CHAPTER TEN : SAMPLING AND ESTIMATION

Introduction

- To learn the population characteristics we have to use samples. A sample is a proportion or sub-set of the population, and is used to predict the characteristics of the population. Based on the principle of statistical regularity and the principle of "Inertia of large numbers", if the collected sample is not biased and is large enough, to a certain extent the sample results will reveal the characteristics of the population.
- Population parameters can be predicted from samples. We can collect a sample from the population and compute appropriate sample statistics. The sample statistics is used to predict the value of the population parameters. This is known as e*stimation. Estimation* can be defined as " techniques used to establish a value for an unknown population parameter using the respective sample statistics".

Population
Sample
Census
Random Sampling
Attribute Sampling

Population Parameter is the value obtained from a set of data which represents all the observations in the designated population.

Sample Statistic is the value that describes the sample.

The table below shows the designated statistical value, the symbol used to describe a sample and a population.

Statistical value	Symbols Used	
	Sample statistic	Population parameter
Mean	x	μ
Standard Deviation	s	σ
Variance	s^2	σ^2
Proportion	p	π

Standard Error (SE)

Standard error measures the sampling error associated with a sampling distribution. The larger the size of the sample taken, the smaller the standard error.

SE for sampling distribution of the mean

$SE = \frac{\sigma}{\sqrt{n}}$ where σ: population standard deviation and *n*: sample size used

If 'σ' is not known, then the sample standard deviation , *s*, can be used as follows:

$$SE = \frac{s}{\sqrt{n}}$$

If the population is finite, with the size *N*, then the standard error would be:

$$SE = \frac{s}{\sqrt{n}} . \frac{\sqrt{N-n}}{\sqrt{N-1}}$$

SE for sampling distribution of the proportion

$SE = \sqrt{\frac{p(1-p)}{n}}$ where *p*: sample proportion

If the population is finite, with the size *N*, then the standard error would be:

$$SE = \sqrt{\frac{p(1-p)}{n}} . \frac{\sqrt{N-n}}{\sqrt{N-1}}$$

Central Limit Theorem

The Central Limit Theorem states that "as the sample size increases, the sampling distribution of the sample mean approaches the normal distribution in shape, regardless of the shape of the population".

For samples taken from a non-normal population with mean μ and variance σ^2, by the Central Limit Theorem, $\bar{X}$ is approximately normal.

Point Estimate

A point estimate for a population parameter is the value of the sample statistic that is directly used to estimate the population parameter.

Example 1

A survey was conducted to determine the average amount of money spent weekly by students in a particular college. Data collected from the sample of 40 students (20 male and 20 female students) are given below:

25, 15, 30, 40, 25, 20, 25, 40, 20, 25, 50, 25, 20, 15, 50,30, 20, 20, 25, 30, 35,
40, 100, 50, 25, 20, 25, 25, 40, 100, 30, 20, 20, 25, 30, 20, 20, 25, 50, 30

The mean amount of money spent = RM32. This value is directly used to make a conclusion about the mean amount of money spent by all the students in the college. Hence, using the point estimate, we say that the mean amount of money spent by students in the college (population) is **RM32**.

Interval Estimate

An interval estimate of a population parameter consists of an interval of numbers obtained from a point estimate of the population parameter together with a percentage that specifies how confident we are that the population parameter lies in the interval.

Population parameter = Sample statistic $\pm$ *Sampling Error*
Population parameter = Sample statistic $\pm$ *confidence limit* x *standard error*

Confidence interval	95% (0.95)	99% (0.99)
α = 1 – confidence interval	0.05	0.01
Two sided test (α/2)	0.025	0.005

Estimating the population mean for a large sample

A sample is said to be a large sample if its size is equal to or greater than 30. A sample of this size, based on the central limit theorem, can be said to be normally distributed.

$$\mu = \bar{x} \pm confidence\lim it \times s\tan darderror$$

$$\mu = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

Estimating the population mean for a small sample

If the sample size, *n*, is small, i.e. *n<30*, then the t-distribution is used to estimate the population parameters.

$$\mu = \bar{x} \pm confidence\lim it \times s\tan darderror$$

$$\mu = \bar{x} \pm t_{\alpha/2, n-1} \frac{\sigma}{\sqrt{N}}$$

degrees of freedom , v = n – 1

Estimating the population proportion

The population proportion can be calculated using the following mathematical equation:

$$\pi = p \pm confidence\lim it \times satndarderror$$

$$\pi = p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Determining the sample size

By increasing the sample size, we can reduce the sampling error.

$$\text{Sampling error, } r = \text{confidence interval} \times \frac{s}{\sqrt{n}}$$

Population Mean

$$r = 1.96\frac{s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1.96s}{r}$$

$$n = \left(\frac{1.96s}{r}\right)^2$$

Population Proportion

$$r = 1.96\sqrt{\frac{p(1-p)}{n}}$$

$$r^2 = (1.96)^2 \frac{p(1-p)}{n}$$

$$n = \frac{(1.96)^2 p(1-p)}{r^2}$$

Learning Outcomes

- Students should be able to understand the concepts of point estimation, sampling distribution, interval estimation, confidence estimation.
- Students should be able to estimate the population mean based on a large sample.
- Students should be able to estimate the population mean based on a small sample.
- Students should be able to estimate the sample size.

Basic Reading

1. Saravanan Kullandavelli (1994) LCCI Business Statistics; 5th ed. Malaysia; Stamford College Group Publishing.
2. A. Francis (1995) Business Mathematics and Statistics; 4th ed. London DP Publications Ltd.

Revision Questions

1. From a random sample of 420 computer chips produced on a production line, it was found that on average there were 6 defects with a standard deviation of 1.5. What are the confidence limits for the production as a whole at:
   a. 95% confidence interval
   b. 99% confidence interval

2. With a sample size of 196 the calculated standard error is 4 with a mean of 90.
   a. What is the sample standard deviation?
   b. What are the 99% confidence limits for the population mean?
   c. If we wish to be within ±5 of the population mean, at a 95% confidence limit, what sample size should we use?

3. In order to calculate a realistic price for a repair contract, a manager recorded the time taken for a random sample of 8 recent repairs. The times (in minutes) were 54, 78, 43, 61, 92, 48, 65 and 71. Construct a 95% confidence interval for the true mean repair time.