Chapter 19
NONPARAMETRIC TESTS

19.1. Introduction

Traditionally, statistical inference deals with two types of problems. They are (i) problems of estimation and (ii) Test of hypothesis. Problems of estimation were briefly discussed in section 15.10 of Chapter 15.

Broadly speaking, there are two types of tests. They are

1) Parametric test and
2) Nonparametric test.

Most widely used parametric tests were discussed in Chapter-16. All the parametric tests are based on normal distribution, assuming either that the samples come from normal population or that they are large enough to justify normal approximation. Some nonparametric alternatives to those parametric tests, which do not require knowledge about the population or populations from which the samples are obtained will be taken up in this chapter. Most of the parametric tests have two features in common.

i) The form of the parent population from which the samples have been drawn is assumed to be known, and

ii) They were concerned with testing statistical hypothesis about the parameters of the population or estimating its parameter.

For example, almost all the exact (small) sample tests of significance are based on the fundamental assumptions that the parent population is normal and are concerned with testing or estimating the means and variances of the populations. Such tests, which deal with the parameters of the population, are known as parametric test.

Definition. A parametric test is a test whose model specifies certain conditions about the parameters of the population from which the samples are drawn.

Definition. A nonparametric test is a test that does not depend on the particular form of the basic population from which the samples are drawn. That is, nonparametric test does not make any assumption regarding the form of the population from which the samples are drawn.

Assumptions associated with nonparametric tests are (i) Sample observations are independent; (ii) The variable under study is continuous; (iii) The probability density function is continuous; and (iv) Lower order moments exist.

Obviously, these assumptions are fewer and much weaker than those associated with parametric test.

19.2. Advantages and Disadvantages of Nonparametric Methods Over Parametric Methods

Advantages

i) Nonparametric methods are readily comprehensible, very simple and easy to apply and do not require complicated sample theory.

ii) Nonparametric technique will apply to the data which are mere classification (i.e., which are measured in nominal scale) while nonparametric methods exist to deal with such data. Nonparametric tests are found applications in Psychometric, Sociology and Educational statistics.

iii) Nonparametric tests are available to deal with the data which are given in ranks or whose seemingly numerical scores have strength of ranks. For instance, no parametric test can be applied if the scores are given in grades A^+, A, B^+, etc.

Disadvantages

i) If all the assumptions in the data are met, then the use of nonparametric tests wastes information.

ii) There is no nonparametric test for any interaction effects in the analysis of variance.

iii) The arithmetical calculation of these test procedures is simple but sometimes very laborious and tedious.

iv) For conclusions, large number different types of tables are required.

Remarks

- Since no assumption is made about parent distribution, the nonparametric methods are sometimes referred to as Distribution Free methods. These tests are based on the "Order Statistics" theory. In these tests we shall be using median, range, quartiles, inter-quartile range, etc., for which ordered sample is desired.
- The whole structure of the nonparametric methods rests on a simple but fundamental property of order statistic.

In many nonparametric tests the computational burden is so light that they come under the heading of quick-and-easy or shortcut techniques. Party for these reason, nonparametric tests have become very popular and extensive literature is devoted to this theory and applications.

19.3. Some Important Nonparametric Test Procedures

Some widely used nonparametric tests will be discussed in this section. There exists a nonparametric test for each parametric test. Usually, z or t test is used to test the mean of a population when a sample is taken from a normal population in case of small sample. But median is better than mean as a measure of central tendency in case of a skewed distribution. So, for a small sample when sample is not drawn from a normal population, median is better than mean to test as a measure of central tendency. The simple nonparametric test to test the median of a population is the sign test.

19.3.1. One sample Sign test. The sign test is often used as a nonparametric alternative to the one-sample Z or t test. It is the oldest of all nonparametric tests and it is so called as we usually convert the data for analysis to a series of plus and minus signs. The test statistic consists of either the number of plus or the number of minus signs.

Assumptions

i) The sample values $x_1, x_2, \ldots\ldots, x_n$ will be continuous.

ii) The observations may be considered as a random sample from a population with unknown median M.

iii) The variable of interest is measured at least on ordinal scale.

Hypotheses

i) Two tailed test $H_0: M = M_0,\ H_1: M \neq M_0$;

ii) Right tailed test $H_0: M = M_0,\ H_1: M > M_0$;

iii) Left tailed test $H_0: M = M_0,\ H_1: M < M_0$; and

iv) We select α as the level of significance.

- ***Test statistic.***

First we have to subtract each value of the sample from the median M_0 as defined in null hypothesis. We replace each sample value exceeding M_0 as plus sign and each value less than M_0 with a minus sign. That is for n sample values, we put $x_i > M_0 = +$, $x_i < M_0 = -$ for $I = 1, 2, ..., n$. But we put zero for $x_i = M_0$ $(I = 1, 2, ..., n.)$. In this case, we omit the observation from the sample. If the number of observations $x_i = M_0$ $(I=1, 2,....,n.)$ is k, then the valid number of observations in the sample is $n' = n-k$.

We reject the null hypothesis in (i) if we observe a sufficiently small number of plus or minus signs. In this case the test statistic is the number of plus or minus signs whichever is minimum. Suppose the number is S, then S is the test statistic. If the null hypothesis is true, then S is a binomial variate with parameters n and $p = 0.5$ So the test statistic S in (i) is the number of plus or minus signs whichever is minimum. If we observe a sufficiently small number of minus signs, we reject H_0 in (ii). So the test statistic S in case (ii) is the number of minus signs. If we observe a sufficiently small number of plus signs, we reject H_0 in (iii). So the test statistic S in case (iii) is the number of plus signs.

Note. If $x_i = M_0$, that item is omitted from the sample size as well as analysis.

Conclusion can be made as follows :

a) Reject H_0 in case (i) at α level of significance, if $P[S \leq s] \leq \frac{\alpha}{2}$. Here S is a binomial variate with parameters n and $p = 0.5$ and S is the number of plus or minus signs whichever is minimum.

b) Reject H_0 in case (ii) at α level of significance, if $P[S \leq s] \leq \alpha$. Here S is the number of minus signs.

c) Reject H_0 in case (iii) at α level of significance, if $P[S \leq s] \leq \alpha$. Here S is the number of plus signs.

Note. To perform a sign test when the sample size is very small, we refer directly to a table of binomial probabilities such as Table 9; when the sample size is large, we use the normal approximation to the binomial distribution. In that case the test statistic is $Z = \frac{S - 0.5n}{0.5\sqrt{n}}$; which follows standard normal distribution. Since under the null hypothesis mean equals to $np = 0.5n$ and standard deviation $= \sqrt{np(1-p)} = \sqrt{n}(0.5)$.

Example 19.1. The following data relate to the waiting time in minutes of 20 patients to meet a doctor for a particular disease.

17 25 26 35 32 20 24 9 25 12 18 27 15 35 23 21 28 20 29 16.

The doctor claims that on average no patients wait more than 20 minutes to meet him. Justify the claim of the doctors at 5% level of significant with the help of sign test.

Solution. We set up (i) null hypothesis $H_0: M \leq 20$, (ii) Alternative hypothesis $H_1: M > 20$, (iii) $\alpha = 0.05$, (iv) S = test statistic = Number of minus signs. (v) Put plus sign when x is greater than 20, minus sign when x is less than 20 and zero when $x = 20$. That is, when $x_i > 20 = +$, $x_i < 20 = -$ and zero when $x_i = 20$. The calculation table will take the following form.

Values of x:	17	25	26	35	32	20	24	9	25	12	18	27	15	25	23	21	28	20	29	16
Sign of $x - 20$:	–	+	+	+	+	0	+	–	+	–	–	+	–	+	+	+	+	0	+	–

Here $s = 6$, $n' = n - k = 20 - 2 = 18$. $P[S < 6 \mid 18, 0.5] = 0.119 > 0.05$.

Conclusion. The calculated probability is greater than 0.05. So, we have no reason to reject the null hypothesis. This means we may accept the claim of the doctor.

Example 19.2. The following are the measurements of the breaking strength of a certain kind of inch cotton ribbon in pounds

163 165 160 189 161 171 158 151 169 162 163 139 172 165 148 166 172 163 187 173

Use the sign test to test the null hypothesis $M = 160$ against the alternative hypothesis $M > 160$ at 0.05 level of significance.

Solution.

Sign of $x - 160$:																			
+	+	0	+	+	+	–	–	+	+	+	–	+	+	–	+	+	+	+	+

Here $s = 4$, $n' = n - k = 20 - 1 = 19$. $P[S < 4 \mid 19, 0.5] = 0.01 > 0.025 = \frac{\alpha}{2}$.

Conclusion. Here we have two sided alternatives. The calculated probability 0.01 is less 0.025, we may reject the null hypothesis at 5% level of significance and we conclude that the breaking strength of the given kind of ribbon is not equal to 160 pounds.

