Measures of Shape: Skewness and Kurtosis

Summary: You've learned numerical measures of center, spread, and outliers, but what about **measures of shape**? The histogram can give you a general idea of the shape, but two numerical measures of shape give a more precise evaluation: **skewness tells you the amount and direction of skew** (departure from horizontal symmetry), and **kurtosis tells you how tall and sharp the central peak is**, relative to a standard bell curve.

Why do we care? One application is **testing for normality**: many statistics inferences require that a distribution be normal or nearly normal. A normal distribution has skewness and excess kurtosis of 0, so if your distribution is close to those values then it is probably close to normal.

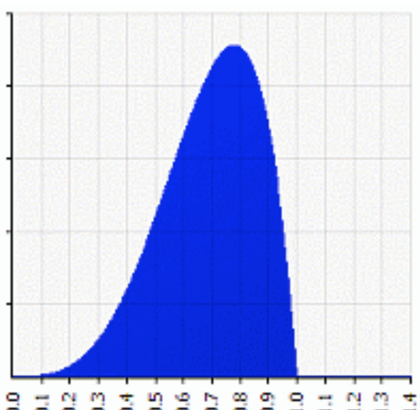
See also: MATH200B Program — Extra Statistics Utilities for TI-83/84 has a program to download to your TI-83 or TI-84. Among other things, the program computes all the skewness and kurtosis measures in this document, except confidence interval of skewness and the D'Agostino-Pearson test.

Contents:

Skewness

The first thing you usually notice about a distribution's shape is whether it has one mode (peak) or more than one. If it's **unimodal** (has just one peak), like most data sets, the next thing you notice is whether it's **symmetric or skewed** to one side. If the bulk of the data is at the left and the right tail is longer, we say that the distribution is **skewed right or positively skewed**; if the peak is toward the right and the left tail is longer, we say that the distribution is **skewed left or negatively skewed**.

Look at the two graphs below. They both have μ = 0.6923 and σ = 0.1685, but their shapes are different.



Beta(α=4.5, β=2)
skewness = −0.5370

1.3846 − Beta(α=4.5, β=2)
skewness = +0.5370

The first one is moderately skewed left: the left tail is longer and most of the distribution is at the right. By contrast, the second distribution is moderately skewed right: its right tail is longer and most of the distribution is at the left.

You can get a general impression of skewness by drawing a histogram

(MATH200A part 1), but there are also some common numerical measures of skewness. Some authors favor one, some favor another. This Web page presents one of them. In fact, these are the same formulas that Excel uses in its "Descriptive Statistics" tool in Analysis Toolpak.

You may remember that the mean and standard deviation have the same units as the original data, and the variance has the square of those units. However, the skewness has no units: it's a pure number, like a z-score.

Computing

The **moment coefficient of skewness** of a data set is

$$\text{skewness: } g_1 = m_3 / m_2^{3/2}$$

where (1)

$$m_3 = \sum(x-\bar{x})^3 / n \quad \text{and} \quad m_2 = \sum(x-\bar{x})^2 / n$$

$\bar{x}$ is the mean and n is the sample size, as usual. m_3 is called the **third moment** of the data set. m_2 is the **variance**, the square of the standard deviation.

You'll remember that you have to choose one of two different measures of standard deviation, depending on whether you have data for the whole population or just a sample. The same is true of skewness. If you have the whole population, then g_1 above is the measure of skewness. But **if you have just a sample**, you need the **sample skewness**:

$$\text{sample skewness: } G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 \qquad (2)$$

source: D. N. Joanes and C. A. Gill. "Comparing Measures of Sample Skewness and Kurtosis". *The Statistician* 47(1):183–189.

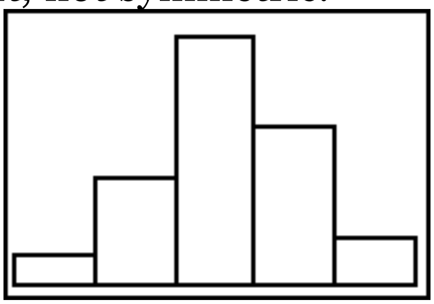
Excel doesn't concern itself with whether you have a sample or a population: its measure of skewness is always G_1.

Example 1: College Men's Heights

Here are grouped data for heights of 100 randomly selected male students, adapted from Spiegel & Stephens, *Theory and Problems of Statistics* 3/e (McGraw-Hill, 1999), page 68.

