

Practical Exercise -1

TEST FOR THE SIGNIFICANCE OF SINGLE MEAN

1. Suppose a random sample of size 37 was taken and the mean was found to be 61.4. Can we consider that the sample came from a population with mean 61 and standard deviation 1.01. Test the hypothesis at 1% level of significance.
2. A sample of 900 members has a mean 3.4 cms and standard deviation 2.61 cms. Is the sample from a large population of mean 3.25 cms and standard deviation 2.61 cms. Is the sample drawn from the given population or not?. Also find the 95% and 99% confidence limits for the population mean.
3. An insurance agent has claimed that the average age of policyholders who insured him is less than the average for all agents, which is 35 years. A random sample of 100 policyholders who had insured through him gave the following age distribution.

Age	No.of Persons
16-20	12
21-25	22
26-30	20
31-35	30
36-40	16

Calculate the arithmetic mean and standard deviation and use these values and test the claim at 5% level of significance.

4. A sample of 100 items drawn from the universe with mean value 64 and standard deviation 3 has a mean value 63.5. Is the difference in mean significant?

Aim:

To test the significance of single mean (Use Z- test)

Formula and Procedure:

Consider a population with finite mean μ and variance σ^2 . Our interest is to test the null hypothesis H_0

H_0 : The population mean $\mu = \mu_0$ (a given value)

Against the alternative hypothesis H_1

H_1 : The population mean $\mu \neq \mu_0$ or $\mu < \mu_0$ or $\mu > \mu_0$ (based on given population)

Let the level of significance be α

Suppose the X_1, X_2, \dots, X_n is a random sample of size n drawn from the above population. Let \bar{x} be the sample mean and s^2 be the sample variance. When n is large then,

$$x \sim N(\mu, \sigma^2)$$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Under the null hypothesis,

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

If the alternative hypothesis is two sided,

$$|z_{cal}| \leq z_{\alpha/2} \quad \text{Do not reject nul hypothesis, otherwise reject}$$

If the alternative hypothesis is one / single sided,

$|z_{cal}| \leq z_{\alpha}$ Do not reject nul hypothesis, otherwise reject

Note: 1. If Population variance σ^2 is not known, we can use sample variance s^2 . In such a case the test statistic is,

Under the null hypothesis,

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sim N(0, 1)$$

Note 2: 95 % Confidence limits for population mean μ

$$Prob\left\{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right\} = (1 - \alpha)$$

The $(1-\alpha)\%$ confidence interval for the population mean is

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Calculation :

Conclusion:

Practical Exercise -2

Test of significance for Difference of Means

1. The means of two single two large samples of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can samples be regarded as drawn from the same population of standard deviation 2.5 inches? Test at 5 percent level of significance.
2. In a survey of buying habits 400 women shoppers are chosen at random in super market A located in a certain locality of the city. Their average weekly food expenditure is Rs. 250 with a standard deviation of Rs 40. For 400 women shoppers chosen at random in super market B in another locality of the city, the average weekly food expenditure is Rs. 220 with a standard deviation of Rs. 55. Test at 1 % level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.
3. The following table presents data on the values of harvested crop stored in a open and inside a godown:

	Sample size	Mean	$\sum (x - \bar{x})^2$
Outside	40	117	8.685
Inside	100	132	27.315

Assuming that the two samples are random and they have been drawn from normal distribution with equal variances, examine the mean value of a harvested crop affected by weather conditions.

4. The mean height of 50 male students who showed above average participation in college athletics was 68.2 inches with standard deviation of 2.5 inches; while 50 male students who showed no interest in such participation had mean height of 67.5 inches with a standard deviation of 2.8 inches.

Test the hypothesis that male students who participate in college athletics are taller than other male students.

Aim: To test of significance for Difference of Means (you have to write aim as per the problem)

Formula and Procedure:

Let X and Y follows Normal distribution of parameters (μ_1, σ_1^2) and (μ_2, σ_2^2) respectively.

