Module-4

Statistical Inference - II

Statistics:
Any function of the sample values is known as statistics.

Sampling distribution: It is a distribution of a statistic overall possible samples. ie sample distribution is the probability distribution of statistics.

Sampling Variables: it is a process used to predict the value of a specific variable within a population.
ex: A limited sample size can be used to compute the average accounts receivable balance, as well as a statistical derivation of '+' & '-' range of total receivable value.

Central limit theorem

A sample of size 'n' is selected from a population that has mean μ & S.D σ, then let $x_1, x_2, \ldots x_n$ be the 'n' observation, they are independent and identically distribution with mean $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$

ie $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. where μ, σ are mean and S.D of population from where the sample was select and sample size become large $(n \geq 30)$

Degree of freedom :

Refers to maximum number of logically independent values, which may vary in a data sample. They are calculated by $(n-1)$.

Description	population	Sample
Size	N	n
Mean	μ	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Variance	σ^2	$s^2 = \frac{1}{n-1} \sum (x-\bar{x})^2$
Standard Deviation.	σ	$s = \sqrt{\frac{1}{n-1} \sum (x-\bar{x})^2}$

Confidence Interval :

Estimate an actual population mean μ obtain $\bar{x}$ (mean of a sample randomly selected from population..

ie lower value < population mean μ < upper value

then $C.I = \text{Mean} \pm Z\,(S.D / \sqrt{\text{Sample size}})$

$$= \mu \pm Z \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} = \mu \pm Z \frac{s}{\sqrt{n}}.$$

Confidence level.	99%	98%	95%	90%	50%
Z	2.58	2.33	1.96	1.645	0.6748

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 size 100 from an infinite population with mean $\mu = 53$ and variance $\sigma^2 = 400$.

Soln:- Central limit theorem: states that "The sample mean $\bar{x}$ follows approximately the normal distribution mean μ & S.D $\frac{\sigma}{\sqrt{n}}$ (S.E) ie $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$ where μ, σ are mean and standard deviation of the population from sample.

given: Sample size $n = 100$.

Mean of population $\mu = 53$

Variance $\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$.

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

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

$\therefore$ W.K.T $\quad Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{X} - 53}{20/\sqrt{100}} = \dfrac{\bar{X} - 53}{2}$

at $\bar{X} = 50 \Rightarrow Z = \dfrac{50 - 53}{2} = -1.5 = Z_1$

at $\bar{X} = 56 \Rightarrow Z = \dfrac{56 - 53}{2} = 1.5 = Z_2$

18   8

$\therefore P(50 < \bar{x} < 56) = P(-1.5 < z < 1.5)$

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

$= 2A(1.5)$

$= 2 \times 0.4332$

$\boxed{P(50 < \bar{x} < 56) = 0.8664}$

2 An unknown distribution has a mean of 90 and a S.D of 15. Samples of size $n = 25$ are drawn randomly from the population. find the probability that the sample mean is b/w 85 and 92.

Sol: given

sample size $n = 25$

mean $= \mu = 90$

$S.D = \sigma^2 = 225$

Variance $= \sigma = 15$

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

$= \bar{X} \sim N\left(90, \frac{\cancel{9}15}{\sqrt{25}}\right) = \bar{X} \sim N\left(90, \frac{15}{5}\right) = \bar{X} \sim N(90, 3)$

$\therefore$ W.K.T $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 90}{\frac{15}{5}} = \frac{\bar{X} - 90}{3}$

at $\bar{X} = 85$

$Z = \frac{85 - 90}{3} = \frac{-5}{3} = -1.66$

at $\bar{X} = 92$

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

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

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

$= 0.4515 + 0.2454$

$= 0.6965$

$\boxed{P(85 < \bar{x} < 92) = 0.6965}$

(5) A random sample of size 64 is taken from an infinite population having mean 112 and variance 144. Using C.L.T find the probability of getting sample mean $\bar{x}$ greater than 114.5.

Soln:- given

Sample size $= n = 64$

mean $\mu = 112$

Variance $= \sigma^2 = 144 \Rightarrow \sigma = 12$

$\therefore \bar{x} \sim \left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \bar{x} \sim \left(112, \frac{12}{\sqrt{64}}\right)$

$\bar{x} \sim \left(112, \frac{12}{8}\right) = \bar{x} \sim (112, 1.5)$

$\Rightarrow z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 112}{1.5} = \frac{114.5 - 112}{1.5} = 1.66 \quad \bar{x} = 114.5$

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

$= 0.5 - 0.4515$

$\boxed{P(z > 1.66) = 0.0489}$

that the stimulus administration. (15)

(8)

4) Let $\bar{x}$ denotes 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:- Sample size $n = 100$.

