**Skewness, Kurtosis, and the Normal Curve[[1]](#footnote-1)©**

**Skewness**

In everyday language, the terms “skewed” and “askew” are used to refer to something that is out of line or distorted on one side. When referring to the shape of frequency or probability distributions, “skewness” refers to asymmetry of the distribution. A distribution with an asymmetric tail extending out to the right is referred to as “positively skewed” or “skewed to the right,” while a distribution with an asymmetric tail extending out to the left is referred to as “negatively skewed” or “skewed to the left.” Skewness can range from minus infinity to positive infinity.

Karl Pearson (1895) first suggested measuring skewness by standardizing the difference between the mean and the mode, that is, . Population modes are not well estimated from sample modes, but one can estimate the difference between the mean and the mode as being three times the difference between the mean and the median (Stuart & Ord, 1994), leading to the following estimate of skewness: . Many statisticians use this measure but with the ‘3’ eliminated, that is, . This statistic ranges from -1 to +1. Absolute values above 0.2 indicate great skewness (Hildebrand, 1986).

Skewness has also been defined with respect to the third moment about the mean: , which is simply the expected value of the distribution of cubed *z* scores. Skewness measured in this way is sometimes referred to as “Fisher’s skewness.” When the deviations from the mean are greater in one direction than in the other direction, this statistic will deviate from zero in the direction of the larger deviations. From sample data, Fisher’s skewness is most often estimated by: . For large sample sizes (*n* > 150), *g1* may be distributed approximately normally, with a standard error of approximately . While one could use this sampling distribution to construct confidence intervals for or tests of hypotheses about γ1, there is rarely any value in doing so.

The most commonly used measures of skewness (those discussed here) may produce surprising results, such as a negative value when the shape of the distribution appears skewed to the right. There may be superior alternative measures not commonly used (Groeneveld & Meeden, 1984).

It is important for behavioral researchers to notice skewness when it appears in their data. Great skewness may motivate the researcher to investigate outliers. When making decisions about which measure of location to report (means being drawn in the direction of the skew) and which inferential statistic to employ (one which assumes normality or one which does not), one should take into consideration the estimated skewness of the population. Normal distributions have zero skewness. Of course, a distribution can be perfectly symmetric but far from normal. Transformations commonly employed to reduce (positive) skewness include square root, log, and reciprocal transformations.

Also see [Skewness and the Relative Positions of Mean, Median, and Mode](http://core.ecu.edu/psyc/wuenschk/StatHelp/Skew.htm)

**Kurtosis**

Karl Pearson (1905) defined a distribution’s degree of kurtosis as , where , the expected value of the distribution of *Z* scores which have been raised to the 4th power. *β2* is often referred to as “Pearson’s kurtosis,” and *β2* ‑ 3 (often symbolized with *γ2* ) as “kurtosis excess” or “Fisher’s kurtosis,” even though it was Pearson who defined kurtosis as *β2* ‑ 3. An unbiased estimator for *γ2* is . For large sample sizes (*n* > 1000), *g2* may be distributed approximately normally, with a standard error of approximately  (Snedecor, & Cochran, 1967). While one could use this sampling distribution to construct confidence intervals for or tests of hypotheses about γ2, there is rarely any value in doing so.

Pearson (1905) introduced kurtosis as a measure of how flat the top of a symmetric distribution is when compared to a normal distribution of the same variance. He referred to more flat-topped distributions (*γ2* < 0) as “platykurtic,” less flat-topped distributions (*γ2* > 0) as “leptokurtic,” and equally flat-topped distributions as “mesokurtic” (*γ2* ≈ 0). Kurtosis is actually more influenced by scores in the tails of the distribution than scores in the center of a distribution (DeCarlo, 1967). Accordingly, it is often appropriate to describe a leptokurtic distribution as “fat in the tails” and a platykurtic distribution as “thin in the tails.”

Student (1927, *Biometrika*, *19*, 160) published a cute description of kurtosis, which I quote here: “Platykurtic curves have shorter ‘tails’ than the normal curve of error and leptokurtic longer ‘tails.’ I myself bear in mind the meaning of the words by the above *memoria technica*, where the first figure represents platypus and the second kangaroos, noted for lepping.” [See Student’s drawings](http://core.ecu.edu/psyc/wuenschk/stathelp/kurtosis-Student1927.gif).

