

UNIT-4 » APPLIED STATISTICS**❖ INTRODUCTION:**

- ✓ Many problems in engineering required that we decide which of two competing claims for statements about parameter is true. Statements are called Hypotheses, and the decision making procedure is called hypotheses testing. This is one of the most useful aspects of statistical inference, because many type of decision making problems, tests or experiments in the engineering world can be formulated as hypotheses testing problems.

❖ POPULATION OR UNIVERSE

- ✓ An aggregate of objects (animate or inanimate) under study is called population or universe. It is thus a collection of individuals or of their attributes (qualities) or of results of operations which can be numerically specified.
- ✓ A universe containing a finite number of individuals or members is called a finite universe. For example, the universe of the weights of students in a particular class or the universe of smokes in Rothay district.
- ✓ A universe with infinite number of members is known as an infinite universe. For example, the universe of pressures at various points in the atmosphere.
- ✓ In some cases, we may be even ignorant whether or not a particular universe is infinite, e.g., the universe of stars.
- ✓ The universe of concrete objects is an existent universe. The collection of all possible ways in which a specified event can happen is called a hypothetical universe. The universe of heads and tails obtained by tossing an infinite number of times is a hypothetical one.

❖ SAMPLING

- ✓ A finite sub-set of a universe or population is called a sample. A sample is thus a small portion of the universe. The number of individuals in a sample is called the sample size. The process of selecting a sample from a universe is called sampling.
- ✓ The theory of sampling is a study of relationship between a population and samples drawn from the population. The fundamental object of sampling is to get as much information as possible of the whole universe by examining only a part of it.

- ✓ Sampling is quite often used in our day-to-day practical life. For example, in a shop we assess the quality of sugar, rice or any commodity by taking only a handful of it from the bag and then decide whether to purchase it or not.

❖ TEST OF SIGNIFICANCE

- ✓ An important aspect of the sampling theory is to study the test of significance. Which will enable us to decide, on the basis of the results of the sample. Whether
- ✓ The deviation between observed sample statistic and the hypothetical parameter value
- ✓ The deviation between two samples statistics is significant or might be attributed due to chance or the fluctuations of the sampling.
- ✓ For applying the tests of significance, we first set up a hypothesis which is a definite statement about the population parameter called null hypothesis denoted by H_0 .
- ✓ Any hypothesis which is complementary to the null hypothesis (H_0) is called an alternative hypothesis denoted by H_1 .
- ✓ For example, if we want to test the null hypothesis that the population has a specified mean μ_0 , then we have $H_0 : \mu = \mu_0$
- ✓ Alternative hypothesis will be
 - $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$) (Two tailed alternative hypothesis).
 - $H_1 : \mu > \mu_0$ (Right tailed alternative hypothesis or single tailed).
 - $H_1 : \mu < \mu_0$ (Left tailed alternative hypothesis or single tailed).
- ✓ Hence alternative hypothesis helps to know whether the test is two tailed or one tailed test.

❖ STANDARD ERROR

- ✓ The standard deviation of the sampling distribution of a statistic is known as the standard error.
- ✓ It plays an important role in the theory of large samples and it forms a basis of testing of hypothesis. If t is any statistic, for large sample. Then

$$z = \frac{t - E(t)}{S.E(t)}$$

is normally distributed with mean 0 and variance unity.

- ✓ For large sample, the standard errors of some of the well-known statistic are listed below:

No.	Statistic	Standard error
1	\bar{x}	$\frac{\sigma}{\sqrt{n}}$
2	S	$\sqrt{\frac{\sigma^2}{2n}}$
3	Difference of two sample means $\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
4	Difference of two sample standard deviation $s_1 - s_2$	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$
5	Difference of two sample proportions $p_1 - P_2$	$\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$

❖ ERRORS IN SAMPLING

- ✓ The main aim of the sampling theory is to draw a valid conclusion about the population parameters. On the basis of the same results. In doing this we may commit the following two type of errors:

- Type I error: When H_0 is true, we may reject it.

$$P(\text{Reject } H_0 \text{ when it is true}) = P\left(\text{Reject } \frac{H_0}{H_0}\right) = \alpha$$

Where α is called the size of the type I error also referred to as product's risk.

- Type II error: When H_0 is wrong we may accept it.

$$P(\text{Accept } H_0 \text{ when it is wrong}) = P\left(\text{Accept } \frac{H_0}{H_1}\right) = \beta$$

Where β is called the size of the type II error, also referred to as consumer's risk.

❖ **STEPS FOR TESTING OF STATISTICAL HYPOTHESIS:**

- ✓ **Step 1:** Null hypothesis.
 - Set up H_0 in clear terms.(Always in Equality.)

- ✓ **Step 2:** Alternative hypothesis.
 - Set up H_1 , so that we could decide whether we should use one-tailed or two-tailed test.(Always less than or greater than or not equal.)

- ✓ **Step 3:** Level of significance.
 - Select appropriate level of significance in advance depending on reality of estimates.

- ✓ **Step 4:** Critical region.
 - Given in data.

- ✓ **Step 5:** Test statistic.
 - Under null hypothesis compute the test statistic

$$z = \frac{t - E(t)}{S. E(t)}$$

- ✓ **Step 6:** Conclusion.
 - Compare the computed value of z with critical value z_α at the level of significance (α).
 - If $|z| > z_\alpha$, we reject H_0 and conclude that there is significant difference.
 - If $|z| < z_\alpha$, we accept H_0 and conclude that there is no significant difference.
 - It means if test statistic value belongs to critical Region, then we reject H_0 otherwise we accept H_0 .