19.3.1a. The Sign test for large sample. For large sample when the value of p is not so small. Then binomial distribution tends to normal distribution. For testing median of a distribution, we take $p = 0.5$ as null hypothesis. The case in which n is greater than 10, then the value of np and nq are greater than 5. As a result, $n > 10$, binomial distribution reduces to normal distribution. In *such case* we can safely convert S into standard normal variate Z and Z can be used as a test statistic. The test statistic is

$$Z = \frac{S - np}{\sqrt{np(1-p)}} = \frac{S - 0.5n}{0.5\sqrt{n}}; \quad \text{Here } p = 0.5.$$

Example 19.3. The following data, in tons, are the amounts of sulfur oxides emitted by a large industrial plant in 40 days.

17 15 20 29 19 18 22 25 27 9 24 20 17 6 24 14 15 23 24 26 19 23
28 19 16 22 24 17 20 13 19 10 23 18 31 13 20 17 24 14.

Use the sign test to test the null hypothesis $M = 21$ against the alternative hypothesis $M < 21$ at the 0.01 level of significance.

Solution. We set up

i) Null hypothesis $H_0: M = 21$, ii) Alternative hypothesis $H_1: M < 21$, iii) $\alpha = 0.01$.

iv) Test statistic, Z is defined by $Z = \dfrac{S - 0.5n}{0.5\sqrt{n}}$.

v) Reject the null hypothesis if $Z = -z_{.01} = -2.33$.

Now we put $x_i > 21 = +$, $x_i < 21 = -$ and zero when $x_i = 21$ for all the 40 observations, we get the following signs.

−, −, −, +, −, −, +, +, +, −, +, −, −, −, +, −, −, +, +, +, −,
+, +, −, −, +, +, −, −, −, −, −, +, −, +, −, −, −, +, −.

Here number of plus signs is 16 and the number of minus signs is 24. The value of the test statistic is

$$Z = \frac{S - 0.5n}{0.5\sqrt{n}} = \frac{16 - 0.5 \times 40}{0.5\sqrt{40}} = \frac{-4}{3.16} = -1.26.$$

Conclusion. Since the calculated value of Z is greater than - 2.33, the null hypothesis cannot be rejected. That means the industrial plant on average emits daily 21 tons sulfur oxides may be accepted.

19.3.2. Paired sample sign test or the sign test for paired sample. The sign test can also be used when we deal with paired data. In such problem, each pair of sample values is replaced by a plus sign if the difference between the paired observations is positive and by a minus sign if the difference between the paired observations is negative, and it is discarded if the difference is zero. To test the null hypothesis that two continuous symmetrical populations have equal means or that two continuous populations have equal median even the populations are not symmetrical. We can use the sign test with these kinds of problems which is called paired sign test.

Example 19.4. To determine the effectiveness of a new traffic-control system, the number of accidents that occurred at 12 dangerous intersections during four weeks before and four weeks after the installation of the new system was observed, and the following data were obtained.

3 and 1, 5 and 2, 2 and 0, 3 and 2, 3 and 2, 3 and 0, 0 and 2, 4 and 3, 1 and 3, 6 and 4, 4 and 1, 1 and 0.

Use the paired-sample sign test at the 0.05 level of significance to test the null hypothesis that the new traffic- control system is only as effective as the old system.

Solution. Suppose M_1 and M_2 are the number of accidents before 4 weeks and after four weeks respectively. We set up

i) Null hypothesis $H_0: M_1 = M_2$, ii) Alternative hypothesis $H_1: M_1 > M_2$ and,

iii) $\alpha = 0.01$, iv) Use the test statistic S, the observed number of minus signs.

v) Replacing each positive difference by a plus sign and each negative difference by a minus sign, we get

+, +, +, +, +, +, +, +, −, +, −, +, +, +.

Here $n = 12$ and $S = 2$, From Table 9, we find that $P[S \le 2 \mid 12; 0.5] = 0.0192$.

Conclusion. The calculated probability, 0.091, is less than 0.05. The null hypothesis must be rejected, and we conclude that the new traffic-control system is effective in reducing the number of accidents at dangerous intersections.

Remark. If n is greater than 25, the normal approximation to binomial can be used as before.

Example 19.5. The following data relate the marks obtained by 10 students in two class test:

Class test-1 (x) :	19	15	25	19	17	13	23	21	13	25
Class test-2 (y) :	7	21	13	17	15	17	13	15	7	15

Use the paired-sample sign test to test the null hypothesis that the two tests have the same medians at the 0.05 level of significance.

Solution. We set up

i) Null hypothesis $H_0 : M_1 = M_2$, ii) Alternative hypothesis $H_1 : M_1 \neq M_2$,

iii) $\alpha = 0.05$, iv) Use the test statistic S, the observed number of minus signs.

v) Replacing each positive difference by a plus sign and each negative difference by a minus sign, we get

$$+,\ -,\ +,\ +,\ +,\ -,\ +,\ +,\ +,\ +.$$

Here $n = 10$ and $S = 2$. From Table 9, we find that $P[S < 2 \mid 10; 0.5] = 0.055 > 0.025 = \frac{\alpha}{2}$.

Conclusion. The calculated probability, 0.0.55, is greater than 0.025. The null hypothesis cannot be rejected. That is, two paired samples may have come from two populations which have the same median.

19.3.3. Wilcoxon Signed-Rank Test. As we saw before, the sign test is easy to perform, but since we utilize only the signs of the differences between the observations and M_0 in the one-sample case, or the signs of the differences between the pairs of observations in the paired-sample case, it tends to be wasteful of information. An alternative nonparametric test, the Wilcoxon sign-rank test, is less wasteful in that it takes into account also the magnitude of the differences. In this test, we rank the differences without regard to their signs, assigning rank 1 to the smallest difference in absolute value, rank 2 to the second smallest difference in absolute value ,...., and rank n to the largest difference in absolute value. Zero differences are again discarded, and if the absolute values of two or more differences are the same, we assign each one the mean of the ranks which they jointly occupy. Then the signed-rank test is based on T^+, the sum of the ranks assigned to the positive differences, T^-, the sum of the ranks assigned to the negative differences. Since the population considered here is symmetric, the conclusion about the population median may be applied to the population mean.

Assumptions. (i) The sample values $x_1, x_2, \ldots\ldots, x_n$ of size n is a random sample from a population with unknown median M; (ii) The variable of interest is continuous; (iii) The variable of interest is measured on at least interval scale; (iv) The population from the sample is drawn is symmetric and (v) The observations are independent.

Hypotheses. a) Two sided test $H_0: M = M_0$, $H_1: M \neq M_0$; b) Right tailed test $H_0: M = M_0$, $H_1: M < M_0$; c) Left tailed test $H_0: M = M_0$, $H_1: M > M_0$.

We select as α the level of significance.

Test statistic. To obtain the test statistic, we follow the following steps

i) Subtract the median value given in the null hypothesis from each observation i.e. find $D_i = x_i - M_0$.. If $x_i = M_0$ (I = 1, 2, ..., n) eliminate these values from calculations and thus the sample size is reduced.

ii) Find the absolute values of D_i i.e. $|D_i|$.

iii) Rank the $|D_i|$ values from the smallest to the largest. If two or more $|D_i|$'s are equal, assign each tied value the mean of the rank positions occupied by the differences that are tied. For Example, if the three smallest differences are all equal, rank them, 1, 2 and 3 but assign each a rank of (1+2+3)/3 = 2.

iv) The sign of the differences are attached to the corresponding ranks.

v) Obtain the sum of the rank values with + signs and call it T^+ and also obtain the sum of the rank with - signs and call it T^-. If any one of T^+ or T^- is obtained the other can be calculated from the relation $T^+ + T^- = n(n+1)/2$, then $T^+ = \{n(n+1)\}/2 - T^-$.

vi) For case (a); the test statistic T which is the minimum value of either T^+ or T^-.That is $T = \min(T^+, T^-)$. Reject H_0 at the α level of significance if $T \leq T_\alpha$ where value of T_α is obtained from the Table 10 in the appendix.

vii) For case (b); the test statistic is T^-. Reject H_0 at the α level of significance if $T^- \leq T_{2\alpha}$ where value of $T_{2\alpha}$ is obtained from the Table 10 in the appendix.

viii) For case (c); the test statistic is T^+. Reject H_0 at the α level of significance if $T^+ \leq T_{2\alpha}$ where value of $T_{2\alpha}$ is obtained from the Table 10 in the appendix.

It is noted that the test statistic will be T, T^- and T^+ according to the alternative hypothesis. We have to be careful, though, to use the right statistic and the right critical value, as summarized in the following table, where in each case the level of significance is α:

Alternative hypothesis	Reject the null hypothesis if
$H_1: M \neq M_0$	$T \leq T_\alpha$
$H_1: M > M_0$	$T^- \leq T_{2\alpha}$
$H_1: M < M_0$	$T^+ \leq T_{2\alpha}$

The Critical values in the right-hand column of this table, T_α or $T_{2\alpha}$ are the largest values for which the corresponding p-values do not exceed α or 2α, respectively. They may be obtained from the Table 10 for values of n not exceeding 25. Note that the same critical values can serve for tests at different levels of significance depending on whether the alternative hypothesis is one -sided or two sided. For instance, $T_{0.02}$ can serve as a critical value at the 0.02 level of significance when the alternative hypothesis is two-sided and at the 0.01 level of significance when the alternative hypothesis is one-sided. This may seem confusing, but it is how these critical values are tabulated in some texts.

