SECTION B

14.6 Non-parametric tests

Most of the statistical tests that we have discussed in the previous section are based on the assumption that the forms of the population distributions are known and the tests are concerned with hypotheses about parameters of the population distributions. For example, in all the exact tests in Sec.14.4.3, the population distributions are taken to be normal and the tests relate to the means and variances of these distributions. Such tests may be termed parametric tests.

A non-parametric test, on the otherhand, is a test that does not depend on the particular forms of the population distributions, i.e. here we do not make any assumption about the forms of the parent distributions from which the random samples are drawn. However, we assume that the population distributions are continuous and the sample observations, as in the case of parametric tests, are independent. It may be noted that Karl Pearson's x^2-test for goodness of fit, discussed in Sec. 14.5, is a non-parametric test.

Sometimes the term 'distribution free' is used instead of 'non-parametric'. Actually the two terms are not synonymous. If the testing procedures depend neither on the form of the parent distributions nor on their parameters, then the procedures are said to be distribution free. In fact, all non-parametric tests are distribution free but all distribution free tests are not necessarily non-parametric.

Non-parametric tests have certain advantages in that they require less assumptions, are simple and easy to apply and can be used even in situations where actual measurements are unavailable and the data are obtained only as ranks.

The main disadvantage of non-parametric tests is that they can be used only if the measurements are nominal or ordinal. Even in that case, if a parametric test exists, it is more powerful than the non-parametric test.

Now we consider some non-parametric tests.

14.6.1 One-sample sign test

It is a test for the location parameter (median) of a population.

Suppose θ is the unknown median of a continous distribution and we want to test the null hypothesis $H_0 : \theta = \theta_0$, against one-sided alternative $H : \theta < \theta_0$ (or $H : \theta > \theta_0$) or two-sided alternative $H : \theta \neq \theta_0$, on the basis of independent and random sample observations x_1, x_2, ..., x_n.

Now, if the sample comes from a distribution with median θ_0, then we can expect that, on the average, half of the sample values shall the greater than θ_0 and half smaller than θ_0. Each of the values that are greater than θ_0 is replaced by a plus sign (+) and each of the values that are less than θ_0 by a minus sign (–). Sample values equal to θ_0 may be ignored as they have zero probability due to continuity of the population distribution. Let the total number of plus signs be r and total number of minus signs be s, where $r + s \leq n$. Then the distribution of r, given $r + s$, is binomial with $p = P(x > \theta_0)$. The number r is used to test

$$H_0 : \theta = \theta_0 \text{ (which is equivalent to } H_0 : p = \tfrac{1}{2}).$$

The critical region for testing H_0 at level α against two-sided alternative is given by

$$r \geq r_{\alpha/2} \text{ and } r \leq r'_{\alpha/2},$$

where $r_{\alpha/2}$ is the smallest and $r'_{\alpha/2}$ is the largest integer satisfying

$$\sum_{x=r_{\alpha/2}}^{r+s} \binom{r+s}{x} \frac{1}{2^{r+s}} \leq \frac{\alpha}{2} \quad \text{and} \quad \sum_{x=0}^{r'_{\alpha/2}} \binom{r+s}{x} \frac{1}{2^{r+s}} \leq \frac{\alpha}{2}.$$

The critical region for testing H_0 against one sided alternative $H : \theta > \theta_0$ is $r \geq r_\alpha$ where r_α is the smallest integer such that

$$\sum_{x=r_\alpha}^{r+s} \binom{r+s}{x} \frac{1}{2^{r+s}} \leq \alpha,$$

and the critical region for testing H_0 against the alt. $H : \theta < \theta_0$ is $r \leq r'_\alpha$ where r'_α is the largest integer such that

$$\sum_{x=0}^{r'_\alpha} \binom{r+s}{x} \frac{1}{2^{r+s}} \leq \alpha.$$

When $r + s > 25$, one may use the normal approximation to the binomial. The probability of r or fewer plus signs among $r + s$ plus and minus signs will, under H_0, be approximately given by $\Phi(t)$ where

$$t = \frac{r - (r+s)/2}{\sqrt{(r+s)/4}} = \frac{r-s}{\sqrt{r+s}}.$$

Illustration 14.30

To determine the mileage of a type of truck, 6 trucks were run and the mileage of each obtained with a gallon of gasoline was as follows :

21, 19, 22, 18, 20, 24.

