Sign Test

In testing hypotheses about a population mean we have used the z- and t-statistics. When we had a large ($n \geq 30$) random sample from a general population, we used the z-statistic. When we had a small ($n < 30$) random sample from a normal distribution, we used the t-statistic. We now consider the situation where we have a small sample from a population that is not (or cannot be assumed to be) normal. In cases like this we will use a ***distribution-free*** or ***nonparametric test.***

The ***sign test*** may be used to test hypotheses about the central tendency of a nonnormal probability distribution. The sign test provides inferences about the population median and not the population mean μ. We denote the population median by η ($\hat{e}ta$). Of course η is also the 50th percentile of the distribution and is less affected by the skewness of the distribution and outliers than is μ.

Example 1: A potential employee submitted to a test for substance abuse. The results of an analysis of eight independent measurements of a sample presented by the employee are as follows:

0.78 0.51 3.79 0.23 0.77 0.98 0.96 0.89

On the scale used by the testing lab, measurement values 1.00 indicate "normal" ranges and values greater than or equal to 1.00 are indicators of potential substance abuse. The lab reports a normal result as long as the median level for an individual is less than 1.00. Our objective is to determine whether the population median η (the true median level if an infinite number of measurements were made on this individual's sample) is less than 1.00. We identify our hypotheses as follows:

H_0: $\eta = 1.00$
Ha: $\eta < 1.00$

The one-tailed sign test is performed by first counting the number of sample measurements "favoring" the alternative hypothesis. (In this case 7 of the 8 measurements are less than 1.00) If H_0 is true we expect about half the measurements to fall on each side of the hypothesized median; if the alternative is true, we expect significantly more than half to fall below 1.00.

Test statistic: S = Number of measurements less than 1.00 (null hypothesized median)

Suppose we wish to test at the $\alpha = 0.05$ level of significance. The rejection region can be expressed in terms of the observed significance level, or p-value, of the test.

Rejection region: p-value ≤ 0.05

We observe that if H_0 is true the number of measurements less than 1.00 is a binomial random variable with $n = 8$ and $p = 0.5$. If H_0 is true what is the probability that 7 or more of 8 measurements will result is Success (be less than 1.00)? Binomial Table II in Appendix A (with $n = 8$ and $p = 0.5$) shows that

$P(x \geq 7) = 1 - P(x \leq 6) \approx 1 - 0.965 \approx 0.035$.

So, the probability that at least 7 or the 8 measurements would be less than 1.00 if the true median were 1.00 is 0.035. Hence the p-value of the test is 0.035 which is less than $\alpha = 0.05$. We have sufficient evidence to reject H_0. The laboratory can conclude at the $\alpha = 0.05$ level of significance that the true median level for the individual tested is less than 1.00.

Example 2: A manufacturer of CD players has found that the median time to failure for its players is 5,250 hours of use. The manufacturer obtains a sample of 20 CD players from a competitor and those players are tested. The failure times in the sample range from 5 hours to 6,575 hours and 14 of the 20 have failure times exceeding 5,250 hours. Is there evidence that the median failure time of the competitor's CD players differs from 5,250? (Use the $\alpha = 0.10$ level of significance.)

The null and alternative hypotheses are

$$H_0: \eta = 5{,}250 \text{ hours}$$
$$H_a: \eta \neq 5{,}250 \text{ hours}$$

We conduct a two-tailed test in this case.

Test statistic: S = the larger of S_1 and S_2 where S_1 is the number of measurements less than 5,250 and S_2 is the number of measurements greater than 5,250.

Rejection region: Reject if p-value $\leq \alpha = 0.10$

Observed significance level: p-value $= 2P(x \geq 14)$; x is a binomial distribution with $n = 20$ and $p = 0.5$

p-value $= 2[1 - P(x \leq 13] \approx 2[1 - 0.942] \approx 0.116$

The p-value is not in the rejection region, so we cannot reject the null hypothesis at the $\alpha = 0.10$ level of significance. The manufacturer should not conclude, on the basis of this sample, that the competitor's CD players have a median failure time that differs from 5,250.

See Example 6.9 in the textbook (p. 339) for an alternative approach to Example 2. The alternative approach is explained at the bottom of page 338.