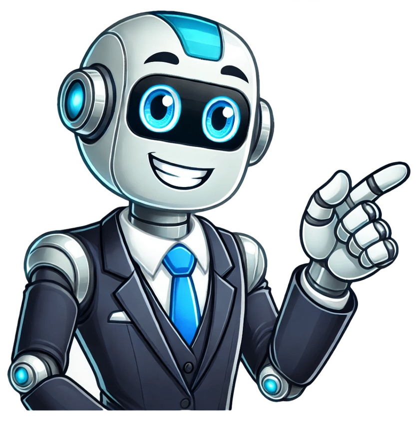

How to test for normal distribution

Use this normality test calculator to easily assess if the normality assumption can be applied to your data by using a battery of mis-specification tests. Currently supports: Shapiro-Wilk test / Shapiro-Francia test (n < 50 / n > 50), Anderson-Darling test, Jarque & Bera test, Cramer-von Mises test, d'Agostino-Pearson test. Plots a histogram of the data with a normal distribution overlay. Quick navigation: A test of normality in statistics and probability theory is used to quantify if a certain sample was generated from a population with a normal distribution via a process that produces independent and identically-distributed values. Normality tests can be based on the 3-rd and 4-th central moments (skewness and kurtosis), on regressions/correlations stemming from P-P and Q-Q plots or on distances defined using the empirical cumulative distribution functions (ecdf). The Null hypothesis can generally be stated as: "data can be modelled using the normal distribution", but since some normality tests also check if the data is independent and identically distributed (IID) a low p-value from these tests may be either due to a non-normal distribution or due to the IID assumption not holding. Separate tests for independence and heterogeneity can be performed to rule out those possibilities. Tests for normality like the Shapiro-Wilk are useful since many widely used statistical methods work under the assumption of normally-distributed data and may require alteration in order to accommodate non-normal data. Using a statistical test designed under the assumption of Normal or NIID data when the data is not normal renders the statistical model inadequate and the results meaningless, regardless if one is dealing with experimental or observational data (regressions, correlations, etc.). Normality tests such as those implemented in our normality test calculator should be run on the full data without removing any outliers, unless the reason for the outlier is known and its removal from the analysis as a whole can be readily justified (e.g. erroneously recorded data, data from source later proven to be unreliable, etc.). Interpreting the outcome of tests for normality The outcomes generated by our normality calculator consist of the p-value from each test and the test statistic (e.g. W, JB, K2). A lower p-value is a stronger signal for a discrepancy. Conventionally values under 0.05 are considered strong evidence for departure from normality (or IID, for some tests). Since the null is that the data is normal, the alternative is that it is not normal, but note that these tests do not point to a particular alternative distribution. However, the opposite is not necessarily true: a high p-value, say 0.3, might be due to the low sensitivity of the test relative to the number of data points you have entered and the type of distribution. That said, with a sufficiently large sample size a high p-value can be treated as evidence for lack of discrepancy. See "Supported Tests" below for a brief discussion on the relative sensitivity of some of the tests. With smaller sample sizes and/or distributions close to the normal it is expected to see some tests detect a departure from normality (very low p-values) while others show much higher p-values. This is most likely due to different sensitivity of the various tests towards different types and sizes of departures. If even a single mis-specification test results in a low p-value the normality assumption should be reconsidered, usually through re-specification. Switching to non-parametric tests is generally not recommended as it leads to loss of specificity and thus to more vague statistical inferences. Supported tests This online normality calculator currently supports the following tests: Shapiro-Wilk / Shapiro-Francia, Anderson-Darling, Cramer-von Mises, d'Agostino-Pearson and the Jarque & Bera test. The following tests are not supported since they have significantly inferior sensitivity: Kolmogorov-Smirnov test, Ryan-Joiner test, Lilliefors-van Soest test. While most who want to check their data for normality would search for the Shapiro-wilk test online, Mbah & Paothong (2014) [1] demonstrate via a comparison of several of the most-widely used tests across nine of the most-popular tests for normality that the Shapiro-Francia test is generally the most powerful, followed by the Shapiro-Wilk test. The Anderson Darling test is most sensitive under certain conditions, followed by the D’Agostino and Pearson. The Jarque-Bera test outperforms all against several distributions but with considerably high sample sizes (hundreds of data points). More on each of the supported tests below. The Shapiro-Wilk test / Shapiro-Francia test The Shapiro-Wilk test is a regression/correlation-based test using the ordered sample. It results in the W statistic which is scale and origin invariant and can thus test the composite null hypothesis of normality. It was devised in 1965 by Samuel Shapiro and Martin Wilk who tabulated linear coefficients for computing W for samples of up to 50 data points [2]. The test is consistent against all alternatives. Shapiro in collaboration with Francia proposed an extension of the method for handling samples with more than 50 data points in 1972 [3]: the Shapiro-Francia test, which is what our Shapiro-Wilk test calculator uses automatically if you supply it with more than 50 data points. Some people incorrectly refer to this test as the Shapiro-Wilk test, but it is different and in fact performs better than the Shapiro-Wilk test as it is more sensitive against most distributions even for sample sizes smaller than 50 [1]. In computing the W statistic we employ the Royston method [4] with a maximum sample size of 5,000. The test assumes a random sample and thus a violation of the IID assumption may result in a low p-value even if the