ASSIGNMENT
OF
DTNS

दक्षिण बिहार केन्द्रीय विश्वविद्यालय

CENTRAL UNIVERSITY OF SOUTH BIHAR

ज्ञान सेवा विमुक्तये

DEPARTMENT OF STATISTICS

SUBMITTED BY:

NAME-SURESH KUMAR PRAJAPATI

ENROLMENT NO.-CUSB2302222008

RUN TEST

Q.1- The win-loss record of a certain basketball team for their 50 consecutive games was as follows :-

6W L 6W L W L 3W 2L 4W L 3W 2L 6W 2L 2W 3L 2W L 3W

Apply the run test to test that the sequence of win and loss is random.

Solution:-

Here we set the null and alternative hypothesis as:

Ho: The sequence of win and loss is random

H1: The sequence of win and loss is not random.

Observation:-

n1	36
n2	14
R	19
μ	21.16
σ^2	7.882971429
σ	2.807662983
Zo	-0.769323103
|Zo|	0.769323103
F(|Zo|)	0.779149237
p*	**0.441701526**
α	**0.05**

Conclusion:-

Here, we can conclude that the value of p* is greater than the value of α (p*> α). So, we accept the null hypothesis and we can say that the sequence of win and loss is random.

SIGN TEST

Q-2. The minimum temperature of a place has been recorded for fourteen successive days. Test if the successive changes (positive and negative) in the minimum temperatures are random.

Day	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Temperature(°C)	**20.1**	**20.3**	**20.1**	**20.4**	**20.5**	**20.6**	**20.4**	**20.3**	**20.7**	**20.7**	**21.0**	**21.6**	**21.5**	**21.0**

Solution:-

Here we set the null and alternative hypothesis as

H_o: The successive positive and negative changes in minimum temperatures are random.

H_1: The successive positive and negative changes in minimum temperatures are not random.

We find the following positive and negative changes in temperatures with respect to day 1:

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Temperature(°C)	20.1	20.3	20.1	20.4	20.5	20.6	20.4	20.3	20.7	20.7	21.0	21.6	21.5	21.0
Difference		+	-	+	+	+	-	-	+	0	+	+	-	-

Observation:-

n	12
n1	7
n2	5
R	6
k	3
p*	**0.227273**
α	**0.05**

Conclusion:-

Here, we can conclude that the value of p* is greater than the value of α (p*> α). So, we accept the null hypothesis and we can say that the successive positive and negative changes in minimum temperatures are random.

SIGN TEST FOR LARGE SAMPLE

Q-3. Let us consider motor trained car road test data which consist 32 observations & 11 variables by using the dataset "mtcars" in R. Perform a sign test to test mileage (mileage per gallan) varies from its median mpg or not.

Solution:-

H_0: The population median is = 19.2

H_1: The population median is not equal to 19.2

```
> x=mtcars
> x
                    mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82  0  0    3    4
Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
```

```
Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
> length(x)
[1] 11
> y=x[,1]
> y
 [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
10.4 14.7
[18] 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7 15.0 21.4
```
> median(y)

[1] 19.2

```
> z=y-median(y)
> z
 [1]  1.8  1.8  3.6  2.2 -0.5 -1.1 -4.9  5.2  3.6  0.0 -1.4 -2.8 -1.9 -4.0 -8.8 -8.8 -4.5
[18] 13.2 11.2 14.7  2.3 -3.7 -4.0 -5.9  0.0  8.1  6.8 11.2 -3.4  0.5 -4.2  2.2
> which(y>19.2)
 [1]  1  2  3  4  8  9 18 19 20 21 26 27 28 30 32
> q=which(y>19.2)
```
> length(q)

[1] 15

```
> w=which(y<19.2)
```
> length(w)

[1] 15

```
> r=which(y==19.2)
```
> length(r)

[1] 2

```
>
q=choose(30,0)+choose(30,1)+choose(30,2)+choose(30,3)+choose(30,4)+choos
e(30,5)+choose(30,6)+choose(30,7)+choose(30,8)+choose(30,9)+choose(30,10)
+choose(30,11)+choose(30,12)+choose(30,13)+choose(30,14)+choose(30,15)
```

```
> q
[1] 614429672
> l=2^30
> l
[1] 1073741824
> p=q/l
```
> p
[1] 0.5722322