Chisquare distribution $X \sim \chi^2(50)$ d.f = 50

mean of chisquare is $\mu = 50$

$\therefore \sigma^2 = 2 \times d.f = 2 \times 50 = 100$, $\sigma = 10$

Sample mean of chisquare distribution follows normal distribution with mean μ & S. error $\frac{\sigma}{\sqrt{n}}$

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

$$\tilde{\bar{X}} \sim N\left(50, \frac{10}{\sqrt{100}}\right) \sim N\left(50, \frac{10}{10}\right) = N(50, 1)$$

$$\therefore Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 50}{1} = \bar{x} - 50$$

$\therefore$ @ $\bar{x} = 49 \Rightarrow z = \frac{49 - 50}{1} = \frac{-1}{1} = -1$

@ $\bar{x} = 51 \Rightarrow z = \frac{51 - 50}{1} = \frac{1}{1} = 1$

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

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

$= 2A(1)$

$= 2 \times 0.3416 = 0.6826$

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

5) An Electric 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 less than 775 hours.

Sol:

Total no of bulbs $n = 16$

An average life of a bulb $\mu = 800$

S.D of bulbs $\Rightarrow \sigma = 40$

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \sim N\left(800, \frac{40}{\sqrt{16}}\right) \sim N\left(800, \frac{40}{4}\right)$$

$$\bar{x} \sim N(800, 10)$$

WKT $$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 800}{\frac{40}{4}} = \frac{\bar{x} - 800}{10}$$

$\therefore$ at $\bar{x} = 775$, $$z = \frac{775 - 800}{10} = \frac{-25}{10} = -\frac{5}{2} = -2.5$$

$$\therefore P(\bar{x} < 775) = P(z < -2.5)$$

$$= P(z > 2.5)$$

$$= 0.5 - P(0 < z < 2.5)$$

$$= 0.5 - A(2.5)$$

$$= 0.5 - 0.4938$$

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

(6) The heights of a random sample of 50 college students showed a mean of 174.5 cm and a S.D of 6.9 cm. Construct a 99% confidence interval for the mean height of all college students.

Soln: Sample size $n = 50$

Average height of students (mean) $= \mu = 174.5$ cm.

S.D of the students $\sigma = 6.9$ cm

confidence interval $z = 2.5176$

W.K.T

C.I = Mean $\pm$ z(S.D / $\sqrt{\text{Sample size}}$)

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

C.I $= 174.5 \pm (2.576)\left(\frac{6.9}{\sqrt{50}}\right)$

C.I $= 174.5 \pm 2.5136$

$\therefore$ C.I at Lower end is C.I $= 174.5 - 2.5136 = 171.9864$

C.I at upper end is C.I $= 174.5 + 2.5136 = 177.0136$

$\therefore$ The C.I with 99% the mean height of all college students is b/w 171.9864 to 177.0136 cm.

(7) The mean and S.D of the diameter of all rivet heads manufactured by a company are 7.2642 mm and 0.0058 mm respectively find

a) 99% b) 98% c) 95% d) 90% e) 50%

Confindence limit for the mean diameter of all the

the rivet heads manufactured by the Company.

① Soln. given sample size. $n = 250$

Mean of diameter $\mu = 7.2642$ mm.

S.D of diameter $\sigma = 0.0058$ mm.

W.K.T Confidence level at

1) $99\% = 2.58$ 2) $98\% = 2.33$ 3) $95\% = 1.96$

4) $90\% = 1.645$ 5) $50\% = 0.6745$

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

1) @ 99% $CI = 7.2642 \pm \left(\frac{2.58 \times 0.0058}{\sqrt{250}}\right) = (7.26326, 7.26514)$

2) @ 98% $C.I = 7.2642 \pm \left(\frac{2.33 \times 0.0058}{\sqrt{250}}\right) = (7.26334, 7.26504)$

3) @ 95% $C.I = 7.2642 \pm \left(\frac{1.96 \times 0.0058}{\sqrt{250}}\right) = (7.26347, 7.26493)$