underlying distribution is normal, therefore additional tests for independence and heterogeneity are recommended if only the Shapiro-Wilk or Shapiro-Francia test results in a p-value below the desired significance threshold. The Cramer-von Mises test The Cramer-von Mises goodness-of-fit test is based on the empirical distribution and an ordered statistic [5,6]. It is distribution-free (can be used for other distributions as well) omnibus test alternative to the Kolmogorov-Smirnov test (also ecdf-based). The p-value is based on the largest discrepancy between the empirical distribution and the hypothetical (normal, in this case) distribution. In terms of power against commonly-encountered alternatives it doesn't shine compared to the rest of the test in our goodness-of-fit calculator, but it is still widely used. The Anderson-Darling test The Anderson-Darling normality test [7] is a modification of the Cramer-von Mises approach and is thus a distance-test based on the empirical cumulative distribution function and distribution-free in its generic form. Compared with the Cramer–von Mises distance, the Anderson–Darling distance places more weight on observations in the tails of the distribution. It shows decent sensitivity against a variety of distributions, most notably the Laplace and Uniform distribution. The d'Agostino-Pearson test The d'Agostino-Pearson test a.k.a. as the D'Agostino's K-squared test is a normality test based on moments [8]. More specifically, it combines a test of skewness and a test for excess kurtosis into an omnibus skewness-kurtosis test which results in the K2 statistic. Due to its reliance on moments this test is generally less powerful than the SW/SF tests above as it ignores not just the dependence between the moments themselves, but also any existing higher-order moments making it lose all power if a distribution is non-normal but shows little deviation in terms of skewness and kurtosis. However the test has really good power against data from a uniform distribution which is why we have included it. The K2 statistic is only approximately X2-distributed due to the dependence between the two moments used so p-values may in fact be rough approximations at small sample sizes. The Jarque & Bera test The Jarque-Bera test [9] is another normality test based on moments our normality calculator supports. It is one of the simplest, combining the skewness and kurtosis into a single JB statistic which is asymptotically X2 distributed. This asymptotic property is why it performs poorly with small sample sizes, but can be the most sensitive test against a number of alternatives such as the uniform, logistic, Laplace and t-distribution given the sample size is in the hundreds or small thousands. The Jarque-Bera test may have zero power to detect departures towards distributions with 0 skewness and kurtosis of 3 (excess kurtosis of 0) like the Tukey λ distribution for certain values of λ. Practical examples Let us see how the normality test calculator works in practice. Clicking on this link will reload the page with a set of example data in the tool and the results from the battery of normality tests supported. It should look something like so (the histogram was generated in an external tool): As we can see the data resembles normal, but it has a rather heavy right tail and two of the tests: the Shapiro-Francia and the Anderson-Darling are sensitive enough to infer this from the limited sample size. The Cramer-von Mises test with a p-value of 0.075 is a close third. Given these statistics we have more than enough evidence to rule out normality for most practical purposes and to seek a different distribution which more appropriately fits the empirical data. We might also want to check the independence and heterogeneity of the data. 1 Mbah A.K. & Paothong A. (2014) "Shapiro-Francia test compared to other normality test using expected p-value", Journal of Statistical Computation and Simulation, 85:3002-3016; DOI: 10.1080/00949655.2014.947986 2 Shapiro S.S. & Wilk M.B. (1965) "An analysis of variance test for normality (complete samples)", Biometrika, 52:591–611 3 Shapiro S.S. & Francia R.S. (1972) "An approximate analysis of variance test for normality", Journal of the American Statistical Association, 67:215–216. 4 Royston P. (1993) "A Pocket-Calculator Algorithm for the Shapiro-Francia Test for Normality - An Application to Medicine", Statistics in Medicine, 12(2):181-184; DOI: 10.1002/sim.4780120209 5 Cramer H. (1928) "On the composition of elementary errors" Skandinavisk Aktuarietidskrift, 11:13–74, 141–180 6 von Mises R. (1931) "Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik" Julius Springer 7 Anderson T.W. & Darling D.A. (1954) "A Test of Goodness of Fit", Journal of the American Statistical Association 49:765-769 8 D'Agostino R.B., Pearson E.S. (1973) "Tests for Departure from Normality", Biometrika 60:613-622. 9 Jarque C.M., Bera A.K. (1987) "A test for normality of observations and regression residuals" International Statistical Review 55(2):163–172 You can test the hypothesis that your data were sampled from a Normal (Gaussian) distribution visually (with QQ-plots and histograms) or statistically (with tests such as D'Agostino-Pearson and Kolmogorov-Smirnov). However, it’s rare to need to test if your data are normal. Most likely you’re fitting some type of statistical model to your data such as ANOVA, linear regression, and nonlinear regression. In these cases, the assumption is that the residuals, the deviations between the model predictions and the observed data, are sampled from a normally distribution. The residuals need to be approximately normally distributed to get valid statistical inference such as confidence intervals, coefficient estimates, and p values. This means that the data don’t necessarily need to be normally distributed, but the residuals do. In this article, we will take a deeper dive into the