UNIT - III . TESTING OF HYPOTHESIS - I 22/02/18.

Testing of Hypothesis or Significance :-

The procedure which enable us to decide on the basis of sample results whether a hypothesis is true or not is called testing of hypothesis or significance

NULL HYPOTHESIS :- (H_0)

A hypothesis of definite statement about the population parameter is called Null hypothesis

NOTE :-

Generally we define null hypothesis with respect to given problem.

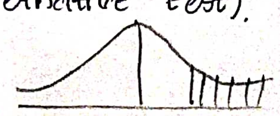
ALTERNATIVE HYPOTHESIS :- (H_1)

Any hypothesis which contradicts null hypothesis is called alternative hypothesis

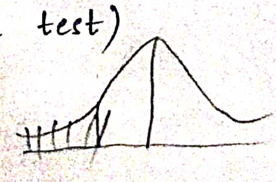
Eg :- To test null hypothesis of the population parameter mean μ_0

(i) i.e H_0 ; $\mu = \mu_0$, then the alternative hypothesis can be $\mu \neq \mu_0$ (two - tailed alternative test)

(ii) H_1 ; $\mu > \mu_0$ (right - tailed alternative test).

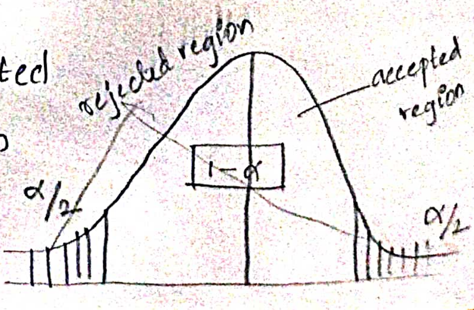


(iii) H_1 ; $\mu < \mu_0$ (left - tailed alternative test)



LEVEL OF SIGNIFICANCE (α) :-

The level of significance denoted by 'α', is the confidence with we reject or accept the null



hypothesis (H_0)

NOTE :- In practice we take either 5% or 1% is level of significance

5% level of significance means out of 100 cases we accept H_0 95 cases, reject H_0 5 cases.

Type 1. ERROR :-

Reject H_0 when it is true

i.e the null hypothesis H_0 is true but it is rejected by test procedure. It is called Type-I error.

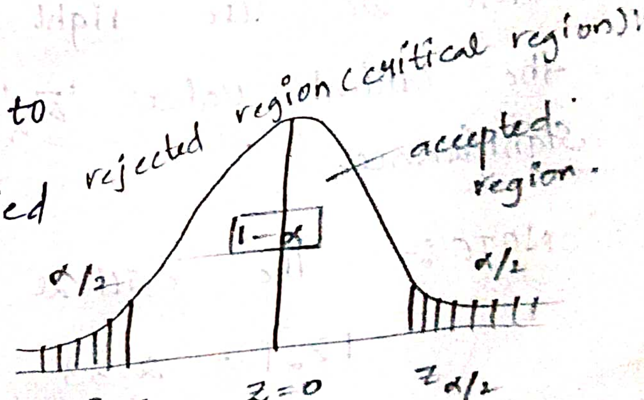
Type 2 ERROR :-

Accept H_0 when it is false

i.e the null hypothesis H_0 is false but it is accepted by test procedure. It is called Type-II error.

CRITICAL REGION :-

The region which leads to reject the hypothesis is called critical region.



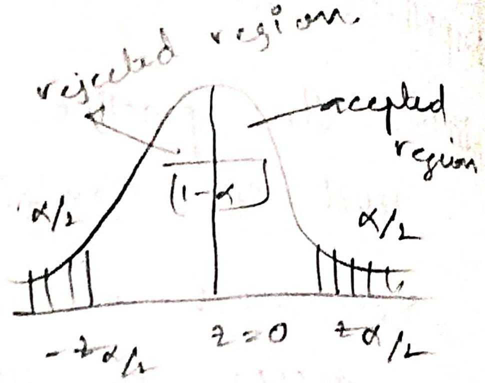
Similarly, the region which leads to accept the region is called accepted region

NOTE :-

The value of 'z' which separate the critical region and accepted region is called critical value

TWO- TAILED ALTERNATIVE TEST :-

We have null hypothesis H_0, $\mu = \mu_0$ and alternative hypothesis H_1, $\mu \neq \mu_0$, ($\mu > \mu_0$ or $\mu < \mu_0$), Here we use two- tailed alternative test, and the critical value is $Z_{\alpha/2}$ corresponding to 'α' level of significance.



~~RIGHT- TAILED~~

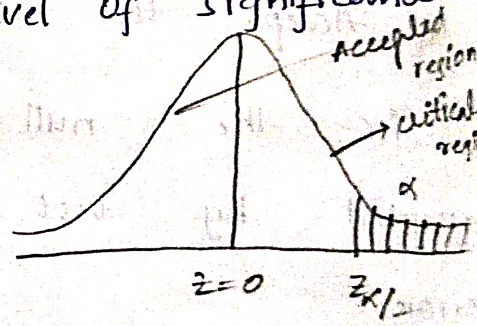
NOTE :- The critical values for two tailed test are

$|Z_{\alpha/2}| = 2.58$ if 1% level of significance.

$|Z_{\alpha/2}| = 1.96$ if 5% level of significance

$|Z_{\alpha/2}| = 1.645$ if 10% level of significance

RIGHT- TAILED ALTERNATIVE TEST :-



Null hypothesis H_0; $\mu = \mu_0$,

Alternate hypothesis H_1; $\mu > \mu_0$

Here we use right tailed test

the critical value 'Z_α' corresponding to 'α' level of significance

NOTE :- The critical values for right- tailed test

$|Z_\alpha| = 2.33$ if 1% level of significance

$|Z_\alpha| = 1.645$ if 5% level of significance

$|Z_\alpha| = 1.28$ if 10% level of significance

Similarly we can define left tailed alternative test

Procedure to test hypothesis :-

Step 1 :- Null hypothesis

$H_0 : \mu = \mu_0$

define Null hypothesis H_0 taking consideration the nature of the problem

Step 2 :- Alternative hypothesis.

define the alternative hypothesis H_1, so that we could decide whether we use One tailed test or two tailed test

Step 3 :- Critical value

By using 'α' level of significance find critical value $Z_{\alpha/2}$ or Z_α

Note :- If the significance level is not mentioned we have to consider 5% level of significance

Step 4 :- Test statistic.

Compute test statistic by using

$$Z = \frac{t - E(t)}{S.E}$$

where

$E(t)$ is population estimate

't' is sample statistic

$S.E$ standard error.

Step 5 :- Conclusion

we compare the computed value 'Z' w.r.to with the critical value Z_α or $Z_{\alpha/2}$ at the level of 'α' significance

i.e $|Z| < \{Z_{\alpha/2}\}$ (or) $|Z| < Z_\alpha$

We accept Null hypothesis 'H_0'

$|z| > z_\alpha$ (or) $|z| > z_{\alpha/2}$

We reject null hypothesis H_0.

MODEL 1 :- Testing of hypothesis or significance for single mean

Here the test statistics $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

n - sample size $\qquad \bar{x}$ - sample mean

μ - population density $\qquad \sigma$ - sample s.D

1. A sample of 64 students have mean wt of 70kgs can this be regarded as a sample from population having mean 65 kgs and s.D 25kgs

Sol:- Given that

Sample size $n = 64$

Sample mean wt of students

$\bar{x} = 70$ kgs

population mean of students

$\mu = 65$ kgs.

S.D of students

$\sigma = 25$ kgs.

Assume level of significans α is 5%.

Step 1 :-

Null hypothesis $H_0 : \mu = 65$ kgs

i.e population mean can be considered as 65kgs.

Step 2 :- Alternative hypothesis.

$H_1 : \mu \neq 65$ kgs. (Two-tailed test)

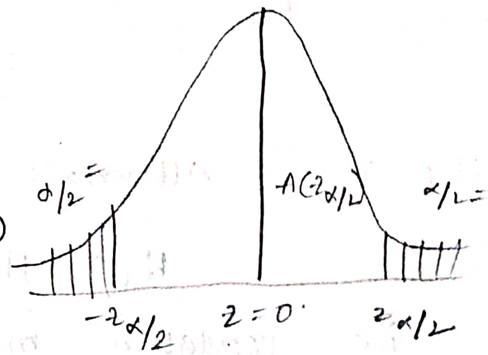
i.e population mean cannot be considered as 65 kgs (two tailed test) step 3: critical value $(Z_{\alpha/2})$

$\alpha = 0.05$; $\alpha/2 = 0.025$

$\alpha/2 = 0.5 - A(Z_{\alpha/2})$

$A(Z_{\alpha/2}) = 0.5 - \frac{\alpha}{2} = 0.5 - 0.025$

$A(Z_{\alpha/2}) = 0.475 = A(1.96)$



$\boxed{Z_{\alpha/2} = 1.96}$

step 4: Test statistic

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

$$= \frac{70 - 65}{\frac{25}{\sqrt{64}}}$$

$$= 1.6.$$

step 5: conclusion:

$|Z| = 1.6 < 1.96 = Z_{\alpha/2}$

$|Z| < Z_{\alpha/2}$ Accept H_0.