4) @ 90% $C.I = 7.2642 \pm \left(\frac{1.645 \times 0.0058}{\sqrt{250}}\right) = (7.26359, 7.26481)$

5) @ 50% $C.I = 7.2642 \pm \left(\frac{0.6745 \times 0.0058}{\sqrt{250}}\right) = (7.26395, 7.26445)$

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

Soln: given Sample size $n = 25$

Mean of a Sample $\bar{x} = 78.3$

S.D $\sigma = 2$

W.K.T C.I of 99%. Z is 2.58

$\therefore$ C.I = μ = Mean $\pm$ Z (S.D / $\sqrt{\text{Sample size}}$)

$\mu = 78.3 \pm 2.56 (2/\sqrt{25})$

C.I of $\mu = 78.3 \pm 2.56 (2/5)$

$\boxed{\text{C.I of } \mu = 77.268, 79.332}$

(9) Let the observed value of mean $\bar{x}$ of a random sample of size 20 from a normal distribution with mean μ & variance $\sigma^2 = 80$ be 81.2. find 90% & 95% confidence interval for μ.

Soln:- given sample size $n = 20$.

mean of Sample $\mu = 81.2$

Variance of $\sigma^2 = 80$

$S.D = \sigma = \sqrt{80} = 8.9442.$

W.K.T C.I of 95%, 90% are $z_{95\%} = 1.96$, $z_{90\%} = 1.645$

by N.D.T for 95%. C.I of μ = mean $\pm$ z $\left(\frac{S.D}{\sqrt{n}}\right)$

C.I of $\mu = 81.2 \pm 1.96 \times \frac{8.9442}{\sqrt{20}}$

Sampling Distribution

Students t distribution

Let μ be the mean of population, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the mean and $S = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ be the standard deviation of a sample, then Students t distribution is defined as

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

Alternative formulae for t-test distribution of 2 samples is

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

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

or

$$S = \sqrt{\frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2\right]}$$

$\therefore$ C.I of μ @ 95% is

$$\boxed{C.I = (77.28, 85.12)}$$

For 90%.

$$C.I \text{ of } \mu = 81.2 \pm \left(\frac{1.645 \times 8.9442}{\sqrt{20}}\right)$$

$$\boxed{C.I \text{ of } \mu = 77.91, 84.89}$$

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

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

$\therefore$ sample size $n = 4$

Mean $\bar{x} = 14.25$

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

WKT C.I of 95%. $Z = 1.96$.

$\therefore$ C.I $= \mu =$ Mean $\pm Z (S.D/\sqrt{n})$

$$\mu = 14.25 \pm 1.96 (2.5/\sqrt{4})$$

$$\boxed{C.I \text{ of } \mu = 11.80, 16.70}$$

Chi-Square distribution

Let O_i $(i = 0, 1, 2, 3 \ldots n)$ and E_i $(i = 1, 2, 3, \ldots n)$ be the set of observed and Expected values (frequencies) respectively. then Chi-Square distribution is defined as.

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \ldots + \frac{(O_n - E_n)^2}{E_n}$$

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

F-distribution

It is useful in hypothesis testing. Hypothesis testing is used ~~by~~ in comparing data from 2 or more populations. It is used to find whether the F-value for a study. indicating that statistically significant difference b/w 2 populations.

F ratio is $F = \frac{\sigma_1^2}{\sigma_2^2}$ where $\sigma^2 = \frac{1}{n} \sum (x - \mu)^2$

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

σ_1 = S.D of population 1

σ_2 = S.D of population 2

S_1 = S.D of sample ~~mean~~ 1

S_2 = S.D of sample ~~mean~~ 2

Note:- for 2 Samples drawn from normal distribution

i) $F_{ratio} = \frac{S_1^2}{S_2^2}$ where $S_1^2 = \frac{1}{n_1 - 1} \Sigma (x - \bar{x})^2$

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

ii) $F = \frac{\text{larger variance}}{\text{Smaller variance}}$

n_1 = d.f for sample having larger variance

n_2 = d.f for Sample having Smaller variance.

iii) Expected value of F:-

$F_E = \frac{S_2^2}{S_1^2}$, $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ d.f.

Problems

1) A certain stimulus administered to each of the 12 patients resulted in the following changes in the blood pressure 5, 2, 8, −1, 3, 0, 6, −2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase the b.p? (Note: $t_{0.05}$ for 11 d.f = 2.2) blood pressure.

