UNIT-IV

Null Hypothesis (H_0) **:** It is denoted by H_0, is a statement about the population parameter which is to be actually tested for acceptance or rejection.

Alternative Hypothesis (H_1) **:** It is denoted by H_1, is the opposite statement of null hypothesis.

Types of errors in test of hypothesis:

Type I error: The rejection of null hypothesis when it is true and should be accepted.

Type II error: The acceptance of null hypothesis when it is false and should be rejected.

	Accept H_0	Reject H_0
H_0 is true	Correct Decision	Type I error
H_0 is false	Type II error	Correct Decision

Level Of Significance (L.O.S.): It is denoted by α, is the probability of committing type I error. Thus L.O.S. measures the amount of risk or error associated in taking decisions. L.O.S. is expressed in percentage. Thus L.O.S. $\alpha = 5\%$ means that there are 5 chances in 100 that null hypothesis is rejected when it is true.

$$\alpha = probability\ of\ committing\ \ type\ I\ \ error = P(reject\ \ H_0 / H_0)$$
$$\beta = probability\ of\ committing\ \ type\ II\ \ error = P(reject\ \ H_0 / H_1)$$

Critical Region (C.R.): In any test of hypothesis, a test statistic S^*, calculated from the sample data is used to accept or reject the null hypothesis. Consider the area under the probability curve of the sampling distribution of the test statistic S^*. This area under the probability curve is divided into two regions, namely the region of rejection where N.H. is rejected and the region of acceptance where N.H. is accepted. Thus critical region is the region of rejection of N.H. The area of the critical region equals to the level of significance α. Note that C.R. always lies on the tail of the distribution.

One tailed test and two tailed test:

Right tailed test: When the alternative hypothesis H_1 is of the greater than type $i.e., H_1 : \mu > \mu_0\ or\ H_1 : \sigma_1^2 > \sigma_2^2\ etc.$ Then the entire critical region of area α lies on the right side of the curve as shown shaded in the fig. In such case the test of hypothesis is known as right tailed test.

Left tailed test: When the alternative hypothesis H_1 is of the less than type $i.e., H_1 : \mu < \mu_0\ or\ H_1 : \sigma_1^2 < \sigma_2^2.\ etc$ Then the entire critical region of area α lies on the left side of the curve as shown shaded in the fig. In such case the test of hypothesis is known as left tailed test.

Two tailed test: When the alternative hypothesis H_1 is of the Not equals type $i.e., H_1 : \mu \neq \mu_0\ or\ H_1 : \sigma_1^2 \neq \sigma_2^2\ etc.$ Then the entire critical region of area α lies on the both sides of the curve as shown shaded in the fig. In such case the test of hypothesis is known as two tailed test.

Critical Values of z

Level of Significance α	1%	5%	10%
Critical values for two-tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Critical values for right-tailes test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Critical values for left-tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

TEST OF SIGNIFICANCE FOR SMALL SAMPLES

(1) Test of significance for single mean (Students's t- test):

(i) Null Hypothesis (H_0) **:** $\bar{x} = \mu$ i.e., "there is no significance difference between the sample mean and population mean" or "the sample has been drawn from the population"

(ii) Alternative Hypothesis (H_1) **:** (i) $\bar{x} \neq \mu$ or (ii) $\bar{x} < \mu$ or (iii) $\bar{x} > \mu$

(iii) Level of Significance (α) **:** Set a level of significance

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$

(v) Conclusion: (i) If $|t| < t_\alpha$ we accept the Null Hypothesis H_0

(ii) If $|t| > t_\alpha$ we reject the Null Hypothesis H_0 i.e., we accept the Alternative Hypothesis H_1

(1) A random sample of 155 members has a mean 67 and S.D. 5.2. Is this sample has been taken from a large population of mean 70?

Solution:

Given $n =$, $\mu =$, $\bar{x} =$ *mean of the sample* $=$ *and* $s =$ *S.D. of the sample* $=$

(i) Null Hypothesis (H_0) *:*

(ii) Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$

(v) Conclusion: Degrees of freedom= $n-1=$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(2) A random sample of 10 boys had the following I.Q.'s: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100. Do these data support the assumption of a population I.Q. of 100 at a 5% L.O.S.? Also find the 95% confidence limits.

Solution:

$n =$, $\mu =$,

Given $\bar{x} =$ *mean of the sample* $= \dfrac{\sum x_i}{n} =$ *and*

$s =$ *S.D. of the sample* $= \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}} =$

(i) Null Hypothesis (H_0) *:*

(ii) Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

(v) Conclusion: Degrees of freedom= $n-1=$

Tabulated value of t_α =

Calculated value of $|t_\alpha|$ =

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(3) Producer of gutkhs, claims that the nicotine content in his gutkhs on the average is 1.83mg. Can this claim accepted if a random sample of 8 gutkha of this type have the nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.0, 1.6 mg? Use a 0.05 L.O.S.

Solution:

$n =\quad , \mu = \quad ,$

Given $\bar{x} = \text{mean of the sample} = \dfrac{\sum x_i}{n} =$ and

$s = \text{S.D. of the sample} = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}} =$

(i) Null Hypothesis (H_0) :

(ii) Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}} =$

(v) Conclusion: Degrees of freedom= $n-1=$

Tabulated value of t_α =

Calculated value of $|t_\alpha|$ =

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(4) The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data.

Item	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Life in 1000hrs	*1.2*	*4.6*	*3.9*	*4.1*	*5.2*	*3.8*	*3.9*	*4.3*	*4.4*	*5.6*

Can we accept the hypothesis that the average life time of bulbs is 4000hrs?

