NON-PARAMETRIC TESTS

Lalmohan Bhar
I.A.S.R.I., Library Avenue, New Delhi- 110 012
lmbhar@iasri.res.in

Testing (usually called 'hypothesis testing') play a major role in statistical investigation. In statistical testing, we are concerned with examining the truth or otherwise, of hypothesis (assumptions, claims, guesses, etc.) about some feature(s) of one or more populations. Almost all large and small sample tests such as t, F and χ^2 are based on the assumptions that the parent population (from which the sample is drawn) has a specific distribution, such as normal distribution. The distributions are usually defined through some parameters. Nonparametric tests do not require such assumption. Hence nonparametric tests are also known as distribution free tests. The term nonparametric refers to the fact that there are no parameters involved in the traditional sense of the term parameter used generally. Nonparametric test statistics utilize some simple aspects of sample data such as the signs of measurements, order relationships or category frequencies. Therefore, stretching or compressing the scale does not alter them. As a consequence, the null distribution of the nonparametric test statistic can be determined without regard to the shape of the parent population distribution.

The inferences drawn from tests based on the parametric tests such t, F and χ^2 may be seriously affected when the parent population distributions is not normal. These effects could be more if when sample size is small. Thus when there is doubt about the distribution of the parent population, a nonparametric method should be used. In many situations particularly in social and behavioral sciences observations are difficult or impossible to take on numerical scales. Nonparametric tests are well suited under such situations.

First step in statistical testing is formulation of a hypothesis. A hypothesis is a statement about the population. Its plausibility is evaluated on the basis of information obtained by sampling from the population. A test generally involves two hypotheses. An assertion about the population in faviour of the 'existing' situations is taken as null hypothesis and denoted as H_0. The negation of the null hypothesis is known as alternative hypothesis and denoted as H_1. H_1 plays a decisive role in classifying a test as one-sided or two sided. We first develop a statistic T(say) on the basis of the sample observations. The statistic T decides whether to reject or accept the null hypothesis. Usually T follows some distribution. Based on this distribution the range of T is divided into two groups; the critical region and the region of acceptance. If the sample point falls in critical region , we reject null hypothesis. The size of the critical region depends the risk we wish to incur which ultimately gives the significance level of the test. It is denoted by α. It represent the probability of rejecting the null hypothesis when it is triue, also known as Type I error. Type II error is the probability of rejecting H_0 when H_1 is true and denoted by β. Commonly used significance levels are 5% and 1% (α = .05 and .01). Finally, a conclusion is drawn on the basis of the value of T falling or not falling in the critical region. Some commonly used nonparametric tests are discussed in the sequel.

1. Run Test for Randomness

Run test is used for examining whether or not a set of observations constitutes a random sample from an infinite population. Test for randomness is of major importance because the assumption of randomness underlies statistical inference. In addition, tests for randomness are important for time series analysis. Departure from randomness can take many forms.

H_0: Sample values come from a random sequence

H_1: Sample values come from a non-random sequence

Test statistic: Let r be the number of runs (a Run is a sequence of sign of same kind bounded by signs of other kind). For finding the number of runs, the observations are listed in their order of occurrence. Each observation is denoted by a '+' sign if it is more than the previous observation and by a '-' sign if it is less than the previous observation. Total number of runs up (+s) and down (-) is counted. Too few runs indicate that the sequence is not random (has persistency) and too many runs also indicate that the sequence is not random (is zigzag).

Critical value: Critical value for the test is obtained from the table for a given value of n and at desired level of significance (α). Let this value is r_c.

Decision rule: If r_c (lower) $\leq r \leq r_c$ (upper) accept H_0. Otherwise reject H_0.

Tied values: If an observation is equal to its preceding observation denote it by zero. While counting the number of runs ignore it and reduce the value of n accordingly.

Large sample sizes: When sample size is greater than 25 the critical value r_c can be obtained using a normal distribution approximation.

The critical values for two-sided test at 5% level of significance are

r_c (lower) $= \mu - 1.96\,\sigma$; r_c (upper) $= \mu + 1.96\,\sigma$

For one-sided tests, these are

r_c (left tailed) $= \mu - 1.65\,\sigma$, if $r \leq r_c$, reject H_0

r_c (right tailed) $= \mu + 1.65\,\sigma$, if $r \geq r_c$, reject H_0

where $\mu = \dfrac{2n-1}{3}$ and $\sigma = \sqrt{\dfrac{16n-29}{90}}$

Example 1: Data on value of imports of selected agricultural production inputs from U.K. by a county (in million dollars) during recent 12 years is given below: Is the sequence random?

