MODULE-4. STATISTICAL INFERENCE - II

CONTENTS.

SAMPLING DISTRIBUTIONS

The probability distribution of a statistic is called a sampling distribution.

Sampling distribution of Means

Consider the mean $\overline{X}$. Suppose that a random sample of n observations is taken from a normal population with mean μ and variance σ^2. Each observation X_i, $i=1,2,\ldots n$ of the random sample will then have the same normal distribution as the population being sampled.

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots\ldots + X_n)$$

has a normal distribution with mean

$$\mu_{\bar{x}} = \frac{1}{n}(\underbrace{\mu+\mu+\cdots\cdots+\mu}_{n \text{ terms}}) = \mu \quad \text{and}$$

$$\text{Variance } \sigma^2_{\bar{x}} = \frac{1}{n^2}(\underbrace{\sigma^2+\sigma^2+\cdots\cdots+\sigma^2}_{n \text{ terms}}) = \frac{\sigma^2}{n}$$

If we are sampling from a population with unknown distribution, either finite or infinite the sampling distribution of $\bar{x}$ will be normal mean μ and variance σ^2 provided sample size is large n.

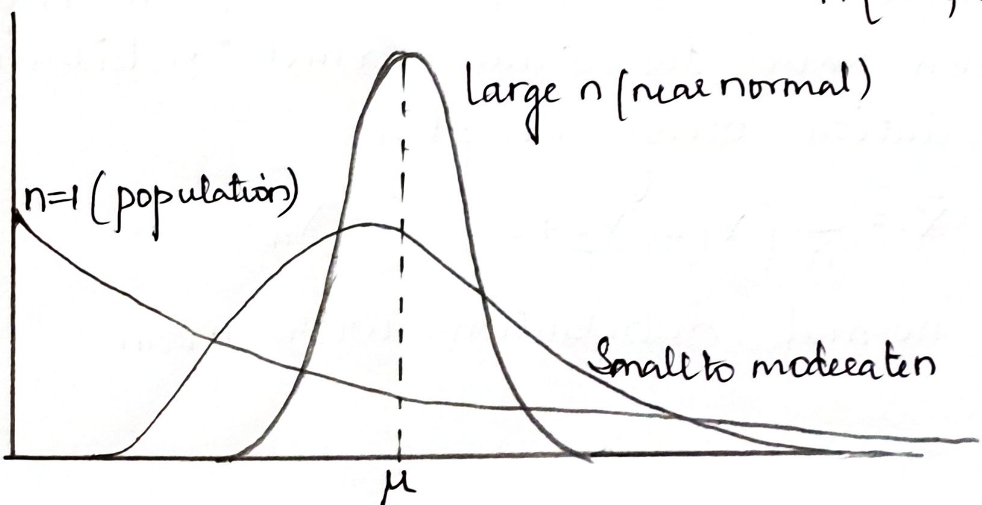
The above result consequence called as Central limit theorem.

CENTRAL LIMIT THEOREM.

Statement : If $\bar{x}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

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

, as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$



CENTRAL LIMIT THEOREM.

STATEMENT: The central limit theorem states that the sample mean $\bar{x}$ follows approximately the normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$,
ie $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, $\Rightarrow z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ where μ, σ are mean and standard deviation of the population from the sample.

PROBLEMS.

1. State Central limit theorem. Use the theorem to evaluate $P(50 < \bar{x} < 56)$ where $\bar{x}$ represents the mean of a random sample of size 100 from an infinite population with mean $\mu = 53$ and variance $\sigma^2 = 400$.

Soln: 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 also called standard error), ie, $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ, σ are mean and standard deviation of the population from where the sample.

Given,

Sample size $n = 100$

Mean of the population $\mu = 53$

Variance of the population $\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$

$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{x} \sim N\left(53, \frac{20}{\sqrt{100}}\right) \Rightarrow \bar{x} \sim N(53, 2)$

We know that $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-53}{20/\sqrt{100}}$

$$z = \frac{\bar{x}-53}{2}$$

At $\bar{x} = 50$, $z = \frac{50-53}{2} = \frac{-3}{2} = -1.5 = z_1$

At $\bar{x} = 56$, $z = \frac{56-53}{2} = \frac{3}{2} = 1.5 = z_1$

$$\therefore P(50 < \bar{x} < 56) = P(-1.5 < z < 1.5)$$
$$= P(-1.5 < z < 0) + P(0 < z < 1.5)$$
$$= P(0 < z < 1.5) + P(0 < z < 1.5)$$
$$= 2P(0 < z < 1.5)$$
$$= 2\phi(1.5)$$
$$= 2 \times 0.4332$$
$$P(50 < \bar{x} < 56) = 0.8664$$

2. An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size n=25 are drawn randomly from the population. find the probability that the sample mean is between 85 and 92.

