Test Of Hypothesis

Module – V

Sample:

A finite sub set of the population is called a sample

Ex: selecting 2 students in a class of 60 students is a sample

sample size:

The no. of items in a sample is called sample size & it is denoted by n.

Sampling:

The process of selecting samples is called sampling

Small Sample: A sample with less than 30 ($n < 30$) items is called small sample or ~~exact~~ exact sample

large sample: A sample with greater than or equal to 30 items ($n \geq 30$) is called large sample

Random sample of finite population:

Selection of sample from finite population is known as random sample of finite population

Random sample of infinite population:

Selection of sample from infinite population is known as random sample of infinite population.

Note:

- The population measures mean, SD etc called as parameters.

The measures obtain from the sample are called as statistics

Parameters of Statistics :

The measures of population namely mean(μ) SD(σ), variance (σ^2) proportions(p) known as population parameters.

The measures computed from the sample observations namely mean ($\bar{x}$) Variance (x^2 or s^2) SD (s), proportions (p) are known as sample parameters or sample statistics or Statistics.

If $x_1, x_2 \dots x_n$ are elements of a random sample then sample mean $\bar{x} = \frac{\Sigma x}{n}$ where x is a element of given sample. n is no. of elements in a sample

Sample Variance x^2 or $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

where $\bar{x} = \frac{\Sigma x}{n}$, n = size of sample

standard deviation $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

Sampling Distribution:

The prob distribution of sample is called a sampling distribution.

Sampling distribution of Mean (σ known):

The prob distribution of $\bar{X}$ is called the sampling distribution of mean, the sampling distri of statistics depends on the size of population, size of sample, the method of choosing the samples

Case i: Sampling is done with replacement [suppose the samples are drawn from infinite population]

If the population size is N and sample size is n then the possible no. of samples are N^n.

pling

ple

al

Scanned with ACE Scanner

(a) Mean of sampling distri of means is given by

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

$$\therefore \mu = \mu_{\bar{x}}.$$

(b) Variance of sampling distri of means is given

$$V(\bar{x}) = \sigma_{\bar{x}}^2$$

$$= \frac{\sigma^2}{n}$$

where n is sample size

(c) SD of sampling distri of means,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Case 2: Sampling is done without replacement then (finite population). If the population size is N and sample size is n then the possible no of samples are NC_n.

(a) $\mu_{\bar{x}} = \mu$

(b) $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

(c) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$

Standard Error:

The SD. of $\bar{x}$ is also called as the standard error of mean & it is denoted by $\sigma_{\bar{x}}$.

$$SE = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Finite population multiplier ~~are~~ (or) finite population correction factor.

Here $\left(\frac{N-n}{N-1}\right)$ is known as finite population multi correction factor

Procedure for testing Of Hypothesis

Step 1: Statement or assumption Of hypothesis.

There are 2 types.

1. Null hypothesis
2. Alternative hypothesis

Null hypothesis

For applying the test of significance we step up a hypothesis a definite statement about the population parameter such hypothesis is called null hypothesis it is denoted by H_0. It is in the form of H_0 such that $\mu = \mu_0$

$$\boxed{H_0 : \mu = \mu_0}$$

Where μ_0 is the value which is assumed or claimed for the population characteristic

Alternative hypothesis

Any hypothesis which contradicts the null hypothesis is called alternative hypothesis. It is denoted by H_1

It is in the form of (i) $H_1 : \mu \neq \mu_0$

(ii) $H_1 : \mu > \mu_0$

(iii) $H_1 : \mu < \mu_0$.

The alternative hypothesis ~~One~~ (i) is known as a 2 tailed alternative hypothesis

Alternative hypothesis (ii) is known as right tailed and (iii) is known as left tailed.

The setting of alternative hypothesis is very imp to decide wheather we have to use a single tailed or 2 tailed test

Step 2: Specification Of the level Of significance

The level of significance denoted by α is the confidence with which we rejects or accepts the null hypothesis H_0 i.e is the max possible probability with which we are willing to risk an error in rejecting H_0 when it is true.

Level of significance is also known as the size of the test.

Step 3: Indentification of the test statistic.

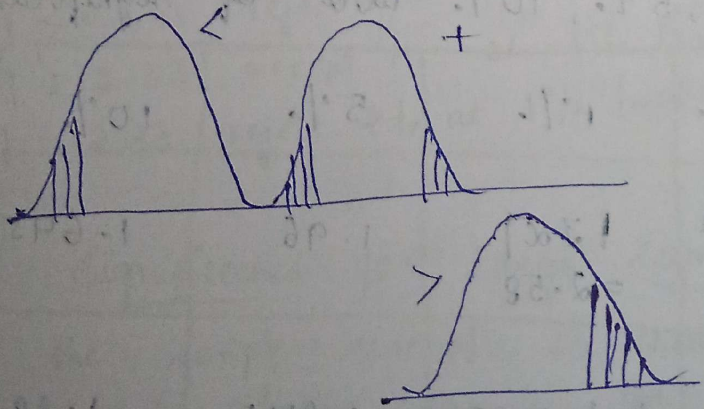
There are several tests of significance z, t, f. First we have to select the right test depending on the nature of the information given in the problem then we construct the test criteria & select the appropriate probability distribution.

