

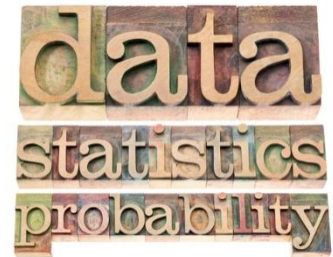
Are Skewness and Kurtosis Useful Statistics?

Note: This article was originally published in April 2008 and was updated in February 2016. The original article indicated that kurtosis was a measure of the flatness of the distribution – or peakedness. This is technically not correct (see below). Kurtosis is a measure of the combined weight of the tails relative to the rest of the distribution. This article has been revised to correct that misconception. Additional information on both skewness and kurtosis has also been added.

In this Issue

- [Introduction](#)
- [Skewness](#)
- [Kurtosis](#)
- [Our Population](#)
- [The Impact of Sample Size](#)
- [Conclusions](#)
- [Quick Links](#)

You have a set of samples. Maybe you took 15 samples from a batch of finished product and measured those samples for density. Now you are armed with data you can analyze. And your software package has a feature that will generate the descriptive statistics for these data. You enter the data into your software package and run the descriptive statistics. You get a lot of numbers – the sample size, average, standard deviation, range, maximum, minimum and a host of other numbers. You spy two numbers: the skewness and kurtosis. What do these two statistics tell you about your sample?



This month's newsletter covers the skewness and kurtosis statistics. These two statistics are called "shape" statistics, i.e., they describe the shape of the distribution. What do the skewness and kurtosis really represent? And do they help you understand your process any better? Are they useful statistics? Let's take a look

Introduction

Skewness and kurtosis are two commonly listed values when you run a software's descriptive statistics function. Many books say that these two statistics give you insights into the shape of the distribution.

Skewness is a measure of the symmetry in a distribution. A symmetrical data set will have a skewness equal to 0. So, a normal distribution will have a skewness of 0. Skewness essentially measures the relative size of the two tails.

Kurtosis is a measure of the combined sizes of the two tails. It measures the amount of probability in the tails. The value is often compared to the kurtosis of the normal distribution, which is equal to 3. If the kurtosis is greater than 3, then the data set has heavier tails than a normal distribution (more in the

tails). If the kurtosis is less than 3, then the data set has lighter tails than a normal distribution (less in the tails). Careful here. Kurtosis is sometimes reported as “excess kurtosis.” Excess kurtosis is determined by subtracting 3 from the kurtosis. This makes the normal distribution kurtosis equal 0. Kurtosis originally was thought to measure the peakedness of a distribution. Though you will still see this as part of the definition in many places, this is a misconception.



Skewness and kurtosis involve the tails of the distribution. These are presented in more detail below.

Skewness

Skewness is usually described as a measure of a data set’s symmetry – or lack of symmetry. A perfectly symmetrical data set will have a skewness of 0. The normal distribution has a skewness of 0.

The skewness is defined as (Advanced Topics in Statistical Process Control, Dr. Donald Wheeler, www.spcpress.com):

$$a_3 = \sum \frac{(X_i - \bar{X})^3}{ns^3}$$

where n is the sample size, X_i is the i^{th} value, \bar{X} is the average and s is the sample standard deviation. Note the exponent in the summation. It is “3”. The skewness is referred to as the “third standardized central moment for the probability model.”

Most software packages use a formula for the skewness that takes into account sample size:

$$Skewness = \frac{n}{(n - 1)(n - 2)} \sum \frac{(X_i - \bar{X})^3}{s^3}$$

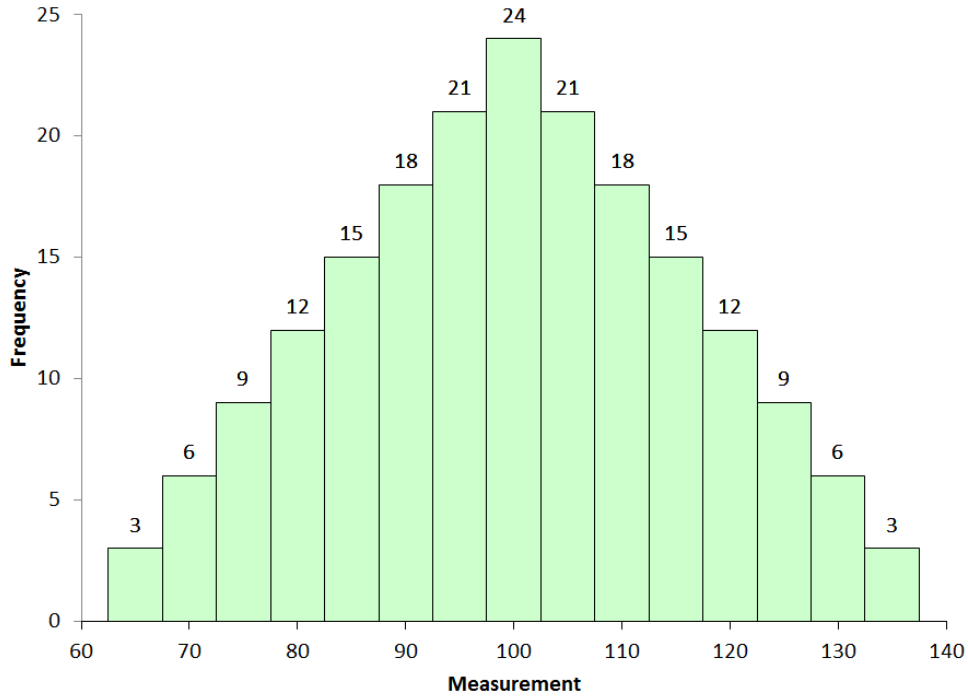
This sample size formula is used here. It is also what Microsoft Excel uses. The difference between the two formula results becomes very small as the sample size increases.

Figure 1 is a symmetrical data set. It was created by generating a set of data from 65 to 135 in steps of 5 with the number of each value as shown in Figure 1. For example, there are 3 65’s, 6 65’s, etc.



A truly symmetrical data set has a skewness equal to 0. It is easy to see why this is true from the skewness formula. Look at the term in the numerator after the summation sign. Each individual X value is subtracted from the average. So, if a set of data is truly symmetrical, for each point that is a distance “d” above the average, there will be a point that is a distance “-d” below the average.

Figure 1: Symmetrical Data Set with Skewness = 0



Consider the value of 65 and value of 135. The average of the data in Figure 1 is 100. So, the following is true when X = 65:

$$\frac{(X_i - \bar{X})^3}{s^3} = \frac{(65 - 100)^3}{s^3} = \frac{(-35)^3}{s^3} = \frac{-4278}{s^3}$$

For X = 135, the following is true:

$$\frac{(X_i - \bar{X})^3}{s^3} = \frac{(135 - 100)^3}{s^3} = \frac{(35)^3}{s^3} = \frac{4278}{s^3}$$

So, the -4278 and +4278 even out at 0. So, a truly symmetrical data set will have a skewness of 0.

To explore positive and negative values of skewness, let's define the following term:

$$S_{\text{above}} = |\sum(X_i - \bar{X})^3| \text{ if } X_i \text{ is above the average}$$

$$S_{\text{below}} = |\sum(X_i - \bar{X})^3| \text{ if } X_i \text{ is below the average}$$

So, S_{above} can be viewed as the “size” of the deviations from average when X_i is above the average. Likewise, S_{below} can be viewed as the “size” of the deviations from average when X_i is below the average. Then, skewness becomes the following:

$$Skewness = \frac{n}{(n-1)(n-2)} \sum \frac{(X_i - \bar{X})^3}{s^3} = \frac{n}{s^3(n-1)(n-2)} (S_{above} - S_{below})$$

If S_{above} is larger than S_{below} , then skewness will be positive. This typically means that the right-hand tail will be longer than the left-hand tail. Figure 2 is an example of this. The skewness for this data set is 0.514. A positive skewness indicates that the size of the right-handed tail is larger than the left-handed tail.

