Comparing Four Normality Tests

Suppose you have a set of data, and you want to apply some statistical tests to the data. You read up on some of the tests, and you discover that the test assumes the data are normally distributed.

How do you find out if your data are normally distributed? You can make a histogram and see if it looks



like that familiar bell-shaped curve. If it does, you can assume that the data are normally distributed. You can also construct a normal probability plot to see if the data falls along a straight line for the most part. If it does, you can assume that the data are normally distributed.

But some of us like numbers. Is there a statistic that will tell us if the data are normally distributed? Yes, several, when used in combination with a p-value.

This publication looks at four different normality tests: the Anderson-Darling (AD) test, the Kolmogorov-Smirnov (KS) test, the Lilliefors test, and the Shapiro-Wilk (SW) test. An explanation of how each test works is given, along with the steps in using each test.

In this publication:

- <u>The Hypotheses</u>
- Example Data
- Anderson-Darling (AD) Test
- Kolmogorov-Smirnov (KS) Test
- Lilliefors Test
- Shapiro-Wilk (SW) Test
- Which is Best?
- <u>Summary</u>
- Quick Links

All four normality tests are included in the <u>SPC for Excel software</u>.

The Hypotheses

The two hypotheses for the four normality tests are given below:

 $H_{\ensuremath{\tiny 0}}$: The data follows the normal distribution

 $H_1:$ The data do not follow the normal distribution

Our approach for each test will be to calculate a statistic for the example data and then determine the p-value for that statistic to help us determine if the data are normally distributed or not. Remember the p ("probability") value is the probability of getting a result that is more extreme if the null hypothesis is true. If the p-value is small (e.g., <0.05), you conclude that the data do not follow the normal distribution. If the p-value is large (e.g., > 0.05), you conclude that the data are normally distributed.

Remember that you choose the significance level even though many people just use 0.05 the vast majority of the time. The significance level is often called alpha (α). We will use alpha = 0.05 in this publication.

Example Data

Suppose we have a dataset of 50 samples. We want to know if this data is normally distributed. The data are shown in Table 1.

0.35	0.18	
0.25	0.37	
0.42	0.20	
0.16	0.19	
0.55	0.28	
0.11	0.10	
0.51	0.21	
0.18	0.31	
0.20	0.04	
0.38	0.12	
0.54	0.21	
0.08	0.36	
0.36	0.65	
0.27	0.16	
0.13	0.07	
0.35	0.08	
0.13	0.17	
0.33	0.54	
0.17	0.06	
0.21	0.24	
0.26	0.31	
0.39	0.15	
0.15	0.39	
0.31	0.48	
0.20	0.10	

Table 1: Example Data

One way of seeing if you have a normal distribution is to construct a histogram. The histogram for these 50 samples is shown in Figure 1.



Figure 1: Example Data Histogram

Does the data in the histogram look like a normal distribution? With just 50 data points, it is sometimes hard to know for sure. It looks like it might be a stretch to call this a normal distribution.

So, we can make a normal probability plot and see if the points fall along a straight line. Figure 2 shows the normal probability plot for the data.

Do the points tend to lie along the straight line? Some do, some don't. It is probably questionable if this comes from a normal distribution.

But to confirm it further, we can use one of the four tests above. In fact, we will use all four to see if there are differences, starting with the Anderson-Darling statistic.





Anderson-Darling (AD) Test

The Anderson-Darling test was developed in 1952 by Theodore Anderson and Donald Darling. This test checks whether a dataset comes from a normal distribution (it works with other distributions as well) by comparing the actual distribution of the data to what you would expect if the data were normally distributed.

The test involves calculating the AD statistic and then determining the p-value for that AD value.

The steps in the calculation of AD are given below.

- 1. Sort the n samples from smallest to largest
- 2. Calculate the average \overline{X} and the standard deviation (s) of the data
- 3. Standardized the data using:

$$z_i = \frac{X_i - \bar{X}}{s}$$

4. Calculate the cumulative distribution function (CDF) for the normal distribution for each z value: F(z_i). This calculates how far off the data are from the expected values of a normal distribution (using the mean and standard deviation of the sample data).