Step 4: Critical Reg Region.

The critical region is formed based on the following factors

a) Distribution of statistics i.e wheather the statistic follows the normal, t, X^2

b) Form of alternative hypothesis if the form has not equal to X the critical region is divided equally in the left & right tails.



Step 5: Making Decision

We compute the value of the appropiate statistic the vals and as certain wheather computed value false in acceptance or rejection region depend on the specified level of significance

If the computed value < critical value we accept H_0 otherwise we reject H_a

Errors In Sampling:

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results in sam practice we decide to accept or to reject the lot after examine a sample from it. as such we have 2 types of errors

① **Type One Error:**

Reject H_0 when it is true it is denoted by α

Type Two Error:

Accept H_0 when it is wrong it is called type 2 error denoted by β

Critical values of Z for 2 tail & 1 tail test at 1%, 5%, 10% level of significance

Level of significance	1%	5%	10%
Critical values for 2 tailed test	$\lvert Z\alpha \rvert = 2.58$	1.96	1.645
Critical values for right tailed test	$\lvert Z\alpha \rvert = 2.33$	1.645	1.28
left tailed	-2.33	-1.645	-1.28

Procedure for Testing Of Hypothesis

Various steps involved in the testing of hypothesis

1.) Null Hypothesis

2) Alternative Hypothesis

3) Level Of Significance

4) Test statistic

Compute the test statistic $Z = \frac{t - E(t)}{SE \text{ of } t}$

under the null hypoth

5) conclusion:

For 2 tailed Test

If $|z| \geq 1.96$ reject ~~accept~~ H_0 at 5% level of significance ~~in~~

$|z| < 1.96$ accept H_0 at 5% level of significance

If $|z| < 2.58$ accept H_0 at 1% level of significance

If $|z| > 2.58$ reject H_0 at 1% level of significance

for single tailed (Right or left Test):

- If $|z| < 1.645$ accept H_0 at 5% level of significance
- If $|z| > 1.645$ rejects H_0 at 5% level of significance
- If $|z| < 2.33$ accept H_0 at 1% level of significance.
- If $|z| > 2.33$ reject H_0 at 1% level of significance

Test Of Significance for large Sample:

If the sample size $N \geq 30$ then we consider such samples as large samples. The test of significance used in large samples are diff from those used in small samples because small samples fail to satisfy the assumption under which large sample analysis is done.

If n is large the distributions binomial poison, chi square etc are closely approximated by normal distribution.

$\therefore$ for large samples the sampling distribution of a satistic approximately the normal distribution.

Suppose we wish to test the hypothesis that the probability of success in such big trial p assuming it to be true the mean μ and S.D σ of the sampling distribution of no. of successes are np and ~~$\sqrt{p}$~~ $\sqrt{npq}$ respectively.

If 'x' be the observed no. of successes in the sample and 'z' is the standard normal statistic $z = \frac{x-\mu}{\sigma}$

Thus we have the following test of significance

1. If $|z| < 1.96$ the difference b/w the observed and expected no. of successes is not significant.
2. If $|z| > 1.96$ the difference is significant at 5% level of significance.
3. If $|z| > 2.58$ the difference is significant at 1% level of significance.

Assumption for Large Samples:

1. The random sampling distribution of statistic has the properties of the normal curve. This may be not whole good in the case of ~~sa~~ small samples.
2. Values (~~#~~ i.e statistic) given by the samples are

sufficiently close to the population values & can be used in its place for calculating the S.E of the estimate.

Q.) A coin was tossed 960 times and returned heads 183 times test the hypothesis ~~then~~ that the coin is unbiased used at 0.05 level of significance.

Sol: Given $n = 960$

$x = 183$

P = prob of getting head

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

$q = \frac{1}{2}$ $(\because P+q=1)$.

$\mu = np$

$\mu = 960\left(\frac{1}{2}\right) = 480$.

$\sigma = \sqrt{npq}$

$= \sqrt{960\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$

$= \cancel{15.25}\ 15.49$

null hypothesis (H_0):

The coin is unbiased.

Alternative hypothesis (H_1):

The coin is biased.

Level of significance (α) = 0.05

The test of statistic $Z = \frac{x-\mu}{\sigma} = \frac{183-480}{15.49}$

$Z = -19.17$

$|Z| = 19.17$.

