Measuring Skewness: A Forgotten Statistic?

David P. Doane
Oakland University

Lori E. Seward
University of Colorado

Key Words: Moment Coefficient; Pearson 2; Monte Carlo; Type I Error; Power; Normality Test.

Abstract

This paper discusses common approaches to presenting the topic of skewness in the classroom, and explains why students need to know how to measure it. Two skewness statistics are examined: the Fisher-Pearson standardized third moment coefficient, and the Pearson 2 coefficient that compares the mean and median. The former is reported in statistical software packages, while the latter is all but forgotten in textbooks. Given its intuitive appeal, why did Pearson 2 disappear? Is it ever useful? Using Monte Carlo simulation, tables of percentiles are created for Pearson 2. It is shown that while Pearson 2 has lower power, it matches classroom explanations of skewness and can be calculated when summarized data are available. This paper suggests reviving the Pearson 2 skewness statistic for the introductory statistics course because it compares the mean to the median in a precise way that students can understand. The paper reiterates warnings about what any skewness statistic can actually tell us.

1. Introduction

In an introductory level statistics course, instructors spend the first part of the course teaching students three important characteristics used when summarizing a data set: center, variability, and shape. The instructor typically begins by introducing visual tools to get a "picture" of the data. The concept of *center* (also *location* or *central tendency*) is familiar to most students and they can easily see the "middle" or "typical" data values on a graph such as a histogram. The

concept of *variability* (also *dispersion* or *spread*) is less familiar, but when shown histograms or dot plots of different data sets on the same scale, students can usually identify which data sets have more variability and which have less. The concept of *shape* is even less familiar than variability, but visual tools are again useful for comparing symmetric and asymmetric distributions. Instructors can use examples familiar to students, such as ordering time at Starbuck's or professional athletes' salaries, as data sets that clearly exhibit asymmetrical distributions.

Studies have shown (delMas, Garfield, Ooms, and Chance 2007) that students' abilities to describe and interpret a variable's distribution from a histogram, in the context of the data, is quite high even before taking a first course in statistics. While qualitative descriptions of a distribution are helpful for summarizing a data set, students eventually will be asked to use statistics to numerically describe a distribution in terms of center, variability, and shape. Without difficulty, they can see how the *mean*, *median*, and *mode* can indicate the center, and how *standard deviation* and *range* can describe variability. But the terms *skewness* and *kurtosis* are non-intuitive. Worse, skewness and kurtosis statistics and formulas are opaque to the average student, and lack concrete reference points.

Cobb and Moore (1997, p. 803) note that "In data analysis, context provides meaning." Realizing this, over the past several decades, more and more instructors are using sample data arising from real (or realistic) scenarios. One result is that students are learning that perfectly symmetrical graphical displays are hard to find. Even with the ability to verbally describe a distribution from a visual display, researchers have found (delMas et al. 2007) that students cannot translate their understanding of shape when asked to compare numerical statistics such as the mean and median. Hence, measures of skewness are becoming more important (although many instructors may reasonably conclude that kurtosis does not deserve extended discussion in a basic statistics class). To answer this need, our paper suggests reviving an intuitive skewness statistic that compares the mean to the median in a precise way that students can understand.

2. Visual Displays

A textbook discussion would typically begin by showing the relative positions of the mean, median, and mode in smooth population probability density functions, as illustrated in Figure 1. The explanation will mainly refer to the positions of the mean and median. There may be comments about tail length and the role of extreme values in pulling the mean up or down. The mode usually gets scant mention, except as the "high point" in the distribution.
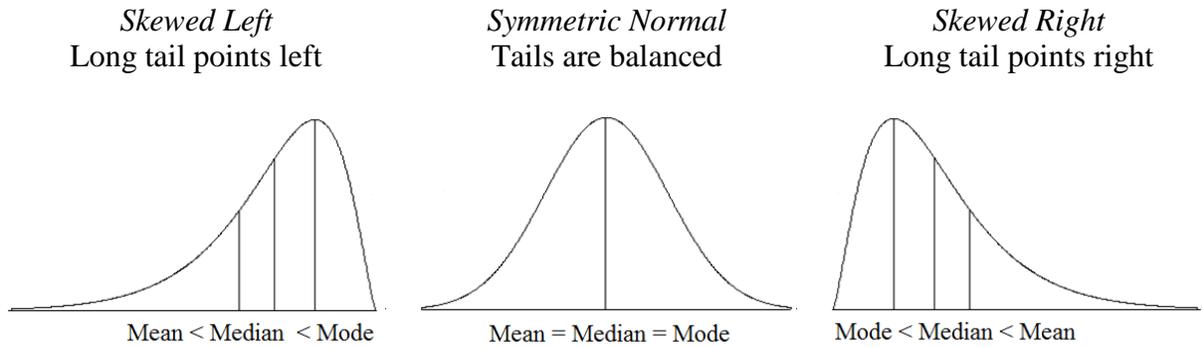
Skewed Left
Long tail points left

Symmetric Normal
Tails are balanced

Skewed Right
Long tail points right

Mean < Median < Mode

Mean = Median = Mode

Mode < Median < Mean

Figure 1. Sketches showing general position of mean, median, and mode in a population.

