Estimation and Testing of Hypothesis

Introduction: Sometimes the area of investigation is very large. It becomes difficult to study the entire population under investigation. In such case a smaller group called a 'sample' is taken from the population. The necessary data is obtained from the sample and then the results are generalized to the entire population.

Sampling distribution of sample mean:

Consider a bag containing three tickets with 1, 3 and 5. Suppose we draw two tickets with replacements one after the other from them. We see that sample mean ($\bar{x}$) changes as the sample changes and thus, we can consider sample mean as the random variable. The mean and standard deviation of sample mean are calculated as follows.

Mean of $\bar{x} = \frac{\sum \bar{x}}{n}$ and

S.D. of $\bar{x} = \sqrt{\frac{\sum(\bar{x})^2}{n} - \left(\frac{\sum \bar{x}}{n}\right)^2}$

Let us take the sample for the above data and find the mean and S.D.

Sample no.	No. on the ticket in the sample	Sample mean	$(\bar{x})^2$
1	1 , 1	1	1
2	1 , 3	2	4
3	1 , 5	3	9
4	3 , 1	2	4
5	3 , 3	3	9
6	3 , 5	4	16
7	5 , 1	3	9
8	5 , 3	4	16
9	5 , 5	5	25
		$\sum \bar{x} = 27$	$\sum(\bar{x})^2 = 93$

Thus, for the sample,

Mean of $\bar{x} = \frac{\sum \bar{x}}{n} = \frac{27}{9} = 3$

S.D. of $\bar{x} = \sqrt{\frac{\sum(\bar{x})^2}{n} - \left(\frac{\sum \bar{x}}{n}\right)^2} = \sqrt{\frac{93}{9} - \left(\frac{27}{9}\right)^2} = \sqrt{\frac{93}{9} - (3)^2} \sqrt{\frac{31}{3} - 9} = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}}$

Next, consider the population 1, 3 and 5 and let us find the mean and S. D.

We have, n = 3, $\sum x = 9$ and $\sum(x)^2 = 35$

Population mean = = $\frac{\sum x}{n} = \frac{9}{3} = 3$ and

S.D. of Population = $\sqrt{\sum \frac{(x)^2}{n} - \left(\sum \frac{n}{n}\right)^2}$

$$= \sqrt{\frac{35}{3} - (3)^2}$$
$$= \sqrt{\frac{35}{3} - 9}$$
$$= \sqrt{\frac{35 - 27}{3}}$$
$$= \sqrt{\frac{8}{3}}$$
$$= \sqrt{\frac{4 * 2}{3}}$$
$$= 2 * \sqrt{\frac{2}{3}}$$

Note that men of $\bar{x} = 3$ = Men of the population.

S.D. of sample, $\bar{x} = \frac{2}{\sqrt{3}} = \frac{1}{\sqrt{2}} * \frac{2 * \sqrt{2}}{\sqrt{3}} = \frac{S.D. of \ Population}{\sqrt{2}} = \frac{S.D. of \ Population}{\sqrt{sample \ \ size}}$

Thus, in general,

For population mean = μ
Population S. D. = σ and
Sample size, = n , we have,

1. Mean of sample mean ($\bar{x}$) = μ
2. S.D. of ($\bar{x}$) = $\frac{\sigma}{\sqrt{n}}$

If the population values follows Normal distribution with mean μ and S.D. = σ then distribution of $\bar{x}$ IS ALSO Normal with mean μ and SD = $\frac{\sigma}{\sqrt{n}}$

<u>Note</u>:

1. For a distribution with very large population and sample size $n \geq 30$, the distribution is considered as a Normal distribution by a "Central Limit theorem".
2. If we considered mean distribution as a Normal distribution than, s.n.v. z can be written as,

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

<u>Definition</u>:

The distribution of $\bar{x}$ is known as it sampling distribution.
The standard deviation of $\bar{x}$ is known as its '**Standard Error**'.

Note: We solve the following assuming that the $\bar{x}$ behave as normal.

Ex.1. A large population has a mean height of 150 cms and a std deviation of 20 cm. A random sample of size 100 is taken from this population. Find the probability that the sample mean will

(i) Exceed 151 cms

(ii) lie between 148 cm and 155 cm.