5. Calculate the AD statistic as shown below. The value of AD summarizes how different the sample data is from a true normal distribution. The higher the value, the less likely the data is to be from a normal distribution.

$$AD = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) [lnF(z_i) + \ln(1 - F(z_{n-i+1}))]$$

The equation above for AD puts more weight on the differences between the sample data and a true normal distribution in the tails. It does this through these two terms in the equation: $lnF(X_i)$ and $ln(1 - F(X_{n-i+1}))$. These terms become large when they are close to 0 and 1, i.e., in the tails. So, the AD test is sensitive to differences from a normal distribution in the tails.

For the data in Table 1, the calculated value of AD is 0.2993.

The value of AD needs to be adjusted for small sample sizes. The adjusted AD value is given by:

$$AD^* = AD(1 + \frac{.75}{n} + \frac{2.25}{n^2})$$

For the data in Table 1, $AD^* = 0.8791$.

Does the value of AD* imply the data are normally distributed? You don't know using the value of AD* by itself. This is determined by the p-value.

The calculation of the p-value for the AD test is not straightforward. The reference most people use is R.B. D'Augostino and M.A. Stephens, Eds., 1986, *Goodness-of-Fit Techniques*, Marcel Dekker.

There are different equations depending on the value of AD*. These are given by:

- If AD*=>0.6, then p = exp(1.2937 5.709(AD*)+ 0.0186(AD*)²)
- If 0.34 < AD* < .6, then p = exp(0.9177 4.279(AD*) 1.38(AD*)²)
- If 0.2 < AD* < 0.34, then p = 1 exp(-8.318 + 42.796(AD*)- 59.938(AD*)²)
- If AD* <= 0.2, then p = 1 exp(-13.436 + 101.14(AD*)- 223.73(AD*)²)

The value of AD* for our example falls in the range of the first bullet above. Substituting in the value of AD* into the equation gives the p-value = 0.0245. Since this probability is low (2.45%), we assume that the null hypothesis is false, and the data are not normally distributed.

Kolmogorov-Smirnov (KS) Test

Two mathematicians, Andrey Kolmogorov and Nikolai Smirnov, developed this test in the 1930s. This test is a non-parametric test used to determine if a sample comes from a normal distribution. It can be used with other distributions also. The KS test compares the empirical distribution function (EDF) of the sample with the cumulative distribution function (CDF) of the normal distribution.

The steps in the calculation of KS are given below.

- 1. Sort the n samples from smallest to largest.
- 2. Calculate the average \overline{X} and the standard deviation (s) of the data
- 3. Standardized the data using:

$$z_i = \frac{X_i - \bar{X}}{s}$$

4. Calculate the empirical CDF:

$$F_n(z_i) = \frac{i}{n}$$

- 5. Calculate the theoretical normal CDF: $F(z_i)$; you can use the Excel function NORM.S.DIST for this calculation.
- 6. Calculate D+ and D-:

$$D += max\left(\frac{i}{n} - F(z_i)\right)$$
$$D -= max\left(F(z_i) - \frac{i-1}{n}\right)$$

7. Calculate D, which is the maximum of D+ and D-.

Note that D+ is the maximum positive deviation between the empirical CDF, $F_n(z_i)$, and the theoretical CDF, $F(z_i)$. D- is the maximum negative deviation. D is the larger D+ and D-.

The value of KS for the data in Table 1 is D = 0.1515. You can't simply compare the normality statistics to find anything out. You have to go to the p-values each time. There are tables of critical values for D to compare the calculated D value to. You can also use the equations below to determine the p-value.

$$\lambda = \left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right)D$$
$$p(\lambda) = 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2\lambda^2}$$

© 2025 BPI Consulting, LLC www.spcforexcel.com The p-value from the above equation converges rapidly usually. The p-value for the example data in Table 1 is 0.1844. Note that it is larger than 0.05, which would imply that the data are normally distributed.

This conflicts with the Anderson-Darling results. What is going on? There is a problem with using the p-values calculated as shown above for the KS test. It is conservative when used with the average and standard deviation calculated from the sample. The KS test is really designed to be used with known average and known standard deviation from the distribution population – the "true" values – which we seldom know. This leads to the KS test failing to reject that the data comes from a normal distribution when it should reject it.

You should use a modified KS test with the Lilliefors p-value.

Lilliefors Test

The Lilliefors test was developed in 1967 by Hubert W. Lilliefors, an American statistician. Lilliefors adjusted the critical values for D to account for using the sample averages and standard deviations.

The Lilliefors test accounts for the fact that the mean and standard deviation are estimated from the data, which affects the distribution of the D statistic. The calculation of the D statistic is the same as for the KS test.