Example 19.6. The following are 15 observations of the octane rating of a certain kind of gasoline.

97.5, 95.2, 97.3, 6.0, 96.8, 100.3, 97.4, 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5 and 94.9.

Use the Wilcox on signed-rank test at the 0.05 level of significance to test whether or not the mean octane rating of the given kind of gasoline is 98.5.

Solution.

1) $H_0: M = 98.5$, 2) $H_1: M \neq 98.5$, 3) $\alpha = 0.05$
4) Reject the null hypothesis if $T \leq T_{0.05}$ where $T_{0.05}$ must be read from the Table 10 for the appropriate value of n.
5) Subtracting 98.5 from each value and ranking the differences without regard to their sign, we get;

Observation: x	x − 98.5	|x−98.5|	Rank
97.5	−1.0	1.0	4
95.2	−3.3	3.3	12
97.3	−1.2	1.2	6
96.0	−2.5	2.5	10
96.8	−1.7	1.7	7
100.3	1.8	1.8	8
97.4	−1.1	1.1	5
95.3	−3.2	3.2	11
93.2	−5.3	5.3	14
99.1	0.6	0.6	2
96.1	−2.4	2.4	9
97.6	−0.9	0.9	3
98.2	−0.3	0.3	1
98.5	0.0	0.0	
94.9	−3.6	3.6	13

So that $T^- = 4 + 12 + 6 + 10 + 7 + 5 + 11 + 14 + 9 + 3 + 1 + 13 = 95$,

$T^+ = 8 + 2 = 10$ and $T = 10$. From the Table 10; we find that $T_{0.05} = 21$ for $n = 14$.

Conclusion. Since $T = 10$ is less than $T_{0.05} = 21$, the null hypothesis must be rejected; then mean octane rating of the given kind of gasoline is not 98.5. Wilcoxon sign-rank test can also be used for paired-sample data. In that case, we test the null hypothesis $M_1 = M_2$ with the alternative hypotheses $M_1 \neq M_2$, $M_1 > M_2$ or, $M_1 < M_2$.

Example 19.7. The following data relate the scores obtained by 10 students in two class tests.

Class test-1 (x):	16	12	22	16	14	11	20	18	10	22
Class test-2 (y):	4	18	10	14	12	14	10	12	4	12

Use the sign-rank test to test the null hypothesis that the two tests have the same medians at the 0.05 level of significance.

Solution. We set up;

i) Null hypothesis $H_0: M_1 = M_2$, ii) Alternative hypothesis $H_1: M_1 \neq M_2$, iii) $\alpha = 0.05$,

iv) Reject the null hypothesis if $T \leq T_{0.05}$, where $T_{0.05}$ must be read from the Table 10 for the appropriate value of n.

v) Subtracting x from y for each pair values and ranking the differences without regard to their sign, we get

x	y	D = |x−y|	|D|	Rank of |D|	R (+)	R(−)
16	4	12	12	9.5	9.5	
12	18	−6	6	5		5
22	10	12	12	9.5	9.5	
16	14	2	2	1.5	1.5	
14	12	2	2	1.5	1.5	
10	14	−4	4	3		3
20	10	10	10	7.5	7.5	
18	12	6	6	5	5	
10	4	6	6	5	5	
22	12	10	10	7.5	7.5	

So that $T^- = 5 + 3 = 8$, $T^+ = 9.5 + 9.5 + 1.5 + 1.5 + 7.5 + 5 + 5 + 7.5 = 47$, and $T = \min(8, 47) = 8$. From the Table 10; we find that $T_{0.05} = 8$ for $n = 10$.

Conclusion. Since $T = 8$ is equal to $T_{0.05} = 8$, the null hypothesis may be rejected; the median scores of the two tests are different at the 5% level of significance.

19.3.3(a). Large Sample Approximation. When the sample size, n is greater than 25 we cannot use Table 10 for finding the critical value of T. In practice for $n > 14$, it is considered reasonable to assume that T^+ is a value of a random variable having approximately a normal distribution. To perform the signed-rank test based on the assumption, we need the following results, which apply regardless of whether the null hypothesis is $M = M_0$ or $M_1 = M_2$.

Theorem 19.1. Under the assumptions required by the signed-rank test, T^+ is a value of a random variable with the mean

$$\mu = \frac{n(n+1)}{4}, \text{ variance } \sigma^2 = \frac{n(n+1(2n+1)}{24}$$

then $Z = \dfrac{T^+ - \mu}{\sigma}$ is a standard normal variate. This Z is used as a test statistics to test the null hypothesis.

Conclusion. For two-sided test, Z has the critical value ± 1.96 at $\alpha = 0.05$ and ± 2.58 at $\alpha = 0.01$. Therefore the conclusion can be made as usual.

Example 19.8. The following are weights in kilograms, before and after, of 16 persons who stayed on a certain reducing diet for four weeks :

Before diet (x) : 147.0 183.5 232.1 161.6 197.5 206.3 177.0 215.4 147.7 208.1 166.8 131.9 150.3 197.2

After diet (y) : 137.9 176.2 219.0 163.8 193.5 201.4 180.6 203.2 149.0 195.4 158.5 134.4 149.3 189.1

Before diet (x) : 159.8 171.7

After diet (y) : 159.1 173.2

Use the signed-rank test to test at the 0.05 level of significance whether the weights reducing diet is effective.

Solution. 1) $H_0 : M_1 = M_2$, 2) $H_1: M_1 > M_2$, 3) $\alpha = 0.05$,

4) Reject the null hypothesis if $Z > Z_{0.05} = 1.645$, where $Z = \frac{T^+ - \mu}{\sigma}$.

5) The differences between the respective pairs, their absolute values and ranks are :

D	9.1	7.3	13.1	−2.2	4.0	4.9	−3.6	12.2	−1.3	12.7	8.3	−2.5	1.0	8.1	0.7	−1.5
|D|	9.1	7.3	13.1	2.2	4.0	4.9	3.6	12.2	1.3	12.7	8.3	2.5	1.0	8.1	0.7	1.5
Rank	13	10	16	5	8	9	7	14	3	15	12	6	2	11	1	4

Thus $T^+ = 13 + 10 + 16 + 8 + 9 + 14 + 15 + 12 + 2 + 11 + 1 = 111$.

Here $\mu = \frac{n(n+1)}{4} = \frac{16\times17}{4} = 68$ and variance $(\sigma^2) = \frac{n(n+1(2n+1)}{24} = \frac{16\times17\times33}{24} = 374$

We get $Z = \frac{T^+ - \mu}{\sigma} = \frac{111-68}{\sqrt{374}} = 2.22$

Conclusion. Since $z = 2.22$ exceeds $z_{0.05} = 1.645$, the null hypothesis must be rejected; we conclude that the diet is , indeed, effective in reducing weight.

19.3.4. Run Test. In many situations we are interest to know whether the data collected for statistical analysis is random or not. In all most all statistical inference, we assume that the sample drawn from the population is random. If the randomness of a sample is suspected, we test its randomness before going to statistical analysis in detail. There are many situations where we may investigate the assumption of randomness. The data of statistical quality control analysis regression analysis, estimation and test of hypothesis are some of the important situations in which we must be sure about randomness.

Procedures for investigating the randomness are based on the number and nature of the runs present in a data of interest. A run is defined as a sequence of like events or items or symbols of the sample that is preceded and followed by an event, item or symbol of a different type. The number of items or events in a run is called the length of the run. We doubt on randomness of a sample if there are too many or too few runs. For example, consider the following arrangement of defective, d, and no defective, n, pieces produced in a given order by a certain machine

nnnnn dddd nnnnnnnnnn dd nn dddd n dd nn.

We find that there is first a run of five n's, then a run of four d's, then a run of ten n's,........ and finally a run of two n's; in all there are nine runs of varying lengths. The total number of runs appearing in an arrangement of this kind is often a good indication of a possible lack of randomness. If there are few runs, we might suspect a definite grouping or clustering, or perhaps a trend; if there are too many runs, we might suspect some sort of repeated alternating pattern. In our example, there seems to be a definite clustering, the defective pieces seem to come in groups, but it remains to be seen whether this is significant or whether it can be attributed to chance. The following procedure will give us a guideline to decide whether the sequence of events or items in a sample is the result of a random process.

Assumption. The data given for statistical analysis consist of a sequence of observations recorded in the order of their occurrence which can be classified into two mutually exclusive types. Let us consider n = the sample size, n_1 = number of observations in one type and n_2 = number of observations the other type.

Hypotheses. Two sided H_0 : The pattern of occurrence of the two types of observations is determined by a random process; H_1 = The pattern of occurrence is not random.

(a) Two sided H_0 : The pattern of occurrence of the two types of observations is determined by a random process; H_1 = The pattern of occurrence is not random.