<u>By 2nd method,</u>

```
> signtest(y,m=median(y),alternative="less")
```

Large Sample Approximation for the Sign Test

H_0: The population median is = 19.2
H_1: The population median is less than 19.2

B = 15

Significance Level = 0.05
The p-value is 0.572433929707647
There is not enough evidence to conclude that the population median is less than 19.2 at a significance level of 0.05

```
> signtest(y,m=median(y),alternative="greater")
```
Large Sample Approximation for the Sign Test
H0: The population median is = 19.2
HA: The population median is greater than 19.2
B = 15
Significance Level = 0.05
The p-value is 0.572433929707647
There is not enough evidence to conclude that the population median is greater than 19.2 at a significance level of 0.05
```
> signtest(y,m=median(y))
```
Large Sample Approximation for the Sign Test
B = 15
Significance Level = 0.05
The p-value is 0.855132140584706

Conclusion:

There is not enough evidence to conclude that the population median is different than 19.2 at a significance level of 0.05

SIGN TEST FOR TWO SAMPLE

Q-4 A random sample of size 10 of diabetic patients is selected at random to see the effect of reducing blood sugar level of a new drug their fasting blood sugar level are measured before and after using the drug for 3 months. Perform a test at 1% level of significance to see if the drug is effective. given that $S_{(L,alpha)}$=0 for n=2 and alpha=0.01.

Solution:-

H_0:After 3 month the drug is effective

H_1:After 3 month the drug is not effective

```
Time.1<-c(125,132,137,128,140,160,155,145,132,152)
Time.2<-c(110,135,125,132,140,142,148,127,132,140)
SIGN.test(x = Time.1,
+          y = Time.2,
+          alternative = "two.sided",
+          conf.level = 0.99)
            Dependent-samples Sign-Test
data:  Time.1 and Time.2
S = 6, p-value = 0.2891
alternative hypothesis: true median difference is not equal to 0
99 percent confidence interval:
 -3.588 18.000
sample estimates:
median of x-y
        9.5
Achieved and Interpolated Confidence Intervals:
              Conf.Level L.E.pt U.E.pt
Lower Achieved CI     0.9785 -3.000     18
Interpolated CI       0.9900 -3.588     18
Upper Achieved CI     0.9980 -4.000     18
```

Conclusion:

After compare with p value we conclude that after 3 month the drug is effective

WILCOXON SIGN RANK TEST

Q-5: In a certain experiment to compare two types of animal foods A and B the following results of increase in weights were observed in animals:

Animal number		1	2	3	4	5	6	7	8	Total
Increase weight in lb	Food A	49	53	51	52	47	50	52	53	407
	Food B	52	55	52	53	50	54	54	53	423

(I) Assuming that the two samples of animals are independent, can we conclude that food B is better than food A?

Solution:

H_0:Food B is batter than A

H_1:Food B is not batter than A

> x=c(49,53,51,52,47,50,52,53)

> x

[1] 49 53 51 52 47 50 52 53

> y=c(52,55,52,53,50,54,54,53)

> y

[1] 52 55 52 53 50 54 54 53

> z=x-y

> z

[1] -3 -2 -1 -1 -3 -4 -2 0

> wilcox.test(x,y,paired=T)

Wilcoxon signed rank test with continuity correction

data: x and y

V = 0, p-value = 0.02178

alternative hypothesis: true location shift is not equal to 0

Warning messages:

1: In wilcox.test.default(x, y, paired = T) :

cannot compute exact p-value with ties

2: In wilcox.test.default(x, y, paired = T) :

cannot compute exact p-value with zeroes

Conclusion:

Here the p value is less than significance level so that the H_0 hypothesis is accept.

WILCOXON SIGN RANK TEST FOR SMALL SAMPLE

Q-6.Let us consider the follwing Data which shows IQ score of 15 employs selected at

random given as *:*
99 ,100 ,90 ,94,135,108,107,111,119,104,127,109,117,105,125.

