Practical No. 11

Aim: Testing of hypothesis using R: Chi-square distribution, Sign test, Wilcoxon sign rank test.

Chi-square distribution:

Code: chisq.test(y, conf.level=C, correct=F), where C is the confidence level.

Example 9.25 : *Use the following data to test whether the attributes condition of home and condition of child are independent.*

Condition of home → Condition of child	Clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

Solution : R-Code with output

```
> cat("Enter observed frequencies columnwise \n")
> x=scan()
> m = 3; n = 2
> y=matrix(x,nrow=m,ncol=n)
> cat("H0 : The two attributes are independent\n")
> cat("H1 : The two attributes are dependent\n")
> rt =chisq.test(y,correct=F)
>rt
  Pearson's Chi-squared test
data: y
X-squared = 25.6463, df = 2, p-value = 2.698e-06
```

Here given level of significance α = 0.05 is less than p-value = 0.416. Hence we accept H_0. Therefore the attributes condition of home and condition of child are independent.

Questions for practice:

Example 9.26 : *Following table provides data with regard to stature of the fathers and their first SO4 at the age of 25 years.*

Stature of Sons	Stature of Fathers	
	Tall	Short
Tall	8	2
Short	7	6

Test whether the stature of sons is independent of the stature of the fathers.

Contingency table of the handedness of a sample of Americans and Canadians

	Right-handed	Left-handed
American	236	19
Canadian	157	16

A chi-square test (a test of independence) can test whether these observed frequencies are significantly different from the frequencies expected if handedness is unrelated to nationality.

Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to determine their preferred political party. The results of the survey are shown in the table below:

	Republican	Democrat	Independent	Total
Male	100	70	30	200
Female	140	60	20	220
Total	240	130	50	440

To see if gender is linked to political party preference, perform a Chi-Square test

	Obese	Nonobese	Total
Exercise regularly	7	20	27
Don't exercise regularly	15	8	23
Total	22	28	50

A psychologist wishes to test if preference of method of learning differs with gender. He asks a group of 146 individuals their preferred method of learning. Below is a table of the results.

Perform a test to see if a relationship exists.

	Male	Female	TOTAL
Visual	23	17	40
Auditory	13	35	48
Kinaesthetic	30	28	58
TOTAL	66	80	146

NON PARAMETRIC TESTS:

Sign Test and Wilcoxon Test :

11.3 Sign Test

This is the oldest non-parametric test particularly useful in situations in which quantitative measurement is impossible or infeasible, but it is possible to determine whether the observation is greater or smaller than specified value of median.

Consider a random sample of size n drawn from a distribution with distribution function $F_x(.)$. The median M of the distribution of X is unknown and it is required to test $H_0: M = M_0$, where M_0 is specified value of median, against one of the alternative hypotheses $M > M_0$, $M < M_0$ or $M \neq M_0$. While applying sign test, we replace each observation by '+' or '–' sign depending on whether the observation is greater than or less than M_0 respectively.

S+ = The number of positive signs.

S– = The number of negative signs.

Under H_0, S+ has binomial distribution with parameters n and $p = 0.5$.

Test procedure :

(1) If alternative hypothesis H_1 is $M < M_0$, then reject H_0 if $S^+ \le k_\alpha$, where, k_α is largest integer satisfying $P[B(n,1/2) \le k_\alpha] \le \alpha$, where α is specified level of significance.

(2) If alternative hypothesis H_1 is $M > M_0$, then reject H_0 if $S^- \le k_\alpha$

(3) If alternative hypothesis H_1 is $M \neq M_0$, then reject H_0 if $S+ \le k_{\alpha/2}$ or $S- \le k_{\alpha/2}$

Example 11.3 : *Following are the amounts of sulphur oxides (x) (in tons) emitted by large industrial plant in 20 days.*

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26.

Apply sign test to test the hypothesis that population median of X is 21.5 against the alternative hypothesis that it is less than 21.5 at 0.05 level of significance.

Solution : R-Code with output

```
> x=c(17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20,17, 6, + 24, 14, 15, 23, 24, 26)
> m0=21.5
> sp=length(x[x>m0])
> sn=length(x[x<m0])
> n=sp+sn
> pv=pbinom(sp,n,0.5)
> sp
> [1] 9
> pv
  [1] 0.4119015
```

Here sp is the value of s+ and pv is the p-value. If given level of significance is less than p-value, we accept H_0. here $\alpha = .05$ hence accept H_0. So population median of X is 21.5.

Example 11.4 : *Following are data on ten randomly selected specimen of a certain material subjected to stress and the fatigue lifes (in kilocycles) 612, 619, 631, 628, 643, 640, 655, 649, 670, 663.*

Apply sign test to test the hypothesis that population median fatigue life is 625 against the alternative hypothesis that it is greater than 625 at 5% level of significance.