Assume that the population of height is normal and that the sampling is with replacement.

(Given for s.n.v. Z area between (i) Z = 0 to Z = 0.5 is 0.1915

(ii) Z = 0 to Z = 1 is 0.3413

(iii) Z = 0 to Z = 2.5 is 0.4938)

Soln: Let $\bar{x}$ be the variate that follows normal distribution.

We have,

$\mu = 150$

$\sigma = 20$

$n = 100$

For normal distribution, s.n.v. z is given by,

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 1570}{\frac{20}{\sqrt{100}}} = \frac{\bar{x} - 150}{2}$$

<u>To find</u>: $P(\bar{x}\ exceeds\ 151)$ i.e. P($\bar{x}$ > 151)

Now , when $\bar{x}$ = 151, s.n.v. Z is given by,

$$z = \frac{151 - 150}{2}$$

$$= \frac{1}{2}$$

$$= 0.5$$

$\therefore P(\bar{x} > 151) = P(z > 0.5)$

= Area to the right of z = 0.5

= 0.5 - Area(bet z = 0 to z = 0.5)

= 0.5 – 0.1915

= 0.3085

Next,

<u>To find</u>: $P(148 < \bar{x} < 155)$

Now , when $\bar{x}$ = 148, s.n.v. Z is given by,

$$z = \frac{148 - 150}{2}$$

$$= \frac{-2}{2}$$

$$= -1$$

Next, when $\bar{x}$ = 155, s.n.v. Z is given by,

$$z = \frac{155 - 150}{2}$$

$$= \frac{5}{2}$$

$= -2.5$

$\therefore P(148 < \bar{x} < 155) = P(-1 < z < 2.5)$

= Area to the left of z = - 1 to z = 2.5

= Area (z= -1 to z = 0) + Area (z = 0 to z = 2.5)

= Area (z = 0 to z = 1) + Area (z = 0 to z = 2.5)

= 0.3413 + 0.4938

= 0.8351

Ex.2. The mean life of a large set of fluorescent tubes is 1570 hour with a standard deviation of 150 hours. A sample of 100 tubes is drawn from it with replacement. Find the probabilities that the mean life of these tubes will-

(i) exceed 1600 hrs.

(ii) not exceed 1540 hrs.

(Given: For s.n.v. z, area between (i) z = 0 to z = 2 is 0.4772

(ii) z = 0 to z = 1.33 is 0.482)

Soln: Let $\bar{x}$ be the variate that follows normal distribution.

We have,

$\mu = 1570$

$\sigma = 150$

For normal distribution, s.n.v. z is given by,

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 1570}{\frac{150}{\sqrt{100}}} = \frac{\bar{x} - 1570}{15}$$

<u>To find</u>: $P(\bar{x} > 1600)$

Now , when $\bar{x}$ = 1600, s.n.v. Z is given by,

$$z = \frac{1600 - 1570}{15}$$

$$= \frac{30}{15}$$

$$= 2$$

$\therefore P(\bar{x} > 1600) = P(z > 2)$

= Area to the right of z = 2

= 0.5 - Area(bet z = 0 to z = 2)

= 0.5 – 0.4772

= 0.0228

Next,

<u>To find</u>: $P(\bar{x} < 1540)$

Now , when $\bar{x}$ = 1540, s.n.v. Z is given by,

$$z = \frac{1540 - 1570}{15}$$

$$= \frac{-30}{15}$$

$$= -2$$

$$\therefore P(\bar{x} < 1540) = P(z < -2)$$

= Area to the left of z = -2

= Area to the right of z = 2

= 0.5 - Area(bet z = 0 to z = 2)

= 0.5 – 0.4772

= 0.0228

Sampling Distribution of Sample Proportion:

Suppose 15 % of the adults staying in a city wear glasses. A sample of 100 adults is taken and 14 % of them were found wearing glasses. Than the proportion of adults who wear glasses in the city is 15/100. It is called "Population Proportion" (of adults who wear glasses) and is denoted by P.

Similarly the proportion of adults wearing glasses from the sample is 14/100. It is known as "sample proportion" and is denoted by p.