LARGE SAMPLE ($n \geq 30$)

❖ TEST FOR SINGLE PROPORTION

- ✓ This test is used to find the significant difference between proportion of sample & population.
- ✓ Let X be number of successes in n independent trials with constant probability P of success for each trial.

➤ $E(X) = nP$; $V(X) = nPQ$; $Q = 1 - P =$ Probability of failure.

➤ $E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{nP}{n} = P$

➤ $V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{1(PQ)}{n} = \frac{PQ}{n}$

➤ $S.E.(p) = \sqrt{\frac{PQ}{n}}$; $Z = \frac{p-E(p)}{S.E(p)} \sim N(0,1)$

$$\text{i.e. } z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

- ✓ This z is called test statistics which is used to test the significant difference of sample and population proportion.
- ✓ The probable limit for the observed proportion of successes are $p \pm z_{\alpha} \sqrt{\frac{PQ}{n}}$, where z_{α} is the significant value at level of significance α .
- ✓ If P is not known, the limits for proportion in the population are $p \pm z_{\alpha} \sqrt{\frac{pq}{n}}$, $q = 1 - p$.
- ✓ If α is not known, we can take safely 3σ limits.
- ✓ Hence, confidence limits for observed proportion p are $p \pm 3 \sqrt{\frac{PQ}{n}}$.
- ✓ The confidence limits for the population proportion p are $p \pm \sqrt{\frac{pq}{n}}$.

METHOD – 1: TEST FOR SINGLE PROPORTION

C	1	A political party claims that 45% of the voters in an election district prefer its candidate. A sample of 200 voters include 80 who prefer this candidate. Test if the claim is valid at the 5% significance level. ($z_{0.05} = 1.96$) Answer: The party's claim might be valid.	
H	2	In a sample of 400 parts manufactured by a factory, the number of defective parts found to be 30. The company, however, claims that only 5% of their product is defective. Is the claim tenable? (Take level of significance 5%) ($z_{0.05} = 1.645$) Answer: The claim of manufacturer is not tenable (acceptable).	
C	3	A certain cubical die was thrown 9000 times and 5 or 6 was obtained 3240 times. On the assumption of certain throwing, do the data indicate an unbiased die? ($z_{0.05} = 1.96$) Answer: The die is unbiased.	
H	4	A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased. Answer: The coin is unbiased. ($z_{0.05} = 1.96$)	

❖ TEST FOR DIFFERENCE BETWEEN PROPORTIONS

- ✓ Consider two samples X_1 and X_2 of sizes n_1 and n_2 respectively taken from two different population. To test the significance of the difference between sample proportion p_1 & p_2 .
- ✓ The test statistic under the null hypothesis H_0 , that there is no significant difference between the two sample proportions, we have

$$z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\text{where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \text{ and } Q = 1 - P.$$

METHOD – 2: TEST FOR DIFFERENCE BETWEEN PROPORTIONS

C	1	In a certain city A, 100 men in a sample of 400 are found to be smokers. In another city B, 300 men in a sample of 800 are found to be smokers. Does this indicate that there is greater proportion of smokers in B than in A? ($z_{0.05} < -1.645$) Answer: The proportion of smokers is greater in city B than in A.
H	2	500 Articles from a factory are examined and found to be 2% defective. 800 Similar articles from a second factory are found to have only 1.5% defective. Can it reasonably concluded that the product of first factory are inferior than those of second? ($z_{0.05} > 1.645$) Answer: Products do not differ in quality.
H	3	Before an increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the duty, 800 persons were known to be tea drinkers in a sample of 1200 persons. Do you think that there is a significant decrease in the consumption of tea after the increase in the excise duty? ($z_{0.05} > 2.33$) Answer: There is significant decrease in consumption of tea.
C	4	A question in a true-false is considered to be smart if it discriminates between intelligent person (IP) and average person (AP). Suppose 205 out of 250 IP's and 137 out of 250 AP's answer a quiz question correctly. Test of 0.01 level of significance whether for the given question, proportion of correct answers can be expected to be at least 15% higher among IP's than among the AP's. ($z_{0.05} < -1.645$) Answer: Proportion of correct answer by IP's is 15% more than those by AP's.

❖ TEST FOR SINGLE MEAN

- ✓ To test whether the difference between sample mean and population mean is significant or not.
- ✓ Let X_1, X_2, \dots, X_n be a random sample of size n from a large population X_1, X_2, \dots, X_N of size N with mean μ and variance σ^2 . Therefore the standard error of mean of a random sample of size n from a population with variance σ^2 is $\frac{\sigma}{\sqrt{n}}$.

- ✓ To test whether given sample of size n has been drawn from a population with mean μ i.e., to test whether the difference between the sample mean and population mean is significant or not. Under the null hypothesis that there is no difference between the sample mean and population mean.
- ✓ The test statistic is

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

where σ is the standard deviation of the population.

