

1 SAMPLE SIGN TEST

A non-parametric equivalent of the *1 SAMPLE T-TEST*.

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: 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, do you want to test whether the median value of the data set is equal to some hypothesized value ($H_0: \eta = \eta_o$), or do you want to test whether it is greater (or lesser) than that value ($H_0: \eta >$ or $< \eta_o$). Likewise, you can choose to set confidence intervals around η at some α level and see whether η_o falls within the range. This is equivalent to the significance test, just as in any *T-TEST*.

In English, 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 WORKINGS: Very simple. Under the hypothesis that the sample median (η) is equal to some hypothesized value (η_o , so $H_0: \eta = \eta_o$), then you would expect half the data set S of sample size n to be greater than the hypothesized value η_o . If $S > 0.5n$ then $\eta > \eta_o$, and if $S < 0.5n$ then $\eta < \eta_o$. 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 ($H_0: \eta >$ or $< \eta_o$), 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 MINITAB 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.

MINITAB EXAMPLE:

Suppose we have a data set of twenty-six observations (N=26) and we want to see if a value of 30 is a reasonable estimate of the mean:

1	1	2	2	3	3	4	5	5
6	7	7	8	10	20	22	25	27
33	40	42	50	55	75	80		

To test for normality we print out the descriptive statistics from MINITAB and run an Anderson-Darling normality test.

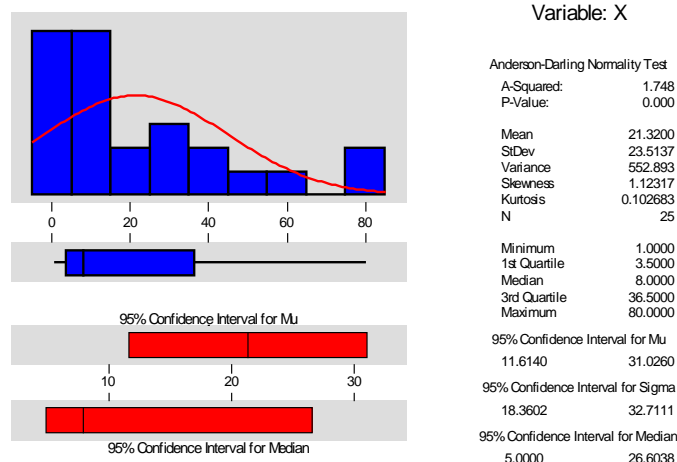
For the descriptive stats the MINITAB procedure is:

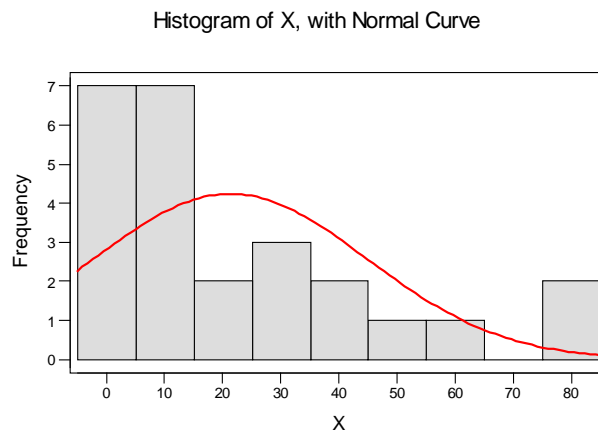
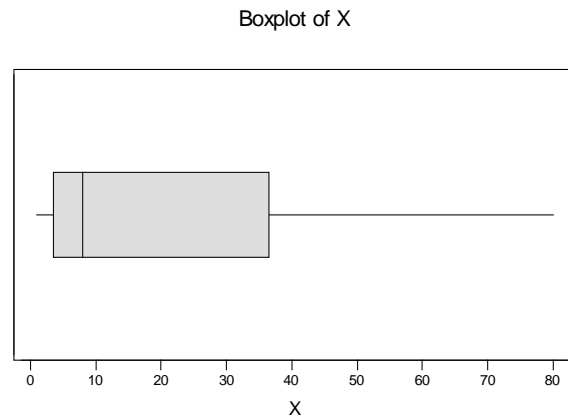
```

>STAT
  >BASIC STATISTICS
    >DESCRIPTIVE STATISTICS
      >Double click on the column your data is entered
        >GRAPHS: Click on BOXPLOT, GRAPHICAL SUMMARY,
and HISTOGRAM WITH NORMAL CURVE
          >OK
            >OK
  
```

We get the three following graphics.