$n =\quad , \mu = \quad ,$

Solution:

Given $\bar{x} = \text{mean of the sample} = \dfrac{\sum x_i}{n} =$ and

$s = \text{S.D. of the sample} = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}} =$

(i) Null Hypothesis (H_0) :

(ii) Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

(v) Conclusion: Degrees of freedom= $n-1=$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(2) Test of significance for difference of means (Students's t- test):

***(i) Null Hypothesis*(H_0):** $\bar{x}=\mu$ i.e., "there is no significance difference between the sample mean and population mean" or "the sample has been drawn from the population"

***(ii)Alternative Hypothesis*(H_1):** (i) $\bar{x} \neq \mu$ or (ii) $\bar{x} < \mu$ or (iii) $\bar{x} > \mu$

***(iii) Level of Significance*(α):** Set a level of significance

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ *where* $s^2 = \dfrac{\sum(x_i-\bar{x})^2+(y_i-\bar{y})^2}{n_1+n_2-2}$

(v) Conclusion: (i) If $|t| < t_\alpha$ we accept the Null Hypothesis H_0

(ii) If $|t| > t_\alpha$ we reject the Null Hypothesis H_0 i.e., we accept the Alternative Hypothesis H_1

(1) The nicotine in milligrams of two samples of tobacco were found to be as follows. Test whether there is a significant difference between the two samples.

Sample A	*24*	*27*	*26*	*23*	*25*	*--*
Sample B	*29*	*30*	*30*	*31*	*24*	*36*

Solution: Given $n_1 =$ $n_2 =$

Calculations for means and Variances of samples					
Sample A			**Sample B**		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
24			29		
27			30		
26			30		
23			31		
25			24		
--			36		
$\bar{x} = \dfrac{\sum x}{n_1} =$		$\sum(x-\bar{x})^2 =$	$\bar{y} = \dfrac{\sum y}{n_2} =$		$\sum(y-\bar{y})^2 =$

$$s^2 = \frac{\sum(x_i-\bar{x})^2+(y_i-\bar{y})^2}{n_1+n_2-2} =$$

***(i) Null Hypothesis*(H_0):**

***(ii)Alternative Hypothesis*(H_1):**

***(iii) Level of Significance*(α):**

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} =$

(v) Conclusion: Degrees of freedom= $n_1 + n_2 - 2 =$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(2) 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.

Solution: Given $n_1 =$ $n_2 =$

Calculations for means and Variances of samples					
Sample A			Sample B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
28 30 32 33 33 29 34			29 30 30 24 27 29 --		
$\bar{x} = \frac{\sum x}{n_1} =$		$\sum(x-\bar{x})^2 =$	$\bar{y} = \frac{\sum y}{n_2} =$		$\sum(y-\bar{y})^2 =$

$$s^2 = \frac{\sum (x_i - \bar{x})^2 + (y_i - \bar{y})^2}{n_1 + n_2 - 2} =$$

(i) Null Hypothesis (H_0):

(ii) Alternative Hypothesis (H_1):

(iii) Level of Significance (α):

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} =$

(v) Conclusion: Degrees of freedom= $n_1 + n_2 - 2 =$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

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

Solution: Given $n_1 =$ $n_2 =$

Calculations for means and Variances of samples					
Sample A			Sample B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
117			106		
105			98		
97			87		
105			104		
123			116		
109			95		
86			90		
78			69		
103			108		
107			85		
$\bar{x} = \frac{\sum x}{n_1}$ =		$\sum(x-\bar{x})^2$ =	$\bar{y} = \frac{\sum y}{n_2}$ =		$\sum(y-\bar{y})^2$ =

$$s^2 = \frac{\sum (x_i-\bar{x})^2 + (y_i-\bar{y})^2}{n_1+n_2-2} =$$

(i) Null Hypothesis (H_0):

(ii) Alternative Hypothesis (H_1):

(iii) Level of Significance (α):

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} =$

(v) Conclusion: Degrees of freedom= $n_1+n_2-2=$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(4) Ten 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*
Scores after	*70*	*38*	*58*	*58*	*56*	*67*	*68*	*75*	*42*	*38*

Solution: Given $n_1 =$ $n_2 =$

Calculations for means and Variances of samples					
Sample A			Sample B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
67			70		
24			38		
57			58		
55			58		
63			56		

54 56 68 33 43			67 68 75 42 38		
$\bar{x} = \frac{\sum x}{n_1}$ =		$\sum(x-\bar{x})^2$ =	$\bar{y} = \frac{\sum y}{n_2}$ =		$\sum(y-\bar{y})^2$ =

$$s^2 = \frac{\sum (x_i-\bar{x})^2 + (y_i-\bar{y})^2}{n_1+n_2-2} =$$

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} =$

(v) Conclusion: Degrees of freedom= $n_1+n_2-2=$

Tabulated value of $t_\alpha =$

Calculated value of $|t_\alpha| =$

Calculated value of $|t_\alpha|$ Tabulated value of t_α

(3) <u>SNEDECOR'S F-TEST OF SIGNIFICANCE</u>

(i) Null Hypothesis $(H_0): \sigma_1^2 = \sigma_2^2$ *or* $s_1^2 = s_2^2$ i.e., the variances of the two populations are same.

(ii)Alternative Hypothesis $(H_1): \sigma_1^2 \neq \sigma_2^2$

(iii) Level of Significance (α) *:* set a lelvel of significance

(iv)Test Statistic: The test statistic

$$F = \frac{larger\ variance}{smaller\ variance}, \quad where\ s_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1},\ s_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$$

(v) Conclusion: Degrees of freedom = $(n, m) = (n_1-1, n_2-1)$

(i) If Calculated value of F < Tabulated value of F, we accept H_0

(ii) If Calculated value of F > Tabulated value of F, we reject H_0

(1) The time taken by the 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?

Solution: Given $n_1 =$ $n_2 =$

Calculations for means and Variances of samples					
Sample A			Sample B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
20 16 26 27 23 22 –			27 33 42 35 32 34 38		
$\bar{x}=\frac{\sum x}{n_1}$ =		$\sum(x-\bar{x})^2$ =	$\bar{y}=\frac{\sum y}{n_2}$ =		$\sum(y-\bar{y})^2$ =

$$s_1^2=\frac{\sum(x_i-\bar{x})^2}{n_1-1}=$$

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

(i) Null Hypothesis (H_0) :

(ii)Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $F=\frac{larger\ variance}{smaller\ variance}=$

(v) Conclusion: Degrees of freedom = $(n,m)=(n_1-1,n_2-1)$

Tabulated value of $F=$

Calculated value of $F=$

Calculated value of F Tabulated value of F

(2) The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal population a t10% significant level, test whether the two populations 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*

