Sign Test

DR.E.S.JEEVANAND
ASSOCIATE PROFESSOR

U.C.COLLEGE, ALUVA

What is Sign Test

This is a nonparametric analogue of Student's t test for the mean(i.e. a location parameter) of the population. The t-test is based on the assumption of normality of the underlying population. Sign test is a nonparametric alternative to t test. This test does not require the assumption of normality. It also provides a test of location but uses quantiles of the distribution as the location parameter. Moreover, sign test is based on only the continuity of the underlying population.

The Sign Test for Median

Sign test is a non-parametric test for testing hypotheses concerning the location parameters of one or more populations. If the population is normally distributed the location parameter is the arithmetic mean. But we are case where we have no idea of the distribution of the population and we did not make any assumption about other than it is continuous. So we take the median as the location parameter.

The Sign Test for Median

Recall that for a continuous random variable X, the median is the value M such that 50% of the time X lies below M and 50% of the time X lies above M *or* $P(X \leq M) = P(X \geq M) = \frac{1}{2}$. Two problem usually come up in this connection (1) to test $\boldsymbol{H_0{:}M = M_0}$ where $\boldsymbol{M_0}$ is a value suggested theoretically or from experience and (2) to test whether the Median of two population are the same or not. i.e. $\boldsymbol{H_0{:}M_1 = M_2}$.

One Sample Sign Test for Median

Throughout our discussion, and as the above illustration suggests, we'll assume that our random variable X is a continuous random variable with unknown median m. Upon taking a random sample X_1, X_2, ..., X_n, we'll be interested in testing whether the median M takes on a particular value M_0. That is, we'll be interested in testing the null hypothesis:

$$H_0: M = M_0$$

against any of the possible alternative hypotheses

$H_1: M < M_0$	$H_1: M > M_0$	$H_1: M \neq M_0$
One-tailed (lower-tail)	One-tailed (upper-tail)	Two-tailed

One Sample Sign Test for a Median
Steps in testing the hypothesis

1. Calculate $X_i - M_0$ for $i = 1, 2, ..., n$.
2. Define $N-$ = the number of negative signs obtained upon calculating $X_i - M_0$ for $i = 1, 2, ..., n$.
3. Define $N+$ = the number of positive signs obtained upon calculating $X_i - M_0$ for $i = 1, 2, ..., n$.

Then, if the null hypothesis is true, that is, $M = M_0$, then $N-$ and $N+$ both follow a binomial distribution with parameters N the sum of $N-$ and $N+$ (If all $X_i - M_0$ are non zero then N=n other wise N= n- number of zeros) and $p = 1/2$. The test statistics is

T = $N+$ = the number of positive signs obtained which has binomial $\mathrm{B}\left(N, \frac{1}{2}\right)$ distribution

One Sample Sign Test for Median Critical Region

For the Alternative Hypothesis $\boldsymbol{H_1 : M < M_0}$ $\boldsymbol{(lower\ tail)}$

The rejection rule is:

- Reject H_0 if $\mathrm{T} \leq \mathrm{t}_\alpha$

Where for a given α, t_α (Left tail value) is obtained from the binomial table $\mathrm{B}\left(N, \frac{1}{2}\right)$ such that $P(X \leq t_\alpha) \leq \alpha$.

For the Alternative Hypothesis $\boldsymbol{H_1 : M > M_0}$ $\boldsymbol{(upper\ tail)}$

The rejection rule is:

- Reject H_0 if $\mathrm{T} \geq \mathrm{t}'_\alpha$

Where for a given α, t'_α(Right tail value) is obtained from the binomial table $\mathrm{B}\left(N, \frac{1}{2}\right)$ such that $P(X \geq t'_\alpha) \leq \alpha$.

One Sample Sign Test for Median
Critical Region

For the Alternative Hypothesis $\boldsymbol{H_1: M \neq M_0}$ $\boldsymbol{(Two\ tail)}$

The rejection rule is:

- Reject H_0 if either $T \leq t_{\alpha/2}$ or $T \geq t'_{\alpha/2}$

Where for a given α, $t_{\alpha/2}$ (Left tail value) and $t'_{\alpha/2}$(Right tail value) is obtained from the binomial table $B\left(N, \frac{1}{2}\right)$ such that $P(X \leq t_{\frac{\alpha}{2}}) \leq \frac{\alpha}{2}$ and $P(X \geq t'_{\frac{\alpha}{2}}) \leq \frac{\alpha}{2}$.

