Testing of hypothesis

Basic definition:

Population: A population is used to refer any collection of individual it may be finite or infinite.

Sample: A sample is a small proportion selected from the population and the process of drawing a sample from a population is called sampling.

Sample size: The number of individual in a selected sample is called the sample size.

Parameter and statistics: Any statistical method computed from population data is known as parameter and any statistical method computed from sample data is known as statistics.

Notations:

Measure	Population	Sample
size	N	n
Mean	μ	$\bar{x}$
Standard deviation	σ	s
Proportion	P	P'
variance	σ^2	s^2

Sampling distributions: The various value of statistics so obtained may be arranged as a frequency distribution which is known as sampling distributions.

Standard error: The standard deviation of sampling distribution of a statistic is known as its standard error, abbreviated as S.E. (ie average amount of variability from the observation of a sampling distribution).

Statistical Hypothesis: In attempting to reach decision about population on the basis of sample observations, we make assumptions about population, which are not necessarily true are called statistical hypothesis.

Null hypothesis: Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true and is denoted by H_0.

Alternative hypothesis: A hypothesis that is complementary to null hypothesis is called alternative hypothesis and is denoted by H_1.

A procedure for deciding whether to accept or reject the null hypothesis is called the test of hypothesis.

Level of Significance: It is the probability level below which the null hypothesis is rejected. Generally 5% and 1% level of significance are used.

Critical region (or) Region of rejection: The critical region of a test of statistical hypothesis is that region which leads to the rejection of null hypothesis H_0. Those region which leads to the acceptance of H_0 is called acceptance region.

Error in Sampling: Errors are Type I, Type - II errors.

Type I error: Reject H_0 when it is true

Type II error: Accept H_0 when it is false.

$P(\text{Type I error}) = \alpha$ and $P(\text{Type II error}) = \beta$.

One tail & Two tail test: If μ_0 is population parameter & μ is the sample statistics, then the null hypothesis is given by $H_0: \mu = \mu_0$

Alternative hypothesis is given by,

$H_1 : \mu \neq \mu_0$ (two-tailed)

$H_1 : \mu > \mu_0$ (Right tailed) (One-tail test)

$H_1 : \mu < \mu_0$ (Left tailed) (one-tail test)

Procedure for testing of hypothesis:

1. Formulate H_0 and H_1
2. Choose the level of significance α
3. Compute the test statistic using data available.
4. Pick out the critical value from the tabulation.
5. Conclusion: Compare the computed value of the test statistic with the critical value at the ~~origin~~ given level of significance.

Sampling

- Large samples ($n \geq 30$)
  - Test for single mean
  - Test for difference of mean
  - Test for single proportion
  - Test for difference of proportion
  - Test for standard mean
- Small samples ($n < 30$)
  - t-Test
    - single mean
    - Difference of mean
  - F-test
  - χ^2-test

Large samples ($n \geq 30$):

Critical values (or) significant values: The sample values of the statistics beyond which the null hypothesis will be rejected are called critical values or significant values.

Nature of test	Level of significance 1%	5%	10%
Two-tailed test (z_α)	2.58	1.96	1.645
One-tailed test (z_α)	2.33 −2.33	1.645 −1.645	1.28 (right) −1.28 (left)

Test of significance of Large samples:

Test for single mean

Null hypothesis : $H_0 : \mu = \mu_0$

Test statistic, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (or) $z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

1. A sample of 900 members is found to have a mean of 3.4cm and S.D of 2.61cms. Is the sample from a large population of mean 3.25cm and SD 2.61cms. If the population is normal and its mean is unknown find the 95% confidence limits of true mean.

$n = 900$, $\bar{x} = 3.4$, $S = 2.61$, $\mu = 3.25$, $\sigma = 2.61$

Null hypothesis H_0 : Assume that the sample is drawn from the population with $\mu = 3.25$ (ie) $H_0 : \mu = 3.25$

Alternative hypothesis H_1 : $\mu \neq 3.25$ [2-tailed test]

Level of significance :

Level of significance $\alpha = 5\% = 0.05$

Test statistic : $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}}$

$= 1.724$

Critical value : Critical value at 5% is $z_\alpha = 1.96$

conclusion : $\because$ $|z| = 1.724 < 1.96 = z_\alpha$.

