

Chapter 5:
Chapter 5:



Chapter Five

Hypothesis Testing and Estimation
Hypothesis Testing and Estimation



There are two main purposes in statistics;

- (Chapter 1 & 2) \Rightarrow Organization & ummarization of the data

[Descriptive Statistics]

- (Chapter 5) \Rightarrow Answering research questions about some population parameters

[Statistical Inference]

Statistical Inference \Rightarrow (1) Hypothesis Testing:
Answering questions about the population parameters

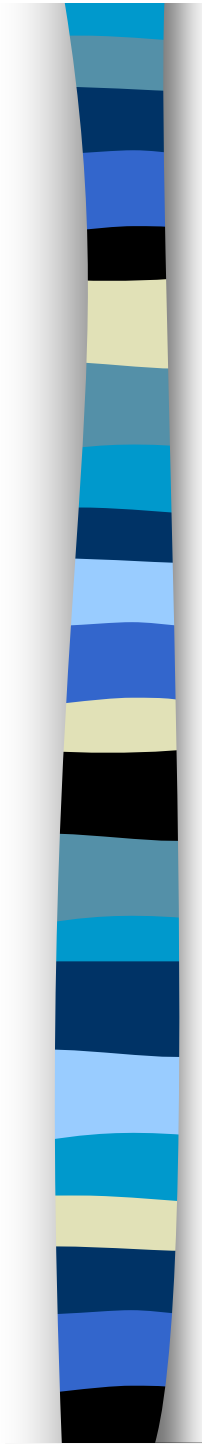
\Rightarrow (2) Estimation:

Approximating the actual values of Parameters;

Ø Point Estimation

Ø Interval Estimation

(or Confidence Interval)



We will consider two types of population parameters:
(1) Population means (for quantitative variables):
 μ = The average (expected) value of some quantitative variable.

Example:

- The mean life span of some bacteria.
- The income mean of government employees in Saudi Arabia.

(2) Population proportions (for qualitative variables)

$$\pi = \frac{\text{no. of elements in the Population with some specified characteristic}}{\text{Total no. of elements in the Population}}$$



Example:

- The proportion of Saudi people who have some disease.
- The proportion of smokers in Riyadh
- The proportion of females in Saudi Arabia

Estimation of Population Mean :-

Population (distribution)

Population mean = μ

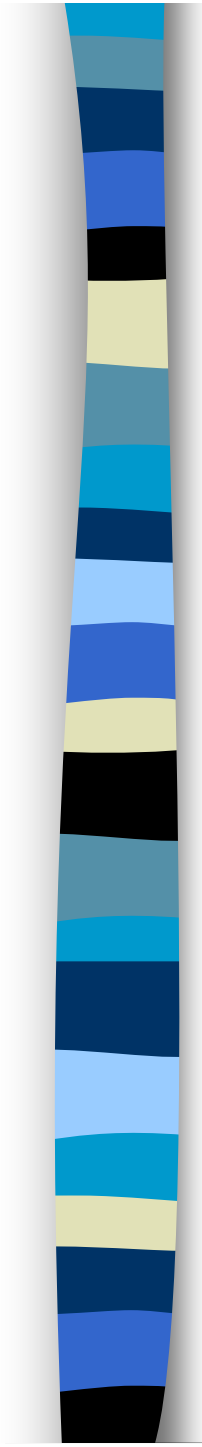
Population Variance = σ^2

Random
Sample

of size
 n

X_1, X_2, \dots, X_n
Sample mean = \bar{X}

Sample Variance = S^2


$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\sum X_i^2 - n\bar{X}^2}{n-1}$$

We are interested in estimating the mean of a population (μ)

(I) Point Estimation:

A point estimate is a single number used to estimate (approximate) the true value of .

• Draw a random sample of size n from the population:

$$X_1, X_2, \dots, X_n$$

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ is used as a point estimator of .



(II) Interval Estimation:

An interval estimate of μ is an interval (L, U) containing the true value of μ “ with probability $1 - \alpha$ “

$1 - \alpha$ is called the confidence coefficient

L = lower limit of the confidence interval

U = upper limit of the confidence interval

- Draw a random sample of size n from the population .
 X_1, X_2, \dots, X_n



Result:

If X_1, X_2, \dots, X_n is a random sample of size n from a distribution with mean μ and variance σ^2 , then:

A $(1-\alpha)$ 100% confidence interval for μ is :

(i) if σ is known:

$$\left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad \text{OR} \quad \bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

(ii) if σ is unknown.

$$\left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right) \quad \text{OR} \quad \bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$



Example: [point estimate of is $\bar{X} = 26.2$]

Variable = blood glucose level (quantitative variable)

Population = diabetic ketoacidosis patients in Saudi Arabia of age 15 or more

parameter = μ = the average blood glucose level

$$n = 123 \text{ (large)}$$

$$\bar{X} = 26.2$$

$$S = 3.3 \text{ (}\sigma^2 \text{ is unknown)}$$

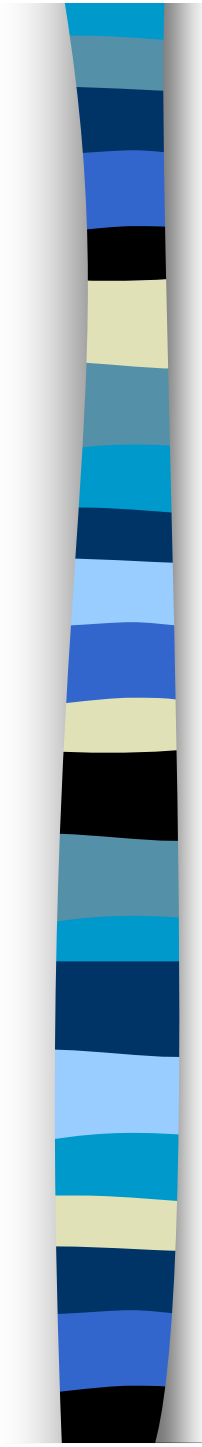
(i) Point Estimation:

We need to find a point estimate for μ .