Solution: Given $n_1=$ $n_2=$

Calculations for means and Variances of samples					
Sample A			Sample B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
14.1 10.1 14.7 13.7			14.0 14.5 13.7 12.7		

14.0 $\bar{x}=\frac{\sum x}{n_1}$ =		$\sum(x-\bar{x})^2$ =	14.1 $\bar{y}=\frac{\sum y}{n_2}$ =		$\sum(y-\bar{y})^2$ =

$$s_1^2=\frac{\sum(x_i-\bar{x})^2}{n_1-1}= \qquad s_2^2=\frac{\sum(y_i-\bar{y})^2}{n_2-1}=$$

(i) Null Hypothesis (H_0) *:*

(ii) Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $F=\frac{larger\ variance}{smaller\ variance}=$

(v) Conclusion: Degrees of freedom = $(n,m)=(n_1-1,n_2-1)$

Tabulated value of F =

Calculated value of F =

Calculated value of F Tabulated value of F

(3) In two independent samples of sizes 8 and 10 the sum of squares of deviations of the sample values from the respective means were 84.4 and 102.6. Test whether the difference of variences of the population is significant or not. Use a 0.05 level of significance.

Solution:

Given $n_1=$ $\quad n_2=$ $\quad \sum(x-\bar{x})^2=$ $\quad \sum(y-\bar{y})^2=$

$$s_1^2=\frac{\sum(x_i-\bar{x})^2}{n_1-1}= \qquad s_2^2=\frac{\sum(y_i-\bar{y})^2}{n_2-1}=$$

(i) Null Hypothesis (H_0) *:*

(ii) Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $F=\frac{larger\ variance}{smaller\ variance}=$

(v) Conclusion: Degrees of freedom = $(n,m)=(n_1-1,n_2-1)$

Tabulated value of F =

Calculated value of F =

Calculated value of F Tabulated value of F

(4) <u>CHI- SQUARE TEST (χ^2) FOR GOODNESS OF FIT</u>

(i) Null Hypothesis (H_0) ***:*** There is no significant difference between expected frequency and observed frequency

(ii) Alternative Hypothesis (H_1) ***:*** There is a significant difference between expected frequency and observed frequency

(iii) Level of Significance (α) ***:*** set a lelvel of significance

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

(v) Conclusion: Degrees of freedom = $n-1$

(i) If Calculated value of $\chi^2 <$ Tabulated value of χ^2, we accept H_0

(ii) If Calculated value of $\chi^2 >$ Tabulated value of χ^2, we reject H_0

(1) A die is thrown 264 times with the following results. Show that the die is biased.

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

Solution: Given n=

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
40		
32		
28		
58		
54		
52		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n-1 =$

Calculated value of χ^2 =

Tabulated value of χ^2 =

Calculated value of χ^2 Tabulated value of χ^2

(2) The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory.

Solution: Given n=

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$

1026 1107 997 966 1075 933 1107 972 964 853		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n-1 =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(3) A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had scored a third class, 90 were placed in second class and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio 4:3:2:1 for the various categories respectively.

Solution: Given n=

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n-1 =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(4) A pair of dies are thrown 360 times and the frequency of each sum is indicated below:

$X = x_i$	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 fair on the basis of the chi-square test at 0.05 level of significance?

Solution: Given n=

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$p(x_i)$											
Expected Frequencies = $360\,p(x_i)$											

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
8 24 35 37 44 65 51 42 26 14 14		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n - 1 =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(5) 4 coins were tossed 160 times and the following results were obtained.

No. of Heads	0	1	2	3	4
Observed Frequency	17	52	54	31	6

Under the assumption that coins are balanced, find the expected frequencies of 0,1,2,3 or 4 heads, and test the goodness of fit at $\alpha = 0.05$

Solution: No. of coins =

Probability to get a head p = , q = 1 – p =

$X = x_i$	0	1	2	3	4

$p(x_i)$					
Expected Frequencies $= 160\,p(x_i)$					

Given n=

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
17 52 54 31 6		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n-1 =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(6) A survey of 240 families with 4 children each revealed the following distribution.

Male Births	***4***	***3***	***2***	***1***	***0***
Observed Frequencies	***10***	***55***	***105***	***58***	***12***

Can we accept that the male and female births are equally distributed?

Solution: No. of families = , No. of children =

Probability to have a male birth p = , q = 1 – p =

$X = x_i$	***4***	***3***	***2***	***1***	***0***
$p(x_i)$					
Expected Frequencies $= 240\,p(x_i)$					

Given n=

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
10 55 105 58 12		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= n-1 =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(5) CHI-SQUARE TEST FOR INDEPENDENT OF ATTRIBUTES

Let us consider two attributes A and B, and they are divided into two classes. The various frequencies can be expressed as follows:

A	a	b
B	c	d

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d=N$

The expected frequencies are given given by:

$E(a) = \frac{(a+c)(a+b)}{N}$	$E(b) = \frac{(b+d)(a+b)}{N}$	$a+b$
$E(a) = \frac{(a+c)(c+d)}{N}$	$E(b) = \frac{(b+d)(c+d)}{N}$	$c+d$
$a+c$	$b+d$	$a+b+c+d=N$

Degrees of freedom $= (n-1)(m-1)$

(1) The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	**Unstable**	**Totla**
Males	**40**	**20**	**60**
Females	**10**	**30**	**40**
Totla	**50**	**50**	**100**

Solution:

(i) Null Hypothesis (H_0) *:*

(ii)Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$E(a) = \frac{(a+c)(a+b)}{N} =$	$E(b) = \frac{(b+d)(a+b)}{N} =$
$E(a) = \frac{(a+c)(c+d)}{N} =$	$E(b) = \frac{(b+d)(c+d)}{N} =$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
40 20 10 30		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= (n-1)(m-1) =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(2) Given the following contingency table for hair colour and eye colour. Find the value of χ^2 ***Is there good association between the two?***

		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***

Solution:

(i) Null Hypothesis (H_0) :

(ii)Alternative Hypothesis (H_1) :

(iii) Level of Significance (α) :