$\therefore$ population mean can be considered as 65kgs.

@ 5% level of significance

A random sample of 60 workers the avg time taken by them to complete the work is 33.8 min with S.D 6.1 min. Is it possible population mean time more than 32.6 mins at 1% level of significance

Sol:- Given $n = 60$

$\bar{x} = 33.8$ min. $\quad$ S.D $= 6.1$ m

$\mu = 32.6$ mins $\quad$ $\alpha = 1\%$

Step 1 :- Null hypothesis

$H_0 : \mu = 32.6$ mins

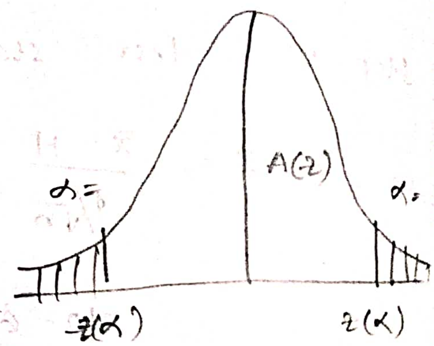
i.e population mean tym can be considered as 32.6 mins.

Step 2 :- Alternate hypothesis

$H_1 : \mu > 32.6$ mins (right tailed test)

i.e population mean tym is more than 32.6 mins

Step 3 :- critical value (Z_α)



$\alpha = 0.01$

$\alpha = 0.5 - A(Z)$

$A(Z_\alpha) = 0.5 - 0.01$

$A(2.33) = 0.4900$.

$Z_\alpha = 2.33$.

Step :- 4 :- test statistic

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

$$= \frac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}}$$

$$Z = 1.5237$$

Step 5 :- conclusion

$|Z| = 1.52 < 2.33 = |Z_\alpha|$

$|Z| < |Z_\alpha|$

Accept H_0

population mean time can be taken as 32.6 mins

∴ population mean tym is not more than 32.6 min @ 1% level of significance

A sample of 105 iron bars whose mean length is 10ft. Is it drawn from a population whose mean is 12ft and S.D 4ft at 2% level of significance

Sol:- Given that

sample size $n = 105$

sample mean length $\bar{x}$ is $= 10ft$

population mean length $\mu = 12ft$

S.D of population $\sigma = 4ft$.

@ 2% level of significance

$\alpha = 2.$

Step. 1:- Null hypothesis.

$H_0 : \mu = 12ft$.

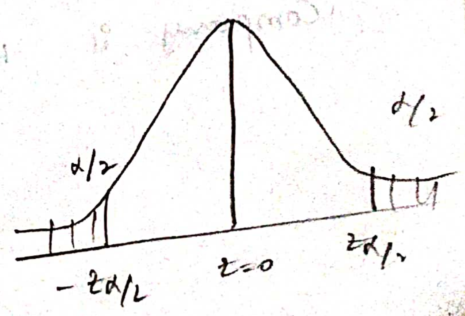
ie population mean length can be taken as 12fts

Step 2:- Alternative hypothesis

$H_1 : \mu \neq 12ft$

i.e population mean length cannot be taken as 12fts. (two tailed test)

Step 3:- Critical value. $(Z_{\alpha/2})$



$\alpha/2 = 0.02$

$\alpha/2 = 0.01$

$\alpha/2 = 0.5 - A(z_{\alpha/2})$

$A(z_{\alpha/2}) = 0.5 - 0.01$

$$A(z_{\alpha/2}) = 0.49$$

$$z_{\alpha/2} = 2.33.$$

Step 4:- test statistics

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

$$= \frac{10 - 12}{\frac{4}{\sqrt{105}}}$$

$$= -5.1235.$$

Step 5:- conclusion

$$|z| = |-5.1235| > 2.33 = z_{\alpha/2}$$

$$|z| > z_{\alpha/2}$$

reject 'H_0'

$\therefore$ population mean length is cannot be taken as 12fts. @ 2% level of significance (or) 98% confidence

4. The mean life time of a sample of 100 light tubes produced by a company is found to be 1560 hrs with a population mean of 90 hrs. Test the hypothesis that the mean life time of the tubes produced by the company is 1580 hrs at 2%. level of significance

Sample size $n = 100$

Sample mean $\bar{x} = 1560$ hrs

population mean $\mu = 1580$ hrs

S.D of population. $\sigma = 90$ hrs

3% level of significance.

Step 1:- Null hypothesis.

$H_0 : \mu = 1580.$

i.e population mean can be taken as 1580

Step 2:- Alternate hypothesis

$H_1 : \mu \neq 1580.$

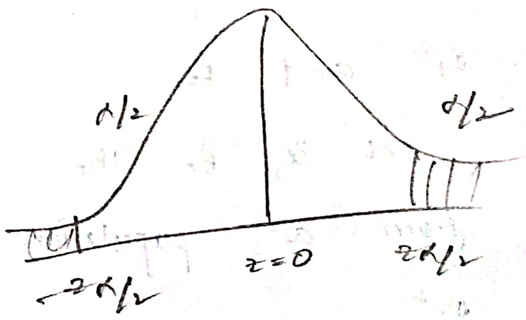
i.e population mean cannot be taken as 1580.

two tailed test

Step 3:- critical value $(z_{\alpha/2})$

$\alpha = 0.03$

$\frac{\alpha}{2} = 0.015$



$\alpha/2 = 0.5 - A(z_{\alpha/2})$

$A(z_{\alpha/2}) = 0.5 - 0.015$

$A(2.17) = 0.485$

$z_{\alpha/2} = 2.17.$

Step 4:- test statistic

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

$$= \frac{1560 - 1580}{\frac{90}{\sqrt{100}}}$$

$$= -2.22$$

Step 5 :- Conclusion

$$|z| = 2.22 > 2.17 = z_{\alpha/2}$$

reject ~~Accept~~ H_0.

Population mean life time cannot be taken as 1580 @ 3% level of significance with 97% confidence

MODEL - 2 :-

Testing of hypothesis for difference of means

H_1 σ_1 $\longrightarrow$ $\bar{x}_1 \to n_1$

H_2 σ_2 $\longrightarrow$ $\bar{x}_2 \to n_2$

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population having mean H_1 with variance σ_1^2 and be

Let $\bar{x}_2$ be the mean of a sample of size n_2 from a population having mean H_2 with variance σ_2^2

To test difference b/w population means

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

Example :- Test the significance of the difference b/w. population mean w.r. to given samples at 4% level of significance

	Size of sample	mean	S.D.
Sample A	100	61	4
Sample B	200	63	6

Given that the sample means

$\bar{x}_1 = 61$ $\quad$ $\bar{x}_2 = 63$.

~~Given~~ Sample sizes are

$n_1 = 100$ $\quad$ $n_2 = 200$.

Sample S.D's

$s_1 = \sigma_1 = 4$ $\quad$ $s_2 = \sigma_2 = 6$.

Significance level $\alpha = 4\%$.

Let μ_1, μ_2 are means of population of samples A and B respectively.

Step 1:- Null hypothesis

$H_0 : \mu_1 = \mu_2$

i.e The two samples are drawn from same population.

Step 2:- Alternative hypothesis

$H_1 : \mu_1 \neq \mu_2$

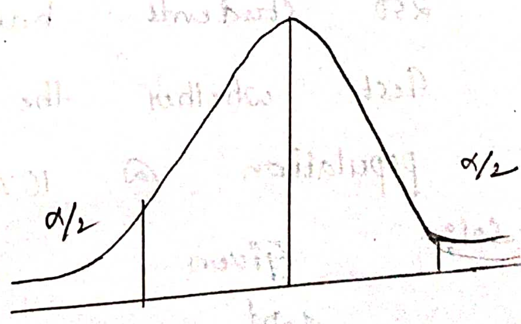
i.e The two samples are not drawn from same population (Two - tailed test)

Step 3:- Critical value.

$\alpha = 0.04$

$\alpha/2 = 0.02$.

$\alpha/2 = 0.5 - A(z_{\alpha/2})$



$A(z_{\alpha/2}) = 0.5 - 0.02$

$A(2.06) = 0.48$

$\boxed{z_{\alpha/2} = 2.06.}$

STEP 4:- Test statistic.

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

$$= \frac{61 + 63}{\sqrt{\frac{16}{100} + \frac{36}{200}}}$$

$$= -3.43.$$

STEP 5:- Conclusion.

$$|z| = 3.43 > 2.06 = z_{\alpha/2}$$

$|z| > z_{\alpha/2}$ reject H_0 (or)

Accept H_1

$\therefore$ The sample A and B are drawn from different popupation @ 4% level of significance with 96% confidence

2. In the 1st group there are 150 students having IQ 75 with S.D 15. In the 2nd group there are 250 students having mean IQ 70 with S.D 20. Test whether the groups have came from same population @ 10% level of significance

Sol:- Given.