- ✓ If σ is not known, we use test statistic $z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$, where s is standard deviation of the sample.
- ✓ If the level of significance is α and z_α is the critical value $-z_\alpha < |z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| < z_\alpha$
- ✓ The limit of the population mean μ are given by $\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$.
- ✓ Confidence limits:
 - At 5% of level of significance, 95% confidence limits are $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$
 - At 1% of level of significance, 99% confidence limits are $\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$

METHOD - 3: TEST FOR SINGLE MEAN

C	1	<p>Let X be the length of a life of certain computer is approximately normally distributed with mean 800 days and standard deviation 40 days. If a random sample of 30 computers has an average life of 788 days, test the null hypothesis that $\mu \neq 800$ days at (a) 0.5 %, (b) 15% level of significance. ($z_{0.05} = 1.96$, $z_{0.15} = 1.44$)</p> <p>Answer: (a) Accept null hypothesis, (b) Reject null hypothesis.</p>	
---	---	---	--

H	2	The mean weight obtained from a random sample of size 100 is 64 gms. The S.D. of the weight distribution of the population is 3 gms. Test the statement that the mean weight of the population is 67 gms. at 5% level of significance. ($z_{0.05} = 1.96$) Answer: The mean weight of the population is not 67 gms.	
C	3	A college claims that its average class size is 35 students. A random sample of 64 students from class has a mean of 37 with a standard deviation of 6. Test at the $\alpha = 0.05$ level of significance if the claimed value is too low. ($z_{0.05} > 1.645$) Answer: The true mean class size is likely to be more than 35.	
H	4	Sugar is packed in bags by an automation machine with mean contents of bags as 1.000 kg. A random sample of 36 bags is selected and mean mass has been found to be 1.003 kg. If a S.D. of 0.01 kg is acceptable on all the bags being packed, determine on the basis of sample test whether the machine requires adjustment. ($z_{0.05} = 1.96$) Answer: The machine does not require any adjustment.	

❖ TEST FOR DIFFERENCE BETWEEN MEANS

- ✓ Let \bar{x}_1 be the mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2 . Let \bar{x}_2 be the mean of an independent sample of size n_2 from another population with mean μ_2 and variance σ_2^2 . The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- ✓ Under the null hypothesis that the samples are drawn from the same population where $\sigma_1 = \sigma_2 = \sigma$ i.e., $\mu_1 = \mu_2$ the test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- ✓ If σ_1, σ_2 are not known and $\sigma_1 \neq \sigma_2$ the test statistic in this case is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

✓ If σ is not known and $\sigma_1 = \sigma_2$ we use $\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$ to calculate σ ,

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

METHOD - 4: TEST FOR DIFFERENCE BETWEEN MEANS

C	1	<p>In a random sample of 100 light bulbs manufactured by a company A, the mean lifetime of light bulb is 1190 hours with standard deviation of 90 hours. Also, in a random sample of 75 light bulbs manufactured by company B, the mean lifetime of light bulb is 1230 hours with standard deviation of 120 hours. Is there a difference between the mean lifetime of the two brands of light bulbs at a significance level of (a) 0.05, (b) 0.01? ($z_{0.05} = 1.96$, $z_{0.01} = 2.58$)</p> <p>Answer: (a) There is difference between the mean lifetimes. (b) There is no difference between the mean lifetimes.</p>	
H	2	<p>A company A manufactured tube lights and claims that its tube lights are superior than its main competitor company B. The study showed that a sample of 40 tube lights manufactured by company A has a mean lifetime of 647 hours of continuous use with a standard deviation of 27 hours, while a sample of 40 tube lights manufactured by company B had a mean lifetime of 638 hours of continuous use with a standard deviation of 31 hours. Does this substantiate the claim of company A that their tube lights are superior than manufactured by company B at (a) 0.05, (b) 0.01 level of significance? ($z_{0.05} > 1.645$, $z_{0.01} > 2.33$)</p> <p>Answer: (a) The claim of company A is not valid. (b) The claim of company A is not valid.</p>	

C	3	<p>For sample I, $n_1 = 1000, \sum x = 49,000, \sum(x - \bar{x})^2 = 7,84,000$.</p> <p>For sample II, $n_2 = 1500, \sum x = 70,500, \sum(x - \bar{x})^2 = 24,00,000$.</p> <p>Discuss the significance of the difference of the sample means. ($z_{0.05} = 1.96$)</p> <p>Answer: No significant difference between the sample means.</p>										
H	4	<p>A company claims that alloying reduces resistance of electric wire by more than 0.050 ohm. To test this claim samples of 32 standard wire and alloyed wire are tested yielding the following results. ($z_{0.05} > 1.645$)</p> <table border="1" style="margin: 10px auto;"> <thead> <tr> <th>Type of wire</th> <th>Mean resistance (ohms)</th> <th>S.D. (ohms)</th> </tr> </thead> <tbody> <tr> <td>Standard</td> <td>0.136</td> <td>0.004</td> </tr> <tr> <td>Alloyed</td> <td>0.083</td> <td>0.005</td> </tr> </tbody> </table> <p>At the 0.05 level of significance, does this support the claim?</p> <p>Answer: The data supports the claim.</p>	Type of wire	Mean resistance (ohms)	S.D. (ohms)	Standard	0.136	0.004	Alloyed	0.083	0.005	
Type of wire	Mean resistance (ohms)	S.D. (ohms)										
Standard	0.136	0.004										
Alloyed	0.083	0.005										

❖ TEST FOR DIFFERENCE BETWEEN STANDARD DEVIATIONS

- ✓ If s_1 and s_2 are the standard deviations of two independent samples, then under the null hypothesis $H_0: \sigma_1 = \sigma_2$, i.e., the population standard deviation don't differ significantly, the static is

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

where σ_1 and σ_2 are population standard deviations.