$\therefore$ H_0 is accepted at 5% level of significance

$\therefore$ The sample is taken from population where mean is 3.25cm

Confidence limits : $z_\alpha = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$\Rightarrow \mu = \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96\left(\frac{2.61}{\sqrt{900}}\right) = 3.4 \pm 0.17$

(ie) $3.23 < \mu < 3.57$

2. A random sample of $\boxed{200}$ employees at a large corporation showed this mean average to be 42.8 years with S.D of 6.89 years. Test the hypothesis $H_0: \mu = 40$, $H_1: \mu > 40$ at $\alpha = 0.01$ L. of S.

Given: $n = 200$, $\bar{x} = 42.8$, $\mu = 40$, $\sigma = 6.89$

Null hypothesis: $H_0: \mu = 40$

Alternative hypothesis $H_1: \mu > 40$ [1-tail right]

Level of significance: $\alpha = 0.01 = 1\%$.

Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{42.8 - 40}{6.89 / \sqrt{200}}$

$= 5.747$

Critical value: Critical value at 1% [1-tailed right] is

$z_\alpha = 2.33$

Conclusion: $\because |z| = 5.747 > 2.33 = z_\alpha$

H_0 is rejected at 1% level of significance

$\therefore$ Alternative hypothesis $H_1: \mu > 40$ is accepted.

3. The mean height of college students in a city are normally distributed with S.D 6 cms. A sample of 100 students has mean height 158cms. Test the hypothesis that the mean height of college students in the city is 160cm. Also obtain 99% confidence limits for the true mean.

Given: $n = 100$, $\sigma = 6$, $\bar{x} = 158$, $\mu = 160$

Null hypothesis: Difference is not significant, $H_0: \mu = 160$

Alternative hypothesis: $H_1: \mu \neq 160$ [2-tailed]

Level of significance: $\alpha = 1\% = 0.01$

Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{158 - 160}{6 / \sqrt{100}} = 3.33$

Critical value: Critical value at 1%. [2-tailed] is $z_\alpha = 2.58$

Conclusion: $\because |z| = 3.33 > 2.58 = z_\alpha$

Null hypothesis is rejected at 1% level of significance

∴ The mean height of the college students in the city is 160cms is not true.

Confidence limits:-

$$\mu = \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}} = 158 \pm 2.58 \frac{6}{\sqrt{100}} = 158 \pm 1.548$$

$$= 156.452, 159.548.$$

ie, $156.452 \nless \mu \nless 159.548$. Here $\mu = 160$ doesn't lie in the interval.

4. A sample of 100 students is taken from a large population. The mean height of the students in the sample is 160cms. Can it be reasonably regarded that this sample is from the population of mean 165cms and s.D 10cm? Also find the 95% confidence limits for the mean.

Given: $n = 100$, $\bar{x} = 160$ cm, $\mu = 165$ cm, $\sigma = 10$ cm

Null hypothesis: The sample is from the population of mean 165cm

$H_0: \mu = 165$

Alternative hypothesis: $H_1: \mu \neq 165$ [2-tailed]

Level of significance: $\alpha = 5\% = 0.05$

Test statistic: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{160 - 165}{10/\sqrt{100}}$

$$= -5$$

Critical value: z_α at 5% (2-tailed) is 1.96

Conclusion: $|z| = 5 > 1.96 = z_\alpha$.

∴ H_0 is rejected at 5% level of significance.

∴ The sample is not from the population of mean 165cm.

Confidence limits: $z_\alpha = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$$\Rightarrow \mu = \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}} = 160 \pm 1.96 \frac{10}{\sqrt{100}} = 160 \pm 1.96$$

$$\Rightarrow 158.04 \nless \mu \nless 161.96$$

ie, $\mu = 165$ cm doesn't lie in this interval.