1st & 2nd group of sample sizes are

$n_1 = 150$ $n_2 = 250$

Sample IQ's.

$\bar{x}_1 = 75$ $\bar{x}_2 = 70$

Sample S.D's

$\sigma_1 = 15$ $\qquad$ $\sigma_2 = 20.$

@ 10% level of significance $\alpha = 0.1$

Step 1:- Null hypothesis

$$H_0 : \mu_1 = \mu_2$$

i.e Both the samples are taken from same mean population.

Step 2:- Alternative hypothesis.

$$H_1 : \mu_1 \neq \mu_2$$

i.e Both the samples are not taken from same mean population (Two tailed test)

Step 3:- Critical value

$$\alpha = 0.1$$

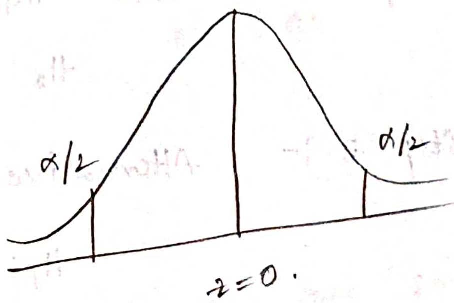
$$\alpha/2 = 0.05.$$

$$\alpha/2 = 0.5 - A(z_{\alpha/2})$$

$$A(z_{\alpha/2}) = 0.5 - 0.05$$

$$A(1.65) = 0.450$$

$$z_{\alpha/2} = 1.65$$



Step 4:- test statistic

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

$$= \frac{75 - 70}{\sqrt{\frac{225}{150} + \frac{400}{250}}}$$

$$z = 2.8398$$

$|z|$

Step 5 :- Conclusion

$$|z| = 2.8398 > 1.65 = z_{\alpha/2}$$

$$|z| > z_{\alpha/2}$$

Reject H_0 / Accept H_1

i.e. Both the samples are not taken from same mean population

MODEL :- 3.

Testing of hypothesis for single proportion

Suppose a large random sample of size 'n', as a sample proportion small 'p', where the sample has drawn from population having proportion 'P_0'. To test the hypothesis for single proportion.

Step 1 :- Null hypothesis

$$H_0 : P = P_0$$

Step 2 :- Alternative hypothesis

$$H_1 : P \neq P_0$$

Two tailed test

$H_1 : P > P_0$ - right tailed test

$H_1 : P < P_0$ - left tailed test

Step 3 :- Critical value

Step 4 :- Test statistic. $z = \dfrac{p - P_0}{\sqrt{\frac{PQ}{n}}}$ $\because P + Q = 1.$

Step 5 :- Conclusion

Note :- The confidence interval for population proportion P, for large sample @ 'α' level of significance $p - z_{\alpha/2}\sqrt{\frac{PQ}{n}} < P < p + z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

Maximum error

$$E = z_{\alpha/2} \sqrt{\frac{PQ}{n}}.$$

1. In a big city 325 men out of 600 were found to be smokers. Thus this information support the conclusion that the majority of men in the city are smokers.

Sol: Given that sample size

$n = 600 \geq 30$ large sample

Given that ~~3~~ 325 men are smokers out of 600

$p = 325/600$

$\boxed{p = 0.542.}$

Assume population proportion of smokers

$P_0 = \frac{1}{2}.$

Assume significance level 5% ; $\alpha = 0.05.$

Step 1:- Null hypothesis

$H_0 : P_0 = P_0 = \frac{1}{2}.$

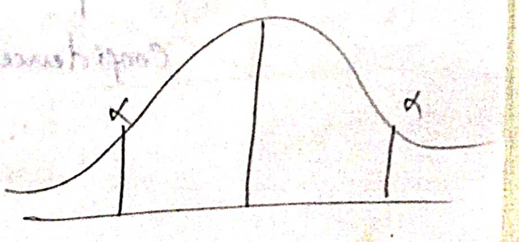
i.e Half of the men in the city are smokers

Step 2:- Alternative hypothesis

$H_1 : P > \frac{1}{2}.$

i.e Majority of the men in the city are smokers (Right-tailed test)

Step 3:- Critical value (Z_α)



$\alpha = 0.05$

$\alpha = 0.5 - A(z_\alpha)$

$A(z_\alpha) = 0.5 - 0.05$

$$\boxed{Z_\alpha = 1.65}$$

Step 4 :- test statistic

$$Z = \frac{\cancel{p} - P_0}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{\cancel{\tfrac{1}{2}}\ 0.542 - 0.5}{\sqrt{\frac{1}{600 \times 4}}} \qquad P = \frac{1}{2} \ ; \ P = \frac{1}{2}.$$

$$Z = 2.0575.$$

Step 5 :- conclusion

$$|Z| \simeq 2.06 > 1.65 = Z_\alpha$$

$$|Z| > Z_\alpha.$$

Accept H_1

$\therefore$ The majority of the men in the city are smokers. @ 5% level of significance.

2. Among 900 people in a state, 90 are found to be B. tech students construct 94% confidence interval for population / true proportion.

Sol:- Given that

Sample proportion of b. tech students

$$p = \frac{9\cancel{0}}{90\cancel{0}} = \frac{1}{10}.$$

Sample size $n = 900$

Confidence $(1-\alpha)\,100\% = 94\%.$

$$\therefore \quad 1-\alpha = 0.94$$

$$1 - 0.94 = \alpha$$

$$\boxed{\alpha = 0.06}$$

confidence interval for single proportion.

$$p - z_{\alpha/2}\sqrt{\frac{PQ}{n}} < P < p + z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

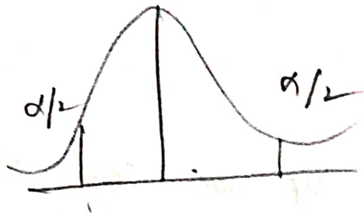
let $P = p = \frac{1}{10}$ $\quad Q = \frac{9}{10}$.

$\alpha/2 = 0.5 - A(z_{\alpha/2})$ $\qquad \alpha/2 = \frac{0.06}{2} = 0.03$

$A(z_{\alpha/2}) = 0.5 - 0.03$

$A(1.89) = 0.47$

$z_{\alpha/2} = 1.89$.



Confidence interval.

$$\frac{1}{10} - 1.89\sqrt{\frac{(0.1)(0.9)}{900}} < P < 0.1 + 1.89\sqrt{\frac{(0.1)(0.9)}{900}}$$

$0.0811 < P < 0.1189$

3. what can we say with 96% confidence about the maximum error if 50 students gets job among 90 students ?

A. Given.

sample size(n) = 90

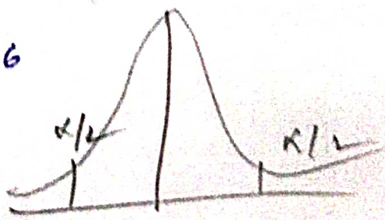
Sample proportion $= p = \frac{50}{90} = \frac{5}{9}$.

confidence $(1-\alpha)100 = 96\%$.

$(1-\alpha) = 0.\cancel{0}96$

$\alpha = 1 - 0.96$

$\alpha = 0.04$



$\alpha/2 = 0.02$.

$\alpha/2 = 0.5 - \cancel{0.02}\, A(z_{\alpha/2})$

$A(z_{\alpha/2}) = 0.5 - 0.02$.

$$A(z_{\alpha/2}) = 0.48$$

$$z_{\alpha/2} = 2.06.$$

$$\text{Max error} = \epsilon = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

$$= 2.06\sqrt{\frac{\frac{5}{9}\times\frac{4}{9}}{90}}$$

$$\epsilon = 0.1079.$$

4. If we can about with 93% confidence that the maximum error is 0.05 and $p = 0.3$. Find sample size ?

Sol:- Given

$$\epsilon = 0.05$$

$$p = 0.3 \qquad Q = 1-0.3$$

$$= 0.7$$

$$n = ?$$

$$\epsilon = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

$$n = \left[\frac{\cancel{(\epsilon/\times z_{\alpha/2})}}{\cancel{PQ}}\right]^2 \qquad n = \frac{(z_{\alpha/2})^2\, PQ}{(\epsilon)^2}$$

$$\text{confidence}\ (1-\alpha)100 = 93\%.$$

$$1-\alpha = 0.93$$

$$\alpha = 0.07$$

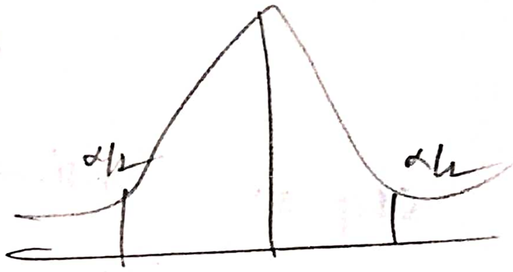
$$\boxed{\alpha/2 = 0.035}$$

$$\alpha/2 = 0.5 - A(z_{\alpha/2})$$

$$A(z_{\alpha/2}) = 0.5 - 0.035$$

$$A(1.7) = 0.465$$

$$z_{\alpha/2} = 1.7$$



$$n = \frac{[0.05(1.7)]^2}{(0.3)(0.7)}$$

$$n = \frac{(1.82)^2 (0.3 \times 0.7)}{(0.05)^2}$$

$$n = 278.7416.$$

$$\boxed{n \approx 278}$$

5. A coin is tossed 10,000 times and it turns up head 5195 times. Discuss whether the coin is biased or unbiased at $\alpha = 9\%$ level of significance

Sol: Given that

the sample proportion of getting heads

$$p = \frac{5195}{10000} = 0.5195$$

Sample size $n = 10,000$

Assume population proportion getting head

$$P = \frac{1}{2} \qquad Q = \frac{1}{2}$$

Given Significance level $\alpha = 9\%$.