Our interest is to test the null hypothesis H_0

H_0 : There is no significant difference between the sample means

i.e., $H_0 : \mu_1 = \mu_2$ (or sample means)

Against the alternative hypothesis H_1

H_1 : There is a significant difference between the sample means

i.e., $\mu_1 \neq \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$

Suppose the X_1, X_2, \dots, X_{n_1} is a random sample of size n_1 drawn from the above population X and $Y_1, Y_2, Y_3, \dots, Y_{n_2}$ another random sample of size n_2 . Since sample sizes are large,

$$\bar{x} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$y \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

and also

Also, $(\bar{x} - \bar{y})$ being the difference of two independent normal variates also a normal variate. The standard normal variate Z corresponding to $(\bar{x} - \bar{y})$ is given by

$$Z = \frac{(\bar{x} - \bar{y}) - E(\bar{x} - \bar{y})}{S.E.(\bar{x} - \bar{y})} \sim N(0, 1)$$

Under the null Hypothesis,

$$Z = \frac{(\bar{x} - \bar{y})}{S.E.(\bar{x} - \bar{y})} \sim N(0, 1)$$

Case (i) : If the population variances known to us, then test statistic under null hypothesis,

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Case (ii) : If Population variances are not known to us i.e., if the sample have been drawn from the population with common standard deviation σ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then under null hypothesis

$$Z = \frac{(\bar{x} - \bar{y})}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Case (iii) : If Population variances are not known to us $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then the estimates of the population variances are used. If the sample sizes are not sufficiently large then the unbiased estimate of σ^2 is given by

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

Then under null hypothesis,

$$Z = \frac{(\bar{x} - \bar{y})}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Case (ii) : However, if $\sigma_1^2 \neq \sigma_2^2$ and σ_1 and σ_2 are not known, then they are estimated from the sample values.

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

Calculation :

Conclusion :

Practical Exercise -3

Test of significance for Single Proportion

1. A manufacturer of scientific calculators claims that 65 percent of the students use calculators of this company. In a sample of 34, there are 19 students using the calculators of that company. Do you accept the manufacturer's claim at 1% level of significance?
2. Past experience reveals that 6% of the items produced by a machine are defective. In a sample of size 34 there are 5 defective items found. Can we claim that there is a deterioration of machine performance?
3. In a sample of 1000 people in Maharashtra, 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat eaters are equally popular in this state at 1% level of significance.
4. 40 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease is 85% in favour of the hypothesis that it is more, at 5% level.

Aim:

To test whether the population proportion value is equal to a specified value or not.

Formula and Procedure:

If X is the number of successes in n independent trials with constant probability P of success for each trial. Consider a random sample of size n and there are x successes and x follows binomial distribution with parameter n and p . Then

$$E(X) = nP \text{ and } Var(X) = nPQ$$

Where $Q=1-P$, is the probability of failure

Null Hypothesis : The population Proportion is equal to given value

$$\text{i.e., } H_0 : P = P_0$$

Alternative Hypothesis: The population Proportion is not equal to the given value.

$$\text{i.e., } H_1 : P \neq P_0 \text{ or } P < P_0 \text{ or } P > P_0$$

Let consider sample proportion $p = \frac{X}{n}$ then $E(p) = P$ and $Var(p) = \frac{PQ}{n}$

$$\text{Hence for large } n, \quad p \sim N\left(P, \frac{PQ}{n}\right)$$

$$Z = \frac{p - E(p)}{\sqrt{var(p)}} \sim N(0, 1)$$

Under the null hypothesis,

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0, 1) \text{ or } \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} \sim N(0, 1)$$

If the alternative hypothesis is two sided,

$$|z_{cal}| \leq z_{\alpha/2} \quad \text{Do not reject nul hypothesis, otherwise reject}$$

If the alternative hypothesis is one / single sided,

$$|z_{cal}| \leq z_{\alpha} \quad \text{Do not reject nul hypothesis, otherwise reject}$$

Calculation:

Conclusion:

Practical Exercise -4
Test of significance for difference of Proportions

1. Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that the proportions of men and women in favour of the proposal, are same against that they are not at 5% level.
2. In a simple random sample of 600 men taken from big city 400 are found to be smokers. In another sample of 900 men taken from another city 450 are smokers. Do the data indicate that there is a significant difference in the habit of smoking in the two cities?
3. Before the increase in excise duty on tea 400 people out of a sample of 500 persons were found to be tea drinkers. After an increase in excise duty 400 persons were found to be tea drinkers in a sample of 600 people. Do you think that there has been a significant decrease in the consumption of tea after the increase in the excise duty?
4. 500 units from a factory are inspected and 12 are found to be defective. 800 units from another factory are inspected and 12 are found to be defective. Can it be concluded at a 5% level of significance that the production at the second factory is better than in the first factory?

