

Module - 4

Statistical Inference - II

* t-test (Student)

* Chi-Square Distribution

* F-test

* Central limit theorem problems

→ Chi-Square Distribution:

Denoted by (χ^2) . Hence $\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$ when

i varies from 0 to n .

O_i = Observed frequency.

E_i = Estimated frequency.

→ Degree of freedom (df) $\rightarrow n-1$

Problems:

1. 4 coins are tossed hundred times and the following results were obtained, fit a binomial distribution for the data and test the goodness of fit. ($\chi^2 = 9.49$ for 4 df, 0.05).

⇒

No of heads	0	1	2	3	4
frequency (f)	5	29	36	25	5

Given frequency is 5, 29, 36, 25, 5 called observed frequency. Then we need to calculate estimated frequency.

given,

$n=4$, WKT formula for mean $\Rightarrow \text{mean} = \frac{\sum fx}{\sum f}$

$$\text{mean} = \frac{0 \times 5 + 1 \times 29 + 2 \times 36 + 3 \times 25 + 4 \times 5}{5 + 29 + 36 + 25 + 5}$$

$$\boxed{\text{mean} = 1.96}$$

In binomial distribution, $\text{mean} = np$
 $1.96 = 4(p)$

$$\boxed{p = 0.49}$$

$$p + q = 1$$

$$q = 1 - p$$

$$\boxed{q = 0.51}$$

let $f(x) = 100 p(x)$

$$f(x) = 100 [{}^n C_x p^x q^{n-x}]$$

$$f(x) = 100 [{}^4 C_x (0.49)^x (0.51)^{4-x}]$$

when $x=0$

$$f(0) = 100 [{}^4 C_0 (0.49)^0 (0.51)^{4-0}]$$

$$f(0) = 100 \cdot 6.765 \approx 7$$

when $x=1$

$$f(1) = 100 [{}^4 C_1 (0.49)^1 (0.51)^{4-1}]$$

$$f(1) = 25.99 \approx 26$$

when $x=2$

$$f(2) = 100 [{}^4 C_2 (0.49)^2 (0.51)^{4-2}]$$

$$f(2) = 37.47 \approx 38$$

when $x=3$

$$f(3) = 24.00 \approx 24$$

when $x=4$

$$f(4) = 5.76 \approx 6$$

	0	1	2	3	4
O_i	5	29	36	25	5
E_i	7	26	38	24	6

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

E_i

$$\chi^2 = \frac{(5-7)^2}{7} + \frac{(29-26)^2}{26} + \frac{(36-38)^2}{38} + \frac{(25-24)^2}{24} + \frac{(5-6)^2}{6}$$

$$\chi^2 = 1.23$$

$H_0 \rightarrow$ test the goodness of fit

$$\chi^2_{0.05} = 9.49 \text{ for 4 df}$$

$1.23 < 9.49$ H_0 is accepted

2. 5 dice were thrown 96 times and the no of heads odd faces 1, 3, 5 appearing on the face follow the frequency distribution as follows.

no of faces (1, 3, 5)	5	4	3	2	1	0
frequency	7	19	35	24	8	3

test the goodness of fit by using the binomial distribution
($\chi^2_{0.05} = 11.07$ of 5 df)

$$\text{Mean} = n = 5 \quad p = 1/2 \quad q = 1/2$$

mean = np In binomial distribution

$$= 5 \left(\frac{1}{2} \right)$$

$$\text{mean} = 2.5$$

$$\text{let } f(x) = 100 p(x)$$

$$f(x) = 100 {}^n C_x p^x q^{n-x}$$

when $x=0$

$$f(0) = 100 {}^5 C_0 \left(\frac{1}{2} \right)^0 \left(\frac{1}{2} \right)^{5-0}$$

$$f(0) = 3.125 \times 100 = 312.5$$

when $x=1$

$$f(1) = 15.25 \approx$$

when $x=5$

$$f(5) = 3$$

when $x=2$

$$f(2) = 30.25 \approx 30$$

when $x=3$

$$f(3) = 27.25 \approx 30$$

when $x=4$

$$f(4) = 15.25 \approx 15$$

	5	4	3	2	1	0
O_i	7	19	35	24	8	3
E_i	3	15	30	30	15	3

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

$$\chi^2 = \frac{(7-3)^2}{3} + \frac{(19-15)^2}{15} + \frac{(35-30)^2}{30} + \frac{(24-30)^2}{30} + \frac{(8-15)^2}{15} + \frac{(3-3)^2}{3}$$

$$\chi^2 = 11.7$$

H₀ → test for goodness of fit

$$\chi^2_{0.05} = 11.07 \text{ for } 5df$$

As $11.7 > 11.07$ H₀ rejected.

3. 4 coins were tossed 160 times and the following results were obtained.

no of heads	0	1	2	3	4
frequency	17	52	54	31	6

Test the goodness of fit using Binomial Distribution
($\chi^2_{0.05} = 9.49$ for 4df)

$$\Rightarrow n=4$$

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{0 \times 17 + 1 \times 52 + 2 \times 54 + 3 \times 31 + 4 \times 6}{160}$$

