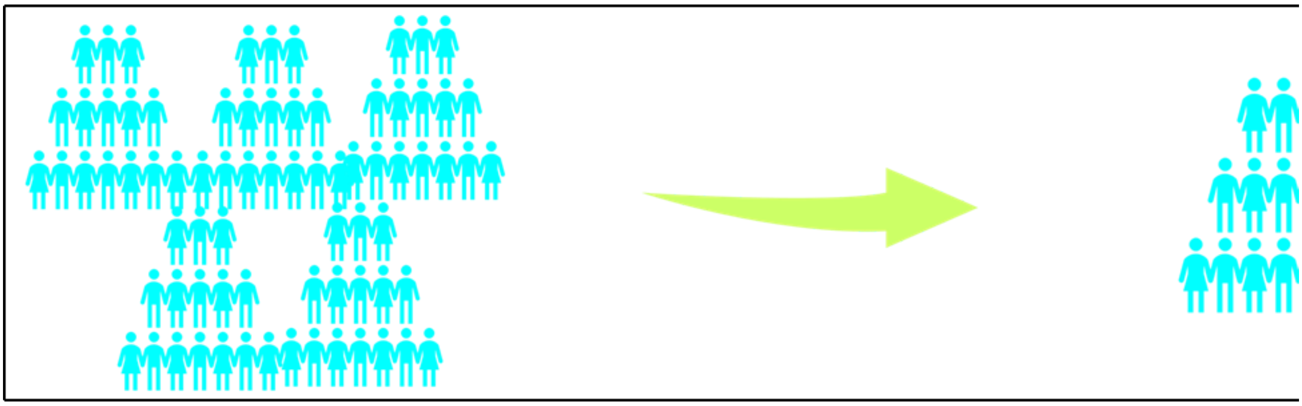
Examrace

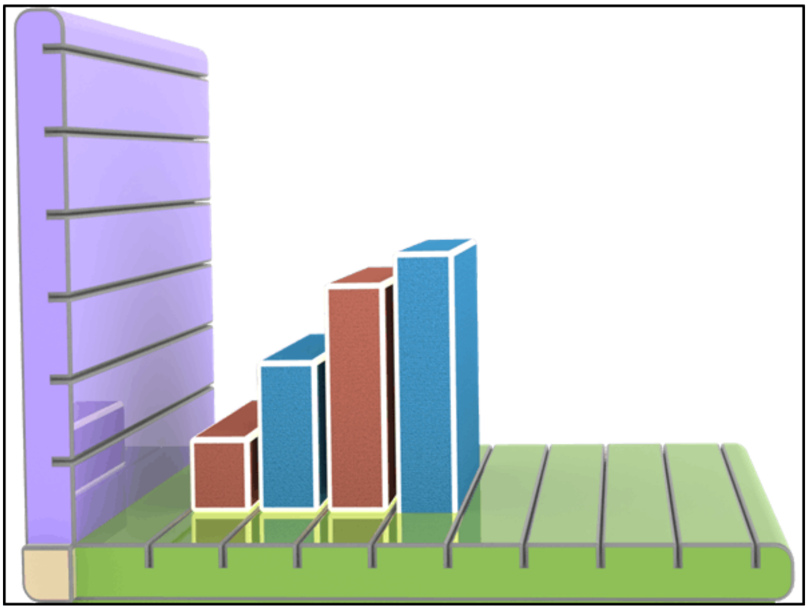
Testing for Normality, Tests for Normality, Graphical Methods YouTube Lecture Handouts