Descriptive Statistics

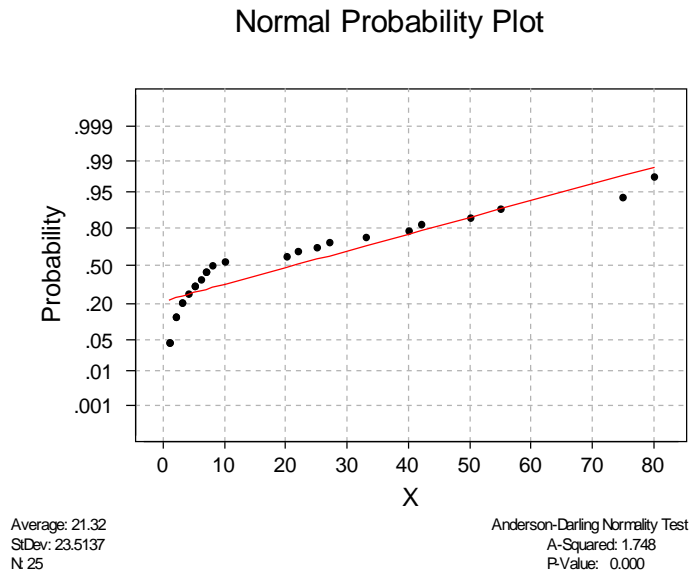




These data do not look normal, so we run a formal normality test (Anderson-Darling). The MINITAB procedure is:

```
>STAT
  >BASIC STATISTICS
    >NORMALITY TEST
      >Double click on the column your data is in
        >Choose ANDERSON-DARLING
          >OK
```

You will get the following output:



You see that the $p = 0.000$, meaning that the data is highly non-normal; this is obviously verified by looking at the above three plots from the descriptive stats.

From here we decide to perform a non-parametric *Sign test* to see if 30 is a reasonable estimate of the mean, that is to say if a value of 30 could belong to this distribution:

```

>STAT
  >NON-PARAMETRICS
    >1 SAMPLE SIGN
      >Double click on the column your data is in: for the Confidence level
click on CONFIDENCE INTERVAL and leave it at the default value (0.95)
        >OK
          >For a hypothesis test click on TEST MEDIAN, and enter the
hypothesized value 30 and leave it on the default value (not equal)
            >OK
  
```

And we get the following output:

Sign Confidence Interval

Sign confidence interval for median

	N	Median	Achieved Confidence	Confidence Interval	Position
X	25	8.00	0.8922	(5.00, 25.00)	9
			0.9500	(5.00, 26.60)	NLI
			0.9567	(5.00, 27.00)	8

This output gives you three CIs. If you remember, we are interested in the *NLI* output (non-linear interpolation). This says that our median is 8, with confidence intervals bounding it at 5 and 26.6. As our hypothesized value 30 lies outside this range we reject the null hypothesis in favor of the alternative and conclude that at the $\alpha = 0.05$ level we reject the hypothesis that a value of 30 is a reasonable estimate of the median.

For the hypothesis test:

Sign Test for Median

Sign test of median = 30.00 versus not = 30.00

	N	Below	Equal	Above	P	Median
X	25	18	0	7	0.0433	8.000

This output states the hypothesis at the top and then states that there are 18 observations below the hypothesized median, and 7 above (remember it should be around 50/50). The $p = 0.0433$, less than the stated α , therefore we conclude that at the $\alpha = 0.05$ level, we reject the null hypothesis that 30 is a reasonable estimate of the mean and accept the alternative hypothesis.

The results of the hypothesis test are obviously the same as the CI test. Which method you choose to use is dependent on personal preference and the exact type of question you are interested in answering.