Soln:- given the change in

x: 5, 2, 8, −1, 3, 0, 6, −2, 1, 5, 0, 4.

$\therefore \bar{x} = \frac{1}{n} \Sigma x = \frac{31}{12} = 2.5833$

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

$s^2 = \frac{1}{11} \{ (5 - 2.582)^2 + (2 - 2.582)^2 + (8 - 2.582)^2 + (-1 - 2.582)^2 + (3 - 2.58$

$+ (0 - 2.582)^2 + (6 - 2.582)^2 + (-2 - 2.582)^2 + (1 - 2.582)^2 + (5 - 2.582)$

$+ (0 - 2.582)^2 + (4 - 2.582)^2 \}$

$s^2 = 9.538$

$\boxed{s = 3.088}$

Let us suppo... with ...
is not accompanied with ...

$\therefore \mu = 0$

$$t = \frac{\bar{x} - \bar{u}}{S/\sqrt{n}} = \frac{2.5833 - 0}{\left(\frac{3.088}{\sqrt{12}}\right)}$$

$t = 2.8979 \simeq 2.97$ 2.201

Hence the hypothesis is rejected at 5% level of Sign...
we conclude with 95% confidence that the stimulus
general is accompanied with increase of blood pres...

12) A random sample of 10 boys had the followin...
IQ: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does this data s...
the assumption of a population mean IQ of 100 at 5% (...
of significance? (Note: $t_{0.05} = 2.262$ for 9 d.f.).

Soln: Given IQ of 10 boys.

ie x: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

$$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{972}{10} = 97.2.$$

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

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

$$S^2 = 203.7333$$

$$S = 14.2735$$

given mean population $\mu = 100$, $n = 10$

we have $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{97.2 - 100}{\frac{14.2738}{\sqrt{10}}} = \underline{-0.6203 < 2.262}$

(13) Ten 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 universe is 66 inches ($t_{0.05} = 2.262$ for 9 d.f.)

Soln: given heights of the population in inches.

$x: 63, 63, 66, 67, 68, 69, 70, 70, 71, 71.$

$n = 10 \quad \Sigma x = 678.$

$\therefore \quad \bar{x} = \frac{1}{n}\Sigma x = \frac{1}{10} 678 = 67.8$

$\bar{x} = 67.8$

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

$s^2 = \frac{1}{9}\{[(63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2$

$+ (69-67.8)^2 + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2]\}$

$s^2 = 9.069, \quad s = 3.011$

given mean population $\mu = 66$

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{67.8 - 66}{\frac{3.011}{\sqrt{10}}} = 1.8979 \simeq 1.9 < 2.262.$

$\therefore$ The hypothesis is accepted at 5% level of significance.

(14) The nine items of a sample have the following value $x: 45, 47, 50, 52, 48, 47, 49, 53, 51.$ Does the mean of these differ significantly from the assumed mean of 47.5 at 5% level of significance.

Soln:- Given sample values: 45, 47, 50, 52, 48, 47, 49, 53, 51,

$\therefore n = 9$; = sample size

population mean $\mu = 47.5$, $\Sigma x = 442$

$\therefore$ Sample mean. $\bar{x} = \frac{1}{n} \Sigma x = \frac{442}{9} = 49.11$

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

$s^2 = \frac{1}{8} \{ (45 - 49.11)^2 + (47 - 49.11)^2 + (50 - 49.11)^2 + (52 - 49.11)^2$

$+ (48 - 49.11)^2 + (47 - 49.11)^2 + (49 - 49.11)^2 + (53 - 49.11)^2 + (51 - 49.11)^2 \}$

$s^2 = \frac{54.9}{8} = 6.8625 \Rightarrow s = \sqrt{6.8625} = 2.6196$

$\therefore$ Null hypothesis $H_0: \mu = 47.5$

$\therefore t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{49.11 - 47.5}{\frac{2.6196}{\sqrt{9}}} = 1.8437$

$\therefore \boxed{t = 1.8437}$

$\therefore$ at 5% level of significance for $V = 9 - 1 = 8$ d.f is 2.3060.

Since $t = 1.8432 < 2.3060$

hence Null hypothesis is accepted.

