E-Learning Module on One sample Sign**Test**

Learning Objectives

At the end of this session, you will be able to know

- About One sample sign test
- Principle of the test
- Assumptions and Procedure of the test
- Applications of the test

Introduction

- The sign test is the simplest of the nonparametric tests, and is similar to testing if a two-sided coin is fair. Its name comes from the fact that it is based on the direction (or signs for + and -) of a pair of observations and not on their numerical magnitude. Count the number of +ve values (larger than hypothesized median), the number of -ve values (smaller than the hypothesized median), and test whether there are significantly more positives (or negatives) than expected.

The One Sample Sign Test is a Nonparametric equivalent to the parametric One Sample t-Test.

The one-sample sign test is used to test the null hypothesis that the probability of a random value from the population being above the specified value is equal to the probability of a random value being below the specified value.

Nonparametric tests do not make assumptions that the population is from a specific distribution. Therefore its results are more robust than a parametric test when such assumptions are violated.

Test	**Null Hypothesis, H_0**	**Alternate Hypothesis, H_1**
One-tailed	$M = M_0$	$M < M_0$ or $M > M_0$
Two-tailed	$M = M_0$	$M \neq M_0$

where M is the population median and M_0 is the hypothesized population median.

ASSUMPTIONS:

Data is non-normally distributed, even after log transforming.

This test makes no assumption about the shape of the population distribution, therefore this test can handle a data set that is non-symmetric, that is skewed either to the left or the right.

The point to be noted:

The SIGN TEST simply computes a significance test of a hypothesized median value for a single data set. Like the 1 sample t-test you can choose whether you want to use a one-tailed or two-tailed distribution based on your hypothesis; basically, what do you want to test .

To properly analyze and interpret results of the *one-sample sign test*, you should be familiar with the following terms and concepts:

- One sample problem
- Independent samples
- Violation of test assumptions
- Distribution free tests
- Rank tests

One should be familiar with these terms and concepts, in order to apply the nonparametric tests. Failure to understand and properly apply the *one-sample sign test* or any nonparametric test may result in drawing erroneous conclusions from the data collected.

Additionally, we may have to check our data and take decision whether to go for t test or sign test preferably based on Normality test

Whether the median value of the data set is equal to some hypothesized value (H0: M = Mo), or do you want to test whether it is greater (or lesser) than that value (H0: M > or < Mo). Likewise, you can choose to set confidence intervals around M at some α level and see whether M0 falls within the range. This is equivalent to the significance test, just as in any t-TEST.

The kind of question you would be interested in asking is whether some observation is likely to belong to some data set you already have defined. That is to say, whether the observation of interest is a typical value of that data set.

Formally, you are either testing to see whether the observation is a good estimate of the median, or whether it falls within confidence levels set around that median.

The Principle:

Under the hypothesis that the sample median (m) is equal to some hypothesized value (Mo , so H0: M = Mo), then you would expect half the data set S of sample size n to be greater than the hypothesized value Mo. If S > 0.5n then M > Mo, and if S < 0.5n then M< Mo. The SIGN TEST simply computes whether there is a significant deviation from this assumption, and gives you a p value based on a binomial distribution.

If you are only interested in whether the hypothesized value is greater or lesser than the sample median (H0: M > or < Mo), the test uses the corresponding upper or lower tail of the distribution.

The workings for calculating the confidence intervals (CI) get complicated to do manually as you have to use a technique called non-linear interpolation that is about as fun to do as it sounds, therefore, let any software do it for you.

Basically, it is doing the same thing as in the parametric tests, setting a CI around a measure of central tendency, however, because we are now dealing with discrete data it isn't always possible to calculate exactly the CI corresponding to your assigned α level: Non-linear interpolation is simply a method of getting as close to that value as possible.

Procedure:

The procedure is very simple. Let a random sample x1,x2,...,xn of size n be drawn from a population with distribution function F(x) where F(x) is assumed to be continuous in close vicinity of a n average (say Median). Suppose the Median of F(x) is M then P[X=M]=0

To test H0:M=M0 against the alternative hypothesis H1: M≠M0 where M0 is the given value of the population median. We know that P[X>M0]=P[X<M0]=0.5.

Hence a null hypothesis under test is equivalent to H0:P[X>M0]=P[X<M0] against

H1: P[X>M0]≠P[X<M0]

To perform the sign test we take the differences Xi–M0 for i=1,2,...,n and consider the signs. Let the number of positive signs be U and negative signs (n-U) .

For the test we consider only the number of positive signs. In this way the data have been dichotomized which consists of number of positive signs and negative signs. The distribution of U given n is a Binomial distribution with p=P[X>M0].

Thus the null hypothesis H0 changes to H0:p=0.5.

So now we test H0:p=0.5 against H0:p ≠0.5.