Height (inches)	Class Mark, x	Frequency, f
59.5–62.5	61	5
62.5–65.5	64	18
65.5–68.5	67	42
68.5–71.5	70	27
71.5–74.5	73	8

A histogram shows that the data are skewed left, not symmetric.



But **how highly skewed** are they, compared to other data sets? To answer this question, you have to compute the skewness.

Begin with the sample size and sample mean. (The sample size was given, but it never hurts to check.)

$$n = 5+18+42+27+8 = 100$$
$$\bar{x} = (61\times5 + 64\times18 + 67\times42 + 70\times27 + 73\times8) \div 100$$
$$\bar{x} = 9305 + 1152 + 2814 + 1890 + 584) \div 100$$
$$\bar{x} = 6745\div100 = 67.45$$

Now, with the mean in hand, you can compute the skewness. (Of course in real life you'd probably use Excel or a statistics package, but it's good to know where the numbers come from.)

Class Mark, x	Frequency, f	xf	$(x-\bar{x})$	$(x-\bar{x})^2f$	$(x-\bar{x})^3f$
61	5	305	-6.45	208.01	-1341.68
64	18	1152	-3.45	214.25	-739.15
67	42	2814	-0.45	8.51	-3.83
70	27	1890	2.55	175.57	447.70
73	8	584	5.55	246.42	1367.63

Σ	6745	n/a	852.75	−269.33
$\bar{x}$, m_2, m_3	67.45	n/a	8.5275	−2.6933

Finally, the skewness is

$$g_1 = m_3 / m_2^{3/2} = -2.6933 / 8.5275^{3/2} = -0.1082$$

But wait, there's more! That would be the skewness if the you had data for the whole population. But obviously there are more than 100 male students in the world, or even in almost any school, so what you have here is a sample, not the population. You must compute the **sample skewness**:

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 = [\sqrt{(100 \times 99)} / 98] [-2.6933 / 8.5275^{3/2}] = -0.1098$$

Interpreting

If skewness is positive, the data are positively skewed or skewed right, meaning that the right tail of the distribution is longer than the left. If skewness is negative, the data are negatively skewed or skewed left, meaning that the left tail is longer.

If skewness = 0, the data are perfectly symmetrical. But a skewness of exactly zero is quite unlikely for real-world data, so **how can you interpret the skewness number**? Bulmer, M. G., *Principles of Statistics* (Dover, 1979) — a classic — suggests this rule of thumb:

- If skewness is less than −1 or greater than +1, the distribution is **highly skewed**.
- If skewness is between −1 and −½ or between +½ and +1, the distribution is **moderately skewed**.
- If skewness is between −½ and +½, the distribution is **approximately symmetric**.

With a skewness of −0.1098, the sample data for student heights are approximately symmetric.

Caution: This is an interpretation of the data you actually have. When you have data for the whole population, that's fine. But when you have a sample, the sample skewness doesn't necessarily apply to the whole population. In that case the question is, from the sample skewness, can you conclude anything about the population skewness? To answer that question, see the next section.

Inferring

Your data set is just one sample drawn from a population. Maybe, from ordinary sample variability, your sample is skewed even though the population is symmetric. But if the sample is skewed too much for random chance to be the explanation, then you can conclude that there is skewness in the population.

But what do I mean by "too much for random chance to be the explanation"? To answer that, you need to divide the sample skewness G_1 by the **standard error of skewness (SES)** to get the **test statistic**, which measures how many standard errors separate the sample skewness from zero:

$$\text{test statistic: } Z_{g1} = G_1/SES \quad \text{where} \quad SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{3}$$

This formula is adapted from page 85 of Cramer, Duncan, *Basic Statistics for Social Research* (Routledge, 1997). (Some authors suggest $\sqrt{(6/n)}$, but for small samples that's a poor approximation. And anyway, we've all got calculators, so you may as well do it right.)

The critical value of Z_{g1} is approximately 2. (This is a two-tailed test of skewness ≠ 0 at roughly the 0.05 significance level.)

- **If $Z_{g1} < -2$**, the population is very likely skewed negatively (though you don't know by how much).
- **If Z_{g1} is between −2 and +2**, you can't reach any conclusion about the skewness of the population: it might be symmetric, or it might be skewed in either direction.
- **If $Z_{g1} > 2$**, the population is very likely skewed positively (though you don't know by how much).

Don't mix up the meanings of this test statistic and the amount of skewness. The amount of skewness tells you how highly skewed your sample is: the bigger the number, the bigger the skew. The test statistic tells you whether the whole population is probably skewed, but not by how much: the bigger the number, the higher the probability.