Solution : R-Code with output

```
> x=c(612, 619, 631, 628, 643, 640, 655, 649, 670, 663)
> m0=625
> sp=length(x[x>m0])
> sn=length(x[x<m0])
> n=sp+sn
> pv=pbinom(sn,n,0.5)
  sn
[1] 2
> pv
[1] 0.0546875
```

Here sn is the value of s- and pv is the p-value. If given level of significance is less than p-value, we accept H_0. Here $\alpha = 0.05$ hence accept H_0. So population median of X is 625.

11.4 Sign Test For Paired Data

Sign test can also be applied to paired sample data by taking the difference of observation in each pair. Suppose (x_i, y_i), $i = 1, 2, \ldots, n$ be n paired observations. Let $D_i = x_i - y_i$, $i = 1, 2, \ldots, n$. We consider D_i, $i = 1, 2, \ldots, n$ a random sample from the distribution of $D = x - y$. Then we test the null hypothesis $H_0 : M_D = 0$ against one of the alternatives. $H_1 : M_D \neq 0$, $M_D < 0$ or $M_D > 0$, where, M_D is the median of the distribution of D. The test procedure is similar to that of sign test described earlier for testing $H_0 : M = M_0$.

Example 11.5 : *An I.Q. test was administered to 5 persons before and after they were trained. The results are given below :*

Candidate	*I.Q before training*	*I.Q After training*
1	*110*	*120*
2	*120*	*118*
3	*123*	*125*
4	*132*	*136*
5	*125*	*121*

Use sign test to test whether there is increase in I.Q after the training programme at 5% level of significance.

Solution : Here we want to test

$H_0 : M_D = 0$ against $H_1 : M_D < 0$
where, $D = x - y$
x = IQ before training
y = IQ after training

```
> x=c(110,120,123,132,125)
> y=c(120,118,125,136,121)
> d=x-y
> sp=length(d[d>0])
> sn=length(d[d<0])
> n=sp+sn
> pv=pbinom(sp,n,0.5)
> sp
[1] 2
> pv
[1] 0.5
```

Here sp is the value of s^+ and pv is the p-value. If given level of significance is less than p-value, we accept H_0. Here $\alpha = 0.05$ hence accept H_0. So there is no increase in I.Q after training program.

11.5 Wilcoxon's Signed Rank Test

The sign test for location uses only the signs of differences of observations from hypothesized median without considering their magnitude. If the information regarding magnitude is used, then that test procedure is expected to give better performance. Wilcoxon's signed rank test is based on these considerations. However, this test requires additional assumption of symmetry of the population about true median.

Suppose $x_1, x_2, \ldots\ldots, x_n$ is a random sample of size n from the distribution of random variable X. Let Fx(.) be the distribution function and M be the median of X. To test the hypothesis $H_0 : M = M_0$ against alternative $H_1 : M > M_0$ or $M < M_0$ or $M \neq M_0$, we use R command wilcox.test.

Example 11.6 : *The following are the measurements of the breaking strength (X) (in pounds) of a certain kind of 2-inch cotton ribbon .*

163, 165, 160, 189, 161, 171, 158, 151, 169, 162, 163, 139, 172, 165, 148, 166, 172, 163, 187, 173. Test the null hypothesis that population median of X is 160 against the alternative that it is greater than 160 at 0.05 level of significance using Wilcoxon signed rank test.

Solution : R-Code with output

```
> x=c(163,165,160,189,161,171,158,151,169,162,163,139,172,
+ 165,148,166,172,163,187,173)
```

```
> wilcox.test(x,alter="greater",mu=160)
Wilcoxon signed rank test with continuity correction
data: X
V = 146, p-value = 0.02095
Alternative hypothesis: true location is greater then 160
warning messages:
1:cannot compute exact p-value with ties in: wilcox.test.default(x, alter = "greater", mu = 160)
2:cannot compute exact p-value with zeroes in:wilcox.test.default(x, alter = "greater", mu = 160)
```

Here V denotes the value of test statistics in Wilcoxon signed rank test.

If given level of significance is less than p-value, we accept H_0, here $\alpha = 0.05$. Hence we reject H_0. So breaking strength of a certain kind of cotton ribbon may be greater then 160.

11.6 Wilcoxon's Signed Rank Test For Paired Data

Suppose (x_i, y_i) $i = 1, 2, \ldots, n$ be a random sample from bivariate distribution of (x, y). Let $D_i = x_i - y_i$, $i = 1, 2, \ldots, n$. We consider D_i, $i = 1, \ldots, n$ a random sample from the distribution of $D = X - Y$. Let $F_D(.)$ be the distribution function of D and M_D be the median of the distribution of D. We test the null hypothesis $H_0 : M_D = 0$ against one of the alternatives $H_1 : M_D > 0$ or $M_D < 0$ or $M_D \neq 0$, using R command wilcox.test.

Example 11.7 : *The following are the weights in pounds of 16 persons, before and after a certain weight reducing diet programme of four weeks.*

Use Wilcoxon's signed rank test to test whether the weight reducing diet is effective at 0.01 level of significance.