Use the sign test to examine whether the average number of miles run with a gallon of gasoline by trucks of this type is 20, the alternative hypothesis being that it is greater than 20.

Solution : Let the population median be θ miles. Then we are to test $H_0 : \theta = 20$ against the alternative $H : \theta > 20$.

We replace each sample value by a plus or a minus sign according as the value is greater than 20 or less than 20. Values equal to 20 are ignored.

Values :	21	19	22	18	20	24
Signs :	+	–	+	–	ignored	+

∴ The number of plus signs $(r) = 3$ and the number of minus signs $(s) = 2$. So, $r + s = 5$.

The critical region for testing H_0 against H at level 0.05 is $r \geq r_{0.05}$ where $r_{0.05}$ is the smallest integer satisfying

$$\sum_{x=r_{0.05}}^{5} \binom{5}{x} \cdot \frac{1}{2^5} \leq 0.05.$$

This gives $r_{0.05} = 5$. Since for the given problem $r = 3$, the null hypothesis is accepted at 5% level. (The value of $r_{0.05}$ can be obtained from Table VI in the Appendix.)

Note **:** The above procedure of one-sample sign test can be applied to test the hypothesis regarding the median of the distribution of differences of two variables in the population if paired sample data are available. Suppose we are given a random sample of n pairs of values $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ of variables x and y. Assuming that the population distribution of difference $x - y$ is continuous and has median θ, one can test the hypothesis relating to θ by taking the difference $x_i - y_i = d_i$ (say) in place of x_i everywhere in the one-sample sign test.

Illustration 14.31

The weights (in kg.) of 12 persons before they were subjected to a change of diet and after a lapse of six months are given below :

Serial No.	:	1	2	3	4	5	6	7	8	9	10	11	12
Weight (in kg.)													
Before	:	62	58	49	64	42	51	62	40	38	62	47	60
After	:	66	65	51	68	39	54	63	38	41	71	46	64

Test whether there has been any significant gain in weight from the change of diet.

Solution **:** Let x and y denote, respectively the weight after and before the change of diet, and θ be the median of the distribution of $d = x - y$. We are to test the null hypothesis $H_0 : \theta = 0$, against the alternative hypothesis $H : \theta > 0$.

We attach a plus sign or a minus sign to each sample value $d_i = x_i - y_i$ according as $d_i > 0$ or $d_i < 0$.

Value of d_i	:	4	7	2	4	– 3	3	1	– 2	3	9	– 1	4
Sign	:	+	+	+	+	–	+	+	–	+	+	–	+

The number of plus signs $(r) = 9$ and the number of minus signs $(s) = 3$, so that $r + s = 12$.

The critical region for testing H_0 against H at level 0.05 is $r \geq r_{0.05}$ where $r_{0.05}$ is the smallest integer satisfying

$$\sum_{z=r_{0.05}}^{12} \binom{12}{z} \cdot \frac{1}{2^{12}} \leq 0.05.$$

This gives $r_{0.05} = 10$. Since for our problem $r = 9$, we accept the null hypothesis at 5% level. (For value of $r_{0.05}$, one can consult Table VI in the Appendix).

14.6.2 One-sample Wilcoxon signed-rank test

It is a test for the location parameter (median) of a distribution which is assumed to be continuous and symmetric. This test is more efficient than the corresponding sign test, since here both the signs and the magnitudes (in the form of ranks) are taken into consideration.

Suppose θ is the unknown median of the population and we are required to text the null hypothesis $H_0 : \theta = \theta_0$, against one-sided alternative $H : \theta < \theta_0$ (or $H : \theta > \theta_0$) or two-sided alternative $H : \theta \neq \theta_0$, on the basis of n independent and random sample observations $x_1, x_2, ..., x_n$. Now, under H_0, the differences $d_i = x_i - \theta_0$ are independent and come from a population which is continuous and symmetric about zero. So, positive and negative differences of the same numerical value have equal probabilities to occur. We rank the absolute differences, $|d_i|$'s, in increasing order, ignoring 0's (if there be any), since they have zero probability to occur. Tied ranks are given the average value of the ranks in case there had been no ties. Let the sum of ranks of positive d_i's be T^+ and that of negative d_i's be T^-. Then $T^+ + T^- = \frac{m(m+1)}{2}$, where m is the number of non-zero d_i's and $m \leq n$. The distributions of T^+ and T^-, under H_0, are identical, each being symmetric about $\frac{m(m+1)}{4}$ and having range from 0 to $\frac{m(m+1)}{2}$. Again, since $T^+ + T^- = \frac{m(m+1)}{2}$, a constant, the test statistics based on T^+ and T^- are related and they provide equivalent test criteria.