Figure 2: Data Set with Positive Skewness

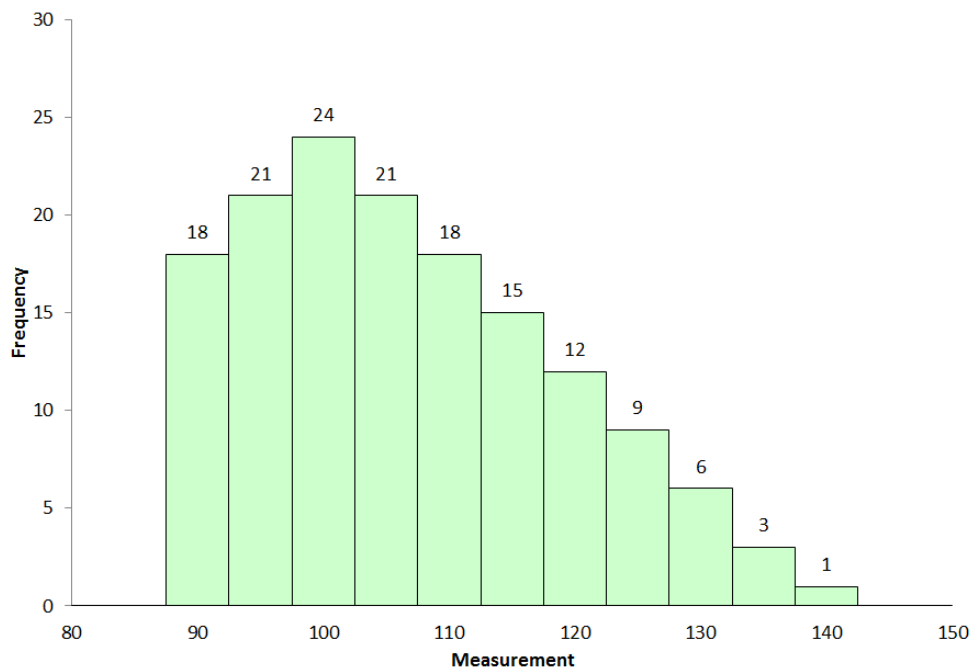
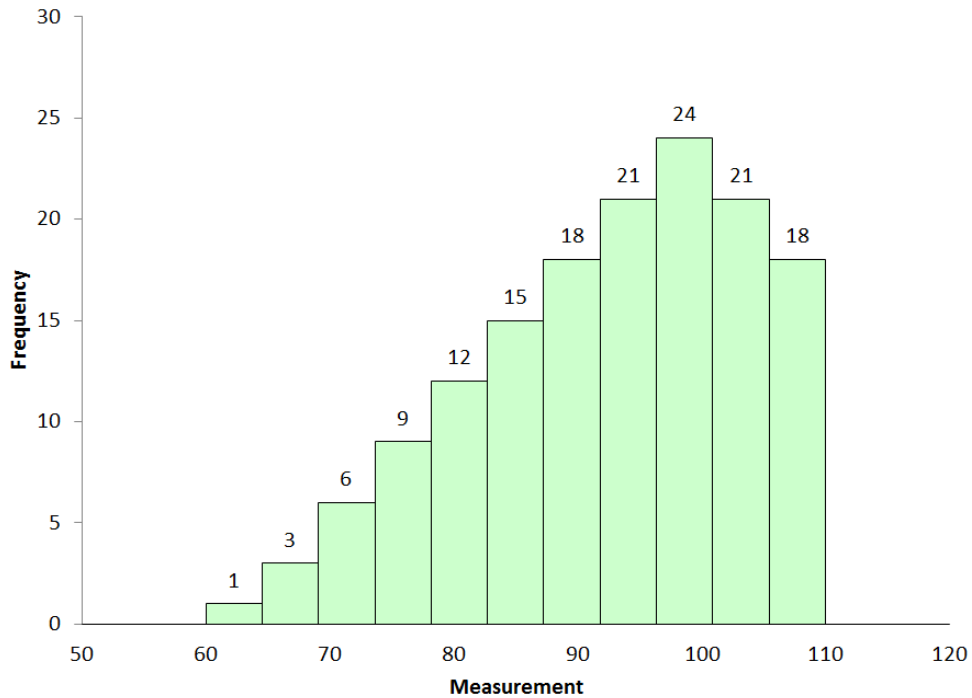