Estimating

GraphPad suggests a **confidence interval for skewness**:

$$95\% \text{ confidence interval of population skewness} = G_1 \pm 2 \text{ SES} \qquad (4)$$

I'm not so sure about that. Joanes and Gill point out that sample skewness is an unbiased estimator of population skewness for normal distributions, but not others. So I would say, compute that confidence interval, but take it with several grains of salt — and the further the sample skewness is from zero, the more skeptical you should be.

For the college men's heights, recall that the sample skewness was $G_1 = -0.1098$. The sample size was n = 100 and therefore the standard error of skewness is

$$\text{SES} = \sqrt{[\, (600\times99) / (98\times101\times103) \,]} = 0.2414$$

The test statistic is

$$Z_{g1} = G_1/\text{SES} = -0.1098 / 0.2414 = -0.45$$

This is quite small, so **it's impossible to say whether the population is symmetric or skewed.** Since the sample skewness is small, a confidence interval is probably reasonable:

$$G_1 \pm 2 \text{ SES} = -.1098 \pm 2\times.2414 = -.1098\pm.4828 = -0.5926 \text{ to } +0.3730.$$

You can give a 95% confidence interval of skewness as about −0.59 to +0.37, more or less.

Kurtosis

If a distribution is symmetric, the next question is about the central peak: is it high and sharp, or short and broad? You can get some idea of this from the histogram, but a numerical measure is more precise.

The **height and sharpness of the peak** relative to the rest of the data are measured by a number called kurtosis. **Higher values indicate a higher, sharper peak; lower values indicate a lower, less distinct peak.** This occurs because, as Wikipedia's article on kurtosis explains, higher kurtosis means more of the variability is due to a few extreme differences from the mean, rather than a lot of modest differences from the mean.

Balanda and MacGillivray say the same thing in another way: **increasing kurtosis is associated with the "movement of probability mass from the shoulders of a distribution into its center and tails**." (Kevin P. Balanda and H.L. MacGillivray. "Kurtosis: A Critical Review". *The American Statistician* 42:2 [May 1988], pp 111–119, drawn to my attention by Karl Ove Hufthammer)

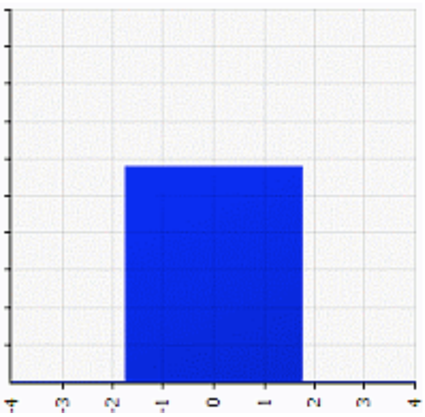
You may remember that the mean and standard deviation have the same units as the original data, and the variance has the square of those units. However, the kurtosis has no units: it's a pure number, like a z-score.

The reference standard is a normal distribution, which has a kurtosis of 3. In token of this, often the **excess kurtosis** is presented: excess kurtosis is simply **kurtosis−3**. For example, the "kurtosis" reported by Excel is actually the excess kurtosis.

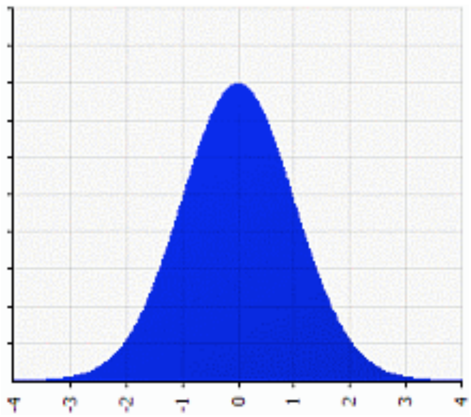
- A normal distribution has kurtosis exactly 3 (excess kurtosis exactly 0). Any distribution with kurtosis ≈3 (excess ≈0) is called **mesokurtic**.
- A distribution with kurtosis <3 (excess kurtosis <0) is called **platykurtic**. Compared to a normal distribution, its central peak is lower and broader, and its tails are shorter and thinner.
- A distribution with kurtosis >3 (excess kurtosis >0) is called **leptokurtic**. Compared to a normal distribution, its central peak is higher and sharper, and its tails are longer and fatter.