We note that

$$
\begin{aligned}
P[T^+ \geq k \mid \theta = \theta_0] &= P\left[T^+ \leq \frac{m(m+1)}{2} - k \mid \theta = \theta_0\right] \\
&= P\left[T^- \leq \frac{m(m+1)}{2} - k \mid \theta = \theta_0\right] \\
&= P[T^- \geq k \mid \theta = \theta_0].
\end{aligned}
$$

In practice, we work using T, the smaller of the two sums, and Table VII of Appendix gives the left-hand critical values for the random variable T(which may be T^+ or T^-). If T_α is such that $P[T \leq T_\alpha] = \alpha$, the approximate critical regions for test of H_0, at level α, against different alternatives are as follows :

Alternative	Critical region
$H : \theta > \theta_0$	$T^- \leq T_\alpha$
$H : \theta < \theta_0$	$T^+ \leq T_\alpha$
$H : \theta \neq \theta_0$	$T^+ \leq T_{\alpha/2}$ or, $T^- \leq T_{\alpha/2}$.

One can see, for instance, that in case of alternative $H : \theta > \theta_0$, we should reject $H_0 : \theta = \theta_0$ if T^+ is too large or, equivalently, T^- is too small. This is the rational behind the above test rules.

For $m > 25$, under H_0, T is approximately normally distributed with $E(T) = \frac{m(m+1)}{4}$ and

$$\text{Var}(T) = \frac{m(m+1)(2m+1)}{24}$$

Note : We can apply the above procedure of one-sample Wilcoxon signed-rank test for testing hypothesis about median of the distribution of difference of two variables if paired sample data are available. Here we are to take $d_i = x_i - y_i - \theta_0$ in place of $d_i = x_i - \theta_0$ in the one-sample test.

Illustration 14.32

Test the hypothesis that the median length of ear-head of a variety of wheat is 9.0 cm. against the alternative that it is not equal to 9.0 cm. at level 0.05, on the basis of the following 16 sample ear-head measurements (in cm.) :

8.3, 9.0, 8.8, 10.5, 10.7, 8.3, 9.7, 10.1,
7.9, 10.0, 11.1, 8.6, 8.9, 9.3, 8.5, 9.4.

Solution : Let θ be the median length of ear-head of the variety of wheat. We are to test the null hypothesis $H_0 : \theta = 9.0$ against the alternative hypothesis $H : \theta \neq 9.0$.

The deviations of the sample observations from 9.0 are – 0.7, 0, – 0.2, 1.5, 1.7, – 0.7, 0.7, 1.1, – 1.1, 1.0, 2.1, – 0.4, – 0.1, 0.3, – 0.5, 0.4 Now, we arrange the deviations according to their absolute values in increasing order and assign ranks.

Values	:	0	– 0.1	– 0.2	0.3	– 0.4	0.4	– 0.5	– 0.7
Ranks	:	(ignored)	1	2	3	4.5	4.5	6	8
Values	:	– 0.7	0.7	1.0	– 1.1	1.1	1.5	1.7	2.1
Ranks	:	8	8	10	11.5	11.5	13	14	15

Here sum of the ranks of positive deviations is $T^+ = 79$ and that of the negative deviations is $T^- = 41$, so that the smaller of the two sums is $T = 41$.

From Table VII of Appendix, for $n = 15$ (number of non-zero derivations) and $\alpha = 0.05$ (for two-sided test), we get $T_\alpha = 25$. Since T^+ and T^{-1} are both greater than T_α, we have not sufficient evidence to reject H_0.