Soln : Given, sample size $n = 25$
Mean of the population $\mu = 90$
Variance of the population $\sigma = 15$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-90}{15/\sqrt{25}} \Rightarrow z = \frac{\bar{x}-90}{3}$$

At $\bar{x} = 85 \Rightarrow z = \frac{85-90}{3} = \frac{-5}{3} = -1.5$

$\bar{x} = 92 \Rightarrow z = \frac{92-90}{3} = \frac{2}{3} = 0.66$

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

$$p(-1.66 < z < 0.66) = p(0 < z < 1.66) + p(0 < z < 0.66)$$

$$= 0.4515 + 0.2454$$

$$\Rightarrow p(-1.66 < z < 0.66) = 0.6965$$

3. A random sample of size 64 is taken from an infinite population having mean 112 and variance 144. Using central limit theorem, find the probability of getting the sample mean $\bar{x}$ greater than 114.5

Soln: Given, sample size $n=64$
Mean of the population $\mu=112$
Variance of the population $\sigma^2=144$, $\sigma=12$

$$Z = \frac{\bar{x} - 112}{12/\sqrt{64}} = \frac{\bar{x} - 112}{1.5}$$

At $\bar{x} = 114.5 \Rightarrow z = \dfrac{114.5 - 112}{1.5} = 1.66$.

$$p(\bar{x} > 114.5) = p(z > 1.66)$$

$$\Rightarrow p(z > 1.66) = 0.5 - p(0 < z < 1.66)$$

$$= 0.5 - 0.4515$$

$$p(z > 1.66) = 0.0489$$

4. Let $\bar{x}$ denote the mean of a random sample of size 100 from a distribution, that is $\chi^2(50)$. Compute an approximate value of $p(49 < \bar{x} < 51)$.

Soln: The sample size n is = 100

chi-square distribution $\chi^2(50)$, $\chi^2(d.f)$

where d.f is 50,

The mean and variance of chi-square distribution is given as $\mu = 50$

$\therefore \sigma^2 = 2 \times d.f = 2 \times 50 = 100 \Rightarrow \sigma^2 = 100$

$\sigma = \sqrt{100}$

$\boxed{\sigma = 10}$

$\therefore Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - 50}{10/\sqrt{100}} = \dfrac{\bar{x} - 50}{1}$

$\therefore$ At $(\bar{x} = 49) \Rightarrow z = \dfrac{49 - 50}{1} = \dfrac{-1}{1} = -1 = z_1$

$\bar{x} = 51 \Rightarrow z = \dfrac{51 - 50}{1} = \dfrac{1}{1} = 1 = z_2$

$p[49 < \bar{x} < 51) = p(-1 < z < 1)$

$= p(-1 < z < 0) + p(0 < z < 1)$

$= p(0 < z < 1) + p(0 < z < 1)$

$= 2p(0 < z < 1)$

$= 2 \times 0.3416$

$p(49 < \bar{x} < 51) = 0.6826$

5. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distribute with mean 800 hours. and a standard deviation of 40 hours. find the probability that a random sample of 16 bulbs will have an average life of life of less than 775 hours.

Soln: Total no. of bulbs $n = 16$
An average life of bulbs $\mu = 800$
Std deviation of the bulbs $\sigma = 40$

At $\bar{x} = 775$, $\Rightarrow z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{775 - 800}{40/\sqrt{16}} = \frac{775 - 800}{10}$

$\therefore z = \frac{-25}{10} = -2.5$

At $p(\bar{x} < 775) = p(z < -2.5)$
$= p(z > 2.5)$
$= 0.5 - \phi(2.5)$
$= 0.5 - 0.4938$
$p(\bar{x} < 775) = 0.0062$

CONFIDENCE INTERVAL.

1. A random sample of size 25 from a normal distribution $(\sigma^2 = 4)$ yields, sample mean $\bar{x} = 78.3$. obtain a 99% confidence interval interval for μ.

Soln: Given the sample size $n = 25$
Mean of sample $\bar{x} = 78.3$
Std Deviation $\sigma = 2$.

We know, Confidence level of 99%, the corresponding z value is 2.58. This is determined from the normal distribution table.

Confidence Interval $C.I = \mu = \text{Mean} \pm z(\text{Std Dev}/\sqrt{\text{sample size}})$

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

$$C.I = \mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

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

$$= 78.3 \pm 1.032$$

$$C.F = (78.3 - 1.032 ,\ 783 + 1.032)$$

$$C.F = (77.268 ,\ 79.332)$$

2. Suppose that 10, 12, 16, 19 is a sample taken from a normal population with variance 6.25. find at 95% confidance Intuval for the population mean.

Soln: Given, samples are 10, 12, 16 and 19.

Sample size $n = 4$

Mean, $\bar{x} = 14.25$

Variance, $\sigma^2 = 6.25 \Rightarrow \sigma = \sqrt{6.25} = 2.5$

Confidence level of 95%, Corresponding z value is 1.96

$C.I = \mu =$ mean $\pm$ z (std Dev / $\sqrt{\text{Samplesize}}$

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

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

$$C.F = 14.25 \pm 2.45$$

$$\Rightarrow C.F = (14.25 - 2.45 ,\ 14.25 + 2.45) = (11.80 ,\ 16.70)$$

Test of Significance for Small Samples

MODULE-03

Student's "t" distribution

A test of significance for small sample

Consider a small sample of size n, drawn from a normal population with mean μ and standard deviation σ. If '$\bar{x}$' and 's' be the sample mean and sample standard deviation, then the statistic, 't' is defined as

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

Where, $\bar{x} \rightarrow$ sample mean

$\mu \rightarrow$ population mean

$s \rightarrow$ standard deviation of sample

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

't' test of two samples is

$$t = \frac{(\bar{x}_2-\bar{x}_1)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \quad \text{Where, } s^2 = \frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2-2}$$

$$\text{or} \quad s^2 = \frac{\sum(x-\bar{x})^2+\sum(y-\bar{y})^2}{n_1+n_2-2}$$

Confidence limits for 't' distribution.