Next, a textbook might present stylized sample histograms, as in Figure 2. Figures like these allow the instructor to point out that (a) "symmetric" need not imply a "bell-shaped" distribution; (b) extreme data values in one tail are not unusual in real data; and (c) real samples may not resemble any simple histogram prototype. The instructor can discuss causes of asymmetry (e.g., why waiting times are exponential, why earthquake magnitudes follow a power law, why home prices are skewed to the right) or the effects of outliers and extreme data values (e.g., how a customer with a complicated order affects the queue at Starbuck's, how one heart transplant affects the health insurance premiums for a pool of employees).
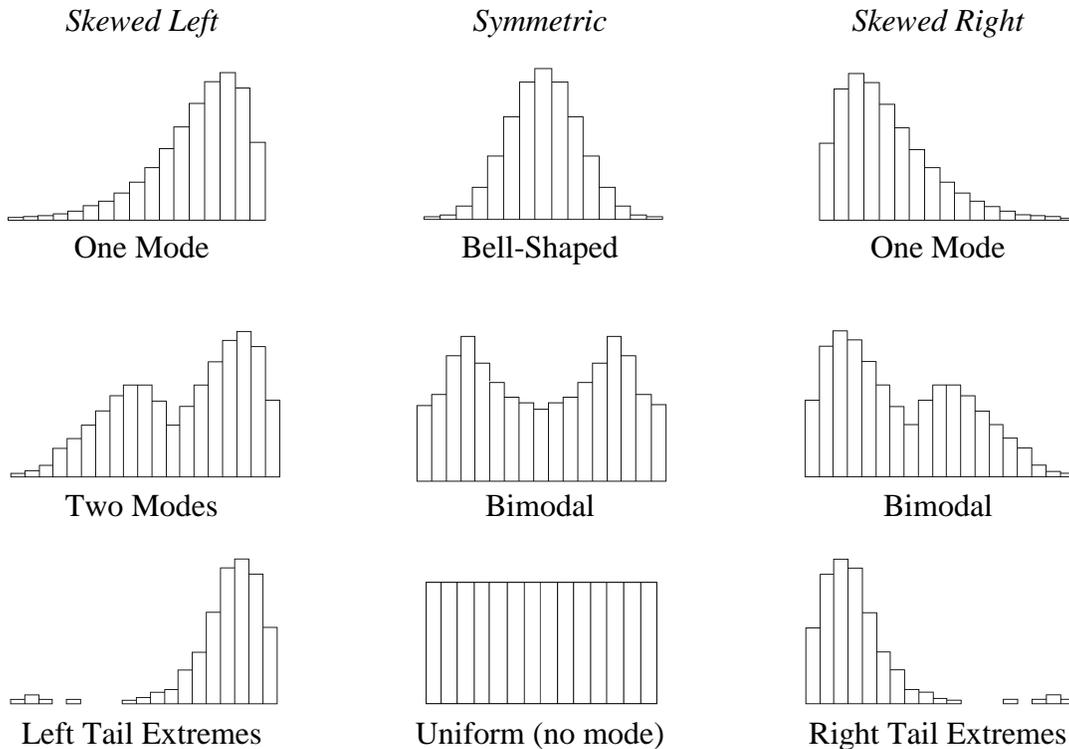
Skewed Left

Symmetric

Skewed Right

One Mode

Bell-Shaped

One Mode

Two Modes

Bimodal

Bimodal

Left Tail Extremes

Uniform (no mode)

Right Tail Extremes

Figure 2. Illustrative prototype histograms.

Examples are essential. For example, Figure 3 shows 1990 data on death rates in 150 nations. Depending on the binning, one may gain varying impressions of skewness. The instructor can call attention, in a general way, to the fact that a larger sample size (when available) is more likely to yield a histogram that reflects the true population shape (and allows for more bins in a histogram).
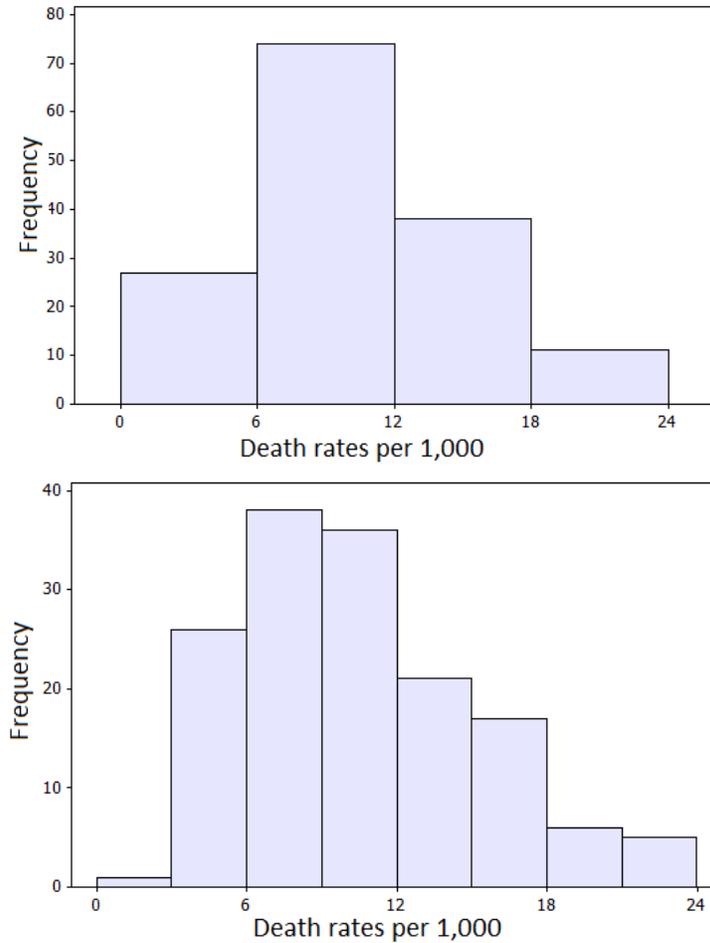

Figure 3. Effect of histogram binning on perceived skewness ($n = 150$).