$\bar{X} = 26.2$ is a point estimate for μ .

(ii) Interval Estimation (Confidence Interval):

We need to find 90% confidence interval for μ .


$$90\% = (1 - \alpha) \quad 100\%$$

$$1 - \alpha = 0.9 \Leftrightarrow \alpha = 0.1$$

$$\frac{\alpha}{2} = 0.05 \quad 1 - \frac{\alpha}{2} = 0.95$$

$$Z_{1 - \frac{\alpha}{2}} = Z_{0.95} = 1.645$$

\therefore 90% confidence interval for μ is:

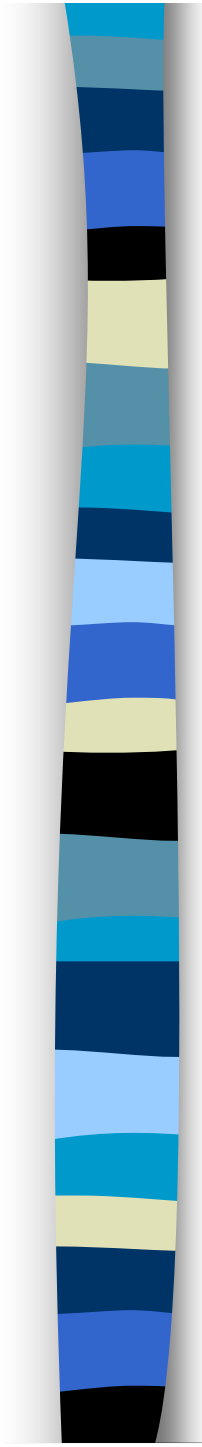
$$\left(\bar{X} - Z_{1 - \frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + Z_{1 - \frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

or

$$\left(26.2 - (1.645) \frac{3.3}{\sqrt{123}}, 26.2 + (1.645) \frac{3.3}{\sqrt{123}} \right)$$

$$(26.2 - 0.4894714, 26.2 + 0.4894714)$$

or $(25.710529, 26.689471)$



we are 90% confident that the mean μ lies in
(25.71,26.69) or

$$25.72 < \mu < 26.69$$

5.4. Estimation for a population proportion:-

- The population proportion is

$$\pi = \frac{N(A)}{N} \quad (\pi \text{ is a parameter})$$

where

$N(A)$ = number of elements in the population with a specified characteristic “A”

N = total number of element in the population
(population size)

- The sample proportion is


$$p = \frac{n(A)}{n} \quad (p \text{ is a statistic})$$

where

$n(A)$ = number of elements in the sample with the same characteristic “A”

n = sample size

Result: For large sample sizes ($n \geq 30$), we have

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \approx N(0,1)$$

Estimation for π :

(1) Point Estimation:

A good point estimate for π is p .

(2) Interval Estimation (confidence interval):

A $(1 - \alpha)100\%$ confidence interval for π is



or

$$p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$
$$\left(p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right),$$



Example 5.6 (p.156)

variable = whether or not a women is obese (qualitative variable)

population = all adult Saudi women in the western region seeking

care at primary health centers

parameter = π = The proportion of women who are obese

$n = 950$ women in the sample

$n(A) = 611$ women in the sample who are obese

$$\therefore p = \frac{n(A)}{n} = \frac{611}{950} = 0.643$$



is the proportion of women who are obese in the sample.

(1) A point estimate for π is $p = 0.643$

(2) We need to construct 95% C.I. about π .

$$95\% = (1 - \alpha)100\% \quad \Leftrightarrow 0.95 = 1 - \alpha$$

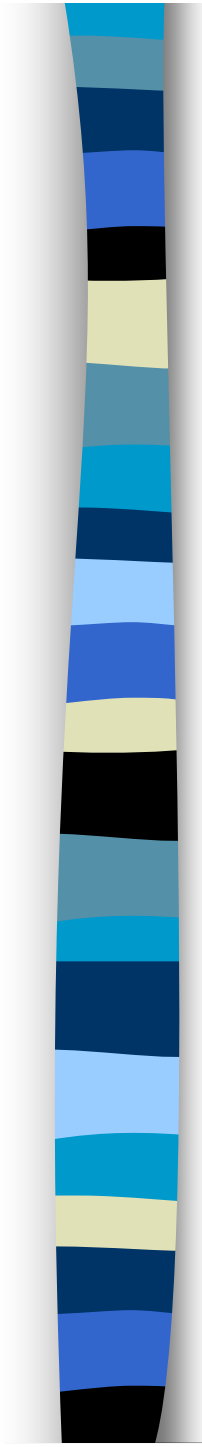
$$\Leftrightarrow \alpha = 0.05$$

$$\Leftrightarrow \frac{\alpha}{2} = 0.025$$

$$\Leftrightarrow 1 - \frac{\alpha}{2} = 0.975$$

$$\therefore z_{1 - \frac{\alpha}{2}} = z_{0.975} = 1.96$$

\therefore 95% C.I. about π is


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

or

$$0.643 \pm (1.96) \sqrt{\frac{(0.643)(1-0.643)}{950}}$$

$$0.643 \pm (1.96)(0.01554)$$

or

$$0.643 \pm 0.0305$$

or

$$(0.6127, 0.6735)$$

We can 95% confident that the proportion of obese women, π , lies in the interval (0.61, 0.67) or

$$0.61 < \pi < 0.67$$