As $|Z|$ is > 1.96 the null hypothesis H_0 has to be rejected at 5% level of significance and the coin is biased.

29) A dice is tossed 960 times and it falls with 5 upwards 184 times is the die unbiased at a level of significance of 0.01

Sol: $n = 960$

$x = 184$, p = prob of getting 5

$p = \frac{1}{6}$

$p + q = 1$

$q = \frac{5}{6}$

$\mu = np$

$= 960\left(\frac{1}{6}\right) = 160$

$\sigma = \sqrt{npq}$

$= \sqrt{960\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = 11.55$

① null hypothesis (H_0) = the die is unbiased

② Alternative hypothesis (H_1) = the die is biased

③ Level of significance = 0.01 = 1%

④ the test of statistic (z) = $\frac{x - \mu}{\sigma} = \frac{184 - 160}{11.55}$

$z = 2.07$

As $|z| < 2.58$ the null hypothis H_0 accepted at 1% level of signifance. & the die is unbiased.

Under large sample Test we will see 4 imp tests to test the significance.

① Test of significance for single mean

② Test of significance for diff of mean

③ Test of significance for single proportion

④ Test of significance for difference of proportion.

① Test of Significance for a single mean: (large sample)

~~At~~ Let a random sample of size

$N \geq 30$ has the mean $\bar{x}$ of μ be the population mean also the popu mean μ has a specified value μ_0.

working Rule:

(1) Null hypothesis: H_0 is $\bar{x} = \mu$. i.e there is no significance diff b/w the sample mean and popu mean or the sample mean has been drawn from the parent population.

(2) Alternative hypothesis H_1:

(1) $\bar{x} \neq \mu$

(2) $\bar{x} > \mu$. $(\mu > \mu_0)$

(3) $\bar{x} < \mu$ $(\mu < \mu_0)$

since n is large the sampling distribution of $\bar{x}$ is approximately normal.

(3) Level of significance α:

(4) The test statistic: we have the following 2 cases

Case (i): when the S.D σ of population is known in this case S.E of mean

$$S.E = \frac{\sigma}{\sqrt{n}}$$

where n = sample size, σ = S.D

The test statistic is $Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - \mu}{S.E}$

Case (ii): when the S.D σ of popu is not known in this case $S.E = \dfrac{s}{\sqrt{n}}$

Hence the test statistic is $Z = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

(5) Find the critical value $Z\alpha$ at the level of significance α.

(6) Decision:

(i) If $|Z| \leq Z\alpha$ we accept H_0

(ii) If $|Z| > Z\alpha$ we reject H_0

Note:

We reject null hypothesis when $|z| > 3$ without mentioning any level of significance. The test statistic is $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

- when σ is known.

- The population SD σ is not known then we use the statistic $z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$

The values $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ is called 95% confidence limits.

- $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ are called 99% confidence limits
- $\bar{x} \pm 2.33 \frac{\sigma}{\sqrt{n}}$ are called 98% confidence limits

Q) According to the norms ~~est~~ established for a mechanical aptitude test persons who are 18% years old have an average height of 73.2 with a standard deviation of 8.6 if 4 randomly ~~to~~ selected persons of that age averaged 76.7 test the hypothesis $\mu = 73.2$ against for alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance

Sol: Given $\mu = 73.2$

mean of sample $\bar{x} = 76.7$

$n = 4$

$SD = 8.6$

(i) null hypothesis $H_0 : \mu = 73.2$

(ii) Alternative hypothesis $H_1 : \mu > 73.2$

(iii) Level of significance $\alpha : 0.01$

(iv) ~~stat~~ test statistic

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{76.7 - 73.2}{\frac{8.6}{\sqrt{4}}} = 0.8139 \approx 0.814$$

$Z_c = 2.33$ at 0.01 (99%)

$|z| < Z_c = 2.33$

we accept H_0 i.e $\bar{x}$ and μ donot differ significantly.

Q) A sample of 64 students have a mean weight of 70 Kgs can this be regarded as a sample from population with mean weight 56 Kgs. and standard deviation 25 Kgs.

given: $n = 64$

$\mu = 56.$

$\bar{x} = 70$

a sample of 64 students with mean weight 70 Kgs can be regarded as a sample from a population with mean weight 56 Kgs and $\sigma = 25$ Kgs population with mean weight sample cannot be regarded as I coming from the population.

level of significance $\alpha = 0.05$ (assumption)

Test statistic ; $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$= \dfrac{70 - 56}{\frac{25}{\sqrt{64}}} = 4.48.$

we reject null hypothesis H_0.

$\therefore$ Sample cannot be regarded as I coming from population.

Q) In a random sample of 60 workers the avg time taken by them to get to work is 33.8 mins. with a S.D of 6.1 min. Can we reject the null hypothesis $\mu = 32.6$ mins in favour of alternative hypothesis $\mu > 32.6$ at 0.025 level of significance.