Step 1:- Null hypothesis

$$H_0 : P = \frac{1}{2}.$$

i.e the given coin is unbiased

Step 2 :- Alternate hypothesis

$$H_1 : P \neq \frac{1}{2}$$

i.e the given ~~point~~ coin is biased

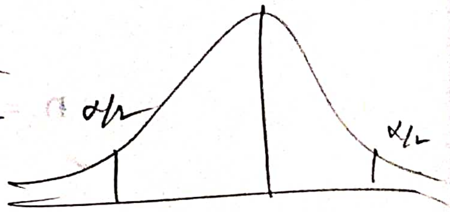
(Two tailed biased)

Step 3 :- critical value ($z_{\alpha/2}$)

$$\alpha/2 = 0.5 - A(z_{\alpha/2})$$

$$0.045 = 0.5 - A(z_{\alpha/2})$$

$$A(z_{\alpha/2}) = 0.5 - 0.045$$



$$A(1.69) = 0.455$$

$$\boxed{z_{\alpha/2} = 1.70}$$

Step 4 :- test statistics.

$$z = \frac{p - P_0}{\sqrt{\frac{PQ}{N}}}$$

$$= \frac{0.5195 - 0.5}{\sqrt{\frac{1}{10000 \times 4}}}$$

$$\boxed{z = 3.9}$$

Step 5 :- Conclusion

$$|z| = 3.9 \geq 1.7 = z_{\alpha/2}$$

$$|z| > z_{\alpha/2}$$

reject H_0

Accept H_1

The coin is biased

A random sample of 64 Oranges revealed that 14 oranges were bad. It is reasonable to assume that 20% of the oranges were bad? at 6% level.

Sample size $n = 64$ Oranges.

Sample proportion $p = \frac{14}{64} = 0.21875$.

Assume population proportion getting good

$P_0 = \cancel{\frac{1}{2}} 0.2 \quad Q = \cancel{\frac{1}{2}} 0.8$

$\therefore$ Assume that 20% are bad $= P_0 = \frac{20}{100} = 0.2$

Given significance level $\alpha = 0.06$

Step 1:- Null hypothesis

$H_0 : P = \cancel{\frac{1}{2}} 0.2$.

ie The oranges revealed are ~~good~~ bad

Step 2:- Alternate hypothesis

$H_1 : P \neq \cancel{\frac{1}{2}} 0.2$ (Two - tailed test)

The given oranges are not assumed to be bad.

Step 3:- Critical value $(z_{\alpha/2})$

$\alpha = 0.06$.

$\frac{\alpha}{2} = 0.03$.

$\alpha/2 = 0.5 - A(z_{\alpha/2})$

$A(z_{\alpha/2}) = 0.5 - 0.03$

$A(1.89) = 0.47$

$z_{\alpha/2} = 1.89$.

Step 4:- test statistic

$$z = \frac{P - P_0}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.21875 - 0.2}{\sqrt{\frac{0.2 \times 0.8}{64 \times 4}}}$$

$$= 0.375$$

Step 5 :- Conclusion.

$$|Z| = 0.375 < 1.89 = Z_{\alpha/2}$$

$$|Z| < Z_{\alpha/2}.$$

Reject H_1. Accept H_0

∴ The Oranges assumed that 20% of the oranges are bad.

7. A Manufacturer claimed thatleast 98% of audio equipments will meet the required specification. A sample of 500 equipments was tested and 30 were defective. Test his claim at a level of 2%.

Sol:- Given.

Sample size $n = 500$

Sample proportion $p = \frac{30}{500} = 0.06$.

The audio equipments are claimed atleast 98%

i.e $P = \frac{98}{100} = 0.98$ $Q = 0.02$

Significance level $\alpha = 0.02$.

Step 1 :- Null hypothesis

$H_0 : P = 0.98$

i.e the Manufactures claimed ~~atleast~~ is equal 98% of audio equipments step 2:- Alternate hypothesis

$H_1 : P \geq 0.98$

i.e the manufacturer claimed atleast 98%. of audio equipments (Right tailed test)

step 3:- Critical value Z_α.

$\alpha = 0.02.$

$\alpha = 0.5 - A(Z_\alpha)$

$A(Z_\alpha) = \cancel{\alpha} 0.5 - \alpha$

$= 0.5 - 0.02$

$A(2.06) = 0.480$

$$\boxed{Z_\alpha = 2.06}$$

step 4:- test statistic.

$$Z = \frac{p - P_0}{\sqrt{PQ/n}}$$

$$= \frac{0.06 - 0.98}{\sqrt{\frac{(0.98)(0.02)}{500}}}$$

$Z = -146.94$

step 5:- conclusion

$|Z| = 146.94 > 2.06 = Z\alpha/2$

$|Z| > Z\alpha/2$ H_0 is rejected.

i.e the manufacturer claimed is alteast 98%. of audio equipments

8. Electric bulbs company claims that the percentage defectives in his product doesn't exceed '6'. A sample of 40 bulbs is found to contain '5' defectives. If would you consider the claims justified.

Sol:- Given

Sample size $n = 40$.

Sample proportion $p = \frac{5}{40} = 0.125$

The ~~and~~ % of defectives in his product doesn't exceed '6'

i.e $P = \frac{6}{100} = 0.06$

Assume the level of significance is 5%.

$\alpha = 0.05$.

Step 1 :- Null hypothesis

$H_0 : P = 0.06$.

i.e The % of defectives in his product is equal to 6.

Step 2 :- Alternate hypothesis

$H_1 : p \not< 0.06$

i.e the % of defectives in his product must not exceed '6'. (One tail test)

Step 3 :- Critical value (z_α)

$\alpha = 0.05$

$\alpha = 0.5 - A(z_\alpha)$

$A(z_\alpha) = 0.5 - 0.05$

$A(1.65) = 0.450$

$\boxed{z_\alpha = 1.65}$

Step 4:- Test statistic.

$$z = \frac{P - P_0}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.125 - 0.06}{\sqrt{\frac{0.06 \times 0.94}{40}}}$$

$$z = 1.7310$$

Step 5:- Conclusion

$$|z| = 1.7310 > 1.65 = (z_\alpha).$$

$$|z| > (z_\alpha)$$

$\therefore$ Reject H_0

i.e Accept H_0

i.e the % of defectives in his product is equal to '6' and must exceed 6 also @ 5% level of significance.

MODEL- 4 :- Test of hypothesis for difference of proportions

Let p_1 and p_2 with proportions of 2 Large samples with sample sizes n_1 and n_2 are drawn from population having proportions P_1 and P_2.

To test significance difference b/w population proportions take $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Where

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

$$q = 1 - p.$$

1. In a city 'A' 20% of a random sample 900 school boys have a health problem ~~out~~ of. In another city 'B' 18.5% of random sample 1600 school boys have a same effect. Is the difference b/w proportions are significant @ 3% level.

Given that from city 'A'

20% having health problem $p_1 = \frac{20}{100} \times 900 = 180$

$$p_1 = \frac{20}{100} = 0.2.$$

$$n_1 = 900$$

In city 'B' 18.5% having heath problem $p_2 = \frac{18.5}{100} = 0.185$

$$n_2 = 1600.$$

Significance level $\alpha = 0.03$.

Assume P_1, P_2 are proportion of students hav[e] math problem from city A and city B. respectively.