Other tools of exploratory data analysis (EDA) such as the boxplot or dotplot may be used to assess skewness visually. The less familiar beam-and-fulcrum plot (Doane and Tracy 2001) reveals skewness by showing the mean in relation to tick marks at various standard deviations from the mean, e.g., $\bar{x} \pm 1s$, $\bar{x} \pm 2s$, and $\bar{x} \pm 3s$. But the boxplot and beam-and-fulcrum displays do not reveal sample size. For that reason, the dotplot is arguably a more helpful visual tool for assessing skewness. In Figure 4, all three displays suggest positive skewness.
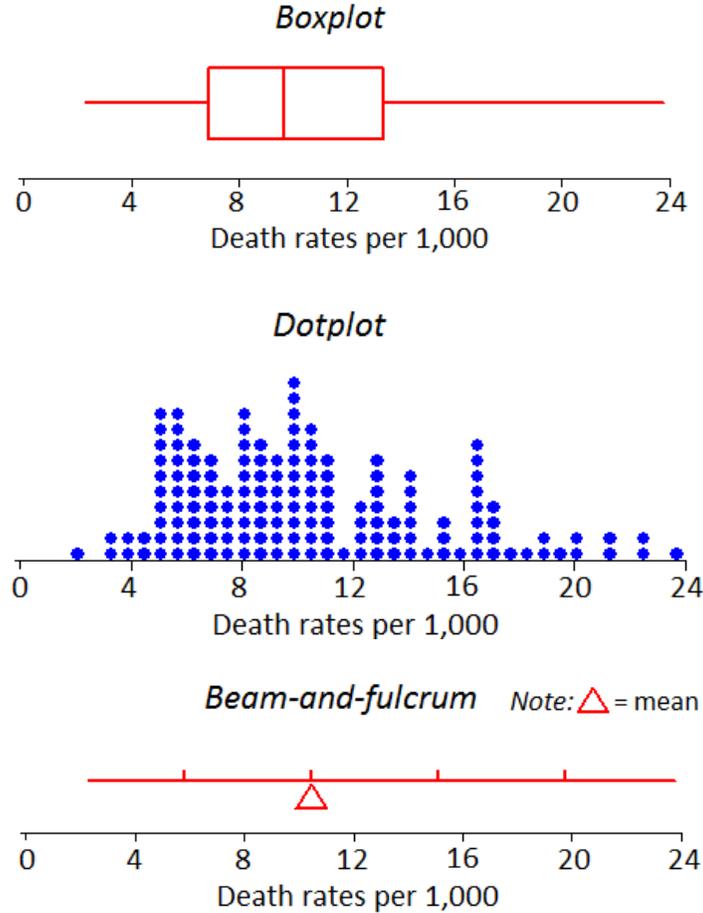
Boxplot

Dotplot

Beam-and-fulcrum *Note:* △ = mean

Figure 4. EDA plots can reveal skewness.

More sophisticated visual tests for symmetry and normality, such as the empirical cumulative distribution function (ECDF) and normal probability plot (e.g., D'Agostino and Stephens 1986) usually are covered later in the semester (if at all). An instructor who does not want to develop the idea of statistical inference at this point can simply say that the current learning objective is to understand the concept of skewness, and to recognize its symptoms in a general way. Exam questions can then be based mostly on visual displays. That may be the end of the story.

Unfortunately, most students will say that the "data are skewed" if there is even the slightest difference between the sample mean and sample median, or if the histogram is even slightly asymmetric. They are thinking that the population cannot be symmetric if *any* differences exist in the sample. Can the instructor allow such statements to pass without comment or correction? Hoping to avoid a deep dive into statistical inference, the instructor could explain that, even in samples from a symmetric population, we do not expect the sample mean generally to be *exactly* equal to the median, or the histogram to be *exactly* like the population. Simulation can be used to illustrate, perhaps by computing the mean and median from samples generated from the Excel function =NORMINV(RAND(), μ, σ). The instructor can also remind the students that a histogram's

appearance varies if we alter the bin limits, so any single histogram may not give a definitive view of population shape. But students like clear-cut answers. The next question is likely to be:

"OK, but how big a difference must we see between the mean and median to say that the population is skewed? Isn't there some kind of test for skewness?"

Students who notice the skewness statistic in Excel's Descriptive Statistics may ask more specific questions. For example:

"My sample skewness statistic from Excel is –0.308. So can I say that my sample of 12 items came from a left-skewed population?"

"In my sample of 12 items, the sample mean 56.56 exceeds the sample median 53.83, yet my skewness statistic is negative –0.308. How can that be?"

3. Skewness Statistics

Since Karl Pearson (1895), statisticians have studied the properties of various statistics of skewness, and have discussed their utility and limitations. This research stream covers more than a century. For an overview, see Arnold and Groenveld (1995), Groenveld and Meeden (1984), and Rayner, Best and Matthews (1995). Empirical studies have examined bias, mean squared error, Type I error, and power for samples of various sizes drawn from various populations. A recent study by Tabor (2010) ranked 11 different statistics in terms of their power for detecting skewness in samples from populations with varying degrees of skewness. MacGillivray (1986) concludes that "…the relative importance of the different orderings and measures depends on circumstances, and it is unlikely that any one could be described as most important …". He notes that describing skewness is really a special case of comparing distributions. This key point is perhaps a bit subtle for students. Students (and instructors) merely need to bear in mind that we are not testing for symmetry in general. Rather, the (often implicit) null hypothesis must refer to a specific symmetric population. Because the most common reference point is the normal distribution (especially in an introductory statistics class) we will limit our discussion accordingly.

Mathematicians discuss skewness in terms of the second and third moments around the mean, i.e., $m_2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $m_3 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$. Mathematical statistics textbooks and a few software packages (e.g., Stata, Visual Statistics, early versions of Minitab) report the traditional *Fisher-Pearson coefficient of skewness*:

$$[1a] \qquad g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{3/2}}.$$

Following Pearson's notation, this statistic is sometimes referred to as $\sqrt{\beta_1}$ which is awkward because g_1 can be negative. Pearson and Hartley (1970) provide tables for g_1 as a test for departure from normality (i.e., testing the sample against one particular symmetric distribution). Although well documented and widely referenced in the literature, this formula does not correspond to what students will see in most software packages nowadays. Major software packages available to educators (e.g., Minitab, Excel, SPSS, SAS) include an adjustment for sample size, and provide the *adjusted Fisher-Pearson standardized moment coefficient*[1]:

$$[1b] \qquad G_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s} \right)^3 .$$

In large samples, g_1 and G_1 will be similar. Few students will be aware of this formula because it is buried within the help files for the software. The formula for G_1 is probably not even in the textbook unless the student is studying mathematical statistics. However, this statistic is included in Excel's Data Analysis > Descriptive Statistics and is calculated by the Excel function =SKEW(Array), so it will be seen by millions of students. If students look up "skewness" in Wikipedia, they will find a different-looking but equivalent formula:

$$[1c] \qquad G_1 = \frac{\sqrt{n(n-1)}}{n-2} \left[\frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^3}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{3/2}} \right] .$$

This alternate formulation (1c) has the attraction of showing that the adjustment for sample size approaches unity as n increases. Joanes and Gill (1998) compare bias and mean squared error (MSE) of different measures of skewness in samples of various sizes from normal and skewed populations. G_1 is shown to perform well, for example, having small MSE in samples from skewed populations.