Sol:

$n = 60$

$\mu = 32.6$, $\bar{x} = 33.8$

null hypothesis $H_0 \rightarrow \mu = 32.6$

Alternative hypothesis $H_1 \rightarrow \mu > 32.6$

level of significance $\rightarrow \alpha = 0.025$

test statistic $Z, = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}}$

$= 1.523$

$Z_\alpha = 2.58$

$Z < Z_\alpha$

accept H_0

Q.) A sample of 400 items is taken from a population whose SD is 10 the mean of sample is 40 test wheather the sample has come from a population with mean 38 also calculate 95% confidence interval for the population.

Given: sample size $n = 400$

population mean $\mu = 38$, $\sigma = 10$

$\bar{x} = 40$

null hypothesis $H_0 \to \mu = 38$

Alternative hypothesis H_1

$\mu \neq 38$.

level of significance $\to \alpha = 0.05 (95\%)$

Test statistic $Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{40 - 38}{\frac{10}{\sqrt{400}}}$

$= 4.$

$Z_\alpha = 1.96$ $\quad Z > 3$

$\therefore$ we reject null hypothesis H_0.

Confidence interval $= \left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right)$

$= \cancel{1.96}$

$\left[40 - 1.96\left(\frac{10}{\sqrt{400}}\right), 40 + 1.96\left[\frac{10}{\sqrt{400}}\right]\right]$

$= (39.02, 40.98)$

Method - II

Test of significance (or) Test for equality of 2 means (large samples)

Let the null hypothesis H_0 is $\mu_1 = \mu_2$

the alternative hypothesis H_1 $\mu_1 \neq \mu_2$.

S.E of $\bar{x}_1 - \bar{x}_2$ is equal to $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Where σ_1, σ_2 are SD.

To test whether there is any significant difference b/w $\bar{x}_1$ & $\bar{x}_2$ we have to use the statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{S.E(\bar{x}_1 - \bar{x}_2)} \Rightarrow \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$(\Delta = \mu_1 - \mu_2)$

if $\Delta = 0$ then the 2 populations have the same means.

If $\Delta \neq 0$ the 2 populations are different under $H_0: \mu_1 = \mu_2$ then Δ becomes 0. the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ when σ is known.

(or)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$ when σ is not known

where S_1^2, S_2^2 are sample variances.

Rejection Rule for H_0:

- If $|z| > 1.96$ reject H_0 at 5% level of significance
- If $|z| > 2.58$ reject H_0 at 1% level of signific.
- If $|z| > 1.645$ reject H_0 at 10% level of significa.
- If $|z| > 3$ reject then either samples have not been drawn from the sample population or the sampling is not simple.

otherwise Accept H_0

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

If σ is not known we can estimate σ^2 given by $\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$

Q) The means of two large samples of sizes 1000 & 2000 members are 67.5 inches and 68.0 inchs. respectively. Can the samples be regarded as drawn from the sample population of SD 2.5 inchs.

Sol: given:

Let μ_1 & μ_2 be the means of two populations and $\bar{x}_1$ & $\bar{x}_2$ be the means of 2 samples.

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68.0$

$n_1 = 1000$, $n_2 = 2000$

$\sigma = 2.5$

null hypothesis $H_0 \rightarrow \mu_1 = \mu_2$

alternative hypothesis $H_1 \rightarrow \mu_1 \neq \mu_2$

level of significance

test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1^2} + \dfrac{\sigma^2}{n_2^2}}}$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1^2} + \dfrac{1}{n_2^3}}}$$

$$= \frac{67.5 - 68.0}{2.5\sqrt{\dfrac{1}{(1000)^3} + \dfrac{1}{(2000)^2}}}$$

$$= -\cancel{178.8} \quad -5.16$$

$|z| = 5.16 > 3$

no need to consider level of significance

∴ We reject null hypothesis Ho and we can conclude that $\mu_1 \neq \mu_2$

∴ samples are not drawn from same population.

Q) Samples of students were drawn from 2 universities and from their weights in kgs, mean & SD are calculated & shown below make a large sample test to test the significance of the difference b/w the means.

	mean	S.D.	Size of Sample
uni A	55	10	400
uni B	57	15	100

Sol: given $\bar{x}_1 = 55$, $\sigma_1 = 10$, $n_1 = 400$

$\bar{x}_2 = 57$, $\sigma_2 = 15$, $n_2 = 100$

null hypothesis $H_0 \rightarrow \bar{x}_1 = \bar{x}_2$ (i.e no diff)

Alternative hypothesis $H_1 \rightarrow \bar{x}_1 \neq \bar{x}_2$

~~LoS~~

test statistic is $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$= \frac{55 - 57}{\sqrt{\frac{10^2}{400} + \frac{15^2}{100}}}$$

$$= -1.26$$

assume $\alpha = 0.05$ then $Z_\alpha = 1.96$

$|z| < Z_\alpha$ accept H_0.

∴ We can conclude that there is no diff in sample means.

Q) The research investigator is interested in studying wheather there is a significant difference in the salaries of MBA grades in 2 metro polytin cities. A random sample of size 100 from mumbai yields on avg income of rupees 20,150 from another random sample of chennai results in an avg income of 20,250 the variances of both the population gives as $S_1^2 = 40000$, $S_2^2 = 32400$ respectively.

given: $n_1 = 100$, $n_2 = 60$, $\bar{x}_1 = 20150$, $\bar{x}_2 = 20250$

$S_1^2 = 40000$, $S_2^2 = 32400$.

null hypothesis $H_0 \rightarrow \mu_1 = \mu_2 \rightarrow$ diff of mean is significant

Alternative hypothesis $H_1 \rightarrow \mu_1 \neq \mu_2$

Level of significance $\alpha = 0.05$ (assume)

test statistic is $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

Scanned with ACE Scanner

$$z = \frac{20150 - 20250}{\sqrt{\frac{40000}{100} + \frac{32400}{60}}}$$

$$= 3.26.$$

$|z| = 3.26 > z_\alpha$

reject H_0 at 5% level of significance & conclude that there is a significant diff b/w 2 means. ~~of~~

Test of significance ~~per~~ for single proportion (large samples)

Method 3:

Suppose a large sample of size n has a sample proportion p of members possessing a certain attribute (i.e. proportion of successes).

To test the hypothesis that the proportion P in the population has a specified value P_0. Let us set a null hypothesis H_0 is $P = P_0$ (P_0 is particular value of P) then

alternative hypothesis $H_1 \rightarrow$

(i) $P \neq P_0$

(ii) $P > P_0$

(iii) $P < P_0$

Since n is large the sampling distribution of p is approx normal

$\therefore$ If H_0 is true the test statistic

$$z = \frac{p - P_0}{S.E \text{ of } p}$$

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

where $\hat{p}$ is the sample proportion is approx normally distributed.

Q) In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice & wheat are equally popular in this state at 1% level of significance.

Sol: $n = 1000$

$x = 540$

$$p = \frac{x}{n} = \frac{540}{1000} = 0.54$$

$$P = \frac{1}{2} = 0.5 \ (P_0)$$

1. $H_0 : P = 0.5$

2. $H_1 : P \neq 0.5$ (two tailed)

3. $\alpha = 1\%$ ($Z_\alpha = 2.58$)

4. $$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5) \times (0.5)}{1000}}}$$

$$Z = 2.53$$

$$Z_\alpha = 2.58$$

$$\therefore |z| < Z_\alpha$$

We can conclude that we accept H_0.

Scanned with ACE Scanner

Q) A manufacture claimed that atleast 95% of the equipment which he supplied to a factory conformed ~~the~~ to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

Sol: given: $n = 200$ P = population proportion.

$x = 200 - 18 = 182$

$p = \frac{182}{200} = 0.91$

$P = \frac{95}{100} = 0.95$

1. $H_0 : P = P_0 = 0.95$ (95%)

2. $H_1 : P < 0.95$

3. Level of significance (α):

$\alpha = 5\% = 0.05$

$(Z_\alpha = 1.645)$

4. $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{0.95(0.05)}{200}}}$

$= -2.59$

$|Z| > Z_\alpha$

$\therefore$ We reject H_0.

Q) A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Find the percentage of bad pineapples in the consignment.

Sol: given: $n = 500$

$x = 65$

p = sample proportion

$$p = \frac{x}{n} = \frac{65}{500} = 0.13.$$

$$q = 1 - p = 0.87.$$

WKT, limits for population proportion P are given

$$p \pm 3\sqrt{\frac{pq}{n}}$$

$$= 0.13 \pm 3\sqrt{\frac{0.13(0.87)}{500}}$$

$$= 0.13 \pm 0.045$$

$$= (0.084, 0.0175).$$

Scanned with ACE Scanner

Test for significance of difference of proportion

Let P_1 & P_2 be sample proportion in 2 large random sample of sizes n_1, n_2 drawn from the 2 population properties P_1, P_2. $H_0 : P_1 = P_2$

alternative hypothesis $H_1 : P_1 \neq P_2$ or $P_1 > P_2$ or $P_1 < P_2$

Level of signi α :

Test statistic $z = \left(\dfrac{p_1 - p_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}} \right)$

also $\dfrac{p_1 - p_2}{S.E(p_1 - p_2)}$.

when the population proportion P_1 & P_2 are not known but sample proportion p_1 & p_2 are known.

Method of substitution

p_1 & p_2 are substituted for P_1 and P_2.