Moors (1986) demonstrated that. Accordingly, it may be best to treat kurtosis as the extent to which scores are dispersed away from the shoulders of a distribution, where the shoulders are the points where *Z2* = 1, that is, *Z* = ±1. Balanda and MacGillivray (1988) wrote “it is best to define kurtosis vaguely as the location- and scale-free movement of probability mass from the shoulders of a distribution into its centre and tails.” If one starts with a normal distribution and moves scores from the shoulders into the center and the tails, keeping variance constant, kurtosis is increased. The distribution will likely appear more peaked in the center and fatter in the tails, like a [Laplace distribution](http://mathworld.wolfram.com/LaplaceDistribution.html) () or [Student’s *t*](http://mathworld.wolfram.com/Studentst-Distribution.html) with few degrees of freedom ().

Starting again with a normal distribution, moving scores from the tails and the center to the shoulders will decrease kurtosis. A [uniform distribution](http://mathworld.wolfram.com/UniformDistribution.html) certainly has a flat top, with , but *γ2* can reach a minimum value of −2 when two score values are equally probably and all other score values have probability zero (a [rectangular U distribution](http://core.ecu.edu/psyc/wuenschk/Gifs/Bin_n1_p.5.gif), that is, a binomial distribution with *n* =1, *p* = .5). One might object that the rectangular U distribution has all of its scores in the tails, but closer inspection will reveal that it has no tails, and that all of its scores are in its shoulders, exactly one standard deviation from its mean. Values of *g2* less than that expected for an uniform distribution (−1.2) may suggest that the distribution is bimodal (Darlington, 1970), but bimodal distributions can have high kurtosis if the modes are distant from the shoulders.

One leptokurtic distribution we shall deal with is Student’s ***t*** distribution. The kurtosis of *t* is infinite when *df* < 5, 6 when *df* = 5, 3 when *df* = 6. Kurtosis decreases further (towards zero) as *df* increase and *t* approaches the normal distribution.

Kurtosis is usually of interest only when dealing with approximately symmetric distributions. Skewed distributions are always leptokurtic (Hopkins & Weeks, 1990). Among the several alternative measures of kurtosis that have been proposed (none of which has often been employed), is one which adjusts the measurement of kurtosis to remove the effect of skewness (Blest, 2003).

There is much confusion about how kurtosis is related to the shape of distributions. Many authors of textbooks have asserted that kurtosis is a measure of the peakedness of distributions, which is not strictly true.

It is easy to confuse low kurtosis with high variance, but distributions with identical kurtosis can differ in variance, and distributions with identical variances can differ in kurtosis. Here are some simple distributions that may help you appreciate that kurtosis is, in part, a measure of tail heaviness relative to the total variance in the distribution (remember the “*σ4*” in the denominator).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1.**  **Kurtosis for 7 Simple Distributions Also Differing in Variance** | | | | | | | |
| X | freq A | freq B | freq C | freq D | freq E | freq F | freq G |
| 05 | 20 | 20 | 20 | 10 | 05 | 03 | 01 |
| 10 | 00 | 10 | 20 | 20 | 20 | 20 | 20 |
| 15 | 20 | 20 | 20 | 10 | 05 | 03 | 01 |
| Kurtosis | -2.0 | -1.75 | -1.5 | -1.0 | 0.0 | 1.33 | 8.0 |
| Variance | 25 | 20 | 16.6 | 12.5 | 8.3 | 5.77 | 2.27 |

Platykurtic Leptokurtic

When I presented these distributions to my colleagues and graduate students and asked them to identify which had the least kurtosis and which the most, all said A has the most kurtosis, G the least (excepting those who refused to answer). But in fact A has the least kurtosis (−2 is the smallest possible value of kurtosis) and G the most. The trick is to do a mental frequency plot where the abscissa is in standard deviation units. In the maximally platykurtic distribution A, which initially appears to have all its scores in its tails, no score is more than one *σ* away from the mean - that is, it has no tails! In the leptokurtic distribution G, which seems only to have a few scores in its tails, one must remember that those scores (5 & 15) are much farther away from the mean (3.3 *σ* ) than are the 5’s & 15’s in distribution A. In fact, in G nine percent of the scores are more than three *σ* from the mean, much more than you would expect in a mesokurtic distribution (like a normal distribution), thus G does indeed have fat tails.

If you were you to ask SAS to compute kurtosis on the A scores in Table 1, you would get a value less than −2.0, less than the lowest possible population kurtosis. Why? SAS assumes your data are a sample and computes the *g2* estimate of population kurtosis, which can fall below −2.0.