Figure 3 is an example of data set with negative skewness. It is the mirror image essentially of Figure 2. The skewness is -0.514. In this case, S_{below} is larger than S_{above} . The right-hand tail will typically be longer than the left-hand tail.

So, when is the skewness too much? The rule of thumb seems to be:

- If the skewness is between -0.5 and 0.5, the data are fairly symmetrical
- If the skewness is between -1 and -0.5 or between 0.5 and 1, the data are moderately skewed
- If the skewness is less than -1 or greater than 1, the data are highly skewed

Figure 3: Data Set with Negative Skewness



Kurtosis

How to define kurtosis? This is really the reason this article was updated. If you search for definitions of kurtosis, you will see some definitions that includes the word “peakedness” or other similar terms. For example,

- “Kurtosis is the degree of peakedness of a distribution” – [Wolfram MathWorld](#)
- “We use kurtosis as a measure of peakedness (or flatness)” – [Real Statistics Using Excel](#)

You can find other definitions that include peakedness or flatness when you search the web. The problem is these definitions are not correct. Dr. Peter Westfall published an article that addresses why kurtosis does not measure peakedness ([link to article](#)). He said:

“Kurtosis tells you virtually nothing about the shape of the peak – its only unambiguous interpretation is in terms of tail extremity.”

Dr. Westfall includes numerous examples of why you cannot relate the peakedness of the distribution to the kurtosis.

Dr. Donald Wheeler also discussed this in his [two-part series on skewness and kurtosis](#). He said:

“Kurtosis was originally thought to be a measure the “peakedness” of a distribution. However, since the central portion of the distribution is virtually ignored by this parameter, kurtosis cannot be said to measure peakedness directly. While there is a correlation between peakedness and kurtosis, the relationship is an indirect and imperfect one at best.”



Dr. Wheeler defines kurtosis as:

“The kurtosis parameter is a measure of the combined weight of the tails relative to the rest of the distribution.”

So, kurtosis is all about the tails of the distribution – not the peakedness or flatness. It measures the tail-heaviness of the distribution.

Kurtosis is defined as:

$$a_4 = \sum \frac{(X_i - \bar{X})^4}{ns^4}$$

where n is the sample size, X_i is the i^{th} X value, \bar{X} is the average and s is the sample standard deviation. Note the exponent in the summation. It is “4”. The kurtosis is referred to as the “fourth standardized central moment for the probability model.”

Here is where it gets a little tricky. If you use the above equation, the kurtosis for a normal distribution is 3. Most software packages (including Microsoft Excel) use the formula below.

$$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \frac{(X_i - \bar{X})^4}{s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

This formula does two things. It takes into account the sample size and it subtracts 3 from the kurtosis. With this equation, the kurtosis of a normal distribution is 0. This is really the excess kurtosis, but most software packages refer to it as simply kurtosis. The last equation is used here. So, if a data set has a positive kurtosis, it has more in the tails than the normal distribution. If a data set has a negative kurtosis, it has less in the tails than the normal distribution.

Since the exponent in the above is 4, the term in the summation will always be positive – regardless of whether X_i is above or below the average. X_i values close to the average contribute very little to the kurtosis. The tail values of X_i contribute much more to the kurtosis.

Look back at Figures 2 and 3. They are essentially mirror images of each other. The skewness of these data sets is different: 0.514 and -0.514. But the kurtosis is the same. Both have a kurtosis of -0.527. This is because kurtosis looks at the combined size of the tails.