The difference is in determining the p-value associated with that D value. The easiest way to get this p-value is from a table of Lilliefors critical D values. You look up the critical D value in the table based on n = sample size and alpha = significance level and then decide based on the following:

 $D < D_{\mbox{\tiny critical}}$: Assume that the data does come from a normal distribution

 $D > D_{critical}$: Assume that the data does not come from a normal distribution

You can also use interpolation to find the p-value. Doing this gives a p-value < 0.01 for the data in Table 1. This definitely says that the data are not normally distributed.

You can download a table of the critical D values at this link.

Shapiro-Wilk (SW) Test

The Shapiro-Wilk test was developed in 1965 by Samuel Shapiro and Martin Wilk to test for normality. It is only used to determine if the data is normally distributed or not. It is not used with other distributions. The steps in performing the SW test are given below.

- 1. Sort the n samples from smallest to largest.
- 2. Determine the values of m_i using:

$$m_i = \Phi^{-1} \left(\frac{i - 0.375)}{n + .25} \right)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution (NORM.S.INV in Excel) and i is the ith number in sorted data.

3. Calculate m.

$$m = \sum_{i=1}^{n} m_i^2$$

4. Calculate the "a" coefficients.

$$u = \frac{1}{\sqrt{n}}$$

 $a_n = -2.706056u^5 + 4.434685u^4 - 2.07119u^3 - 0.147981u^2 + 0.22117u + m_m m^{-0.5}$

$$\begin{aligned} a_{n-1} &= -3.582633u^5 + 5.682633u^4 - 1.752461u^3 - 0.293762u^2 + 0.042981u \\ &+ m_{m-1}m^{-0.5} \end{aligned}$$

$$\begin{aligned} a_i &= \frac{m_i}{\sqrt{\varepsilon}} \text{ for } 2 < i < n-1 \\ a_2 &= -a_{n-1} \\ a_1 &= -a_n \\ \varepsilon &= \frac{m - 2m_n^2 - 2m_{n-1}^2}{1 - 2a_n^2 - 2a_{n-1}^2} \end{aligned}$$

5. Calculate the W test statistic:

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

6. Determine the p-value by taking the standard normal distribution of the z value using the following:

$$\mu = 0.0038915(LN n)^3 - 0.083751(LN n)^2 - 0.31082(Ln n) - 1.5861$$

$$\sigma = \exp((0.0030302(LN n)^2 - 0.082676(LN n) - 0.4803)$$

$$z = \frac{\ln(1-W) - \mu}{\sigma}$$

The p-value is 1 - NORMSDIST(z).

For the data in Table 1, W = 0.9438. W is a measure of how close the data is to a normal distribution. The p-value associated with that value of W is 0.0191. This is less than 0.05, so we conclude that the data does not come from a normal distribution.

Which is Best?

The results for the data in Table 1 are shown in Table 2 below.

Normality Test	Statistic	p-Value	Normal?
Anderson-Darling	0.08791	0.0245	No
Kolmogorov-Smirnov	0.1515	0.1844	Yes
Lilliefors	0.1515	<0.01	No
Shapiro-Wilk	0.9438	0.0191	No

Table 2: Normality Results

The Kolmogorov-Smirnov test incorrectly assumes that the data are normally distributed. The other three tests correctly point out that the data are not normally distributed.

Of course, one data set doesn't tell us for sure which tests are better. From the literature, it appears that the following is true for detecting normality.

The Shapiro-Wilk test has the best power for detecting distributions that are not normal. It is good for sample sizes less than 5000.

The Anderson-Darling test is next. It is very powerful especially in the tails of the distribution.

Next is the Lilliefors test. It is more powerful than the KS test since it takes into account calculating the average and standard deviation from the sample.

And last is the Kolmogorov-Smirnov test. It is only good if you know the true average and standard deviation of the population.

Summary

This publication reviewed four different normality tests. The steps in using each test were covered. A statistic for each test is calculated with a p-value calculated to determine if the data are normal or not. For normality studies, the Shapiro-Wilk and the Anderson-Darling test are the best.

Quick Links

Visit our home page <u>SPC for Excel Software</u> <u>Download SPC for Excel Demo</u> <u>SPC Training</u> <u>SPC Consulting</u> <u>SPC Knowledge Base</u> <u>Ordering Information</u>

Thanks so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese