6

Nonparametric Methods for Data Analysis

Eldho Varghese[1] and Cini Varghese[2]
[1]ICAR-Central Marine Fisheries Research Institute, Kochi
[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
eldho.varghese@icar.gov.in ; cini.varghese@icar.gov.in

1. Introduction

A parametric test specifies certain conditions about the distribution of responses in the population from which the research sample was drawn. The meaningfulness of the results of a parametric test depends on the validity of these assumptions. A nonparametric test is based on a model that specifies very general conditions and none regarding the specific form of the distribution from which the sample was drawn. Hence nonparametric tests are also known as distribution free tests. Certain assumptions are associated with most nonparametric statistical tests, but these are fewer and weaker than those of parametric tests.

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. These tests have the obvious advantage of not requiring the assumption of normality or the assumption of homogeneity of variance. They compare medians rather than means and, as a result, if the data have one or two outliers, their influence is negated.

Advantages of nonparametric tests

- Non-parametric methods are used with all scales
- When sample size is very small, there may be no alternative to use a nonparametric test unless the population distribution is known exactly
- They are easier to learn and compute
- Fewer assumptions are made
- Due to the reliance on fewer assumptions, non-parametric methods are more robust
- Need not involve population parameters
- Results may be as exact as parametric procedures

Disadvantages of nonparametric tests

- There may be wastage of information
- Parametric models are more efficient if data permit.
- It is difficult to compute by hand for large samples
- Tables are not widely available
- In cases where a parametric test would be appropriate, nonparametric tests have less power. In other words, a larger sample size can be required to draw conclusions with the same degree of confidence.

The inferences drawn from tests based on the parametric tests such as t, F and χ^2 may be seriously affected when the parent population's distribution is not normal. The adverse effect could be more 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 and a suitable nonparametric test is an alternative under such situations. Some commonly used nonparametric tests are discussed in the sequel.

2. Binomial Test

It tests whether the proportion of people in one of two categories is different from a specified amount. For example, if we ask people to select one of two pets, either a cat or a dog, we could determine if the proportion of people who selected a cat is different from 0.5.

i.e., the proportion of people who selected a cat is different from the proportion of people who selected a dog.

Binomial test using R: binom.test(x, n, p = 0.5)

Suppose for example x= 8 and n=38, the test is as follows:

```
> binom.test(8, 38, p = 0.5,)


Exact binomial test
data:  8 and 38
number of successes = 8, number of trials = 38, p-value = 0.000472
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.09554112 0.37318828
sample estimates:
```

```
probability of success
0.2105263
```

3. Chi-square Test

A random sample of students enrolled in Statistics 101 at ABC University was taken. It consists of the following: there are 25 freshman in the sample, 32 sophomores, 18 juniors, and 20 seniors. Test the null hypothesis that freshman, sophomores, juniors, and seniors are equally represented among students signed up for Stat 101.

This is a goodness of fit test with equal expected frequencies.

Chi-square test using R:

The "p" vector does not need to be specified, since equal frequencies is the default...

```
> chisq.test(c(25,32,18,20))

Chi-squared test for given probabilities


data:  c(25, 32, 18, 20)
X-squared = 4.9158, df = 3, p-value = 0.1781
```

Now the expected frequencies are no longer equal, so a "p" vector must be specified. A little algebra leads us to the expected probabilities of 1/3, 1/3, 1/6, and 1/6.

```
> ofs <- c(25,32,18,20)
>null.probs <- c(1/3,1/3,1/6,1/6)
> chisq.test(ofs, p=null.probs)

Chi-squared test for given probabilities

data:  ofs
X-squared = 2.8, df = 3, p-value = 0.4235
```

4. 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 signs 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 (+) 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 be 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 4.1: {Example 1 in the E-book available at http://iasri.res.in/ebook/EBADAT**}** 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	-	+	+

Here n = 11, the number of runs r = 7. Critical n values 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, H_0 is accepted. The sequence is random.

Run test using R
First download and install the package named *lawstat* . Then use

```
runs.test(x)

> y<-c(5.2, 5.5,3.8,2.5,8.3,2.1,1.7,10,10,6.9,7.5,10.6)
> runs.test(y)

Runs Test - Two sided

data:  y
Standardized Runs Statistic = -1.8166, p-value = 0.06928
```

5. 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 denote 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 $r \leq r_c$, 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 $\mu = 1 + \frac{2n_1n_2}{n_1+n_2}$ and $\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$

Example 5.1: {Example 1 in the E-book available at http://iasri.res.in/ebook/EBADAT**}** 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 and conclude that x and y have identical distribution.