- ✓ When population standard deviations are not known then

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

where s_1 and s_2 are sample standard deviations.

METHOD – 5: TEST FOR DIFFERENCE BETWEEN STANDARD DEVIATIONS

C	1	<p>Random samples drawn from two countries gave the following data relating to the heights of adult males:</p> <table border="1"> <thead> <tr> <th></th> <th>Country A</th> <th>Country B</th> </tr> </thead> <tbody> <tr> <td>Standard deviation</td> <td>2.58</td> <td>2.5</td> </tr> <tr> <td>Number in samples</td> <td>1000</td> <td>1200</td> </tr> </tbody> </table> <p>Is the difference between the standard deviation significant? ($z_{0.05} = 1.96$)</p> <p>Answer: The sample standard deviations do not differ significantly.</p>		Country A	Country B	Standard deviation	2.58	2.5	Number in samples	1000	1200	
	Country A	Country B										
Standard deviation	2.58	2.5										
Number in samples	1000	1200										
H	2	<p>Intelligence test of two groups of boys and girls gives the following results:</p> <table border="1"> <thead> <tr> <th></th> <th>n</th> <th>S.D.</th> </tr> </thead> <tbody> <tr> <td>Girls</td> <td>121</td> <td>10</td> </tr> <tr> <td>Boys</td> <td>81</td> <td>12</td> </tr> </tbody> </table> <p>Is the difference between the standard deviations significant?</p> <p>($z_{0.05} = 1.96$)</p> <p>Answer: The sample standard deviations do not differ significantly.</p>		n	S.D.	Girls	121	10	Boys	81	12	
	n	S.D.										
Girls	121	10										
Boys	81	12										
C	3	<p>The mean yield of two plots and their variability are as given below:</p> <table border="1"> <thead> <tr> <th></th> <th>40 plots</th> <th>60 plots</th> </tr> </thead> <tbody> <tr> <td>S.D.</td> <td>34</td> <td>28</td> </tr> </tbody> </table> <p>Check whether the difference in the variability in yields is significant. ($z_{0.05} = 1.96$)</p> <p>Answer: The sample standard deviations do not differ significantly.</p>		40 plots	60 plots	S.D.	34	28				
	40 plots	60 plots										
S.D.	34	28										
H	4	<p>The yield of wheat in a random sample of 1000 farms in a certain area has a S.D. of 192 kg. Another random sample of 1000 farms gives a S.D. of 224 kg. Are the S. Ds significantly different? ($z_{0.05} = 1.96$)</p> <p>Answer: The sample standard deviations are significantly different.</p>										

SMALL SAMPLE ($n < 30$)

❖ T-TEST FOR SINGLE MEAN

- ✓ To test whether the mean of a sample drawn from a normal population deviates significantly from a stated value when variance of the population is unknown.
- ✓ H_0 : There is no significant difference between the sample mean \bar{X} and the population mean μ i. e., we use the static

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

$$\text{where } s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- ✓ This test static is known as one sample t-test.

METHOD – 6: T-TEST FOR SINGLE MEAN

C	1	A random sample of size 16 has 53 as mean. The sum of squares of the derivation from mean is 135. Can this sample be regarded as taken from population having 56 as mean? (value of t for 15 degree of freedom at 5% level of significance is 2.131) Answer: The sample mean has not come from a population mean.	
H	2	A machine is designed to produce insulting washers for electrical devices of average thickness of 0.025 cm. A random sample of 10 washers was found to have an average thickness of 0.024 cm with S.D. of 0.002 cm. Test the significance of the deviation.(value of t for 9 degree of freedom at 5% level of significance is 2.262) Answer: There is no significant difference between population mean and sample mean.	

C	3	Ten individuals were chosen random from a normal population and their heights were found to be in inches 63,63,66,67,68,69,70,70,71 and 71. Test the hypothesis that the mean height of the population is 66 inches. (value of t for 9 degree of freedom at 5% level of significance is 2.262) Answer: There is no significant difference between population mean and sample mean.
H	4	The 9 items of a sample have the values 45,47,50,52,48,47,49,53,51. Does the mean of these values differ significantly from assumed mean 47.5? (value of t for 9 degree of freedom at 5% level of significance is 2.262) Answer: The mean of given values does not differ significantly from assumed mean 47.5.