The kurtosis decreases as the tails become lighter. It increases as the tails become heavier. Figure 4 shows an extreme case. In this data set, each value occurs 10 times. The values are 65 to 135 in increments of 5. The kurtosis of this data set is -1.21. Since this value is less than 0, it is considered to be a “light-tailed” data set. It has as much data in each tail as it does in the peak. Note that this is a symmetrical distribution, so the skewness is zero.

Figure 4: Negative Kurtosis Example

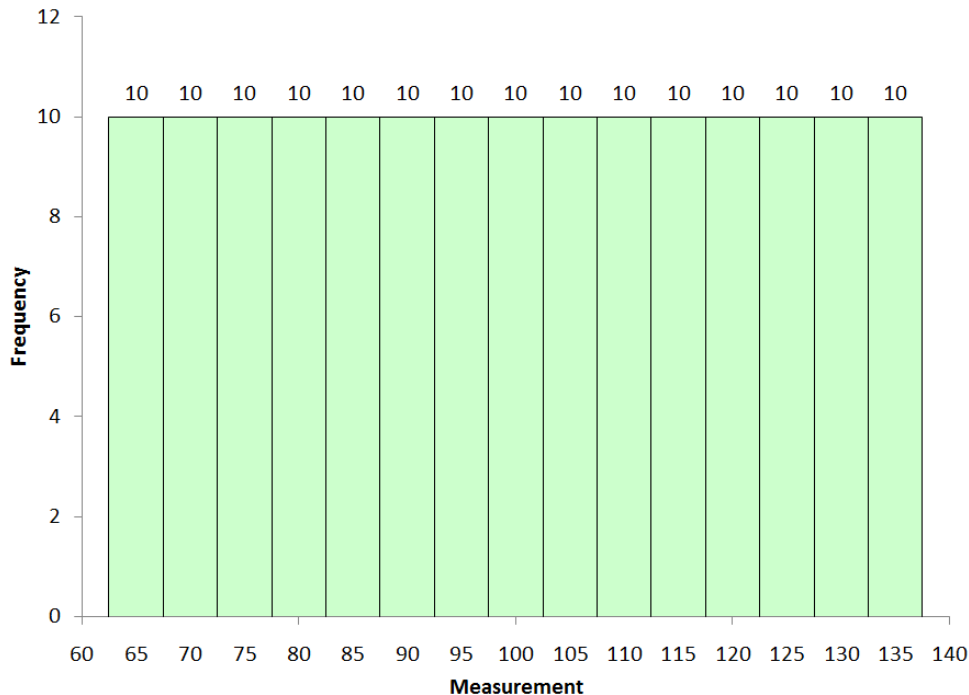


Figure 5 shows a data set with more weight in the tails. The kurtosis of this data set is 1.86.

Most often, kurtosis is measured against the normal distribution. If the kurtosis is close to 0, then a normal distribution is often assumed. These are called mesokurtic distributions. If the kurtosis is less than zero, then the distribution is light tails and is called a platykurtic distribution. If the kurtosis is greater than zero, then the distribution has heavier tails and is called a leptokurtic distribution.

The problem with both skewness and kurtosis is the impact of sample size. This is described below.

Our Population

Are the skewness and kurtosis any value to you? You take a sample from your process and look at the calculated values for the skewness and kurtosis. What can you tell from these two results? To explore this, a data set of 5000 points was randomly generated. The goal was to have a mean of 100 and a standard deviation of 10. The random generation resulted in a data set with a mean of 99.95 and a standard deviation of 10.01. The histogram for these data is shown in Figure 6 and looks fairly bell-shaped.

