt{\hat{p}(1-\hat{p})/n}$$
$$.65 \pm 2.58 \times \sqrt{(.65(.35)/200)}$$
$$99\ CI : 0.5623, 0.7377$$

Stated in sentence:

"We are 99% confident that the population proportion *p* is between 56.23% and 73.77%."

or in journal presentation form:

65.0% (95% CI: 56.23%, 73.77%)

Hypothesis Testing

Hypothesis (Hx)

- Types of hypothesis:
  - Research Hx – "is the conjecture or supposition that motivates the research" (Daniel, 1995).
  - Statistical Hx – the hypothesis that is stated in a way that is possible to evaluate by appropriate statistical analysis.

- Types of statistical hypothesis:
  - Null Hx (H_0) – Hx of no difference/agreement with population of interest.
  - Alternative Hx (H_A) – Inverse of H_0. Disagreement with population of interest.

- H_0 and H_A statements are 3 ways based on H_A (for example testing mean A and mean B):

1. Different (two-tailed)*:

Mean of population A is **not different** from B vs Mean of population A is **different** from B.
H_0: $\mu_A = \mu_B$ vs H_A: $\mu_A \neq \mu_B$

*Most applicable in health sciences, examples would be based on this Hx statements.

2. More (one-tailed):

Mean of population A is **less than or similar** to B vs Mean of population A is **more** than B.
H_0: $\mu_A \leq \mu_B$ vs H_A: $\mu_A > \mu_B$

3. Less (one-tailed):

Mean of population A is **more than or similar** to B vs Mean of population A is **less** than B.
H_0: $\mu_A \geq \mu_B$ vs H_A: $\mu_A < \mu_B$

- Types of errors related to testing statistical hypothesis:

    - Type I error (α)
        - Erratically rejects True H_0.
        - Erratically fail to reject False H_A.
        - False Positive (Significant) result.

    - Type II error (β)
        - Erratically fail to reject False H_0.
        - Erratically rejects True H_A.
        - False Negative (Insignificant) result.

- Relationship between statistical testing and null Hx can be viewed as relationship between diagnostic test and its gold standard test (Gaddis and Gaddis, 1990) as shown in table below:

		H_0	
		False (Difference +)	True (Difference –)
Test statistic	Significant (Test +)	True Positive $1 - \beta$ Sensitivity/Power	False Positive α Type I Error
	No significant (Test –)	False Negative β Type II Error	True Negative $1 - \alpha$ Specificity

* Familiarize yourself with diagnostic test.

Test statistic

- It is the *statistic* obtained from our calculation using appropriate formula e.g. value of z.
- In general formula for Hx testing is given by,

$$test\ statistic = \frac{relevant\ statistic - hypothesized\ parameter}{standard\ error\ of\ relevant\ statistic}$$

- It is then compared against a preset value e.g. preset value of z (i.e. based on *significance level*)
- Or converted first into probability value called *p-value* and compared against preset *significance level*.

Significance level, α

- It "is a probability of rejecting a true null hypothesis" (Daniel, 1995).
- A preset value of acceptable significance level (0.05, 0.01, 0.001).
- Acceptable level to commit type I error.
- Compared with *p-value* or converted to corresponding z value for comparison (as preset value as mentioned before).

p-value

- It is the probability of getting a test statistic as ***EXTREME*** as the one that we computed.
- If we get very low *p-value*, thus it means that the probability of getting a value similar or better than the one that we have is very unlikely.
- Think of anyone beating Usain Bolt for his world record of 9.58secs in 100m sprint, would you think anyone else can easily get similar or even better record than what he did?
- We reject such thinking/hypothesis when p-value $< \alpha \rightarrow$ Significant result i.e. such thing would not happen easily by chance/very unlikely.

References

Daniel, W. W. (1995). *Biostatistics: A foundation for analysis in the health sciences* (6th ed.). USA: John Wiley & Sons.

Gaddis, G. M., Gaddis, M. L. (1990). Introduction to biostatistics: Part 3, sensitivity, specificity, predictive value, and hypothesis testing. *Annals of Emergency Medicine*, **19(5)**, 145-151.