❖ T-TEST FOR DIFFERENCE BETWEEN MEANS

- ✓ This test is used to test whether the two samples $x_1, x_2, x_3, \dots, x_{n_1}$ and y_1, y_2, \dots, y_{n_2} of sizes n_1 and n_2 have been drawn from two normal populations with mean μ_1 and μ_2 respectively under the assumption that the population variance are equal. ($\sigma_1 = \sigma_2 = \sigma$)
- ✓ H_0 : The samples have been drawn from the normal population with means μ_1 and μ_2
- ✓ i.e., $H_0 : \mu_1 = \mu_2$.
- ✓ Let \bar{X}, \bar{Y} be their means of the two samples.
- ✓ Under this H_0 the test static t is given by

$$t = \frac{(\bar{X} - \bar{Y})}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- ✓ If the two sample standard deviations s_1, s_2 are given then we have

$$\sigma^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

METHOD – 7: T-TEST FOR DIFFERENCE BETWEEN MEANS

C	1	<p>Two sample of 6 and 5 items, respectively, gave the following data.</p> <table border="1"> <thead> <tr> <th></th> <th>1st sample</th> <th>2nd sample</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>40</td> <td>50</td> </tr> <tr> <td>S.D.</td> <td>8</td> <td>10</td> </tr> </tbody> </table> <p>Is the difference of the means significant?(Test at 5% level of significance) (The value of t for 9 degree of freedom at 5% level is 2.262)</p> <p>Answer: There is no significant difference between two population means.</p>		1st sample	2nd sample	Mean	40	50	S.D.	8	10														
	1st sample	2nd sample																							
Mean	40	50																							
S.D.	8	10																							
H	2	<p>Two sample of 10 and 14 items, respectively, gave the following data.</p> <table border="1"> <thead> <tr> <th></th> <th>1st sample</th> <th>2nd sample</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>20.3</td> <td>18.6</td> </tr> <tr> <td>S.D.</td> <td>3.5</td> <td>5.2</td> </tr> </tbody> </table> <p>Is the difference of the means significant?(Test at 5% level of significance) (The value of t for 22 degree of freedom at 5% level is 2.0739)</p> <p>Answer: There is no significant difference between two population means.</p>		1st sample	2nd sample	Mean	20.3	18.6	S.D.	3.5	5.2	W 2019 (3)													
	1st sample	2nd sample																							
Mean	20.3	18.6																							
S.D.	3.5	5.2																							
C	3	<p>A large group of teachers are trained, where some are trained by institution A and some are trained by institution B. In a random sample of 10 teachers taken from a large group, the following marks are obtained in an appropriate achievement test.</p> <table border="1"> <tbody> <tr> <td>Institution A</td> <td>65</td> <td>69</td> <td>73</td> <td>71</td> <td>75</td> <td>66</td> <td>71</td> <td>68</td> <td>68</td> <td>74</td> </tr> <tr> <td>Institution B</td> <td>78</td> <td>69</td> <td>72</td> <td>77</td> <td>84</td> <td>70</td> <td>73</td> <td>77</td> <td>75</td> <td>65</td> </tr> </tbody> </table> <p>Test the claim that institute B is more effective. (The value of t for 18 degree of freedom at 5% level is 1.734)</p> <p>Answer: The claim is valid.</p>	Institution A	65	69	73	71	75	66	71	68	68	74	Institution B	78	69	72	77	84	70	73	77	75	65	
Institution A	65	69	73	71	75	66	71	68	68	74															
Institution B	78	69	72	77	84	70	73	77	75	65															

H	4	<p>Random samples of specimens of coal from two mines A & B are drawn and their heat producing capacity (in millions of calories per ton) were measured yielding the following result:</p> <table border="1"> <tr> <td>Mine A</td> <td>8260</td> <td>8130</td> <td>8350</td> <td>8070</td> <td>8340</td> <td>—</td> </tr> <tr> <td>Mine B</td> <td>7950</td> <td>7890</td> <td>7900</td> <td>8140</td> <td>7920</td> <td>7840</td> </tr> </table> <p>Test whether the difference between the means of these two samples is significant. (The value of t for 9 degree of freedom at 5% level is 2.262)</p> <p>Answer: The average heat producing capacity of coal from two mines is not same.</p>	Mine A	8260	8130	8350	8070	8340	—	Mine B	7950	7890	7900	8140	7920	7840									
Mine A	8260	8130	8350	8070	8340	—																			
Mine B	7950	7890	7900	8140	7920	7840																			
H	5	<p>The following figures refer to observations in live independent samples:</p> <table border="1"> <tr> <td>Sample I</td> <td>25</td> <td>30</td> <td>28</td> <td>34</td> <td>24</td> <td>20</td> <td>13</td> <td>32</td> <td>22</td> <td>38</td> </tr> <tr> <td>Sample II</td> <td>40</td> <td>34</td> <td>22</td> <td>20</td> <td>31</td> <td>40</td> <td>30</td> <td>23</td> <td>36</td> <td>17</td> </tr> </table> <p>Analyze whether the samples have been drawn from the population of equal means. [t at 5% level of significance for 18 d.f. is 2.1] Test whether the means of two populations are same at 5% level (t at 0.05=2.0739).</p> <p>Answer: Samples have not been drawn from population with equal mean. Also, means of two populations are not same.</p>	Sample I	25	30	28	34	24	20	13	32	22	38	Sample II	40	34	22	20	31	40	30	23	36	17	W 2019 (7)
Sample I	25	30	28	34	24	20	13	32	22	38															
Sample II	40	34	22	20	31	40	30	23	36	17															

❖ T-TEST FOR CORRELATION COEFFICIENTS

- ✓ Consider a random sample of n observations from a bivariate normal population. Let r be the observed correlation coefficient and ρ be the population correlation coefficient.
- ✓ Under the null and alternative hypothesis as follows,
- ✓ $H_0 : \rho = 0$ (There is no correlation between two variables)
- ✓ $H_1 : \rho \neq 0$ or $\rho > 0$ or $\rho < 0$ (There is correlation between two variables)
- ✓ The test static t is given by

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$$

with $v = n - 2$ degrees of freedom.