95% Confidence limits for mean 'μ' in 't' distribution given by $\bar{x} \pm t_{0.05}\left(\frac{s}{\sqrt{n}}\right)$.

Degree of freedom: Degrees of freedom is nothing but the number of given values minus 1 ie, $(n-1)$

Note
* If $t_{cal} < t_{tab}$ then we accept the hypothesis.
* If $t_{cal} > t_{tab}$ then we reject the hypothesis.

$t_{cal} \rightarrow$ calculated values
$t_{tab} \rightarrow$ tabulated values.

Problems

1) A certain stimulus administered to each of the 12 patients resulted in the following change in blood pressure 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase the blood pressure? Given $t_{0.05} = 2.201$ for 11 degree of freedom.

Soln. Given, $n = 12$, Assume $\mu = 0$ ($\because$ population mean not given)

We know that, 't' distribution

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

Let H_0: stimulus will not increase the blood pressure.

Now, $\bar{x} = \frac{\Sigma x}{n} = \frac{31}{12} = 2.58$

$$\boxed{\bar{x} = 2.58}$$

To find 's' we have $s^2 = \frac{1}{n-1}\Sigma(x - \bar{x})^2$

$$= \frac{1}{11}\, 104.96$$

$$s^2 = 9.54 \Rightarrow \boxed{s = 3.08}$$

x	$(x-\bar{x})^2$
5	5.8564
2	0.3364
8	29.3764
-1	12.8164
3	0.1764
0	6.6564
6	11.6964
-2	20.9764
1	2.4964
5	5.8564
0	6.6564
4	2.0164
31	104.96

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

$$s^2 = \frac{1}{11}\left[(5-2.58)^2+(2-2.58)^2+(8-2.58)^2+(-1-2.58)^2+(3-2.58)^2+(0-2.58)^2+(6-2.58)^2+(-2-2.58)^2+(1-2.58)^2+(5-2.58)^2+(0-2.58)^2+(4-2.58)^2\right]$$

$$= \frac{1}{11}\times 104.96$$

$$s^2 = 9.54 \Rightarrow \boxed{s = 3.08}$$

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

$$= \frac{2.58-0}{3.08}\sqrt{12}$$

$$\boxed{t_{cal} = 2.89} > t_{tab} = 2.201$$

Since, $t_{cal} > t_{tab}$

$\therefore$ Null Hypothesis (H_0) is rejected.

2) 10 individuals are choosen 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. ($t_{0.05} = 2.262$ for 9 d.f).

Soln Given, $n = 10 \rightarrow$ sample size

$\mu = 66 \rightarrow$ population mean

H_0: let $\mu = 66$ inches.

Wkt, $t_{cal} = \frac{\bar{x}-\mu}{s}\cdot\sqrt{n}$

We have, $\bar{x} = \frac{\sum x}{n} = 67.8$

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

$= \frac{1}{9}\left[(63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + (69-67.8)^2 + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2\right]$

$= \frac{1}{9} \times 81.6$

$s^2 = 9.06 \Rightarrow \boxed{s = 3.01}$

$t_{cal} = \frac{67.8-66}{3.01}\sqrt{10} = 1.89 < 2.262$

Given $t_{0.05} = 2.262$ (t_{tab}) $\Rightarrow t_{cal} < t_{tab}$

$\therefore$ Null Hypothesis (H_0) is accepted.

3) The nine item of a sample have the following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differ significantly from the assumed mean of 47.5?

Sol^n Given, $n = 9$, $\mu = 47.5$

H_0: let $\mu = 47.5$, $H_1: \mu \neq 47.5$

Wkt, $t_{cal} = \frac{\bar{x}-\mu}{s}\sqrt{n}$

We have, $\bar{x} = \frac{\sum x}{n} = \frac{442}{9} = 49.11$

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

$= \frac{1}{8}\left[(45-49.11)^2 + (47-49.11)^2 + (50-49.11)^2 + (52-49.11)^2 + (48-49.11)^2 + (47-49.11)^2 + (49-49.11)^2 + (53-49.11)^2 + (51-49.11)^2\right]$

$= \frac{54.9}{8} = 6.8625 \Rightarrow \boxed{s = 2.6196}$

Now, $t_{cal} = \frac{49.11 - 47.5}{2.6196}\sqrt{9} = 1.84 < 2.306$

Since, $t_{cal} < t_{tab}$

$\therefore$ The null hypothesis is accepted.

