Mean, Median, and Skew: Correcting a Textbook Rule

Paul T. von Hippel

Mean, Median, and Skew: Correcting a Textbook Rule

Paul T. von Hippel
The Ohio State University

Abstract

Many textbooks teach a rule of thumb stating that the mean is right of the median under right skew, and left of the median under left skew. This rule fails with surprising frequency. It can fail in multimodal distributions, or in distributions where one tail is long but the other is heavy. Most commonly, though, the rule fails in discrete distributions where the areas to the left and right of the median are not equal. Such distributions not only contradict the textbook relationship between mean, median, and skew, they also contradict the textbook interpretation of the median. We discuss ways to correct ideas about mean, median, and skew, while enhancing the desired intuition.

1. A Rule of Thumb

Among the eighteen introductions to data analysis that I have examined, fourteen give a rule of thumb relating skew to the positions of the median and mean.

> "In a skewed distribution, the mean is farther out in the long tail than is the median." (Moore and McCabe 2003, p. 43)

> "For skewed distributions, the mean lies toward the direction of skew (the longer tail) relative to the median." (Agresti and Finlay 1997, p. 50)

Five textbooks extend the rule to cover the mode as well.

> "In a skewed distribution..., the mean is pulled in the direction of the extreme scores or tail (same as the direction of the skew), and the median is between the

mean and the mode." ([Thorne and Giessen 2000, pp. 81-82](#))

"[T]he mode, median, and mean do not coincide in skewed distributions, *although their relative positions remain constant* - moving away from the `peak' and toward the `tail,' the order is always from mode, to median, to mean." ([Levin and Fox, 2003, p. 85](#); also [Levin and Fox 2004, p. 56](#))

The relationship between skew and measures of center is often illustrated with an idealized graph like [Figure 1](#).
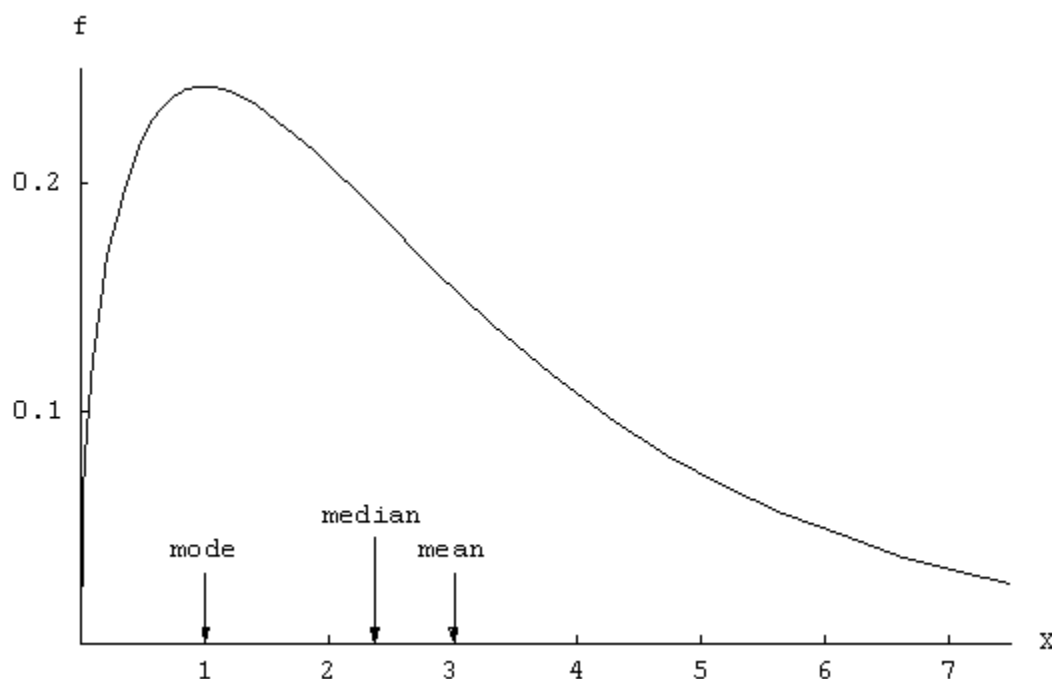

Figure 1. Classic illustration of the relationship between skew, mean, median, and mode. The skew is to the right, the mean is right of the median, and the median is right of the mode. The density shown is the chi-square with 3 degrees of freedom.

