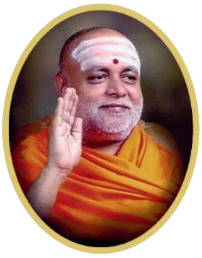
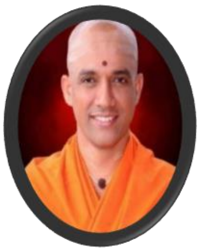




||Jai Sri Gurudev||

S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR

DEPARTMENT OF MATHEMATICS

MATHEMATICS-3 FOR COMPUTER SCIENCE (BCS301)

MODULE - 4

STATISTICAL INFERENCE -2

Prepared by:
Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in
Phone: 7338005244

Statistics: Any function of the sample values is known as a statistics.
Eg: Sample mean, Sample median, Sample variance etc. are all statistics.

Sampling Distribution: A sampling distribution is a distribution of a statistic over all possible samples. That is sampling distribution is the probability distribution of the statistics.

Sampling Variables: Variables sampling is the process used to predict the value of a specific variable within a population. For example, a limited sample size can be used to compute the average accounts receivable balance, as well as a statistical derivation of the plus or minus range of the total receivables value that is under review.

The Central Limit Theorem: Suppose that a sample of size n is selected from a population that has mean μ and the standard deviation σ, then Let $x_1, x_2, x_3, x_4 \ldots\ldots, x_n$ be the n observations, they are independent and identically distributed with mean $\bar{X} = \frac{x_1+x_2+x_3+\ldots\ldots x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$, 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 also called Standard error) , i.e. $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ , σ are mean and standard deviation of the population from where the sample was selected and the sample size becomes large ($n \geq 30$).

Degrees of freedom: Degrees of freedom refer to the maximum number of logically independent values, which may vary in a data sample. Degrees of freedom are calculated by subtracting one from the number of items within the data sample (n -1).

Description	Population notation	Sample Notation
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 Intervals:

Suppose we want to estimate an actual population mean μ. As you know, we can only obtain $\bar{x}$, the mean of a sample randomly selected from the population of interest. We can use $\bar{x}$ to find a range of values:

Lower value < population mean μ < Upper value

That we can be really confident contains the population mean μ. The range of values is called a "**confidence interval**."

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

Confidence interval $C.I. == Mean \pm Z(Standard\ Deviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$ or $\bar{X} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$

Confidence Level	99%	98%	95%	90%	50%
Z	2.58	2.33	1.96	1.645	0.6745

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 μ = 53 and variance σ^2 = 400.

Sol.

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), i.e., $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, 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 $\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$

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

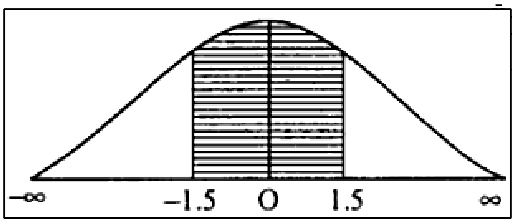
$$\Rightarrow \bar{X} \sim N\left(53, \frac{20}{\sqrt{100}}\right)$$

$$\Rightarrow \bar{X} \sim N(53,2)$$

$$\therefore \text{ we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

53

$\sqrt{100}$



$$\Rightarrow Z = \frac{\bar{X}-53}{2}$$

$$\therefore \text{ At } \bar{X}=50 \Rightarrow Z = \frac{50-53}{2} = -\frac{3}{2} = -1.5 = z_1$$

$$\text{At } \bar{X}=56 \Rightarrow Z = \frac{56-53}{2} = \frac{3}{2} = 1.5 = z_2$$

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

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

$$= 2A(1.5)$$

$= 2 \times 0.4332$

$\therefore P(50 < \bar{X} < 56) = 0.8664$

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.

Sol.