5.2	5.5	3.8	2.5	8.3	2.1	1.7	10.0	10.0	6.9	7.5	10.6

H_0: Sequence is random. H_1: Sequence is not random.

5.2	5.5	3.8	2.5	8.3	2.1	1.7	10.0	10.0	6.9	7.5	10.6
	+	-	-	+	-	-	+	0	-	+	+

Herte n = 11, the number of runs r = 7. Critical nvalues for α = 5% (two sided test) from the table are r_c (lower) = 4 and r_c (upper) = 10. Since r_c (lower) $\leq r \leq r_c$ (upper), i.e., observed r lies between 4 and 10 the H_0 is accepted. The sequence is random.

2. WALD-WOLFOWITZ Two-Sample Run Test

Wald –wolfowitz run test is used to examine whether two random samples come from populations having same distribution. This test can detect differences in averages or spread or any other important aspect between the two populations. This test is efficient when each sample size is moderately large (greater than or equal to 10).

H_0: Two sample come from populations having same distribution
H_1: Two sample come from populations having different distributions

Test statistic: Let r denotes the number of runs. To obtain r, list the $n_1 + n_2$ observations from two samples in order of magnitude. Denote observations from one sample by x's and other by y's. Count the number of runs.

Critical Value: Difference in location results in few runs and difference in spread also result in few number of runs. Consequently, critical region for this test is always one-sided. The critical value to decide whether or not the number of runs are few, is obtained from the table. The table gives critical value r_c for n_1 (size of sample 1) and n_2 (size of sample 2) at 5% level of significance.

Decision rule: If $r \le r_c$ reject H_0.

Tie: In case x and y observations have same value place the observation $x(y)$ first if run of $x(y)$ observation is continuing.

Large sample sizes: For sample sizes larger than 20 critical value r_c is given below.

$r_c = \mu - 1.96\,\sigma$ at 5% level of significance

where $\mu = 1 + \dfrac{2n_1n_2}{n_1+n_2}$ and $\sigma = \sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$

Example 2: To determine if a new hybrid seeding produces a bushier flowering plant, following data was collected. Examine if the data indicate that new hybrid produces larger shrubs than the current variety?

Shrubs Girth (in inches)

Hybrid	***x***	31.8	32.8	39.2	36.0	30.0	34.5	37.4
Current variety	***y***	35.5	27.6	21.3	24.8	36.7	30.0	

H_0: x and y populations are identical
H_1: There is some difference in girth of x and y shrubs.

Consider the combined ordered data.

21.3	24.8	27.6	30.0	30.0	31.8	32.8	34.5	35.5	36.0	36.7	37.4	39.2
y	*y*	*y*	*y*	*x*	*x*	*x*	*x*	*y*	*x*	*y*	*x*	*x*

Test statistic r =6 (total number of runs). For n_1 = 7 and n_2 = 6, critical value r_c at 5% level of significance is 3. Since $r > r_c$, we accept H_0 that x and y have identical distribution.

3. Median Test for Two Samples

To test whether or not two samples come from same population median test is used. It is more efficient than the run test but each sample should be of size 10 at least. In this case, the hypothesis to be tested is

H_0 : Two samples come from populations having same distribution.

H_1 : Two samples come from populations having different distribution.

Test Statistic: χ^2 (Chi-square). To test the value of test statistics two samples of sizes n_1 and n_2 are combined. Median M of the combined sample of size $n = n_1 + n_2$ is obtained. Number of observations below and above the median M for each sample is determined. This is then analyzed as a 2 × 2 contingency table in the manner given below.

	Number of observations		
	Sample 1	**Sample 2**	**Total**
Above Median	a	b	a+b
Below Median	c	d	c+d
	a+c= n_1	b+d = n_2	n = a+b+ c+d

Test Statistic: $\chi^2 = \dfrac{(ad-bc)^2(a+b+c+d)}{(a+c)(b+d)(a+b)(c+d)}$

Decision rule: if $\chi^2 \geq \chi_c^2$ reject H_0 otherwise accept it.

Tie: ties are ignored and *n* is adjusted accordingly.

Note: This test can be extended to *k* samples. Number of observations below and above the combined median *M* from a 2 × *k* contingency table.

Example 3: Perform a median test on the problem of example 1 for the testing that the two samples come from same population.

H_0 : *x* and *y* populations are identical.

H_1 : There is some difference in girth of *x* and *y* shrubs.

Seventh value 32.8 is the median of combined ordered sequence.