Authors typically state this rule without qualification, and some, like Levin and Fox above, indicate that it "always" holds. In follow-up exercises, some authors ask in what direction the mean or skew would "usually" or "probably" lie, but almost no author indicates what unusual or improbable circumstances might change the picture. ([Ritchey 2000](#) mentions bimodal distributions, but does not elaborate.)

In this paper, we demonstrate that violations are not at all unusual if the distribution is discrete. Continuous densities seem much better behaved, though continuous violations can also be found or constructed. We discuss the reasons for these violations, and propose ways that teachers can allow for violations while continuing to develop students' basic intuition.

2. Breaking the Rule

It is helpful to look at illustrative violations. We begin with empirical violations appropriate for a basic, algebra-based course in data analysis. We then continue with theoretical violations at the level of a calculus-based course in mathematical statistics.

2.1. Empirical violations

In a data analysis course, skew is often defined informally in terms of tail length or extreme values. If a numeric value is required, it is usually calculated using the third-moment formulas favored by data-analysis software – e.g. $\frac{1}{n}\sum\left(\frac{(x_i - \bar{x})}{s}\right)^3$ or $\frac{n}{(n-1)(n-2)}\sum\left(\frac{(x_i - \bar{x})}{s}\right)^3$ where n is the sample size, $\bar{x}$ is the sample mean and s is the sample standard deviation.

Under these definitions, discrete distributions can easily break the rule. For example, in the General Social Survey, respondents are asked how many people older than 18 live in their household. Figure 2 gives the responses for 2002 (1996 was similar). The skew is clearly to the right, yet the mean is left of the median and mode.
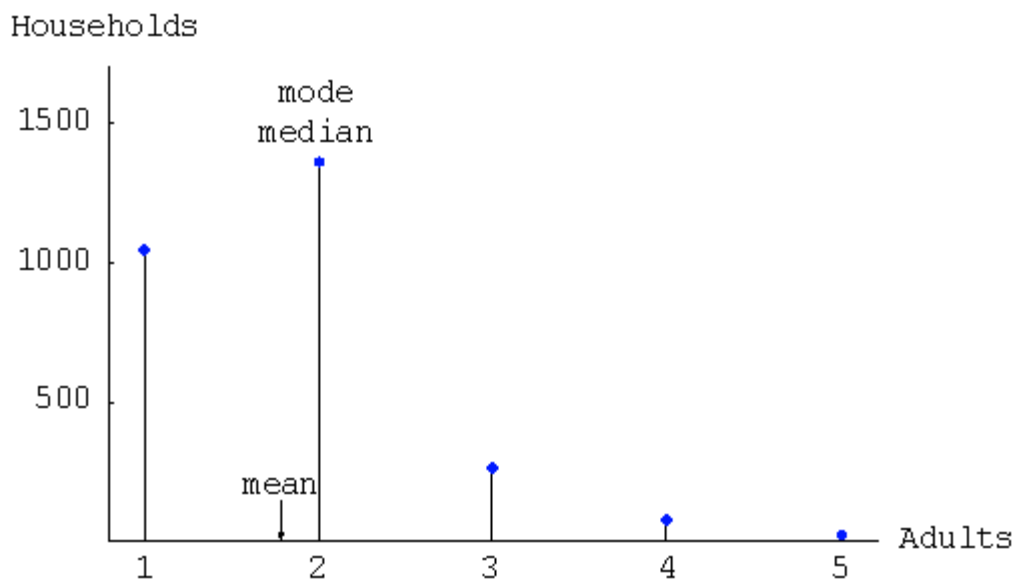

Figure 2. Distribution of adult residents across US households. The skew is to the right (1.11), yet the mean is left of the median and mode.

The key feature of Figure 2 is that there are substantially more cases on one side of the median than on the other. This is typical of discrete violations. In Figure 2, 38% of the cases are left of the median, 49% coincide with the median, and 13% are right of the median. The mean, or center of gravity, sits in the heavier left tail, but the longer right tail determines the skew. The rightmost values affect the skew more than the mean, because extreme values are cubed in the skew formula.