METHOD – 8: T-TEST FOR CORRELATION COEFFICIENT

C	1	<p>The correlation coefficient between income and food expenditure for sample of 7 household from a low income group is 0.9. Using 1% level of significance, test whether the correlation coefficient between incomes and food expenditure is positive. Assume that the population of both variables are normally distributed.</p> <p>(The value of t for 5 degree of freedom at 1% level is 4.032)</p> <p>Answer: There is correlation between incomes and food expenditure.</p>	
H	2	<p>A random sample of fifteen paired observations from a bivariate population gives a correlation coefficient of -0.5. Does this signify the existence of correlation in the sample population? (The value of t for 13 degree of freedom at 5% level is 2.160)</p> <p>Answer: The sample population is uncorrelated.</p>	
C	3	<p>A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0.6. Is this significant of correlation in the population?</p> <p>(The value of t for 25 degree of freedom at 5% level is 2.06)</p> <p>Answer: The sample population is correlated.</p>	
H	4	<p>A coefficient of correlation of 0.2 is derived from a random sample of 625 pairs of observations. Is this value of r significant? (The value of t for 623 degree of freedom at 5% level is 1.96)</p> <p>Answer: It is highly significant.</p>	

❖ F-TEST FOR RATIO OF VARIANCES

- ✓ Let n_1 and n_2 be the sizes of two samples with variance s_1^2 and s_2^2 . The estimate of the population variance based on these samples are $s_1 = \frac{n_1 s_1^2}{n_1 - 1}$ and $s_2 = \frac{n_2 s_2^2}{n_2 - 1}$. The degrees of freedom of these estimates are $v_1 = n_1 - 1, v_2 = n_2 - 1$.
- ✓ To test whether these estimates are significantly different or if the samples may be regarded as drawn from the same population or from two populations with same variance σ^2 . We setup the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$

✓ So, the test static is

$$F = \frac{(s_1)^2}{(s_2)^2}, \text{ where } s_1^2 > s_2^2.$$

METHOD – 9: F-TEST FOR RATIO OF VARIANCES

C	1	<p>In two independent samples of sizes 8 and 10 the sum of squares of derivations of the samples values from the respective sample means were 84.4 and 102.6. Test whether the difference of variances of the populations is significant or not. (F for 7 and 9 d.f. = 3.29)</p> <p>Answer: There is no significant difference between the variances of two populations.</p>																		
H	2	<p>Two random samples reveal the following data:</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Sample no.</th> <th>Size</th> <th>Variance</th> </tr> </thead> <tbody> <tr> <td>I</td> <td>16</td> <td>40</td> </tr> <tr> <td>II</td> <td>25</td> <td>42</td> </tr> </tbody> </table> <p>Test whether the samples come from the same normal population. (F for 8 and 7 d.f. = 3.73)</p> <p>Answer: The population variances are equal.</p>	Sample no.	Size	Variance	I	16	40	II	25	42									
Sample no.	Size	Variance																		
I	16	40																		
II	25	42																		
C	3	<p>Two random samples drawn from 2 normal populations are as follows:</p> <table border="1" style="margin-left: 20px;"> <tbody> <tr> <td>A</td> <td>17</td> <td>27</td> <td>18</td> <td>25</td> <td>27</td> <td>29</td> <td>13</td> <td>17</td> </tr> <tr> <td>B</td> <td>16</td> <td>16</td> <td>20</td> <td>27</td> <td>26</td> <td>25</td> <td>21</td> <td>–</td> </tr> </tbody> </table> <p>Test whether the samples are drawn from the same normal population. (F for 7 and 6 d.f. = 1.19)</p> <p>Answer: The population variances are equal.</p>	A	17	27	18	25	27	29	13	17	B	16	16	20	27	26	25	21	–
A	17	27	18	25	27	29	13	17												
B	16	16	20	27	26	25	21	–												
H	4	<p>Two independent sample of size 7 and 6 had the following values:</p> <table border="1" style="margin-left: 20px;"> <tbody> <tr> <td>A</td> <td>28</td> <td>30</td> <td>32</td> <td>33</td> <td>31</td> <td>29</td> <td>34</td> </tr> <tr> <td>B</td> <td>29</td> <td>30</td> <td>30</td> <td>24</td> <td>27</td> <td>28</td> <td>–</td> </tr> </tbody> </table> <p>Examine whether the samples have been drawn from normal populations having the same variance. (F for 5 and 6 d.f. = 4.39)</p> <p>Answer: Samples have been drawn from the normal populations with same variance.</p>	A	28	30	32	33	31	29	34	B	29	30	30	24	27	28	–		
A	28	30	32	33	31	29	34													
B	29	30	30	24	27	28	–													