Aim:

Formula and Procedure:

Let X_1 and X_2 be the number of persons possessing the given attribute A in a random sample of sizes n_1 and n_2 from the two populations respectively. Then sample proportions are

$$p_1 = \frac{X_1}{n_1} \text{ and } p_2 = \frac{X_2}{n_2}$$

If P_1 and P_2 are the Population proportions then,

$$E(p_1) = P_1, E(p_2) = P_2$$

$$\text{and } \text{Var}(p_1) = \frac{P_1 Q_1}{n_1} \text{ and } \text{Var}(p_2) = \frac{P_2 Q_2}{n_2}$$

Null Hypothesis: There is no significant difference between the two population proportions.

$$\text{i.e., } H_0: P_1 = P_2$$

Alternative Hypothesis: There is a significant difference between the two population proportions.

$$\text{i.e., } H_1: P_1 \neq P_2 \text{ or } P_1 < P_2 \text{ or } P_1 > P_2$$

Now the test statistic for the difference of two proportions,

$$Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\text{Var}(p_1 - p_2)} \sim N(0, 1)$$

Under the null hypothesis the test statistic becomes,

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0, 1)$$

Case (i): If $P_1 = P_2 = P$ (say) and $Q_1 = Q_2 = Q$ (say) then the test statistic becomes,

$$Z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)$$

Case(ii): In general, we do not have any information as to the proportions of A's in the populations from which the samples have been drawn. The unbiased estimate of population proportion is

$$\widehat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Under the null hypothesis, the test statistic becomes

$$Z = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

If the alternative hypothesis is two sided,

$$|z_{cal}| \leq z_{\alpha/2} \quad \text{Do not reject nul hypothesis, otherwise reject}$$

If the alternative hypothesis is one / single sided,

$$|z_{cal}| \leq z_{\alpha} \quad \text{Do not reject nul hypothesis, otherwise reject}$$

Conclusion:

Practical Exercise -5
Test for significance of Single Standard Deviation

1. A random sample of size 44 was taken from a normal population. The sample standard deviation was found to be 3.07. Test the hypothesis that the sample came from a normal population with standard deviation 3.
2. A random sample of size 38 was taken from a normal population. The sum of squares of the observations from their arithmetic mean was found to be 985. Test the hypothesis that the population standard deviation is 5.
3. A random sample of size 50 has standard deviation 11.8 drawn from the normal population. Can we assume that the sample has been drawn from the population with standard deviation 10.
4. A random sample of 200 items, drawn from a population with standard deviation with 0.8, the sample standard deviation is 0.7. Can we conclude that the sample SD is less than the population S.D at 1% level of significance?

Practical Exercise -6

Test for significance of difference of two Standard Deviations

1. Two independent random samples of sizes 37 and 44 were taken from two independent normal distributions. The sample standard deviations were found to be 2.53 and 2.75. Test the hypothesis that the standard deviations of the two normal populations are equal or not.
2. Two independent random samples of sizes 40 and 48 were taken from two independent normal distributions. The sum of squares of deviations from their arithmetic means were found to be 4620 and 5160. Can we assume that variability is more in the second population.
3. Random samples drawn from two countries gave the following details relating to the heights of adult males. Test is there any difference between the standard deviations significant.

	Country A	Country B
No.of Samples	1000	1200
Standard deviation in inches	2.58	2.5

4. Two random samples of sizes 100 each have drawn from two populations with the standard deviations 2.823 and 1.548. Test the significant difference between the sample standard deviations, if the population standard deviation is 2.

Practical Exercise -7

TEST FOR THE SIGNIFICANCE OF CORRELATION COEFFICIENT

1.

2.

3. The correlation coefficient of Physics and Statistics of 45 boys is 0.45 and the same of 39 girls in a class is 0.38. Test the significant difference between the correlation coefficients.

4. The correlation coefficient of a bivariate sample of size 50 is 0.36. Can this be regarded as drawn from the normal population with the correlation coefficient 0.5?