(b) Left tailed test H_0 : The pattern of occurrence of the two types of observations is determined by a random process; H_1 = The pattern of occurrence is not random because there are too few runs to be attributed to chance.

(c) Right tailed test H_0 : The pattern of occurrence of the two types of observations is determined by a random process; H_1 = The pattern of occurrence is not random because there are too many runs to be attributed to chance.

Test Statistic. The test statistic is r, the total number of runs. Conclusion can be made as follows.

(a) Reject H_0, if r is less than or equal to the lower critical value obtained from Table 11 or greater than or equal to the upper critical value obtained from the Table 12.

(b) Enter Table 11 with n_1 and n_2. If r is less than or equal to the tabulated value of the test statistic, reject H_0 at 0.025 level of significance.

(c) Enter Table 12 with n_1 and n_2. If r is greater than or equal to the tabulated value of the test statistic, reject H_0 at 0.025 level of significance.

Note. Table 11 and Table 12 give lower and upper critical values respectively of the test statistic for the 0.05 level of significance and values of n_1 and n_2 through 20.

Example 19.9. Checking on Mehgani trees that were planted few years ago along a road side of Chittagong University campus, a county official obtained the following arrangement of healthy, H, and diseased, D, trees.

HHHHH DD HH DDDD HHHH DDD.

Test at the 0.05 level of significance whether this arrangement may be regarded as random.

Solution. H_0 : Arrangement is random; H_1: Arrangement is not random; $\alpha = 0.05$.

Here in the sequence, first five observations HHHHH constitute the first run, the second, third, fourth, fifth and sixth runs are DD, HH, DDDD, HHHH, DDD respectively. The total number of runs in the sample is $r = 6$. The sample size is 20 in which $n_1 = 11$ and $n_2 = 9$. From Table 11 and 12 for $n_1 = 11$ and $n_2 = 9$ lower critical value is 6 and the upper critical value is 16.

Conclusion. Since the number of runs 6, is equal to the lower critical value of the test statistic, 6. We must reject the null hypothesis at 5% level of significance; the arrangement of healthy and diseased Mehgani trees is not random. It appears that the diseased trees come in cluster.

19.3.4(a). Large Sample Approximation. When either n_1 or n_2 is greater than 20, we cannot use Table 10 and Table 11 to test the above hypothesis. Even n_1 and n_2 are both greater than or equal to 10, it is considered reasonable to assume that the distribution of the random variable r is approximately normal with mean

$$\mu=\frac{2n_1n_2}{n_1+n_2}+1 \quad \text{and} \quad \text{variance}(\sigma^2)=\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)}$$

then, $Z=\frac{r-\mu}{\sigma}$ is the test statistic to test the null hypothesis.

Example 19.10. The following is an arrangement of men, M, and women, W, lined up to purchase tickets for a rock concert.

M W M W MMM W M W MMM WW MMMM WW M W
MM W MMM WWW M W MMM W M W MMMM WW M

Test for randomness at the 0.05 level of significance.

Solution.

1) H_0: Arrangement is random;
2) H_1: Arrangement is not random;
3) $\alpha=0.05$.
4) Reject the null hypothesis if $Z<-1.96$ or $Z>1.96$ where $Z=\frac{r-\mu}{\sigma}$.
5) Here $n_1=30$, $n_2=18$, and $r=27$, we get

$$\mu=\frac{2n_1n_2}{n_1+n_2}+1=\frac{2\times30\times18}{30+18}+1=23.5$$

and standard deviation,

$$(\sigma)=\sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2}{(n_1+n_2)^2(n_1+n_2-1)}}=\sqrt{\frac{2\times30\times18(2\times30\times18-30-18}{(30+18)^2(30+18-1)}}=3.21$$

and, hence the value of Z is: $z=\frac{27-23.5}{3.21}=1.09.$

Conclusion. Since $z=1.09$ falls between -1.96 and 1.96, the null hypothesis cannot be rejected; in other words, there is no real evidence to indicate that the arrangement is not random.

The method we have discussed in this section is not limited to tests of the randomness of series of attributes. Any sample which consists of numerical measurements or observations can be treated similarly by using letters a and b to denote, respectively, values falling above and below the median of the sample. Numbers equaling median are omitted. The resulting series of a's and b's can be tested for randomness on the basis of the total number of runs of a's and b's namely, the total number of runs above and below the median.

Example 19.11. The following are the speeds in km per hour at which every fifth passenger car was timed at a certain checkpoint of Dhaka-Chittagong high way.

41, 43, 62, 40, 54, 61, 48, 46, 66, 51, 70, 67, 56, 59, 60, 59, 58, 52, 46, 52, 76, 67, 65, 65, 69, 53, 62, 45, 52; 39, 73, 69, 66, 63, 70, 67, 62, 47, 56, 42, 59, 57, 72, 63m 63, 59, 63, 67, 77, 42.

Test the null hypothesis of randomness at the 0.05 level of significance.

...tion. 1) H_0: Arrangement is random; 2) H_1: Arrangement is not random;

3) $\alpha = 0.05$. 4) Reject the null hypothesis if $Z < -1.96$ or $Z > 1.96$ where $Z = \frac{r-\mu}{\sigma}$

where r is the number of runs above and below the median.

5) For finding median, we arrange the observations in ascending order of magnitude. The ordered observations are

39 40 41 42 42 43 45 46 46 47 48 51 52 52 52 53 54 56 56 57 58 59 59 59 59 60 61 62 62 62 63 63 63 63 65 65 66 66 67 67 67 67 69 69 70 70 72 73 76 77.

Here $n = 50$. The median is the mean of the 25th and 26th observations of the ordered speeds of the cars which is 59.5. Since the median of the speeds is 59.5, we get the following arrangement of a's and b's

bb a b aa bb a bbbb a b aaaaa bbbb a bb a bbb aaa b aaa bbbb aaaaaa

6) Here $n_1 = 25$, $n_2 = 25$, and $r = 20$, we get

$$\mu = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2\times25\times25}{25+25} + 1 = 26$$

and standard deviation,

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2\times25\times25(2\times25\times25-25-25}{(25+25^2(25+25-1)}} = \sqrt{122} = 3.493$$

and, hence the value of Z is: $z = \frac{20-26}{3.49} = -1.72$.

...clusion. Since $z = -1.72$ falls between -1.96 and 1.96, the null hypothesis cannot be rejected; ...her words, there is no real evidence to indicate that the arrangement is not random.

...mple 19.12. The following is an arrangement of men, M and women, W. lined up to purchase ...ets for Subarea Express for traveling from Dhaka to Chittagong.

M FF MMM F M F MM FF MMM FF M F MMM F MMM Ff M FFF MMMM Ff MM F

Test for randomness at the 0.05 level of significance.

...ution. 1) H_0 : Arrangement is random; 2) Arrangement is not random; 3) $\alpha = 0.05$

4) Reject the null hypothesis if $Z \le -1.96$ *or* $Z \ge 1.96$; where $Z = \frac{r-\mu}{\sigma}$.

5) Here r = number of runs = 22, $n_1 = 24$, $n_2 = 18$, we get

$$\mu = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2\times24\times18}{24+28} + 1 = 21.57$$

and standard deviation,

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2\times24\times18(2\times24\times18-24-18}{(24+18)^2(24+18-1)}} = \sqrt{\frac{864\times822}{1764\times41}} = \sqrt{\frac{710208}{72324}}$$

$= \sqrt{9.82} = 3.13$.

The value of Z is: $z = \frac{22-21.57}{3.13} = 0.14$. The critical value of $Z = \pm1.96$.

Conclusion. Since the calculated value of Z, 0.14 falls within –1.96 to +1.96; we cannot reject the null hypothesis. That means the arrangement of the passengers in the line can be considered as random.

19.3.5. Rank -Sum Tests: The U test or the Wilcoxon test or the Mann-Whitney test. Now we shall present a nonparametric alternative to the two-sample t test, which is called the U test, the Wilcoxon test, or the Mann-Whitney test, named after the statistician who contributed to its development. Without having to assume that the two populations sampled have normal distributions, we will be able to test the null hypothesis that we are sampling identical continuous populations against the alternative that the two populations have different location parameters. Mann-Whitney proposed the test for two samples of unequal sizes. Wilcoxon gave this type of test for two samples of equal sizes and used the rank sum as the test statistic. This test is, therefore sometimes known as the Mann-Whitney-Wilcoxon test.

Assumptions

i) The data consist of a random sample of observations $x_1, x_2, \ldots, x_{n1}$ of size n_1 from population 1 and another random sample of observations $y_1, y_2, \ldots, y_{n2}$ of size n_2 from population 2.
ii) The two samples are independent.
iii) The variables are of continuous type.
iv) The distribution functions of the two populations differ only with respect to location parameters if they differ at all.
v) The variable of interest is measured at least on ordinal scale.

Hypotheses

a) Two tailed test H_0 : The two populations have identical distributions; H_1 : The two populations differ with respect to location parameter.
b) Right tailed test H_0 : The two populations have identical distributions; H_1 : The x's tend to be greater the y's.
c) Left tailed test H_0 : The two populations have identical distributions; H_1 : The x's tend to be less than y's.