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}$$

→ **Method of pooling:**

In this method the estimated value of 2 populations proportions is obtained by pooling of 2 sample proportion p_1 & p_2 into a single proportion p.

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{x_1 + x_2}{n_1 + n_2}$$

$p = \dfrac{x}{n}$

$x = np$

$$z = \frac{p_1 - p_2}{SE(p_1 - p_2)} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Scanned with ACE Scanner

If sample proportion are not known

$$|z| = \frac{|P_1 - P_2|}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Q) Random Sample of 400 men and 600 women where asked wheather they would like to have flyover beside their residence 200 men 325 women were in favour of propotion. Test the hypothesis that the proportion of men & women in favour of proportion at same at 5% level.

Sol: given: $n_1 = 400$, $n_2 = 600$

$x_1 = 200$, $x_2 = 325$

$p_1 = \frac{x_1}{n_1} = \frac{200}{400}$, $p_2 = \frac{325}{600}$

$= 0.5$ $= 0.54$.

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{200 + 325}{400 + 600} = 0.525$$

$$q = 1 - P \rightarrow 1 - 0.525 = 0.475$$

1. $H_0 : P_1 = P_2$.
2. $H_1 : P_1 \neq P_2$
3. α : $\alpha = 5\%$ ($Z_\alpha = 1.96$)
4. $$Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.5 - 0.54}{\sqrt{(0.525)(0.475)\left[\frac{1}{400} + \frac{1}{600}\right]}}$$

$$Z = -1.24$$

$$|Z| < Z_\alpha$$

We accept H_0.

$\therefore$ We conclude that the proportion of men and women in the favour of proposal are same.

Q.) A manufacture of electronic equipment subjects samples of two completing brands of transistors to an accelerated performance. If 45 of 180 transistor of the first kind and 34 of 120 transistors of the second kind of fail the test, what can he conclude at the level of significance $\alpha = 0.05$ about the diff b/w corresponding sample proportions?

Sol: $n_1 = 180$ $\quad x_1 = 45$ $\quad P_1 = 0.25$

$n_2 = 120$ $\quad x_2 = 34$ $\quad P_2 = 0.283$

$$P_1 = \frac{x_1 + x_2}{n_1 + n_2} = \frac{45 + 34}{180 + 120} = 0.26$$

$$q = 1 - p$$
$$= 1 - 0.26$$
$$= 0.737$$

1. $H_0 : P_1 = P_2$
2. $H_1 : P_1 \neq P_2$
3. $\alpha : \alpha = 0.05 \quad (Z_\alpha = 1.96)$
4. $$Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.25 - 0.283}{\sqrt{(0.26)(0.737)\left[\frac{1}{180} + \frac{1}{120}\right]}}$$

$$= 0.639.$$

$|Z| < Z_\alpha$

accept H_0.

We conclude there is no diff.

Q) In two large proportions there are 30% and 25% respectively of fair haired people. Is this diff lighty to be hidden in samples of 1200 and 900 respectively from the two populations.

Sol: $n_1 = 1200,\ n_2 = 900.$

$$P_1 = \frac{30}{100} = 0.3 \ ; \ P_2 = \frac{25}{100} = 0.25$$

1) $H_0 : P_1 = P_2$

2) $H_1 : P_1 \neq P_2$

3) $\alpha : \alpha = 0.05$ (assume)

4) $$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Q)

$Q_1 = 1 - P_1 = 0.7$ $\quad Z_\alpha = (1.96)$

$Q_2 = 1 - P_2 = 0.75$

$$z = \frac{0.3 - 0.25}{\sqrt{\frac{0.3(0.7)}{1200} + \frac{0.25(0.75)}{900}}} = 2.55$$

$\therefore |z| > Z_\alpha$

we reject H_0..

$\therefore$ There is

Q) A cigrette - manufacturing firm claims that its brand A line of cigrettes outsells is brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B, test whether the 8% diff is a valid claim.

→ $n_1 = 200$ $\quad x_1 = 42$ $\quad \therefore P_1 - P_2 = 8\%$

$n_2 = 100$ $\quad x_2 = 18$

$P_1 = \frac{x_1}{n_1} = 0.21$, $P_2 = \frac{x_2}{n_2} = 0.18$

$\cancel{P = \frac{x_1 + x_2}{n_1 + n_2} = 0.2}$ $\quad p = \frac{x_1 + x_2}{n_1 + n_2} = 0.2$

$q = 1 - 0.2 = 0.8$

1.) Null hypothesis $H_0 \to P_1 - P_2 = 8\%$

2) H_1 : $P_1 - P_2 \neq 8\%$

3) α ; $\alpha = 5\%$ $(Z_\alpha = 1.96)$

4) $$z = \frac{(P_1 - P_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.21 - 0.18) - 0.08}{\sqrt{0.2(0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$= 1.02$

we accept H_0

$|z| < Z_\alpha$

Test of Hypothesis for small samples.