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$E(a) = \frac{(a+d+g)(a+b+c)}{N} =$	$E(b) = \frac{(b+e+h)(a+b+c)}{N} =$	$E(b) = \frac{(c+f+i)(a+b+c)}{N} =$
$E(c) = \frac{(a+d+g)(d+e+f)}{N} =$	$E(b) = \frac{(b+e+h)(d+e+f)}{N} =$	$E(b) = \frac{(c+f+i)(d+e+f)}{N} =$
$E(c) = \frac{(a+d+g)(g+h+i)}{N} =$	$E(b) = \frac{(b+e+h)(g+h+i)}{N} =$	$E(b) = \frac{(c+f+i)(g+h+i)}{N} =$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
15 5 20 20 10 20 25 15 20		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= (n-1)(m-1) =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(3) From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

Employees / *Soft drinks*	*Clerks*	*Teachers*	*Officers*
Pepsi	*10*	*25*	*65*
Thums Up	*15*	*30*	*65*
Fanta	*50*	*60*	*30*

Solution:

(i) Null Hypothesis (H_0) *:*

(ii) Alternative Hypothesis (H_1) *:*

(iii) Level of Significance (α) *:*

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$E(a) = \frac{(a+d+g)(a+b+c)}{N} =$	$E(b) = \frac{(b+e+h)(a+b+c)}{N} =$	$E(b) = \frac{(c+f+i)(a+b+c)}{N} =$
$E(c) = \frac{(a+d+g)(d+e+f)}{N} =$	$E(b) = \frac{(b+e+h)(d+e+f)}{N} =$	$E(b) = \frac{(c+f+i)(d+e+f)}{N} =$
$E(c) = \frac{(a+d+g)(g+h+i)}{N} =$	$E(b) = \frac{(b+e+h)(g+h+i)}{N} =$	$E(b) = \frac{(c+f+i)(g+h+i)}{N} =$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
10 25 65 15 30 65 50 60 30		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) Conclusion: Degrees of freedom $= (n-1)(m-1) =$

Calculated value of $\chi^2 =$

Tabulated value of $\chi^2 =$

Calculated value of χ^2 Tabulated value of χ^2

(4) 1000 students at college level were graded according to their I.Q. and the economic conditions of their home. Use χ^2 test to find out whether there is any association between condition at home and I.Q. Use 0.05 L.O.S.

I.Q. / *Economic Condition*	*High*	*Low*	*Total*
Rich	***460***	***140***	***600***
Poor	***240***	***160***	***400***
Total	***700***	***300***	***1000***

Solution:

(i) Null Hypothesis(H_0):

(ii)Alternative Hypothesis(H_1):

(iii) Level of Significance(α):

(iv) Test Statistic: The test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$E(a) = \frac{(a+c)(a+b)}{N} =$	$E(b) = \frac{(b+d)(a+b)}{N} =$
$E(a) = \frac{(a+c)(c+d)}{N} =$	$E(b) = \frac{(b+d)(c+d)}{N} =$

Observed Frequency (O_i)	Expected Frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
460 140 240 160		
		$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} =$

(v) conclusion: Degrees of freedom $= (n-1)(m-1) =$

Calculated value of χ^2 =

Tabulated value of χ^2 =

Calculated value of χ^2 Tabulated value of χ^2

CHAPTER 7

TEST OF SIGNIFICANCE

7.1 INTRODUCTION

In the earlier chapter, we considered certian tests of significance based on the theory of the normal distribution. The assumptions made in deriving those tests will be valid only for large samples. When the sample is small ($n < 30$), we can use normal distribution to test for a specified population mean or difference of two population means as in large sample tests only when the sample is drawn from a normal population whose S.D., σ is known. If σ is not known, we cannot proceed as above. If a population is normally distributed, the sampling distribution of the sample mean for any sample size is also normally distributed whether σ is known or not.

7.2 DEGREE OF FREEDOM (d.f.)

The number of independent variates which make up the statistic is known as the degree of freedom (d.f.) and it is denoted by ν (the letter 'Nu' of the Greek alphabet). In other words, it is the number of values in a set of data which may be assigned arbitrarily or, it refers to the number of "independent constraints" in a set of data.

Definition : It is a number which indicates howmany of the values of a variable may be independently (or freely) chosen. For example, if we have to choose any four numbers freely, then we may choose 11,6,14,28 or any other set of four numbers. As in this case all the four numbers have freedom to vary, we say that the degrees of freedom is 4. If we impose a restriction on the numbers, say the sum is 50, then we can choose first three numbers freely and the fourth number is such that the sum is 50. Thus, the three variables are free and independent choices for finding the fourth. Hence these are the degrees of freedom. In this case, the degrees of freedom is 3, *i.e.*, $4-1$.

In general, the number of degrees of freedom is equal to the total number of observations less the number of independent constraints imposed on the observations. For example, in a set of data of *n* observations, if *k* is the number of independent constraints then $\nu = n-k$.

7.3 *t*- DISTRIBUTION (OR) STUDENT'S *t*- DISTRIBUTION

It is used for testing of hypothesis when the sample size is small and population S.D., σ is not known.

Definition : If $\{x_1, x_2, ..., x_n\}$ be any random sample of size n drawn from a normal (or approximately normal) population with mean μ and variance σ^2, then the test statistic 't' is defined by $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ where $\bar{x}$ = sample mean and

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of σ^2. The test statistic $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ is a random variable having the t-distribution with $\nu = n - 1$ degrees of freedom and with probabilty density function $f(t) = y_0\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ where $\nu = n-1$ and y_0 is a constant given by $\int_{-\infty}^{\infty} f(t)\,dt = 1$. This is known as "Student's t - distribution" or simply "t distribution".

The above result is more general than the central limit theorem, since it does not require the knowledge of σ, and since population is assumed to be normal it is less general than the central limit theorem.

Properties of t - Distribution :

1. The shape of t-distribution is **bell-shaped,** which is similar to that of a normal distribution and is symmetrical about the mean.

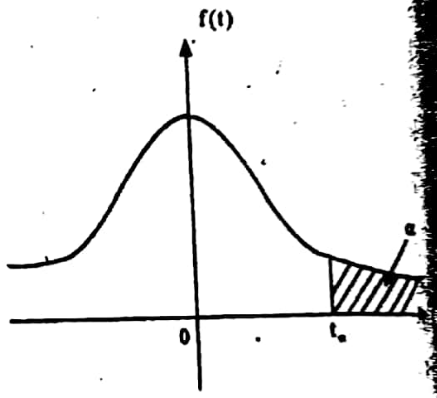

Fig. t-distribution

2. The t - distribution curve is also asymptotic to the t - axis, *i.e.*, the two tails of the curve on both sides of $t = 0$ extends to infinity.