Given,

Sample size n=25

Mean of the population $\mu = 90$

Variance of the population $\Rightarrow \sigma = 15$

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

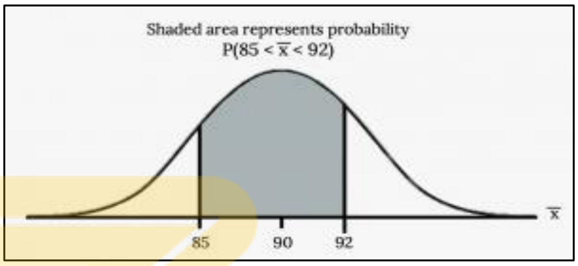
$$\Rightarrow \bar{X} \sim N\left(90, \frac{15}{\sqrt{25}}\right)$$

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

$$\therefore \text{ we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

$$\Rightarrow Z = \frac{\bar{X}-90}{15/\sqrt{25}}$$

$$\Rightarrow Z = \frac{\bar{X}-90}{3}$$



$\therefore$ At $\bar{X} = 85 \Rightarrow z = \frac{85-90}{3} = -\frac{5}{3} = -1.66$

$\therefore$ At $\bar{X} = 92 \Rightarrow z = \frac{92-90}{3} = \frac{2}{3} = 0.66$

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

$\Rightarrow P(-1.66 < z < 0.66) = P(0 < z < 1.66) + P(0 < z < 0.66)$

$= 0.4515 + 0.2454$

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

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.

Sol.

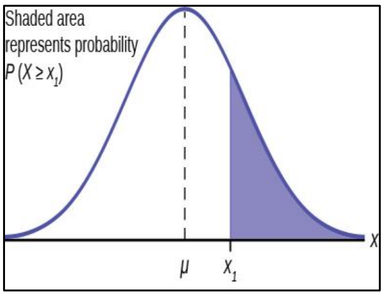
Given,

Sample size n=64

Mean of the population $\mu = 112$

Variance of the population $\Rightarrow \sigma^2 = 144 \Rightarrow \sigma = 12$

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



$$\Rightarrow \bar{X} \sim N\left(112, \frac{12}{\sqrt{64}}\right)$$

$$\Rightarrow \bar{X} \sim N(90,1.5)$$

$$\therefore \text{ we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

$$\Rightarrow Z = \frac{\bar{X}-112}{12/\sqrt{64}}$$

$$\Rightarrow Z = \frac{\bar{X}-112}{1.5}$$

$\therefore$ At $\bar{X} = 114.5 \Rightarrow z = \frac{114.5-112}{1.5} = 1.66$

$$\therefore P(\bar{X} > 114.5) = P(z > 1.66)$$
$$\Rightarrow P(z > 1.66) = 0.5 - P(0 < z < 1.66)$$
$$= 0.5 - 0.4515$$
$$\Rightarrow P(z > 1.66) = 0.0489$$

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).

Sol.

The sample size n is=100

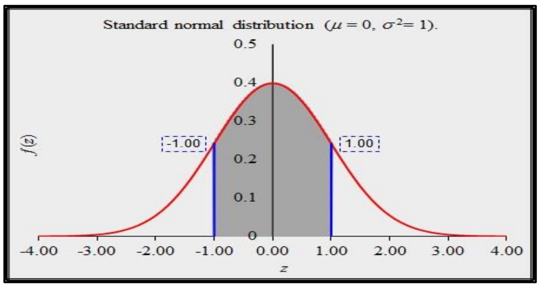
The chi-square distribution is given as ,X~ $\chi^2(50)$, where d.f.=50

The mean and variance of chi-square distribution is given as, μ = 50

Therefore $\Rightarrow \sigma^2 = 2 \times d.f. = 2 \times 50 = 100 \Rightarrow \sigma = 10$

The sample mean of chi-square distribution follows normal distribution with mean and standard error $\frac{\sigma}{\sqrt{n}}$.

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
$$\Rightarrow \bar{X} \sim N\left(50, \frac{10}{\sqrt{100}}\right)$$
$$\Rightarrow \bar{X} \sim N(50,1)$$
$$\therefore \text{we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$



$$\therefore \text{At } \bar{X}\text{=50} \Rightarrow Z = \frac{49-50}{1} = -\frac{1}{1} = -1 = z_1$$
$$\text{At } \bar{X}\text{=51} \Rightarrow Z = \frac{51-50}{1} = \frac{1}{1} = 1 = z_2$$
$$\therefore P(49 < \bar{X} < 51) = P(-1 < z < 1)$$
$$= 2P(0 < z < 1)$$
$$= A(1)$$
$$= 2 \times 0.3416$$
$$\therefore P(50 < \bar{X} < 56) = 0.6826$$

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 less than 775 hours.

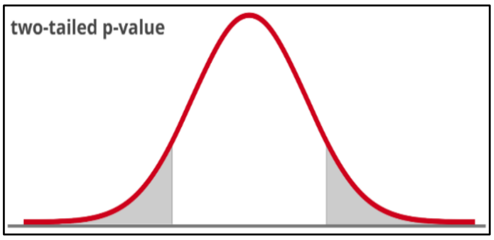
Sol.