It may be noted that in case of one-sided alternative $H : \theta > 9.0$ $(H : \theta < 9.0)$, we have to compare $T^- = 41(T^+ = 79)$ with the critical value $T_a = 25$, at level $\alpha = 0.025$, and arrive at the same conclusion that there is no evidence for rejecting H_0 since $T > T_\alpha$.

14.6.3 Two-sample Wilcoxon rank-sum test

Here our task is to test whether the location parameters (medians) of two distributions are same. For this test we assume that the distributions are continuous and that they may differ only in their locations (medians).

Suppose x and y are the variables of the two distributions having medians θ_1 and θ_2, and we are to test the null hypothesis $H_0 : \theta_1 = \theta_2$ against one-sided alternative $H : \theta_1 > \theta_2$ (or $H : \theta_1 < \theta_2$) or two-sided alternative $H : \theta_1 \neq \theta_2$, on the basis of independent sample observations $x_1, x_2, ..., x_{n_1}$ and $y_1, y_2, ..., y_{n_2}$ from the two distributions.

We arrange the $n_1 + n_2 = N$(say) observations in increasing order of magnitude and rank them from 1 to N; in case of ties, we allot average ranks. The sum of the ranks assigned to the values of y is denoted by W, which is the Wilcoxon rank-sum statistic. When the two distributions are identical, i.e., when $\theta_1 = \theta_2$, we can expect that the sample values of x and y will be thoroughly mixed. But when $\theta_1 < \theta_2$, most of the higher ranks will be occupied by y-values and hence W will be large. Similarly, a small value of W will indicate $\theta_1 > \theta_2$, whereas too small or too large values of W will indicate $\theta_1 \neq \theta_2$.

Since the sum of the ranks assigned to N sample values is $N(N + 1)/_2$, the sum of the ranks assigned to x-values will be $N(N + 1)/2 - W$. When H_0 holds, the distribution of W is symmetric about its mean $n_2(N + 1)/2$.

While testing $H_0 : \theta_1 = \theta_2$ against the alternative $H : \theta_1 < \theta_2$ at level α, we reject H_0 if $W > \omega_{\alpha;\, n_1,\, n_2}$, where $\omega_{\alpha;\, n_1,\, n_2}$ is the upper α-point of the distribution of W (under H_0). Against alternative $H : \theta_1 > \theta_2$, we reject H_0 if $W \leq n_2(N + 1) - \omega_{\alpha;\, n_1,\, n_2}$. Again, for the two-sided alternative $H : \theta_1 \neq \theta_2$, we reject H_0 at level α if $W \geq \omega_{\alpha_2;\, n_1,\, n_2}$ or $W \leq n_2(N + 1) - \omega_{\alpha_1;\, n_1,\, n_2}$, where $\alpha_1 + \alpha_2 = \alpha$. [Values of $\omega_{\alpha;\, n_1,\, n_2}$ are given in the book 'Non-parametric Statistical Methods" by M. Hollander and D.A. Wolfe.]

For large samples, the test statistic

$$W^* = \frac{W - n_2(N+1)/2}{\sqrt{n_1 n_2 (N+1)/12}},$$

which follows, under H_0, an asymptotic N(0, 1) distribution as $\min(n_1, n_2) \to \infty$, is taken to perform the normal deviate test.

For testing $H_0 : \theta_2 - \theta_1 = \theta_0$, a given non-zero value, we are to compute W using x_i and $y_i - \theta_0$ in place of x_i and y_i, respectively, and then proceed as before.

14.6.4 Two-sample Mann-Whitney U-test

The problem is exactly same as in the two-sample Wilcoxon rank-sum test, but here we have a different test-statistic, though the tests are equivalent.

For testing $H_0 : \theta_1 = \theta_2$, Mann and Whitney defined the statistic

$$U = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \phi(x_i, y_j) \text{ where } \phi(x_i, y_j) = \begin{cases} 1 \text{ if } x_i < y_j \\ 0 \text{ otherwise} \end{cases}$$

So, U is the number of times x_i precedes y_j, $\forall i, j$ ($i = 1, 2, \ldots, n_1$, and $j = 1, 2, \ldots, n_2$). In case there are no ties, it can be shown that $W = U + n_2(n_2 + 1)/_2$, n_2 being size of the larger sample. It may be noted that, if U' denotes the number of times $x_i > y_j$, $\forall$ i, j then $U + U' = n_1 n_2$, provided there are no $x = y$ ties.