6. 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

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+c)(b+d)(a+b)(c+d)}$$

Decision Rule: if $\chi^2 \geq \chi^2_c$, reject H_0 otherwise accept it.

Tie: Ties are ignored and n is adjusted accordingly.

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

Example 6.1: **{**Example 1 in the E-book available at http://iasri.res.in/ebook/EBADAT**}** Perform a median test on the problem of Example 3.1 for 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 χ^2 =1.33 < χ_c^2 =3.84, H_0 is accepted. It is concluded that two samples come from the same population. There is no significant difference in the girth of hybrid and current variety of shrub.

*Remark***:** This example is for demonstrating the test procedure. In real situation n should be at least 20 and each cell frequency at least 5.

<u>R code for Median test</u>

```
# Median comparison test: count samples above and not above the joint median.
> x1 <- c(1.1, 2.1, 1.7, 1.6, 1.9, 1.3)
> x2 <- c(1.0, 1.2, 0.7, 0.6, 0.9, 0.5)
> m <- median(c(x1,x2))     # joint median
> f11 <- sum(x1>m)          # Pop.1 samples above median
> f12 <- sum(x2>m)
> f21 <- sum(x1<=m)         # Pop.1 samples below or at median
> f22 <- sum(x2<=m)
> table <- matrix(c(f11,f12,f21,f22), nrow=2,ncol=2)  # 2x2 contingency table
> chisq.test(table)
```

```
Pearson's Chi-squared test with Yates' continuity correction
data:  table
X-squared = 3, df = 1, p-value = 0.08326

Warning message:

In chisq.test(table) : Chi-squared approximation may be incorrect
```

7. 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 i^{th} pair of observation 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, otherwise accept H_0.

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

Remark: In case, if the 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. Critical region is two sided and significance level is given by α_2 for finding S_c.

Example 7.1: 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.

R code for Sign test (Here different data set has been used)

```
> #Sign test
> x <- c(9, 5, 9 ,10, 13, 8, 8, 13, 18, 30)
> y <- c(10, 6, 9, 8, 11, 4, 1, 3, 3, 10)
> 2*(1-pbinom(7-1,9,0.5))
[1] 0.1796875
```

Hence accept the null hypothesis

8. 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 number of pairs. Significance level is given by α_1 as critical region is one sided.

Decision Rule: $T \le T_c$ 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 8.1: {Example 1 in the E-book available at http://iasri.res.in/ebook/EBADAT**}** Blood pressure reading of ten patients before and after medication for reducing the blood pressure are as follows:

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	-	+	+	-	+

Test the null hypothesis of no effect against the alternative that medication is effective.

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.

R code for Wilcoxon Signed Rank test

First load the package named *'exactRankTests'*

```
> b<-c(80,80,92,79,92,82,88,89,92,83)
> a<-c(86,84,78,90,92,77,89,90,90,86)
> wilcox.exact(a,b, paired = TRUE, alternative = "two.sided")
```

```
Exact Wilcoxon signed rank test
data:  a and b
V = 27, p-value = 0.6328
alternative hypothesis: true mu is not equal to 0
```

9. 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 significant 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 $D_{m,n}$ = $\sup|F_m(x) - G_n(x)|$, 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 levels of significance and is given in Table 4 for n number 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 9.1: {Example 1 in the E-book available at http://iasri.res.in/ebook/EBADAT**}** 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 as follows:

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

$D_{6,6}$ = $\sup|F_6(x) - G_6(x)|$ = 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, H_0 is not rejected and it is concluded that the average length of life for two brands is the same.

R code for KS two sample test

```
> a<-c(40,30,40,45,55,30 )
> b<-c(50,50,45,55,60,40 )
> ks.test(a,b)
```

```
Two-sample Kolmogorov-Smirnov test
data:  a and b
D = 0.5, p-value = 0.4413
alternative hypothesis: two-sided
Warning message:
In ks.test(a, b) : cannot compute exact p-value with ties
```

10. Kruskal-Wallis test

This test is appropriate for use under the following circumstances: (a) If somebody wants to compare three or more conditions; (b) each condition is performed by a *different* group of participants; i.e. you have an independent-measures design with three or more conditions. (c) data do not meet the requirements for a parametric test. (i.e. use it if the data are not normally distributed; if the variances for the different conditions are markedly different; or if the data are measurements on an ordinal scale).