Continuous variables are less likely to break the rule, because the median of a continuous density must

divide the area in half. But continuous violations can occur. For example, the Early Childhood Longitudinal Study (Kindergarten cohort) administered reading tests to 14,455 first graders in the spring of 2000. The distribution of scores is given in Figure 3. The skew is slightly to the left, yet the mean is just right of the median, and the median is right of the primary mode.
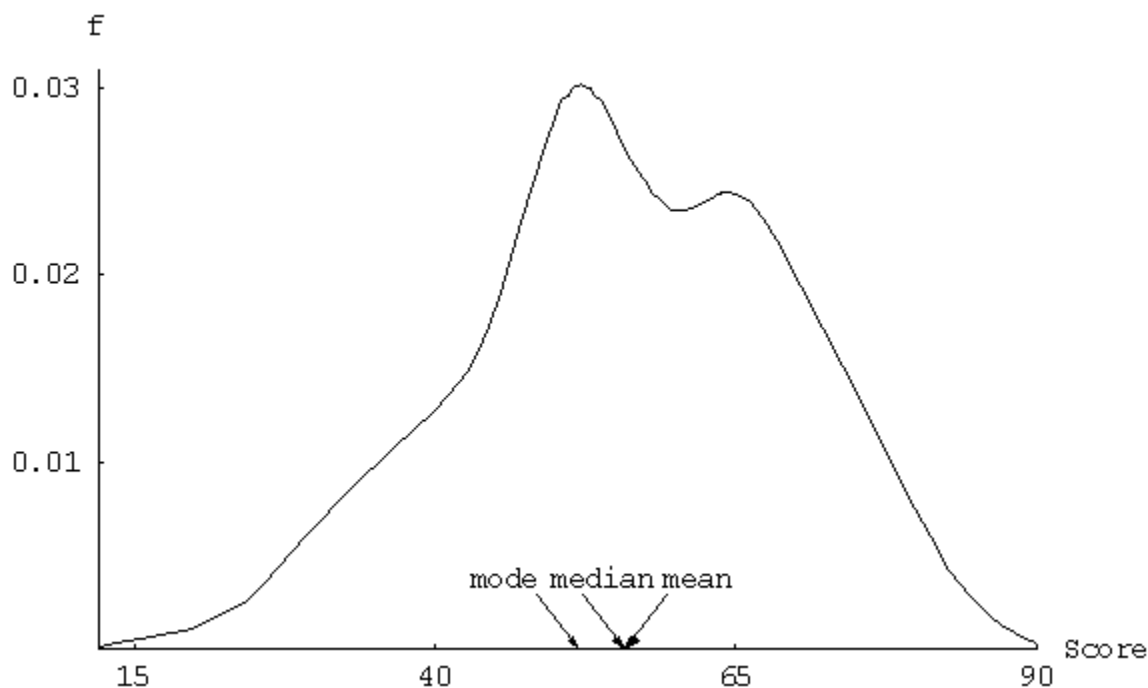

Figure 3. Spring 2000 reading scores from the Early Childhood Longitudinal Study (Kindergarten cohort). The skew is slightly to the left (-0.22), yet the mean is just right of the median, and the median is right of the primary mode. (The density was estimated using a Epachenikov kernel and a Silverman bandwidth (Silverman 1986).)

The continuous violation in Figure 3 is milder than the discrete violation in Figure 2. But in one respect the violations are similar: both figures have greater area in one tail, but greater length in the other. In Figure 3, the long tail is to the left of the primary mode, and the heavy tail is to the right. In addition, Figure 3 is slightly bimodal; we will discuss bimodal and multimodal distributions in Section 2.2.

The question arises whether better results could be obtained using an alternative definition of skew. An obvious attempt is the old "Pearson" formula $3\dfrac{(\bar{x} - m)}{s}$ where m is the median (e.g., Knoke, Bornstedt, and Mee 2002, p. 53). The Pearson formula makes a tautology of the relationship between skew, median, and mean – but it also has the counterintuitive implication that Figure 2, despite its long right tail, has negative skew.