Visualizing

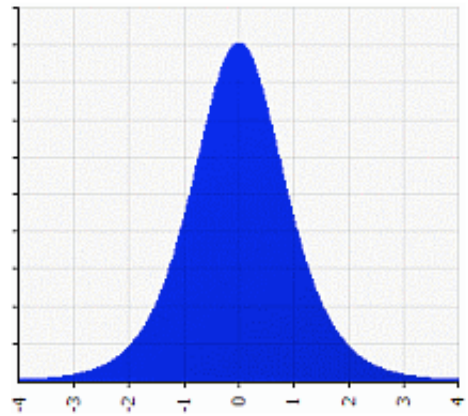
Kurtosis is unfortunately harder to picture than skewness, but these illustrations, suggested by Wikipedia, should help. All three of these distributions have mean of 0, standard deviation of 1, and skewness of 0, and all are plotted on the same horizontal and vertical scale. Look at the progression from left to right, as kurtosis increases.



Uniform(min=–√3, max=√3)
kurtosis = 1.8, excess = –1.2



Normal(μ=0, σ=1)
kurtosis = 3, excess = 0

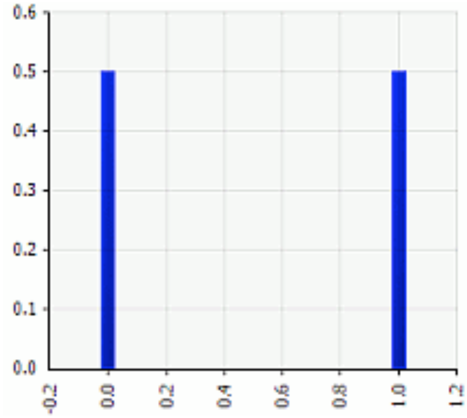


Logistic(α=0, β=0.55153)
kurtosis = 4.2, excess = 1.2

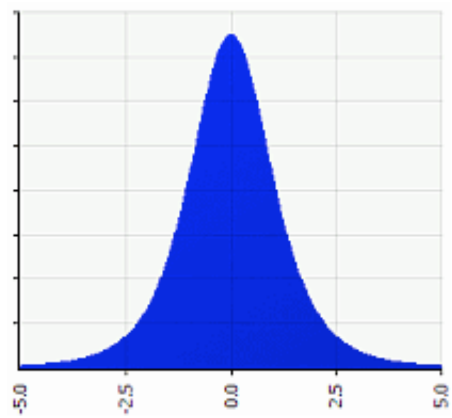
Moving from the illustrated uniform distribution to a normal distribution, you see that the “shoulders” have transferred some of their mass to the center and the tails. In other words, the intermediate values have become less likely and the central and extreme values have become more likely. The kurtosis increases while the standard deviation stays the same, because more of the variation is due to extreme values.

Moving from the normal distribution to the illustrated logistic distribution, the trend continues. There is even less in the shoulders and even more in the tails, and the central peak is higher and narrower.

How far can this go? What are the **smallest and largest possible values of kurtosis**? The smallest possible kurtosis is 1 (excess kurtosis –2), and the largest is ∞, as shown here:



Discrete: equally likely values
kurtosis = 1, excess = −2



Student's t (df=4)
kurtosis = ∞, excess = ∞

A discrete distribution with two equally likely outcomes, such as winning or losing on the flip of a coin, has the **lowest possible kurtosis.** It has no central peak and no real tails, and you could say that it's "all shoulder" — it's as platykurtic as a distribution can be. At the other extreme, Student's t distribution with four degrees of freedom has **infinite kurtosis**. A distribution can't be any more leptokurtic than this.

Computing

The **moment coefficient of kurtosis** of a data set is computed almost the same way as the coefficient of skewness: just change the exponent 3 to 4 in the formulas:

$$\text{kurtosis: } a_4 = m_4 / m_2^2 \quad \text{and} \quad \text{excess kurtosis: } g_2 = a_4 - 3$$

where (5)

$$m_4 = \sum(x-\bar{x})^4 / n \quad \text{and} \quad m_2 = \sum(x-\bar{x})^2 / n$$

Again, the excess kurtosis is generally used because the excess kurtosis of a normal distribution is 0. $\bar{x}$ is the mean and n is the sample size, as usual. m_4 is called the **fourth moment** of the data set. m_2 is the **variance**, the square of the standard deviation.

Just as with variance, standard deviation, and kurtosis, the above is the final computation if you have data for the whole population. But **if you have data for only a sample**, you have to compute the sample excess kurtosis using this formula, which comes from Joanes and Gill:

sample excess kurtosis: $G_2 = \frac{n-1}{(n-2)(n-3)}[(n+1)g_2+6]$ (6)

Excel doesn't concern itself with whether you have a sample or a population: its measure of kurtosis is always G_2.