When the alternative hypothesis $H : \theta_1 < \theta_2$ is true, U will tend to be larger than U'. So while testing $H_0 : \theta_1 = \theta_2$, against $H : \theta_1 < \theta_2$, we reject H_0 for large values of U (or equivalently, for large values of W) or small values of U'. Similarly, while testing $H_0 : \theta_1 = \theta_2$, against $H_0 : \theta_1 > \theta_2$, we reject H_0 for small values of U. Again, for two-sided alternative $H : \theta_1 \neq \theta_2$, we reject H_0 for small values of min (U, U′). For appropriate critical regions at specified level of significance, we use Table VIII of Appendix for n_2 (size of the larger sample) between 9 and 20 and $n_1 \leq 20$. If the computed value of appropriate U [i.e. U for $H : \theta_1 > \theta_2$, U' for $H : \theta_2 > \theta_1$ and min(U, U′) for $H : \theta_1 \neq \theta_2$] is less than or equal to the tabulated value, we reject $H_0 : \theta_1 = \theta_2$ at the stated level of significance. For small values of n_1, n_2 (none larger than 8), Mann and Whitney have given a table of exact probabilities.

For large samples, the statistic

$$U^* = \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (N+1)/12}},$$

which has, under H_0, an asymptotic N(0, 1) distribution as $\min(n_1, n_2) \to \infty$, is taken to perform the normal deviate test.

Illustration 14.33

The following are the Rockwell hardness numbers obtained for five aluminium die castings randomly selected from production lot A and nine from production lot B :

Production lot A :	75, 56, 63, 70, 58
Production lot B :	63, 85, 77, 80, 86, 76, 72, 82, 74

Use the U-test at the 0.05 level of significance to test whether the castings of production lot B are on the average equally hard or whether they are harder than those of production lot A.

Solution : Let θ_1 and θ_2 be the medians of the Rockwell hardness numbers for Production lot A and Production lot B, respectively. We are to test the null hypothesis $H_0 : \theta_1 = \theta_2$, against the alternative hypothesis $H : \theta_2 > \theta_1$ at 0.05 level of significance.

Arrangement of the combined sample data in increasing order is

56,	58,	63,	63,	70,	72,	74,	75,	76,	77,	80,	82,	85,	86.
(A)	(A)	(A)	(B)	(A)	(B)	(B)	(A)	(B)	(B)	(B)	(B)	(B)	(B)

The lot to which a number belongs has been mentioned.

So, U = 9 + 9 + 8 + 8 + 6 = 40 and U′ = 1 + 3 = 4.

From Table VIII of Appendix, we find that for $n_2 = 9$ and $n_1 = 5$ for a one-tail test at the level 0.05, the critical value is 9. Since 4 (the value of U′) is smaller than 9, we reject H_0 and conclude that castings of production lot B are on the average harder than those of production lot A.

14.6.5 Wald-Wolfowitz run test

Suppose we have two independent random samples from two continuous distributions. On the basis of the samples we want to test the null hypothesis H_0 : The two population distributions are identical, against the alternative hypothesis H : The two distributions differ (in any manner).

Let n_1 and n_2 be the sizes of the two samples. We arrange the combined $n_1 + n_2$ observations in order of magnitude. Denoting the observations of the first sample by x's and those of the second by y's, we might have an arrangement of the type

$$y_{(1)}\ x_{(1)}\ x_{(2)}\ x_{(3)}\ y_{(2)}\ y_{(3)}\ x_{(4)}\ \cdots$$

Next, we count the number of runs in the arrangement. A run is a sequence of identical letters (or other kind of symbols) preceded and followed by different letters or no letters at all. Thus, in the above sequence, we have a run of one y followed by a run of three x's, which in turn is followed by a run of two y's, and so on. Let r be the total number of runs in the group of $n_1 + n_2$ observations. Now, if the two distributions are identical, then there would be thorough mingling of x's and y's and consequently r would be large; whereas r would be relatively small if the distributions are not the same. So, we reject H_0 when r is very small.