4) A sample of 10 measurments of the diameter of a sphere gave a mean of 12cm and S.D 0.15cm Find the 95% confidence limits for the actual diameter. ($t_{0.05} = 2.262$)

Soln Given, $n = 10$, $\bar{x} = 12$, $s = 0.15$

$n \to$ sample size
$\bar{x} \to$ sample mean
$s \to$ sample S.D

WKT, the 95% confidence limits for mean: $\bar{x} \pm t_{0.05}\left(\frac{s}{\sqrt{n}}\right)$

i.e $\bar{x} - t_{0.05}\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + t_{0.05}\left(\frac{s}{\sqrt{n}}\right)$

$$12 - 2.262\left(\frac{0.15}{\sqrt{10}}\right) \leq \mu \leq 12 + 2.262\left(\frac{0.15}{\sqrt{10}}\right)$$

$$\boxed{11.893 \leq \mu \leq 12.107}$$

5) A sample of 11 rats from a central population had an average blood viscosity of 3.92 with a S.D of 0.61. On the basis of this sample, establish 95% confidence limits for μ, the mean blood viscosity of the central population. ($t_{0.05} = 2.228$ for 10 degree of freedom).

Soln Given, $n = 11$, $\bar{x} = 3.92$, $s = 0.61$

WKT, 95% confidence limits for μ is $\bar{x} \pm t_{0.05}\left(\frac{s}{\sqrt{n}}\right)$

i.e, $\bar{x} - t_{0.05}(s/\sqrt{n}) \leq \mu \leq \bar{x} + t_{0.05}\left(\frac{s}{\sqrt{n}}\right)$

$$3.92 - 2.228\left(\frac{0.61}{\sqrt{11}}\right) \leq \mu \leq 3.92 + 2.228\left(\frac{0.61}{\sqrt{11}}\right)$$

$$\boxed{3.51 \le \mu \le 4.33}$$

6) A random sample of 10 boys had the following IQ 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean IQ of 100 at 5% level of significance? (Given $t_{0.05} = 2.262$ for 9 d.f).

Soln. $n=10$, let $H_0: \mu = 100$ (null hypothesis)
$H_1: \mu \neq 100$ (Alternative hypothesis)

We have, $\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$

$$s^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2 = \frac{1}{9}\left[(70-97.2)^2 + (120-97.2)^2 + (110-97.2)^2 + (101-97.2)^2 + (88-97.2)^2 + (83-97.2)^2 + (107-97.2)^2 + (95-97.2)^2 + (98-97.2)^2 + (100-97.2)^2\right]$$

$$s^2 = \frac{1}{9} \times 1833.6$$

$$s^2 = 203.73 \Rightarrow \boxed{s = 14.27}$$

Now, $t_{cal} = \frac{\bar{x}-\mu}{s}\sqrt{n} = \frac{97.2-100}{14.27}\sqrt{10}$

$= -0.6205 < 2.262$

Since, $t_{cal} < t_{tab} \Rightarrow H_0$ is accepted.

7). Two types of batteries are tested for their length of life and the following results are obtained:

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs, $\sigma_1^2 = 100$

Battery B : $n_2 = 10$, $\bar{x}_2 = 500$ hrs, $\sigma_2^2 = 121$

Compute student's 't' and test whether there is a significant difference in the two means (Given $t_{0.05} = 2.101$ for 18 d.f)

Sol^n Given

Battery A: $n_1 = 10$, $\bar{x}_1 = 500$, $\sigma_1^2 = 100$

B: $n_2 = 10$, $\bar{x}_2 = 560$, $\sigma_2^2 = 121$

W.K.T
$$s^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2-2}$$

$$= \frac{10(100) + 10(121)}{10+10-2}$$

$$s^2 = 122.78$$

$$\boxed{s = 11.0805}$$

We have,
$$t_{cal} = \frac{(\bar{x}_2 - \bar{x}_1)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{560-500}{11.0805\sqrt{0.1+0.1}}$$

$$t_{cal} = 12.1081 \simeq 12.11$$

The value of t_{cal} is greater than the t_{tab} (2.101) for 18 d.f at all levels of significance.

8). A group of boys and girls were given an intelligent test. The mean score, S.D score and numbers 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 20 d.f)

Sol^n

Given, $n_1 = 12$, $\bar{x}_1 = 74$, $\sigma_1 = 8$ (boys)

$n_2 = 10$, $\bar{x}_2 = 70$, $\sigma_2 = 10$ (girls)

WKT, $s^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2-2} = \frac{(12\times 8^2 + 10\times 10^2)}{12+10-2} = 88.4$

$\boxed{S = 9.4}$

we have, $t = \frac{(\bar{x}_2 - \bar{x}_1)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{74-70}{9.4\sqrt{\frac{1}{12}+\frac{1}{10}}}$

$t_{cal} = 0.99 < 2.086$

Since, $t_{cal} < t_{tab}$, Thus, the hypothesis that there is a difference b/w the means of two groups is accepted at 5% level of significance.

9) Two horses A and B were tested according to the time 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.