Example: Let's continue with the example of the college men's heights, and compute the kurtosis of the data set. n = 100, $\bar{x}$ = 67.45 inches, and the variance m_2 = 8.5275 in² were computed earlier.

Class Mark, x	Frequency, f	$x-\bar{x}$	$(x-\bar{x})^4 f$
61	5	-6.45	8653.84
64	18	-3.45	2550.05
67	42	-0.45	1.72
70	27	2.55	1141.63
73	8	5.55	7590.35
	∑	n/a	19937.60
	m_4	n/a	199.3760

Finally, the kurtosis is

$a_4 = m_4 / m_2^2 = 199.3760/8.5275^2 = 2.7418$

and the excess kurtosis is

$g_2 = 2.7418-3 = -0.2582$

But this is a sample, not the population, so you have to compute the sample excess kurtosis:

$G_2 = [99/(98\times97)]\ [101\times(-0.2582)+6)] = -0.2091$

This sample is **slightly platykurtic**: its peak is just a bit shallower than the peak of a normal distribution.

Inferring

Your data set is just one sample drawn from a population. How far must the excess kurtosis be from 0, before you can say that the **population** also has nonzero excess kurtosis?

The answer comes in a similar way to the similar question about skewness. You divide the sample excess kurtosis by the **standard error of kurtosis (SEK)** to get the **test statistic**, which tells you how many standard errors the sample excess kurtosis is from zero:

test statistic: $Z_{g2} = G_2 / SEK$ where $SEK = 2(SES)\sqrt{\frac{n^2-1}{(n-3)(n+5)}}$ (7)

The formula is adapted from page 89 of Duncan Cramer's *Basic Statistics for Social Research* (Routledge, 1997). (Some authors suggest $\sqrt{(24/n)}$, but for small samples that's a poor approximation. And anyway, we've all got calculators, so you may as well do it right.)

The critical value of Z_{g2} is approximately 2. (This is a two-tailed test of excess kurtosis ≠ 0 at approximately the 0.05 significance level.)

- **If $Z_{g2} < -2$**, the population very likely has negative excess kurtosis (kurtosis <3, platykurtic), though you don't know how much.
- **If Z_{g2} is between −2 and +2**, you can't reach any conclusion about the kurtosis: excess kurtosis might be positive, negative, or zero.
- **If $Z_{g2} > +2$**, the population very likely has positive excess kurtosis (kurtosis >3, leptokurtic), though you don't know how much.

For the sample college men's heights (n=100), you found excess kurtosis of $G_2 = -0.2091$. The sample is platykurtic, but is this enough to let you say that the whole population is platykurtic (has lower kurtosis than the bell curve)?

First compute the standard error of kurtosis:

$$SEK = 2 \times SES \times \sqrt{[\,(n^2-1) / ((n-3)(n+5))\,]}$$

n = 100, and the SES was previously computed as 0.2414.

$$SEK = 2 \times 0.2414 \times \sqrt{[\,(100^2-1) / (97\times105)\,]} = 0.4784$$

The test statistic is

$$Z_{g2} = G_2/SEK = -0.2091 / 0.4784 = -0.44$$

You can't say whether the kurtosis of the population is the same as or different from the kurtosis of a normal distribution.

Assessing Normality

There are many ways to assess normality, and unfortunately none of them are without problems. Graphical methods are a good start, such as plotting a histogram and making a quantile plot. (You can find a TI-83 program to do those at MATH200A Program — Statistics Utilities for TI-83/84.)

The University of Surrey has a good survey or problems with normality tests, at How do I test the normality of a variable's distribution? That page recommends using the test statistics individually.

One test is the **D'Agostino-Pearson omnibus test**, so called because it uses the test statistics for both skewness and kurtosis to come up with a single p-value. The test statistic is

$$DP = Z_{g1}^2 + Z_{g2}^2 \text{ follows } \chi^2 \text{ with df=2} \qquad (8)$$

You can look up the p-value in a table, or use `χ²cdf` on a TI-83 or TI-84.

