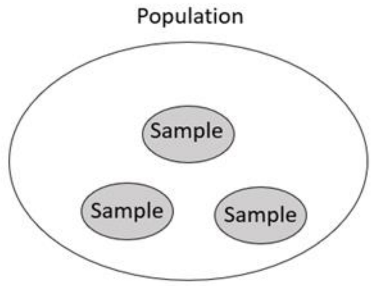
**Mathematics Methods**

Unit 4

**Sample proportion**

**1. Population, sample and sampling techniques**




*Population:* Population can be defined as the set of all eligible members of a group that is intended to be studied.

*Sample:* Sample can be defined as the subset of the population in a smaller and manageable size.

Why a sample is selected rather than dealing with population?

- Size of population is too large
- Difficulty in accessing the population
- Data collection consume a huge amount of time, near impossible to encompass the entire population

There are many ways to select sampling which falls under probabilistic and non-probabilistic sampling. However, the general principle of sampling selection is that the sampling method should not be bias (favour and disfavour) any subgroup of population. One of the ways is random sampling (probabilistic).

Random sampling is the sampling process whereby each individual of the subset has an equal probability of being chosen.

Ways to sample randomly:

- Use random number generator (without repetition of any numbers) to generate random integer
- Use lottery method by writing numbers (eg: 1 to 10) on individual papers and draw the numbers randomly.

**2. Population and sample proportion**

**(a) Population proportion**

Definition: Population proportion is a parameter that describes a percentage value associated with a population.

Formula:

$$p = \frac{X}{N}$$

Example 1:
There are a total of 30,000 people living in a particular village. A researcher found out that out of the 30,000 people, 500 people possess a certain unique DNA. Find the population proportion.

$$p = \frac{X}{N}$$
$$= \frac{500}{30,000}$$
$$= \frac{1}{60}$$

Example 2:
A lucky spinner divided into four sections was spun. What is the population proportion of a section is obtained through spinning the lucky spinner.

$$p = \frac{X}{N}$$
$$= \frac{1}{4}$$
$$= 0.25$$

Example 3:
In a study, it is identified that 70% of the nation owns a mobile phone. Find the population proportion.

$$p = 70\%$$
$$= 0.7$$

**(b) Sample proportion**

Definition: Sample proportion is the proportion of individuals in a sample sharing a certain trait.

Formula:

$$\hat{p} = \frac{x}{n}$$

Assumption: samples are selected randomly

Example 1:
In a sample of 120 students, it is found that 99 of them passed the recent Mathematics Methods examination. Find the sample proportion.

$$\hat{p} = \frac{x}{n}$$
$$= \frac{99}{120}$$
$$= 0.825$$

Example 2:

A lucky spinner shown below was spun 100 times, the occurrence that the pointer lands on the grand prize is 12 occasions. Find the sample proportion of the pointer landing on a grand prize for the sample spin of 100 times.

$$\hat{p} = \frac{x}{n}$$
$$= \frac{12}{100}$$
$$= 0.12$$


**(i)** **Sampling distribution of sample proportion**

Definition: Sampling distribution of sampling proportion is the distribution of a statistic which is calculated from a sample.

**(ii)** **Sampling distribution of the sample proportion, $\hat{p}$ when population, $n$ is small**

Probabilities associated with the possible values of the sample proportion can be calculated by:

- direct consideration of sample outcomes
- using knowledge of selections

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

**(iii)** **Sampling distribution of the sample proportion, $\hat{p}$ when population, $n$ is large**

Probabilities associated with the possible values of the sample proportion:

- is assumed to remain constant
- can be calculated by applying binomial distributions

  Recall:

$$X \sim Bin(n, p)$$
$$\mu = np$$
$$\sigma^2 = np(1-p)$$

Mean,

$$E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{1}{n}(np) = p$$

Variance,

$$Var(\hat{p}) = Var\left(\frac{x}{n}\right) = \left(\frac{1}{n}\right)^2 Var(x) = \frac{1}{n^2}[np(1-p)] = \frac{p(1-p)}{n}$$

Standard deviation,

$$Std(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

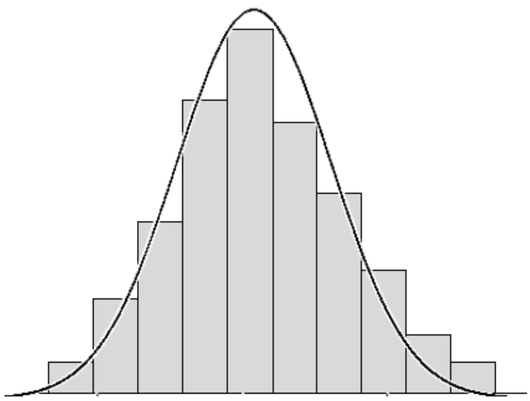
**3. Central limit theorem**

Definition:
Central limit theorem (CLT) established that when independent random variables are added, their normalised sum tends towards a normal distribution.