Soln Given, $x: 28, 30, 32, 33, 33, 29, 34, \Rightarrow \bar{x} = \frac{1}{n_1}\sum x_i = 31.30$

$y: 29, 30, 30, 24, 27, 29 \Rightarrow \bar{y} = \frac{\sum y_i}{n_2} = 28.20$

$\sum(x-\bar{x})^2 = (28-31.3)^2 + (30-31.3)^2 + (32-31.3)^2 + (33-31.3)^2 + (33-31.3)^2 + (29-31.3)^2 + (34-31.3)^2 = 31.4$

$\sum(y-\bar{y}_i)^2 = (29-28.2)^2 + (30-28.2)^2 + (30-28.2)^2 + (24-28.2)^2 + (27-28.2)^2 + (29-28.2)^2 = 26.84$

$\therefore s^2 = \frac{\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2}{n_1+n_2-2} = \frac{31.4+26.84}{7+6-2} = 5.2973$ $\boxed{S = 2.3016}$

Now, $t_{cal} = \frac{(\bar{x}-\bar{y})}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.30-28.2}{2.3016\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.42 \begin{cases} > t_{0.05} = 2.2 \\ < t_{0.02} = 2.72 \end{cases}$

IV) Eleven school boys were given a test in drawing. further they were given a month's tuition and a second test of equal difficulty was held at the end of it. Do the marks give the evidence that Students have benefitted by extra coaching? for d.f = 10, 0.05 = 2.228. Boys

Boys	1	2	3	4	5	6	7	8	9	10	11
I-test	23	20	19	21	18	20	18	17	23	16	19
II-test	24	19	22	18	20	22	20	20	23	20	17

Soln:- we have, $\bar{x} = \frac{\Sigma x_i}{n_i} = \frac{214}{11} = 19.5$

$$\bar{y} = \frac{\Sigma y_i}{n_i} = \frac{225}{11} = 20.5$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
23	3.5	12.25	24	3.5	12.25
20	0.5	0.25	19	-1.5	2.25
19	-0.5	0.25	22	1.5	2.25
21	1.5	2.25	18	-2.5	6.25
18	-1.5	2.25	20	-0.5	0.25
20	0.5	0.25	22	1.5	2.25
18	-1.5	2.25	20	-0.5	0.25
17	-2.5	6.25	20	-0.5	0.25
23	3.5	12.25	23	2.5	6.25
16	-3.5	12.25	20	-0.5	0.25
19	-0.5	0.25	17	-3.5	12.25
		$\Sigma(x_i - \bar{x})^2 = 50.75$			$\Sigma(y_i - \bar{y})^2 = 44.75$

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

$$= \frac{50.75 + 44.75}{11 + 11 - 2}$$

$$s^2 = 4.775$$

$$\boxed{s = 2.18}$$

W.K.T

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (or)$$

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

$$= \frac{20.5 - 19.5}{2.18\sqrt{\frac{1}{11} + \frac{1}{11}}}$$

$$\boxed{t = 1.073}$$

$t_{cal} = 1.073 < t_{tab} = 2.228$

$\therefore t_{cal} < t_{tab}$.

It is accepted.

i) From a random sample of 10 pigs fed on diet A, the increases in weight in a certain period were 10, 6, 16, 17, 13, 12, 8, 14, 15, 9 lbs for another random sample of 12 pigs fed on diet B, the increases in the sam period were 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17 lbs. Test whether diets A and B differ significantly as regards their effect on increases in weight? ($t_{0.05} = 2.09$)

Soln:- We have, $\bar{x} = \frac{\sum x_i}{n_i} = \frac{120}{10} = 12$

x_i $\quad x_i - \bar{x}$

$\bar{y} = \frac{\sum y_i}{n_i} = \frac{180}{12} = 15$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
10	-2	4	7	-8	64
6	-6	36	13	-2	4
16	4	16	22	7	49
17	5	25	15	0	0
13	1	1	12	-3	9
12	0	0	14	-1	1
8	-4	16	18	3	9
14	2	4	8	-7	49
15	3	9	21	6	36
9	-3	9	23	8	64
			10	-5	25
			17	2	4
$\sum \bar{x} = 120$		$\sum (x_i - \bar{x})^2 = 120$	$\sum \bar{y} = 180$		$\sum (y_j - \bar{y})^2 = 314$

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

$$= \frac{120+314}{10+12-2}$$

$$S^2 = 21.7$$

$$\boxed{S = 4.65}$$

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

$$= \frac{15-12}{4.65\sqrt{\frac{1}{10}+\frac{1}{12}}}$$

$$\boxed{t = 1.6}$$

$$t_{cal} = 1.6 < t_{tab} = 2.09$$

$$\therefore t_{cal} < t_{tab}$$

$\therefore$ It is accepted.

Chi-square distribution

Chi-square distribution measure the correspondence between the theoretical frequencies and observed frequencies.

If O_i $(i=1,2,3,\ldots n)$ and E_i $(i=1,2,3\ldots n)$ respectively represents the observed and estimated frequencies then the statistic "chi-square" denoted by 'χ' is defined as follows

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

Where degree of freedom is $n-1$

Note:-

* The theoretical frequencies are found by prob-ability distribution of Binomial and Poisson distribution.

* If the calculated value of 'χ^2' is less than the tabulated value of "χ^2" at specific level of significance then hypothesis is accepted. otherwise hypothesis is rejected.

Problems

1) The theory predicts the proportion of beans in the four groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beans, the number in the four groups were 882, 313, 287 and 118. The goodness of fit χ^2 value of above data is approximately equal to ?

or

Does the experimental result support the theory (at 5% level of significance for 3, $\chi^2_{0.05} = 7.815$)

Sol^n. Total number of beans:

$\Rightarrow 882 + 313 + 287 + 118 = 1600$.