	Number of observations		
	x	*y*	Total
Above M	4	2	6
Below M	2	4	6
	6	6	12

$$\chi^2 = \frac{12(16-4)^2}{6.6.6.6} = \frac{4}{3} = 1.33.$$

Since $\chi^2 = 1.33 < \chi_c^2 = 3.84$, H_0 is accepted. It is concluded that two samples come from the same population. There is no significance difference in the girth of hybrid and current variety of shrub.

Note: This example is simple to demonstrate test procedure. In real situation *n* should be at least 20 and each cell frequency at least 5.

4. Sign Test for Matched Pairs

In many situations, comparison of effect of two treatments is of interest but observations occur in pairs. Thus the two samples are not truly random. Because of such pair-wise dependence ordinary two sample tests are not appropriate. In such situations when one member of the pair is associated with the treatment A and the other with treatment B, nonparametric sign test has wide applicability. It can be applied even when qualitative data are available. As the name suggests it is based on the signs of the response differences D_i. If ith pairs of observations is denoted by (x_i, y_i) where x is the effect of treatment A and y to B then $D_i = x_i - y_i$. The hypothesis to be tested is

H_0 : No difference in the effect of treatments A and B.
H_1 : A is better than B.

Test Statistic: Let S be the number of '-' signs.

Critical value: Critical value S_c corresponding to *n* the number of pairs is given in Table 3. Significance level is given by α_1 as critical region is one sided (left tailed).

Decision rule: If $S \leq S_c$ reject H_0, other wise accept H_0.

Tie: In case two values of a pair are equal, reject that pair and reduce the number of observations accordingly.

Note: In case alternative H_1 is that there is some difference in effect of A and B, S represents either the number of negative signs or the number of positive signs whichever turn out to be smaller. A critical region is two sided and significant level is given by α_2 for finding S_c.

Example 4: In a market study, two brands of lemonade were compared. Each of 50 judges tasted two samples, one of brand A and one of brand B with the following results. 35 preferred brand A, 10 preferred B, and 5 could not tell the difference. Thus n = 45 and S = 10. Assuming α_1 = 5%, critical value S_c = 16 from Table 3. Since $S < S_c$, we reject H_0 of no difference in favour of the alternative H_1 that the brand A is preferred.

5. WILCOXON Signed Rank Test for Matched Pairs

In situations where there is some kind of pairing between observations in the two samples ordinary two sample tests are not appropriate. Signed rank tests are useful in such situations. When observations are measured data, signed rank test is more efficient than sign test as it takes account of the magnitude of the observed differences, if the difference between the response of the two treatments A and B is to be tested the test hypothesis is

H_0 : No difference in the effect of treatments A and B.
H_1 : Treatment A is better than B.

Test Statistic: T represents the sum of ranks with negative signs. For calculating T, obtain the differences $D_i = x_i - y_i$ where x_i's are response of treatment A and y_i's of treatment B. Rank the absolute values of differences. Smallest give rank 1. Ties are assigned average ranks. Assign to each rank sign of observed difference. Obtain the sum of negative ranks.

Critical value: T_c is given in Table 4 for *n* no. of pairs. Significance level is given by α_1 as critical region is one sided.

Decision rule: $T \leq T_c$ reject H_0, other wise accept it.

Tie: Discard the pair for which difference = 0 and reduce *n* accordingly. Equal differences are assigned average ranks.

Example 5: Blood pressure reading of ten patients before and after medication for reducing the blood pressure are as follows. Test the null hypothesis of no effect against the alternative that medication is effective.

Patient		1	2	3	4	5	6	7	8	9	10
Before treatment	x	86	84	78	90	92	77	89	90	90	86
After treatment	y	80	80	92	79	92	82	88	89	92	83
Differences		6	4	-14	11	0	-5	1	1	-2	3
Rank		7	5	9	8	Discard	6	1.5	1.5	3	4
Sign		+	+	-	+	Discard	-	+	+	-	+

Rank sum of negative differences = 3+6+9 = 18. therefore value of test statistic T= 18. for n = 9 and α_1 =5 % T_c = 8 from table 4. Since $T>T_c$ null hypothesis of no effect of medication is accepted.

6. KOLMOGOROV-SMIRNOV Test

In situations where there is unequal number of observations in two samples Kolmogorov-Smirnov test is appropriate. This test is used to test whether there is any significance difference between two treatments A and B (say). The test hypothesis is

H_0 : No difference in the effect of treatments *A* and *B*.

H_1 : There is some difference in the effect of treatments *A* and *B*.

Test Statistic: The test statistic is $D_{m,n} = \sup\left|F_m(x) - G_n(x)\right|$, *F* and *G* are the sample empirical distributions of sample observations of two samples respectively with respective sample sizes *m* and *n*. $F(x_i)$ is calculated as the average number of sample observations of the first sample that are less than x_i. Similarly $G(x_i)$ is calculated. $D_{m,n}$ is largest value of the absolute difference between F(x) and G(x).