Note: As the binomial distribution is a discrete distribution we cannot get the critical region such that the significance level is exactly equal to the given α (0.01 or 0.05)

Example I

Recent studies of the private practices of physicians who saw no Medicaid patients suggested that the median length of each patient visit was 22 minutes. It is believed that the median visit length in practices with a large Medicaid load is shorter than 22 minutes. A random sample of 20 visits in practices with a large Medicaid load yielded, in order, the following visit lengths:

9.4 13.4 15.6 16.2 16.4 16.8 18.1 18.7 18.9 19.1
19.3 20.1 20.4 21.6 21.9 23.4 23.5 24.8 24.9 26.8

Based on these data, is there sufficient evidence to conclude that the median visit length in practices with a large Medicaid load is shorter than 22 minutes?

We are interested in testing the null hypothesis $\boldsymbol{H_0: M = 22}$ against the alternative hypothesis $\boldsymbol{H_1: M < 22}$. To do so, we first calculate $x_i - 22$, for $i = 1, 2, \ldots, 20$.

Example I

X_i	X_i-M	X_i	X_i-M
9.4	-12.6	19.3	-2.7
13.4	-8.6	20.1	-1.9
15.6	-6.4	20.4	-1.6
16.2	-5.8	21.6	-0.4
16.4	-5.6	21.9	-0.1
16.8	-5.2	23.4	+1.4
18.1	-3.9	23.5	+1.5
18.7	-3.3	24.8	+2.8
18.9	-3.1	24.9	+2.9
19.1	-2.9	26.8	+4.8

Here $N- = 15$ and $N+ = 5$ and N=20

the test statistics is

T= $N+$ (the number of positive signs)

i.e. T= 5

From the Binomial table for a α=0.05, t_α=6

Since $T < t_\alpha$ we reject H_0 and conclude that M<22.

Example II

The height of 20 plants of a certain variety were measured one month after plating them and are given below and are given below. Test whether their median height is 30 inches

31.5, 28.8, 29.3, 32.3, 33.5, 31.0, 29.5, 34.2, 31.0, 33.5,

30.0, 30.1, 30.0, 32.3, 35.6, 32.7, 30.1, 29.1, 33.2, 30.0

We are interested in testing the null hypothesis $\boldsymbol{H_0}$: $\boldsymbol{M = 30}$ against the alternative hypothesis $\boldsymbol{H_1}$: $\boldsymbol{M \neq 30}$. To do so, we first calculate $x_i - 30$, for $i = 1, 2, ..., 20$.

Example II

X_i	X_i-M	X_i	X_i-M
31.5	+1.5	30.0	0.0
28.8	-1.2	30.1	+0.1
29.3	-0.7	30.0	0.0
32.3	+2.3	32.3	+2.3
33.5	+3.5	35.6	+5.6
31.0	+1.0	32.7	+2.7
29.5	-0.5	30.1	+0.1
34.2	+4.2	29.1	-0.9
31.0	+1.0	33.2	+3.2
33.5	+3.5	30.0	0.0

Here $N- = 4$ and $N+ = 13$ Here three values give $x_i - 30=0$, we have N=17

the test statistics is

T= $N+$ (the number of positive signs)

i.e. T= 13

From the Binomial table for a α=0.05, $t_{\alpha/2}$=4 and $t'_{\alpha/2}$ =13

Since $T \geq t'_{\alpha/2}$ we reject H_0 and conclude that M≠30.

Example III

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 $H_0 : M = 160$ against the alternative $H_1 : M > 160$

To do so, we first calculate $x_i - 160$, for $i = 1, 2, ..., 20$.

Example III

X_i	X_i-M	X_i	X_i-M
163	+3	163	+3
165	+5	139	-21
160	0	172	+12
189	+29	165	+5
161	+1	148	-12
171	+11	166	+6
158	-2	172	+12
151	-9	163	+3
169	+9	187	+27
162	+2	173	+13

Here $N- = 4$ and $N+ = 15$ Here there values give x_i – 160=0, we have N=19

the test statistics is

T= $N+$ (the number of positive signs)

i.e. T= 15.

From the Binomial table for a α=0.05, t'_α =14.

Since T $\geq t'_\alpha$ we reject H_0 and conclude that M>160.