Test statistic. To calculate the test statistic, we first combine the two samples and rank all observations from the smallest to the largest. We assign tied observations the mean of the rank positions they would have occupied. The sum of the rank values in sample 1 is obtained and let it be S. Then the test is: $U = S - \frac{n_1(n_1+1)}{2}$.

Conclusion can be made as follows:

1. For case (a), From Table 13 we note down $W_{\alpha/2}$ and also calculate $W_{(1-\alpha/2)} = n_1 \times n_2 - W_{\alpha/2}$. Reject H_0 if the calculated value of U is less than $W_{\alpha/2}$ or greater than $W_{(1-\alpha/2)}$.
2. For case (b), reject H_0 if the calculated value of U is less than W_α .
3. For case (c), we calculate the value of $W_{(1-\alpha)}$ where $W_{(1-\alpha)} = n_1 \times n_2 - W_\alpha$. Reject H_0 if the calculated value of U is greater than $W_{(1-\alpha)}$.

Example 19.13 . The following data refer to the nicotine contain in milligram by two brands of cigarettes.

Brand A : 5.4 4.8 6.8 3.3 6.1 3.7 2.1 4.0 6.3

Brand B : 2.5 4.0 4.1 0.6 3.1 6.2 1.6 2.2 1.9 5.4

Use the U test at the 0.05 level of significance to test the null hypothesis that the two populations sampled are identical against the alternative hypothesis that the two populations differ with respect location parameters.

Solution.

Two tailed test H_0 : The two populations have identical distributions; H_1 : The two populations differ with respect to location parameter.

First, we combine all the observations of the two samples and then arrange them in order of magnitude. Ranking the data jointly according to size and add all the ranks of the sample 1 which are shown in the following table.

Observations in ascending order	Ranks of sample 1	Ranks of sample 2
0.6		1
1.6		2
1.9		3
2.1	4	
2.2		5
2.5		6
3.1		7
3.3	8	
3.7	9	
4.0	10.5	
4.0		10.5
4.1		12
4.8	13	
5.4	14.5	
5.4		14.5
6.1	16	
6.2		17
6.3	18	
6.4	19	

The sum of the ranks of the first sample

$$= W = 4 + 8 + 9 + 10.5 + 13 + 14.5 + 16 + 18 + 19 = 112.$$

The value of the test statistic is: $U = W - \frac{n_1(n_1+1)}{2} = 112 - \frac{9(10)}{2} = 112 - 45 = 67.$

From the Table 13, we have for

$n_1 = 9$ and $n_2 = 10$, $\alpha/2 = 0.025$, $W_{0.025} = 19$ and $W_{0.975} = 9(10) - 21 = 69.$

Conclusion. Since the calculated value of U, 67 is less than critical value, 69, we have no reason to reject the null hypothesis. That means the two brands of cigarettes on average contains the same amount of nicotine.

19.3.5(a). Mann-Whitney Test for Large sample. When the sample sizes either n_1 or n_2 are greater than 20, we cannot use Table 13. In that case the distribution of U is approximately normal with mean $= \mu = \frac{n_1 \times n_2}{2}$ and variance $= \sigma^2 = \frac{n_1 \times n_2(n_1 + n_2 + 1)}{12}$. For large samples, the test statistic of the null hypothesis is $Z = \frac{U - \mu}{\sigma}$ which is a standard normal variate. But in practice when both n_1 and n_2 are greater than 8, the distribution of U follows approximately normal.

Since both n_1 and n_2 are greater than 8, we can safely use the test statistic Z to test the null hypothesis of example 20.13.

Example 19.14. With reference to the data of Example 20.13 test the null hypothesis that the two populations sampled are identical against the alternative hypothesis that the two populations differ with respect location parameters at the 0.05 level of significance by using Z test since the sample sizes are greater than 8.

Solution. Here $n_1 = 9$, $n_2 = 10$, $S = 112$, and $U = 67$.

The mean and variance of U are

$$\mu = \frac{n_1 \times n_2}{2} = \frac{9 \times 10}{2} = 45, \text{ and } \sigma^2 = \frac{n_1 \times n_2(n_1 + n_2 + 1)}{12} = \frac{9 \times 10(9 + 10 + 1)}{12} = 150.$$

Hence the standard error of U is $\sigma = \sqrt{150} = 12247$. The value of the test statistic is

$$z = \frac{67 - 45}{12247} = 1.80.$$

Conclusion. The critical value of Z is ±1.96. The calculated value of Z is 1.80 which is less than 1.96; we cannot reject the null hypothesis. That means on average the brands of cigarettes contain the same amount of nicotine.

Nonparametric test of Spearman's Rank correlation coefficient. The simple correlation coefficient measures the strength of linear relationship between two quantitative variables and test statistic t is used to test the null hypothesis $H_0 : \rho = 0$ which was discussed in Chapter 16.

The assumptions required for t test are

i) The bivariate sample must come from a bivariate normal distribution.
ii) The relationship between the variables is linear.
iii) The variables must be measured at least in interval scale.
iv) The null hypothesis is $H_0 : \rho = 0$. That means the null hypothesis is that the two variables are linearly independent.

When the above-mentioned assumptions are not satisfied, a nonparametric alternative to measure the association is the rank correlation coefficient, also called Spearman's rank correlation coefficient.

The rank correlation coefficient is measured and tested under the following situations.

i) Sometimes two variables are not measured numerically but can be ranked according quality.
ii) When the relationship between the variables is linear or not linear.
iii) When the bivariate population is not normal.
iv) The variables are measured at least in ordinal scale.

For a given set of paired data $\{(x_i, y_i); i = 1, 2, \ldots\ldots, n\}$, it is obtained by ranking the x's among themselves and also the y's, both from low to high or from high to low, the formula for finding rank correlation coefficient is

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

where d_i is the difference between the ranks assigned to x_i and y_i.

Hypothesis. $H_0: \rho = 0$; that means x and y are independent; $H_1: \rho \neq 0$; x and y are not independent.

Here ρ is the population rank correlation coefficient between the variables x and y. r_s is taken as the test statistic to test the null hypothesis.

Conclusion. If the calculated value of r_s is greater than the critical value for the sample size n at the α level of significance which is obtained from the Table14, we must reject the null hypothesis. In that case the variables x and y are independent.

Example 19.15. The following data relate to the efficiency ranking of eight primary school teachers given by two inspectors of school.

Serial no. teachers :	1	2	3	4	5	6	7	8
Inspector 1 X :	7	4	2	6	1	3	8	5
Inspector 2 Y :	1	5	3	4	8	7	2	6

Use 0.05 level of significance to test the null hypothesis that there is no association between judgments of the two inspectors.

Solution. Table for calculation.

Rank of 1st inspector	Rank of 2nd Inspector	d	d^2
7	1	6	36
4	5	-1	1
2	3	-1	1
6	4	2	4
1	8	-7	49
3	7	-4	16
8	2	6	36
5	6	-1	1
			144

The rank correlation coefficient is

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times144}{8\times63} = 1 - \frac{864}{504} = 1 - 1.714 = -0.714.$$

$H_0: \rho = 0$; $H_1: \rho \neq 0$; $\alpha = 0.05$ and $\alpha/2 = 0.025$.

Conclusion. Since it is a two tailed test, we take the critical value at the 0.025 level of significance. For $n = 8$, it is 0.738 which is obtained from the Table 14 of Appendix. But our calculated value, −0.714 is less than the critical value. The null hypothesis cannot be rejected; in other words, there exist a no association between judgments of the two inspectors.

Remarks. We take the absolute value of the rank correlation coefficient in taking the decision since no negative critical values are given in the Table 14.

Example 19.16. The following are the number of hours which 10 students studied for an examination and the scores they obtained.

Number of hours studied (x) :	8	5	11	13	10	5	18	15	2	8
Scores obtained (y) :	56	44	79	72	70	54	94	85	33	65

Calculate the Spearmen's rank correlation coefficient and test its significance at 0.05 level.

Solution. Hypothesis. $H_0: \rho = 0$; that mean x and y are independent.

$H_1: \rho \neq 0$; x and y are not independent.

Here ρ is the population rank correlation coefficient between the variables x and y. Here $\alpha = 0.05$,

Ranking the x's and the y's and proceeding as in the following table, we get

Rank of x	Rank of y	d	d^2
6.5	7	-0.5	0.25
8.4	9	-0.5	0.25
4	3	1.0	1.00
3	4	1.0	1.00
5	5	0.0	0.00
8.5	8	0.5	0.25
1	1	0.0	0.00
2	2	0.0	0.00
10	10	0.0	0.00
6.5	6	0.5	0.25
			3.00

The rank correlation coefficient: $r_s = 1 - \dfrac{6 \times 3}{10 \times 99} = 0.98$.

This shows that there exist a very strong association between x and y.