Total number of bulbs n=16

An average life of bulbs μ = 800

Standard deviation of the bulbs $\Rightarrow \sigma = 40$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
$$\Rightarrow \bar{X} \sim N\left(800., \frac{40}{\sqrt{16}}\right)$$
$$\Rightarrow \bar{X} \sim N(800,10)$$
$$\therefore \text{We know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-800}{10}$$



$$\therefore \text{At } \bar{X}\text{=775} \Rightarrow Z = \frac{775-800}{10} = -\frac{25}{10} = -2.5.$$
$$\therefore P(\bar{X} < 775) = P(z < -2.5)$$
$$\Rightarrow P(z < -2.5) = P(z > 2.5)$$
$$\Rightarrow P(z < -2.5) = 0.5 - P(0 < z < 2.5)$$
$$\Rightarrow P(z < -2.5) = 0.5 - A(2.5)$$

$$\Rightarrow P(z < -2.5) = 0.5 - 0.4938$$
$$\Rightarrow P(z < -2.5) = 0.0062$$

6. The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Construct a 99% confidence interval for the mean height of all college students.

Sol.

Given the sample size n=50

Average height of Students (Mean) $\mu = 174.5 c.m.$

Standard deviation of the Students $\sigma = 6.9 c.m.$

We know that, Confidence level of 99%, the corresponding z value is 2.576. This is determined from the normal distribution table.

Confidence interval $C.I. == Mean \pm Z(S\tan d\, ardDeviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = 174.5 \pm \left(2.576 \times \frac{6.9}{\sqrt{50}}\right)$$
$$\Rightarrow C.I. = 174.5 \pm (2.576 \times 0.9758)$$
$$\Rightarrow C.I. = 174.5 \pm 2.5136$$

The lower end of the confidence interval is $= 174.5 - 2.5136 = 171.9864$

The upper end of the confidence interval is $= 174.5 + 2.5136 = 177.0136$

Therefore, with 99% confidence interval, the mean height of all college students is between 171.9864 centimeters and 177.0136 centimeters.

7. The mean and SD of the diameters of a sample of 250 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%

Confidence limits for the mean diameter of all the rivet heads manufactured by the company.

Sol.

Given the sample size n=250

Mean of a diameter $\mu = 7.2642 mm.$

Standard deviation of the diameter $\sigma = 0.0058 mm$

We know that,

Confidence Level	99%	98%	95%	90%	50%
Z	2.58	2.33	1.96	1.645	0.6745

Confidence interval $\boldsymbol{C.I. = Mean \pm Z(S\tan d\, ardDeviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}}$

Confidence Level	$C.I. = \mu \pm Z\frac{\sigma}{n}$	Final C.I.	Interval
99%	$7.2642 \pm \left(2.58 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00094	(7.26326 , 7.26514)

98%	$7.2642 \pm \left(2.33 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00086	(7.26334 , 7.26504)
95%	$7.2642 \pm \left(1.96 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00073	(7.26347 , 7.26493)
90%	$7.2642 \pm \left(1.645 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00061	(7.26359 , 7.26481)
50%	$7.2642 \pm \left(0.6745 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00025	(7.26395 , 7.26445)

8. 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 for μ.

Sol.

Given the sample size n=25

Mean of sample $\bar{X} = 78.3$

Standard 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 $C.I. = \mu = Mean \pm Z(Standard Deviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = \mu = 78.3 \pm \left(2.58 \times \frac{2}{\sqrt{25}}\right)$$

$$\Rightarrow \mu = 78.3 \pm 1.032$$

$$\Rightarrow C.I \Rightarrow (78.3 - 1.032\,, 78.3 + 1.032) = (77.268\,, 79.332)$$

9. Let the observed value of the mean $\bar{X}$of a random sample of size 20 from a normal distribution with *mean* μ and variance $\sigma^2 = 80$ be 81.2. Find a 90% and 95% confidence intervals for μ.

Sol.

Given the sample size n=20

Mean of sample $\bar{X} = 81.2$

Variance $\sigma^2 = 80 \Rightarrow \sigma = \sqrt{80} = 8.9442$

We know, Confidence level of 95%, 90% the corresponding z values are 1.96 , 1.645. This is determined from the normal distribution table.