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

Battery A :- $n_1 = 10$, $x_1 = 500$ hrs , $\sigma_1^2 = 100$

B :- $n_2 = 10$ $x_2 = 560$ hrs , $\sigma_2^2 = 121$

Compute ~~data~~ Students t and test whether there is a significance differences in the 2 means. (18)

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

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

$\bar{x}_1 = 500$ hrs

$\bar{x}_2 = 560$ hrs

$$S^2 = 122.78$$

$$\boxed{S = 11.0805}$$

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

$$t = \frac{560-500}{11.0805\sqrt{\frac{1}{10}+\frac{1}{10}}}$$

$$t = 12.108 \simeq 12.11$$

$\therefore$ The value of t is greater than the table value of t. for 18 d.f ~~at~~ all level of significance.

(16) Two horses A & B 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 determinate b/w 2 horses

Soln:- Let the variable x & y respectively corresponds A & B.

x: 28, 30, 32, 33, 33, 29, 34

y: 29, 30, 30, 24, 27, 29, -.

$\therefore \bar{x} = \frac{1}{n_1} \sum x_i = \frac{219}{7}$ $\therefore \bar{y} = \frac{1}{n_2} \sum y_i = \frac{169}{6} = 28.20$

$= 31.30$

$\sum (x-\bar{x})^2 = (28-31.30)^2 + (30-31.3)^2 + (32-31.3)^2 + (33-31.3)^2$
$+ (29-31.3)^2 + (34-31.3)^2 = 31.4$

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

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

$S^2 = 5.2973$

$\boxed{S = 2.3016}$

We have

$t = \frac{|\bar{x}_2 - \bar{x}_1|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{|31.3 - 28.2|}{2.3016\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.42 \begin{cases} > t_{0.05} \\ < t_{0.01} \end{cases}$

17) 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 goodness of fit $\chi^2_{0.05} = 9.49$ for 4 d.f.

Soln. Given 4 coins are tossed 100 times.

The probability of getting head is $p = 0.5$, $q = 0.5$.

The probability mass function of a binomial distribution is

$$P(X=x) = 4C_x (0.5)^x (0.5)^{4-x}$$

$$P(x=0) = 4C_0 (0.5)^0 (0.5)^{4-0} = (0.5)^4 = 0.0625$$

$$P(x=1) = 4C_1 (0.5)^1 (0.5)^3 = 4(0.5)^4 = 0.25$$

$$P(x=2) = 4C_2 (0.5)^2 (0.5)^2 = 6(0.5)^4 = 0.375$$

$$P(x=3) = 4C_3 (0.5)^3 (0.5)^1 = 4(0.5)^4 = 0.25$$

$$P(x=4) = 4C_4 (0.5)^4 (0.5)^0 = (0.5)^4 = 0.0625$$

$\therefore$ $E_0 = 100 \times 0.0625 = 6.25$

$E_1 = 100 \times 0.25 = 25$

$E_2 = 100 \times 0.375 = 37.5$

$E_3 = 100 \times 0.25 = 25$

$E_4 = 100 \times 0.0625 = 6.25$

where 100 is sum of frequency.

O_i	5	29	36	25	5
E_i	6.25	25	37.5	25	6.25

$$\chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] = 0.25 + 0.64 + 0.06 + 0.25$$

$$\chi^2 = 1.2 < 9.49$$

Hence fitness is good.

(18) A die thrown 264 times and the number appearing on face (x) follows the following frequency (f) distribution. Calculate the value of χ^2 (25)

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

Soln The given frequency is Observed frequency. assuming that dice is unbiased the expected frequency for numbers 1, 2, 3, 4, 5, 6 to appear on face is $\frac{264}{6} = 44$ each.

Now the data is

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

$$\therefore \chi^2 = \sum_i \left(\frac{(O_i - E_i)^2}{E_i}\right)$$

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

$$\chi^2 = \frac{1}{44}[16 + 144 + 256 + 196 + 100 + 264]$$

$$\boxed{\chi^2 = \frac{968}{4} = 22}$$

(22)

(19) A survey of 320 families with 5 children each revealed the following distribution.

No of boys	5	4	3	2	1	0
No of girls	0	1	2	3	4	5
No of families	14	56	110	88	40	12

Is the result consistent with the hypothesis that male and female births are equally probable at 5% level of significance?

Soln:- Given

No of families selected for survey = 320.

The probability of female and male birth is equal.

$p = \frac{1}{2} = 0.5$, $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2} = 0.5$.