Sune Karlsson, of the Stockholm School of Economics, has provided me with the following modified example which holds the variance approximately constant, making it quite clear that a higher kurtosis implies that there are more extreme observations (or that the extreme observations are more extreme). It is also evident that a higher kurtosis also implies that the distribution is more ‘single-peaked’ (this would be even more evident if the sum of the frequencies was constant). I have highlighted the rows representing the shoulders of the distribution so that you can see that the increase in kurtosis is associated with a movement of scores away from the shoulders.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.**  **Kurtosis for Seven Simple Distributions Not Differing in Variance** | | | | | | | |
| X | Freq. A | Freq. B | Freq. C | Freq. D | Freq. E | Freq. F | Freq. G |
| −6.6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| −0.4 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| 1.3 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| 2.9 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| 3.9 | 0 | 0 | 20 | 0 | 0 | 0 | 0 |
| 4.4 | 0 | 20 | 0 | 0 | 0 | 0 | 0 |
| **5** | 20 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 10 | 20 | 20 | 20 | 20 | 20 |
| **15** | 20 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15.6 | 0 | 20 | 0 | 0 | 0 | 0 | 0 |
| 16.1 | 0 | 0 | 20 | 0 | 0 | 0 | 0 |
| 17.1 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| 18.7 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| 20.4 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| 26.6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Kurtosis | −2.0 | −1.75 | −1.5 | −1.0 | 0.0 | 1.33 | 8.0 |
| Variance | 25 | 25.1 | 24.8 | 25.2 | 25.2 | 25.0 | 25.1 |

While is unlikely that a behavioral researcher will be interested in questions that focus on the kurtosis of a distribution, estimates of kurtosis, in combination with other information about the shape of a distribution, can be useful. DeCarlo (1997) described several uses for the *g2* statistic. When considering the shape of a distribution of scores, it is useful to have at hand measures of skewness and kurtosis, as well as graphical displays. These statistics can help one decide which estimators or tests should perform best with data distributed like those on hand. High kurtosis should alert the researcher to investigate outliers in one or both tails of the distribution.

**Tests of Significance**

Some statistical packages (including SPSS) provide both estimates of skewness and kurtosis and standard errors for those estimates. One can divide the estimate by it’s standard error to obtain a z test of the null hypothesis that the parameter is zero (as would be expected in a normal population), but I generally find such tests of little use. One may do an “eyeball test” on measures of skewness and kurtosis when deciding whether or not a sample is “normal enough” to use an inferential procedure that assumes normality of the population(s). If you wish to test the null hypothesis that the sample came from a normal population, you can use a chi-square goodness of fit test, comparing observed frequencies in ten or so intervals (from lowest to highest score) with the frequencies that would be expected in those intervals were the population normal. This test has very low power, especially with small sample sizes, where the normality assumption may be most critical. Thus you may think your data close enough to normal (not significantly different from normal) to use a test statistic which assumes normality when in fact the data are too distinctly non-normal to employ such a test, the nonsignificance of the deviation from normality resulting only from low power, small sample sizes. SAS’ PROC UNIVARIATE will test such a null hypothesis for you using the more powerful [Kolmogorov](http://www.nytimes.com/1987/10/23/obituaries/an-kolmogorov-dies-at-84-top-russian-mathematician.html)-Smirnov statistic (for larger samples) or the Shapiro-Wilks statistic (for smaller samples). These have very high power, especially with large sample sizes, in which case the normality assumption may be less critical for the test statistic whose normality assumption is being questioned. These tests may tell you that your sample differs significantly from normal even when the deviation from normality is not large enough to cause problems with the test statistic which assumes normality. Furthermore, with small sample sizes these tests have little power but the normality-assuming procedures are less robust to the assumption. In this case the test of normality may have so little power that it fails to detect significant non-normality even when the population is distinctly non-normal.

I recommend using measure of skewness and kurtosis and plots to evaluate whether or not there is problem with non-normality. For most procedures I am comfortable when both *g1* (skewness estimate) and *g2* (kurtosis estimate) have absolute values less than one. If I am uncomfortable with the *g1* statistics I will try order-preserving nonlinear transformations (such as square root and log) to reduce the skewness to an acceptable level.