$$\text{mean} = 1.75$$

$$\text{let } f(x) = 160 p(x) \\ = 160 {}^n C_x p^x q^{n-x}$$

mean = np In binomial distribution

$$1.73 = 4(p)$$

$$\frac{1.73}{4} = p$$

$$\boxed{p = 0.43}$$

$$p + q = 1$$

$$q = 1 - p$$

$$\boxed{q = 0.56}$$

$$f(x) = 160 {}^n C_x p^x q^{n-x}$$

$$f(x) = 160 {}^4 C_x (0.43)^x (0.56)^{4-x}$$

when $x = 0$

$$f(0) = 160 {}^4 C_0 (0.43)^0 (0.56)^{4-0}$$

$$f(0) = 15.73 \approx 16$$

when $x = 1$

$$f(1) = 48.32 \approx 48$$

when $x = 2$

$$f(2) = 55.66 \approx 56$$

when $x = 3$

$$f(3) = 28.49 \approx 29$$

when $x = 4$

$$f(4) = 5.47 \approx 6$$

	0	1	2	3	4
O_i	17	52	54	31	6
E_i	16	48	56	29	6

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

$$\chi^2 = \frac{(17-16)^2}{16} + \frac{(52-48)^2}{48} + \frac{(54-56)^2}{56} + \frac{(31-29)^2}{29} + \frac{(6-6)^2}{6}$$

$$\chi^2 = 0.6051$$

4. 5 dice were thrown fit a poisson distribution for the following data and test the goodness of fit. Given that $\chi^2_{0.05} = 7.815$ for 3df.

x	0	1	2	3	4
f	122	60	15	2	1

$$n = 5 \quad \text{mean } \mu = \frac{\sum fx}{\sum f} = \frac{0 \times 122 + 1 \times 60 + 2 \times 15 + 3 \times 2 + 4 \times 1}{122 + 60 + 15 + 2 + 1}$$

$$\mu = \text{mean}(\mu) = 0.5$$

$$\text{mean} = np$$

$$0.5 = 5(p)$$

$$p = 0.1$$

$$p + q = 1$$

$$0.1 + q = 1$$

$$q = 0.9$$

$$\text{let } f(x) = 200 p(x)$$

$$f(x) = \frac{200 (0.5)^x e^{-0.5}}{x!}$$

$$\frac{200 m^x e^{-m}}{x!} = p(x)$$

$$\text{when } x = 0$$

$$f(0) = 121.30 \approx 121$$

$$f(1) = 60.65 \approx 61$$

$$f(2) = 15.16 \approx 15$$

$$f(3) = 2.52 \approx 3$$

$$f(4) = 0.31 \approx 0$$

estimated frequency.

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

$$\chi^2 = \frac{(122 - 121)^2}{121} + \frac{(60 - 61)^2}{61} + \frac{(15 - 15)^2}{15} + \frac{(2 - 3)^2}{3} + \frac{(1 - 0)^2}{0}$$

O_i	122	60	15	2	1
E_i	121	61	15	3	0

$$\chi^2 = 0.85$$

$$\chi^2 = 0.024$$

$$\chi^2 = 0.025$$

H_0 is accepted.

5. Fit a poisson distribution for the following and test the goodness of fit. $\chi^2_{0.05} = 9.49$ for 4df

x	0	1	2	3	4
f	419	352	154	56	19

$$n = 5 \quad \text{mean} = \frac{\sum fx}{\sum f} = \frac{0 \times 419 + 1 \times 352 + 2 \times 154 + 3 \times 56 + 4 \times 19}{419 + 352 + 154 + 56 + 19}$$

$$\text{mean} = 0.904$$

$$\text{let } f(x) = 1000 p(x) \quad p(x) = \frac{m^x e^{-m}}{x!}$$

$$1000 \times \frac{(0.904)^x e^{-0.904}}{x!}$$

when

$$x=0 \quad f(0) = 404.94 = 405$$

$$x=1 \quad f(1) = 366.07 = 366$$

$$x=2 \quad f(2) = 165.46 = 165$$

$$x=3 \quad f(3) = 49.85 = 50$$

$$x=4 \quad f(4) = 11.86 = 11$$

K_i

O_i	419	352	154	56	19
E_i	405	366	165	50	11

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

$$\chi^2 = 8.3$$

$$\chi^2_{0.05} = 9.49$$

Hence H_0 accepted

6. The no of accidents per day (x) as recorded in a textile industry over a period of 400 days is given below. Test the goodness of fit in respect of poisson distribution. The given $\chi^2_{0.05} = 9.49$ for 4 df

x	0	1	2	3	4	5
f	173	168	37	18	3	1

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{0 \times 173 + 1 \times 168 + 2 \times 37 + 3 \times 18 + 4 \times 3 + 5 \times 1}{173 + 168 + 37 + 18 + 3 + 1}$$

$$\text{mean} = 0.7825$$

$$\text{let } f(x) = 400 p(x)$$

$$p(x) = \frac{m^x e^{-m}}{x!}$$

$$400 \times \frac{(0.7825)^x e^{-0.7825}}{x!}$$

when

$$x=0 \quad f(0) = 182.90 \approx 183$$

$$x=1 \quad f(1) = 143.12 \approx 143$$

$$x=2 \quad f(2) = 55.99 \approx 56$$

$$x=3 \quad f(3) = 14.60 \approx 15$$

$$x=4 \quad f(4) = 2.85 \approx 3$$

$$x=5 \quad f(5) = 0.44 \approx 0$$

O_i	173	168	37	18	3 + 1 = 4
E_i	183	143	56	15	3 + 0 = 3

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

$$= \frac{(173-183)^2}{183} + \frac{(168-143)^2}{143} + \frac{(37-56)^2}{56} + \frac{(18-15)^2}{15} + \frac{(4-3)^2}{3}$$

notesforvtu.in $\chi^2 = 12.89$ is rejected.