3. It is symmetrical about the line $t = 0$.

4. The form of the probability curve varies with degrees of freedom i.e., with sample size.

5. It is unimodel with Mean = Median = Mode.

6. The *mean* of standard normal distribution and as well as t-distribution is zero but the variance of t-distribution depends upon the parameter ν which is called the degrees of freedom.

7. The **variance** of t-distribution exceeds 1, but approaches 1 as $n \to \infty$. Infact the t-distribution with ν-degrees of freedom approaches standard normal distribution as $\nu = (n-1) \to \infty$.

The selected values of t_α for various values of ν can be obtained from the tables of t-distribution, where t_α denotes the area under t-distribution to its right is equal to α. These values are tabulated in tables. In tables the left - hand column contains values of ν, the column heading are areas α in the right-hand tail of the t-distribution, and the entries are values of t_α. It is not necessary to tabulate values of t_α for α < 0.50, as it follows from the symmetry of the t-distribution that $t_{1-\alpha} = -t_\alpha$ *i.e.*, the t - value leaving an area of $1-\alpha$ to the right and therefore an area α to its left, is equal to the negative t - value which leads an area α in the right. The bottom row of the table of t-distribution correspond to the values of Z that cut off right hand tails of area α under the standard normal curve. We use the notation Z_α for such a value of Z, it can be seen for example, that $Z_{0.025} = 1.96 = t_{0.025}$ for $\nu = \infty$.

This result should really have been expected, since the t-distribution approaches the standard normal distribution as $\nu \to \infty$. In fact, if we observe, we find that the value of t_α for 29 or more degrees of freedom are close to the corresponding values of Z; we conclude that the standard normal distribution provides a good approximation to the t-distribution for samples of size $n \geq 30$.

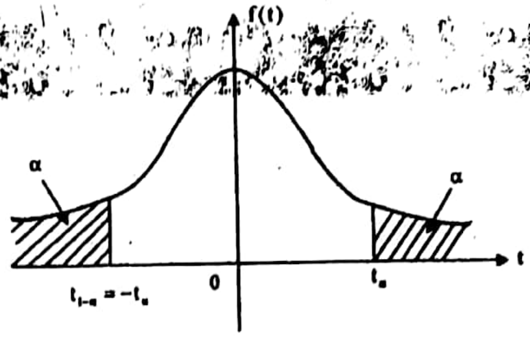

Fig. Symmetry of t-distribution

The t-distribution is extensively used in hypothesis about one mean, or about equality of two means when σ is unknown.

7.4 APPLICATIONS OF THE t - DISTRIBUTION

The t - distribution has a wide number of applications in Statistics, some of them are given below :

(1) 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 between two sample means or to compare two samples.

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

7.5 CHI-SQUARE (χ^2) DISTRIBUTION

Chi-squared distribution is a continuous probability distribution of a continuous random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} \cdot e^{-x/2}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

where ν is a positive integer is the only single parameter of the distribution, also known as "degrees of freedom (dof)".

χ^2 - distribution was extensively used as a measure of goodness of fit and to test the independence of attributes.

Properties of χ^2 - distribution :

1. χ^2 - distribution curve is not symmetrical, lies entirely in the first quadrant, and hence not a normal curve, since χ^2 varies from 0 to ∞.

2. It depends only on the degrees of freedom ν.

3. If χ_1^2 and χ_2^2 are two independent distributions with ν_1 and ν_2 degrees of freedom, the $\chi_1^2 + \chi_2^2$ will be chi-squared distribution with $(\nu_1 + \nu_2)$ degrees of freedom. That is, it is additive.

4. Here α denotes the area under the chi-square distribution to the right of χ_α^2.

So, χ_α^2 represents the χ^2- value such that the area under the chi-square curve to its (χ_α^2) right is equal to α. Chi-square distribution is very important in estimation and hypothesis testing. χ^2-distribution is used in sampling distribution, analysis of variance, mainly, it is used as a measure of goodness of fit and in analysis of $r \times c$ tables.

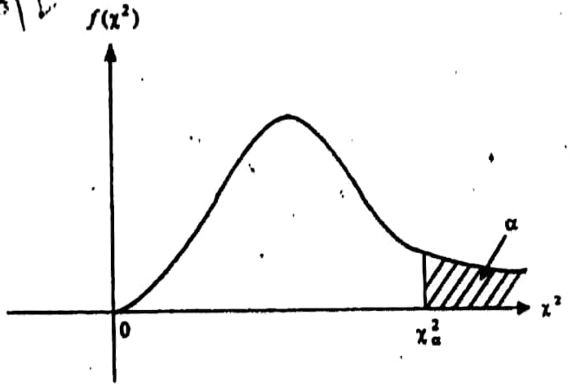

Fig. χ^2 - distribution

For various values of α and ν, the values of χ_α^2 are tabulated in tables.

In χ^2 table, the left hand column contains values of ν (degrees of freedom), the column headings are areas α in the right hand tail of χ^2 - distribution curve, the entries are χ^2 values. It is necessary to calculate values of χ_α^2 for $\alpha > 0.50$ since χ^2 curve or distribution is not symmetrical.

5. Mean $= \nu$ and variance $= 2\nu$

Sampling distribution of variance S^2 :

The theoretical sampling distribution of the sample variance for random samples from normal population is related to the chi-squared distribution as follows :

Let S^2 be the variance of a random sample of size n, taken from a normal population having the variance σ^2.

$$\text{Then } \chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma^2} \quad \left[\because S^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}\right]$$

is a value of a random variable having the χ^2-distribution with $\nu = n - 1$ degrees of freedom.

Exactly 95% of χ^2-distribution lies between $\chi^2_{0.975}$ and $\chi^2_{0.025}$ and when σ^2 is too small, χ^2 falls to the right of $\chi^2_{0.025}$ and when σ_1^2 is too large, χ^2 falls to the left of $\chi^2_{0.975}$. Thus when σ^2 is correct, χ^2-values are to the left of $\chi^2_{0.975}$ or to the right of $\chi^2_{0.025}$.