Compute

$$P(X \le U) = \sum^{U} n_{c_x} (\frac{1}{2})^x (\frac{1}{2})^{n-x}$$

and this value is compared with α/2, the level of significance. If the calculated value of P(X≤U) is less than the predicted α/2, the null hypothesis is rejected

If n≥25, the value of Z is computed and the normal test is applied to decide about H0 where

$$Z = \frac{(U + 0.5) - np_0}{\sqrt{np_0(1 - p_0)}}$$ when $U<np_0$

$$Z = \frac{(U - 0.5) - np_0}{\sqrt{np_0(1 - p_0)}}$$ when $U>np_0$

where $p_0 = 0.5$

If $Z \geq Z_{\alpha/2}$ or $Z \leq Z_{1-\alpha/2}$. a null hypothesis is rejected which implies $|Z| \geq Z_{\alpha/2}$ a null hypothesis is rejected

One sided tests:

To test H0:p=0.5 against H0:p >0.5. Compute the value of Z and the normal test is applied to decide about H0 where Z is obtained using the above formula. If $Z > Z_{\alpha}$ a null hypothesis is rejected

To test H0:p=0.5 against H0:p <0.5. Compute the value of Z as above and the normal test is applied to decide about H0

If $Z < -Z_{\alpha}$, a null hypothesis is rejected

Sign test concepts

1Make null hypothesis about true median
2. Let S = number of values greater than median
3.Each sampled item is independent
4.If null hypothesis is true, S should have binomial distribution with success probability 0.5

The 1-Sample Sign is a nonparametric test of population location (median) and also calculates the corresponding point estimate and confidence interval. Its parametric counterpart is the 1-sample Z and 1-sample t tests.

Application -1:

A large group of senior citizens enrolled for adult evening classes at a university. To get a quick check on whether there has been an increase from last year's average age of 65.4 years, the director of the program takes a random sample of 15 of the enrollees, getting 68, 62, 70, 64, 61, 58, 65, 86, 88, 62, 60, 71, 60, 84, and 61 years.

Solution:

Here the null hypothesis

H0: the average age is 65.4

and the alternative hypothesis

H1: the average age is>65.4.

The level of significance $\alpha=0.05$.

Here n=15.

By subtracting 65.4 from each of the observations. The number of plus signs and minus signs are noted as follows

+ , - , +, -, - , - , -, + , + , - , - , + , - , + , -

We observe that the number of positive signs U=6.
Number of negative signs are 9.
We need to use the Binomial distribution with n=15 , p=0.5 (assuming + and - are equally likely to occur if there is no difference).

P(U≥6)=P(getting 6 or more pluses) = 0.849
Since this is more than 0.05 (level of significance), we do not reject the null hypothesis that there has been no increase since last year's average of 65.4.

Application -2

The PQR campany claims that the life time of a type of battery that it manufactures is more than 250 hours. A consumer advocate wishing to determine whether the claim is justified measures the life times of 24 of the company's batteries. Assuming the sample to be random, determine whether the company's claim is justified at the 0.05 significance level. Observations are :

271, 230, 198, 275, 282, 225, 284, 219
253, 216, 262, 288, 236, 291, 253, 224
264, 295, 211, 252, 294, 243, 272, 268

Solution:

Let H0 be the hypothesis that the company's batteries have a life time equal to 250 hours and let H1 be the hypothesis that they have a life time greater than 250 hours. To test H0 against H1 we can use sign test.

To do this we substract 250 from each entry of the above table and record the signs of differences as shown in the table below. We see that there are 15 plus signs and 9 minus signs

Using a one tailed test at the 0.05 significance level , we would reject H0 if the Z score were greater than 1.645. Since the Z score using a correction for continuity is (Since U > np_0 that is 15 > (24)(0.5))

$$Z = \frac{(U - 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{(15 - 0.5) - (24)(0.5)}{\sqrt{(24)(0.5)(0.5)}} = 1.02$$

Since Z is less than Z_α at 5% level of significance we cannot accept the alternative hypothesis. The company's claim cannot be justified at the 0.05 level

Consider a large population of elements each of which has a measurable value. Suppose that the distribution of the population values is continuous and that we are interested in testing hypothesis concerning the median or middle value of this distribution.

To test the hypothesis H0: M=M0 against H1: M≠M0, we can use an alternate method for the sign test as follows

Choose a random sample. Discard any values equal to M0. Let n be the number of values that remain. Let the test statistic be the number of values that are less than M0. If there are i such values then p value is

P value = 2 Min ($P[X \leq i], P[X \geq i]$) where X is a Binomial random variable with parameters n and ½. The null hypothesis is then rejected at all significance levels greater than or equal to the p value.

To find p value it is not necessary to compute both P[X≤i] and P[X≥i]. Rather we only need to compute the smaller of these two probabilities. Thus from a practical point of view the p can be expressed as

$$pvalue = \begin{Bmatrix} 2P\{X \le i\] if\ i \le n/2 \\ 2P\{X \ge i\] if\ i \ge n/2 \end{Bmatrix}$$

Where X is the Binomial variate with parameters n and 1/2

Application 3:

The inventory ordering policy of a particular shoe store is partly based on the belief that the median foot size of teenage boys is 10.25 inches. To test this hypothesis the foot size of each of the random sample of 50 boys was determined. Suppose that 36 boys had sizes in excess of 10.25 inches .

Does this disprove the hypothesis that the median size is 10.25?

Solution:

Let X be a Binomial random variable with parameters 50 and ½. Since 36 is larger than 50* (½)= 25, we see that the p value is

P value = 2P{X≥36]

We can now see the Normal approximation

E(X)= 50* (1/2)=25 V(X)= 12.5

P value = 2P{X≥36]= 2P{X≥35.5](the continuity correction)

$$pvalue = 2P[Z \geq 2.97] = 0.0030$$

Thus the belief that the median shoe size is 10.25 inches is rejected even at the 1% level of significance. There appears to be a strong evidence that the median shoe size is greater than 10.25 inches

Non-parametric tests are normally based on ranks of the data samples, and test hypotheses relating to quintiles of the probability distribution representing the population from which the data are drawn.