Figure 5: Positive Kurtosis Example

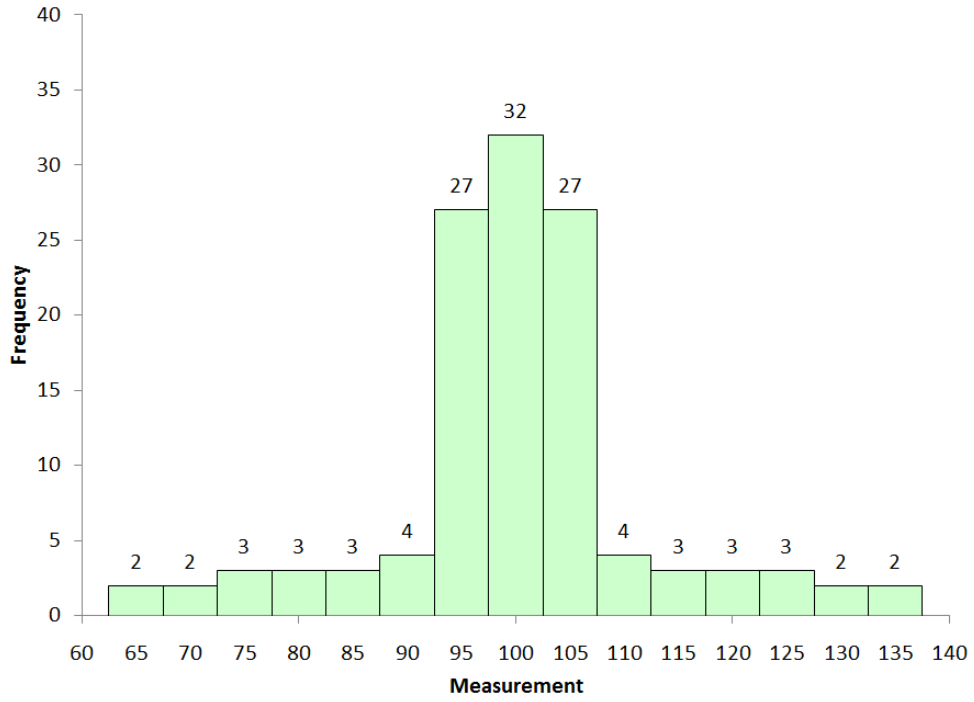
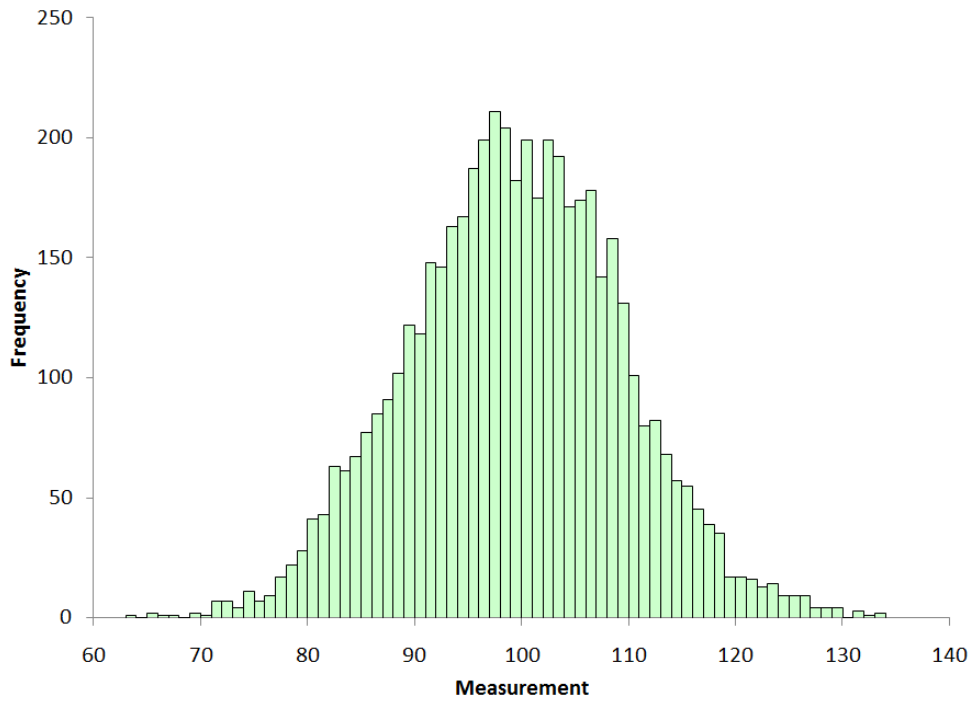


Figure 6: Population Histogram



The skewness of the data is 0.007. The kurtosis is 0.03. Both values are close to 0 as you would expect for a normal distribution. These two numbers represent the "true" value for the skewness and kurtosis since they were calculated from all the data. In real life, you don't know the real skewness and kurtosis because you have to sample the process. This is where the problem begins for skewness and kurtosis. Sample size has a big impact on the results.