Solution:

H_0:The data of IQ score of employs selected at random

H_1:The data of IQ score of employs not selected at random

```
> x=c(99,100,90,94,135,108,107,111,119,104,127,109,117,105,125) #assigned vector
> median(x)
[1] 108
> y=x-median(x)
> y
 [1]  -9  -8 -18 -14  27   0  -1   3  11  -4  19   1   9  -3  17
> 3
[1] 3
> a=which(y>0) #extract the y>0 values
> b=which(y<0) #extract the y<0 values
> a
[1]  5  8  9 11 12 13 15
> b
[1]  1  2  3  4  7 10 14
> length(a)
[1] 7
> length(b)
[1] 7

>
> t.test(x,mu=107)

        One Sample t-test
```

```
data:  x
t = 0.92227, df = 14, p-value = 0.372
alternative hypothesis: true mean is not equal to 107
95 percent confidence interval:
 103.0234 116.9766
sample estimates:
mean of x
     110

> wilcox.test(x,mu=107,alternative ="less")

        Wilcoxon signed rank test with continuity correction

data:  x
V = 64.5, p-value = 0.7837
alternative hypothesis: true location is less than 107

Warning messages:
1: In wilcox.test.default(x, mu = 107, alternative = "less") :
  cannot compute exact p-value with ties
2: In wilcox.test.default(x, mu = 107, alternative = "less") :
  cannot compute exact p-value with zeroes

> t.test(x,mu=107,alternative = "g")

        One Sample t-test

data:  x
t = 0.92227, df = 14, p-value = 0.186
alternative hypothesis: true mean is greater than 107
95 percent confidence interval:
 104.2707      Inf
sample estimates:
mean of x
     110
```

Conclusion:

Here the value of p is greater than significance value so that the alternative

Hypothesis is accept so that here the data is not selected randomly.

WILCOXON TEST FOR LARGE SAMPLE

Q-7 Let us consider the follwing Data which shows IQ score of 15 employs selected at random given as:99,100,90,94,135,108,107,111,119,104,127,109,117,105,125,98,112,85,92,140,111,115,137,117,123,132,83,120,82,106,147,110

Solution:

H_0: The data of IQ score of employs selected at random

H_1: The data of IQ score of employs not selected at random

```
>f=c(99,100,90,94,135,108,107,111,119,104,127,109,117,105,125,98,112,85,9
2,140,111,115,137,117,123,132,83,120,82,106,147,110)
> f
 [1]  99 100  90  94 135 108 107 111 119 104 127 109 117 105 125  98 112  85
92
[20] 140 111 115 137 117 123 132  83 120  82 106 147 110
> length(f)
[1] 32
> x=median(f)
> x
[1] 110.5
> q=which(f>110.5)
> q
 [1]  5  8  9 11 13 15 17 20 21 22 23 24 25 26 28 31
> length(q)
[1] 16
> p=which(f<110.5)
> p
 [1]  1  2  3  4  6  7 10 12 14 16 18 19 27 29 30 32
> length(p)
[1] 16
> t=which(f==110.5)
```

```
> t
integer(0)
> length(t)
[1] 0
> sum_q=sum(q)
> sum_q
[1] 298
> n=32
> E_q=((n*(n+1))/4)
> E_q
[1] 264
> var_q=((n*(n+1)*(2*n+1))/24)
> var_q
[1] 2860
> sqrt_var_q=sqrt(var_q)
> sqrt_var_q
[1] 53.47897
> Zo=((sum_q-E_q)/sqrt_var_q)
> Zo
[1] 0.635764
> pnorm(Zo)
[1] 0.7375349
> P*=2*(1-pnorm(Zo))
> P*
[1] 0.5249303
```

Conclusion:

Here p value is greater than significance value so that the H_0 is rejected so the data are not selected randomly.