Unfortunately, none of these formulas is likely to convey very much to a student. An ambitious instructor can dissect such formulas to impart grains of understanding to the best students, while the rest of the class groans as the discussion turns to second and third moments around the mean. The resulting insights are, at best, likely to be short-lived. Yet once students realize that there is a formula for skewness and see it in Excel, they will want to know how to interpret it. The instructor must decide what to say about a statistic such as G_1 without spending more time than the topic is worth. A minimalist might say that

- Its sign reflects the direction of skewness.
- It compares the sample with a normal (symmetric) distribution.

---

[1] Not many years ago, computer packages reported g_1 without an adjustment for sample size. Although the adjustment is now incorporated in software packages, textbooks (with a few exceptions) do not report an adapted version of the Pearson-Hartley tables. For that matter, many textbooks show no table of critical values at all. Without a table, why even mention the sample skewness statistic?

- Values far from zero suggest a non-normal (skewed) population.
- The statistic has an adjustment for sample size.
- The adjustment is of little consequence in large samples.

Because it is used in Excel (millions of customers) let us look more closely at G_1. An ambitious instructor could introduce Table 1, showing critical values, here called a "90 percent range" to avoid introducing formal hypothesis testing terminology. Students can learn to use such a table to decide whether or not the sample statistic is far enough from zero to conclude that the sample probably did not come from a normal population. Because the table starts at $n = 25$, the instructor can point out that skewness is hard to judge in smaller samples (a wide expected range for G_1). The instructor can also use the table to explain that, in samples from a normal population, the expected range of G_1 decreases as sample size increases (and conversely). The instructor should also mention that G_1 is *not* a general test of symmetry, because the table refers only to a normal population.

Table 1. 90% range for sample skewness coefficient G_1.

n	Lower Limit	Upper Limit	n	Lower Limit	Upper Limit
25	-0.726	0.726	90	-0.411	0.411
30	-0.673	0.673	100	-0.391	0.391
40	-0.594	0.594	150	-0.322	0.322
50	-0.539	0.539	200	-0.281	0.281
60	-0.496	0.496	300	-0.230	0.230
70	-0.462	0.462	400	-0.200	0.200
80	-0.435	0.435	500	-0.179	0.179

Source: David P. Doane and Lori E. Seward (2011), *Applied Statistics in Business and Economics*, 3e, (McGraw-Hill), p. 155. The table is adapted from E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, 3rd Edition, Cambridge University Press, 1970, page 207 using an adjustment for sample size. Values outside this range would suggest a non-normal population. Table used with permission.

For example, G_1 tells us that the death rate data for150 nations do not seem to be from a normal distribution, because $G_1 = 0.735$ is well outside the 90 percent range –0.322 to +0.322 for a sample of $n = 150$. The value of G_1 is shown in Excel's descriptive statistics output shown in Table 2. The G_1 statistic is helpful in this example, because the visual displays (Figures 3 and 4) do not appear strongly asymmetric. Further, the median (9.65) is arguably close to the mean (10.47). Without G_1 we might not detect skewness in the sample[2].

---

[2] Students who ask whether this is a sample or a population can be reminded that death rates for a nation are measured only at a given point in time, and so a given year's data may be affected by transient factors. In this sense, it resembles a sample from some kind of true, long-run steady state.

Table 2. Excel's descriptive statistics.

Death Rate Per 1,000	
Mean	10.468
Standard Error	0.3787
Median	9.65
Mode	9.7
Standard Deviation	4.6381
Sample Variance	21.512
Kurtosis	-0.03876
Skewness	0.73465
Range	21.5
Minimum	2.3
Maximum	23.8
Sum	1570.2
Count	150

A word about kurtosis is in order. Horswell and Looney (1993, p. 437) note that "The performance of skewness tests is shown to be very sensitive to the kurtosis of the underlying distribution." Few instructors say much about kurtosis, partly because it is difficult explain, but also because it is difficult to judge from histograms. Kurtosis is essentially a property of symmetric distributions (Balanda and MacGillivray 1988). Data sets containing extreme values will not only be skewed, but also generally will be leptokurtic. We cannot therefore speak of non-normal skewness as if it were separable from non-normal kurtosis. The best we can do is to focus on the skewness statistic simply as one test for departure from the symmetric normal distribution[3].

Because G_1 is a common, well-documented statistic, why look further? There are three reasons:

- The mathematical form of this statistic is likely to be non-intuitive to a student, and may be intimidating because of its mathematical complexity.
- Its mathematical form fails to build on what was said previously about comparing the mean and median, which in effect makes it seem to be a new topic entirely.
- The formula only works if we have raw data $x_1, x_2, \ldots, x_n$. What if we only have summarized data (mean, median, standard deviation) as in many textbook problems?

Older mathematical statistics textbooks (e.g., Yule and Kendall 1950; Kenney and Keeping 1954; Clark and Schkade 1974) refer to skewness measures that directly compare either the mean and mode, or the mean and median. For empirical calculations, Yule and Kendall (1950, p. 161) recommend using a statistic that compares the mean ($\bar{x}$) and median (m):

---

[3] If we are mainly interested in testing for non-normality, there is a simple measure that can be computed from summarized data: the studentized range $(x_{max}-x_{min})/s$. This test is attractive because it does not require raw data (Tracy and Doane, 2005) and has good power.

$$[2] \qquad Sk_2 = 3\frac{(\bar{x} - m)}{s} \quad .$$

The attraction of this statistic (henceforward the *Pearson 2 skewness coefficient*) is that it is consistent with the intuitive approaches developed earlier. You can see its sign at a glance. It shows how many standard deviations apart the two measures of center are. Hotelling and Solomons (1932) first showed that the statistic $(\bar{x} - m)/s$ will lie between -1 and $+1$, so Sk_2 will lie between -3 and $+3$ (although, in practice, it rarely approaches these limits.)