H	5	Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in kg): <table border="1" style="margin: 10px auto;"> <tr> <td>Sample I</td> <td>9</td> <td>11</td> <td>13</td> <td>11</td> <td>15</td> <td>9</td> <td>12</td> <td>14</td> </tr> <tr> <td>Sample II</td> <td>10</td> <td>12</td> <td>10</td> <td>14</td> <td>9</td> <td>8</td> <td>10</td> <td>—</td> </tr> </table> <p>Do the two estimates of population variance differ significantly? Given that for (7,6) d.f. the value of F at 5% level of significance is 4.20 nearly.</p> <p>Answer: There is no significant difference between the variances of two population.</p>	Sample I	9	11	13	11	15	9	12	14	Sample II	10	12	10	14	9	8	10	—	W 2019 (4)
Sample I	9	11	13	11	15	9	12	14													
Sample II	10	12	10	14	9	8	10	—													
H	6	Two samples of size 9 and 8 give the sum of squares of deviations from their respective means equal 160 inches and 91 inches respectively. Can they be regarded as drawn from two normal populations with the same variance? (F for 8 and 7 d.f. = 3.73). <p>Answer: There is no significant difference between the variances of the population.</p>	W 2019 (3)																		

❖ CHI-SQUARE TEST FOR GOODNESS OF FIT

✓ Pare-1

- Find the expected frequencies using general probability considerations or specific probability model (Poisson, binomial, normal) given in the problem itself.

✓ Part-2

- Testing under the null and alternative hypothesis as follows.
- H_0 : Given probability distribution fits good with the given data; that is ,there is no significant difference between observed frequencies (O_i) and expected frequencies (e_i)
- H_1 : Given probability distribution does not fit good with the given data; that is , there is significant difference between observed frequencies (O_i) and expected frequencies (e_i)

✓ The test static given by

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} \quad (\text{with } v = k - m \text{ degree of freedom})$$

- ✓ Note that the value of degree of freedom v for binomial, exponential and normal distribution is $n - 1$, $n - 2$ and $n - 3$, respectively.

METHOD-10: CHI-SQUARE TEST FOR GOODNESS OF FIT

C	1	<p>Suppose that a die is tossed 120 times and the recorded data is as follows:</p> <table border="1"> <tr> <td>Face Observed(x)</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> </tr> <tr> <td>Frequency</td> <td>20</td> <td>22</td> <td>17</td> <td>18</td> <td>19</td> <td>24</td> </tr> </table> <p>Test the hypothesis that the die is unbiased at $\alpha = 0.05$. $[\chi^2$ at 5% level of significance for 5 df is 11.070]</p> <p>Answer: The die is unbiased.</p>	Face Observed(x)	1	2	3	4	5	6	Frequency	20	22	17	18	19	24						
Face Observed(x)	1	2	3	4	5	6																
Frequency	20	22	17	18	19	24																
H	2	<p>The following table gives the number of accidents that took place in an industry during various days of the week. Test if accidents are uniformly distributed over the week.</p> <table border="1"> <tr> <td>Day</td> <td>Mon</td> <td>Tue</td> <td>Wed</td> <td>Thus</td> <td>Fri</td> <td>Sat</td> </tr> <tr> <td>No. of accidents</td> <td>14</td> <td>18</td> <td>12</td> <td>11</td> <td>15</td> <td>14</td> </tr> </table> <p>$[\chi^2$ at 5% level of significance for 5 df is 11.09]</p> <p>Answer: The accidents are uniformly distributed over the week.</p>	Day	Mon	Tue	Wed	Thus	Fri	Sat	No. of accidents	14	18	12	11	15	14						
Day	Mon	Tue	Wed	Thus	Fri	Sat																
No. of accidents	14	18	12	11	15	14																
C	3	<p>The following table indicates (a) the frequencies of a given distribution with (b) the frequencies of the normal distribution having the same mean, standard deviation and the total frequency as in (a).</p> <table border="1"> <tr> <td>(a)</td> <td>1</td> <td>5</td> <td>20</td> <td>28</td> <td>42</td> <td>22</td> <td>15</td> <td>5</td> <td>2</td> </tr> <tr> <td>(b)</td> <td>1</td> <td>6</td> <td>18</td> <td>25</td> <td>40</td> <td>25</td> <td>18</td> <td>6</td> <td>1</td> </tr> </table> <p>Apply the χ^2-test of goodness of fit. $[\chi^2$ at 5% level of significance for 4 df is 9.488]</p> <p>Answer: This normal distribution provides a good fit.</p>	(a)	1	5	20	28	42	22	15	5	2	(b)	1	6	18	25	40	25	18	6	1
(a)	1	5	20	28	42	22	15	5	2													
(b)	1	6	18	25	40	25	18	6	1													

H	4	<p>Suppose that during 400 five-minute intervals the air-traffic control of an airport received 0,1,2, ... , or 13 radio messages with respective frequencies of 3,15,47,76,68,74,46,39,15,9,5,2,0 and 1. Test at 0.05 level of significance, the hypothesis that the number of radio messages received during 5 minute interval follow Poisson distribution with $\lambda = 4.6$.</p> <p>$[\chi^2$ at 5% level of significance for 8 df is 15.507]</p> <p>Answer: Poisson distribution with $\lambda = 4.6$ provides a good fit.</p>																			
C	5	<p>Records taken of the number of male and female births in 830 families having four children are as follows:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>No. of male births</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>No. of female births</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> <td>0</td> </tr> <tr> <td>No. of families</td> <td>32</td> <td>178</td> <td>290</td> <td>236</td> <td>94</td> </tr> </table> <p>Test whether data are consistent with hypothesis that the binomial law holds and the chance of male birth is equal to that of female birth, namely $p = q = \frac{1}{2}$. $[\chi^2$ at 5% level of significance for 4 df is 9.49]</p> <p>Answer: The data are not consistence with the hypothesis.</p>	No. of male births	0	1	2	3	4	No. of female births	4	3	2	1	0	No. of families	32	178	290	236	94	W 2019 (7)
No. of male births	0	1	2	3	4																
No. of female births	4	3	2	1	0																
No. of families	32	178	290	236	94																
H	6	<p>A die is thrown 276 times and the results of these throws are given below:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>Number appeared on the die</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> </tr> <tr> <td>Frequency</td> <td>40</td> <td>32</td> <td>29</td> <td>59</td> <td>57</td> <td>59</td> </tr> </table> <p>Test whether the die is biased or not.</p> <p>$[\chi^2$ at 5% level of significance for 5 df is 11.09]</p> <p>Answer: The die is biased.</p>	Number appeared on the die	1	2	3	4	5	6	Frequency	40	32	29	59	57	59	W 2019 (4)				
Number appeared on the die	1	2	3	4	5	6															
Frequency	40	32	29	59	57	59															