Step 1:- Null hypothesis

$$H_0: P_1 = P_2$$

i.e no. significant difference

Step 2:- Alternate hypothesis

$$H_1: P_1 \neq P_2.$$

i.e There significance difference

(Two tailed test)

Step 3:- Critical value $(z_{\alpha/2})$

$$\alpha = 0.03$$

$$\alpha/2 = 0.015$$

$$\alpha/2 = 0.5 - A(z_{\alpha/2})$$

$$A(z_{\alpha/2}) = 0.5 - 0.015$$

$$A(2.17) = 0.485$$

$$\boxed{z_{\alpha/2} = 2.17}$$

Step 4:- Test statistic

$$z = \frac{P_1 - P_2}{\sqrt{pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$p = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2}$$

$$= \frac{0.2 - 0.185}{\sqrt{(0.19)(0.81)\left[\frac{1}{900} + \frac{1}{1600}\right]}}$$

$$p = 0.19$$

$$q = 0.81$$

$$z = 0.9176$$

Step 5 :- Conclusion.

$$|Z| = 0.9176 \not> 2.17 = Z_{\alpha/2}$$

$$|Z| < Z_{\alpha/2}$$

H_0 is accepted.

There is no significant difference @ 3% level of significance

2. A Machine puts out 21 defective articles in a sample of 500 articles another machine gives 3 defective articles in a sample of 100 articles. Are the two machines differente significantly in their performance

Sol:- Given.

Sample proportion of defective articles in

Machine 1 $p_1 = \frac{21}{500}$

$= 0.042$

Sample size $= n_1 = 500$.

Sample proportion of defective articles in

Machine 2 $p_2 = \frac{3}{100}$

$= 0.03$

Sample size $n_2 = 100$

By default, significance level is $\alpha = 0.05$

Step 1 :- Null hypothesis

$H_0 : P_1 = P_2$.

No significance difference

Step 2 :- Alternate hypothesis $H_1 : P_1 \neq P_2$

i.e there is significance difference

Step :- 3 :-
critical value $(z_{\alpha/2})$

$$\alpha/2 = 0.5 - A(z_{\alpha/2})$$

$$A\{z_{\alpha/2}\} = 0.5 - 0.025$$

$$A(1.96) = 0.475$$

$$z_{\alpha/2} = 1.96.$$

step :- 4 :- test statistic

$$z = \frac{p_1 - p_2}{\sqrt{pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

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

$$= \frac{500(0.042) + 100(0.03)}{600}$$

$$p = 0.04$$

$$q = 0.96$$

$$= \frac{0.042 - 0.03}{\sqrt{(0.04)(0.96)\left[\frac{1}{500} + \frac{1}{600}\right]}}$$

$$z = \cancel{0.22360}\ 0.5590$$

step 5 :- Conclusion.

$$|z| = \overset{0.5590}{\cancel{0.22360}} \not< 1.96 = z_{\alpha/2}$$

$$|z| \not< z_{\alpha/2}$$

H_0 is ~~rejected~~. accepted.

∴ There is no significance difference @ 5% Level of significance

Unit-V Small Samples

Test of Significance
A very important aspect of the sampling theory is the study of tests of significance which enable us to decide on the basis of the sample results if

(i) The deviation b/w the observed sample statistic and the hypothetical parameter value is significant.

(ii) The deviation b/w two sample statistics is significant.

Test of significance for small samples;

Def: When the size of the sample (n) is less than 30, then that sample is called small sample.
i.e., when $n<30$, the sample is called a small sample

For "expensive!" population such as ~~statistic~~ satelites, aeroplanes, nuclear reactors, super computers, etc. $n \geq 30$ (uneconomical) (impracticable) (Time consuming)

In all such cases, the size of the sample, drawn is small (i.e., $n<30$)

The following some important tests for small samples
(i) Student's 't' test (Sir William Gossett 1905)
(ii) Snedecor's F-Test (R.A. Fisher 1926)
(iii) χ^2-Test (Karl Pearson)

Scanned by TapScanner

Student's 't' - Test:

The student's t is defined by the statistic

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

where $\bar{x} = \frac{\Sigma x_i}{n}$ = sample mean.

μ = Population mean

n = sample size

$(n-1)$ degree of freedom.

$$S.D. \text{ 'S' } = \sqrt{\frac{1}{n-1} \Sigma (x_i - \bar{x})^2}$$

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

Confidence or Fiducial limit for μ.

(i) If $t_{0.05}$ is the table value of t for $(n-1)$ dof at 5% los then 95% confidence limit for μ is given by $\boxed{\bar{x} \pm t_{\frac{0.05}{2}} \frac{S}{\sqrt{n}}}$ 0.025

(ii) 99% Confidence limit

$$\boxed{\bar{x} \pm t_{\frac{0.01}{2}} \frac{S}{\sqrt{n}}}$$ 0.005

(general $\bar{x} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$)

(iii) Maximum Error (E) $t_{\alpha/2} \frac{S}{\sqrt{n}}$

Application of t-distribution:

(1) To test the significance of the mean of a small random sample from a normal population (diff b/w two means)

(2) To test the significance of a Rank correlation ... co-effici...

Applications of t-distribution:

1) The t-distribution to test the significance of the sample mean when population variance is not given.

2) To test the significance of the mean of the sample i.e., to test if the sample mean differs significantly from the population mean.

3) To test the significance of the difference b/w two small means or to compare two samples.

4) To test the significance of an observed sample correlation coefficient and sample regression coefficient.

5) The deviation b/w the observed sample statistic and the hypothetical parameter value is significant.

6) Deviation b/w two small sample statistic is significant

Assumption for the students t-test:

1) Sample size $n<30$

2) Parent population from which sample is drawn is normal.

3) The population standard deviation is unknown

4) The sample observations are independent i.e; the sample is random.

where $\bar{x} = \frac{}{n}$

Testing of hypothesis about t-test for single mean.

Let a random sample of size 'n' ($n<30$), has a sample mean $\bar{x}$ to test the hypothesis that the population mean μ has a specified value μ_0 when population standard deviation σ is not known.

Step 1: N.H.: $\mu = \mu_0$

Step 2: A.H.: $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$

Step 3: L.O.S: α with degrees of freedom $\nu = n-1$

Step 4: Case (i), Test of significance $t_{cal} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

s small

(If the sample standard deviation is given directly to find the test statistic then we can use, t_{cal})

Case (ii) If the sample standard deviation is not given directly to find the test statistic we can use the following formula

$$t_{cal} = \frac{\bar{x}-\mu}{S/\sqrt{n}} \quad \text{where } S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{(n-1)}}$$

Capital

Step 5: Conclusions $|t_{cal}| > |t_{tab}|$ Reject H_0

... pool from ...

(11) (1) 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 hrs. Is the sample not upto the standards.

30

5.1%.

Sol^n: sample size $n = 26 < 30$ (small samples)

Sample mean $\bar{x} = 990$

Population mean $\mu = 1000$ and S.D. $s = 20$

Degrees of freedom $= n-1 = 26-1 = 25$

Here we know $\bar{x}$, μ, S.D. and n.

$\therefore$ we use students 't' test.

1) N.H. $H_0: \mu = 1000$
2) A.H. $H_1: \mu < 1000$ (L.T.T.
3) L.of.S. $(\alpha) = 0.05$
4) T.S. $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}} = -2.5$ $\quad |t| = 2.5$.

Tabulated value of 't' at 5%. level with 25 degrees of freedom for L.t.t. is 1.708

$|t| > t_\alpha$ Reject H_0.

(III) The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?

Sol^n: $n = 14$, $\bar{x} = 17.85$, $S.D = 1.955$, $\mu = 18.5$

$\mu \neq 18.5$

$\mu = 18.5$, D of $n-1 = 13$

$|t| = \underline{1.199}$

Accept

(IV) H.w A sample of 15 members has a mean 67 and S.D. 5.2. Is this sample has been taken from a large population of mean 70.

$n = 15$, $\bar{x} = 67$, $\sigma = S.D.\ 5.2$

$\mu = 70$

(1) A random sample of six steel beams has a mean compressive strength of 58,392 p.s.i. (pounds per square inch) with a standard deviation of 648 p.s.i. Use this information and the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality.

Sol: We have n = sample size (no. of steel beams) = 6 < 30 (S.S).

$\bar{x}$ = sample mean (Avg S) = 58,392 p.s.i

s = S.D of six beams = 648 psi.

Degree of freedom (dof) = $n-1 = 6-1 = 5$

σ is unknown and $n < 30$.

Hence we use t-distribution.

① Null hypothesis $H_0: \mu = 58000$

② Alternative hypothesis $H_1: \mu \neq 58000$. (Two-Tailed)

③ level of significance: $\alpha = 0.05$

④ Test statistic $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} = \dfrac{58,392 - 58000}{648/\sqrt{5}}$

$t = 1.482$ 1.353

$|t| = 1.482$

$t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.57$

from table with dof

② A random sample from a company very extensive files shows that the orders for a certain kind of machinery were filled, respectively in 10, 12, 19, 14, 15, 18, 11 and 13 days. Use the level $\alpha = 0.01$ to test the claim that on the avg such orders are filled in 10.5 days. Choose the alternative hypothesis so that rejection of null hypothesis $\mu = 10.5$ days implies that it takes longer than indicated.