Person	Weight before	Weight After	person	Weight Before	Weight After
1	147	137.9	9	147.7	149
2	183.5	176.2	10	208.1	195.4
3	232.1	219	11	166.8	158.5
4	161.6	163.8	12	131.9	134.4
5	197.5	193.5	13	150.3	149.3
6	206.3	201.4	14	197.2	189.1
7	177	180.6	15	159.8	159.1
8	215.4	203.2	16	171.7	173.2

Solution : Here we want to test

Ho : MD = 0 against H1 : MD > 0 Where, D = x - y,

x = Weight before diet

y = Weight after diet

```
> x=c(147,183.5,232.1,161.6,197.5,206.3,177,215.4,147.7,208.1,
+ 166.8,131.9,150.3)
> x=c(x,197.2,159.8,171.7)
> y=c(137.9,176.2,219,163.8,193.5,201.4,186.6,203.2,149,195.4,
+158.5,134.4,149.3)
> y=c(y,189.1,159.1,173.2)
> d=x-y
> wilcox.test(d,mu=0,alter="greater")
Wilcoxon signed rank test
data: d
V = 105, p-value = 0.02884
alternative hypothesis : true location is greater then 0
```

Here V denotes the value of test statistics in Wilcoxon signed rank test.
If given level of significance is less than p-value, we accept H0. Here $\alpha = 0.01$.
Hence we accept H0. So weight reducing diet is not effective.
Alternatively we can replace wilcox.test command as given below :

```
> wilcox.test(x,y,paired=T, mu=0,alter="greater")
Wilcoxon signed rank test
data: x and y
V = 105, p-value = 0.02884
Alternative hypothesis: true location shift is greater than 0
```

11.7 Mann-Whitney Test

Questions for Practice:

The following data represents the number of Statistics lectures attended by 18 students : 9, 12, 18, 14, 12, 14, 12, 10, 16, 14, 13, 15, 13, 11, 13, 11, 9, 11. Perform a sign test to test the instructor's claim that the median of number of lectures attended is 12. Use a 2% level of significance.

The PQR company claims that the lifetime of a type of battery that it manufactures is more than 250hrs. A consumer advocate wishing to determine whether the claim is justified measures the lifetimes of 24 of the company's batteries; the results are : $271, 230, 198, 275, 282, 225, 284, 219, 253,$
$216, 262, 288, 236, 291, 253, 224, 264, 295, 211, 252, 294, 242, 272, 268$. Assuming the sample to be random, Using **Sign test**, determine whether the company's claim is justified at the 0.05 significance level.

Test the null hypothesis that the following sample is from a population with median 100 against the alternative the median is greater than 100 i.e., Use normal approximation to Wilcoxon Signed Rank Test (without continuity correction) Assume that the distribution of differences is symmetric. 98.38, 115.33, 98.62, 114.38, 87.79, 84.06, 96.18, 98.74, 91, 107.82, 108.28, 112.62, 124.18, 101.99, 112.51, 75.65, 83.77, 84.91, 109.73, 109.41, 100.4, 95.37, 115.46, 111.78, 86.13, 82.14, 78.47, 98.18

A student tells her parents that the median rental rate for a studio apartment in Portland is $700. Her parents are skeptical and believe the rent is different. A random sample of studio rentals is taken from the internet; prices are listed below. Test the claim that there is a difference using \(\alpha\) = 0.10. Should the parents believe their daughter?

700 650 800 975 855 785 759 640 950 715 825 980 895 1025 850 915 740 985 770 785 700 925

5. The median age of a tourist visiting to certain place is claimed to be 41 years. A random sample of 17 tourists have the ages
   25, 19, 38, 52, 57, 39, 46, 46, 30, 49, 27, 39, 44, 63, 31, 67, 42.
   Use sign test to test the claim at 5% level of significance.
6. To determine the effectiveness of a new traffic control system the number of accidents occured at 12 different locations during four weeks before and four weeks after the installation of new system were observed and data are recorded as follows :

Location	No. of accidents before	No. of accidents after
1	3	1
2	5	2
3	2	0
4	3	2
5	3	2
6	3	0
7	0	2
8	4	3
9	1	3
10	6	4
11	4	1
12	1	0

Use sign test at 5% level of significance to test whether new-traffic control system is effective.

7. The following data represent the number of hours that a rechargeable hedge trimmer operates before a recharge is required.
   1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2, 1.7
   Use Wilcoxn's signed rank test to test the hypothesis that this particular trimmer operates with a median of 1.8 hours before requiring a recharge.
8. The time (in minutes) that a patient has to wait in a consulting room before being seen by doctor is recorded for 12 patients as follows :
   17, 15, 20, 20, 32, 28, 12, 26, 25, 25, 35, 24
   Use Wilcoxn's signed rank test to test whether the median waiting time is not more than 20 minutes, at 5% level of significance.