Conclusion. Since it is two tailed test, we take the critical value at the 0.025 level of significance. For $n = 10$, it is 0.648 which is obtained from the Table 14 of Appendix. But our calculated value, 0.98 is greater is much greater than the critical value. The null hypothesis is rejected; in other words, there exist a strong association between x and y.

19.3.6(a). Test of Rank Correlation Coefficient for Large Sample. For large sample, the distribution of R_s is approximately normal with mean zero and standard error $\dfrac{1}{\sqrt{n-1}}$. In practice, this result is true for $n \geq 10$. In that case the test statistic for testing the null hypothesis is $Z = \dfrac{R-0}{\dfrac{1}{\sqrt{n-1}}} = R\sqrt{n-1}$.

Example 19.17. With reference to Example 20.12, test at the 0.01 level of ignorance whether the value obtained for r_s, 0.98 is significant.

Solution. 1. $H_0: \rho = 0$; 2. $H_1: \rho \neq 0$; 3. $\alpha = 0.01$

Here $n = 10$, we can safely use the test statistic Z defined above to test the null hypothesis.

4. Reject the null hypothesis if $Z \le -2.575$ *or* $Z \ge 2.575$, *where*

$$Z = R_s\sqrt{n-1}.$$

Here $n=10$, and $r_s = 0.98$, then value test statistic $z = 0.98\sqrt{10-1} = 0.98 \times 3 = 2.94$.

nclusion. The calculated value of Z is 2.94 which is greater than 2.575, the null hypothesis ust be rejected: we conclude that there is a real positive relationship between study and scores.

ndal's Coefficient of Concordance. Spearman's rank correlation coefficient is used to measure relationship between two sets of ranks. In practice, we may have to find the association ong m sets of ranks. In this case, we can use Kendal's coefficient of concordance and can test The formula for finding coefficient of concordance is

$$W = \frac{12\Sigma R_i^2 - 3m^2 n(n+1)^2}{m^2 n(n^2-1)};$$

where R_i is the sum of the ranks of the ith element, m is the number of sets and n is the mber of elements in each set.

nclusion. The probability of the value of W for different values of m and n are presented in ble 13. We must reject the null hypothesis if the probability of the calculated value of W is less n or equal to 0.05. In that case it will be conclude that there exists a relationship among the m s of observations.

amples 19.18. The following data relate to the ranking of five competitors by four judges.

Competitors Judges	A	B	C	D	E
1	4	3	5	2	1
2	4	3	1	5	2
3	4	3	2	5	1
4	3	4	5	2	1

Use 0.05 level of significance to test the null hypothesis that there is no relationship among judgments of the four judges.

lution. H_0: There is no associations among the judgments of the judges.

H_1: There exists an association among the judgments of the judges.

Here $R_1 = 15$, $R_2 = 13$, $R_3 = 13$, $R_4 = 14$ and $R_5 = 5$, where R_i is the sum of the ranks for e ith element we have $m=4$, $n=5$. The value of the test statistic under the null hypothesis is

$$W = \frac{12\Sigma R_i^2 - 3m^2 n(n+1)^2}{m^2 n(n^2-1)} = \frac{12(225+169+169+196+25) - 3\times 16\times 5\times 36}{16\times 5\times 24}$$

$$= \frac{9408-8640}{1920} = \frac{768}{1920} = 0.4.$$

Conclusion. From Table 15, it is seen that the probability of W when $m=4$ and $n=5$ is 0.1105 (it is the mean of 0.119 and 0.102). This probability is greater than $\alpha=0.05$. So we have no reason to reject the null hypothesis. That is, there is no association among the judgments of the judges.

19.3.7(a). Coefficient of Concordance for Large Sample. When $m>5$ and $n>7$, we cannot use Table 15 to test the null hypothesis. In that case, $\chi^2=m(n-1)W$ is approximately distributed as chi-square with n-1 degrees of freedom.

19.3.8. Kruskal–Wallis Test or the H Test (One-Way Analysis of Nonparametric Test). The H test, also called the Kruskal- Wallis test is a generalization of the Mann-Whitney which was discuss earlier. In this test we test the null hypothesis that k samples come from identical continuous population. In other words, it is a nonparametric alternative to the one-way analysis of variance. The scores of all samples are combined and the data are ranked jointly from low to high. ties are treated as usual. Then letting R_i be the sum of the ranks of the values of the ith sample, we base the test on the statistic

$$H=\frac{12}{n(n+1)}\Sigma\frac{R_i^2}{n_i}-3(n+1).$$

where n_i is the size of the ith sample, $n=n_1+n_2+\ldots+n_k$ and k is the number of samples.

Conclusion. When we deal with three samples each with 5 or fewer observations, we can compare the calculated value of H with that of given in Table 14 and reject H_0 if the calculated value is equal or smaller than the tabulated value at α level of significance.

For more than three samples with each more than 5 observations we usually compare the calculated value of H with tabulated value of χ^2 with (k–1) degrees of freedom where k is the number of samples considered in the analysis.

Remarks. For large sample theory, the sampling distribution of the random variable corresponding to H can be approximated closely with a chi-square distribution with $k-1$ degrees of freedom.

Example 19.19. The following are the final grades of samples from three groups of students who were taught English by three different methods (classroom instruction and language laboratory, only classroom instruction, and only self-study in language laboratory).

First method : 94, 88, 91, 74, 87, 97
Second method : 85, 82, 79, 84, 61, 72, 80
Third method : 89, 67, 72, 76, 69

Use the Kruskal-Wallis test at the 0.05 level of significance to test the null hypothesis that the three methods are equally effective.

Solution. 1) $H_0:\mu_1=\mu_2=\mu_3$; 2) Are not all equal; 3) $\alpha==0.05$.

4) Reject the null hypothesis if $H\geq 5.991$, where 5.991 is the value of chi-square at 0.05 level of significance when the degree of freedom is 2.

Table for finding H.

Combined observations in ascending order	Rank of the 1st sample	Rank of the 2nd sample	Rank of the 3rd sample
61		1	
67			2
69			3
72		4.5	
72			4.5
74	6		
76			7
79		8	
80		9	
82		10	
84		11	
85		12	
87	13		
88	14		
89			15
91	16		
94	17		
97	18		
	$R_1 = 84$	$R_2 = 55.5$	$R_3 = 31.5$

Here $n_1 = 6,\ n_2 = 7,\ n_3 = 5,\ \ n = 6+7+5 = 18$. The value of the test statistic is

$$H = \frac{12}{n(n+1)} \Sigma \frac{R_i^2}{n_i} - 3(n+1) = \frac{12}{18 \times 19}\left[\frac{(84)^2}{6} + \frac{(55.5)^2}{7} + \frac{(31.5)^2}{5}\right] - 3(19) = 6.67.$$

Conclusion. Since $H = 6.67$ exceeds $\chi^2_{0.05,2} = 5.991$, the null hypothesis must be rejected, we conclude that the three methods are not all equally effective.

19.3.9. Friedman's Nonparametric Test for Randomized Block Design. Friedman's test is a nonparametric alternative of the parametric randomized block design which needs normality assumptions. If there is reason to believe that the data do not satisfy the assumptions of analysis of variance, the Friedman test may be appropriate.

Suppose a randomized block design consists with k number of treatments and r number blocks. In this design, k treatments are randomly given to the experimental units (plots) in such a way that each block contains each treatment only one times. That means each block contains k number of experimental units. The data are then arranged in a 2-way table containing r rows correspond to r blocks and k columns correspond to k treatments. The observations in rows are ordered and ranks are assigned which results to a new set of data of ranks. The ranks of each column are summed. If the samples are from the same population, the ranks in each column *will* be a random arrangement of the number of conditions. If the sums differ significantly *the null* hypothesis that they are from the same population is rejected.

The null and alternative hypotheses of the test are

H_0 : The populations of the k treatments are identical.

H_1 : At least two populations of the treatments are not same.

The test statistic is $\chi^2 = \frac{12}{rk(k+1)} \Sigma R_i^2 - 3r(k+1)$ is a chi-square with $k-1$ degrees of freedom.

Conclusion. The null hypothesis will be rejected at α level of significance if the calculated value of chi-square is greater than the tabulated value of $\chi^2_{\alpha.(k-1)}$ with $k-1$ degrees of freedom.

Example 19.20 Three different levels of a certain fertilizers were tried in a randomized block design with 5 blocks at a certain agricultural farm to study the effects of the levels of fertilizer on cotton crop. The yield per plot in kg for different levels of fertilizer and blocks are given systematically.

Block	Treatment 1	2	3
1	591	682	727
2	818	591	863
3	682	636	773
4	499	625	909
5	648	836	818

Use 0.05 level of significance that the effects of all levels of fertilizer are same.

Solution.

1) H_0 : The effects of all levels of fertilizers are same.

2) H_1 : At least two are not same. 3) $\alpha = 0.05$.

The observations for each treatment are ordered and ranks are assigned The data of the ranks are as follows :

Block	Treatment 1	2	3
1	3	2	1
2	2	3	1
3	2	3	1
4	3	2	1
5	3	1	2
	13	11	6

The test statistic for testing the null hypothesis is $\chi^2 = \frac{12}{rk(k+1)} \Sigma R_i^2 - 3r(k+1)$ is a chi-square with $k-1$ degrees of freedom.