To perform the test at level α, we are to find r_0 such that $P(r \le r_0) = \alpha$ and reject H_0 if the observed value of r does not exceed r_0.

Tables of critical values of r, based on the sampling distribution of r, are given by Swed and Eisenhart. Any value of r which is equal to or smaller than that shown in Table IX of Appendix is significant at 0.05 level.

If both n_1 and n_2 are larger than 10, or either n_1 or n_2 is larger than 20, the sampling distribution of r is approximately normal with

$$E(r) = \frac{2n_1n_2}{N} + 1$$

and $$\operatorname{Var}(r) = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}, \text{ where } N = n_1 + n_2.$$

Hence, in such cases we can perform an approximate test.

It is to be noted that, since the distributions are assumed to be continuous, no ties should occur. But due to approximation in the measurements, ties may be found in practice. Ties within the same sample have no effect on number of runs. But if there are ties among observations from the two samples, one cannot get a unique value of r. In such cases, one has to break ties in all possible ways and find the corresponding values of r. If the different values of r lead to the same conclusion, then there is no problem; otherwise there is difficulty. When the number of ties between observations from the two samples is large, then run test is not to be recommended.

Illustration 14.34

At the beginning of a year a first grade class was randomly divided into two groups. One group was taught to read using a uniform method and the other group was taught to read using an individual method. At the end of the year, each student was given a reading ability test. The results of two independent random sample of students from the two groups are :

First group : 227, 176, 252, 149, 16, 55, 234

Second group : 202, 14, 165, 171, 292, 271.

Use the Wald-Wilfowitz run test for equality of distribution of scores of two groups.

Solution **:** Our null hypothesis is H_0 : The two distributions are identical and the alternative hypothesis is H : The distributions differ.

We arrange the scores of 7 + 6 = 13 students in order of magnitude, noting the group to which a score belongs :

14	16	55	149	165	171	176	202	227	234	252	271	292
S	F	F	F	S	S	F	S	F	F	F	S	S

Here we have $n_1 = 7$, $n_2 = 6$ and 4 runs of S's and 3 runs of F's, giving $r = 7$. The critical value of r at 5% level, from Table IX of Appendix, is 3. Since the observed value 7 is greater than the critical value 3, we accept the null hypothesis that the population score distributions are identical at 5% level.

14.6.6 Run test for randomness

Another application of the theory of runs is in testing the randomness of a given sample. We use the order in which the observations occur. The total number of runs appearing in the sample of a given size indicates whether the sample is random or not. If there are too 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.

We find the number of runs (r) in the group of observations in the sample. The observations may be heads and tails in a coin tossing experiment, the defective and non-defective items in sampling inspection or the measurements below and above the sample median, and so on. The critical values of r at 5% level are given in Table IX of Appendix. If the observed value of r is equal to or smaller than the tabulated value, the hypothesis of randomness is rejected.

Illustration 14.35

The following is the order in which red (R) and black (B) cards were dealt to a bridge player :

B B B R R R R R B B R R R

Test for randomness at the 0.05 level of significance.

Solution : Here number of B's is $n_1 = 5$ and number of R's is $n_2 = 8$; $n_1 + n_2 = 13$. We have 2 runs of B's and 2 runs of R's, giving total number of runs $r = 2 + 2 = 4$. The critical value of r at 5% level, from Table IX of Appendix, is 3. Since the observed value is greater than the critical value, we accept the hypothesis that red and black cards were dealt at random.

14.6.7 Two-sample median test

Here our problem is same as in the U-test. We have two independent random samples from two continuous distributions which may differ only in their locations and we are to test the null hypothesis that the population distributions are identical against the alternative hypothesis that the distributions have different location parameters i.e. medians.

Let n_{10} and n_{20} be sizes of the two samples.