This statistic is no longer seen in textbooks (based on our review of over 30 popular business statistics textbooks with recent copyrights) although it does show up in some Web searches (e.g., http://mathworld.wolfram.com). We can find no tables of critical values. Arnold and Groeneveld (1995) note that Sk_2 has some desirable properties (it is zero for symmetric distributions, it is unaffected by scale shift, and it reveals either left- or right-skewness equally well). But two issues must be examined before suggesting a revival of this intuitively attractive statistic: (1) we need a table of critical values for Sk_2, and (2) we should compare the power of Sk_2 and G_1.

4. Type I Error Simulation

We obtained preliminary critical values for Sk_2 using Monte Carlo simulation with Minitab 16. We drew 20,000 samples of $N(0,1)$ for $n = 10$ to 100 in increments of 10 and computed the sample mean ($\bar{x}$), sample median (m), and sample standard deviation (s). For each sample size, we calculated Sk_2 and its percentiles. Upper and lower percentiles should be the same except for sign, so we averaged their absolute values (effectively 40,000 samples). Table 3 shows the 5% and 10% critical values from our simulation.

Table 3. Monte Carlo estimates of percentiles for Pearson 2 skewness coefficient Sk_2.

Percentile*					Sample Size					
	10	20	30	40	50	60	70	80	90	100
5% Upper*	0.9629	0.7617	0.6433	0.5542	0.5062	0.4632	0.4371	0.4069	0.3851	0.3669
10% Upper*	0.7682	0.5967	0.5058	0.4374	0.3915	0.3605	0.3399	0.3167	0.3005	0.2865

Average of upper and lower absolute percentiles using 20,000 samples from N(0,1) using Minitab 16.

Figure 5 shows the upper percentiles for sample sizes 10 to 100. Sampling variation exists, although the overall pattern is quite stable.
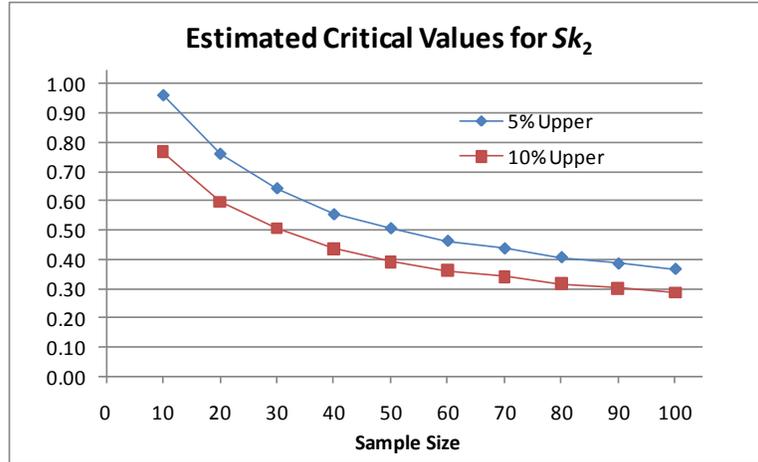
Figure 5. Monte Carlo estimates of percentiles for Pearson 2 skewness coefficient.

To complete the analogy with G_1, Table 4 shows a 90 percent range that makes it easier for students to interpret (cf. Table 1). A footnote is placed in the table to make sure that it is interpreted correctly.

Table 4. 90% expected range for Pearson 2 skewness coefficient Sk_2.

n	Lower Limit	Upper Limit	n	Lower Limit	Upper Limit
10	−0.963	+0.963	60	−0.463	+0.463
20	−0.762	+0.762	70	−0.437	+0.437
30	−0.643	+0.643	80	−0.407	+0.407
40	−0.554	+0.554	90	−0.385	+0.385
50	−0.506	+0.506	100	−0.367	+0.367

Note: If your sample is from a normal population, the skewness coefficient Sk_2 would fall within the stated range 90 percent of the time. Values of Sk_2 outside this range suggest non-normal skewness.

The table for Sk_2 can be turned into a decision diagram that students might find easier to understand, as in Figure 6. We only display sample sizes up to 100 because the diagram narrows sharply and labeling becomes difficult (also because small samples are more common in the classroom). A student can see that inferences for small samples are risky.
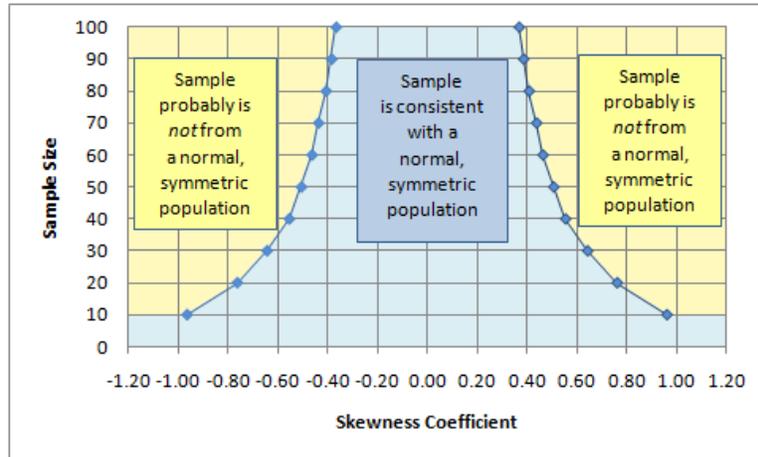
Figure 6. 90% expected range for Pearson 2 skewness coefficient Sk_2.