Testing for Normality



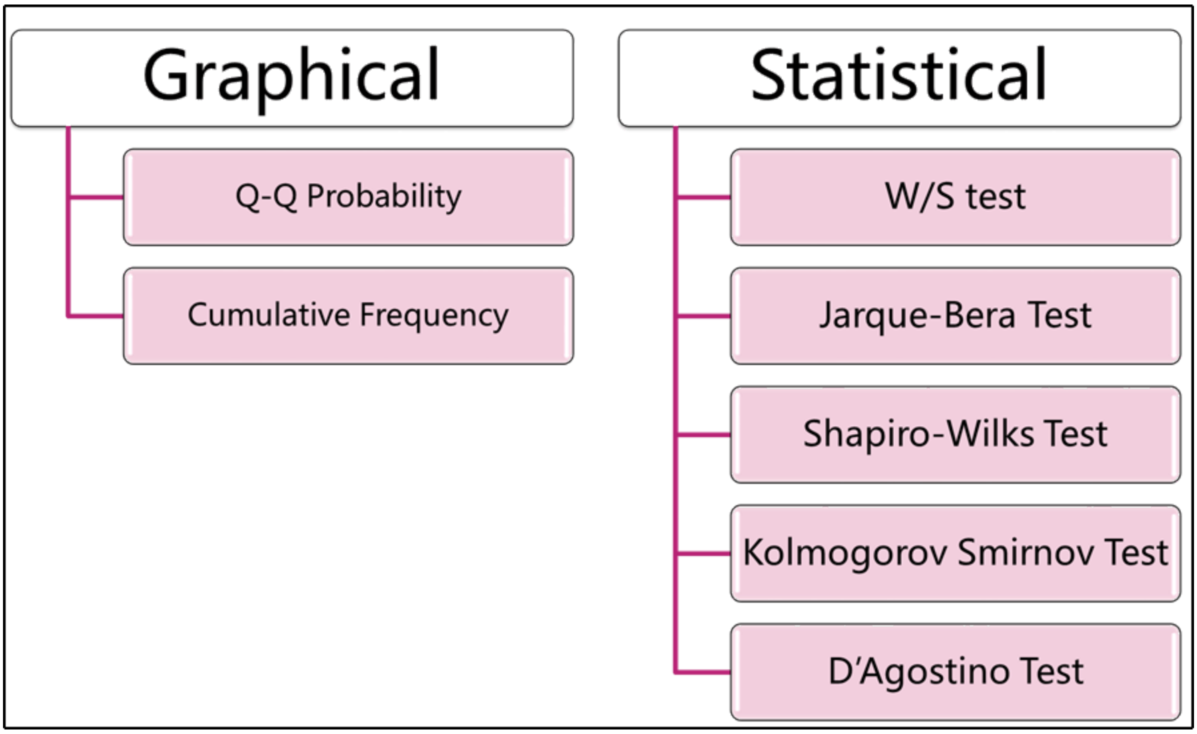
Population vs sample

How we test normality



- Difference between theoretical distribution and actual data – we require tests of normality
- When is non-normality a problem?
- Normality can be a problem when the sample size is small (< 50) .
- Highly skewed data create problems.
- Highly leptokurtic data are problematic, but not as much as skewed data.
- Normality becomes a serious concern when there is "activity" in the tails of the data set.
- Outliers are a problem. (Test used are Grubb's test and Dixon test)
- "Clumps" of data in the tails are worse.
- Final Words Concerning Normality Testing:
- Since it is a test, state a null and alternate hypothesis.
- If you perform a normality test, do not ignore the results.
- If the data are not normal, use non-parametric tests.
- If the data are normal, use parametric tests.
- **If you have groups of data, you MUST test each group for normality**.

Tests for Normality



- Statistical tests for normality are more precise since actual probabilities are calculated.
- Tests for normality calculate the probability that the sample was drawn from a normal population.