Confidence interval $C.I. = \mu = Mean \pm Z(Standard Deviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

For 95%:

$$\therefore C.I. = \mu = 81.2 \pm \left(1.96 \times \frac{8.9442}{\sqrt{20}}\right)$$

$$\Rightarrow \mu = 81.2 \pm 3.92$$

$$\Rightarrow C.I = (81.2 - 3.92, 81.2 + 3.92) = (77.28, 85.12)$$

For 90%:

$$\therefore C.I. = \mu = 81.2 \pm \left(1.645 \times \frac{8.9442}{\sqrt{20}}\right)$$
$$\Rightarrow \mu = 81.2 \pm 3.29$$
$$\Rightarrow C.I = (81.2 - 3.29, 81.2 + 3.29) = (77.91, 84.49)$$

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.

Sol.

Given samples are 10, 12, 16 and 19

Therefore, sample size n=4

Mean $\bar{X}$=14.25

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

We know, Confidence level of 95%, the corresponding z value is 1.96, This is determined from the normal distribution table.

Confidence interval $C.I. = \mu = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = \mu = 14.25 \pm \left(1.96 \times \frac{2.5}{\sqrt{4}}\right)$$
$$\Rightarrow \mu = 14.25 \pm 2.45$$
$$\Rightarrow C.I = (14.25 - 2.45, 14.25 + 2.45) = (11.80, 16.70)$$

SAMPLING DISTRIBUTIONS

Student's t -distribution:

Let μbe 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 the Student's t -distribution is defined as

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

Another formula for t - test of two samples is

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

where, $s^2 = \frac{n_1{\sigma_1}^2 + n_2{\sigma_2}^2}{n_1+n_2-2}$ or $s = \sqrt{\frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2\right]}$

Chi-square distribution:

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

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \frac{(O_3-E_3)^2}{E_3} + \ldots\ldots + \frac{(O_n-E_n)^2}{E_n}$$
$$\Rightarrow \chi^2 = \sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}$$

F-Distribution:

The F-distribution is useful in hypothesis testing. Hypothesis testing is used by scientists to statistically compare data from two or more populations. The F-distribution is needed to determine whether the F-value for a study indicates any statistically significant differences between two populations.

F-test is to determine whether the two independent estimates of population variance differ significantly. In this case, F-ratio is : $F = \frac{{\sigma_1}^2}{{\sigma_2}^2}$ where $\sigma^2 = \frac{1}{n}\sum(x-\mu)^2$ or

$$F = \frac{{s_1}^2/{\sigma_1}^2}{{s_2}^2/{\sigma_2}^2},$$

where σ_1=Standard deviation of population-1
σ_2=Standard deviation of population-2
s_1=Standard deviation of sample-1
s_2=Standard deviation of sample-2

To find out whether the two samples drawn from the normal population have the same variance. In this case, F-ratio is,

$$F = \frac{{s_1}^2}{{s_2}^2} \text{ Where } {s_1}^2 = \frac{1}{n_1-1}\sum(x-\bar{x})^2, {s_2}^2 = \frac{1}{n_2-1}\sum(y-\bar{y})^2$$

It should be noted that numerator is always greater than the denominator in F-ratio

$$F = \frac{Larger\ Variance}{Smaller\ Variance}$$

n_1= d.f for sample having larger variance
n_2 = d.f for sample having smaller variance

Expected value of F:

$$F_E = \frac{{s_2}^2}{{s_1}^2} \text{ follows F- distribution with } v_1 = n_1 - 1,\ v_2 = n_2 - 1 \text{ d.f.}$$

11. A certain stimulus administered to each of the 12 patients resulted in the following change 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 blood pressure? (*Note:* $t_{0.05}$ for 11 d.f. is 2.201).

Sol.

Given the change in blood pressure

x: 5,2,8,-1,3,0,6,-2,1,5,0,4

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

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

$$\Rightarrow s^2 = \frac{1}{11}\begin{Bmatrix}(5-2.58)^2+(2-2.58)^2+(8-2.58)^2+(-1-2.58)^2+(0-2.58)^2+(6-2.58)^2 \\ +(-2-2.58)^2+(1-2.58)^2+(5-2.58)^2+(0-2.58)^2+(4-2.58)^2\end{Bmatrix}$$

$$\Rightarrow s^2 = 9.538 \Rightarrow s = 3.088$$

Let us suppose that the stimulus administration is not accompanied with increase in blood pressure, we can take $\mu = 0$

we have,

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

$$\Rightarrow t = \frac{2.5833-0}{\left(\frac{3.088}{\sqrt{12}}\right)}$$

$$\Rightarrow t = 2.8979 \approx 2.9 > 2.201$$

Hence the hypothesis is rejected at 5%level of significance. We conclude with 95% Confidence that the stimulus in general is accompanied with increase of blood pressure.