7.6 APPLICATIONS OF χ^2 DISTRIBUTION

(1) To test the goodness of fit.

(2) To test the indipendence of attributes.

(3) To test the homogeneity of independent estimation of the population variance.

(4) To test the homogeneity of independent estimation of the population correlation coefficient.

7.7 *F*-DISTRIBUTION (SAMPLING DISTRIBUTION OF THE RATIO OF TWO SAMPLE VARIANCES)

Another important continuous probability distribution which plays an important role in connection with sampling from normal populations is the F-distribution.

Let S_1^2 be the sample variance of an independent sample of size n_1 drawn from a normal population $N(\mu_1, \sigma_1^2)$. Similarly, let S_2^2 be the sample variance in an independent sample of size n_2 drawn from another normal population $N(\mu_2, \sigma_2^2)$.

Thus S_1^2 and S_2^2 are two variances of two random samples of sizes n_1 and n_2 respectively drawn from the normal population, with the variances σ_1^2 and σ_2^2. To determine whether the two samples come from two populations having equal variances.

Consider the sampling distribution of the ratio of the variances of the two independent random samples defined by

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

which follows F-distribution with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom.

F-distribution can be used to test the equality of several population means, comparing sample variances and analysis of variance.

Under the assumption (hypothesis) that two normal populations have the same variance *i.e.*, $\sigma_1^2 = \sigma_2^2$, we have $F = \frac{S_1^2}{S_2^2}$.

F determines whether the ratio of two sample variances S_1 and S_2 is too small or too large.

When *F* is close to 1, the two sample variances S_1 and S_2 are almost same. In practice, it is customary, to take the larger sample variance as the numerator.

F is always a positive number.

The sampling distribution of F is of the form $f(F) = K \dfrac{F^{(\nu_1-2)/2}}{(\nu_1 F + \nu_2)^{(\nu_1+\nu_2)/2}}$ where

ν_1 and ν_2 are two degrees of freedom and K is determined by $\int_0^{\infty} f(F)\, dF = 1$.

Properties of *F*-distribution :

(*i*) F - distribution is free from population parameters and depends upon degrees of freedom only.

(*ii*) *F*-distribution curve lies entirely in first-quadrant.

(*iii*) The *F*-curve depends not only on the two parameters ν_1 and ν_2 but also on the order in which they are stated.

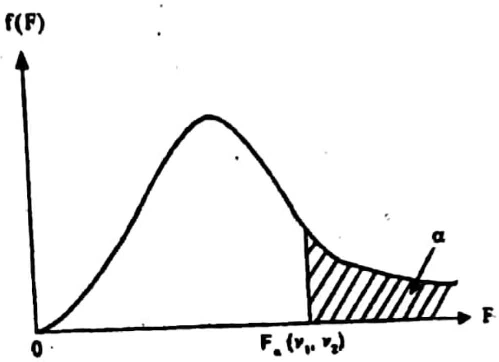

Fig. F - distribution

(*iv*) $F_{1-\alpha}(\nu_1, \nu_2) = \dfrac{1}{F_\alpha(\nu_2, \nu_1)}$

where $F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1 and ν_2 degrees of freedom such that the area under the *F*-distribution curve to the right of F_α is α.

(*v*) The mode of F - distribution is less than unity.

SOLVED EXAMPLES

Example 1 : Find (*a*) $t_{0.05}$ when $\nu = 16$

(*b*) $t_{-0.01}$ when $\nu = 10$

(*c*) $t_{0.995}$ when $\nu = 7$

Solution :

From tables

(*a*) When $\nu = 16$, $t_{0.05} = 1.746$

(*b*) when $\nu = 10$, $t_{-0.01} = -2.764$

(*c*) when $\nu = 7$, $t_{0.995} = t_{1-0.005} = -t_{0.005} = -3.499$

Example 2 : Find (a) $P(t < 2.365)$ when $\nu = 7$

(b) $P(t > 1.318)$ when $\nu = 24$

(c) $P(-1.356 < t < 2.179)$ when $\nu = 12$

(d) $P(t > -2.567)$ when $\nu = 17$

Solution :

(a) When $t < 2.365$, $P(t < 2.365)$ is given by the area to the left of $t = 2.365$

From table $t_\alpha = 2.365$ for $\nu = 7$ degrees of freedom then $\alpha = 0.025$

$\therefore$ $P(t < 2.365) = 1 - 0.025 = 0.975$

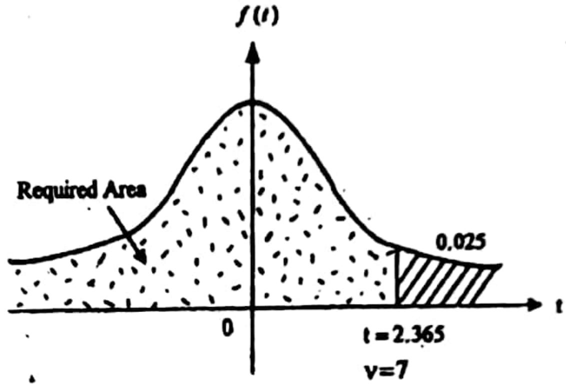

Fig.

(b) When $t > 1.318$

$P(t > 1.318)$ is given by the area to the right of $t = 1.318$

From table $t_\alpha = 1.318$ with $\nu = 24$ d.o.f. then $\alpha = 0.10$

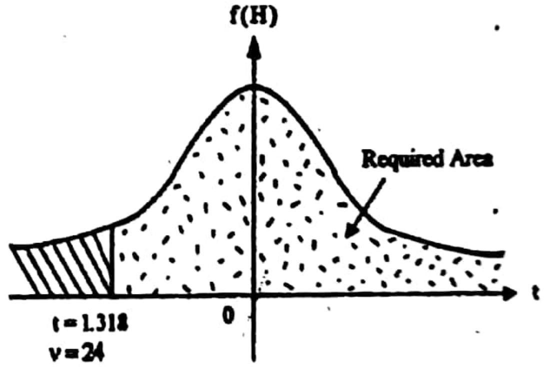

Fig.