7. A die is thrown 264 times and the no number appearing on the face follows the following frequency distribution. Calculate the value of χ^2 .

x	1	2	3	4	5	6
f	40	32	28	58	54	60

⇒ The given frequency is observed frequency.
 We need to calculate Estimated Frequency for 1, 2, 3, 4, 5, 6, appearing on the face. $= \frac{264}{6} = 44$ for each face

Hence, O_i - Observed frequency E_i → Estimated frequency

x	1	2	3	4	5	6
O_i	40	32	28	58	54	60
E_i	44	44	44	44	44	44

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

$$\chi^2 = \frac{(44-40)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(54-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$\chi^2 = 22$$

8. The following table gives the no of aircraft accidents that occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Days	sun	mon	tue	wed	thu	fri	sat	total
no of accidents	14	16	8	12	11	9	14	84

given $\chi^2_{0.05}(6) = 12.59$

The given frequency is observed frequency. Now, we have to calculate estimated frequency.

i.e. total no of accidents = 84
no of days = 7
$$= \frac{84}{7} = 12$$

Days	sun	mon	tue	wed	thu	fri	sat
O_i	14	16	8	12	11	9	14
E_i	12	12	12	12	12	12	12

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

$$\chi^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$\frac{(14-12)^2}{12}$$

$$\chi^2 = 4.16$$

H_0 is accepted

9. In order to test whether a die is biased. It is thrown 72 times and the results are tabulated as follows.

Result of throw	1	2	3	4	5	6	Total
no of throws	8	14	15	9	13	13	72

What is your conclusion?

Given: O_i , we need find E_i

$$E_i = \frac{72}{6} = 12$$

Result of throw	1	2	3	4	5	6	Total
O_i	8	14	15	9	13	13	72
E_i	12	12	12	12	12	12	

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

$$\chi^2 = 3.33$$

10. A sample of examinations of 500 students was made. It was found that 820 students had failed, 170 had secured 3rd class, 90 had secured 2nd class and 20 had secured 1st class. Do these figures support the general examination result which is in the ratio 4:3:2:1 for the respective categories ($\chi^2_{0.05} = 7.81$ for 3df)

O_i	820	170	90	20
E_i	800	150	100	50

$$E_i \frac{4}{10} \times 500 = 200$$

$$\frac{3}{10} \times 500 = 150$$

$$\frac{2}{10} \times 500 = 100$$

$$\frac{1}{10} \times 500 = 50$$

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

$$\chi^2 = \frac{(820 - 800)^2}{800} + \frac{(170 - 150)^2}{150} + \frac{(90 - 100)^2}{100} + \frac{(20 - 50)^2}{50}$$

11. In experiments on pea breeding, the following frequency of seeds were obtained,

Round & yellow 315	wrinkled & yellow 101	Round & green 108
wrinkled & green 32	Total 556	

Theory predicts that the frequencies should be in proportions 9:3:3:1

Examine the correspondence with theory & experiments. $\chi^2_{(0.05)} = 7.815$

$$\frac{9}{16} \times 556 = 312.75 \approx 313$$

$$\frac{3}{16} \times 556 = 104.25 \approx 104$$

$$\frac{1}{16} \times 556 = 34.75 \approx 35$$

O_i	315	101	108	32	Total
E_i	313	104	104	35	

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

$$\chi^2 = \frac{(315-313)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35}$$

$$\chi^2 = 0.51 > 7.815 \quad H_0 \text{ accepted.}$$

→ Student t-distribution (or) t-distribution

For single sample only one sample $t = \frac{\bar{x} - \mu}{s} \sqrt{n}$

Here → μ = mean

s = standard deviation

n = sample size

$\bar{x}$ = sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

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

(or)

$$s^2 = \frac{1}{(n-1)} \left[\sum x^2 - \frac{1}{n} (\sum x)^2 \right]$$

→ Degrees of freedom (df): $df = n-1$

1. A machine is expected to produce nails of length 3 inches. A random sample of 25 nails gave an average length of 3.1 inch with standard deviation 0.3. Said that the machine is producing nails as per specification? (to 0.05 for 24 df = 2.064)

$$\Rightarrow n = 25 \quad v = n-1 = 25-1 = 24$$

H_0 = machine producing the nails.

By the data, for single sample, degree of $v = n-1$

$$\therefore 25-1 = 24 \quad \text{Hence } n = 25$$

$$SD = 0.3 \quad \mu = 3 \text{ inch} \quad \bar{x} = 3.1$$

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

$$= \frac{3.1 - 3}{0.3} (\sqrt{25})$$

$$= \frac{0.1 (5)}{0.3}$$

Thus,

$$t_{0.05} = 1.07 < 2.064 \quad H_0 \text{ accepted.}$$

2. Find the standard deviation for all the following variable values in a sample of 8: -4, -2, -2, 0, 2, 2, 3, 3 taking the mean of the universe 0

$$\sum x = 2 \quad n = 8; \mu = 0$$

$$-4, -2, -2, 0, 2, 2, 3, 3$$

$$\bar{x} = \frac{1}{n} \sum x$$

$$= \frac{1}{n} [x_1 + x_2 + x_3 + \dots + x_n]$$

$$= \frac{1}{8} [-4 - 2 - 2 + 0 + 2 + 2 + 3 + 3]$$

$$\bar{x} = \frac{1}{8} [2]$$

$$\boxed{\bar{x} = 0.25}$$

$$s^2 = \frac{1}{n-1} \left[\sum x^2 - \frac{1}{n} (\sum x)^2 \right]$$

$$\sum x = 2$$

$$\sum x^2 = (-4)^2 + (-2)^2 + (-2)^2 + (0)^2 + (2)^2 + (2)^2 + (3)^2 + (3)^2$$

$$\sum x^2 = 50$$

$$s^2 = \frac{1}{8-1} \left[50 - \frac{1}{8} [2]^2 \right]$$

$$s^2 = 7.07$$

$$s = \sqrt{7.07}$$

$$s = 2.6589$$

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

$$t = \frac{0.25 - 0}{2.6589} \sqrt{8}$$

$$t = 0.266$$

$$0.266 < t_{0.05} (7df) 2.365$$

$$< t_{0.01} (7df) 5.409$$

Ho accepted.