Sol^n:- We have (S.S), $n = 8$ (small S)

$$(S.M)\ \bar{x} = \frac{\sum x_i}{n} = \frac{10+12+19+\cdots+13}{8} = \frac{112}{8} = 14$$

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

$$= \frac{(10-14)^2 + (12-14)^2 - \cdots (13-14)^2}{8-1}$$

$$s^2 = \frac{72}{7} = 10.286$$

$$s = 3.207$$

σ unknown and $n < 30$ Then we use t-distn

① N.H: $H_0 = \mu = 10.5$ day

② A.H: $H_1 = \mu > 10.5$ (R.T.T)

③ L.of S. = $\alpha = 0.01$

④ $T.S = t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{14 - 10.5}{3.207/\sqrt{8}} = 3.087$

⑤ Critical Region: $t_\alpha = t_{0.01} = 2.998$ (From table
dof $n-1 = 7$

Scanned by TapScanner

$\therefore \quad |t| > t_{\alpha}$.

$\therefore$ Null hypothesis is rejected.

(3) Five measurements of tar content of a certain kind of cigarettes yielded 14.5, 14.2, 14.4, 14.3 and 14.6 mg per cigarette. Show that the mean of the sample is $\bar{x} = 14.4$ and the avg tar claimed by the manufacturer $\mu = 14.0$ is significant $\alpha = 0.05$

Sol^n: Given $n = 5 < 30$ (S.S)

$\bar{x} = 14.4$

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{(14.5-14.4)^2 + \cdots + (14.6-14.4)^2}{5-1}$$

$$s^2 = 0.025$$

$$S = 0.158$$

(1) N.H: $H_0 = \mu = 14.0$

(2) A.H. $H_1 = \mu \neq 14.0$ (Two-Tailed)

(3) L.O.S $= \alpha = 0.05$

(4) T.S $= \quad t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{14.4 - 14.0}{0.158/\sqrt{5}} = 5.66$

(5) Critical Region $t_{\alpha/2} = t_{\frac{0.05}{2}} = t_{0.025}$ [dof 4]

$t_{0.025} = 2.776$

$|t| > t_{\alpha/2}$

Null Hypothesis is rejected.

* A sample of size 10 was taken from a population S.D of sample is 0.03. Find the maximum error with 99% confidence

Sol^n:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 3.25 \times \frac{0.03}{3} = 0.0325$$

H.W ① A mechanist is making engine part with axle diameter of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch a S.D. of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification.

Hint $\bar{x} = 0.742$ $\quad t = 3.15$

μ of 0.700 $\quad t_{\alpha/2} = 2.26$.

H_0 is rejected

② It has been suggested that a college teacher in AnPradesh spends an avg of less than 10 hrs in a week on his own academic shedule. The figures for the time spent during a week are given bellow for 12 teachers: 7.1, 13.1, 7.8, 3.6, 8.4, 4.9, 9.6, 3.4, 0.1, 7.2, 20.3, 11.1. Is the claim justified with level of significance of 0.05? Hint $n = 12$, $\mu = 10$

$\bar{x} = \frac{\Sigma x}{n} = 8.05$

$s^2 = 27.663 \Rightarrow s = 5.260$

$t = -1.285 \quad |t| = 1.285$

$t_{\alpha/2} = 2.201$.

H_0 is accepted

③ A random sample of 10 bags of pesticides are taken whose weights are 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 (in kgs) Test whether the avg paking can be taken to be 50 kgs.

Hint $\bar{x} = 47.2$, $s^2 = 7.122 \Rightarrow s = 2.66$, $t = 3.6$, $t_{\alpha/2} = 2.26$

H_0 should be rej

Pbm 2 A sample of 10 cam shafts intended for use in gasoline engines has an average eccentricity of 1.02 and a standard deviation of 0.044 inches. The data may be treated a random sample from a normal population. Determine a 95% confidence interval for the actual mean eccentricity of the cam shaft.

Sol^n:- The confidence interval is $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$

$n = 10$

$s = 0.044$ inches

$t_{\frac{0.05}{2}=0.025}$ (for 9 dof) $= 2.26$ (From tables of t-distribution)

$\bar{x} = 1.02$ inches

$$C.I. \text{ is } \left(1.02 - 2.26 \times \frac{0.044}{\sqrt{10}},\ 1.02 + 2.26 \times \frac{0.044}{\sqrt{10}}\right)$$

$$= (0.99,\ 1.05)$$

H.W. 2 A random sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain the 95% confidence limits of the mean of the population.

Hint $n = 16$

$\bar{x} = 53$

Imp ** $\sum (x_i - \bar{x})^2 = 150$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{150}{15} = 10$

$\mu = 56$

Ans:- $(51.31,\ 54.68)$

Student's t-test for difference of means:-

To test the significance difference b/w two means $\bar{x}_1$ and $\bar{x}_2$ of samples of sizes n_1 and n_2, we use statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ where } s^2 = \frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}$$

(or) $$s^2 = \frac{n_1 \cdot s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

where s_1 and s_2 are sample standard deviation

Degrees of freedom is n_1+n_2-2.

Note: If the difference b/w the means is δ then the test statistic is $$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}.$$

Prob ① Find the maximum difference that we can expect with probability 0.95 b/w the means of samples of sizes 10 and 12 from a normal population if their standard deviations are found to be 2 and 3 respectively.

Sol:- We have

$$n_1 = 10, \quad n_2 = 12, \quad s_1 = 2, \quad s_2 = 3$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{10(2^2)+12(3^2)}{10+12-2} = \frac{40+108}{20} = 7\cdot$$

1. Null Hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

3. Level of significance, $\alpha = 0.05$

4. Test statistic is $t = \dfrac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$|\bar{x} - \bar{y}| = |t| \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (2.086)(2.72)\sqrt{\frac{1}{10} + \frac{1}{12}}$$

(∵ Tabulated value of t for $10+12-2 = 20$ d.o.f at 5% LOS is 2.086.)

$$= (5.674)(0.428)$$

$$= 2.43$$

Hence the max. difference b/w the means is 2.43

Prob ② Measuring specimens of nylon yarn, taken from two machines, it was found that 8 specimens from first machine had a mean denier of 9.67 with a standard deviation of 1.81 while 10 specimens from second machine had a mean denier of 7.43 with a standard deviation of 1.48. Assuming that the proportions are normal, test the hypothesis $H_0: \mu_1 - \mu_2 = 1.5$ against $H_1: \mu_1 - \mu_2 > 1.5$ at 0.05 level of significance

Soln:- $n_1 = 8$, $n_2 = 10$,

$\bar{x} = 9.67$ $\quad \bar{y} = 7.43$

$s_1 = 1.81$, $s_2 = 1.48$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(1.81)^2 + 10(1.48)^2}{8 + 10 - 2}$$

$$s^2 = 3$$

$$s = 1.734$$

Scanned by TapScanner

1. Null Hypothesis $H_0: \mu_1 - \mu_2 = 1.5 = (\delta)$

2. Alternative Hypothesis $H_1: \mu_1 - \mu_2 > 1.5$

3. Level of significance, $\alpha = 0.05$

4. The test statistic is $t = \dfrac{(\bar{x} - \bar{y}) - \delta}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$= \frac{(9.67 - 7.43) - 1.5}{(1.734)\sqrt{\frac{1}{8} + \frac{1}{10}}}$$

$$= 0.9$$

Tabulated value of t for $8 + 10 - 2 = 16$ d.o.f at 5% level of significance is 1.746. Since t < tabulated, we accept H_0.

(3) Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	–

Test whether the two horses have the same running capacity.

Sol^n:- Given $n_1 = 7$, $n_2 = 6$

We first compute the sample mean and standard deviation.

$\bar{x}$ = Mean of 1^{st} sample $= \dfrac{28 + 30 + \cdots + 34}{7} = 31.286$

$\bar{y}$ = Mean " 2^{nd} " $= \dfrac{29 + 30 + \cdots + 29}{6} = 28.16$

Scanned by TapScanner

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219		31.4358	169		26.8336

Now, $$s^2 = \frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}$$

$$= \frac{31.4358 + 26.8336}{11} = 5.29$$

$$s = 2.3$$

1. Null Hypothesis $H_0: \mu_1 = \mu_2$
2. Alternative hypothesis $H_1: \mu_1 \neq \mu_2$
3. Level of signi : $\alpha = 0.05$
4. Test statistic $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{31.286 - 28.16}{(2.3)\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.443$

Tabulated value of t for $7+6-2 = 11$ dof at 5% los is 2.2 since calculated $t >$ Tabulated t, we reject the null hypothesis H_0. That is, both Horses A and B do not have the same running capacity.

(4) To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows:

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of significance of 0.05.