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

Sol.

Given the I.Q. of 10 boys

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

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

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

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

$\Rightarrow s^2 = 203.73333$

$\Rightarrow s = 14.2735$

Given the mean of population $\mu = 100$

We have,

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

$$\Rightarrow t = \frac{97.2-100}{\left(\frac{14.2735}{\sqrt{10}}\right)}$$

$$\Rightarrow t = \frac{-2.8}{4.5136} \approx -0.6203 < 2.262$$

13. Ten individuals are chosen 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 66inches ($t_{0.05}$ =2.262 for 9 d.f.).

Sol.

Given the heights of the population in inches

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

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

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

$\Rightarrow s^2 =$

$$\frac{1}{9}\begin{Bmatrix}(63-67.8)^2+(63-67.8)^2+(66-67.8)^2+(67-67.8)^2+(68-67.8)^2+(69-67.8)^2\\ +(70-67.8)^2+(70-67.8)^2+(71-67.8)^2+(71-67.8)^2\end{Bmatrix}$$

$\Rightarrow s^2 = 9.067 \Rightarrow s = 3.011$

And given the mean of population $\mu = 66$

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

$$\Rightarrow t = \frac{67.8-66}{\left(\frac{3.011}{\sqrt{10}}\right)}$$

We have $\Rightarrow t = 1.8979 \approx 1.89 > 2.262$

Thus, the hypothesis is accepted at 5% level of significance.

14. The nine items 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 at 5% significance level?

Sol.

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

Therefore, sample size n=9

Population Mean $\mu = 47.50$

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

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

$\Rightarrow s^2 =$

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

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

∴Thee Null hypothesis $H_0: \mu = 47.5$

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

$$\Rightarrow t = \frac{49.11-47.5}{\left(2.6196/\sqrt{9}\right)}$$

$$\Rightarrow t = \frac{1.61}{0.8732}$$

$$\Rightarrow t = 1.8437$$

∴ Level of significance =5%

Critical value at 5 % level of significance for v=9-1=8 degrees of freedom is 2.3060.

Since the calculated value 1.8437 is less than the tabulated value 2.3060.

Hence the Null hypothesis is accepted.

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

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

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

Compute Student's t and test whether there is a significant difference in the two means.

Sol.

Given

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

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

We know that,

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

$$\Rightarrow s^2 = \frac{(10\times100)+(10\times121)}{10+10-2}$$

$$\Rightarrow s^2 = 122.78$$

$$\Rightarrow s = 11.0805$$

We have,

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

$$\Rightarrow t = \frac{560-500}{11.0805\sqrt{0.1+0.1}}$$

$$\Rightarrow t = 12.1081 \approx 12.11$$

The value of t is greater than the table value of t for 18d.f.at all levels of significance.

16. A group of boys and girls were given an intelligence test. The mean score , SD score and numbers in each group are as follows.

	Boys	Girls
Mean	74	70
SD	8	10
n	12	10

Is the difference between the means of the two groups significant at 5% level of significance ($t_{0.05} = 2.086$ for 20 $d.f.$)

Sol.

Given

$\bar{x}_1 = 74, \sigma_1 = 8, n_1 = 12\{Boys\}$

$\bar{x}_2 = 70, \sigma_2 = 10, n_2 = 10\{Girls\}$

We know that

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

$$\Rightarrow s^2 = \frac{(12\times64)+(10\times100)}{12+10-2}$$

$$\Rightarrow s^2 = \frac{1768}{20} = 88.4$$

$$\Rightarrow s = 9.402 \approx 9.4$$

We have

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

$$\Rightarrow t = \frac{74-70}{9.4\sqrt{\frac{1}{12}+\frac{1}{10}}} = \frac{4}{9.4\times0.4281} = \frac{4}{4.0244} = 0.9939$$

Thus, the hypothesis that there is a difference between the means of the two groups is accepted at 5% level of significance.

