Estimation Theory

- A researcher may wish or interest in estimating population parameters such as (mean, proportion, variance, etc.)
- Two types of estimate:

1- *A point estimate:* Is a single numerical value used to estimate the corresponding population parameter. For example a sample mean, $\bar{X}$, is unbiased estimate of population mean, (μ), *but it does not necessarily equal to* (μ). This type was pervious explained.

2- <u>An interval estimate</u>: An estimate of population parameter given by two numbers between which the parameter may be considered to lie is called an interval estimate.

<u>Confidence Interval for Mean, μ, when σ is *Known*, (C.I):</u>

In repeated sampling, from a normally distributed population, (1 - α) 100% of all intervals of the form,

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

will in the long run include the population mean, μ.

- The confidence interval is determined by two points, one below and above $\bar{x}$, that the population mean (μ), lies within these limits as determined from the sample.

- A (1 - α) 100% C.I for μ is given by

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

$$\text{or } P\left(\bar{x} - Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

(1 - α) 100%= Coefficient of confidence	90%	95%	99%
Z = Coefficient of reliability	1.645	1.96	2.58

- Example: The population standard deviation of SBP of normal adult males, 25 years old is 15 mm Hg, a sample of 100 had a mean of SBP of 125 mm Hg. Construct 95% C.I for μ, (μ = true population mean of SBP).

- 95% C.I for (μ= true population mean of SBP) is,

$$125 \pm 1.96\left(\frac{15}{\sqrt{100}}\right)$$

$$(122.06 \;,\quad 127.94)$$

$$or\ (122.06 \leq m \leq 127.94)$$

- *Meaning that based on sample statistics we are 95% confident that true population mean of SBP lies within the range of 122.06 to 127.94 mm Hg.*

- Confidence Interval for the Difference between Two Population Means μ_1 and μ_2 ;when σ_1 and σ_2 are Known, (C.I).

- A (1 - a) 100% confidence interval for $\mu_1 - \mu_2$ is given by,

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm Z_{(1-\alpha/2)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\text{or } P\left[\left(\bar{x}_1 - \bar{x}_2\right) - Z_{(1-\alpha/2)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + Z_{(1-\alpha/2)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

- Example: A sample of 100 patients with disease A, admitted to a chronic disease hospital, remained in the hospital, on the average, 35 days. Another sample of 100 patients with disease B stayed, on the average, 28 days. If the population variances are 100 and 225, respectively, find the 99 percent confidence interval for $\mu_A - \mu_B$.
- A 99% C.I. for $\mu_A - \mu_B$ is

$$(35-28) \pm 2.58\sqrt{\frac{100}{100}+\frac{225}{100}}$$

$$(2.348 \quad , \quad 11.652)$$

- *Meaning that based on sample statistics we are 99% confident that the true difference between two population means of the two groups of patients stayed in the hospitals lies within the range of 2.348 to 11.652 days.*

Confidence Interval for a Population Proportion:

Many questions of interest to health worker relate to populations. What proportion of some population has a certain disease?

A (1 - a) 100% confidence interval for P is given by,

$$\hat{p} \pm Z_{(1-a/2)}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

$$P\left(\hat{p} - Z_{(1-a/2)}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq P \leq \hat{p} + Z_{(1-a/2)}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1 - a$$

- Example: A certain drug was found to the effective in the treatment of pulmonary disease in 180 out of 200 cases treated, Construct the 90 percent confidence interval for the population proportion.

$$\hat{p} = 180/200 = 0.9 \qquad , \qquad \hat{q} = 1 - 0.9 = 0.1$$

A 90% C.I. for P is $0.9 \mp 1.645 \sqrt{\frac{(0.9)(0.1)}{200}}$

(0.865 , 0.935)

Or

(0.865 ≤ P ≤ 0.935)

- *Meaning that based on sample statistics we are 90% confident that the true population proportion of* pulmonary disease *lies within the range of 0.865 to 0.935.*

- <u>Confidence Interval for the Difference between Two Population Proportions:</u>
- A (1- a) 100% C.I. for $P_1 - P_2$ is given by,

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{(1-a\ /\ 2)} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

- *Note: The method appropriate when np>10 and npq>10. However if either np or npq is less than 10 the terms in Binomial distribution do not agree well with normal curve.*

- *<u>Example:</u> A study was conducted to look at the effects of oral contraceptives (OC) on heart disease in women 40-44 years of age. It is found that among 5000 current OC users at baseline, 13 women develop a MI over a 3 – year period, while among 10,000 non-OC users, 7 develop an MI over a 3 -year period.*
- *Compute a 95% CI for the difference between the proportion of women who develop MI among OC users and the comparable proportion among OC user's information;*

- n1=5000 $\hat{p}_1 = 13/5000 = 0.0026, \quad \hat{q}_1 = 1 - \hat{p}_1 = 1 - 0.0026 = 0.9974$
- n2=10000 $\hat{p}_2 = 7/10000 = 0.0007, \quad \hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.0007 = 0.9993$
- A (1- a) 00% C.I. for $P_1 - P_2$,

$$0.0026 - 0.0007 \pm 1.96\sqrt{\frac{0.0026 * 0.9974}{5000} + \frac{0.0007 * 0.9993}{10000}} = 0.0019 \pm 0.0015$$

$$(0.0004, 0.0034)$$

- *Meaning that based on sample statistics we are 95% confident that the true difference between the proportion of women who develop MI among OC users and the comparable proportion among OC user's information lies within the range of (0.0004, 0.0034).*