Sum of ratios $= 9+3+3+1 = 16$

$\Rightarrow E(A) = 1600 \times \frac{9}{16} = 900$

$E(B) = 1600 \times \frac{3}{16} = 300$

$E(C) = 1600 * \frac{3}{16} = 300$

$E(D) = 1600 \times \frac{1}{16} = 100$

O_i	882	313	287	118
E_i	900	300	300	100
$(O_i - E_i)^2$	324	169	169	324

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{324}{900} + \frac{169}{300} + \frac{169}{300} + \frac{324}{100}$$

$\chi^2 = 4.72 < 7.815$

Hence, the fitness is good.

2). In experiments on pea breeding the following frequencies of seeds were obtained

Round and Yellow	Wrinkled and Yellow	Round and Green	Wrinkled and Green	Total
315	101	108	32	556

Theory predicts that the frequencies should be in proportions 9:3:3:1. Examine the correspondence between theory and experiment.

Sol^n.

Total number of seeds ⇒ 556

Sum of ratios = 9+3+3+1 = 16

$\Rightarrow E(A) = 556 \times \frac{9}{16} = 312.75$

$E(B) = 556 \times \frac{3}{16} = 104.25$

$E(C) = 556 \times \frac{3}{16} = 104.25$

$E(D) = 556 \times \frac{1}{16} = 34.75$

O_i	315	101	108	32
E_i	312.75	104.25	104.25	34.75
$(O_i-E_i)^2$	5.0625	10.5625	14.0625	7.5625

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i-E_i)^2}{E_i}$$

$$= \frac{5.0625}{312.75} + \frac{10.5625}{104.25} + \frac{14.0625}{104.25} + \frac{7.5625}{34.75}$$

$= 0.47 < 7.815$

Hence, the fitness is good.

3) Five dice were thrown 96 times and the numbers 1, 2 or 3 appearing on the face of the dice follows the frequency distribution as below.

No of dice showing 1, 2, or, 3	5	4	3	2	1	0
Frequency.	7	19	35	24	8	3

test the hypothesis that the data follows a binomial distribution ($\chi^2_{0.05} = 11.07$ for 5 d.f)

Soln Given, that data follows binomial distribution

$n = 5$, p = probability of getting 1, 2, or 3 = $3/6 = 1/2$

$q = 1/2$

Now, Binomial distribution of 'x' successes out of 'n' trails is given by

$P(x) = {}^nC_x \, p^x q^{n-x} = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} = {}^5C_x \left(\frac{1}{2}\right)^5$

Since, five dice were thrown 96 times. expected frequency distribution is given by

$F(x) = 96 \, {}^5C_x \left(\frac{1}{2}\right)^5$

$F(0) = 96 \times {}^5C_0 \left(\frac{1}{2}\right)^5 = 3$

$F(1) = 96 \times {}^5C_1 \left(\frac{1}{2}\right)^5 = 15$

$F(2) = 96 \times {}^5C_2 \left(\frac{1}{2}\right)^5 = 30$

$P(3) = 96 \times {}^5C_3 \left(\frac{1}{2}\right)^5 = 30$

$P(4) = 96 \times {}^5C_4 \left(\frac{1}{2}\right)^5 = 15$

$P(5) = 96 \times {}^5C_5 \left(\frac{1}{2}\right)^5 = 3$

No of dice showing 1, 2/3	0	1	2	3	4	5
O_i	3	8	24	35	19	7
E_i	3	15	30	30	15	3
$(O_i - E_i)^2$	0	49	36	25	16	16

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{0}{3} + \frac{49}{15} + \frac{36}{30} + \frac{25}{30} + \frac{16}{15} + \frac{16}{3}$

$= 11.67 > 11.07$

$\Rightarrow$ Hypothesis that follow binomial distribution is rejected.

4) A dice thrown 264 times and the number appearing on the face (x) follows the following frequency (f) distribution

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

Calculate the value of χ^2.

Soln

The frequencies in the given data are the observed frequencies, assuming that dice is unbiased, the expected number of frequencies for the numbers 1, 2, 3, 4, 5, 6 to appear on the face is $\frac{264}{6} = 44$ each. Now the data is as follows.

x	1	2	3	4	5	6
O_i	40	32	28	58	54	60
E_i	44	44	44	44	44	44
$(E_i - O_i)^2$	16	144	256	196	100	256

$$\chi^2 = \sum_{i=1}^{6} \frac{(E_i - O_i)^2}{E_i} = \frac{16}{44} + \frac{144}{44} + \frac{256}{44} + \frac{196}{44} + \frac{100}{44} + \frac{256}{44}$$

$$\boxed{\chi^2 = 22}$$

5). Four coins are tossed 100 times and the following results were obtained.

No of Heads	0	1	2	3	4
Frequency	5	29	36	25	5

fit a binomial distribution for the data and test the

Sol^n Given, $n=4$, p = probability of getting head = $\frac{1}{2}$

$q = \frac{1}{2}$

W.K.T. binomial distribution of 'x' success out of 'n' trails is given by

$$P(x) = {}^nC_x\, p^x q^{n-x} = {}^4C_x \left(\frac{1}{2}\right)^4 = \frac{{}^4C_x}{16}$$

Since, tossing four coins repeated 100 times Expected frequency distribution is given by

$$F(x) = \frac{100}{16}\, {}^4C_x$$

$F(0) = 6.25$, $F(1) = 25$, $F(2) = 37.5$

$F(3) = 25$, $F(4) = 6.25$.