Caution: The D'Agostino-Pearson test has a tendency to err on the side of rejecting normality, particularly with small sample sizes. David Moriarty, in his StatCat utility, recommends that you **don't use D'Agostino-Pearson for sample sizes below 20.**

For college students' heights you had test statistics $Z_{g1} = -0.45$ for skewness and $Z_{g2} = 0.44$ for kurtosis. The omnibus test statistic is

$$DP = Z_{g1}^2 + Z_{g2}^2 = 0.45^2 + 0.44^2 = 0.3961$$

and the p-value for χ^2(2 df) > 0.3961, from a table or a statistics calculator, is 0.8203. You cannot reject the assumption of normality. (Remember, you never accept the null hypothesis, so you can't say from this test that the distribution *is* normal.) The histogram suggests normality, and this test gives you no reason to reject that impression.

Example 2: Size of Rat Litters

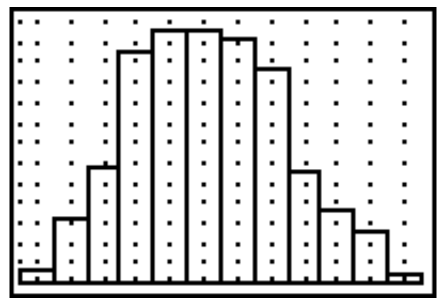
For a second illustration of inferences about skewness and kurtosis of a population, I'll use an example from Bulmer's *Principles of Statistics*:

Frequency distribution of litter size in rats, n=815												
Litter size	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	7	33	58	116	125	126	121	107	56	37	25	4

I'll spare you the detailed calculations, but you should be able to verify them by following equation (1) and equation (2):

$n = 815$, $\bar{x} = 6.1252$, $m_2 = 5.1721$, $m_3 = 2.0316$

skewness $g_1 = 0.1727$ and sample skewness $G_1 = 0.1730$



The sample is roughly symmetric but slightly skewed right, which looks about right from the histogram. The standard error of skewness is

$SES = \sqrt{[(6 \times 815 \times 814) / (813 \times 816 \times 818)]} = 0.0856$

Dividing the skewness by the SES, you get the test statistic

$Z_{g1} = 0.1730 / 0.0856 = 2.02$

Since this is greater than 2, you can say that **there is some positive skewness in the population.** Again, "some positive skewness" just means a figure greater than zero; it doesn't tell us anything more about the magnitude of the skewness.

If you go on to compute a 95% confidence interval of skewness from equation (4), you get $0.1730 \pm 2 \times 0.0856 = 0.00$ to 0.34.

What about the kurtosis? You should be able to follow equation (5) and compute a fourth moment of $m_4 = 67.3948$. You already have $m_2 = 5.1721$, and therefore

kurtosis $a_4 = m_4 / m_2^2 = 67.3948 / 5.1721^2 = 2.5194$

excess kurtosis $g_2 = 2.5194 - 3 = -0.4806$

sample excess kurtosis $G_2 = [814/(813 \times 812)] [816 \times (-0.4806+6) = -0.4762$

So the sample is moderately less peaked than a normal distribution. Again, this matches the histogram, where you can see the higher "shoulders".

What if anything can you say about the population? For this you need equation (7). Begin by computing the standard error of kurtosis, using n = 815 and the previously computed SES of 0.0.0856:

$$SEK = 2 \times SES \times \sqrt{[\ (n^2-1) / ((n-3)(n+5))\ ]}$$

$$SEK = 2 \times 0.0856 \times \sqrt{[\ (815^2-1) / (812\times820)\ ]} = 0.1711$$

and divide:

$$Z_{g2} = G_2/SEK = -0.4762 / 0.1711 = -2.78$$

Since Z_{g2} is comfortably below −2, you can say that **the distribution of all litter sizes is platykurtic**, less sharply peaked than the normal distribution. But be careful: you know that it is platykurtic, but you don't know by how much.

You already know the population is not normal, but let's apply the D'Agostino-Pearson test anyway:

$$DP = 2.02^2 + 2.78^2 = 11.8088$$

$$\text{p-value} = P(\ \chi^2(2) > 11.8088\ ) = 0.0027$$

The test agrees with the separate tests of skewness and kurtosis: sizes of rat litters, for the entire population of rats, is not normally distributed.

What's New

- **26 Apr 2011**: identify the t(4) distribution and the beta distributions in their captions
- **20 Dec 2010**: update citations to textbooks
- **23 Oct 2010**: restore a missing minus sign, thanks to Edward B. Taylor (intervening changes suppressed)
- **13 Dec 2008**: new document
  This page uses some material from the old *Skewness and Kurtosis on the TI-83/84*, which was first created 12 Jan 2008 and replaced 7 Dec 2008 by MATH200B Program part 1; but there are new examples and pictures and

considerable new or rewritten material.

For updates and new info, go to http://www.tc3.edu/instruct/sbrown/stat/