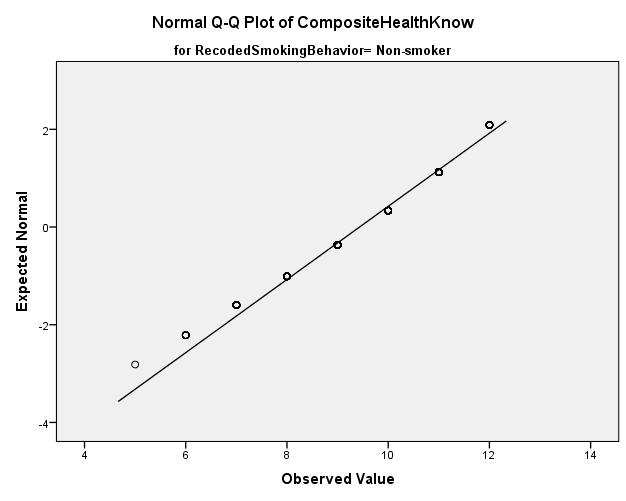
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | | |
|  | RecodedSmokingBehavior | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
|  | Statistic | df | Sig. | Statistic | df | Sig. |
| CompositeHealthKnow | Non-smoker | .174 | 408 | .000 | .938 | 408 | .000 |
| Smoker | .180 | 70 | .000 | .932 | 70 | .001 |
| a. Lilliefors Significance Correction | | | | | | | |

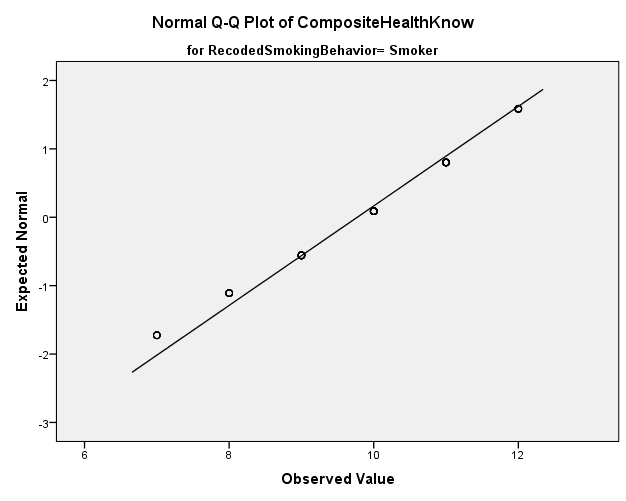
The table above shows that the thesis data are significantly non-normal, but the sample sizes are large. Do the data really differ from normal enough to be worried? Look at the estimates of skewness below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptives** | | | | |
|  | RecodedSmokingBehavior | | Statistic | Std. Error |
| CompositeHealthKnow | Non-smoker | Mean | 9.4338 | .06613 |
| Std. Deviation | 1.33578 |  |
| Skewness | -.420 | .121 |
| Kurtosis | -.090 | .241 |
| Smoker | Mean | 9.7714 | .16427 |
| Std. Deviation | 1.37442 |  |
| Skewness | -.298 | .287 |
| Kurtosis | -.541 | .566 |

IMHO, the values of *g1* and *g2* here are not sufficiently far from 0 to generate any concern.

Graphical procedures can also be helpful in determining whether or not your data seriously violate a normality assumption. I like histograms with a normal curve and a [kernel plot](http://core.ecu.edu/psyc/wuenschk/docs30/Kernal%20Density%20Plots.pdf) superimposed. Stat packages can also produce a normal probability plot. In such a plot if the markers are close to the line, the data are close to normal. Look at the normality plots for these thesis data:





IMHO, no problems here.

**Moments about the Mean**

For a set of *N* scores, the *kth* moment about the mean is . As you already know, the first moment about the mean (the sum of the deviations of scores from their mean) is zero. The second moment about the mean is variance. The third is skewness. The fourth is *β2* kurtosis.

**SAS Exercises**

Go to my StatData page and download the file EDA.dat. Go to my [SAS-Programs page](http://core.ecu.edu/psyc/wuenschk/SAS/SAS-Programs.htm) and download the program file **g1g2.sas**. Edit the program so that the INFILE statement points correctly to the folder where you located EDA.dat and then run the program, which illustrates the computation of *g1* and *g2*. Look at the program. The raw data are read from EDA.dat and PROC MEANS is then used to compute *g1* and *g2*. The next portion of the program uses PROC STANDARD to convert the data to *z* scores. PROC MEANS is then used to compute *g1* and *g2* on the *z* scores. Note that standardization of the scores has not changed the values of *g1* and *g2*. The next portion of the program creates a data set with the *z* scores raised to the 3rd and the 4th powers. The final step of the program uses these powers of *z* to compute *g1* and *g2* using the formulas presented earlier in this handout. Note that the values of *g1* and *g2* are the same as obtained earlier from PROC MEANS.