Impact of Sample Size on Skewness and Kurtosis

The 5,000-point data set above was used to explore what happens to skewness and kurtosis based on sample size. For example, suppose we wanted to determine the skewness and kurtosis for a sample size of 5. 5 results were randomly selected from the data set above and the two statistics calculated. This was repeated for the sample sizes shown in Table 1.

Table 1: Impact of Sample Size on Skewness and Kurtosis

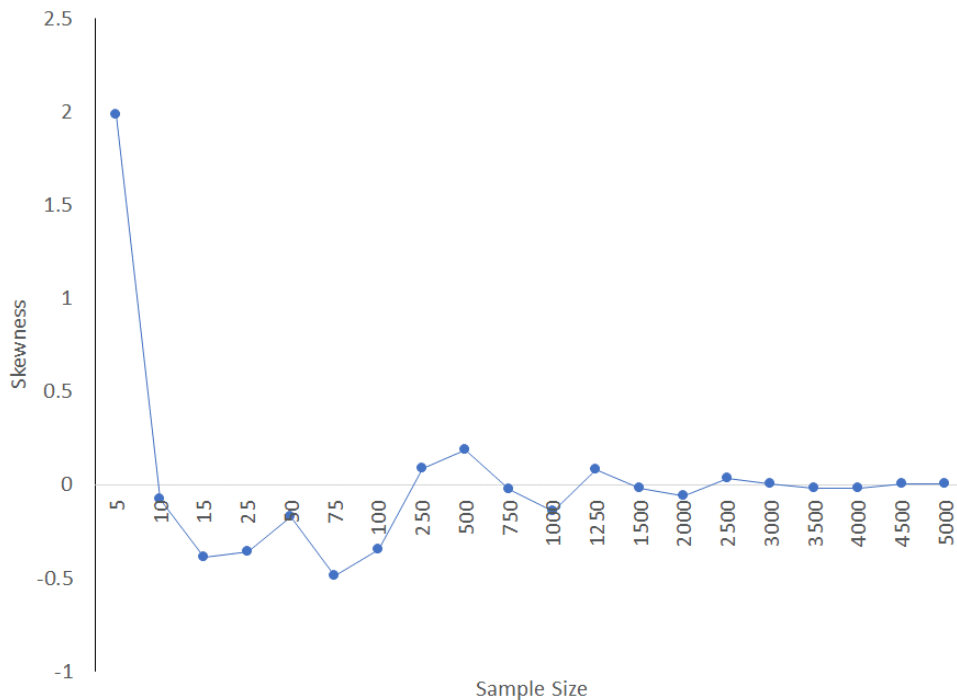
Sample Size	Skewness	Kurtosis
5	1.983	3.974
10	-0.078	-1.468
15	-0.384	0.127
25	-0.356	-0.025
50	-0.169	-0.752
75	-0.489	0.615
100	-0.346	0.671
250	0.089	0.061
500	0.186	0.232
750	-0.02	0.042
1000	-0.138	0.062
1250	0.085	0.079
1500	-0.017	0.001
2000	-0.059	-0.009
2500	0.037	0.096
3000	0.009	0.005
3500	-0.015	0.004
4000	-0.015	-0.009
4500	0.009	0.036
5000	0.007	0.03

Notice how much different the results are when the sample size is small compared to the "true" skewness and kurtosis for the 5,000 results. For a sample size of 25, the skewness was -.356 compared

to the true value of 0.007 while the kurtosis was -0.025. Both signs are opposite of the true values which would lead to wrong conclusions about the shape of the distribution. There appears to be a lot of variation in the results based on sample size.