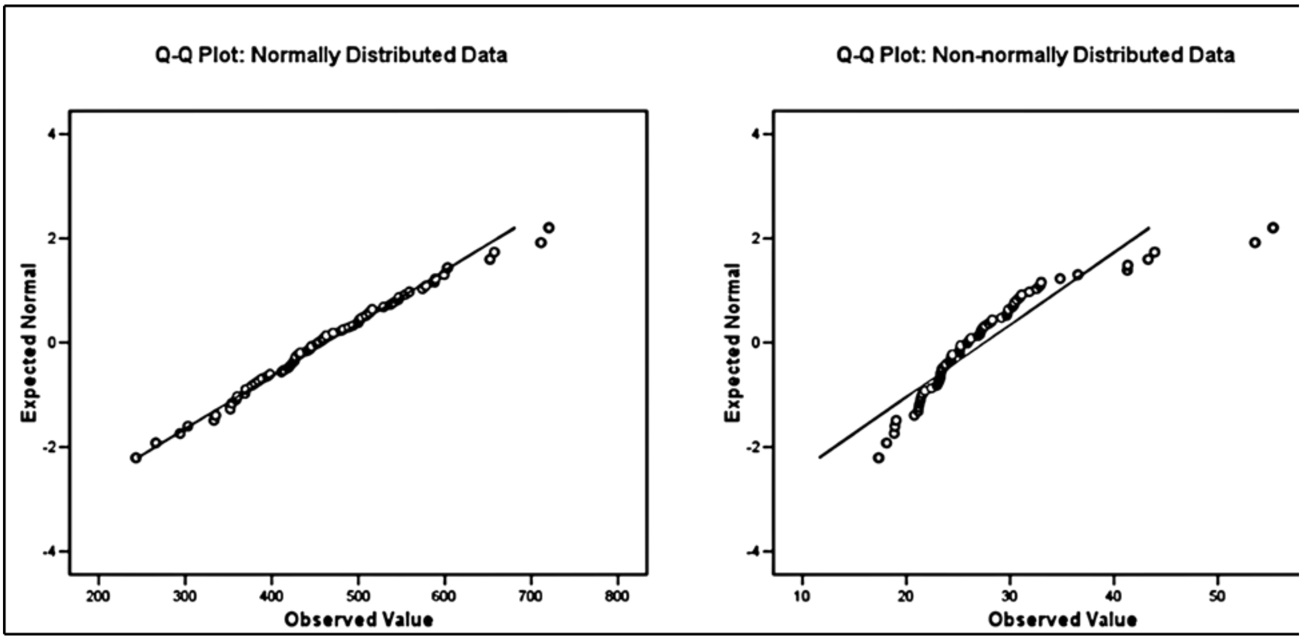
The hypotheses used are:

- Ho: The sample data are not significantly different from a normal population.
- Ha: The sample data are significantly different from a normal population.

When testing for normality:

- Probabilities > 0.05 indicate that the data are normal.
- Probabilities < 0.05 indicate that the data are NOT normal
- SPSS Normality Tests - Kolmogorov-Smirnov and Shapiro-Wilk.
- PAST Normality Tests - Shapiro-Wilk, Anderson-Darling, Lilliefors, Jarque-Bera.

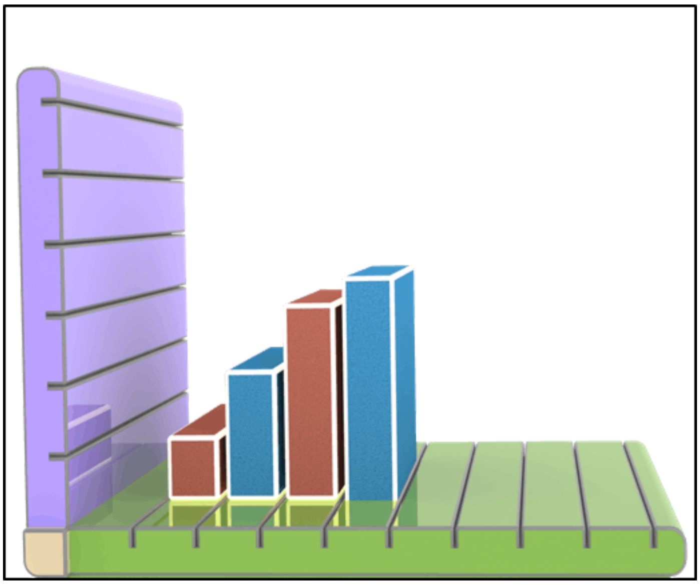
Q-Q Plots



Q-Q plots display the observed values against normally distributed data (represented by the line) .