WILCOXON SIGN RANK TEST

Q-8 Apply the Wilcoxon Signed rank test for the "ToothGrowth" dataset in R, (which contain length (len), supplement (supp) and dose) to test significance difference of supplements for the growing length of the following:

1. **For the supplement "VC",**
   a. **Dose 0.5 is effective to dose 1 or not**
   b. **Dose 1 is effective to dose 1.5 or not**
   c. **Dose 0.5 is effective to dose 1.5 or not**
2. **For the supplement "OJ",**
   d. **Dose 0.5 is effective to dose 1 or not**
   e. **Dose 1 is effective to dose 1.5 or not**
   f. **Dose 0.5 is effective to dose 1.5 or not**

```
x=ToothGrowth
> x
   len supp dose
1   4.2   VC  0.5
2  11.5   VC  0.5
3   7.3   VC  0.5
4   5.8   VC  0.5
5   6.4   VC  0.5
6  10.0   VC  0.5
7  11.2   VC  0.5
8  11.2   VC  0.5
9   5.2   VC  0.5
10  7.0   VC  0.5
11 16.5   VC  1.0
12 16.5   VC  1.0
```

13	15.2	VC	1.0
14	17.3	VC	1.0
15	22.5	VC	1.0
16	17.3	VC	1.0
17	13.6	VC	1.0
18	14.5	VC	1.0
19	18.8	VC	1.0
20	15.5	VC	1.0
21	23.6	VC	2.0
22	18.5	VC	2.0
23	33.9	VC	2.0
24	25.5	VC	2.0
25	26.4	VC	2.0
26	32.5	VC	2.0
27	26.7	VC	2.0
28	21.5	VC	2.0
29	23.3	VC	2.0
30	29.5	VC	2.0
31	15.2	OJ	0.5
32	21.5	OJ	0.5
33	17.6	OJ	0.5
34	9.7	OJ	0.5
35	14.5	OJ	0.5
36	10.0	OJ	0.5
37	8.2	OJ	0.5
38	9.4	OJ	0.5
39	16.5	OJ	0.5
40	9.7	OJ	0.5
41	19.7	OJ	1.0
42	23.3	OJ	1.0
43	23.6	OJ	1.0
44	26.4	OJ	1.0
45	20.0	OJ	1.0
46	25.2	OJ	1.0
47	25.8	OJ	1.0
48	21.2	OJ	1.0
49	14.5	OJ	1.0
50	27.3	OJ	1.0
51	25.5	OJ	2.0
52	26.4	OJ	2.0

```
53 22.4   OJ  2.0
54 24.5   OJ  2.0
55 24.8   OJ  2.0
56 30.9   OJ  2.0
57 26.4   OJ  2.0
58 27.3   OJ  2.0
59 29.4   OJ  2.0
60 23.0   OJ  2.0
> y<-subset(x,supp=="VC")
> y
   len supp dose
1   4.2   VC  0.5
2  11.5   VC  0.5
3   7.3   VC  0.5
4   5.8   VC  0.5
5   6.4   VC  0.5
6  10.0   VC  0.5
7  11.2   VC  0.5
8  11.2   VC  0.5
9   5.2   VC  0.5
10  7.0   VC  0.5
11 16.5   VC  1.0
12 16.5   VC  1.0
13 15.2   VC  1.0
14 17.3   VC  1.0
15 22.5   VC  1.0
16 17.3   VC  1.0
17 13.6   VC  1.0
18 14.5   VC  1.0
19 18.8   VC  1.0
20 15.5   VC  1.0
21 23.6   VC  2.0
22 18.5   VC  2.0
23 33.9   VC  2.0
24 25.5   VC  2.0
25 26.4   VC  2.0
26 32.5   VC  2.0
27 26.7   VC  2.0
28 21.5   VC  2.0
29 23.3   VC  2.0
```

```
30 29.5   VC  2.0
> d1<-subset(y,dose==0.5)
> d1
   len supp dose
1   4.2   VC  0.5
2  11.5   VC  0.5
3   7.3   VC  0.5
4   5.8   VC  0.5
5   6.4   VC  0.5
6  10.0   VC  0.5
7  11.2   VC  0.5
8  11.2   VC  0.5
9   5.2   VC  0.5
10  7.0   VC  0.5
> d2<-subset(y,dose==1)
> d2
   len supp dose
11 16.5   VC    1
12 16.5   VC    1
13 15.2   VC    1
14 17.3   VC    1
15 22.5   VC    1
16 17.3   VC    1
17 13.6   VC    1
18 14.5   VC    1
19 18.8   VC    1
20 15.5   VC    1
> d3<-subset(y,dose==2)
> d3
   len supp dose
21 23.6   VC    2
22 18.5   VC    2
23 33.9   VC    2
24 25.5   VC    2
25 26.4   VC    2
26 32.5   VC    2
27 26.7   VC    2
28 21.5   VC    2
29 23.3   VC    2
30 29.5   VC    2
```

```
> l1<-d1$len
> l1
 [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2 11.2  5.2  7.0
> l2<-d2$len
> l2
 [1] 16.5 16.5 15.2 17.3 22.5 17.3 13.6 14.5 18.8 15.5
> l3<-d3$len
> l3
 [1] 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5
> q<-wilcox.test(l1,l2,paired = T)
> w<-wilcox.test(l1,l3,paired = T)
> e<-wilcox.test(l2,l3,paired = T)
> q

        Wilcoxon signed rank exact test

data:  l1 and l2
V = 0, p-value = 0.001953
alternative hypothesis: true location shift is not equal to 0

> w

        Wilcoxon signed rank test with continuity correction

data:  l1 and l3
V = 0, p-value = 0.005889
alternative hypothesis: true location shift is not equal to 0

> e

        Wilcoxon signed rank exact test

data:  l2 and l3
V = 0, p-value = 0.001953
alternative hypothesis: true location shift is not equal to 0

> p<-wilcox.test(l2,l1,paired = T)
> o<-wilcox.test(l3,l1,paired = T)
> i<-wilcox.test(l2,l3,paired = T)
> p
```

```
        Wilcoxon signed rank exact test

data:  l2 and l1
V = 55, p-value = 0.001953
alternative hypothesis: true location shift is not equal to 0

> o

        Wilcoxon signed rank test with continuity correction

data:  l3 and l1
V = 55, p-value = 0.005889
alternative hypothesis: true location shift is not equal to 0


>
```