Small Sample Tests

Practical Exercise -8

Test for significance of Single Mean and Two Means

1. A manufacturer claims that a special type of projector bulb has an average life of 160 hours. To check this claim an investigator takes a sample of 20 such bulbs, puts them on the test, and obtains an average life of 167 hours with standard deviation 16 hours. Assuming that the lifetime of such bulbs follows normal distribution, does the investigator accept the manufacturer's claim at 5% level of significance?
2. The mean share price of companies in the Pharma sector is Rs.70. The share prices of all companies changed time to time. After a month, a sample of 10 Pharma companies was taken and their share prices were noted as below:
70, 76, 75, 69, 70, 72, 68, 65, 75, 72
Assuming that the distribution of share prices follows normal distribution, test whether mean share price is still the same at 1% level of significance?
3. In a random sample of 10 pigs fed by diet A, the gain in weights (in pounds) in a certain period were
12, 8, 14, 16, 13, 12, 8, 14, 10, 9
In another random sample of 10 pigs fed by diet B, the gain in weights (in pounds) in the same period were
14, 13, 12, 15, 16, 14, 18, 17, 21, 15
Assuming that gain in the weights due to both foods follows normal distributions with equal variances, test whether diets A and B differ significantly regarding their effect on increase in weight at 5% level of significance.
4. The means of two random samples of sizes 10 and 8 drawn from two normal populations are 210.40 and 208.92 respectively. The sum of squares of the deviations from their means is 26.94 and 24.50 respectively. Assuming that the populations are normal with equal variances, can samples be considered to have been drawn from normal populations having equal mean.

Practical Exercise -9

Test for significance of difference between Independent sample means

1. Below are the weights (In lbs) of pigs fed on two diets A and B. Test if two diets differ significantly as regard their effect on their increase in weights due to diet A and B.

+	Diet A	25	32	30	34	24	14	32	24	30	31	35	25						
	Diet B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22			

2. The heights of six randomly selected sailors are in inches: **63, 65, 68, 69, 71, 72**. Those of 10 randomly chosen soldiers are **61, 62, 65, 66, 69, 69, 70, 71, 72, 73**. Discuss the light that these data throw on the suggestion that sailors are on the average taller than soldiers.
3. Ten soldiers visit the rifle range for two consecutive weeks. For the First week their scores are: **67, 34, 57, 55, 63, 54, 56, 68, 33, 43** And during the second week their scores in the same order are: **70, 38, 58, 58, 56, 67, 68, 72, 42, 38**. Examine if there is any significant difference in their performance.
4. Two independent groups of 10 children were tested to find how many digits they could repeat from memory after hearing them the results are as follows:

Group A	8	6	5	7	6	8	7	4	5	6
Group B	10	6	7	8	6	9	7	6	7	7

Practical Exercise -10

Test for significance of difference between Dependent sample means (Paired t -test)

1. GVP college boys were given a test in Statistics. They were given a month's tuition and a second test was held at the end of it. Do the marks give evidence that the students have benefited by the extra coaching?

Marks in First Test	23	20	19	21	18	20	18	17	23	16	19
Marks in Second Test	24	19	22	18	20	22	20	20	23	20	18

2. The following table gives the additional hours of sleep gained by 10 patients in an experiment to test the effect of a drug. Does this data give evidence that the drug provides additional hours of sleep?

Patients	1	2	3	4	5	6	7	8	9	10
Hours gained	0.7	0.1	0.2	1.2	0.31	0.4	3.7	0.8	3.8	2.0

3. In a certain experiment to test the stimulus and the following results of increase in weight were obtained in pigs. Test whether the stimulus helps to increase the weight

Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	53	50	54	54	53

4. A certain stimulus administered to each of the 12 patients results in the following increase in blood pressure:

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

Can it be concluded that the stimulus will, in general, be accompanied by an increase in blood pressure?

Practical Exercise -11
Chi-square test for goodness of fit

1. The following figures show the distribution of digits in numbers chosen at random from the telephone directory: Test whether the digits are equally distributed or not.