Sol^n: We have

$n_1 = 10$, $n_2 = 10$ and

$$\bar{x} = \frac{117+105+\cdots+107}{10} = 103$$

$$\bar{y} = \frac{106+98+\cdots+85}{10} = 95.8$$

Now we compute the standard deviations of both samples

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
117	14	196	106	10.2	104.04
105	2	4	98	2.2	4.84
97	-6	36	87	-8.8	77.44
105	2	4	104	8.2	67.24
123	20	400	116	20.2	408.04
109	6	36	95	-0.8	0.64
86	-17	289	90	-5.8	33.64
78	-25	625	69	-26.8	718.24
103	0	0	108	12.2	148.84
107	4	16	85	-10.8	116.64
1030		1606	958		1679.6

$s^2 = 182.51$

$s = 13.51$

1. Null hypothesis $H_0: \mu_1 = \mu_2$ (i.e, no difference in I.Q.)

2. Alternative Hypothesis $H_1: \mu_1 > \mu_2$ (i.e, husbands are more intelligent than wives) one tailed test, right)

3. LofS: $\alpha = 0.05$

4. Test statistic: $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{103 - 95.8}{(13.51)\sqrt{\frac{1}{10} + \frac{1}{10}}} = 1.19168$

Since $t_{cal} = 1.19168 < t_{tab} = 1.734$, we accept the null hypothesis is H_0 i.e., There is no difference in I.Q's

HW① The soldiers participated in a shooting competition in the first week. After intensive training they participated in the competition in the second week. Their scores before and after training are given as follows:

scores before	67	24	57	55	63	54	56	68	33	43
score after	70	38	58	58	56	67	68	75	42	38

Do the data indicate that the soldiers have been benefited by the training.

Ans: $s^2 = 197$
$s = 14.0357$
$|t| = 0.796$

Alt. $\mu_1 - \mu_2 < 0$ left tailed

$t_{tab} = 1.734$ H_0 accepted

② To compare two kinds of bumper guards, 6 of each kind were mounted on a car and then the car was run into a concrete wall. The following

F-Test:-

To test whether there is any significant difference b/w. two estimates of population variance

(or)

To test if the two samples have came from the same population, we use F-Test.

In this case we set up null hypothesis $H_0: S_1^2 = S_2^2$

i.e., population variances are same.

Under H_0, the test statistic is $\cancel{F = \frac{S_1^2}{S_2^2}}$

where $S_1^2 = \frac{\Sigma (x_1 - \bar{x})^2}{n_1 - 1}$ [n_1 - First sample size]

$S_2^2 = \frac{\Sigma (y_1 - \bar{y})^2}{n_2 - 1}$ [n_2 - Second sample size]

and $S_1^2 > S_2^2$.

The degrees of freedom are $\nu_1 = n_1 - 1$; $\nu_2 = n_2 - 1$

Notes: ① We will take greater of the variances S_1^2 (or) S_2^2 in the numerator and adjust for the degree of freedom accordingly i.e., $\boxed{F = \frac{\text{Greater variance}}{\text{Smaller variance}}}$

$F = \frac{S_1^2}{S_2^2}$ if $S_1^2 > S_2^2$; $F = \frac{S_2^2}{S_1^2}$ if $S_2^2 > S_1^2$

② If sample variance s^2 is given, we can obtain population variances S^2 by using the relation $ns^2 = (n-1)S^2$

Prob ① The measurements of the output of two
Imp** units have given the followings results. Assuming
that both samples have been obtained from the normal
populations at 5% significant level, test whether
the two population have the same variance

Unit - A	14.1	10.1	14.7	13.7	14.0
Unit - B	14.0	14.5	13.7	12.7	14.1

Sol:- $n_1 = 5,\ n_2 = 5$

Now $\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{1}{5}(14.1 + 10.1 + \cdots + 14.0) = 13.32$

$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{1}{5}(14.0 + 14.5 + \cdots + 14.1) = 13.8$

Computing standard deviations of the samples

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
14.1	0.78	0.6084	14.0	0.2	0.04
10.1	−3.22	10.3684	14.5	0.7	0.49
14.7	1.38	1.9044	13.7	−0.1	0.01
13.7	0.38	0.1444	12.7	−1.1	1.21
14.0	0.68	0.4624	14.1	0.3	0.09
66.6		13.488	69		1.84

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{13.488}{4} = 3.372$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{1.84}{4} = 0.46$$

① $H_0: S_1^2 = S_2^2$

② $H_1: S_1^2 \neq S_2^2$

③ $S_1 > S_2$ $\quad F = \frac{S_1^2}{S_2^2} = \frac{3.372}{0.46} = 7.3$

④ 5

Conclusion: $F_{\alpha/2}$ $F_{0.05} = F_{.05}(4,4) = 6.39$

$F = 7.3 \quad F_{\alpha/2} = 6.39$

$|F| > F_{\alpha/2}$

H_0 is rejected

$\therefore$ There is significant difference b/w the variances

(2) The time taken by workers in performing a job by method I and method II is given below:

method I:	20	16	26	27	23	22	—
method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

$n_1 = 6, \; n_2 = 7 \quad \bar{x} = \frac{\Sigma x_i}{n_1} = 22.3 \quad \bar{y} = \frac{\Sigma y_i}{n_2} = 34.4$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
—			38	3.6	12.96
134		81.34	241		133.72

$S_1^2 = \frac{87.34}{5} = 16.26$ and $S_2^2 = \frac{133.72}{6} = 22.29$

Null hypothesis H_0: There is no significant difference btw the variances. Since $S_2^2 > S_1^2$ we use the statistic

$$F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.268} = 1.3691 \approx 1.37$$

Tabulated value $F_{0.05}\,(\overset{\nu_2}{6}, \overset{\nu_1}{5}) = 4.95.$

$|F| < F_\alpha$

$\therefore H_0$ is accepted

There is no significant difference btw the variances of the time distribution by the workers.

③ It is known that the mean diameters of rivets produced by two firms A and B are practically the same but the standard deviations may differ. For 22 rivets produced by firm A, the S.D. is 2.9 mm, while for 16 rivets manufactured by firm B, the S.D. is 3.8. Compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance.

Soln.: Given $n_1 = 22,\ n_2 = 16$

$s_1 = 2.9,\ s_2 = 3.8$

$$n_1 s_1^2 = (n_1 - 1) S_1^2$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{22(2.9)^2}{21} = 8.81$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{16(3.8)^2}{15} = 15.4$$

① Null hypothesis $H_0: S_1^2 = S_2^2$

② Alt " $H_1: S_1^2 \neq S_2^2$ (T.T.T)

Scanned by TapScanner

③ LOS $\alpha = 0.05$

$$F = \frac{S_2^2}{S_1^2} = \frac{15.4}{8.81} = 1.74$$

$F_{0.05}(v_2, v_1) =$ $v_1 = n_1 - 1 = 22 - 1 = 21$ smaller means
$v_2 = n_2 - 1 = 16 - 1 = 15$ means Greater variance

$F_2 (15, 21)$ is 2.18

$|F| < F_2$

$\therefore$ Null Hyp. is accepted.

④ Two random samples gave the following results.

Sample	size	Sample mean	Sum of squares of Deviations from the Mean
1	10 m	15 $\bar{x}$	90 $\Sigma(x_i - \bar{x})^2$
2	12 n	14 $\bar{y}$	108 $\Sigma(y_i - \bar{y})^2$

Test whether the samples came from the same normal population.

Soln:- Note: To test the whether the samples same from same normal population, we have to check whether the given data satisfies student's t-test and F-test

① students t-test:

Null hypothesis: $H_0: S_1 = S_2$

Alt hypothesis $H_1: S_1 \neq S_2$

LOS : 0.05

$t = \frac{\bar{x} - \bar{y}}{\quad}$

Scanned by TapScanner

where $S^2 = \dfrac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2}$

$$S^2 = \frac{90+108}{10+12-2} = 9.9$$

$$S = 3.14$$

$$t = \frac{15-14}{3.14\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.74$$

$$t_{\alpha/2} = t_{0.025} = 2.086$$

d.o.f at 20

$$|t| < t_{\alpha/2}$$

$\therefore H_0$ is accepted.

F-test :

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

$S_2^2 = \dfrac{\Sigma(y_i-\bar{y})^2}{n_2-1}$, $S_1^2 = \dfrac{\Sigma(x_i-\bar{x})^2}{n_1-1}$

$= \dfrac{108}{11} = 9.8$ $\quad = \dfrac{90}{9} = 10$

$$F = \frac{10}{9.8} = 1.02$$

$n_1 = 10$ $\quad v_1 = n_1 - 1 = 9$

$n_2 = 12$ $\quad v_2 = n_2 - 1 = 11$

$F_\alpha = F_{0.05}(v_1, v_2) = 2.90$

9, 4

$\therefore F < F_\alpha$

H_0 is accepted.

H.W. ① Pumpkins were grown under two experimentally conditions. Two random samples of 11 and 9 pumpkins show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal. $S^2 = 0.704$ $\quad F = 2.5$

Model-4: Chi-square (χ^2) test

If set of events $A_1, A_2, \ldots A_n$ are having the observed frequencies $O_1, O_2 \ldots O_n$ respectively then According to probability rules $A_1, A_2 \ldots A_n$ are having the expected frequencies $E_1, E_2, \ldots E_n$. then to check goodness of to fit the data (χ^2) test is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

with $v = (n-1)$ degrees of freedom

i) Chi-square test as test of goodness of fit:

To test the difference between theoritical value and observed value.