Observe that, as sample varies, the sample proportion p varies and therefore p can be considered to be a random variable.

Result: Let,

P = Population proportion
p = Sample proportion
n = Sample size
than

1. Mean of $p = P$
2. Std. deviation of $p = \sqrt{\frac{P\,Q}{n}}$ where $Q = 1 - P$

Note:

1. Sample proportion (p) in general follows Binomial Distribution
2. For a large value of n it can be approximated by Normal Distribution
3. The distribution of p is called the "Sampling distribution" and the std deviation of p is called the "**Standard Error**".
4. Standard Normal Variate (s.n.v.) for p is given by,

$$z = \frac{p-P}{\sqrt{\frac{P\,Q}{n}}}$$

Ex.1. 20 % of a large consignment of apples are found to be bad. Find the probability that at least 25% apples are bad in a sample of size 400 drawn from it. (Given: for s.n.v. z, area between (i) z = 0 to z = 2.5 is 0.4938)

Soln: Given that, for bad apples,

$P = 0.2$

$\therefore Q = 1 - P$

$= 0.8$

$n = 400$

Note that the sample size is large and thus, we assume that the sample proportion of p follows Normal distribution.

Note that, the mean of p = P = 0.2

$\therefore$ S.D. of P = Standard Error, $S.E. = \sqrt{\frac{P * Q}{n}} = \sqrt{\frac{0.2 * 0.8}{400}} = \sqrt{\frac{0.16}{400}} = 0.02$

Now,

When p = 25% = 0.25

Therefore, s.n.v. z is given by,

$$z = \frac{p - P}{\sqrt{\frac{P * Q}{n}}}$$

$$= \frac{0.25 - 0.2}{0.02}$$

$$= \frac{0.05}{0.02}$$

$$= 2.5$$

∴ Prob(p>0.25) = P(z > 2.5)

= Area to the right of 2.5

= 0.5 – 0.4938

= 0.0062

<u>Confidence interval and confidence limits:</u>

<u>For large proportion:</u>

Let a sample size n be drawn from a large population with mean μ and standard deviation σ. Let $\bar{x}$ denote the sample mean of the sample drawn. Than s.n.v. z is given by ,

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

We know that area under the s.n.v. between $z = \pm 1.96$ is 0.95

In other words,

Prob (-1.96 < z < 1.96) = 0.95

i.e. $P\left(-1.96 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right) = 0.95$

i.e. $P\left(-1.96 * \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < 1.96 * \frac{\sigma}{\sqrt{n}}\right) = 0.95$

i.e. $P\left(-1.96 * \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < 1.96 * \frac{\sigma}{\sqrt{n}}\right) = 0.95$

i.e. $P\left(-\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right) = 0.95$

i.e. $P\left(\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right) = 0.95$ (a<x<b then –a>-x>-b)

Hence between $\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x}, +1.96 * \frac{\sigma}{\sqrt{n}}$, we find middle 95 % of the distribution.

Here, $\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}$ are called 95% Confidence limit for the population mean (μ) and $\left(\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} , \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right)$ is called 95% Confidence interval.

On the similar lines we can calculate the 99 % confidence limit and confidence interval interval. The procedure can be repeated for population proportion. The different limits and intervals are written below in the tabular form as follows:

1) For Population mean, $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

l.o.s (α)	Confidence limits	Confidence interval
5 % (0.05)	$\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}$	$\left(\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} , \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right)$
1 % (0.01)	$\bar{x} - 2.58 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 2.58 * \frac{\sigma}{\sqrt{n}}$	$\left(\bar{x} - 2.58 * \frac{\sigma}{\sqrt{n}} , \bar{x} + 2.58 * \frac{\sigma}{\sqrt{n}}\right)$
Certainity	$\bar{x} - 3 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 3 * \frac{\sigma}{\sqrt{n}}$	$\left(\bar{x} - 3 * \frac{\sigma}{\sqrt{n}} , \bar{x} + 3 * \frac{\sigma}{\sqrt{n}}\right)$