Conditions for sampling distribution to be approximately large:

- $n > 30$
- $np > 5$
- $nq > 5$

*(the conditions stated above varies and are not definite)*



$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

Standard score,

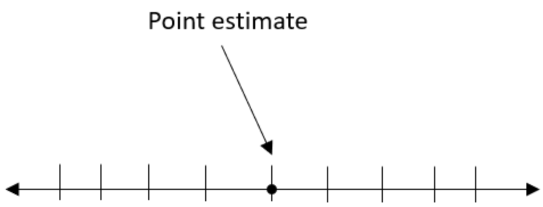
$$Z = \frac{X - \mu}{\sigma}$$

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

**4. Estimates of proportion**

**(a) Point estimate**

Definition: Point estimate is single value estimate where sample proportion $\hat{p}$ can be used to estimate the population proportion $p$.

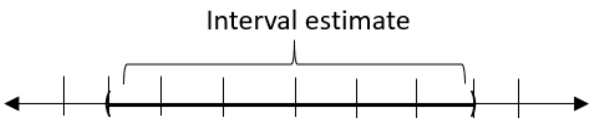



Example:

- Probability of selecting a black marble is 0.6
- Average income of a particular residency is thought to be about $4,000 per month

**(b) Interval estimate**

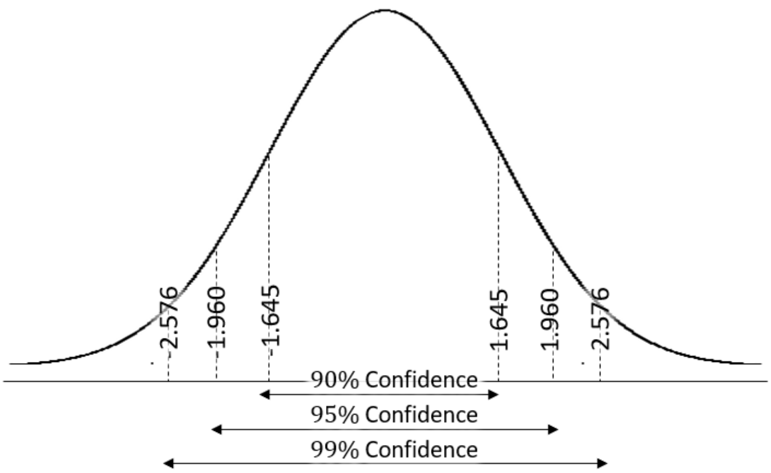
Definition: Interval estimate is an interval estimation within which the value of a parameter of a population proportion has a stated probability of occurring. (known as the "confidence level for population proportion $p$)




*Q: How confident do we want to be that the interval estimate contains the population proportion?*

Example:
- Mean daily expenditure of moderate-income household is between $100 to $250
- Average length of a pen is between 12 cm to 17 cm

Confidence interval




| Percentage % | Number of standard deviation/ $z$ −score |
|---|---|
| 90% | 1.645 |
| 95% | 1.960 |
| 99% | 2.576 |

**take $z$-score to 3 decimal places*

Formula:

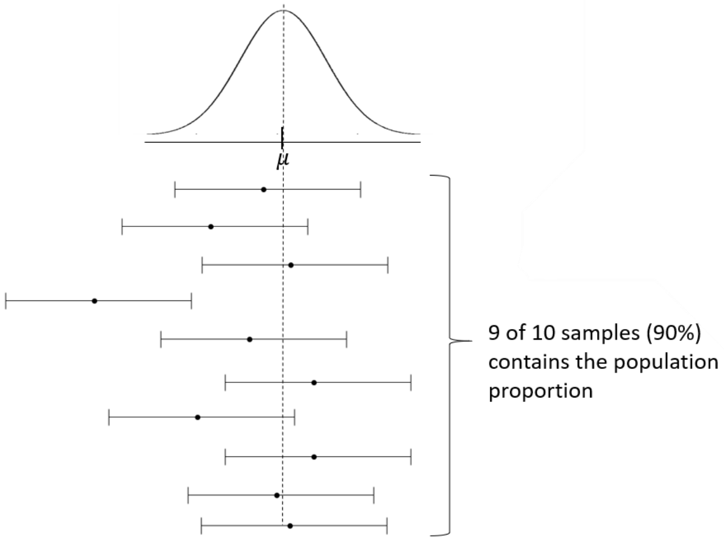
$$\hat{p} \pm z\sigma$$

Represented as:

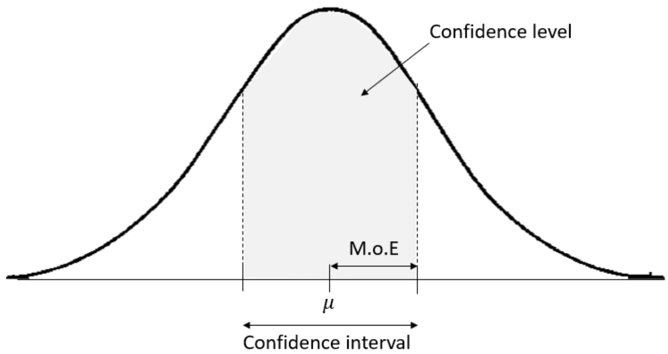
___% confidence interval $(\hat{p} - z\sigma, \hat{p} + z\sigma)$

Interpretation
We could expect that ___% of the ___% confidence interval contains the true/population proportion

Example:
90% confidence level for 10 samples.




<u>Confidence level, confidence interval and margin of error</u>




*M.o.E: Margin of error,* $E = z\sigma$

**(i) Finding confidence level from margin of error**

Example:
What is the confidence level corresponding to the margin of error of 2.23 given that it is based on a standard normal variable?

Normcdf(-2.23, 2.23, 0, 1) = 0.974253
Margin of error of 2.23 corresponds a confidence level of 0.974253,
Confidence level (in %) = 97.4%

**(ii) Finding confidence level from confidence interval**

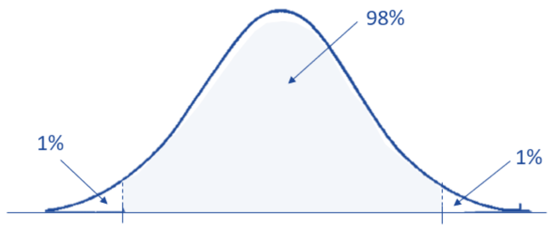
Example:
What is the confidence level corresponding to the confidence interval of $-3.2 < z < 3.2$?

Normcdf(-3.2, 3.2, 0, 1) = 0.998626
Confidence interval of $-3.2 < z < 3.2$ corresponds a confidence level of 0.998626
∴ Confidence level (in %) = 99.9%

**(iii)** **Finding margin of error from confidence level**

Example:
Given confidence level of 98%, find the margin of error in standard normal variable.

Graphical representation:




InvNorm (0.99,0,1) = 2.32635
∴ Margin of error = 2.32635

**(iv)** **Finding margin of error given sample proportion and confidence interval**

Example:
Find the margin of error given $\hat{p} = 0.341$ and confidence interval $0.096 < p < 0.586$

Margin of error = 0.341 – 0.096
= 0.245

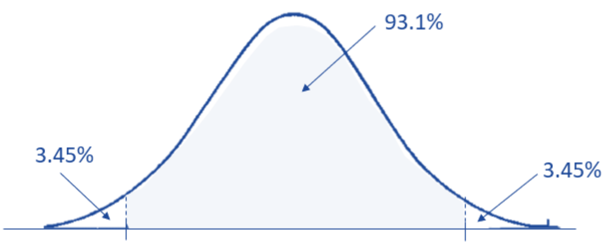
*or*

Margin of error = 0.586 – 0.341
= 0.245

**(v)** **Finding confidence intervals from confidence level**

Example:
Find the confidence interval in standard normal variables given a confidence level of 93.1%.

Graphical representation:




InvNorm (0.9655,0,1) = 1.81842
∴ 93.1% confidence interval (-1.81842, 1.81842)

**END**