subject of normality testing, including: Statistical test for normality with common statistical models How to determine if data is normally distributed using visual and statistical tests Normally distributed data examples What to do if the residuals are not normal How to test for normality with common statistical models Linear and nonlinear regression With simple linear regression, the residuals are the vertical distance from the observed data to the line. In this case, the tests for normality should be performed on the residuals, not the raw data. The same idea applies to nonlinear regression, where the model fits a curve instead of a straight line. The p-values and confidence intervals are based on the assumption that the residuals are normally distributed. Discover the easiest way to test your data using linear regression with a free 30 day trial of Prism. Note the language. The shorthand (used above) is to test the assumption that the residuals are normally distributed. What this really means is testing the assumption that the residuals are sampled from a normal distribution, or are sampled from a population that follows a normal distribution. T tests (paired and unpaired) With t tests and ANOVA models, it appears a little different, but it’s actually the same process of testing the model residuals. With paired t tests, which are used when two measurements are taken on the same data point (for example, before and after measurements for each test subject), the model assumption is that the differences between the two measurements are normally distributed. So in that case, simply test the difference for normality. A common mistake is to test each group as being normally distributed. With unpaired t tests, when comparing if the means between two different independent groups (such as male vs female heights), both columns of data are assumed to be normal, and both should be tested either individually or jointly if you assume equal variance and test the residuals, the difference of each column value minus its respective estimated mean, not the raw data. Are your residuals for t tests clearly deviating a little from normality? Note that t tests are robust to non-normal data with large sample sizes, meaning that as long as you have enough data, only substantial violations of normality need to be addressed. Perform a t test in Prism today. ANOVA with fixed effects In two-way ANOVA with fixed effects, where there are two experimental factors such as fertilizer type and soil type, the assumption is that data within each factor combination are normally distributed. It’s easiest to test this by looking at all of the residuals at once. In this case, the residuals are the difference of each observation from the group mean of its respective factor combination. A common mistake is to test for normality across only one factor. Using the fertilizer and soil type example, the assumption is that each group (fertilizer A with soil type 1, fertilizer A with soil type 2, ...) is normally distributed. It’s not the same thing to test if fertilizer A data are normally distributed, and in fact, if the soil type is a significant factor, then they wouldn’t be. As long as you’re assuming equal variance among the different treatment groups, then you can test for normality across all residuals at once. This is useful in cases when you have only a few observations in any given factorial combination. Test the normality of your data before conducting an ANOVA in Prism. How to test for normality There are both visual and formal statistical tests that can help you check if your model residuals meet the assumption of normality. In Prism, most models (ANOVA, Linear Regression, etc.) include tests and plots for evaluating normality, and you can also test a column of data directly. Visually Q-Q Plot The most common graphical tool for assessing normality is the Q-Q plot. In these plots, the observed data is plotted against the expected quantiles of a normal distribution. It takes practice to read these plots. In theory, sampled data from a normal distribution would fall along the dotted line. In reality, even data sampled from a normal distribution, such as the example QQ plot below, can exhibit some deviation from the line. Frequency distribution You may also visually check normality by plotting a frequency distribution, also called a histogram, of the data and visually comparing it to a normal distribution (overlaid in red). In a frequency distribution, each data point is put into a discrete bin, for example (-10,-5], (-5, 0], (0, 5], etc. The plot shows the proportion of data points in each bin. While this is a useful tool to visually summarize your data, a major drawback is that the bin size can greatly affect how the data look. The following histogram is the same data as above but using smaller bin sizes. With statistical tests There are many statistical tests to evaluate normality, although we don’t recommend relying on them blindly. Prism offers four normality test options: D'Agostino-Pearson, Anderson-Darling, Shapiro-Wilk and Kolmogorov-Smirnov. Each of the tests produces a p-value that tests the null hypothesis that the values (the sample) were sampled from a Normal (Gaussian) distribution (or population). : If the p-value is not significant, the normality test was “passed”. While it’s true we can never say for certain that the data came from a normal distribution, there is not evidence to suggest otherwise. If the p-value is significant, the normality test was “failed”. There is evidence that the data may not be normally distributed after all. If that does not fit with your intuition, remember that the null hypothesis for these tests is that your sample came from a normally distributed population of data. So as with any significant test result, you are rejecting the idea that the data was normally distributed. See our guide for more specific information and background