Here $k=3$, $r=5$, $\alpha=0.05$ and $k-1=3-1=2$. The value of the test statistic is

$$\chi^2 = \frac{12}{rk(k+1)}\Sigma R_i^2 - 3r(k+1) = \frac{12}{5\times3\times4}[13^2 + 11^2 + 6^2] - 3\times5\times4 = 65.2 - 60 = 5.2.$$

Conclusion. The value of chi-square with 2 d.f. at 0.05 level of significance is 5.99. But the calculated value of chi-square with the same d.f. is 5.2 which is less than 5.99, the null hypothesis cannot be rejected This means that the effects of all the levels of fertilizers are same.

19.3.10. McNamara Test for Correlated Proportions in a 2×2 Contingency Table.

McNamara's test assesses the significance of the difference between two correlated proportions, such as might be found in the case where the two proportions are based on the same sample of subjects or on matched-pair samples.

This test is particularly applicable to those "before and after" designs in which each person is used as his own control and in which measurement is in the strength of either nominal or ordinal scale. Thus, it might be used to test the effectiveness of a particular treatment (meeting, newspaper editorial, mailed pamphlets, personal visit etc.) on voter's preference among various candidates. or it might be used to test, say the effect of farm to city moves on people's political affiliations.

To test the significance of any observed change by this method, a four-fold table of frequencies to represent the first and second sets of responses from the same individuals. The general features of such a table are illustrated below, in which + and – signs are used.

		Before	
		+	–
After	–	A	C
	+	B	D

Notice that those cases which show changes between the first and second response appear in cells A and D. An individual is tallied in cell A if he changes from positive to negative while one is tallied in cell D of the changes from negative to positive. if no change is observed he is tallied either in cell B (+ both before and after) or C (– both before and after) Since A +D represent the total number of persons who changed, the expectation under null hypothesis would be 1/2(A+D) cases changed in one direction and 1/2(A+D) cases changed in the other. In other words, 1/2(A+D) is the expected frequency under H_0 in both cells A and D. Then the test statistic is

$$\chi^2 = \Sigma\frac{(O_i - E_i)^2}{E_i};$$

where O_i is the observed number of cases in the ith category and E_i is the expected number of cases in the ith category under H_0. This is distributed as chi-square with 1 degree of freedom. In McNamara test for the significance of changes, we are interested in cells A and D. The test statistic can simplify as follows.

$$\chi^2 = \Sigma\frac{(O_i - E_i)^2}{E_i} = \frac{\left(A - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} + \frac{\left(D - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} = \frac{(A-D)^2}{A+D}.$$

This follows chi-square with 1 degree of freedom.

Yate's Correction for Continuity. The approximation by the chi-square distribution of the sampling distribution above becomes excellent if a correction for continuity is performed. The correction is necessary because continuous distribution, chi-square is used to approximate a discrete distribution. When all the expected frequencies are small, approximation may be a poor one. Yates's correction for continuity is an attempt to remove this source of error after correction for continuity, the test statistic becomes $\chi^2 = \frac{(|A-D|-1)^2}{A+D}$ with 1 d.f..

Example 19.21. The following data relate to the attitude of the farmers on HYV rice cultivation. Plus sign indicates favourable attitude of farmers before and after the workshop and minus sign indicates unfavourable attitudes.

		After	workshop	Total
		–	+	
Before workshop	+	30	150	180
	–	20	80	100
	Total	50	230	280

Solution. H_0 : Workshop has no effect in changing the attitude of the farmers.

H_0 : Workshop has effect in changing the attitude of the farmers.

The value of the test statistic under the null hypothesis is

$$\chi^2 = \frac{(A-D)^2}{A+D} = \frac{(30-80)^2}{30+80} = \frac{2500}{110} = 22.73.$$

Conclusion. The value of chi-square with 1 d.f. at 0.01 level of significance is 10.83. But the calculated value of chi-square with the same d.f. is 22.73 which is greater than 10.83. The value is highly significant. We must reject the null hypothesis. That means the workshop has a positive effect in changing the attitude of the farmer regarding the cultivation of the HYV rice.

Example 19.22. An opinion survey was conducted on 100 persons about a presidential candidate before and after the campaign. The results are as follows :

After the campaign	Before the campaign		
	Support	Non-support	Total
Not - Support	5	20	25
Support	55	20	75
Total	60	40	100

Do you think the campaign was effective? Use $\alpha = 0.05$.

Solution. H_0 : The campaign was not effective.

H_0 : The campaign was effective.

The value of the test statistic under the null hypothesis is

$$\chi^2 = \frac{(|A-D|-1)^2}{A+D} = \frac{(|5-20|-1)^2}{5+20} = \frac{196}{25} = 7.84.$$

Conclusion. The value of chi-square with 1d.f. at 0.05 level of significance is 3.841. But the calculated value of chi-square with the same df. is 7.84 which is greater than 3.841. The value is significant. We must reject the null hypothesis. That means the campaign has a positive impact in changing the attitude of the voters regarding the presidential candidate.

Group-A : Short questions and answers

1. **What is parametric test?**
   Ans. A parametric test is a test whose m specifies certain conditions about the parameters of the population from which the samples are drawn.
2. **What is nonparametric test?**
   Ans. A nonparametric test is a test that does not depend on the particular form of the basic population from which the samples are drawn.
3. **What is a Kolmogorov-Smirnov test?**
   Ans. A nonparametric test, which does not require that data be grouped in any way, for **determining whether there is any significant difference between an observed frequency** distribution and a theoretical frequency distribution.
4. **What is Kruskal-Wallis Test?**
   Ans. A nonparametric method for testing whether three or more independent samples have been drawn from populations with the same distributions. It is a nonparametric counterpart of one-way ANOVA.
5. **What is Mann-Whitney U test?**
   Ans. A nonparametric method used to determine whether two independent samples have been drawn from populations with the same distribution.
6. **What is one-sample run test?**
   Ans. A nonparametric method for determining the randomness with which the items in a sample have been selected.
7. **What is rank sum test?**
   Ans. A family of nonparametric tests that use the order information in a set of data.
8. **What is run?**
   Ans. A sequence of identical occurrence preceded and followed by different occurrences or by nor at all.
9. **What is sign test?**
   Ans. A nonparametric test for the differences between paired observations where + and – signs are substituted for quantitative values.

Group—B & C : Broad questions and problems

1. Define parametric test. What are the advantages of parametric test over non-parametric test? Describe a parametric test.
2. What is a run? How can you test the randomness of a set of data?

3. What is a nonparametric test corresponding to a parametric one sample t-test. Describe a non-parametric test corresponding to completely randomized test.
4. What is Kruskal –Wallis test? Can it be considered as a non-parametric counterpart of a one-way analysis of variance test?
5. Discuss Friedman test. Is it a non-parametric counterpart of a randomized block design?
6. What is Mann-Whitney test? Why it is so called? Can it be considered as non-parametric counterpart of a two-sample t-test?
7. Discuss McNamara test. Cite some examples where you can apply.

Applications

8. The following are the amounts of time, in minutes, which it took a random sample of 20 technicians to perform a certain task.

   18.1, 20.3, 18.3, 15.6, 22.5, 16.8, 17.6, 16.9, 18.2, 17.0, 19.3,
   16.5, 19.5, 18.6, 20.0, 18.8, 19.1, 17.5, 18.5 and 18.0.

   Assuming that this sample came from a symmetrical continuous population, use sign test at the 0.05 level of significance to test the null hypothesis that the mean of this population is 19.4 minutes against the alternative hypothesis that it is not 19.4 minutes.

   Ans. The P-value is 0.0059 the null hypothesis must be rejected.

9. The following are the numbers of passengers carried by two airlines between Dhaka and Chittagong on 12 days.

   232 and 189, 265 and 230 249 and 236 250 and 261 255 and 249 236 and 218
   270 and 258 247 and 253 249 and 251 240 and 233 257 and 254 239 and 249

   Use the sign test at the 0.01 level of significance to test the null hypothesis $\mu_1 = \mu_2$ (that on average the two flights carry equal number of passengers) against the alternative $\mu > \mu_2$.

   Ans. The P-value is 0.1937, the null hypothesis cannot be rejected.

10. In a random sample taken at a public playground, it took 38, 43, 36, 29, 44, 28, 40, 50, 39, and 33 minutes to play a set of tennis. Use the signed-rank test at the 0.05 level of significance to test whether or not it takes on the average 35 minutes to play a set of tennis at that public playground.

    Ans. $T=15$; the null hypothesis cannot be rejected.

11. A sample of 24 suitcase carried by an airline weighed

    32.0, 46.4, 48.1, 27.7, 35.5, 52.6, 66.0, 41.3, 49.9, 36.1, 50.0, 44.7, 48.2, 36.9, 40.8, 35.1,
    63.3, 42.5, 52.4, 40.9, 38.6, 43.2, 41.7 and 35.6 pounds.