Go to my SAS-Programs page and download and run the file **Kurtosis-Uniform.sas**. Look at the program. A DO loop and the UNIFORM function are used to create a sample of 500,000 scores drawn from a uniform population which ranges from 0 to 1. PROC MEANS then computes mean, standard deviation, skewness, and kurtosis. Look at the output. Compare the obtained statistics to the expected values for the following parameters of a uniform distribution that ranges from *a* to *b*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Expected Value |  | Parameter | Expected Value |
| Mean |  |  | Skewness | 0 |
| Standard Deviation |  |  | Kurtosis | −1.2 |

Go to my SAS-Programs page and download and run the file “**Kurtosis‑T.sas**,” which demonstrates the effect of sample size (degrees of freedom) on the kurtosis of the *t* distribution. Look at the program. Within each section of the program a DO loop is used to create 500,000 samples of *N* scores (where *N* is 10, 11, 17, or 29), each drawn from a normal population with mean 0 and standard deviation 1. PROC MEANS is then used to compute Student’s *t* for each sample, outputting these *t* scores into a new data set. We shall treat this new data set as the sampling distribution of *t*. PROC MEANS is then used to compute the mean, standard deviation, and kurtosis of the sampling distributions of *t*. For each value of degrees of freedom, compare the obtained statistics with their expected values.

|  |  |  |
| --- | --- | --- |
| Mean | Standard Deviation | Kurtosis |
| 0 |  |  |

Download and run my program **Kurtosis\_Beta2.sas**. Look at the program. Each section of the program creates one of the distributions from Table 1 above and then converts the data to *z* scores, raises the *z* scores to the fourth power, and computes *β2* as the mean of *z4*. Subtract 3 from each value of *β2*  and then compare the resulting *γ2*  to the value given in Table 1.

Download and run my program **Kurtosis-Normal.sas**. Look at the program. DO loops and the NORMAL function are used to create 100,000 samples, each with 1,000 scores drawn from a normal population with mean 0 and standard deviation 1. PROC MEANS creates a new data set with the *g1* and the *g2* statistics for each sample. PROC MEANS then computes the mean and standard deviation (standard error) for skewness and kurtosis. Compare the values obtained with those expected, 0 for the means, and  and  for the standard errors.

**References**

Balanda & MacGillivray. (1988). Kurtosis: A critical review. *American Statistician*, *42*: 111-119.

Blest, D.C. (2003). A new measure of kurtosis adjusted for skewness. *Australian &New Zealand Journal of Statistics*, *45*, 175-179.

Darlington, R.B. (1970). Is kurtosis really “peakedness?” *The American Statistician*, *24*(2), 19-22.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292-307.

Groeneveld, R.A. & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, *33*, 391-399.

Hildebrand, D. K. (1986). *Statistical thinking for behavioral scientists*. Boston: Duxbury.

Hopkins, K.D. & Weeks, D.L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*, 717-729.

Loether, H. L., & McTavish, D. G. (1988). *Descriptive and inferential statistics: An introduction* , 3rd ed. Boston: Allyn & Bacon.

Moors, J.J.A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician*, *40*, 283-284.

Pearson, K. (1895) Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London*, *186*, 343-414.

Pearson, K. (1905). Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder. *Biometrika*, *4*, 169-212.

Snedecor, G.W. & Cochran, W.G. (1967). Statistical methods (6th ed.), Iowa State University Press, Ames, Iowa.

Stuart, A. & Ord, J.K. (1994). *Kendall’s advanced theory of statistics*. Volume 1. Distribution Theory. Sixth Edition. Edward Arnold, London.

Wuensch, K. L. (2005).  Kurtosis.  In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1028 - 1029).  Chichester, UK:  Wiley.

Wuensch, K. L. (2005).  Skewness.  In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1855 - 1856).  Chichester, UK:  Wiley.

Links

* <http://core.ecu.edu/psyc/wuenschk/StatHelp/KURTOSIS.txt> -- a log of email discussions on the topic of kurtosis, most of them from the EDSTAT list.
* <http://core.ecu.edu/psyc/WuenschK/docs30/Platykurtosis.jpg> -- distribution of final grades in PSYC 2101 (undergrad stats), Spring, 2007.
* [Skewness & Kurtosis](http://core.ecu.edu/psyc/wuenschk/PP/Kurtosis.pptx) – slide show with histograms of the distributions presented in Table 2 above
* [Table 2](http://core.ecu.edu/psyc/wuenschk/SPSS/Kurtosis.sav) – data from Table 2, SPSS format

Copyright 2016, Karl L. Wuensch - All rights reserved.

[Return to My Statistics Lessons Page](http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm)

1. © Copyright 2016, Karl L. Wuensch - All rights reserved. [↑](#footnote-ref-1)