



- 1. [Exploratory Data Analysis](#)
- 1.3. [EDA Techniques](#)
- 1.3.5. [Quantitative Techniques](#)

1.3.5.11. Measures of Skewness and Kurtosis

*Skewness
and
Kurtosis*

A fundamental task in many statistical analyses is to characterize the [location](#) and [variability](#) of a data set. A further characterization of the data includes skewness and kurtosis.

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.

The [histogram](#) is an effective graphical technique for showing both the skewness and kurtosis of data set.

*Definition
of Skewness*

For univariate data Y_1, Y_2, \dots, Y_N , the formula for skewness is:

$$skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3}{(N - 1)s^3}$$

where \bar{Y} is the mean, s is the standard deviation, and N is the number of data points. The skewness for a [normal distribution](#) is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail. Some measurements have a lower bound and are skewed right. For example, in reliability studies, failure times cannot be negative.

For univariate data Y_1, Y_2, \dots, Y_N , the formula for kurtosis is:

Definition of Kurtosis

$$kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)s^4}$$
where \bar{Y} is the mean, s is the standard deviation, and N is the number of data points.

Alternative Definition of Kurtosis

The kurtosis for a [standard normal distribution](#) is three. For this reason, some sources use the following definition of kurtosis (often referred to as "excess kurtosis"):

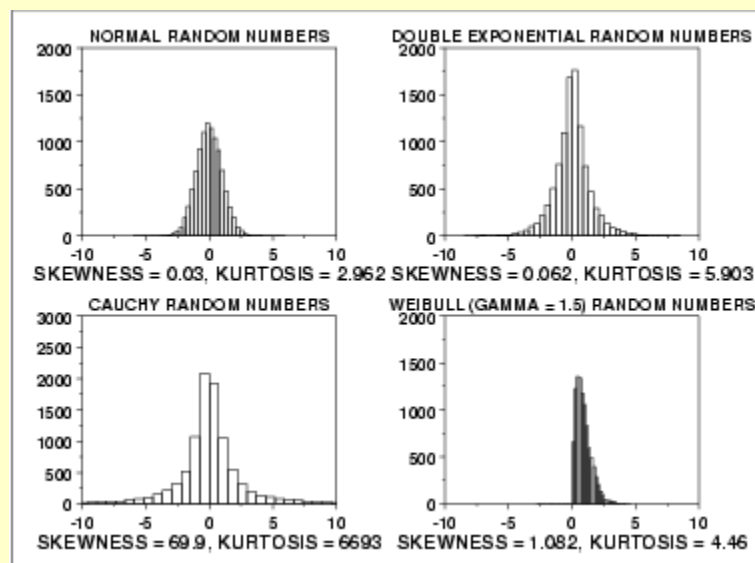
$$kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)s^4} - 3$$

This definition is used so that the standard normal distribution has a kurtosis of zero. In addition, with the second definition positive kurtosis indicates a "peaked" distribution and negative kurtosis indicates a "flat" distribution.

Which definition of kurtosis is used is a matter of convention (this handbook uses the original definition). When using software to compute the sample kurtosis, you need to be aware of which convention is being followed. Many sources use the term kurtosis when they are actually computing "excess kurtosis", so it may not always be clear.

Examples

The following example shows histograms for 10,000 random numbers generated from a normal, a double exponential, a Cauchy, and a Weibull distribution.



<i>Normal Distribution</i>	<p>The first histogram is a sample from a normal distribution.</p> <p>The normal distribution is a symmetric distribution with well-behaved tails. This is indicated by the skewness of 0.03. The kurtosis of 2.96 is near the expected value of 3. The histogram verifies the symmetry.</p>
<i>Double Exponential Distribution</i>	<p>The second histogram is a sample from a double exponential distribution. The double exponential is a symmetric distribution. Compared to the normal, it has a stronger peak, more rapid decay, and heavier tails. That is, we would expect a skewness near zero and a kurtosis higher than 3. The skewness is 0.06 and the kurtosis is 5.9.</p>
<i>Cauchy Distribution</i>	<p>The third histogram is a sample from a Cauchy distribution.</p> <p>For better visual comparison with the other data sets, we restricted the histogram of the Cauchy distribution to values between -10 and 10. The full data set for the Cauchy data in fact has a minimum of approximately -29,000 and a maximum of approximately 89,000.</p> <p>The Cauchy distribution is a symmetric distribution with heavy tails and a single peak at the center of the distribution. Since it is symmetric, we would expect a skewness near zero. Due to the heavier tails, we might expect the kurtosis to be larger than for a normal distribution. In fact the skewness is 69.99 and the kurtosis is 6,693. These extremely high values can be explained by the heavy tails. Just as the mean and standard deviation can be distorted by extreme values in the tails, so too can the skewness and kurtosis measures.</p>
<i>Weibull Distribution</i>	<p>The fourth histogram is a sample from a Weibull distribution with shape parameter 1.5. The Weibull distribution is a skewed distribution with the amount of skewness depending on the value of the shape parameter. The degree of decay as we move away from the center also depends on the value of the shape parameter. For this data set, the skewness is 1.08 and the kurtosis is 4.46, which indicates moderate skewness and kurtosis.</p>
<i>Dealing with Skewness and Kurtosis</i>	<p>Many classical statistical tests and intervals depend on normality assumptions. Significant skewness and kurtosis clearly indicate that data are not normal. If a data set exhibits significant skewness or kurtosis (as indicated by a histogram or the numerical measures), what can we do about it?</p> <p>One approach is to apply some type of transformation to try to make the data normal, or more nearly normal. The Box-Cox</p>

[transformation](#) is a useful technique for trying to normalize a data set. In particular, taking the log or square root of a data set is often useful for data that exhibit moderate right skewness.

Another approach is to use techniques based on distributions other than the normal. For example, in reliability studies, the exponential, Weibull, and lognormal distributions are typically used as a basis for modeling rather than using the normal distribution. The [probability plot correlation coefficient plot](#) and the [probability plot](#) are useful tools for determining a good distributional model for the data.

Software

The skewness and kurtosis coefficients are available in most general purpose statistical software programs.

**NIST
SEMATECH**

HANDBOOK CHAPTERS

- 1. Explore
- 2. Measure
- 3. Characterize
- 4. Model
- 5. Improve
- 6. Monitor
- 7. Compare
- 8. Reliability

HOW TO USE HANDBOOK

TOOLS & AIDS

SEARCH HANDBOOK

DETAILED CONTENTS

ACKNOWLEDGMENTS



**ENGINEERING
STATISTICS**

H A N D B O O K

Welcome! The goal of this handbook is to help scientists and engineers incorporate statistical methods in their work as efficiently as possible.

To reference the Handbook please use a citation of the form:

NIST/SEMATECH e-Handbook of Statistical Methods,
<http://www.itl.nist.gov/div898/handbook/>, date.
(Links to specific pages can also be referenced this way, if suitable.)

A [significant update](#) was made to the Handbook April, 2012

Printer friendly versions of each chapter in the Handbook can be found [here](#).

Feedback on the Handbook sent to handbook@nist.gov is also much appreciated.

[Privacy Policy/Security Notice](#)
[Disclaimer](#) | [FOIA](#)

NIST is an agency of the [U.S. Commerce Department](#).

Date created: 6/01/2003
Last updated: 10/30/2013