If the data meet the requirements for a parametric test, it is better to use a one-way independent-measures Analysis of Variance (ANOVA) because it is more powerful than the Kruskal-Wallis test.

Example 10.1: Does physical exercise alleviate depression? Here, some individuals are randomly allocated to one of three groups: no exercise; 20 minutes of jogging per day; or 60 minutes of jogging per day. At the end of a month, ach individual is asked to rate how depressed they now feel, on a *Likert scale* that runs from 1 ("totally miserable") through to 100 (ecstatically happy").

Rating on depression scale

No exercise	Jogging for 20 minutes	Jogging for 60 minutes
23	22	59
26	27	66
51	39	38
49	29	49
58	46	56
37	48	60
29	49	56
44	65	62

R code for KW one ANOVA

Import the data saved in Excel into R by loading the packages "xlsxjars rjava xlsx".

```
> mydata=read.xlsx("E://kW test.xlsx", sheetName = "Sheet1")
> fix(mydata)
> kruskal.test(response ~ group, data = mydata)
```

```
Kruskal-Wallis rank sum test
data:  response by group
Kruskal-Wallis chi-squared = 7.2903, df = 2, p-value = 0.02612
```

11. Friedman’s test

It is a nonparametric statistical test for testing whether samples originate from the same distribution. It is used for comparing more than two samples that are related. When the Friedman’s test leads to significant results, then at least one of the samples is different from the other samples.

Example 11.1: A researcher wants to examine whether music has an effect on the perceived psychological effort required to perform an exercise session. The dependent variable is "perceived effort to perform exercise" and the independent variable is "music type", which consists of three categories: "no music", "classical music" and "dance music". To test whether music has an effect on the perceived psychological effort required to perform an exercise session, the researcher recruited 12 runners who each ran three times on a treadmill for 30 minutes. For consistency, the treadmill speed was the same for all three runs. In a random order, each subject ran: (a) listening to no music at all; (b) listening to classical music; and (c) listening to dance music. At the end of each run, subjects were asked to record how hard the running session felt on a scale of 1 to 10, with 1 being easy and 10 extremely hard. A Friedman test was then carried out to see if there were differences in perceived effort based on music type.

No Music	Classical Music	Dance Music
8	8	7
7	6	6
6	8	6
8	9	7
5	8	5
9	7	7
7	7	7
8	7	7
8	6	8
7	6	6

7	8	6
9	9	6

R code for Friedman test

```
data<-matrix(c(8,8,7,
        7,6,6,
        6,8,6,
        8,9,7,
        5,8,5,
        9,7,7,
        7,7,7,
        8,7,7,
        8,6,8,
        7,6,6,
        7,8,6,
        9,9,6),
      nrow = 12,
      byrow = TRUE,
      dimnames = list(1 : 12,c("No Music", "Classical Music", "Dance Music")))
friedman.test(data)
```

```
Friedman rank sum test
data:  data
Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
```

It shows a statistically significant difference between the mean ranks of the related groups.

Pairwise comparison

```
install.packages("PMCMRplus")
library(PMCMRplus)
frdAllPairsConoverTest(Yield,Variety,Replication, data = Friedman,p.adjust.method = "BH")
## Eisinga et al. 2017
 frdAllPairsExactTest(y=y, p.adjust = "bonferroni")
 ## Conover's test
 frdAllPairsConoverTest(y=y, p.adjust = "bonferroni")
 ## Nemenyi's test
 frdAllPairsNemenyiTest(y=y)
 ## Miller et al.
```

```
frdAllPairsMillerTest(y=y)

## Siegel-Castellan

frdAllPairsSiegelTest(y=y, p.adjust = "bonferroni")
```

References

- Bhattacharya, G.K. and Johnson, R.A. Statistics concepts and Methods. New York, John Wiley and Sons. 505-521.
- Neave, H.R. and Worthington, P.L. Distribution free tests. London Unwin Hyman, 161-164, 328, 337-341.
- Ostle, B. Statistics in Research. Ames. Iowa, USA. The Iowa State University. 466-473.
- Siegel, Sidney and Castellan, N.J. (1988). Nonparametric Statistics for the Behavioral Sciences (2nd Edn.). Mcgraw-Hill international Edition, USA.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
- http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.tut%2Fintrotut2.htm