on interpreting normality test p-values. Which is better: visual or statistical tests? We recommend both. It’s always a good idea to plot your data, because, while helpful, statistical tests have limitations. This is especially true with medium to large sample sizes (over 70 observations), because in these cases, the normality tests can detect very slight deviations from normality. Therefore, if your data “fail” a normality test, a visual check might tell you that even if the data are statistically not normal, they are practically normal. Get started in Prism with your free 30 day trial today. What if my residuals aren’t normally distributed? If there is evidence your data are significantly different from the expected normal distribution, what can you do? Some models are robust to deviations from normality Depending on the model you are using, it may still provide accurate results despite some degree of non-normality. 1-Way ANOVA, for example, is often robust even if the data are not very close to normal. Transformations In some situations, you can transform your data and re-test for normality. For example, log transformations are common, because lognormal distributions are common (especially in biology) Non-Parametric Tests If your data truly are not normal, many analyses have non-parametric alternatives, such as the one-way ANOVA analog, Kruskal-Wallis, and the two-sample t test analog, Mann-Whitney. These methods don’t rely on an assumption of normality. The downside is that they generally also have less power, so it’s harder to detect statistical differences. Here are some recommendations to determine when to use nonparametric tests. Keywords: normal distribution This quick tutorial will explain how to test whether sample data is normally distributed in the SPSS statistics package. It is a requirement of many parametric statistical tests – for example, the independent-samples t test – that data is normally distributed. There are a number of different ways to test this requirement. We’re going to focus on the Kolmogorov-Smirnov and Shapiro-Wilk tests. Click Analyze -> Descriptive Statistics -> Explore… Move the variable of interest from the left box into the Dependent List box on the right. Click the Plots button, and tick the Normality plots with tests option. Click Continue, and then click OK. Your result will pop up – check out the Tests of Normality section. The Data Our example data, displayed above in SPSS’s Data View, comes from a pretend study looking at the effect of dog ownership on the ability to throw a frisbee. Frisbee Throwing Distance in Metres (highlighted) is the dependent variable, and we need to know whether it is normally distributed before deciding which statistical test to use to determine if dog ownership is related to the ability to throw a frisbee. Test for Normality To begin, click Analyze -> Descriptive Statistics -> Explore… This will bring up the Explore dialog box, as below. The set up here is quite easy. First, you’ve got to get the Frisbee Throwing Distance variable over from the left box into the Dependent List box. You can either drag and drop, or use the blue arrow in the middle. The Factor List box allows you to split your dependent variable on the basis of the different levels of your independent variable(s). In our example, Dog Owner, our independent variable, has two levels – owner and non-owner – so we could add Dog Owner to the Factor List box, and look at our dependent variable split on that basis. However, since we can perfectly well test for normality without adding in this extra complexity, we’ll just leave the box empty. Once you’ve got the variable you want to test for normality into the Dependent List box, you should click the Plots button. The Plots dialog box will pop up. In this box, you want to make sure that the Normality plots with tests option is ticked, and it’s also sensible to select both descriptive statistics options (Stem-and-leaf and Histogram). Now click Continue, which will take you back to the Explore dialog box. This should now look something like this. You’re now ready to test whether your data is normally distributed. Press the OK button. The Result The Explore option in SPSS produces quite a lot of output. Here’s what you need to assess whether your data distribution is normal. SPSS runs two statistical tests of normality – Kolmogorov-Smirnov and Shapiro-Wilk. If the significance value is greater than the alpha value (we’ll use .05 as our alpha value), then there is no reason to think that our data differs significantly from a normal distribution – i.e., we can reject the null hypothesis that it is non-normal. As you can see above, both tests give a significance value that’s greater than .05, therefore, we can be confident that our data is normally distributed. A complication that can arise here occurs when the results of the two tests don’t agree – that is, when one test shows a significant result and the other doesn’t. In this situation, use the Shapiro-Wilk result – in most circumstances, it is more reliable. Q-Q Plot SPSS also provides a normal Q-Q Plot chart which provides a visual representation of the distribution of the data. If a distribution is normal, then the dots will broadly follow the trend line. As you can see above, our data does cluster around the trend line – which provides further evidence that our distribution is normal. Put this Q-Q plot together with the results of the statistical tests, and we’re safe in assuming that our data is normally distributed. This means that at least one of the criteria for parametric statistical testing is satisfied. If you wish to export the SPSS output for your test of normality to another application such as Word, Excel, or PDF, check out our tutorial. *************** Okay, that’s this tutorial over and done with. You should now be able to interrogate your data in order to determine whether it is normally distributed.