There are 3 imp tests for small samples

1) student's t – test
2) student's F – test
3) χ^2 – test

Let $\bar{x}$ = mean of sample
σ = S.D of sample
n = size of sample
μ = mean of the population supposed to be normal.

Then the statistic $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$

If $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n}$ (sample variance)

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{\bar{x} - \mu}{S/\sqrt{n}}.$$

t-Test (single mean):

statistic $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$ where $\bar{x}, \mu, S, n$ have usual meanings

$H_0 \Rightarrow \mu = \mu_0$

$H_1 \Rightarrow \mu \neq \mu_0$

Assuming that H_0 is true, the test statistic is given by

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

Q) A sample of 26 bulbs gives a mean life of 990 hrs with a S.D of 20 hrs. the manufacturer claims that the mean life of bulbs is 1000 hour. Is the sample not upto the standard.

Sol: $\bar{x} = 990$, $\mu = 1000$. SD $\to s = 20$

$n = 26$

student's t - test $\to t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{990 - 1000}{\cancel{4}\ 20/5} = -2.5$

1) $H_0 \to \mu = 1000$

2) $H_1 \to \mu < 1000$ becoz the sample not up to the standard.

$\nu = n-1$

degrees of freedom $= 26 - 1 = 25$

3) $\alpha \to \alpha = 0.05$ (Assume)

4) Test : $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = -2.5$

$|t| = 2.5$

The tabular value 't' at 5% level with degrees of freedom 25 for left tailed is 1.708

$t_{\frac{\alpha}{2}} = 1.708$

$|t| > t_{\frac{\alpha}{2}}$ reject H_0

we conclude that the sample of bulbs are not standard.

Q) A random sample of size 16 values from a normal population showed a mean of 53. sum of stairs of deviation from the mean = 150 Can the sample be regarded as taken from the population having 56 has mean obtain 95% of confident limits of mean of population

Sol: $n = 16$, $\bar{x} = 53$, $\Sigma(x - \bar{x})^2 = 150$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$= \frac{150}{15} = 10 \longrightarrow s = \sqrt{10}$$

~~$\nu = n-1 = 15$~~

1) $H_0 \longrightarrow \mu = 56$

2) $H_1 \longrightarrow \mu \neq 56$

3) $\alpha \longrightarrow \alpha = 0.05$ (Assume)

4) Test $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{53-56}{\sqrt{10}/\sqrt{16}}$

$= -3.79$

$|t| = 3.79$

$t_{\frac{\alpha}{2}} = 2.13$

$t > t_{\alpha/2} = 2.13$

we reject H_0.

Q) Producer of 'gutkha' claims that the ~~m~~ nicotine content in his 'gutkha' on the avg is 83gmg. Can this claim ~~B~~ be accepted if a random sample of 8 gutkhas of this type have the ~~notificatio~~ nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6mg.

Sol: given: $n = 8$, $\mu = 1.83$ mg

$H_0 \rightarrow \mu = 1.83$

$H_1 \rightarrow \mu \neq 1.8\cancel{8}3$

$\alpha \rightarrow \alpha = 0.05$

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} =$ $\qquad \bar{x} = \frac{\Sigma x}{n} = 1.9$

x	2.0	1.7	2.1	1.9	2.2	2.1	2.0	1.6
$x-\bar{x}$	0.05	-0.25	0.15	-0.05	0.25	0.15	0.05	-0.
$\Sigma(x_i-\bar{x})^2$	0.0025	0.0625	0.022	0.0025	0.0625	0.0225	0.0025	0.12

$\Sigma = 0.3025$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{0.3025}{7} = 0.04 \Rightarrow S = 0.02$$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{1.95 - 1.83}{\frac{0.02}{\sqrt{8}}} = 16.97$$

students

- t-Test for diff of means:

Let $\bar{x}$, $\bar{y}$ be the mean of two independent samples of sizes n_1 & n_2 ($n_1 < 30$, $n_2 < 30$) μ_1 & μ_2 are means of 2 population sample. $H_0 \rightarrow \mu_1 = \mu_2$
$H_1 \rightarrow \mu_1 \neq \mu_2$.

If $\sigma_1 = \sigma_2 = \sigma$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$, s_1^2, s_2^2 are the 2 sample variance.

$$S.E \text{ of } (\bar{x} - \bar{y}) = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

assuming that H_0 is true, the test statistic t is given by $t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$.

Scanned with ACE Scanner

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = 9.39$$

Snedecor's F - Test of Significance

When testing the significance of the diff of the means of two samples, we assumed that the 2 samples came from the same population or from population with equal variances.