Digits	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1078	933	1107	972	964	853	10000

2. The following table gives the number of aircraft accidents that occur during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Digits	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Frequency	14	16	8	12	11	9	14

3. Fit a poisson distribution to the following data and test the goodness of fit.

x	0	1	2	3	4	5	6
f	275	72	30	7	5	2	1

4. Fit a binomial distribution for the following data and test the goodness of fit.

x	0	1	2	3	4	5	6
f	5	18	28	12	7	6	4

Practical Exercise -12
Chi-square test for Independence of Attributes

- Two sample polls of votes for two candidates A and B for a public office are taken. One from the residents of rural areas. The results are given in a table. Examine the nature of the area is related to voting preference in this election.

Votes for Area	A	B	Total
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

- Two researchers adopted two different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence levels. The results are as follows:

Researcher	No of Students in each level				Total
	Below Average	Average	Above Average	Genius	
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are significantly different?

3. The following table gives for a sample of married women, the level of education and marriage adjustment score:

Level of Education		Marriage adjustment score			
		Very low	Low	High	Very high
	College	24	97	62	58
	High School	22	28	30	41
	Total	32	10	11	20

Can you conclude from the above, the higher the level of education, the greater the degree of adjustment in marriage?

Practical Exercise -13
Test for Equality of Two Variances (F-test)

1. Two random samples gave the following results:

sample	size	sample mean	sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples came from the normal population with the same variance at 5% level of significance.

2. In one sample of 8 observations, the sum of squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.
3. Two random samples of sizes 10 and 12 are drawn from the normal populations and given the following data. Test the equality of population variances.

Sample I	10	6	16	17	13	12	8	15	9	14		
Sample II	7	13	22	15	12	14	18	8	21	22	10	7

Practical Exercise -14
Nonparametric Tests for Single Sample
(Sign, Run and Wilcoxon Signed Rank Tests)

1. The breaking strength (in pounds) of a random sample of 10 ropes made by a manufacturer is given by 163 165 165 160 171 158 151 162 169 172 Use the sign test to test the manufacturer's claim that the average breaking strength of a rope is greater than 160 pounds at 5% level of significance.

2. Use appropriate non parametric test procedure to test for the randomness of the following 30 two digit numbers

15	17	01	65	69	69	58	41	81	16	16	20	00	84	22
28	26	46	66	36	86	66	17	34	49	85	45	51	40	10

3. A random sample drawn from the continuous distribution. Test the hypothesis that the median of the distribution is 160 or not using wilcoxon signed rank test.

163	165	159	160	188	162	175	158	150	169	162	164	139	177	166
149	161	175	168	188	177									

Practical Exercise -15
Nonparametric Tests for Related Samples
(Sign and Wilcoxon Signed Rank Tests)

1. From a college of 30 students are drawn at random and two talented tests were conducted and the scores are given below. Can we assume that their talents are the same in two tests? (Use Sign test)

Test	133	146	136	172	141	106	159	141	142	140	174	156	82	138	154
I	163	175	150	134	134	160	133	149	129	140	162	115	104	121	181
Test	141	151	99	145	179	161	168	151	132	180	163	133	158	180	204
II	157	215	145	167	137	136	170	157	168	188	188	160	146	123	177

2. The following data relates to the noise levels of two busy junctions during the busy hours and recorded. Using the Wilcoxon signed rank test whether the noise levels are the same at both places or not.

Sample A	69	74	77	59	80	59	71	65	62	79	76	68	62	60
Sample B	59	78	53	63	67	63	59	58	64	74	66	64	63	62

Practical Exercise -16

Nonparametric Tests for Independent Samples

(Median test, Mann Whitney U -Test, Wald Wolfowitz Run Test)

1. Two independent random samples are taken from the two continuous distributions and are given below. Using median test, test the hypothesis that two continuous distributions are identical or not

Sample 1	61	80	83	37	70	98	18			
Sample 2	89	50	47	76	60	55	65	40	98	84

2. A test was conducted on a group of students. The score marks obtained by boys and girls are as follows. Using Mann Whitney test, test the hypothesis that there is no significant difference between boys and girls in their performances.

Boys	13	12	19	20	20	15	12	10	18	19	15				
Girls	20	19	11	10	15	15	12	20	12	10	19	19	12	14	16

3. Two samples are drawn from two continuous distributions. Use Wald Wolfowitz Run Test and test whether two distributions are equal or not.

First Group	227	176	253	149	16	55	234	194	247	92	184	147	88	161	171
Second Group	202	14	165	171	292	271	151	235	147	99	63	284	53	228	271