How do the simulated values of Sk_2 compare with the simulated values of G_1 for the same samples? We used the R language with the CRAN library C1071 to simulate both Sk_2 and G_1 for 50,000 samples of sizes up to $n = 200$, with the results shown in Figure 7. Their similarity shows that Sk_2 is measuring the same thing as G_1. In larger samples, the measures are almost identical. The same stable pattern exists for other percentiles (not shown for simplicity).
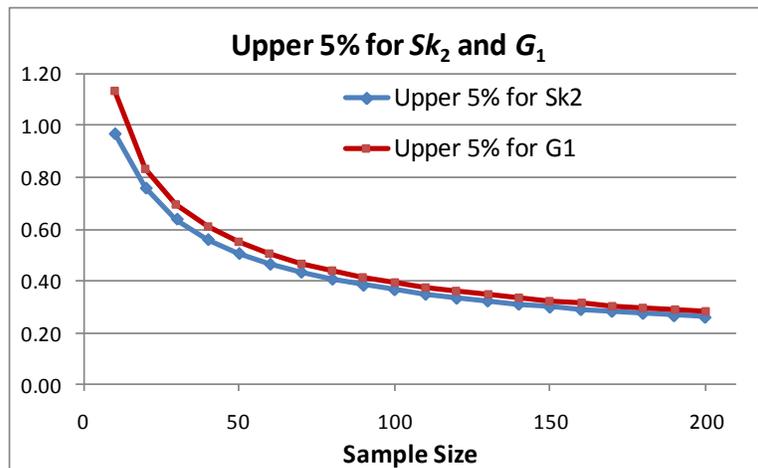

Figure 7. Hartley-Pearson 5% critical values of G_1 and Monte Carlo 5% estimates for Sk_2.

Over many samples (as in our simulation) the tests will agree *on average*. But if G_1 and Sk_2 are so similar, isn't it a tossup which we use? No, because Sk_2 has more variability than G_1. By definition, Sk_2 depends not only on the estimates $\bar{x}$ and s, but also on the sample median m. With more sources of variation, we would therefore expect Sk_2 to have lower power than G_1. We will now show that this is, indeed, the case.

5. Type II Error Simulation

Type II error in this context occurs when a sample from a non-normal skewed distribution does not lead to rejection of the hypothesis of a symmetric normal distribution. There are an infinite number of distributions that could be explored, including "real world" mixtures that do not resemble any single theoretical model. Just to get some idea of the comparative power of Sk_2 and G_1, we will illustrate using samples from two non-normal, unimodal distributions

- a mildly skewed distribution: $\chi^2(5)$, and
- a highly skewed distribution: $\chi^2(2)$.

These two populations are illustrated in Figure 8, along with the normal distribution. The scales do not matter, because our two skewness statistics Sk_2 and G_1 are unit-free measures.
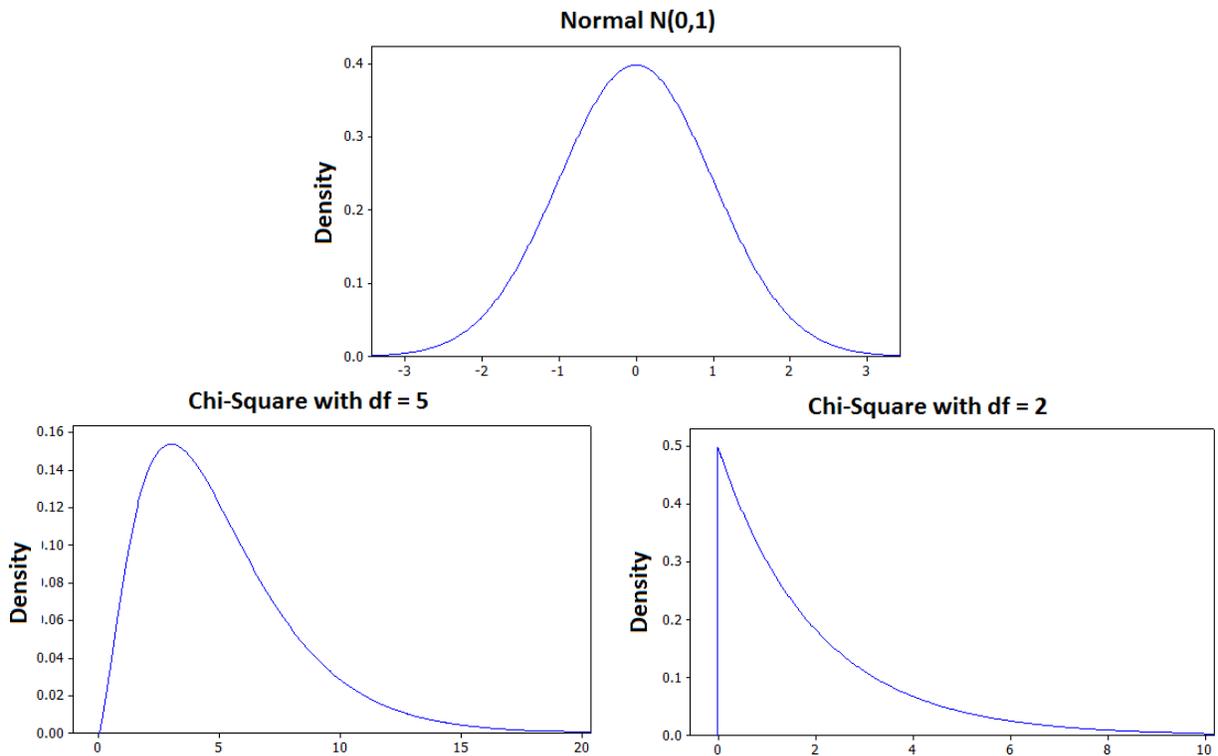

Figure 8. Three populations used in simulations.