(c) When $t < 2.179$ with $\nu = 12$ d.o.f.

Area to the right of 2.179 is 0.025

When $t > -1.356$ with $\nu = 12$, d.o.f.

Area to the left of 1.356 is 0.10

$\therefore$ When $-1.356 < t < 2.179$, the area is $1 - 0.10 - 0.025 = 0.875$

$\therefore$ $P(-1.356 < t < 2.179) = 0.875$

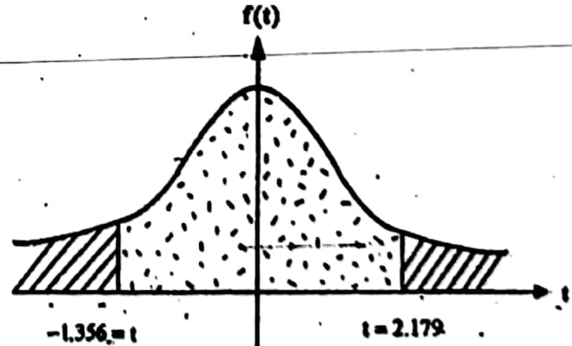

Fig.

(d) When $t > -2.567$

$P(t > -2.567)$ is given by the area to the right of $t = -2.567$

From table.

$t_{-\alpha} = -2.567$ with $\nu = 17$ d.o.f.

Then $\alpha = 1 - 0.01 = 0.99$

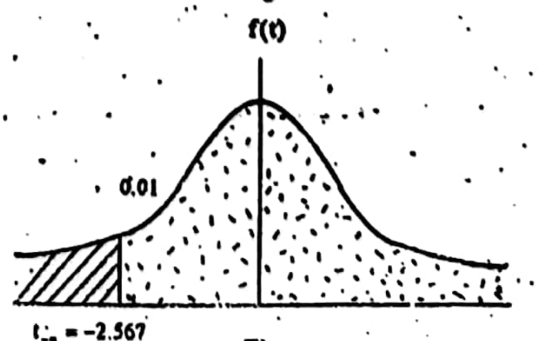

Fig.

Example 3 : A random sample of size 25 from a normal population has the mean $\bar{x} = 47.5$ and the standard deviation $S = 8.4$. Does this information tend to support or refute the claim that the mean of the population is $\mu = 42.5$? [JNTU (K), Nov. 2009 (Set No.3)]

Solution :

Given n = The size of the sample = 25

$\bar{x}$ = The mean of the sample = 47.5

μ = The population mean = 42.5

S = S.D of sample = 8.4

We have t-distribution,

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{47.5-42.5}{8.4/\sqrt{25}} = \frac{5\sqrt{25}}{8.4} = 2.98$$

This value of t has 24 degrees of freedom.

From the table of t-distribution for $\nu = 24$ with $\alpha = 0.005$ is 2.797. We conclude that the information given in the data of this example tend to refuse the claim that the mean of the population is $\mu = 42.5$ (*i.e.* μ cannot be 42.5).

Example 4 : A process for making certain ball bearings is under control if the diameters of the bearings have a mean of 0.5000 cm. If a random sample of 10 of these bearings has a mean diameter of 0.5060 cm and S.D of 0.0040 cm, is the process under control?

Solution :

n = The size of sample = 10

$\bar{x}$ = Sample mean = 0.5060

μ = Population mean = 0.5000

S = Sample S.D = 0.0040

∴ We have t-distribution

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{0.5060-0.5000}{(0.0040)/\sqrt{10}} = 4.7334$$

Here $\nu = n - 1 = 10 - 1 = 9$ degrees of freedom

Since $t = 4.7334 > 3.250 = t_\alpha$ with $\alpha = 0.005$ and $\nu = 9$

The process is not under control.

Example 5 : In 16 one hour test runs, the gasoline consumption of an engine averaged 16.4 gallons with a S.D of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour.

Solution :

Given Sample size, n = 16

Sample mean, $\bar{x}$ = 16.4

Population mean, μ = 12.0

and Sample S.D S = 2.1

From t-distribution, we have

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{16.4-12.0}{2.1/\sqrt{16}} = \frac{4.4\sqrt{16}}{2.1} = \frac{17.60}{2.1} = 8.38$$

The table of t-distribution shows that the probability of getting a value of t greater than 2.947 is 0.005 for 15 degrees of freedom, then the probability of getting a value greater than 8 must be negligible. Hence it would seem reasonable to conclude that the true average hourly gasoline consumption of the engine exceeds 12.0 gallons.

Example 6 : A manufacture claims that any of his list of items cannot have variance more than 1 cm². A sample of 25 items has a variance of 1.2 cm². Test whether the claim of the manufacturer is correct.

Solution :

Given sample size, $n = 25$

Sample variance, $S^2 = 1.2$, population variance, $\sigma^2 = 1$

Now $\chi^2 = \dfrac{(n-1)\,S^2}{\sigma^2} = \dfrac{24\,(1.2)}{1} = 28.8$

For $\alpha = 0.05$, $\chi^2_{0.05,\,24} = 36.41$, $\chi^2_{0.025,24} = 39.36$

Clearly $\chi^2 < \chi^2_{\alpha/2}$.

Hence the claim of the manufacturer is correct.

Example 7 : Suppose that the thickness of a part used in a semiconductor is its critical dimension and that the process of manufacturing these parts is considered to be under control if the true variation among the thickness of the parts is given by a S.D not greater than $s = 0.60$ thousandth of an inch. To keep a check on the process, random samples of size $n = 20$ are taken periodically, and it is regarded to be "out of control" if the probability that S^2 will take on a value greater than or equal to the observed sample value is 0.01 or less (even though $s = 0.60$). What can be concluded about the process if the S.D of such a periodic random sample is S = 0.84, thousandth of an inch ?

Solution :

Given sample size, $n = 20$

Sample S.D , $S = 0.84$

Population mean, $\sigma = 0.60$ and $\alpha = 0.01$.

The process will be declared out of control if the calculated value of χ^2 for 19 degrees of freedom exceeds the table value *i.e.*, $\chi^2_{0.001,19} = 36.191$.