O_i	5	29	36	25	5
E_i	6.25	25	37.5	25	6.25
$(O_i - E_i)^2$	1.5625	16	2.25	0	1.5625

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + \frac{0}{25} + \frac{1.5625}{6.25}$$

$$= 1.2 < 9.49$$

$\Rightarrow$ Binomial distribution fits for given data

6) A ~~fare~~ set of five similar coins is tossed 320 times and the result is $t_{0.05} = 7.$

No. of heads	0	1	2	3	4	5
Frequency	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution at 5% significance level.

7) Fit a binomial distribution for the data

No of Heads	0	1	2	3	4
Frequency	122	60	15	2	1

and also test the goodness of fit given that $\chi^2_{0.05} = 7.815$ for 3 d.o.f.

8). Fit a Poission distribution to the following data and test the goodness of fit at 5% level of significance

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

Given $\chi^2 = 3.82$ with degree of freedom 4.

Soln

$$\text{mean} = m = \frac{\Sigma fx}{x} = \frac{0(419)+1(352)+2(154)+3(56)+4(19)}{1000}$$

$$\boxed{m = 0.904}$$

from poission distribution we have

$$P(x) = \frac{e^{-m} m^x}{x!} = \frac{0.405\,(0.904)^x}{x!}$$

Since, experiment is repeted 1000 times Theoretical frequency is given by

$$F(x) = 1000\left(\frac{0.405\,(0.904)^x}{x!}\right) = \frac{405\,(0.904)^x}{x!}$$

$F(0) = 405$

$F(1) = 366.12$

$F(2) = 165.5$

$F(3) = 49.86$

$F(4) = 11.26$

O_i	419	352	154	56	19
E_i	405	366.12	165.5	49.86	11.28
$(E_i-O_i)^2$	196	199.3	132.25	37.69	59.9

Now, $\chi^2 = \sum_{i=0}^{4} \frac{(O_i - E_i)^2}{E_i}$

$$= \frac{196}{405} + \frac{199.3}{366.12} + \frac{132.25}{165.5} + \frac{37.69}{49.86} + \frac{59.9}{11.28}$$

$$= 7.89 > 3.82$$

$\Rightarrow$ the data does not fit good for poission distribution.

F - test distribution

$\Rightarrow$ F test is a small sample test

$\Rightarrow$ It is used to test $H_0: \sigma_1^2 = \sigma_2^2$

$\Rightarrow$ It is also called as variance Ratio test

$\Rightarrow$ $F_{cal} = \frac{S_1^2}{S_2^2}$ or $\frac{S_2^2}{S_1^2}$ [Here always bigger value in numerator]

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} \qquad S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1}$$

Where :- S_1 and S_2 are standard deviation

n_1 and n_2 are sample size

$\bar{x}_1$ and $\bar{x}_2$ are sample mean

$\Sigma(x_1-\bar{x}_1)^2$ and $\Sigma(x_2-\bar{x}_2)^2$ are sum of squares of deviation from mean

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} \qquad \bar{x}_2 = \frac{\Sigma x_2}{n_2}$$

$\Rightarrow$ Degree of freedom n_1-1, n_2-1

$\Rightarrow$ 5 % level of significance

$F_{0.05} = 3.73$

$\Rightarrow$ 1 % level of significance

$F_{0.01} = 6.84$

Note

1) If $F_{cal} < F_{tab}$ then we accept the hypothesis

2) If $F_{tab} > F_{cal}$ then we reject the hypothesis

$F_{cal} \longrightarrow$ calculated values

$F_{tab} \longrightarrow$ tabulated values.

1) Is there a difference between the variances of the number of weeks on the best seller lists for nonfiction and fiction books? fifteen new york Times bestselling fiction books had a standard deviation of 6.17 weeks on the list. Sixteen new york Times best selling nonfiction books had a standard deviation of 13.12 weeks at the 5% significance level, can we conclude there is a difference in the variances? ($F_{0.05,15,14} = 2.415$) (1)

Soln :-

Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$ (two tail test)

given $n_1 = 15$ $n_2 = 16$

$S_1 = 6.17$ $S_2 = 13.12$

We have

$$F_C = \frac{S_2^2}{S_1^2}$$ (bigger value will be in numerator)

$$= \frac{(13.12)^2}{(6.17)^2}$$

$$= 4.5217$$

$\therefore F_C = 4.5217 > F_{0.05,15,14}$

$\therefore$ we reject the null hypothesis

$\Longrightarrow$ There is a difference in the variance

2) The following table gives the number of units produce per day by 2 workers A and B for some days. Can we say that worker A is more stable than workers B? use 5% level of significance ($F_{0.05,7,5} = 4.88$)