2) For Population mean, $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

l.o.s (α)	Confidence limits	Confidence interval
5 % (0.05)	$p - 1.96 * \sqrt{\frac{PQ}{n}}$ and $p + 1.96 * \sqrt{\frac{PQ}{n}}$	$\left(p - 1.96 * \sqrt{\frac{PQ}{n}}\ ,\ p + 1.96 * \sqrt{\frac{PQ}{n}}\right)$
1 % (0.01)	$p - 2.58 * \sqrt{\frac{PQ}{n}}$ and $p + 2.58 * \sqrt{\frac{PQ}{n}}$	$\left(p - 2.58 * \sqrt{\frac{PQ}{n}}\ ,\ p + 2.58 * \sqrt{\frac{PQ}{n}}\right)$
Certainity	$p - 3 * \sqrt{\frac{PQ}{n}}$ and $p + 3 * \sqrt{\frac{PQ}{n}}$	$\left(p - 3 * \sqrt{\frac{PQ}{n}}\ ,\ p + 3 * \sqrt{\frac{PQ}{n}}\right)$

3) For some problems on sample mean, Population mean μ may be given. We need to take this μ as $\bar{x}$

4) For some problems on sample proportion, Population proportion P should be taken as sample proportion p. (i.e. $P = p$)

Examples: (problems based on Population mean)

Estimation of point: (Population Mean)

Ex.1. A survey of 36 married people yield a mean age of at the time of their marriage as 26 years with a S.D. of 2.4 years. Find 95 5 Confidence limit for the age at the time of marriage.

Soln: Given $n = 36$, $\sigma = 2.4$ years, $\bar{x} = 26$ years and $\alpha = 5$ %

Now, Confidence limits for population mean at 95 % are given by,

$\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}$

$26 - 1.96 * \frac{2.4}{\sqrt{36}}$ and $26 + 1.96 * \frac{2.4}{\sqrt{36}}$

$26 - 1.96 * \frac{2.4}{6}$ and $26 + 1.96 * \frac{2.4}{6}$

$26 - 1.96 * 0.4$ and $26 + 1.96 * 0.4$

$26 - 0.784$ and $26 + 0.784$

25.216 and 26.784

Thus the 95 % confidence limits are 25.216 and 26.784

Ex.2. From a random sample of 64 plastic containers manufactured by a process, the average weight is found to be 50 gms. if the standard deviation of the weight of the containers produced by the machine is 2 gms. Find the limits of weights within which the average weight of the containers manufactured by the process lie with 95 % confidence.

Soln: We proceed as follows.

Given $n = 64$,

$\sigma = 2$ gms,

$\bar{x} = 50$ gms and $\alpha = 5$ %

Now, Confidence limits for population mean at 95 % are given by,

$\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}$

$50 - 1.96 * \frac{2}{\sqrt{64}}$ and $50 + 1.96 * \frac{2}{\sqrt{64}}$

$50 - 1.96 * \frac{2}{8}$ and $50 + 1.96 * \frac{2}{8}$

$50 - 0.49$ and $50 + 0.49$

49.51 and 50.49

Thus the 95 % confidence limits are 49.51 and 50.49

Estimation of interval: (Population Mean)

Ex.1. A sample of 50 bulbs from a large consignments showed a mean life of 52 hours with a standard deviation 4 hours. Find the confidence interval within which the mean life of the bulbs lie almost certainly.

Soln: We Proceeds as follows:

Given that:

$n = 50$

$\bar{x} = 52\ (\mu)$

$\sigma = 4$

To find : Confidence interval for certainty.

Confidence interval for certainty is given by,

$\left(\bar{x} - 3 * \frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + 3 * \frac{\sigma}{\sqrt{n}}\right)$

$= \left(52 - 3 * \frac{4}{\sqrt{50}}\ ,\ 52 + 3 * \frac{4}{\sqrt{50}}\right)$

$= \left(52 - 3 * \frac{4}{7.07}\ ,\ 52 + 3 * \frac{4}{7.07}\right)$

$= (52 - 1.7\ ,\ 52 + 1.7)$

$= (50.3\ ,\ 53.7)$

Thus the Confidence interval for certainty is $(50.3\ ,\ 53.7)$

Exercise:

1. In a study of television viewing habits, in order to obtained an interval estimates of the average number of hours per week that teenager spend watching television programmes, a random sample of 100 teenage children is taken. The sample investigation revealed that a mean of 9.2 hours with S.D. of 3.2 hours. Obtained the desired interval of estimates with confidence coefficient of 0.99. (8.3744, 10.0256)
2. For a given sample of 200 items drawn from a large population, the mean is 65 and the S.D. is 8. Find the 95 % Confidence limits for the population mean. (63.8913 and 66.1087)
3. A hospital is collecting data regarding the number of days spent by a patient in the hospital for typhoid. A sample shows the following:

No. of days spent (x)	3	4	5	6	7	8	9
No. of patients (y)	8	10	15	20	20	15	12

Find the limits within which the mean number of days required to be spent in the hospital by a patient for typhoid lies almost certainty. (5.745 and 6.795)

Estimation of Point: (Sample proportion)

Ex.1. A random sample of 100 balls slected from a large consignment of tennis balls gave 10 % bad balls. Find 99 % confidence limits for the percentage of bad balls in the consignment.

Soln: We Proceeds as follows:

To find: the 99% confidence limits to the percentage of persons who would buy the product.

Given that:

$n = 100$

$p = 10\% = 0.1$

$P = p = 0.1$

$\Longrightarrow$ $Q = 1 - P = 0.9$ and $\alpha = 1\ \%$

99% Confidence limits are given by,

$p - 2.57 * \sqrt{\frac{PQ}{n}}$ and $p - 2.58 * \sqrt{\frac{PQ}{n}}$

$0.1 - 2.58 * \sqrt{\frac{0.1*0.9}{100}}$ and $0.1 + 2.58 * \sqrt{\frac{0.1*0.9}{100}}$

$0.1 - 2.58 * 0.03$ and $0.1 + 2.58 * 0.03$

$0.1 - 0.0744$ and $0.1 + 0.0744$

0.0226 and 0.1744

i.e. 0.0226 and 0.1744

Thus, 99% confidence limits to the percentage of persons are:

2.26 and 17.44

Exercise:

1. In a sample of 1000 T.V. viewer, 340 watch a particular programme. Find 99 % confidence limits for the percentage of all viewers who watch the programme. (30.135% and 37.865%)
2. A department store wants to determine the percentage of shoppers who leaves only after having actually bought at least one item. A random sample of 900 shoppers leaving the store, showed that 750 had brought something ranging from a couple of saop bars to a complete bedroom furniture set. What is 99% confidence interval for the true percentage of buyers? (82%, 84%)
3. A sample of 900 days is taken from a meterological records of a certain districts and 100 of them are found to be foggy. Determine the probable limits for the percentage of foggy days in the district. (7.89% and 14.135).

Testing of Hypothesis:

Hypothesis: Hypothesis is the statement which is believe to be true and an argument generally is based upon.

Null Hypothesis and Alternative Hypothesis:

Null Hypothesis: Null Hypothesis is a assumption which is believe to be true. It is denoted by H_0.

Alternative Hypothesis: A hypothesis which is considered as true in the absence of null hypothesis is called alternative hypothesis. It is denoted by H_1.

Note:

1. Generally, a statistical hypothesis is referred to as a Null hypothesis.
   Example: $H_0 = \mu = Rs. 5000/-$.
2. Hypothesis other than Null hypothesis is called as Alternative hypothesis.
   Example: $H_1 = \mu \neq Rs. 5000/-$
3. Alternative hypothesis is the assumption which can be considered to be the alternative to the null hypothesis.
4. If the Null Hypothesis is not true then what is true is called as Alternative hypothesis.

Testing of Hypothesis and error in decision making:

Testing of Hypothesis:

Testing of hypothesis is the process of accepting or rejecting a null hypothesis based on a chosen sample statistics.Any statistical hypothesis is based on the information supplied by the sample data. The procedure of testing such hypothesis does not guarantee that all decisions are accurate. There is always a chance of some error. There are two types of errors.

Type I error: The process of rejecting a true hypothesis is called type I error

Type II error: The process of accepting a false hypothesis is called type II error.

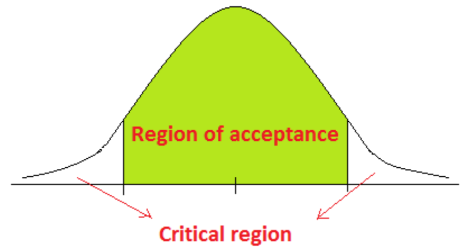
The above error can be represented in a tabular form as follows:

		Null Hypothesis H_0 is	
		True	Not true
Decision	Reject H_0	Type I error	No error
	Do not Reject H_0	No error	Type II error

Note:

1. Type I error is also called as α error.
2. Type II error is also called as β eoor.

Critical region:
The area of the normal curve corresponding to the type I error, is Critical region (also known as Region of rejection)
Critical region is the region which corresponds to the test statistics for which the hypothesis H_0 is to be rejected if the sample point falls in the region. In other words critical region is the region of rejection of null hypothesis.



Acceptance region:
The region that falls outside the region of critical region is called 'Region of acceptance'
Level of significance:
Level of significance is define as the maximum probability of, with which we would be willing to risk a Type I error.
[Alternatively, it is also defined as a test designed so that P(Type I Error) $\leq \propto$, then $\propto$ is called level of significance.] It is denoted by α.
Since $\propto$ is the probability value, it is always less than 1. Generally we take α between1% to 5%. By default we take $\propto$ as 5% (0.05).

Steps involved in testing of hypothesis:
1. State H_0 and H_1
2. Specify the level of significance.
   (If $l.o.s.$ is not mentioned then we take it as 5 %)
3. Mentioned the decision criterion
4. Compute the value of test statistics.
5. Draw the conclusion

Decision criteria based on H_1 and :
Rejection of H_1 can be easily done from the following table. The following table takes into consideration for α = 1% and α = 5 %.

Alternative Hypothesis H_1	Decision criterion: Reject H_0 iff		
	$l.o.s. = 0.05$ $(i.e., \alpha = 5\%)$	$l.o.s. = 0.01$ $(i.e., \alpha = 1\%)$	$l.o.s. = \alpha$
$>$	$z > 1.64$	$z > 2.33$	$z > z_\alpha$
$<$	$z < -1.64$	$z < -2.33$	$z < z_\alpha$
$\neq$	$z < -1.96$ *or* $z > 1.96$	$z < -2.58$ *or* $z > 2.58$	$z > z_\alpha$ or $z > z_\alpha$

Note:

1) Decision criterion is based on:
    i) H_1
    ii) $l.o.s.\ \alpha$ and
    iii) Large Sample size n.
2) However, for a large samples, it will depend on H_1 and α.
3) Here, z is the computed value of test statistics based on selected sample /proportion.

Problems based on Sample mean:

Examples:

Ex. 1. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is found to be 1570 hours with a standard deviation of 120 hours. Test the hypothesis that the mean lifetime of a bulbs produced by the company is 1600 hours against the alternative hypothesis that it is less than 1600 hours at 5% l.o.s.

Soln: We proceed as follows:

Step 1.: H_0 = the mean life of fluorescent bulbs is 1600 hours.
H_1 = the mean life of fluorescent bulbs is less than 1600 hours.
(i.e. $H_0 : \mu = 1600$ hours
$H_1 : \mu < 1600$ hours)

Step 2.: $\alpha = 0.05$ (i.e. 5 %)

Step 3.: Based on H_1 and α and large value of n = 100, we take decision:
Reject H_0 iff $z < -1.64$

Step 4.: For a given sample, given that,
$\bar{x} = 1570$
$\sigma = 120$
$n = 100$
And for population, population mean, $\mu = 1600$

Test statistics z is given by,

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$
$$= \frac{1570 - 1600}{\frac{120}{\sqrt{100}}}$$
$$= \frac{-30}{\frac{120}{10}}$$
$$= \frac{-30}{12}$$
$$= -2.5$$

Step 5: Observe that the computed value of z (i.e. z = - 2.5 < -1.64.
Thus we reject H_0.

Ex.2. An ambulance service claims that it takes on an average 8.9 minutes for an ambulance to reach its destination in emergency calls. To check on this claim, the agency which licenses ambulance service has timed them on 50 emergency calls, getting a mean of 9.3 minutes with a standard deviation of 1.6 minutes. What can they conclude at 5 % l.o.s.

Soln: We proceed as follows:

Step 1.: H_0 = An ambulance can reached its destination in 8.9 minutes.
H_1 = An ambulance cannot reached its destination in 9.8 minutes.
(i.e. $H_0 : \mu = 8.9$ min
$H_1 : \mu \neq 8.9$ min)

Step 2.: $\alpha = 0.05$ (i.e. 5 %)

Step 3.: Based on H_1 and α and large value of n = 50, we take decision:
Reject H_0 iff $z < -1.96$ or $z > 1.96$

Step 4.: For a given sample, given that,
$\bar{x} = 9.3$
$\sigma = 1.6$
$n = 50$
And for population, population mean, $\mu = 8.9$

Test statistics z is given by,

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
$$= \frac{9.3 - 8.9}{\frac{1.6}{\sqrt{50}}}$$
$$= 1.77$$

Step 5: Observe that the computed value of z (i.e. z = 1.77 not > 1.96.
Thus we do not reject H_0.

Testing of hypothesis:(Sample proportion)

Ex.1. A coin is tossed at random 400 times and head turns up 240 times. Can the coin be regarded unbiased? Use 5 % level of significance.
A certain coin showed up head on 240 occasions in 400 tosses. Test the claim that the coin is unbiased.

Soln: We Proceeds as follows:

Step1: H_0 = Selected coin is unbiased.
(The head or tail occurs in the ratio1:1
i.e. $P = \frac{1}{2}$)
H_1 = Selected coin is biased i.e. $P \neq \frac{1}{2}$.

Step 2: Level of significance, $\alpha = 0.05 = 5\%$

Step 3: Decision criteria: Based on H_1 and α (and a large value of n= 400) we arrive at a decision criterion, reject H_0:
if $z < -1.96$ or $z > 1.96$

Step 4: Given that
$$P = \frac{1}{2} = 0.5$$
$$\Rightarrow Q = 1 - P = \frac{1}{2} = 0.5$$
n = 400
$$p = \frac{240}{400} = 0.6$$

Test statistics z is given by,

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.6 - 0.5}{\sqrt{\frac{0.5*0.5}{400}}}$$

$$= \frac{0.1}{\frac{0.5}{20}}$$

$$= \frac{0.1*20}{0.5}$$

$$= 4$$

Step 5: Observe that computed value of z = 4 >1.96.

Thus we reject H_0.

Ex.2. In a sample of 400 residents of a locality, 232 are men. Test the hypothesis that sex ratio is 1:1 at 1 % l.o.s.

Soln: We Proceeds as follows:

Step1: H_0 = Sex ratio is 1:1.

(The man and woman occurs in the ratio1:1 i.e. $P = \frac{1}{2}$)

H_1 = Sex ratio is not 1:1 i.e. $P \neq \frac{1}{2}$.

Step 2: Level of significance, $\alpha = 0.01 = 5\%$

Step 3: Decision criteria: Based on H_1 and α (and a large value of n= 400) we arrive at a decision criterion, reject H_0:

if z < - 2.58 or z > 2.58

Step 4: Given that

$$P = \frac{1}{2} = 0.5 \text{ (P = proportion of men in the population)}$$

$$\Longrightarrow Q = 1 - P = \frac{1}{2} = 0.5$$

n = 400

$$p = \frac{232}{400} = 0.58 \text{ (p = proportion of men in the population)}$$

Test statistics z is given by,

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.58 - 0.5}{\sqrt{\frac{0.5*0.5}{400}}}$$

$$= \frac{0.08}{\frac{0.5}{20}}$$

$$= \frac{0.08*20}{0.5}$$

$$= 3.2$$

Step 5: Observe that computed value of z = 3.2 > 2.58.

Thus we reject H_0.

References:

1. *Gupta S.P. , Statistical Methods, Sultan Chand and sons.*
2. *Gupta C.B., Fundamentals of Statistics, Himalaya Publishing House*
3. *Shah R. J., Statistical Methods.*