    Use Signed-rank test at the 0.05 level of significance to test whether or nor the weight of suitcases carried by the airline is 37.0 pounds.

12. The following is a random sample of the IQ's of husbands and wives: 108 and 103, 104 and 116, 103 and 106, 112 and 104, 99 and 99, 105 and 94, 102 and 110, 112 and 128, 119 and 106, 106 and 103, 125 and 120, 96 and 98, 107 and 117, 115 and 130, 101 and 100, 110 and 101, 103 and 96, 105 and 99, 124 and 120, and 113 and 116. Test at the 0.05 level of significance whether or not bus

bands and wives are on the average equally intelligent. Ans. $T=98.5$; the null hypothesis cannot be rejected.

13. The following are the weight gains (in pounds) of two random samples of young turkeys fed two different diets.

Diet 1 : 16.3, 10.1, 10.7, 13.5, 14.9, 11.8, 14.3, 10.2, 12.0, 14.7, 23.6, 15.1, 14.5, 18.4, 13.2, 14.0.

Diet 2 : 21.3, 23.8, 15.4, 19.6, 12.0, 13.9, 18.8, 19.2, 15.3, 21.1, 14.8, 18.9, 20.7, 21.1, 15.8, 16.2.

Use the Mann-Whitney test at the 0,01 level of significance to test the null hypothesis that the two populations sampled are identical against the alternative hypothesis that on the average the second diet produces a greater gain in weight.

Ans. $Z=-3.11$, the null hypothesis must be rejected.

14. The following are the number of burglaries committed in a city in a random sample of six days in the spring and 8 days in the fall.

Spring : 75, 56, 63, 70, 58, 74.

Fall : 63, 85, 77, 80, 86, 76, 72, 82.

Use the Mann-Whitney test at the 0,05 level of significance to test the claim that on the average there are equally many burglaries per day in the spring as in the fall against the alternative that they are not.

Ans. $U_1=5.5$; the null hypothesis must be rejected.

15. The following are the miles per gallon which a test driver got for 10 tank fuels of each of three kinds of gasoline.

Gasoline A :	20,	31,	24,	33,	23,	24,	28,	16,	19,	26.
Gasoline B :	29,	18,	29,	19,	20,	21,	34,	33,	30,	23.
Gasoline C :	19,	31,	16,	26,	31,	33,	28,	28,	25,	30.

Use the Kruskal-Wallis test at the 0.05 level of significance to test whether or not there is a difference in the actual average mileage yield of the three kinds of gasoline.

Ans. $H=0.86$; the null hypothesis cannot be rejected.

16. The following are data on the breaking strength (in pounds) of random samples of two kinds of 2-inch cotton ribbons.

Type-I ribbons: 144, 181, 200, 187, 169, 171, 186, 194, 176, 182, 133, 183, 197, 165, 180, 198.

Type-II ribbons: 175, 164, 172, 194, 176, 198, 154, 134, 169, 164, 185, 159, 161, 189, 170, 164.

Use the Mann-Whitney test at the 0.05 level of significance to test the claim that Type I ribbon is, on the average, stronger than Type II ribbon.

Ans. $z=1.62$; the null hypothesis cannot be rejected.

17. Checking on elm trees that were planted many years ago along a country road, a county official obtained the following arrangement of health, H and diseased, D, trees:

HHHH DDD HHHHHHH DD HH DDDD

Test at the 0.05 level of significance whether this arrangement may be regarded as random.

Ans. $r=6$ and $r_{0.025}=6$; the null hypothesis must be rejected.

18. The following is the order in which a broker received buy, B and sell, S orders for a certain stock.

BBBBBBBB SS B SSSSSS BBBB

Test for randomness at the 0.05 level of significance.

Ans. $r=5$; the null hypothesis of randomness must be rejected.

19. The following is the order in which red, R and black, B cards were delt to a bridge player:

BBB RRRRR BB RRR

Test for randomness at the 0.05 level of significance.

Ans. $r=4$; the null hypothesis of randomness cannot be rejected.

20. The following are the number of students absent from school on 24 consecutive days.

29, 25, 31, 28, 30, 28, 33, 31, 35, 29, 31, 33, 35, 28, 36, 30, 33, 26, 30, 28, 32, 31, 38 and 27.

Test at the 0.025level of significance whether there might be a trend.

21. The following are the numbers of lunches that an insurance agent claimed as business deductions in 30 consecutive months.

6, 7, 5, 6, 8, 6, 8, 6, 6, 4, 3, 2, 4, 4, 3, 4, 7, 5, 6, 8, 6, 6, 3, 4, 2, 5, 4, 4, 3 and 7.

Use the runs test to test for randomness at the 0.01 level of significance.

Ans. $r=5$; the null hypothesis of randomness must be rejected.

22. Calculate r_s for the following data representing the statistics grades, x, and mathematics grades, y, of 18 students

x : 78, 86, 49, 94, 53, 89, 94, 71, 70, 97, 74, 53, 58, 62, 74, 74, 70, 74.

y : 80, 74, 63, 85, 55, 86, 90, 84, 71, 90, 85, 71, 67, 64, 69, 71, 67, 71.

Test at the 0.05 level of significance whether the value obtained for r_s is significant.

Ans. $z=3.55$; $r_s=0.86$ is significant.

23. The following are scores which 12 students obtained in the midterm and final examinations in a course in statistics:

Midterm examination (x) :	71, 49, 80, 73, 93, 85, 58, 82, 64, 32, 87, 80
Final examination (y)	: 83, 62, 76, 77, 89, 74, 48, 78, 76, 51, 73, 89

Calculate r_s and test the null hypothesis of no correlation at the 0.05 level of significance.

24. The following are the rankings given by three judges to the works of 10 artists:

Judge A:	6, 4, 2, 5, 9, 3, 1, 8, 10, 7
Judge B:	2, 5, 4, 8, 10, 1, 6, 9, 7, 3
Judge C:	7, 3, 1, 2, 10, 6, 4, 9, 8, 5

Calculate the value of W, the coefficient of concordance as a measure of the agreement of the three sets of rankings and test its significance.

15. The following are the cholesterol contents, in milligram per package, which four laboratories obtained for 6-ounce packages of three very similar diet foods :

Laboratory	Diet food A	Diet food B	Diet food C
Laboratory 1	3.4	2.6	2.8
Laboratory 2	3.0	2.7	3.1
Laboratory 3	3.3	3.0	3.4
Laboratory 4	3.5	3.1	3.7

Use the 0.05 level of significance to test the null hypothesis that the three diets are equally effective by applying Friedman's test for randomized block design.

16. The following data relate to the demonstration effect on the changes of farmers attitude towards adoption of a modern cultivation technique for paddy :

	After	demonstration	
Before demonstration	−	+	Total
+	20	100	120
−	30	70	100
Total	50	170	220

Test the significance of demonstration effect using McNamara test at the 0.05 level of significance.

Conclusion. The critical value of Z is ±1.96. The calculated value of Z is 1.80 which is less than 1.96; we cannot reject the null hypothesis. That means on average the brands of cigarettes contain the same amount of nicotine.

Write T for true and F for false of the following

i. Nonparametric methods are more efficient than parametric method.

ii. One advantage of nonparametric methods is that some of the tests do not require us even to rank the observations.

iii. One disadvantage of nonparametric methods is that they tend to ignore a certain amount of information.

iv. The sequence A, A, B, A, B contains four runs.

v. In the Mann-Whitney U test, it is not necessary that the two samples of the same size.

vi. The Kruskal-Wallies test is a nonparametric version of ANOVA.

vii. The Kolmogorov-Smirnov test can be used to measure the goodness-of-fit of a theoretical distribution.

viii. In a one-sample run test, the number of runs is a statistic that has its own sampling distribution .

ix. In a one-sample runs test, the alternative hypothesis is that the sequence of observations is not random.

x. A sign test for paired data is based on the binomial distribution but can often be approximated by

Ans. i. F, ii. T, iii. T, iv. T, v. T, vi. T, vii. T, viii. T, ix. T, x. T

Multiple Choices

1. **What is the maximum number of runs possible in a sequence of length 5 using two symbols?**
   a. 6 b. 3 c. 5 d. 4

2. **The sequence of C, D, C, D, C, D, C, D, C, D would probably be rejected by a test of runs not being truly randomness because:**
   a. The pattern C,D occurs only five times, this is not often enough to guarantee randomness
   b. The sequence contains too many runs
   c. The sequence contains too few runs
   d. The sequence contains only two symbols.

3. **When compared to parametric methods, nonparametric methods are**
   a. Less efficient b. Computationally easier
   c. Require less information d. All of these

4. **For a sample of size greater than 30, the sampling distribution of the rank correlation coefficient is approximately which distribution?**
   a. t b. Binomial c. Normal d. Chi-square

5. **In the Kruskal-Wallis test of k samples, the appropriate number of degrees of freedom is :**
   a. k b. k-1 c. n-1 d. n-k

Answers:

1.	2.	3.	4.	5.
c	b	d	c	b