Step 1: Null hypothesis

$$H_0 : O_i = E_i$$

There is no significance difference between observed value and expected value. Data is goodness of fit

Step 2: Alternate hypothesis

$$H_1 : O_i \neq E_i$$

There is difference between observed value and expected value

Scanned by TapScanner

Step 3 : Critical value

(χ^2_α) with $v = n-1$ degrees of freedom

Step 4 : χ^2 test

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Step 5 : Conclusion

Therefore $\chi^2 < \chi^2_\alpha$ accept H_0

1) The no. of accidents per week in a certain community as follows 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. The accident conditions were the same during this 10 week period.

total no. of accidents in 10 weeks = 100

expected accidents per week $E_i = \frac{100}{10} = 10$

observed accidents (O_i)	Expected accidents (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	10
2	10	-8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	-4	16	1.6
9	10	-1	1	0.1
4	10	-6	36	3.6

$$\sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} = 26.6$$

Step 1: Null hypothesis

$H_0 : O_i = E_i$

The accident conditions were same during this 10 week period.

Step 2: Alternate hypothesis:

$$H_1 : O_i \neq E_i$$

The accident conditions were not same during this 10 week period

Step 3: Critical value

$$\alpha = 5\% = 0.05$$

χ^2_α with $v = (n-1)$ degrees of freedom

$$\chi^2_\alpha = \chi^2_{0.05} = 16.919 \text{ with } v = 9.$$

Step 4:- χ^2 test

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 26.6$$

Step 5: Conclusion

$$\chi^2 = 26.6 > 16.919 = \chi^2_\alpha$$

$$\chi^2 > \chi^2_\alpha \text{ (Reject } H_0)$$

Therefore accident conditions are not same in 10 weeks at 5% significant level.

2) A pair of dice are thrown 360 times and frequency of each sum is given below

sum	2	3	4	5	6	7	8	9	10	11	12
frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are ~~failed~~ fair on the basis of the χ^2 test at 1% level of significance.

If two dice are thrown total no. of outcomes $= 6\times 6 = 36$

$$\left\{\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}\right\}$$

i) Probability of getting sum is "2" when '2' dice are thrown $P(X=2) = \frac{1}{36}$

No. of times getting sum is '2' $= N \times P(X=2)$

$$= 360 \times \frac{1}{36} = 10$$

||ly we get remaining values.

Scanned by TapScanner

$X = r$	O_i's	Probability	No. of times getting sum (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
2	8	$\frac{1}{36}$	$\frac{1}{36} \times 360 = 10$	-2	4	0.4
3	24	$\frac{2}{36}$	$\frac{2}{36} \times 360 = 20$	4	16	0.8
4	35	$\frac{3}{36}$	$\frac{3}{36} \times 360 = 30$	5	25	0.83
5	37	$\frac{4}{36}$	$\frac{4}{36} \times 360 = 40$	-3	9	0.225
6	44	$\frac{5}{36}$	$\frac{5}{36} \times 360 = 50$	-6	36	0.72
7	65	$\frac{6}{36}$	$\frac{6}{36} \times 360 = 60$	5	25	0.42
8	51	$\frac{5}{36}$	$\frac{5}{36} \times 360 = 50$	1	1	0.02
9	42	$\frac{4}{36}$	$\frac{4}{36} \times 360 = 40$	2	4	0.1
10	26	$\frac{3}{36}$	$\frac{3}{36} \times 360 = 30$	-4	16	0.53
11	14	$\frac{2}{36}$	$\frac{2}{36} \times 360 = 20$	-6	36	1.8
12	14	$\frac{1}{36}$	$\frac{1}{36} \times 360 = 10$	4	16	1.6

Step 1: Null hypothesis

$H_0 : O_i = E_i$

The pair of dice are fair

Step 2: Alternate hypothesis

$H_1 : O_i \neq E_i$

The pair of dice are not fair

Step 3: Critical value

given $\alpha = 1\% = 0.01$

$\chi^2_{0.01} = 23.209$ with $\alpha = 10$

Step 4: χ^2 test

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i} =$$

$$\chi^2 = 7.44$$

$$\chi^2 = 7.44 < 23.209 \text{ (i.e)}$$

$$\chi^2 < \chi^2_\alpha$$

16/3/18

ii) Chi-square test for independent of atributes

Definition: An atribute means quality or characteristic.

Ex:- Atributes are honesty, strength, behaviour, beauty, blindness, smoking, drinking, eating. etc.,

Let us consider two atributes 'A' & 'B'. 'A' is divided into two classes, 'B' divided into two classes.

A	a	b
B	c	d

$$\text{Expected frequency } E_i = \frac{\text{Row total} \times \text{column total}}{\text{grand total}}$$

Contingency table for expected frequency:

A $\quad E(a) = \dfrac{(a+b)(a+c)}{N} \qquad E(b) = \dfrac{(a+b)(b+d)}{N}$

B $\quad E(c) = \dfrac{(c+d)(a+c)}{N} \qquad E(d) = \dfrac{(c+d)(b+d)}{N}$

Note:- We calculate critical value χ^2_α at

$\nu = $ (No. of rows -1) (No. of columns -1)

$= (r-1)(c-1)$

1) Given following table for hair colour and eye colour. Is there good relation between two attributes at two % level.

		Hair colour			
		Fair	Brown	Black	Total
Eye colour	Blue	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

ontingency table for expected values

Hair colour / Eye colour	Fair	Brown	Black
Blue	$E(15) = \frac{40 \times 60}{150}$ $E(15) = 16$	$E(5) = \frac{40 \times 30}{150}$ $E(5) = 8$	$E(20) = \frac{40 \times 60}{150}$ $E(20) = 16$
rey	$E(20) = \frac{50 \times 60}{150}$ $E(20) = 20$	$E(10) = \frac{50 \times 30}{150}$ $E(10) = 10$	$E(20) = \frac{50 \times 60}{150}$ $E(20) = 20$
rown	24	12	24.

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
15	16	-1	1	0.0625
5	8	-3	9	1.125
20	16	4	16	1
20	20	0	0	0
10	10	0	0	0
20	20	0	0	0
25	24	1	1	0.0417
15	12	3	9	0.75
20	24	-4	16	0.67
				$\Sigma = 3.64$

$$\sum_{i=1}^{9} \frac{(O_i - E_i)^2}{E_i} = 3.64$$

Step 1: Null hypothesis

$$H_0 : O_i = E_i$$

Atributes are independent

Step 2: Alternate hypothesis

$$H_1 : O_i \neq E_i$$

Atributes are dependent.

Step 3: χ^2_α at $\nu = (r-1)(c-1) = (3-1)(3-1) = 4$

Scanned by TapScanner

$$\chi^2_\alpha = \chi^2_{0.02} = 11.668 \quad \nu = 4.$$

step 4:- χ^2-test:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.64$$

step 5:- Conclusion

$$\chi^2 = 3.64 < 11.668 < \chi^2_\alpha$$

Accept H_0

$\therefore$ Atributes are independent at 2% level.

2) From the following data find whether there is any significant liking in the habbit of taking soft drinks among the employes at 1% level.

Employes/ Softdrinks	Clerks	Teachers	officers	total
Pepsi	10	25	65	100
Thumsup	15	30	65	110
Fanta	50	60	30	140
Total	75	115	160	350

Contingency table for expected values

Employes/ Soft drinks	Clerks	Teachers	officers
Pepsi	$E(10) = \frac{100 \times 75}{350}$ $E(10) = 21.43$	$E(25) = \frac{100 \times 115}{350}$ $E(25) = 32.86$	$E(65) = \frac{100 \times 160}{350}$ $E(65) = 45.71$
Thumsup	23.57	36.14	50.29
Fanta	30	46	64.

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	21	-11	121	5.76
25	33	-8	64	1.94
65	46	19	361	7.85
15	24	-9	81	3.38
30	36	-6	36	1
65	50	15	225	4.5
50	30	20	400	13.33
60	46	14	196	4.26
30	64	-34	1156	18.06

$\Sigma = 60.08$

Step 1: Null hypothesis:

$$H_0 : O_i = E_i$$

There is no significant liking

Step 2: Alternate hypothesis

$$H_1 : O_i \neq E_i$$

There is no significant liking.

Step 3:-

$$\chi^2_\alpha \text{ at } \nu = (r-1)(c-1) = (3-1)(3-1) = 4.$$

$$\chi^2_\alpha = \chi^2_{0.01} = 13.277 \text{ at } \nu = 4.$$

Step 4:-

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 60.08$$

Step 5:- Conclusion:-

$$\chi^2 = 60.08 > 13.277 > \chi^2_\alpha$$

Accept H_1

$\therefore$ There is no significant liking.

Scanned by TapScanner