MANN-WHITNEY U TEST

Q-9. Let the score of two groups of person out of which one under plesibo and other under a new drug given as

X= score under plesibo

Y=score under new drug

X	**10,13,12,15,16,8,6**
Y	**20,14,7,9,17,18,19,25,24**

Test whether the distribution of score of new drug is more suitable then under plesibo.

Given that u_α=16 for this data set, n1=7 & n2=9.

Solution:

```
> x=c(10,13,12,15,16,8,6)

> y=c(20,14,7,9,17,18,19,25,24)
> x

[1] 10 13 12 15 16  8  6

> y
[1] 20 14  7  9 17 18 19 25 24

> n1=7
> n2=9
> n=n1+n2
> n
[1] 16
> m=n1*n2
> m
[1] 63
> z=c(x,y)
> z
 [1] 10 13 12 15 16  8  6 20 14  7  9 17 18 19 25 24
> r=rank(z)
> r
```

```
 [1]  5  7  6  9 10  3  1 14  8  2  4 11 12 13 16 15
> len=length(x)
> sum_rank=sum(r[1:len])
> sum_rank
[1] 41
> s=rank(x)
> s
[1] 3 5 4 6 7 2 1
> l=7
> sum_xrank=sum(s[1:l])
> sum_xrank
[1] 28
> u=sum_rank-sum_xrank
> u
[1] 13
```

Conclusion:

The value of u_α is greater then u, So the distribution of score of new drug is more suitable then under plesibo.

```
> wilcox.test(x,y,alternative ="two.sided")

               Wilcoxon rank sum exact test
data:  x and y
W = 13, p-value = 0.0549
alternative hypothesis: true location shift is not equal to 0
```

Key used:

```
x=c(10,13,12,15,16,8,6)
y=c(20,14,7,9,17,18,19,25,24)
x
y
n1=7
n2=9
n=n1+n2
n
m=n1*n2
m
z=c(x,y)
z
r=rank(z)
r
len=length(x)
sum_rank=sum(r[1:len])
sum_rank
s=rank(x)
```

```
s
l=7
sum_xrank=sum(s[1:l])
sum_xrank
u=sum_rank-sum_xrank
u

wilcox.test(x,y,alternative ="two.sided")
```