For samples from $\chi^2(5)$ (the middle figure) we would anticipate low power for any test of skewness, because the population is only mildly skewed. As illustrated in Figure 9, even a fairly large sample of 50 from $\chi^2(5)$ can produce a histogram that might pass a visual test for normality. However, in samples from $\chi^2(2)$ (an exponential distribution) we would almost always expect rejection of normal skewness (the third figure).
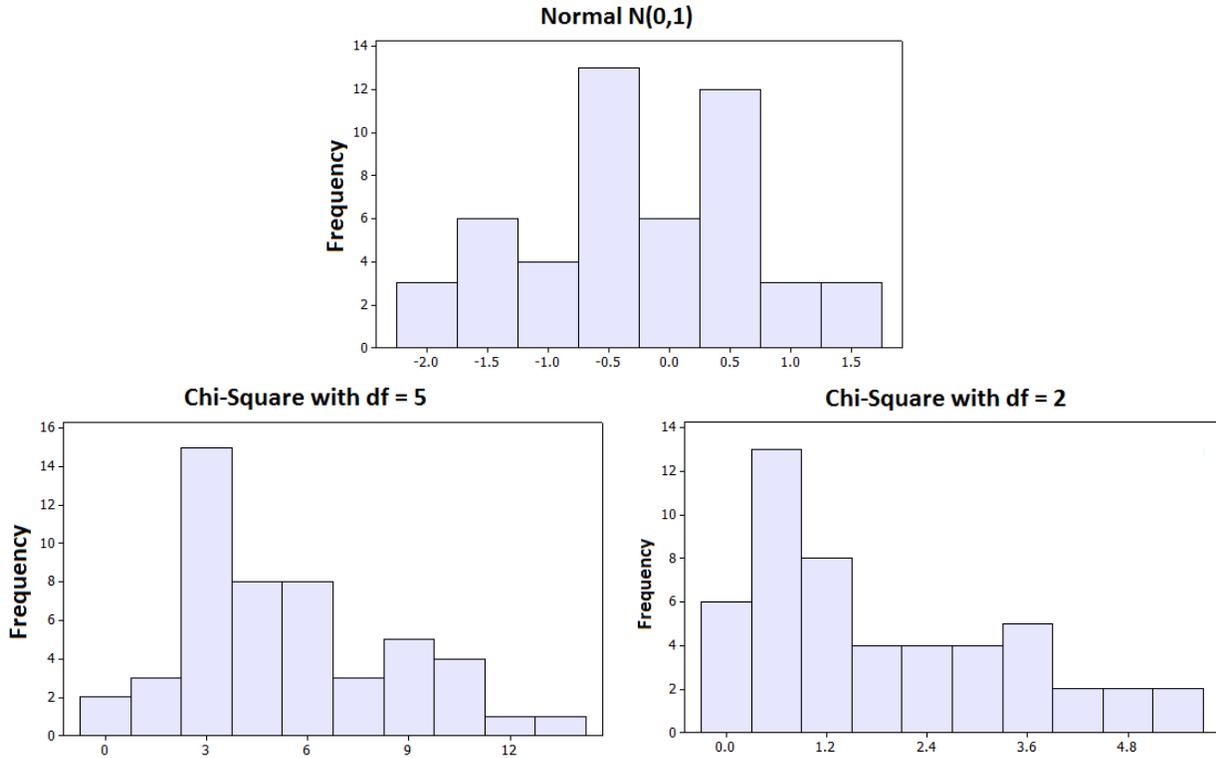
Figure 9. Sample histograms ($n = 50$) from three populations.

Following the spirit of the Type I error simulations, we drew 10,000 samples of each sample size from each non-normal population. For each sample, we calculated both Sk_2 and G_1. Using the previously-calculated percentiles for Sk_2 and the known Hartley-Pearson percentiles for G_1, we counted the number of samples that would lead to rejection of the hypothesis of a symmetric normal population. We then computed the empirical power of each statistic. These results are shown in Table 5. Although our power simulations are not directly comparable to those by Tabor (2010), we did confirm Tabor's 0.64 power for Sk_2 under his experimental setup with $n = 10$ from $\chi^2(1)$ at $\alpha = .05$.

Table 5. Monte Carlo power for Sk_2 and G_1 for two non-normal right-skewed populations.

					Sample Size					
Mildly skewed population: $\chi^2(5)$	10	20	30	40	50	60	70	80	90	100
Sk_2 rejections @ α = .10	2,002	3,327	4,433	5,565	6,395	7,010	7,575	8,047	8,484	8,785
G_1 rejections @ α = .10 (adj)	NA	NA	7,263	8,415	9,119	9,506	9,753	9,874	9,945	9,970
G_1 rejections @ α = .10 (unadj)	NA	NA	7,357	8,457	9,137	9,520	9,759	9,874	9,945	9,970
Strongly skewed population: $\chi^2(2)$										
Sk_2 rejections @ α = .10	3,917	6,331	7,847	8,808	9,398	9,674	9,797	9,902	9,948	9,976
G_1 rejections @ α = .10 (adj)	NA	NA	9,460	9,849	9,969	9,994	9,988	10,000	10,000	10,000
G_1 rejections @ α = .10 (unadj)	NA	NA	9,497	9,856	9,970	9,994	9,989	10,000	10,000	10,000

Based on 10,000 samples using Minitab Version 16. Tests are two-tailed using α = .10, i.e., reject normal (symmetric) distribution if the test statistic exceeds the .05 critical value in either tail.

The adjustment in the Hartley-Pearson critical values makes little difference for the sample sizes we are considering, so Figure 10 only compares power for Sk_2 with power for G_1. In samples from $\chi^2(5)$, neither test performs well in small samples. However, G_1 is clearly superior, its power quickly approaching 1.00, while Sk_2 barely exceeds .80 for the largest sample size shown. In samples from $\chi^2(2)$, both tests perform well beyond n = 50. However, Type II error for G_1 approaches zero for n = 70 or greater while Sk_2 approaches power of 1.00 more slowly.
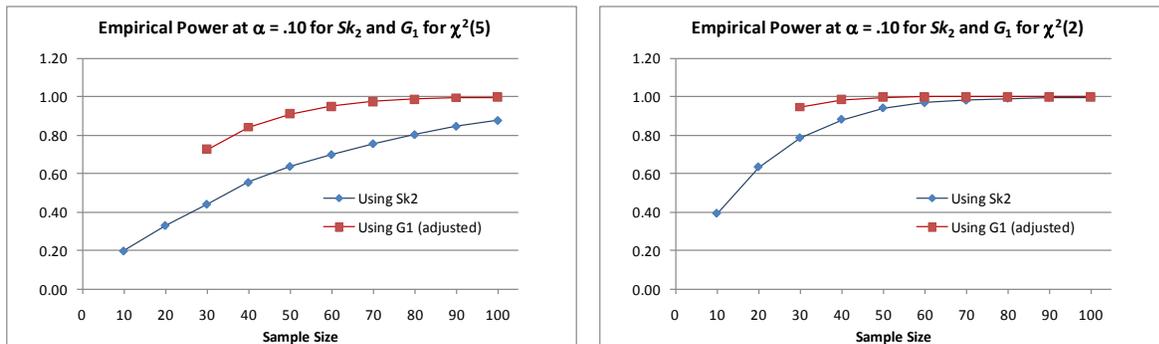

Figure 10. Empirical power based on 10,000 samples from skewed populations.