3. 10 individuals are chosen from at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis that the mean height of the universe is 66 inches.
($t_{0.05} = 2.862$ for 9df)

$$n = 10, \mu = 66$$

$$\bar{x} = \frac{1}{n} \sum x$$

$$\bar{x} = \frac{678}{10}$$

$$= \frac{1}{10} [x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \dots + x_{10}]$$

$$\bar{x} = 67.8$$

$$= \frac{1}{10} [63 + 63 + 66 + 67 + 68 + 69 + 70 + 70 + 71 + 71] \quad \boxed{\sum x = 678}$$

$$\boxed{\bar{x} = 67.8}$$

$$S^2 = \frac{1}{n-1} \left[\sum x^2 - \frac{1}{n} (\sum x)^2 \right]$$

$$= \frac{1}{9} \left[(63^2 + 63^2 + 66^2 + 67^2 + 68^2 + 69^2 + 70^2 + 70^2 + 71^2 + 71^2) \right] - \frac{1}{10} [678^2]$$

$$S^2 = 9.0667$$

$$\boxed{S = 3.011}$$

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

$$t = \frac{67.8 - 66}{3.011} \sqrt{10}$$

$$t = 1.89 < 2.862 \quad H_0 \text{ accepted}$$

4. A certain stimulus administered to each of the 12 patients results in the following change in B.P. 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase the BP? ($t_{0.05} = 2.201$ for 11 df).

$\Rightarrow$ In the given data, mean is not given. Hence we conclude that $\mu = 0$. Here $n = 12$ for single year sample degrees of freedom

$$v = n - 1 = 12 - 1 = 11$$

$$\text{mean} = 0$$

$$\bar{x} = \frac{1}{n} \sum x$$

$$= \frac{1}{12} [5 + 2 + 8 + (-1) + 3 + 0 + 6 + (-2) + 1 + 5 + 0 + 4]$$

$$\boxed{\bar{x} = 2.33} \quad \bar{x} = \frac{31}{12} = 2.5833$$

$$\boxed{\sum x = 31}$$

$$s^2 = \frac{1}{n-1} \left[\sum x^2 - \frac{1}{n} (\sum x)^2 \right]$$

$$\sum x^2 = 5^2 + 2^2 + 8^2 + (-1)^2 + 3^2 + 0^2 + 6^2 + (-2)^2 + 1^2 + 5^2 + 0^2 + 4^2$$

$$\sum x^2 = 185$$

$$s^2 = \frac{1}{n-1} \left[\sum x^2 - \frac{1}{n} (\sum x)^2 \right]$$

$$= \frac{1}{11} \left[185 - \frac{1}{12} (31)^2 \right]$$

$$s^2 = 9.538$$

$$\boxed{s = 3.088}$$

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

$$t = \frac{2.5833 - 0}{3.088} \sqrt{12}$$

notesforvtu.in $t = 2.89 > 2.201$ H_0 rejected.

⇒ For two samples (n_1, n_2)

The required t test is given by $t = \frac{\bar{x} - \bar{y}}{S \sqrt{1/n_1 + 1/n_2}}$

$\bar{x}$ = mean of x : $\bar{y}$ = mean of y

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \quad ; \quad \bar{y} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j$$

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{j=1}^{n_2} (y_j - \bar{y})^2 \right]$$

(or)

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\left(\sum x^2 - \frac{1}{n_1} (\sum x)^2 \right) + \left(\sum y^2 - \frac{1}{n_2} (\sum y)^2 \right) \right]$$

(or)

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

→ df $V = n_1 + n_2 - 2$

1. Two types of batteries are tested for their length of life and the following results were obtained.

Battery A: $n_1 = 10$, $\bar{x}_1 = 460$ hrs ; $\sigma_1^2 = 100$

Battery B: $n_2 = 10$, $\bar{x}_2 = 500$ hrs ; $\sigma_2^2 = 125$

Compute student t test and test whether there is a significant difference in two means.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{1/n_1 + 1/n_2}}$$

$$S^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2 - 2}$$

$$t = \frac{460 - 500}{S/10}$$

$$S/10$$

$$S^2 = 122.75$$

$$S = \sqrt{122.75}$$

$$S = 11.08$$

$$t = -8.072$$

$$|t| = 8.072$$

$$df = n_1 + n_2 - 2$$

$$= 10 + 10 - 2$$

$$df = 20$$

$$t_{0.05} = 2.101$$

2. A group of boys and girls were given an intelligence test. The mean score, standard deviation score, no of in each group are as follows.

	Boys	Girls
mean	74	70
SD	8	10
n	12	10

Is the difference b/w the mean of the two groups significant at 5% level of significance ($t_{0.05} = 2.086$ for 20df)

$$\bar{x} = 74 \quad \bar{y} = 70 \quad \sigma_1^2 = 8 \quad s_1 = 8$$

$$n_1 = 12 \quad n_2 = 10 \quad s_2 = 10$$

$$d = n_1 + n_2 - 2$$

$$= 12 + 10 - 2$$

$$d = 20 \text{ df}$$

$$s^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2 - 2} = \frac{12(8^2) + 10(10^2)}{12 + 10 - 2}$$

$$s^2 = 88.4$$

$$s = 9.4021$$

$$t = \frac{74 - 70}{9.4021 \sqrt{1/12 + 1/10}}$$

$$t = 0.994$$

$t = 0.994 < 2.086$ for H_0 rejected

3. A group of 10 boys fed on diet A and another group of 8 boys fed on different diet B for a period of six months recorded the following increasing weights