❖ **CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES**

✓ Pare-1

- Construct a contingency table on the basis of given information and find expected frequency for each cell using

$$E_{ij} = \frac{\text{column total} * \text{row total}}{\text{grand total}}$$

✓ Part-2

- Testing under the null and alternative hypothesis as follows.
- H_0 : Attributes are independent; that is, there is no significant difference between observed frequencies (O_{ij}) and expected frequencies (E_{ij})
- H_1 : Attributes are dependent; that is, there is significant difference between observed frequencies (O_{ij}) and expected frequencies (E_{ij})

✓ The test static χ^2 for the analysis of $r \times c$ table is given by

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ with degree of freedom } v = (r - 1)(c - 1).$$

✓ Here, the hypothesis H_0 is tested using right one-tailed test.

METHOD-11: CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

C	1	<p>Test the hypothesis at 0.05 level of significance that the presence or absence of hypertension is independent of smoking habits from the following data of 80 persons.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Non smokers</th> <th>Moderate smokers</th> <th>Heavy smokers</th> </tr> </thead> <tbody> <tr> <th>HT</th> <td style="text-align: center;">21</td> <td style="text-align: center;">36</td> <td style="text-align: center;">30</td> </tr> <tr> <th>No HT</th> <td style="text-align: center;">48</td> <td style="text-align: center;">26</td> <td style="text-align: center;">19</td> </tr> </tbody> </table> <p>[χ^2 at 5% level of significance for 2 df is 5.991]</p> <p>Answer: Hypertension and smoking habits are not independent.</p>		Non smokers	Moderate smokers	Heavy smokers	HT	21	36	30	No HT	48	26	19
	Non smokers	Moderate smokers	Heavy smokers											
HT	21	36	30											
No HT	48	26	19											

H	2	<p>A company operates three machines on three different shifts daily. The following table presents the data of the machine breakdowns resulted during a 6-month time period.</p> <table border="1"> <thead> <tr> <th>Shift</th> <th>Machine A</th> <th>Machine B</th> <th>Machine C</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>12</td> <td>12</td> <td>11</td> <td>35</td> </tr> <tr> <td>2</td> <td>15</td> <td>25</td> <td>13</td> <td>53</td> </tr> <tr> <td>3</td> <td>17</td> <td>23</td> <td>10</td> <td>50</td> </tr> <tr> <td>Total</td> <td>44</td> <td>60</td> <td>34</td> <td>138</td> </tr> </tbody> </table> <p>Test hypothesis that for an arbitrary breakdown machine causing breakdown & the shift on which the breakdown occurs are independent. [χ^2 at 5% level of significance for 4 df is 9.488]</p> <p>Answer: Machine causing breakdown and the shift are independent.</p>	Shift	Machine A	Machine B	Machine C	Total	1	12	12	11	35	2	15	25	13	53	3	17	23	10	50	Total	44	60	34	138				
Shift	Machine A	Machine B	Machine C	Total																											
1	12	12	11	35																											
2	15	25	13	53																											
3	17	23	10	50																											
Total	44	60	34	138																											
C	3	<p>From the following data, find whether hair color and gender are associated.</p> <table border="1"> <thead> <tr> <th>Color</th> <th>Fair</th> <th>Red</th> <th>Medium</th> <th>Dark</th> <th>Black</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Boys</td> <td>592</td> <td>849</td> <td>504</td> <td>119</td> <td>36</td> <td>2100</td> </tr> <tr> <td>Girls</td> <td>544</td> <td>677</td> <td>451</td> <td>97</td> <td>14</td> <td>1783</td> </tr> <tr> <td>Total</td> <td>1136</td> <td>1526</td> <td>955</td> <td>216</td> <td>50</td> <td>3883</td> </tr> </tbody> </table> <p>[χ^2 at 5% level of significance for 4 df is 9.488]</p> <p>Answer: The hair color and gender are associated.</p>	Color	Fair	Red	Medium	Dark	Black	Total	Boys	592	849	504	119	36	2100	Girls	544	677	451	97	14	1783	Total	1136	1526	955	216	50	3883	
Color	Fair	Red	Medium	Dark	Black	Total																									
Boys	592	849	504	119	36	2100																									
Girls	544	677	451	97	14	1783																									
Total	1136	1526	955	216	50	3883																									