Worker A	40	30	38	41	38	35		
Worker B	39	38	41	33	32	49	49	34

Soln :- Let

Ho : The stability of both workers is same is $\sigma_1^2 = \sigma_2^2$

H1: Worker A is more stable than worker B is $\sigma_1^2 > \sigma_1^2$

given $n_1 = 6$ $n_2 = 8$

Worker A		Worker B	
x_1	$(x_1-\bar{x})^2$	x_2	$(x_2-\bar{x}_2)^2$
40	9	39	0.1406
30	49	38	1.8906
38	1	41	2.6406
41	16	33	40.6406
38	1	32	54.3906
35	4	49	92.6406
		49	92.6406
222	80	34	28.8906
		315	313.8748

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{222}{6} = 37 \; ; \; \bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{315}{8} = 39.375$$

$$S_1^2 = \frac{\Sigma (x_1 - \bar{x})^2}{n_1 - 1} = \frac{80}{5} = 16$$

$$S_2^2 = \frac{\Sigma (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{313.8748}{7} = 44.8393$$

$$\therefore F_C = \frac{S_2^2}{S_1^2} = \frac{44.8393}{16} = 2.8025$$

$DF = n_2 - 1, n_1 - 1 = 7, 5$

$F_C = 2.8025 < F_{0.05,7,5} = 4.88$

$\therefore$ we accept the null hypothesis.

Thus the stability of both workers is same

3) The following data shows the runs scored by 2 batsman can it be said that the performance of batsman A is more consistent than the performance of batsman B? use 1% level of significance ($F_{0.01,4,7} = 7.85$)

Batsman A	40	50	35	25	60	70	65	55
Batsman B	60	70	40	30	50			

Soln:-

Let $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 > \sigma_2^2$

given $n_1 = 8$ $\quad n_2 = 5$

Batsman A		Batsman B	
x_1	$(x_1-\bar{x}_1)^2$	x_2	$(x_2-\bar{x}_2)^2$
40	100	60	100
50	0	70	400
35	225	40	100
25	625	30	400
60	100	50	0
70	400	250	1000
65	225		
55	25		
400	1700		

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{400}{8} = 50$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{250}{5} = 50$$

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = \frac{1700}{7} = 242.8571$$

$$S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = \frac{1000}{4} = 250$$

$$F_C = \frac{S_2^2}{S_1^2} = \frac{250}{242.8571} = 1.0294$$

$$F_C = 1.0294 < F_{0.01,4,7} = 7.85$$

$\therefore$ We accept the null hypothesis

Thus the performance of batsman is Consistent.

4) Two samples of sizes 9 and 8 give the sum of squares of deviations from their respective means equal to 160 inches² and 91 inches². Can these be regarded as drawn from the same normal population?

Solution :- We have $\Sigma(x-\bar{x})^2 = 160$ and

$$\Sigma(y-\bar{y})^2 = 91$$

$$S_1^2 = \frac{160}{8} = 20$$

$$S_2^2 = \frac{91}{7} = 13$$

Hence $F = \frac{S_1^2}{S_2^2} = \frac{20}{13} = 1.54$

for $n_2 - 1$, $n_2 - 1$ (Here $n_1 = 9$, $n_2 = 8$)

For Degree of freedom $(n_1 - 1, n_2 - 1) = (8, 7)$

We have $F_{0.05} = 3.73$

Since the calculated value of $F_{tab} < F_{0.05}$ the population variances are not significantly different.

5) Measurements on the length of a Copper wire were taken in 2 Experiments A and B as under: [Test whether B's measurements are more accurate than A's. (The reading taken in both Cases being unbiased)]

A's measurements : (mm)	12.29	12.25	11.86	12.13	12.44	12.78	12.77	11.90	12.47
B's measurements : (mm)	12.39	12.46	12.34	12.22	11.98	12.46	12.23	12.06	

<u>Solution</u> :-

Let $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 < \sigma_2^2$

given $n_1 = 9$; $n_2 = 8$

A's measurements		B's measurements	
x_1	$(x_1 - \bar{x}_1)^2$	x_2	$(x_2 - \bar{x}_2)^2$
12.29	0.0009	12.39	0.0169
12.25	0.0049	12.46	0.04
11.86	0.2116	12.34	0.0064
12.13	0.0361	12.22	0.0016
12.44	0.0144	11.98	0.0784
12.78	0.2116	12.46	0.04
12.77	0.2025	12.23	0.0009
11.90	0.1764	12.06	0.04
12.47	0.0225		
$\Sigma x_1 = 110.89$	$\Sigma (x_1 - \bar{x}_1)^2 = 0.8809$	$\Sigma x_2 = 98.14$	$\Sigma (x_2 - \bar{x}_2)^2 = 0.2242$

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{110.89}{9} = 12.32$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{98.14}{8} = 12.26$$

$$S_1^2 = \frac{\sum (x_1-\bar{x}_1)^2}{n_1-1} = \frac{0.8809}{9-1} = 0.1101$$

$$S_2^2 = \frac{\sum (x_2-\bar{x}_2)^2}{n_2-1} = \frac{0.2242}{7} = 0.03202$$

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

$$= \frac{0.1101}{0.03202}$$

$$\boxed{F_c = 3.4406}$$

W.K.T $F_{0.05} = 3.73$ and $F_{0.01} = 6.84$

$\therefore$ Since the Calculated value of F is less than both $F_{0.05}$ and $F_{0.01}$, the result is insignificant at both 5% and 1% level.

Hence there is no reason to say that B's measurements are more accurate than those of A's