Two Sample Sign Test for a Median

This is a modification of the sign test and is used for testing the equality of the location parameters of two population based on paired Samples

That is, we'll be interested in testing the null hypothesis:

$$H_0 : M_1 = M_2$$

against any of the possible alternative hypotheses

$H_a : M_1 < M_2$	$H_a : M_1 > M_2$	$H_a : M_1 \neq M_2$
One-tailed (lower-tail)	One-tailed (upper-tail)	Two-tailed

Two Sample Sign Test for a Median
Steps in testing the hypothesis

Let (x_i, y_i), i =1, 2, …, n be the paired sample.

1. Calculate $d_i = x_i - y_i$ for i =1, 2, …, n.
2. Define $N-$ = the number of negative signs obtained upon calculating d_i for i = 1, 2, …, n.
3. Define $N+$ = the number of positive signs obtained upon calculating d_i for i = 1, 2, …, n.

Then, if the null hypothesis is true, that is, $\boldsymbol{M_1 = M_2}$, then $N-$ and $N+$ both follow a binomial distribution with parameters N the sum of $N-$ and $N+$ (If all $X_i - M_0$ are non zero then N=n other wise N= n- number of zeros) and p = 1/2. The test statistics is

T = $N+$ = the number of positive signs obtained which has binomial $B\left(N, \frac{1}{2}\right)$ distribution

Two Sample Sign Test for Median Critical Region

For the Alternative Hypothesis $\boldsymbol{H_1 : M_1 < M_2 (lower\ tail)}$

The rejection rule is:

- Reject H_0 if $\mathrm{T} \leq \mathrm{t}_\alpha$

Where for a given α, t_α (Left tail value) is obtained from the binomial table $\mathrm{B}\left(N, \frac{1}{2}\right)$ such that $P(X \leq t_\alpha) \leq \alpha$.

For the Alternative Hypothesis $\boldsymbol{H_1 : M_1 > M_2 (upper\ tail)}$

The rejection rule is:

- Reject H_0 if $\mathrm{T} \geq \mathrm{t}'_\alpha$

Where for a given α, t'_α(Right tail value) is obtained from the binomial table $\mathrm{B}\left(N, \frac{1}{2}\right)$ such that $P(X \geq t'_\alpha) \leq \alpha$.

Two Sample Sign Test for Median Critical Region

For the Alternative Hypothesis $\boldsymbol{H_1{:}M_1 \neq M_2 (Two\ tail)}$

The rejection rule is:

- Reject H_0 if either T $\leq$ $t_{\alpha/2}$ or T $\geq t'_{\alpha/2}$

Where for a given α, $t_{\alpha/2}$ (Left tail value) and $t'_{\alpha/2}$(Right tail value) is obtained from the binomial table $B\left(N, \frac{1}{2}\right)$ such that $P(X \leq t_{\frac{\alpha}{2}}) \leq \frac{\alpha}{2}$ and $P(X \geq t'_{\frac{\alpha}{2}}) \leq \frac{\alpha}{2}$.

Note: The difference between the one sample and two sample test is that in one sample test we test the hypothesis $\boldsymbol{H_0{:}M = M_0}$ and in two sample test we test the hypothesis $\boldsymbol{H_0{:}M_1 = M_2}$ or $\boldsymbol{H_0{:}M_1 - M_2}=0$.

Example IV

Is there any difference between the median values for the sets of treatment data for the twelve groups given in the next column?

In this case we want to test the hypothesis $\boldsymbol{H_0{:}M_1 = M_2}$ against $\boldsymbol{H_1{:}M_1 \neq M_2}$

For this first we Calculate

$d_i = x_i - y_i$ for i =1, 2, …, 12.

Group No	Treatment I	Treatment II	Group No	Treatment I	Treatment II
1	2.5	4	7	4.9	6.7
2	3.5	5.6	8	6.6	6
3	2.9	3.2	9	2	3.5
4	2.1	1.9	10	2	4
5	6.9	9.5	11	5.8	8.1
6	2.4	2.3	12	7.5	19.9

Example IV

Group No	x	y	d=x-y
1	2.5	4	-1.5
2	3.5	5.6	-2.1
3	2.9	3.2	-0.3
4	2.1	1.9	+0.2
5	6.9	9.5	-2.6
6	2.4	2.3	+0.1
7	4.9	6.7	-1.8
8	6.6	6	+0.6
9	2	3.5	-1.5
10	2	4	-2
11	5.8	8.1	-2.3
12	7.5	19.9	-12.4

Here $N- = 9$ and $N+ = 3$ and N=12

the test statistics is

T= $N+$ (the number of positive signs)

i.e. T= 3

From the Binomial table for a α=0.05, $t_{\alpha/2}$=2 and $t'_{\alpha/2}$ =10

Since $t_{\alpha/2} \leq T \leq t'_{\alpha/2}$ we accept H_0 and conclude that the median value of the two treatment are same.