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2} = \frac{19\,(0.84)^2}{(0.60)^2} = \frac{19\,(0.7056)}{(0.3600)} = \frac{13.4064}{0.36} = 37.24$$

Now the probability that χ^2 will exceed 36.19 is 0.01 for 19 degrees of freedom then the probability of getting a value greater than 37.24 is even less.

Hence the process is said to be out of control.

Example 8 : If two independent random samples of size $n_1 = 13$ $n_2 = 7$ are taken from a normal population. What is the probability that the variance of the first sample will be atleast four times as large as that of the second sample ?

Solution :

Given $n_1 = 13$ then $\nu_1 = n_1 - 1 = 13 - 1 = 12$

and $n_2 = 7$ then $\nu_2 = n_2 - 1 = 7 - 1 = 6$

and $S_1^2 = 4S_2^2$

i.e. the variance of the first sample will be at least four times as large as that of the second sample.

$$\text{Now, } F = \frac{S_1^2}{S_2^2} = \frac{4S_2^2}{S_2^2} = 4.00$$

$$\therefore \quad F = 4.00 \qquad (\because S_1^2 = 4S_2^2)$$

This value of F follows F-distribution with $\nu_1 = 12$ and $\nu_2 = 6$ degrees of freedom. Hence from the table we get $F_{0.05,\,12,\,6} = 4.00$.

Therefore, the required probability is 0.05.

Example 9 : For an F-distribution, find

(a) $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$

(b) $F_{0.01}$ with $\nu_1 = 24$ and $\nu_2 = 19$

(c) $F_{0.95}$ with $\nu_1 = 19$ and $\nu_2 = 24$

(d) $F_{0.99}$ with $\nu_1 = 28$ and $\nu_2 = 12$ **[JNTU (H) Nov. 2010 (Set No. 1)]**

Solution :

(a) From table, $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$ is 2.71

(b) $F_{0.01}$ with $\nu_1 = 24$, $\nu_2 = 19$ is 2.92

(c) $$F_{0.95}(19,24) = \frac{1}{F_{0.05}(24,19)} = \frac{1}{2.11} = 0.473933$$

(d) $$F_{0.99}(28,\ 12) = \frac{1}{F_{0.01}(12,28)} = \frac{1}{2.90} = 0.34482$$

EXERCISE 7(A)

1. Find the t-value with $\nu = 14$ d.o.f that leaves an area 0.025 to the left.

2. Determine (a) $t_{0.01}$ with $\nu = 18$

   (b) $t_{0.05}$ with $\nu = 12$

   (c) $t_{-0.10}$ with $\nu = 15$

3. Fuses produced by a company will blow in 12.40 minutes as the average when overloaded. Suppose the mean blow time of 20 fuses subjected to overload is 10.63 minutes and S.D 2.48 mts. Does this information tend to support or refuse the claim that the population mean blow time is 12.40 mts ?

4. A random sample of size 25 from a normal population has the mean = 47.5 and S.D = 8.4. Does this information tend to support or refute the claim that the mean of the population is 42.1.

5. The process of making certain bearings is under control if the diameter of the bearings have a mean of 0.5000 cm. What can we say about this process if a sample of 10 of these bearings has a mean diameter of 0.5060 cm and a S.D of 0.0040 cm .

6. Find the value of

   (a) $F_{0.05}$ for $\nu_1 = 15$ and $\nu_2 = 7$

   (b) $F_{0.95}$ for 12 and 15 d.o.f.

7. Find the value of

   (a) $F_{0.95}$ for $\nu_1 = 10$ and $\nu_2 = 20$

   (b) $F_{0.99}$ for $\nu_1 = 6$ and $\nu_2 = 20$

8. If independent random samples of size $n_1 = n_2 = 8$ come from normal populations having the same variance, what is the probability that either sample variance will be atleast seven times as large as the other.

9. Can we conclude that the two population variances are equal for the following data of post graduates passed out from a 'state' and 'private' university.

State :	8350	8260	8130	8340	8070	
Private :	7890	8140	7900	7950	7840	7920

10. The claim that the variance of a normal population is $\sigma^2 = 21.3$ is rejected if the variance of a random sample of size 15 exceeds 39.74.. What is the probability that the claim will be rejected even though $\sigma^2 = 21.3$?

11. The claim that the variance of a normal population is $\sigma^2 = 4$ is to be rejected if the variance of a random sample of size 9 exceeds 7.7535. What is the probability that this claim will be rejected even though $\sigma^2 = 4$?

ANSWERS

1. $t_{-0.025} = -2.145$
2. (a) 2.878 (b) 2.179 (c) 1.753
3. $t = -3.19$ at 19 d.o.f. refutes at a = 0.005
4. does not support the claim
5. $t = 4.74$, Process out of control
6. (a) 3.51 (b) $F_{0.95}(12,15) = \frac{1}{F_{0.05}(15,12)} = \frac{1}{2.62} = 0.38$
7. (a) $F_{0.95}(10,20) = 0.36$ (b) $F_{0.99}(6,20) = 0.135135$
8. $F_{0.01} = 6.99 \approx 7$ [for $\nu_1 = \nu_2 = 7$]
9. $S_1^2 = 15750$, $S_2^2 = 10920$, $F = \frac{S_1^2}{S_2^2} = 1.442$, not significantly different
10. $\chi^2 = 26.120$, $\chi^2_{0.025,\,14} = 26.119$; $\chi^2 > \chi^2_{\frac{0.05}{2},14}$ rejected.
11. $\alpha = 0.05$

7.8 TEST OF SIGNIFICANCE FOR SMALL SAMPLES

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 between the observed sample statistic and the hypothetical parameter value is significant.

(ii) The deviation between two sample statistics is significant.

The following are some important tests for small samples.

(i) Student's 't' test (ii) F-test (iii) χ^2-test

7.9 STUDENTS 't' TEST

Let $\bar{x}$ = mean of a sample

n = size of the sample

σ = Standard deviation of the sample

μ = Mean of the population supposed to be normal.

Then the student's t is defined by the statistic $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$.

If s^2 be the sample variance, $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$

If $S^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$, then $\frac{S^2}{s^2} = \frac{n}{n-1}$ or $S = \sqrt{\frac{n}{n-1}}\, .s$