We arrange the combined $N = n_{10} + n_{20}$ observations in order of magnitude and determine the combined sample median, M(say). Next, we count the number of observations in each sample that are less than M and the number of those that are greater than or equal to M. We can put the numbers in a 2 × 2 contingency table :

	No. of observations		
	$< M$	$\geq M$	Total
Sample 1	n_{11}	n_{12}	n_{10}
Sample 2	n_{21}	n_{22}	n_{20}
Total	n_{01}	n_{02}	N

The exact probability of this configuration is

$$\binom{n_{10}}{n_{11}}\binom{n_{20}}{n_{21}}\Big/\binom{N}{n_{01}}.$$

When the sum of probabilities of the observed configuration and more extreme ones (with fixed marginals) in either direction exceeds the level of significance α, we accept the null hypothesis at level α. The alternative hypothesis may be one-sided, meaning that the median of one population is greater than that of the other. In that case we are to compare the sum of probabilities of the observed configuration and more extreme ones in one direction with the level of significance.

The chief objection to this exact method is the computational labour involved. But when the probability of the observed configuration exceeds the level of significance, we are not required to obtain probabilities of the more extreme cases.

When n_{10} and n_{20}, i.e. sizes of the two samples, are moderately large, say each greater than 10, we can use the frequency χ^2 statistic, where

$$\chi^2 = \frac{(n_{11}n_{22} - n_{21}n_{12})^2 \cdot N}{n_{10}\, n_{20}\, n_{01}\, n_{02}}.$$

This χ^2 has 1 d.f.

If $\chi^2_{\text{observed}} > \chi^2_{\alpha,1}$ we reject the null hypothesis at level α; otherwise we accept it. $\chi^2_{\alpha,1}$ is the upper α-point of the χ^2 distribution with 1 d.f.

Illustration 14.36

The following are the numbers of minutes taken by a random sample of 15 men and 12 women to complete a written test :

Men : 9.9, 7.4, 8.9, 9.1, 7.7, 9.7, 11.8, 9.2, 10.0, 10.2, 9.5, 10.8, 8.0, 11.0, 7.5
Women : 8.6, 10.9, 9.8, 10.7, 9.4, 10.3, 7.3, 11.5, 7.6, 9.3, 8.8, 9.6

Use median test at 0.05 level of significance to decide whether the population distributions of times taken by men and women to complete the test are identical.

Solution : We have to test the null hypothesis H_0 : The population distributions of times taken by men and women are identical against the alternative hypothesis H : The distributions differ in locations.

We arrange the 15 + 12 = 27 observation in order of magnitude :

7.3,	7.4,	7.5,	7.6,	7.7,	8.0,	8.6,	8.8,	8.9,
W	M	M	W	M	M	W	W	M
9.1,	9.2,	9.3,	9.4,	9.5,	9.6,	9.7,	9.8,	9.9,
M	M	W	W	M	W	M	W	M
10.0,	10.2,	10.3,	10.7,	10.8,	10.9,	11.0,	11.5,	11.8
M	M	W	W	M	W	M	W	M

The letters M and W, written below the numbers, indicate whether the time was taken by a man or a woman. From the arrangement we find that combined sample median is 9.5. After counting the number of observations in each of the two samples that lie below 9.5, we prepare the following 2 × 2 contingencey table :

	No. of observations		Total
	< 9.5	≥ 9.5	
Men	7	8	15
Women	6	6	12
Total	13	14	27

Since size of each sample is greater than 10, we use frequency χ^2 statistic to test the null hypothesis H_0. Here

$$\chi^2 = \frac{(7 \times 6 - 6 \times 8)^2 \cdot 27}{15 \times 12 \times 13 \times 14} \doteqdot 0.0297.$$

This χ^2 has 1 d.f. Since χ^2_{observed} (= 0.0297) < $\chi^2_{0.05,1}$(= 3.841), we accept H_0 at level 0.05, and conclude that distributions of times taken by men and women may be taken to be identical.

EXERCISES

1. The following are measurements of the breaking strength of a certain kind of 2-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 that population median of breaking strength is 160 pounds against the alternative hypothesis that it is greater than 160 pounds at 0.05 level of significance.

2. To determine the effectiveness of a new traffic-control system, the number of accidents that occurred at independently and randomly selected 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. (The populations are not continuous, but this does not matter so long as zero differences are discarded.)

3. The following are 15 random sample observations drawn independently from a continuous population : 97.5, 95.2, 97.3, 96.0, 96.8, 100.3, 97.4, 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5, 94.9. Use the signed-rank test at the 0.05 level of significance to test whether or not the median of the population is 98.5.