6. Summary and Conclusions

Visual displays (e.g., histograms) provide easily understood impressions of skewness, as do comparisons of the sample mean and median. However, students tend to take too literal a view of these comparisons, without considering the effects of binning or the role of sample size. The moment coefficient statistic G_1 is widely available, but is not easily interpreted and its tables are not available in textbooks. In contrast, the Pearson 2 skewness statistic Sk_2 has strong pedagogical appeal because it corresponds to the way we like to talk about skewness. It is easy to

calculate and interpret as long as we have just three statistics (the sample mean, median, and standard deviation). Further, Sk_2 is the *only* way to measure skewness when we do not have the original sample data $x_1, x_2, \ldots, x_n$.

We can create tables of critical values for Sk_2 using Monte Carlo simulation to control Type I error at any desired level. With such tables, interpreting Sk_2 is as easy as interpreting G_1. The critical values of Sk_2 and G_1 lead to the same conclusion on average. The weakness of Sk_2 is that it lacks power. Perhaps this is why Sk_2 has fallen out of favor in textbooks, although this argument against Sk_2 has not been forcefully articulated in the literature on teaching statistics. Since many introductory textbooks do not mention these formulas or tables at all, it is hard to argue that G_1 is a well-established benchmark in introductory textbooks, despite its undisputed primacy in mathematical statistics. The tradeoff of lower power against increased comprehension and ease of calculation may be worthwhile for classroom teachers. Regardless which statistic (if any) we use to assess skewness, students should understand that a large skewness statistic casts doubt on the *normality* of the population, and is not merely a test for skewness.

References

Arnold, B. C. and Groeneveld, R. A. (1995), "Measuring Skewness with Respect to the Mode," *The American Statistician*, 49, 34-38.

Balanda, K. P. and MacGillivray, H. L. (1988), "Kurtosis: A Critical Review", *The American Statistician*, 42, 111-119.

Clark, C. T. and Schkade, L. L. (1974), *Statistical Analysis for Administrative Decisions*, 2nd ed., South-Western Publishing Co., 42.

Cobb, G. W. and Moore, D. S. (1997), "Mathematics, Statistics, and Teaching," *The American Mathematical Monthly*, 14, 801-823.

D'Agostino, R. B. and Stephens, M. A. (1986), *Goodness of Fit Techniques*, Marcel Dekker, Inc., 11-12 and 24-33.

delMas, R., Garfield, J., Ooms, A., and Chance, B. (2007), "Assessing Students' Conceptual Understanding After a First Course in Statistics," *Statistics Education Research Journal*, 6, 28-58. http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf

Doane, D. P. (2004), "Using Simulation to Teach Distributions, *Journal of Statistics Education*, 12, 1-21. www.amstat.org/publications/jse/v12n1/doane.html

Doane, D.P. and Seward, L.E. (2011), *Applied Statistics in Business and Economics*, 3rd ed., McGraw-Hill/Irwin,154-156.

Doane, D. P. and Tracy, R. L. (2000), "Using Beam and Fulcrum Displays to Explore Data," *The American Statistician*, 54, 289-290.

Groeneveld, R. A. and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 33, 391-399.

Horswell, R. L. and Looney, S. W. (1993), "Diagnostic Limitations of Skewness Coefficients in Assessing Departures from Univariate and Multivariate Normality," *Communications in Statistics: Simulation and Computation*, 22, 437-459.

Hotelling, H. and Solomons, L. M. (1932), "The Limits of a Measure of Skewness," *The Annals of Mathematical Statistics*, 3, 141-142.

Kenney, J. F. and Keeping, E. S. (1954), *Mathematics of Statistics, Part One*, 3rd Edition, D. Van Nostrand and Company, Inc., 99-103.

Joanes, D. N. and Gill, C. A. (1998), "Comparing Measures of Sample Skewness and Kurtosis," *The Statistician*, 47, Part 1, pp. 183-189.

MacGillivray, H.L. (1986), "Skewness and Asymmetry: Measures and Orderings," *The Annals of Statistics*, 14, 994-1011.

Pearson, K. (1895), "Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material," *Transactions of the Royal Philosophical Society, Series A*, 186, 343-414.

Pearson, E. S. and Hartley, H. O. (1970), *Biometrika Tables for Statisticians*, 3rd Edition, Cambridge University Press, 207.

Rayner, J. C. W., Best, D. J., and Matthews, K. L. (1995), "Interpreting the Skewness Coefficient," *Communications in Statistics – Theory and Methods*, 24, 593-600.

Tabor, J. (2010), "Investigating the Investigative Task: Testing for Skewness - An Investigation of Different Test Statistics and their Power to Detect Skewness," *Journal of Statistics Education*, 18, 1-13. www.amstat.org/publications/jse/v18n2/tabor.pdf

Tracy, R. L. and Doane, D. P. (2005), "Using the Studentized Range to Assess Kurtosis," *Journal of Applied Statistics*, 32, 271-280.

Yule, G. U. and Kendall, M. G. (1950), *An Introduction to the Theory of Statistics*, 3rd edition, Harper Publishing Company, 162-163.

David P. Doane
Department of Decision and Information Sciences
Oakland University
Rochester, MI 48309-4493
Voice: 248.370.4002

Fax: 248.370.4275
mailto:Doane@oakland.edu

Lori E. Seward
Leeds School of Business
University of Colorado
Boulder, CO 80309-0419
Voice: 303.492.3866
mailto:Lori.Seward@colorado.edu