Diet A: 5 6 8 1 12 4 3 9 6 10

Diet B: 2 3 6 8 10 1 2 8

test whether diet A and B differ significantly regarding their increase in weight

$$n_1 = 10 \quad n_2 = 8$$

$$\bar{x} = \frac{1}{n_1} (\sum x_i)$$

$$= \frac{1}{10} (5 + 6 + 8 + 1 + 12 + 4 + 3 + 9 + 6 + 10)$$

$$\boxed{\bar{x} = 6.4}$$

$$\bar{y} = \frac{1}{n_2} (\sum y_i)$$

$$= \frac{1}{8} (2 + 3 + 6 + 8 + 10 + 1 + 2 + 8)$$

$$\boxed{\bar{y} = 5}$$

$$\Rightarrow \frac{1}{10+8-2} \left[512 - \frac{1}{10} (64) + 282 + \frac{1}{8} (40)^2 \right]$$

$$S^2 = 11.525$$

$$S = \sqrt{11.525}$$

$$S = 3.394$$

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

$$t = 0.8696$$

$$t_{0.05} = 0.8696 < 2.120$$

$$t_{0.01} = 0.8696 < 2.921$$

H_0 accepted

4. Two horses were tested according to the time (in seconds) to run a particular race with the following results

Horse A : 28 30 32 33 33 29 34

Horse B : 29 30 30 24 27 29

Test whether you can discriminate b/w the two horses.

$$n_1 = 7 \quad n_2 = 6$$

$$\bar{x} = \frac{1}{n_1} (\sum x_i)$$

$$= \frac{1}{7} (28+30+32+33+33+29+34)$$

$$\bar{x} = 31.28$$

$$\bar{y} = \frac{1}{n_2} (\sum y_i)$$

$$= \frac{1}{6} (29 + 30 + 30 + 24 + 27 + 29)$$

$$\boxed{\bar{y} = 28.16}$$

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\left(\sum x^2 - \frac{1}{n_1} (\sum x)^2 \right) + \left(\sum y^2 - \frac{1}{n_2} (\sum y)^2 \right) \right]$$

$$S^2 = \frac{1}{7 + 6 - 2} \left[(28^2 + 30^2 + 32^2 + 33^2 + 33^2 + 29^2 + 34^2 - \cancel{2(28.28)^2}) - \frac{1}{7} (219)^2 \right] + \left(29^2 + 30^2 + 30^2 + 24^2 + 27^2 + 29^2 - \frac{1}{6} (169)^2 \right)$$

$$S^2 = \frac{1}{11} \left[(6883 - 6851.5) + (4787 - 4760.16) \right]$$

$$S^2 = 5.2966$$

$$S = 2.3014$$

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

$$t = 2.4365 \begin{cases} t_{0.05}(2.001) \approx \text{Ho Rejected} \\ t_{0.01}(3.106) \text{ Ho accepted} \end{cases}$$

5. $\Rightarrow$ Confidence limit of the population of mean (μ)

95% confidence limit for μ is given by

$$95\% = \bar{x} \pm \frac{S}{\sqrt{n}} \quad t_{0.05}$$

$$99\% = \bar{x} \pm \frac{S}{\sqrt{n}} \quad t_{0.01}$$

Confidence limits can be also be written as Fiducial limit

$\rightarrow$ Problems.

1. A sample of 10 measurements of the diameter of the sphere gave a mean of 12 cm and standard deviation of 0.15 cm. Find the 95% and 99% confidence limits for the actual diameter.

$$n = 10$$

$$\bar{x} = 12 \quad \bar{x} = 12$$

$$S = 0.15$$

$$95\% = \bar{x} \pm \frac{S}{\sqrt{n}} \quad t_{0.05}$$

$$= 12 \pm \frac{0.15}{\sqrt{10}} (2.862)$$

$$= 12.10$$

Here single sample is given

$$df = n - 1 = 10 - 1 = 9$$

$$12 \pm \frac{0.15}{\sqrt{10}} (2.262)$$

$$= 11.89$$

$$99\% = \bar{x} \pm \frac{S}{\sqrt{n}} \quad t_{0.01}$$

$$= 12 \pm \frac{0.15}{\sqrt{10}} (3.250)$$

$$= 12.45$$

$$99\% = 12 + \frac{0.15}{\sqrt{10}} (3.250)$$

$$99\% = 12.45$$

2. A sample of 11 rats from a central ~~had~~ population in α had an average blood viscosity of 3.92. With a S.D of 0.61. And the basis of this sample established 95% and 99% confidence limits for the mean blood viscosity of the central population.

$$n = 11$$

$$\bar{x} = 3.92$$

$$S = 0.61$$

$$95\% = \bar{x} \pm \frac{S}{\sqrt{n}} \text{ to } 0.05$$

$$95\% = \bar{x} + \frac{S}{\sqrt{n}} \text{ to } 0.05$$

$$95\% = 3.92 + \frac{0.61}{\sqrt{11}} (2.228)$$

$$95\% = 4.329$$

$$95\% = \bar{x} - \frac{S}{\sqrt{n}} \text{ to } 0.05$$

$$95\% = 3.92 - \frac{0.61}{\sqrt{11}} (2.228)$$

$$95\% = 3.51$$

$$99\% = \bar{x} + \frac{S}{\sqrt{n}} \text{ to } 0.01$$

$$= 3.92 + \frac{0.61}{\sqrt{11}} (3.169)$$

$$99\% = 4.502$$

$$99\% = \bar{x} - \frac{S}{\sqrt{n}} \text{ to } 0.01$$

$$99\% = 3.92 - \frac{0.61}{\sqrt{11}} (3.169)$$

$$99\% = 3.3371$$

→ f-distribution (or) f-test

Test for the equality of variance of two populations

H_0 : The variance of two populations are equal i.e., $\sigma_1^2 = \sigma_2^2$

H_1 : The variance of two populations are not equal i.e., $\sigma_1^2 \neq \sigma_2^2$.

L.S (Level of significance) (at 5% and 1%)

If $S_1^2 > S_2^2$ then $F = \frac{S_1^2}{S_2^2}$ (Large variance / Small variance)

where $S_1^2 = \frac{n_1 \sigma_1^2}{n_1 - 1}$ and $S_2^2 = \frac{n_2 \sigma_2^2}{n_2 - 1}$

Degrees of freedom (df)

df [(n₁-1), (n₂-1)]

df (v₁, v₂)

Suppose $S_2^2 > S_1^2$ then $F = \frac{S_2^2}{S_1^2}$
df [(n₂-1), (n₁-1)] df (v₂, v₁)

1. A sample of 5 observations has the variance 26.4 and another sample of 9 observations has the variance 38.9. Test the equality of variance.

⇒ H₀ : Variance of two populations / samples are equal i.e., $\sigma_1^2 = \sigma_2^2$.

H₁ : Variance of two samples are not equal i.e., $\sigma_1^2 \neq \sigma_2^2$.

By the data, $n_1 = 5$ $\sigma_1^2 = 26.4$
 $n_2 = 9$ $\sigma_2^2 = 38.9$

$$S_1^2 = \frac{n_1 \sigma_1^2}{n_1 - 1} = \frac{5(26.4)}{5 - 1} = 33$$

$$S_2^2 = \frac{n_2 \sigma_2^2}{n_2 - 1} = \frac{9(38.9)}{9 - 1} = 43.76$$

$$S_2^2 > S_1^2$$

$$F = \frac{S_2^2}{S_1^2} = \frac{43.76}{33}$$

$$F = 1.326$$

$$df [(n_2-1), (n_1-1)] = df [8, 4] \quad \begin{matrix} \cancel{df(v_1, v_2)} \\ df(v_1, v_2) \end{matrix}$$

$$= (8, 4)$$

$$df = \begin{matrix} 7.01 \\ (0.01) \\ 14.8 \end{matrix}$$

$$df(0.05) = \begin{matrix} 3.84 \\ 6.04 \end{matrix}$$

$$F = 1.3261 \quad \left\{ \begin{array}{l} < F_{0.05} = 6.04 \\ < F_{0.01} = 14.8 \end{array} \right\} \quad \text{Ho accepted}$$

2. In a study of wheat production in a sample of common 10 common sub division of equal area of 2 agriculture plots. The variance of 0.92 and 0.26. Test at 5% level of significance for equality of variance.

$\Rightarrow H_0$: Variance of two agricultural plots are equal. $\sigma_1^2 = \sigma_2^2$

H_1 : Variance of two agricultural plots are not equal $\sigma_1^2 \neq \sigma_2^2$.

By the data

$$n_1 = 10 \quad n_2 = 10$$

$$\sigma_1^2 = 0.92 \quad \sigma_2^2 = 0.26$$

$$S_1^2 = \frac{n_1 \sigma_1^2}{(n_1 - 1)} = \frac{10(0.92)}{10 - 1} = 1.022$$

$$S_2^2 = \frac{n_2 \sigma_2^2}{(n_2 - 1)} = \frac{10(0.26)}{10 - 1} = 0.288$$

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

$$F = \frac{1.099}{0.288}$$

$$F = 3.548$$

$$df((n_1-1), (n_2-1)) \quad df(9, 9)$$

$$F_{0.05} = 3.18$$

$$F_{0.01} = 5.35$$

$$F = 3.548$$

$$F_{0.05} = 3.18 = H_0 \text{ rejected}$$

3. Two random samples drawn from two normal populations are

Sample I 20 16 26 27 22 23 18 24 19 25

Sample II 27 33 42 35 32 34 38 28 41 43 30 3

Obtain the estimates of variance of the population and test 5% level of significance whether two populations have same the same variance.

$$n_1 = 10 \quad n_2 = 12$$

$$\bar{x} = \frac{1}{n_1} (\sum x_i) = 22$$

$$S_1^2 = \frac{1}{n_1 - 1} \left[\sum x^2 - \frac{1}{n_1} (\sum x)^2 \right]$$

$$= \frac{1}{9} \left[(20^2 + 16^2 + 26^2 + 27^2 + 22^2 + 23^2 + 18^2 + 24^2 + 19^2 + 25^2) - \frac{1}{10} (220)^2 \right]$$

$$= \frac{1}{9} [4960 - \frac{1}{10} (220)^2]$$

$$S_1^2 = 13.33$$

$$\bar{y} = \frac{1}{n_2 - 1} \left[\sum y_j \right]$$

$$\sum y = 420$$

$$\bar{y} = \frac{1}{12} \sum y = 42$$

$$S_2^2 = \frac{1}{n_2 - 1} \left[\sum y_j^2 - \frac{1}{n_2} \left(\sum y \right)^2 \right]$$

$$= \frac{1}{11} \left[15014 - \frac{1}{12} (420)^2 \right]$$

$$S_2^2 = 28.54$$

$$S_2^2 > S_1^2 \quad \text{Then}$$

$$F = \frac{S_2^2}{S_1^2} = \frac{28.54}{13.33} = 2.141$$

$$F_{0.05} = 2.1$$

$$df((n_1 - 1), (n_2 - 1)) = (9, 11) \quad \therefore v = (11, 9)$$

$$= 2.90 \text{ at } 5\%$$

4. Sample I 16 17 18 19 20 21 22 24 26 27
 Sample II 19 22 23 25 26 28 29 30 31 32 35 36
 Obtain the estimates using 5% of LS.

$$n_1 = 10 \quad n_2 = 12$$

$$\bar{x} = \frac{1}{n_1 - 1} \left(\sum x_i \right)$$

$$\sum x = 210$$

$$\bar{x} = \frac{1}{10 - 1} \left(\sum x_i \right)$$

$$\bar{x} = 23.33$$

$$S_1^2 = \frac{1}{n_1 - 1} \left[\sum x^2 - \frac{1}{n_1} (\sum x)^2 \right]$$

$$S_1^2 = \frac{1}{9} \left[16^2 + 17^2 + 18^2 + 19^2 + 20^2 + 21^2 + 22^2 + 24^2 + 26^2 + 27^2 - \frac{1}{10} [210]^2 \right]$$

$$S_1^2 = \frac{1}{9} \left[4536 - \frac{1}{10} (210)^2 \right]$$

$$S_1^2 = 14$$

$$\sum y = 336$$

$$S_2^2 = \frac{1}{n_2 - 1} \left[\sum y^2 - \frac{1}{n_2} (\sum y)^2 \right]$$

$$S_2^2 = \frac{1}{11} \left[9706 - \frac{1}{12} (336)^2 \right]$$

$$S_2^2 = 87.09$$

$$S_2^2 > S_1^2 \text{ then } F = \frac{S_2^2}{S_1^2} = 1.935$$

$$df[(n_2 - 1), (n_1 - 1)]$$

$$df(11, 9) \quad v_2, v_1$$

$$F = 1.935 < 2.90 \text{ Ho accepted}$$

2

→ Central limit theorem:

Suppose a sample of size n is selected from a population that has mean (μ) and standard deviation (σ), then $x_1, x_2, x_3, \dots, x_n$ be the n observations. They are independent and identically distributed with mean ($\bar{x}$) $= \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum x_i$. The central limit theorem states that the sample mean ($\bar{x}$) follows approximately the normal distribution with mean (μ) and standard deviation $\frac{\sigma}{\sqrt{n}}$ (is called standard error) i.e.,

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \mu \text{ and } \sigma \text{ are mean and}$$

standard deviation of the population from where the sample was selected (and the sample size becomes large, $(n > 30)$)

→ Confidence intervals

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

$$CI = \mu \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$$

Confidence level (CI)

99%

2.58

98%

2.33

95%

1.96

90%

1.645

50%

0.6745

→ Problems:

1. State Central limit theorem. Use the theorem to evaluate $P[50 < \bar{X} < 56]$ where $\bar{X}$ mean of a random sample of size 100 from an infinite population with mean 53 and variance 400

↑ mean
 $\mu = 53$; variance $(\sigma^2) = 400$ $\sigma = \sqrt{400}$ $\sigma = 20$

WKT Central Limit Theorem $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$\Rightarrow \bar{X} \sim N\left(53, \frac{20}{\sqrt{100}}\right)$ $\bar{X} \sim N(53, 2)$

WKT $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $Z = \frac{53 - 53}{\frac{20}{\sqrt{100}}}$ $Z = \frac{\bar{X} - 53}{2}$

if $\bar{X} = 50$

$Z = \frac{50 - 53}{2}$ $Z = -1.5$

if $\bar{X} = 56$ $Z = \frac{56 - 53}{2}$ $Z = 1.5$

$P(50 < \bar{X} < 56) = P(-1.5 < Z < 1.5)$

$= P(-1.5 < Z < 0) + P(0 < Z < 1.5)$

$= P(0 < Z < 1.5) + P(0 < Z < 1.5)$

$= 2P(0 < Z < 1.5)$

$= 2\phi(1.5)$

$= 2 \cdot 0.358$

$2(0.4332)$

$= 0.8664$

$1.5 = 1.5$
 0.50
 $1.5 + 0.00$

2. An unknown distribution has a mean of 90 and SD of 15. Sample size 25 drawn randomly for the population. Find the probability that the sample mean is between 85 and 92.

$\Rightarrow$ mean $(\mu) = 90$ $n = 25$

$\sigma = 15$

$P(85 < \bar{X} < 92)$

$\Rightarrow$ Central Limit Theorem $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$$\bar{x} \sim N\left(90, \frac{15}{\sqrt{25}}\right)$$

$$\bar{x} \sim N(90, 3)$$

$$\text{WKT } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 90}{3}$$

when $\bar{x} = 85$,

$$Z = \frac{85 - 90}{3} = \frac{-5}{3} = -1.666$$

when $\bar{x} = 92$

$$Z = \frac{92 - 90}{3} = \frac{2}{3} = 0.66$$

$$P(85 < \bar{x} < 92) = P(-1.66 < Z < 0.66)$$

$$= P(0 < Z < 1.66) + P(0 < Z < 0.66)$$

$$= 0.4452 + 0.2454$$

$$= 0.4452 + 0.2454$$

$$= 0.6906$$

3. A random sample of size 64 is taken from an infinite population with mean 112, variance 144. Using Central Limit Theorem, find the probability of getting sample mean $\bar{x}$ greater than 114.5

$$n=64 \quad \text{mean } (\mu) = 112, \quad \sigma^2 = 144 \quad \sigma = 12$$

$$\text{CLT, } \bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \sim N\left(112, \frac{12}{\sqrt{64}}\right) \quad \bar{x} \sim N(112, 1.5)$$