17. Two horses A and 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 discriminate between the two horses.

Sol.

Let the variables x and y respectively correspond to Horse A and B

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

y: 29,30,30,24,27,29

$\therefore \bar{x} = \frac{1}{n_1}\sum_i x_i = \frac{219}{7} = 31.30$, $\therefore \bar{y} = \frac{1}{n_2}\sum_i y_i = \frac{169}{6} = 28.20$

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

$\sum(y-\bar{y})^2 = (29-28.20)^2 + (30-28.20)^2 + (30-28.20)^2 + (24-28.20)^2 + (27-28.20)^2 + (29-28.20)^2 = 26.84$

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

$$\Rightarrow s^2 = \frac{31.4+26.84}{7+6-2} = 5.2973$$

$$\Rightarrow s = 2.3016$$

We have,

$$t = \frac{|(\bar{x}_2-\bar{x}_1)|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Rightarrow t = \frac{31.30-28.20}{2.3016\sqrt{\frac{1}{7}+\frac{1}{6}}} \Rightarrow t = 2.42 \begin{cases} > t_{0.05} = 2.2 \\ < t_{0.02} = 2.72 \end{cases}$$

18. 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 $\chi 0_{.05}{}^{2}$= 9.49 for 4 d.f.

Sol.

Given the 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

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

$P(0) = 4_{C0}(0.5)^0(0.5)^{4-0} = 0.0625$

$P(1) = 4_{C1}(0.5)^1(0.5)^{4-1} = 0.25$

$P(2) = 4_{C2}(0.5)^2(0.5)^{4-2} = 0.375$

$P(3) = 4_{C3}(0.5)^3(0.5)^{4-3} = 0.25$

$P(4) = 4_{C4}(0.5)^4(0.5)^{4-4} = 0.0625$

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

$E_1 = 100 \times 0.25 = 25$

$E_2 = 100 \times 0.375 = 37.5$

$E_3 = 100 \times 0.25 = 25$

$E_4 = 100 \times 0.0625 = 6.25$

where 100 is the sum of frequency

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

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

$\Rightarrow \chi^2 = \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + 0 + \frac{1.5625}{6.25}$

$\Rightarrow \chi^2 = 0.25 + 0.64 + 0.06 + 0.25$

$\Rightarrow \chi^2 = 1.2 < \chi^2{}_{0.05} = 9.49$

Hence the fitness is good.

19. 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.

Sol.

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

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

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

$\Rightarrow \chi^2 = \frac{1}{44}[16 + 144 + 256 + 196 + 100 + 256]\frac{968}{44}$

$\Rightarrow \chi^2 = 22$

20. A die was thrown 60 times and the following frequency distribution was observed:

Faces	1	2	3	4	5	6
Frequency	15	6	4	7	11	17

Test whether the die is unbiased at 5% significance level.

Sol.

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 $\frac{60}{6} = 10$ each.

Now the data is as follows:

x	1	2	3	4	5	6
O_i	15	6	4	7	11	17
E_i	10	10	10	10	10	10

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

$\Rightarrow \chi^2 = \frac{(15-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(17-10)^2}{10}$

$\Rightarrow \chi^2 = \frac{1}{10}[25 + 16 + 36 + 9 + 1 + 49] = \frac{136}{10}$

$\Rightarrow \chi^2 = 13.6$

21. 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?

Sol.

Given,

Number of families selected for the survey = 320

The probability of female and male birth is equal, $p = \frac{1}{2} = 0.5 \Rightarrow q = 1 - p = 1 - 0.5 = 0.5$

Number of children 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,

H_0: The probability of female and male birth is equal.

H_1: The probability of female and male birth is not equal.

Here Chi square distribution is used to test the hypothesis.

Therefore, by the Binomial distribution.