Figure 7 shows how the skewness changes with sample size. Figure 8 is the same but for kurtosis.

Figure 7: Skewness versus Sample Size



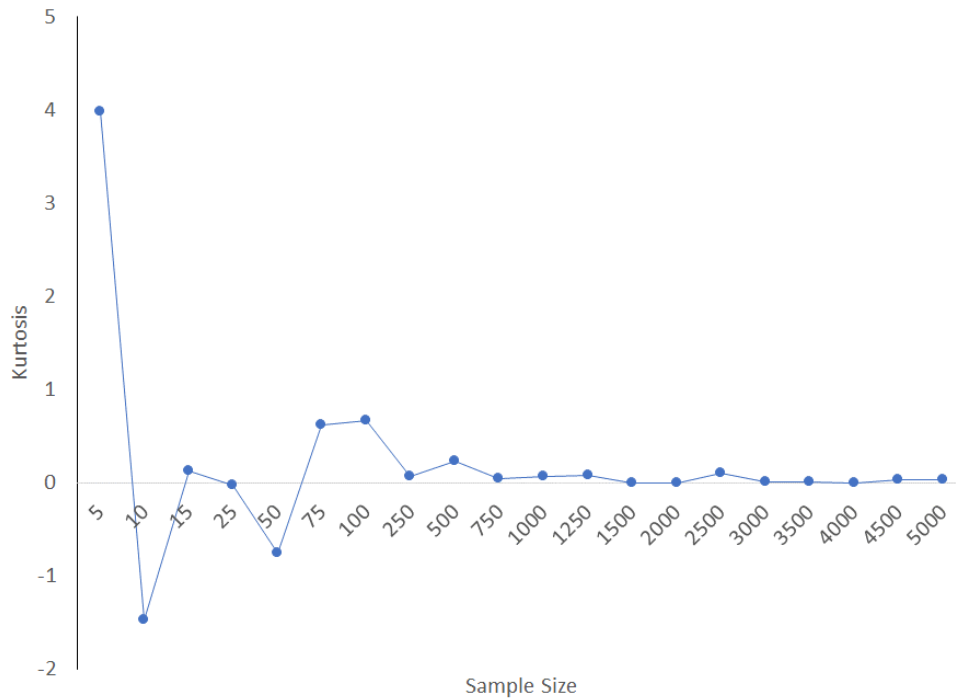
30 is a common number of samples used for process capability studies. A subgroup size of 30 was randomly selected from the data set. This was repeated 100 times. The skewness varied from -1.327 to 1.275 while the kurtosis varied from -1.12 to 2.978. What kind of decisions can you make about the shape of the distribution when the skewness and kurtosis vary so much? Essentially, no decisions.

Conclusions

The skewness and kurtosis statistics appear to be very dependent on the sample size. The table above shows the variation. In fact, even several hundred data points didn't give very good estimates of the true kurtosis and skewness. Smaller sample sizes can give results that are very misleading. Dr Wheeler wrote in his book mentioned above:

"In short, skewness and kurtosis are practically worthless. Shewhart made this observation in his first book. The statistics for skewness and kurtosis simply do not provide any useful information beyond that already given by the measures of location and dispersion."

Figure 8: Kurtosis versus Sample Size



Walter Shewhart was the "Father" of SPC. So, don't put much emphasis on skewness and kurtosis values you may see. And remember, the more data you have, the better you can describe the shape of the distribution. But, in general, it appears there is little reason to pay much attention to skewness and kurtosis statistics. Just look at the histogram. It often gives you all the information you need.

To download the workbook containing the macro and results that generated the above tables, please [click here](#).

Quick Links

[Visit our home page](#)

[SPC for Excel Software](#)

[SPC Training](#)

[SPC Consulting](#)

[SPC Knowledge Base](#)

[Ordering Information](#)

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
BPI Consulting, LLC