No of childrens in the selected families $n = 5$.

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No of families	14	56	110	88	40	12

The statistical hypothesis is

Ho: The probability of female and male birth is equal

H1: The probability of female & male birth is not equal.

∴ By Binomial distribution
we have $p(x) = n_{C_x}\, p^x q^{n-x}$

$$p(x) = 5_{C_x}\,(0.5)^x\,(0.5)^{5-x}$$

$$p(x) = 5_{C_x}\,(0.5)^5$$

The expected frequency can be calculated for 320 families as.

$E(x) = 320 \times P(x) = 320 \times 5_{C_x}\,(0.5)^5$

∴ $E(0) = 320 \times p(0) = 10$

$E(1) = 320 \times p(1) = 50$

$E(2) = 320 \times p(2) = 100$

$E(3) = 320 \times P(3) = 100$

$E(4) = 320 \times P(4) = 50$

$E(5) = 320 \times p(5) = 10.$

∴

No of boys	No of girls	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
5	0	14	10	4	16	1.6
4	1	56	50	6	36	0.72
3	2	110	100	10	100	1
2	3	88	100	-12	144	1.44
1	4	40	50	-10	100	2
0	5	12	10	2	4	0.4

We have χ^2 for 5 degrees of freedom @ 5% is 11.07

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] = 7.16 < 11.07$$

Since χ^2 is < given value of χ^2

$\therefore$ H_0 is accepted

$\therefore$ mean of both male & female are equal.

<u>HW</u>

20: The Theory predicts the proportion of beans in the four groups A, B, C & D should be 9:3:3:1. In an experiment among 1600 beans, the numbers in 4 groups were 882, 313, 287, and 118. The chi square value is approximately equal to.

Ans:- $\chi^2 = 4.72$.

21) Two random samples drawn from 2 normal popu... are

Sample 1	20	16	26	27	22	23	18	24	19	25	–	–
Sample 2	27	33	42	35	32	34	38	28	41	43	30	37

Obtain the estimate of the Variance of the popula... and test 5% level of significance whether 2 po... have the same variance.

Soln:

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

ie The two samples are drawn from 2 populations having the same variance.

Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$.

Given samples

Sample 1	20	16	26	27	22	23	18	24	19	25	-	-
Sample 2	27	33	42	35	32	34	38	28	41	43	30	37

$\bar{x}_1 = \sum_{i=1}^{n} x_i \Rightarrow \bar{x}_1 = \frac{220}{10} = 22$

$\bar{x}_2 = \sum_{i=1}^{n} x_i = \bar{x}_2 = \frac{420}{12} = 35$

x_1	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	x_2	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$
20	-2	4	27	-8	64
16	-6	36	33	-2	4
26	4	16	42	7	49
27	5	25	35	0	0
22	0	0	32	-3	9
23	1	1	34	-1	1
18	-4	16	38	3	9
24	2	4	28	-7	49
19	-3	9	41	6	36
25	3	9	43	8	64
Σ 220	0	120	30	-5	25
			37	2	4
			420	0	314

$\therefore F_0 = \frac{S_1^2}{S_2^2} = \frac{\cancel{13.33}}{\cancel{2}} = \frac{S_2^2}{S_1^2} = \frac{28.54}{13.33} = 2.14$

Since $S_2^2 > S_1^2$, $F_0 = 2.14$

$S_1^2 = \frac{1}{n_1 - 1} \sum (x_1 - \bar{x}_1)^2 = 13.33$

$S_2^2 = \frac{1}{n_2 - 1} \sum (x_2 - \bar{x}_2)^2 = 28.54$

Expected value:

$F_E = \frac{S_2^2}{S_1^2}$, F distribution with d.f below 5% level

$v_1 = n_1 - 1 = 10 - 1 = 9$ } is 3.10

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

Since $F_0 < F_E$, at 5% level of significance.

⇒ Accept the hypothesis ~~and conclude~~

⇒ Two samples may be regarded as drawn from the population having same variance.

22. The table shows the S.D and Sample S.D for both men and women. find the F statistics. Considering the men population in numerator.

Population	Population S.D	Sample S.D
Men	30	35
Women	50	45

Soln: given σ_1 = S.D of population - 1 = 30

σ_2 = S.D of population - 2 = 50.

S_1: S.D of sample 1 = 35

S_2: S.D of sample - 2 = 45

$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = 1.$