We have,

$P(x) = nC_x p^x q^{n-x}$

$\therefore P(x) = 5C_x (0.5)^x (0.5)^{5-x}$

$\Rightarrow P(x) = 5C_x (0.5)^5$

The expected frequencies can be calculated for 320 families as

$E(x) = 320 \times P(x) = 320 \times 5C_x (0.5)^5$

$\therefore E(0) = 320 \times P(0) = 320 \times 5C_0(0.5)^5 = 320 \times (0.5)^5 = 10 = E_0$

$E(1) = 320 \times P(1) = 320 \times 5C_1(0.5)^5 = 320 \times 5 \times (0.5)^5 = 50 = E_1$

$E(2) = 320 \times P(2) = 320 \times 5C_2(0.5)^5 = 320 \times 5C_2 \times (0.5)^5 = 100 = E_2$

$E(3) = 320 \times P(3) = 320 \times 5C_3(0.5)^5 = 320 \times 5C_3 \times (0.5)^5 = 100 = E_3$

$E(4) = 320 \times P(4) = 320 \times 5C_4(0.5)^5 = 320 \times 5C_4 \times (0.5)^5 = 50 = E_4$

$E(5) = 320 \times P(5) = 320 \times 5C_5(0.5)^5 = 320 \times 5C_5 \times (0.5)^5 = 10 = E_5$

No. of Boys	No. of Girls	Total Observed Frequencies (O_i)	Expected Frequencies(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 the Table value of χ^2 for 5 degrees of freedom at level of significance 5% from the chi-square table is 11.07.

$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] \Rightarrow \chi^2 = 7.16 < 11.02$

Since the calculated χ^2 value is less than tabulated χ^2 value then the decision is fail to reject the H_0 (Accept H_0) that means both the male and female birth is equal.

22. 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 four groups were 882, 313, 287 and 118. The chi square value is approximately equal to.

Sol.

Given,

The total number of beans: 882+313+287+118=1600

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

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

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

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

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

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{\sum(O_i - E_i)^2}{E_i}$
882	900	-18	324	0.36
313	300	13	169	0.5633
287	300	-13	169	0.5633
118	100	18	324	3.24

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

$\Rightarrow \chi^2 = 4.72$

23. 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**	**37**

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

Sol.

Set Null Hypothesis: $H_0: {\sigma_1}^2 = {\sigma_2}^2$

i.e., The two samples are drawn from two populations having the same variance.

Alternate Hypothesis: $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$

Given,

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	37

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \Rightarrow \bar{x}_1 = \frac{20+16+26+27+22+23+18+24+19+25}{10} \Rightarrow \bar{x}_1 = \frac{220}{10} \Rightarrow \bar{x}_1 = 22$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_i}{n_2} \Rightarrow \bar{x}_2 = \frac{27+33+42+35+32+34+38+28+41+43+30+37}{12} \Rightarrow \bar{x}_2 = \frac{420}{12} \Rightarrow \bar{x}_2 = 35$$

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
220	**0**	**120**	30	–5	25
			37	2	4
			420	**0**	**314**

∴The statistic F is defined by the ratio:

$$F_0 = \frac{{S_1}^2}{{S_2}^2} ------(1)$$

where ${S_1}^2 = \frac{1}{n_1 - 1}\sum(x_1 - \bar{x}_1)^2 = \frac{120}{9} = 13.33$

${S_2}^2 = \frac{1}{n_2 - 1}\sum(x_2 - \bar{x}_2)^2 = \frac{314}{11} = 28.54$

Since ${S_2}^2 > {S_1}^2$, $F = \frac{L\,arg\,erVariance}{SmallerVariance}$

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

Expected Value:

$F_E = \frac{{S_2}^2}{{S_1}^2}$, follows F-distribution with the degrees of freedom as given below for 5% level of significance:

$v_1 = n_1 - 1 = 10 - 1 = 9, v_2 = n_2 - 1 = 12 - 1 = 11$ is 3.10

Since $F_0 < F_E$ we accept null hypothesis at 5% level of significance and conclude that the two samples may be regarded as drawn from the populations having same variance.

24. The table shows the standard Deviation and Sample Standard Deviation for both men and women. Find the f statistic considering the Men population in numerator.

Population	Population Standard Deviation	Sample Standard Deviation
Men	30	35
Women	50	45

Sol.

Given,

σ_1=Standard deviation of population-1=30

σ_2=Standard deviation of population-2=50

s_1=Standard deviation of sample-1=35

s_2=Standard deviation of sample-2=45

We know that,

$$F = \frac{\frac{{s_1}^2}{{\sigma_1}^2}}{\frac{{s_2}^2}{{\sigma_2}^2}} \Rightarrow F = \frac{\left({35^2}/{30^2}\right)}{\left({45^2}/{50^2}\right)} \Rightarrow F = \frac{\left({1225}/{900}\right)}{\left({2025}/{2500}\right)} \Rightarrow F = \frac{1.3610}{0.81} \Rightarrow F = 1.68$$

***