Normally distributed data fall along the line

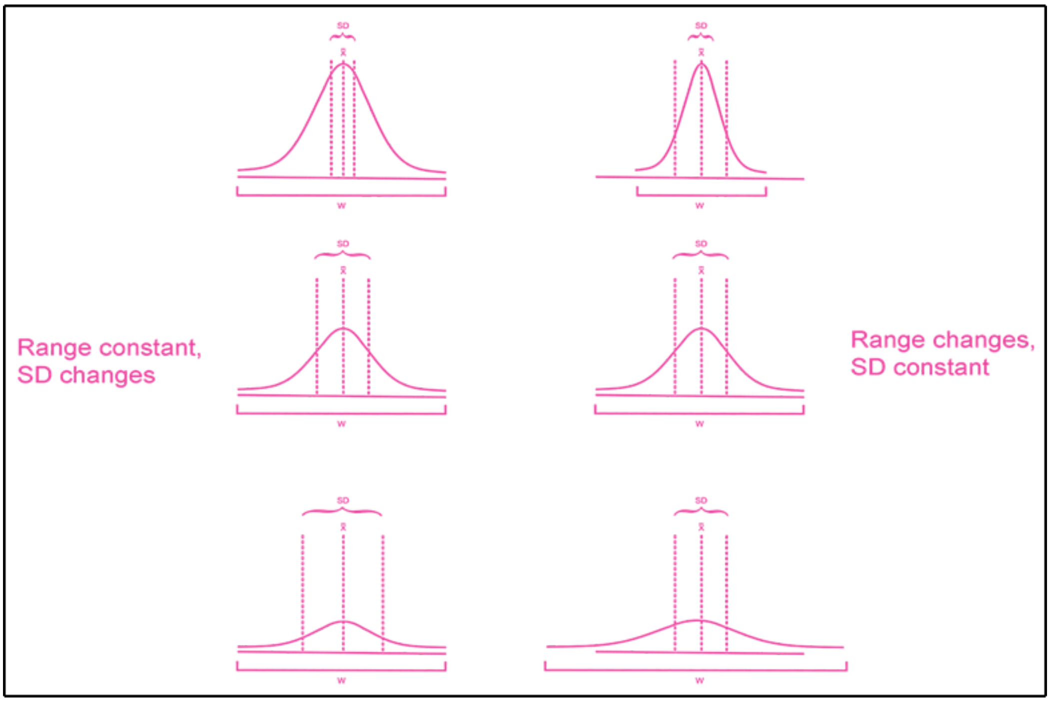
Graphical Methods

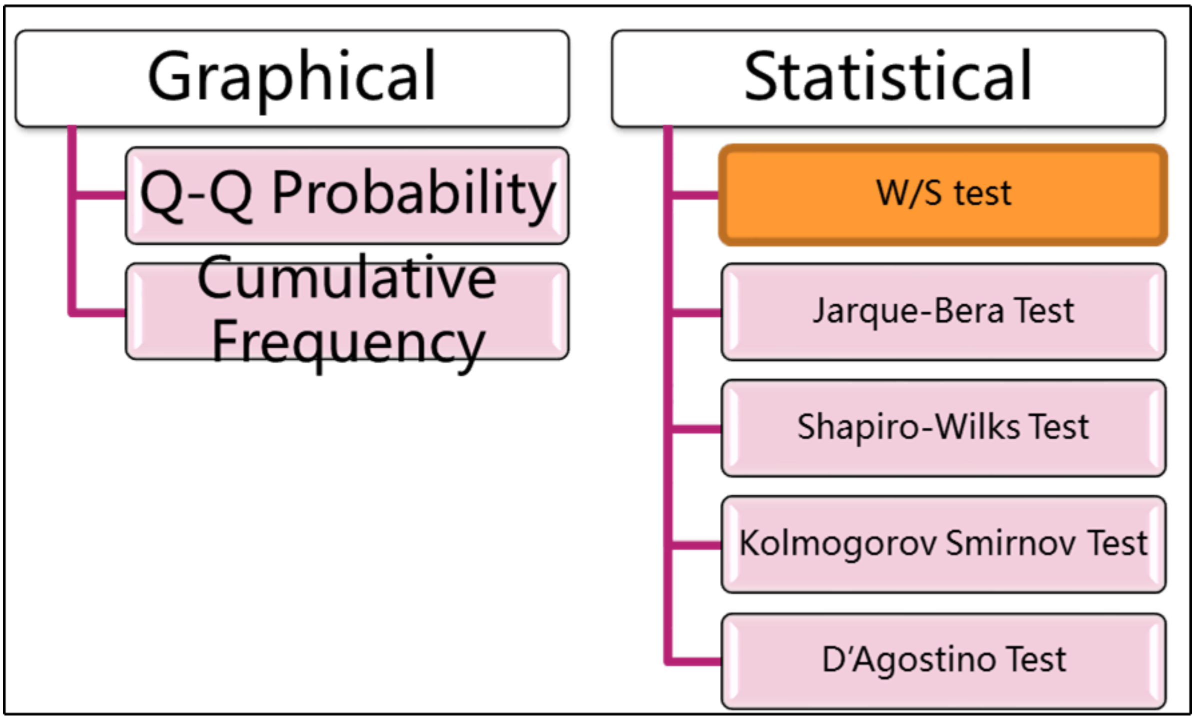


Graphical methods are typically not very useful when the sample size is small. This is a histogram of the last example. These data do not 'look' normal, but they are not statistically different from normal

W/S Test

$$q = \frac{w}{s}$$


SD
x̄
w
Range constant,
SD changes
Range changes,
SD constant



W/S Test for Normality

- A simple test that requires only the sample standard deviation and the data range.
- Should not be confused with the Shapiro-Wilk test.
- Based on the q statistic, which is the 'studentized' (meaning t distribution) range, or the range expressed in standard deviation units.
- Where q is the test statistic, w is the range of the data and s is the standard deviation.
- The test statistic q (Kanji 1994, table 14) is often reported as u in the literature.

Jarque-Bera Test

$$\text{JB} = \frac{n}{6}\left[S^2 + \frac{(K-3)^2}{24}\right]$$

Normality is one of the assumptions for many statistical tests, like the t test or F test; the Jarque-Bera test is usually run before one of these tests to confirm normality. It is usually used for large data sets, because other normality tests are not reliable when n is large

- Where: n is the sample size, S is sample skewness
- K is sample kurtosis

- In general, a large J-B value indicates that errors are **not** normally distributed.
- For sample sizes of 2,000 or larger, this test statistic is compared to a chi-squared distribution with 2 degrees of freedom (normality is rejected if the test statistic is greater than the chi-squared value) .
- The chi-square approximation requires large sample sizes to be accurate. For sample sizes less than 2,000, the critical value is determined via simulation.

Shapiro-Wilks

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

The test gives you a W value; small values indicate your sample is not normally distributed

Where:

- x_i are the ordered random sample values
- a_i are constants generated from the covariances, variances and means of the sample (size n) from a normally distributed sample.
- The test has limitations, most importantly that the test has a bias by sample size. The larger the sample, the more likely you'll get a statistically significant result.
- Univariate continuous data
- Numerator is slope of observed data vs expected normal values
- If H_0 is true then W should be 1.
- Highly sensitive and use graphical method to assess t-test assumptions

Kolmogorov-Smirnov (K-S Test)

$$E_n = \frac{n_i}{N}$$

It compares the observed versus the expected cumulative relative frequencies

Kolmogorov-Smirnov test uses the maximal absolute difference between these curves as its test statistic denoted by D.

