STATISTICS

UNIT 4

1.0 Estimation in statistics

1.1 Point Estimate & Interval Estimate

1.2 Biased & Unbiased Estimate

1.3 Confidence Interval

1.4 Confidence Limits

1.5 Confidence Interval for proportions

1.6 Choice of Samples

1 Estimation in Statistics

1.0 Estimation in statistics

- This refers to the process by which one makes inferences about a population, based on information obtained from a sample; samples means are used to estimate population means; sample proportions to estimate the population proportions.

1.1 Point Estimate Vs. Interval Estimate

- Point Estimate: A point estimate of a population parameter is a single value of a statistic.

  For example, the sample mean $\overline{X}$ is a point estimate of the population mean μ. Similarly, the sample proportion p is a point estimate of the population proportion P.

- Interval Estimate: An interval estimate is defined by two numbers, between which a population parameter is said to lie.

  $a < \overline{X} < b$ *is an interval estimate of the population parameter mean μ*

1.2 Biased & unbiased Estimate

- A statistic is biased if the long term average value of the statistic is not the parameter it is estimating.
- As seen earlier on the sampling distribution of the mean, the mean of the sampling distribution of the sample mean is the population mean (μ). Therefore the sample mean is an unbiased estimate of μ.
- When there is true for any parameter, the statistic is referred to as an unbiased estimator
- $\frac{N-1}{N}\sigma_s^{\,2}$ is an un-biased estimator of the population variance, σ^2
- Here, $\sigma_s^{\,2}$ is a sample variance, or the mean of a number of sample variances and N is the sample size.
- *Consistent estimates of a parameter are estimates which become more accurate as the sample size increases.*

1.3 Confidence Interval

- Statisticians use a confidence interval to express the precision and uncertainty associated with a particular sampling method.
- A confidence interval consists of three parts:

  a. A confidence level
  b. A statistic
  c. A margin of error
- The confidence level describes the uncertainty of a sampling method.
- The statistic and the margin of error define an interval estimate that describes the precision of the method.
- The interval estimate of a confidence level is defined by the sample statistic ± margin of error.

1.3 Confidence Interval – continued

- For example, suppose we compute the interval estimate of a population parameter.
- We might describe this interval estimate as a 95% confidence interval.
- This means that if we used the sampling method to select different samples and compute different interval estimates, the true population parameter would fall within fall within a range defined by the sample statistic ± margin of error 95 % of the time.
- Confidence intervals are preferred to point estimates, because confidence intervals indicate

  a) the precision of the estimate

  b) the uncertainty of the estimate

1.4 Confidence Limits

Confidence Limits

- Confidence limits are the lower and upper boundaries / values of a confidence interval, that is, the values which define the range of a confidence interval.
- The 95% confidence level for a normal curve, the confidence interval is $\mu - 1.96\sigma$ to $\mu + 1.96\sigma$
- These two values are called confidence limits.
- The factor 1.96 is called the confidence coefficient, z_c

Confidence Level	Confidence Coefficient	Confidence Level	Confidence Coefficient	Confidence Level	Confidence Coefficient
99.73%	3.00	96%	2.05	80%	1.28
99%	2.58	95.45%	2.00	75%	1.15
98%	2.33	95%	1.96	68.27%	1.00
97%	2.17	90%	1.64	50%	0.67

Normal Curve

1.4 Confidence Limits – continued

- We have already seen that the mean of the sample means is an unbiased estimator of the population.
- For an infinite population, the population mean is

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

where is the sample mean and σ is the standard deviation of the population.

- However, σ is unknown and we use instead the sample estimate σ_s; N is the sample size.
- **Example**: If the value of the mean of the population, estimated from a sample of size 100 is 8.12 ± 2.5, what would you expect the estimate to be, based on a sample size of 200?

Solution:

$$8.12 \pm \{\sqrt{(100)}/\sqrt{(200)}\} \; X \; 2.5 = 8.12 \pm \{0.7071\} * 2.5$$
$$= 8.12 \pm 1.8$$

1.4 Confidence Limits – continued

- For a finite population, the formula becomes

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

Here N_p is the population size.

- For small sample (N < 30) these formulas will over-estimate the reliability of the estimate and the formula becomes,

$$\bar{X} \pm t_c \frac{\sigma}{\sqrt{N - 1}}$$

Where t is obtained from student's t distribution.

1.5 Confidence Interval for proportions

Confidence interval for proportions

$$P \pm Z_c\sqrt{\frac{pq}{N}}$$

where P is the proportion of success in a sample size N

p is the population proportion of success; q= 1 – p

- For a large sample: $$P \pm Z_c\sqrt{\frac{P(1-P)}{N}}$$
- For a finite population: $$P \pm Z_c\sqrt{\frac{pq}{N}}\sqrt{\frac{N_p - N}{N_p - 1}}$$
- Differences of means: $$X_1 - X_2 \pm Z_c\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$
- Sum of means: $$X_1 + X_2 \pm Z_c\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$
- Standard Deviation: $$s \pm \frac{Z_c \sigma_s}{\sqrt{2N}}$$ *if the population is normally distributed*

1.5 Confidence Interval for proportions

Example: A sample of 200 employees of a large company indicated that 65 % thought that the company's promotion procedures were satisfactory. Find the 50%, 95% and 99% confidence limits for the proportion of all employees who are satisfied with the promotion procedures.

Solution:

$$\sqrt{\frac{pq}{N}} = \text{sq. root } \{(0.65 \times 0.35) / 200\} = 0.034$$

$$P \pm Z_c\sqrt{\frac{pq}{N}}$$

For 50 % confidence limits, we have $0.65 \pm (0.67 \times 0.34) = 0.65 \pm 0.02$

For 95 % confidence limits, we have $0.65 \pm (1.96 \times 0.34) = 0.65 \pm 0.07$

For 99 % confidence limits, we have $0.65 \pm (2.58 \times 0.34) = 0.65 \pm 0.09$

1.6 Choice of sample

Choice of sample

- In planning a sample study, size of the sample depends on
  - ✓ *Confidence interval and confidence limits desired*
  - ✓ *Cost*
  - ✓ *Feasibility*

Example: *The mean and standard deviation of a large population are approximately 10.0 and 1.0 respectively. What sized sample should be used to determine the mean within a confidence interval ± with 90 % confidence.*

Solution:

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

$0.01 = Z_c \ X \ \{1.0 \ / \ \sqrt{N}\}$

For 90% confidence, $Z_c = 1.64$

$0.01 = 1.64 \ X \ \{1.0 \ / \ \sqrt{N}\}$

$\sqrt{N} = (1.64 \ X \ 1.0) \ / \ 0.1 = 16.4$

$N = 16.4^2 = 269$