If the variance of the population are not equal a significance in the means may arise hence before we apply the t-test for significance of diff of 2 means we have to test for equality of population variances using F-test. If S_1^2, S_2^2 are the variances of 2 samples of sizes n_1 & n_2 respectively then population variances are given by:

$$n_1 S_1^2 = (n_1 - 1) s_1^2 \text{ and } n_2 S_2^2 = (n_2 - 1) s_2^2$$

The

If we want to test if these estimate s_1^2 & s_2^2 are significantly diff or if the samples may be regarded as drawn from the same popution or from 2 populations with same variences σ^2

Test for equality for 2 population variances.

Let 2 independent random samples of sizes n_1 & n_2 be drawn from 2 normal populations. To test the hypothesis that 2 population variances σ_1^2, σ_2^2 are equal.

1) null hypothesis $H_0 \rightarrow \sigma_1^2 = \sigma_2^2$ (or) $S_1^2 = S_2^2$

2) Alternative hypothesis $H_1 \longrightarrow \sigma_1^2 \neq \sigma_2^2$

Scanned with ACE Scanner

3) Estimates of σ_1^2, σ_2^2 are given by $S_1^2 = \frac{n_1 s_1^2}{(n_1-1)}$ or $\frac{\Sigma(x_i-\bar{x})^2}{(n_1-1)}$

$$S_2^2 = \frac{n_2 s_2^2}{(n_2-1)} = \frac{\Sigma(y_i-\bar{y})^2}{(n_2-1)}$$

where s_1^2, s_2^2 are variances

Assuming that H_0 is true the test statistic

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

Conclusion: If $F > F_\alpha$ we reject null hypothesis otherwise we accept null hypothesis

Note;

1. If F is close to 1 the two sample variances S_1, S_2 are nearly same.

→ If sample variance s^2 is given we can optain population variance σ^2 by using formula $n\sigma^2 = (n-1)S^2$

→ $F_\alpha(V_1, V_2)$ is the value of F with v_1 and v_2 degrees of freedom such that the area under the F-distribution to the right of F_α is α.

→ In the ~~tab~~ tables F_α tabulated for $\alpha = 0.05$ and $\alpha = 0.01$ for various combination for the degree of freedom V_1 & V_2 Clearly value of F at 5% is lower that at 1%.

Q) Pumpkins which were grown under 2 experimental conditions & random samples of 11 & 9 pumpkins show the sample SD of their weights as 0.8 and 0.5 respectively. assuming that the weight distribution are normal. test hypothesis that the 2 variances are equal.

Sol: given $n_1 = 11$, $S_1 = 0.8$

$n_2 = 9$, $S_2 = 0.8$

→ null hypothesis $H_0 \rightarrow \sigma_1^2 = \sigma_2^2$.

→ . $H_1 \rightarrow \sigma_1^2 \neq \sigma_2^2$

→ level of significance → $\alpha = 0.05$

→ test statistic $F = \frac{S_1^2}{S_2^2}$ or $\frac{S_2^2}{S_1^2}$

degree of freedom

$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11 \times 0.8^2}{11-1} = 0.704$

$\nu_1 = n_1 - 1 = 10$

$\nu_2 = n_2 - 1 = 8$

$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9 \times 0.5^2}{9-1} = 0.21$

$F_\alpha = 3.35$ for 10, 8 at 5% level of significance

$F < F_\alpha$ accept H_0.

∴ we conclude that the variances of 2 populations are same.

Q) In one sample of 8 observations the sum of squares of deviations of the sample values from the sample mean was 84.4 and Another sample of 10 observations it was 102.6 test whether there is any significant difference

Sol: $n_1 = 8$ $n_2 = 10$.

The time taken by the workers in performing a job by method 1 and method 2 is given below.

method 1	20	16	26	27	23	22	–
method 2.	27	33	42	35	32	34	38

Do the data ST the variances of time distribution from population from which the samples are drawn donot differ significantly

$H_0 \rightarrow \sigma_1^2 = \sigma_2^2$

$H_1 \rightarrow \sigma_1^2 \neq \sigma_2^2$

$\bar{x} = 22.3$

$\bar{x}_2 = 34.4$.

level of significance $\alpha = 0.05$ (Assume)

test statistic $\frac{S_1^2}{S_2^2}$ or $\frac{S_2^2}{S_1^2}$

x_1	$x - \bar{x}$	x_2	$x_2 - \bar{x}$	$(x - \bar{x})^2$	$(x_2 - \bar{x})^2$
20	-2.3	27	-7.4	5.29	54.76
16	-6.3	33	-1.4	39.69	1.96
26	3.7	42	7.6	13.69	57.76
27	4.7	35	0.6	22.09	0.36
23	0.7	32	-2.4	0.49	5.76
22	-0.3	34	-0.4	0.09	0.16
		38	3.6	81.34	12.96
					133.72

704

281