Example V

To study the effectiveness of a certain diet in increasing the weight of a person 10 persons of more or less the same age, heritage etc. were selected and the diet administered. Their weights before and after administering the diet are given below. Use the sign test to examine whether the diet is effective

Pearson	A	B	C	D	E	F	G	H	I	J
Wt before	91	95	81	83	76	88	89	97	88	92
Wt after	79	101	85	88	81	92	90	99	97	97

In this case we want to test the hypothesis $\boldsymbol{H_0{:}M_1 = M_2}$ against $\boldsymbol{H_1{:}M_1 < M_2}$ Let y represent the weight before and X represent the weight after so we calculate

$d_i = x_i - y_i$ for i =1, 2, …, 10.

Example V

Person	x	y	d=x-y
A	79	91	-12
B	101	95	+6
C	85	81	+4
D	88	83	+5
E	81	76	+5
F	92	88	+4
G	90	89	+1
H	99	97	+2
I	97	88	+9
J	87	92	-5

Here $N- = 2$ and $N+ = 8$ and N=10

the test statistics is

T= $N+$ (the number of positive signs)

i.e. T= 8

From the Binomial table for a α=0.05, $t_{\alpha/2}$=2 and $t'_{\alpha/2}$ =8

Since $T \geq t'_{\alpha/2}$ we reject H_0 and conclude that the diet is effective.

Normal Approximation with Signe Test

The sign test considered so far is used when the sample is Small . That is when N $\leq$20. A large sample approximation is used for N> 20. If the number of observations N is greater than 20, we can use a normal approximation. In this case T follows normal distribution or Z distribution with

The mean $\mu = \frac{N}{2}$ and the standard deviation $\sigma = \frac{\sqrt{N}}{2}$

In this case the test statistic is $z = \frac{T - \frac{N}{2}}{\frac{\sqrt{N}}{2}}$

Which has standard normal distribution.

Normal Approximation with Wilcoxon Match-Pairs Signed Ranks Test Critical Region

The rejection rule is:

- Lower tail: Reject H_0 if Actual $z \leq$ Critical $-z_\alpha$
- Upper tail: Reject H_0 if Actual $z \geq$ Critical z_α
- Two tail: Reject H_0 if Actual $|z| \geq$ Critical $z_{\alpha/2}$

Where z_α and $z_{\alpha/2}$ are obtained from the standard Normal Table

Example VI

The following gives the pre test value (Y) and post test value (x) of a training given to the 22 students of a class to remember the digit of the numbers. Test whether the training is effective.

x	y	d=x-y	x	y	d=x-y
17	16	+1	35	29	+6
17	18	-1	19	17	+2
20	21	-1	18	16	+2
19	17	+2	19	17	+2
18	16	+2	21	23	-2
19	17	+2	16	17	-1
21	23	-2	17	19	-2
16	17	-1	27	18	+9
17	19	-2	35	36	-1
27	25	+2	19	16	+3
27	25	+2	27	23	+4

Example VI

In this case we want to test the hypothesis $\boldsymbol{H_0{:}M_1 = M_2}$ against

$\boldsymbol{H_1{:}M_1 < M_2}$. Here since N=22. we have to apply the large sample test

Here $N-=9$ and $N+=13$ and N=22 and

T= $N+$ (the number of positive signs) = 13

So $z = \dfrac{T-\frac{N}{2}}{\frac{\sqrt{N}}{2}} = \dfrac{13-\frac{22}{2}}{\frac{\sqrt{22}}{2}} = \dfrac{13-1}{2.3452}$ =0.8528

Here the calculated $z \geq -z_{\alpha} = -1.645$. So we accept H_0 and conclude that the training is not effective.

The End

In the next section we will discuss about the Wilcoxon matched-pairs signed-ranks test

Contact Details esjmaths@gmail.com

9847915101, 8089684892