$$\text{WKT } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 112}{1.5}$$

$$\text{when } \bar{x} = 114.5$$

$$Z = \frac{114.5 - 112}{1.5} = Z = \frac{114.5 - 112}{1.5} \quad Z = 1.66$$

$$\begin{aligned} P(\bar{x} > 114.5) &= 0.5 - \phi(Z_1) & 1.66 \\ &= 0.5 - \phi(1.66) & = 1.6 + 0.00 \\ &= 0.5 - 0.4515 \end{aligned}$$

$$P(\bar{x} > 114.5) = 0.0485$$

4. An electrical firm manufactures light bulbs that have a length of life that is approx normally distribute with mean 800 hrs and SD of 40 hrs. Find the probability that a random sample of 16 bulbs an average life less than 775 hrs

$$\Rightarrow n=16 \text{ mean } (\mu) = 800, \sigma = 40$$

$$\text{LT, } \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{X} \sim N\left(800, \frac{40}{\sqrt{16}}\right) = \bar{X} \sim N(800, 10)$$

$$\text{WKT } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 800}{10}$$

$$\text{when } \bar{x} = 775$$

$$Z = \frac{775 - 800}{10} = -2.5$$

$$\begin{aligned} P(\bar{x} < 775) &= 0.5 + \phi(Z_1) \\ &= 0.5 + \phi(-2.5) \end{aligned}$$

$$= 0.5 - 0.4938$$

$$P(\bar{x} < 775) = 0.0062$$

5. An unknown distribution In a recent study report on the blog, the mean age of tablet users is 34 yrs. Suppose the S.D is 15 yrs, take a sample size 100. Using central limit theorem, find the probability that the sample mean age is more than 30 yrs.

$$\Rightarrow \text{mean}(\mu) = 34 \quad \sigma = 15 \quad n = 100$$

$$p(\bar{x} > 30)$$

$$\text{CLT, } \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \bar{X} \sim N\left(34, \frac{15}{\sqrt{100}}\right)$$

$$\bar{X} \sim N(34, 1.5)$$

$$\Rightarrow Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad Z = \frac{\bar{x} - 34}{1.5}$$

$$\text{when } \bar{x} = 30$$

$$Z = \frac{30 - 34}{1.5} = -2.66$$

$$p(\bar{x} > 30) = 0.5 - \phi(Z_1)$$

$$= 0.5 - \phi(-2.66)$$

$$2.6 + 0.06$$

$$= 0.5 + \phi(2.66)$$

$$= 0.5 + 0.4961$$

$$p(\bar{x} > 30) = 0.9961$$

Problem

1. The height of a random 50 college students showed a mean of 174.5 cm and S.D of 6.9 cm. Construct a 99% CI for the mean height of all college students.

$$\Rightarrow \text{The given data, } n = 50$$

$$SD = 6.9 \text{ cm}$$

$$\text{mean } \mu = 174.5 \text{ cm}$$

WRT, The table value for Z test at 99% is 2.58

$$\text{Confidence Interval CI} = \mu \pm z \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$\text{CI} = 174.5 \pm 2.58 \left(\frac{6.9}{\sqrt{50}} \right)$$

$$\text{CI} = 174.5 \pm 2.5176$$

$$\text{CI} = 2.5176 \quad 174.5 + 2.5176 \quad \text{and} \quad 174.5 - 2.5176$$

CI = 177.01 and 171.98 are the confidence level.

2. A random sample size 25 from a normal distribution variance is 4, sample mean 78.3, obtain a 99% CI for the mean.

$$\Rightarrow n = 25 \quad \mu = 78.3 \quad z = 2.58$$

$$\sigma^2 = 4 \quad \sigma = 2$$

$$\text{CI} = \mu \pm z \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$\text{CI} = 78.3 \pm 2.58 \left(\frac{2}{\sqrt{25}} \right)$$

$$= 78.3 \pm 1.032$$

$$\text{CI} = 78.3 + 1.032 \quad \text{and} \quad \text{CI} = 78.3 - 1.032$$

$$\text{CI} = 79.332 \quad \text{and} \quad \text{CI} = 77.268 \quad \text{are the CI.}$$

3. Suppose that 10, 12, 16, 19 is a sample taken from a normal population with variance 6.25. Find at 95% CI for the population mean.

$$\Rightarrow n = 4 \quad \sigma^2 = 6.25 \quad \sigma = 2.5 \quad z = 1.96$$

$$\mu = 14.25$$

$$\text{CI} = \mu \pm z \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$\text{CI} = 14.25 \pm 1.96 \left(\frac{2.5}{\sqrt{4}} \right)$$

$$CI = 14.25 \pm 2.45$$

$$CI = 14.25 + 2.45$$

$$CI = 16.7$$

$$CI = 14.25 - 2.45$$

$$CI = 11.8 \text{ are the CI}$$

4. The mean of random sample of size 20 from a normal distribution with mean and variance 81.2 and 80, find at 90% and 95% CI for mean.

$$n=20 \quad \mu=81.2 \quad \sigma^2=80 \quad \sigma=8.94$$

for 90%.

$$Z = 1.645$$

$$CI = \mu \pm Z \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$CI = 81.2 \pm 1.645 \left(\frac{8.94}{\sqrt{20}} \right)$$

$$CI = 81.2 \pm 3.288$$

$$CI = 81.2 + 3.288$$

$$CI = 84.488$$

$$CI = 81.2 - 3.288$$

$$CI = 77.912$$

for 95%.

$$Z = 1.96$$

$$CI = \mu \pm Z \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$CI = 81.2 \pm 1.96 \left(\frac{8.94}{\sqrt{20}} \right)$$

$$CI = 81.2 \pm 3.918$$

$$CI = 81.2 + 3.918$$

$$CI = 85.118$$

$$CI = 77.282$$