Critical value: Tabulated value of $D_{m,n}$ is available for different values of m, n and for different level of significance. is given in Table 4 for *n* no. of pairs. Significance level is given by α_1 as critical region is one sided.

Decision rule: If the calculated value of $D_{m,n}$ is greater than the Tabulated value of $D_{m,n}$, H_0 is rejected otherwise it is accepted.

Example 6: The following data represent the lifetimes (hours) of batteries for different brands:

Brand A	40	30	40	45	55	30
Brand B	50	50	45	55	60	40

Are these brands different with respect to average life?

We first calculate the sample empirical distributions of two samples:

x	$F_6(x)$	$G_6(x)$	$\lvert F_6(x)-G_6(x)\rvert$
30	2/6	0	2/6
40	4/6	1/6	3/6
45	5/6	2/6	3/6
50	5/6	4/6	1/6
55	1	5/6	1/6
60	1	1	0

$D_{6,6} = \sup\lvert F_6(x)-G_6(x)\rvert$ = 3/6, from Table the critical value for m = n = 6 at level α = .05 is 4/6. Since the calculated value of $D_{m,n}$ is not greater than the Tabulated value, H0 is not rejected and it is concluded that the average length of life for two brands is the same.

Some Selected references

Bhattacharya, G.K. and Johnson, R.A. *Statistics concepts and Methods.* New York, John Wiley and Sons. pp 505-521.

Neave, H.R. and Worthington, P.L. *Distribution free tests.* London Unwin Hyman, pp 161-164,328,337-341.

Neter, J.W.W. and Whitemore, G.A. *applied Statistics.* London, Allyn and Bacon Inc. pp 360-388.

Ostle, B. *Statistics in Reasersch.* Ames. Iowa, USA. The Iowa State University. pp 466-473.

PRACTICAL EXERCISES

1. Maximum level of a lake each year for a period of 20 years is given below. It is desired to test (a) whether the sequence is generated by a random process, or (b) the process contains a trend. The presence of trend will have significant environmental policy implications.

Year	Level (Above 190 meters)	Year	Level (Above 190 meters)
1	6.6	11	6.0
2	6.5	12	5.8
3	6.4	13	5.9
4	6.5	14	5.6
5	6.4	15	5.5
6	6.4	16	5.3
7	6.3	17	5.1
8	6.2	18	5.3
9	6.1	19	5.4
10	5.9	20	5.2

2. Seasonal rainfall at two meteorological observations of a district is given below. Examine by using Run test and Median test whether the rainfall of two observations can be considered as same.

	Seasonal rainfall(cm) observations	
Year	**A**	**B**
1985	25.34	24.31
1986	49.35	45.13
1987	39.62	42.83
1988	42.90	46.94
1989	57.66	57.50
1990	24.89	30.70
1991	50.63	48.37
1992	38.47	38.45
1993	43.25	44.00
1994	50.83	50.00
1995	22.02	

3. An experiment was performed to determine if self fertilized and cross fertilized plants have different growth rates. Pairs of plants one self and other cross fertilized were planted in 15 pots. Their heights were measured after specified period of time.
   (a) perform the sign test ot determine whether there is any difference in the growth rates of self fertilized and cross fertilized plants.
   (b) Perform Wilcoxon signed rank test to determine of crossed plants jave a higher growth rate.

	Plant height (cms)			Plant height (cms)	
Pair	Crossed fertilized	Self fertilized	Pair	Crossed fertilized	Self fertilized
1	45.5	40.0	9	41.2	41.2
2	40.0	42.3	10	42.7	42.0
3	42.8	41.2	11	43.3	42.0
4	41.6	41.3	12	41.0	40.7
5	37.9	36.7	13	46.0	43.5
6	42.5	38.0	14	39.2	40.6
7	44.1	39.8	15	44.3	42.5
8	40.7	38.9			