- It only applies to continuous distributions.
- It tends to be more sensitive near the center of the distribution than at the tails. Determined by stimulation.
- The Kolmogorov-Smirnov (K-S) test is based on the empirical distribution function (ECDF) . Given N *ordered* data points $Y_1, Y_2, \ldots, Y_N$, the ECDF is defined as

$$\text{En} = \frac{N(I)}{N}$$

- where **n (i)** is the number of points less than Y_i and the Y_i are ordered from smallest to largest value. This is a step function that increases by $\frac{1}{N}$ at the value of each ordered data point.
- Calculated value < critical value (acceptance criteria) default is 0.565 as critical value

D'agostino Test

$$D = \frac{T}{\sqrt{n^3 \text{ SS}}}$$

$$T = \sum \left(i - \frac{n+1}{2}\right) x_i$$

A very powerful test for departures from normality.

Based on the D statistic, which gives an upper and lower critical value.

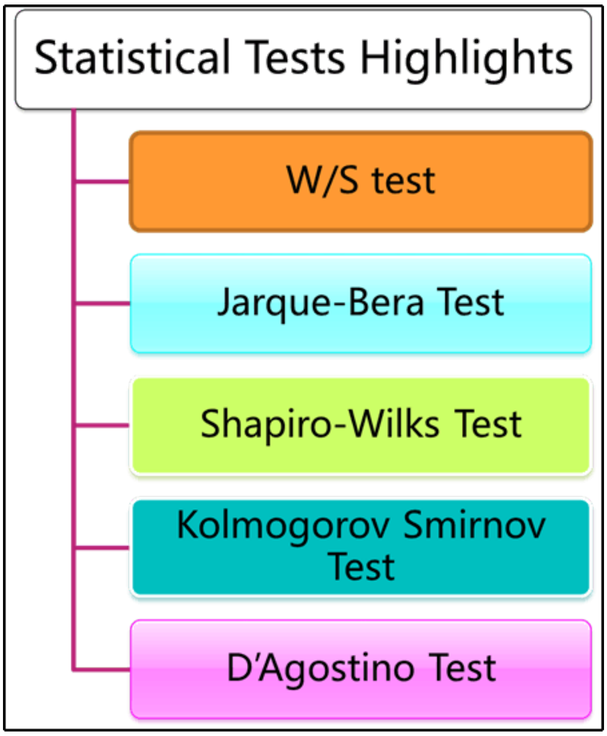
Where D is the test statistic, SS is the sum of squares of the data and n is the sample size, and I is the order or rank of observation x. The df for this test is n (sample size) .

First, the data are ordered from smallest to largest or largest to smallest

$\frac{n+1}{2}$ is middle term of dataset

$i - \frac{n+1}{2}$ is observations' distance from middle

Notice that as the sample size increases, the probabilities decrease. In other words, **it gets harder** to meet the normality assumption as the sample size increases since even small departures from normality are detected.



W/S or studentized range (q) :

- Simple, very good for symmetrical distributions and short tails.
- Very bad with asymmetry.

Shapiro Wilk (W) :

- Powerful omnibus test. Not good with small samples or discrete data.
- Good power with symmetrical, short, and long tails. Good with asymmetry.

Jarque-Bera (JB) :

- Good with symmetric and long-tailed distributions.
- Less powerful with asymmetry, and poor power with bimodal data.

D'Agostino (D or Y) :

- Good with symmetric and very good with long-tailed distributions.
- Less powerful with asymmetry.

Anderson-Darling (A) :

- Similar in power to Shapiro-Wilk but has less power with asymmetry.
- Works well with discrete data.

Distance tests (Kolmogorov-Smirnov, Lillifors, and Chi2) :

- All tend to have lower power. Data have to be very non-normal to reject Ho.
- These tests can outperform other tests when using discrete or grouped data.
- Several goodness-of-fit tests, such as the Anderson-Darling test and the Cramer Von-Mises test, are refinements of the K-S test. As these refined tests are generally considered to be more powerful than the original K-S test, many analysts prefer them. In addition, the advantage for the K-S test of having the critical values be independent of the underlying distribution is not as much of an advantage as first appears.

✍ Manishika