Table 1: Critical values for runs up and down test

	$\alpha_1 = 5$ % $\alpha_2 = 10$ %		$\alpha_1 = 2.5$ % $\alpha_2 = 5$ %		$\alpha_1 = 1$ % $\alpha_2 = 2$ %		$\alpha_1 = 0.5$ % $\alpha_2 = 1$ %	
n	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
3	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-
5	1	-	1	-	-	-	-	-
6	1	-	1	-	1	-	1	-
7	2	-	2	-	1	-	1	-
8	2	-	2	-	2	-	1	-
9	3	8	3	-	3	-	2	-
10	3	9	3	-	3	-	2	-
11	4	10	4	10	3	-	3	-
12	4	11	4	11	4	-	3	-
13	5	12	5	12	4	12	4	-
14	6	12	5	13	5	13	4	13
15	6	13	6	14	5	14	4	14
16	7	14	6	14	6	15	5	15
17	7	15	7	15	6	16	6	16
18	8	15	7	16	7	16	6	17
19	8	16	8	17	7	17	7	18
20	9	17	8	17	8	18	7	18
21	10	18	9	18	8	19	8	19
22	10	18	10	19	9	20	8	20
23	1	19	10	20	10	20	9	21
24	1	20	11	20	10	21	10	22
25	12	21	11	21	11	22	10	22

α_1 : Significance level for one sided test
α_2 : Significance level for two sided test

Source: Distribution Free Tests by H.R. Neave and P.L. Worthington. London, Unwin Hyman.

Table 2: Critical values for the two sample run test.

n_2 \ n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	2	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	2	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Significance level 5%

Source: Statistics in Research by Borten Ostle. Ames. Iowa USA. Iowa State University Press.

Table 3: Critical values for the Sign test (Matched pairs)

n	α_1 / α_2	5 % / 10 %	2.5 % / 5 %	1 % / 2 %	0.5 % / *1 %*	*n*	α_1 / α_2	5 % / 10 %	2.5 % / 5 %	1 % / 2 %	0.5 % / 1 %
1		-	-	-	-	26		8	7	6	6
2		-	-	-	-	27		8	7	7	6
3		-	-	-	-	28		9	8	7	6
4		-	-	-	-	29		9	8	7	7
5		0	-	-	-	30		10	9	8	7
6		0	0	-	-	31		10	9	8	7
7		0	0	0	-	32		10	9	8	8
8		1	0	0	0	33		11	10	9	8
9		1	1	0	0	34		11	10	9	9
10		1	1	0	0	35		12	11	10	9
11		2	1	1	0	36		12	11	10	9
12		2	2	1	1	37		13	12	10	10
13		3	2	1	1	38		13	12	11	10
14		3	2	2	1	39		13	12	11	11
15		3	3	2	2	40		14	13	12	11
16		4	3	2	2	41		14	13	12	11
17		4	4	3	2	42		15	14	13	12
18		5	4	3	3	43		15	14	13	12
19		5	4	4	3	44		16	15	13	13
20		5	5	4	3	45		16	15	14	13
21		6	5	4	4	46		16	15	14	13
22		6	5	5	4	47		17	16	15	14
23		7	6	5	4	48		17	16	15	14
24		7	6	5	5	49		18	17	15	15
25		7	7	6	5	50		18	17	16	15

α_1 : Significance level for one sided test
α_2 : Significance level for two sided test

Source: Distribution Free Tests by H.R. Neave and P.L. Worthington. London, Unwin Hyman.

Table 4 Critical values for the Wilcoxon signed rank test

	α_1	5 %	2.5 %	1 %	0.5 %		α_1	5 %	2.5 %	1 %	0.5 %
n	α_2	10 %	5 %	2 %	*1 %*	*n*	α_2	10 %	5 %	2 %	1 %
1		-	-	-	-	26		110	98	84	75
2		-	-	-	-	27		119	107	92	83
3		-	-	-	-	28		130	116	101	91
4		-	-	-	-	29		140	126	110	100
5		0	-	-	-	30		151	137	120	109
6		2	0	-	-	31		163	147	130	118
7		3	2	0	-	32		175	159	140	128
8		5	3	1	0	33		187	170	151	138
9		8	5	3	1	34		200	182	162	148
10		10	8	5	3	35		213	195	173	159
11		13	10	7	5	36		227	208	185	171
12		17	13	9	7	37		241	221	198	182
13		21	17	12	9	38		256	235	211	194
14		25	21	15	12	39		271	239	224	207
15		30	25	19	15	40		286	264	238	220
16		35	29	23	19	41		302	279	252	233
17		41	34	27	23	42		319	294	266	244
18		47	40	32	27	43		336	310	281	261
19		53	46	37	32	44		353	327	296	276
20		60	52	43	37	45		371	343	312	291
21		67	58	49	42	46		389	361	328	307
22		75	65	55	48	47		407	278	345	322
23		83	73	62	54	48		426	296	362	339
24		91	81	69	61	49		446	415	379	355
25		100	89	76	68	50		466	434	397	373

α_1 : Significance level for one sided test
α_2 : Significance level for two sided test

Source: Distribution Free Tests by H.R. Neave and P.L. Worthington. London, Unwin Hyman.