2.2. Theoretical distributions

In mathematical statistics, the skew is typically defined as the third standardized moment

$$E\left(\frac{(X-\mu)}{\sigma}\right)^3$$, where μ is the mean and σ is the standard deviation.

Under this definition, well-known discrete distributions often put the median on the "wrong" side of the mean. Figure 4 plots the mean, median, mode, and skew of the Poisson distribution as a function of the parameter μ (which is also the mean). All Poisson distributions have an infinite right tail and positive skew (equal to $\mu^{-\frac{1}{2}}$) – yet for more than 30% of parameter values, for example $\mu = .75$ (Figure 5), the mean is less than the median. Since the Poisson is the limiting distribution for the binomial and hypergeometric, it follows that those distributions can break the rule as well. Again, the main reason is that, in discrete distributions, the median can divide the distribution into unequal areas. In Figure 5, for example, 47% of the distribution is left of the median, but only 17% is right of the median; the remaining 35% coincides with the median.

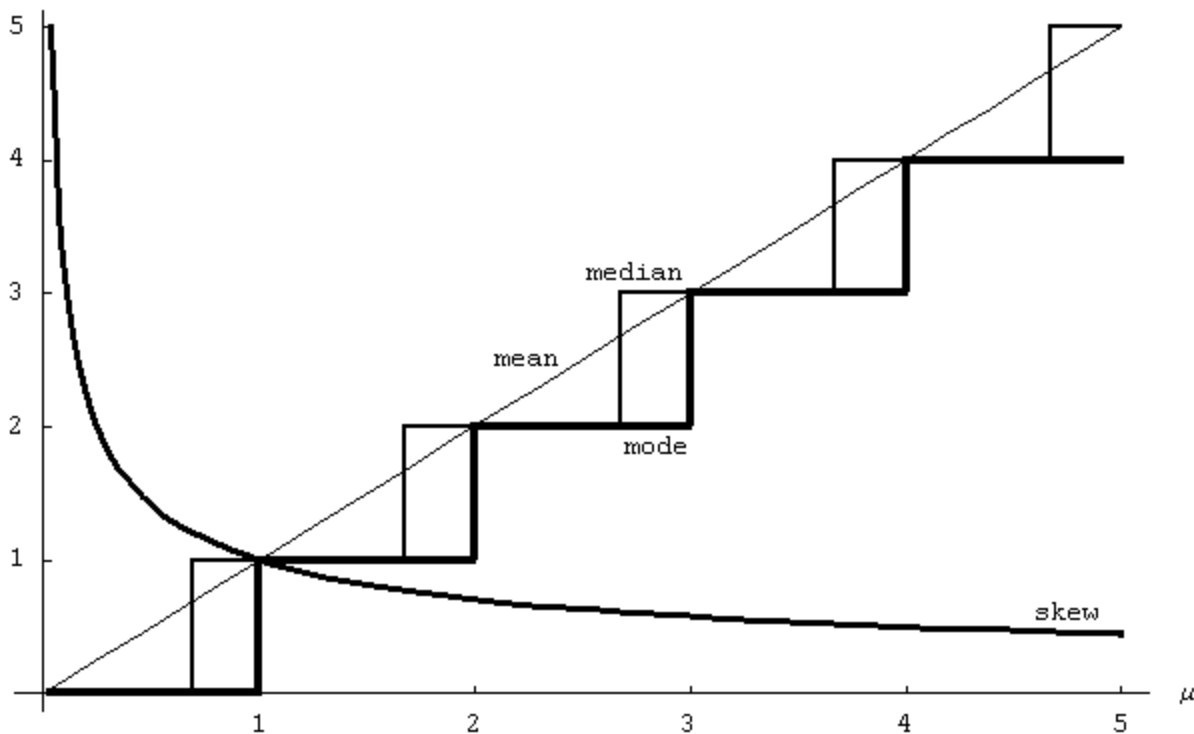Figure 4. The mean, median, mode, and skew of the Poisson distribution, plotted as a function of the parameter μ (μ is also the mean). Although the skew is consistently positive, the mean is less than the median whenever $\mu \bmod 1 > ln(2)$.
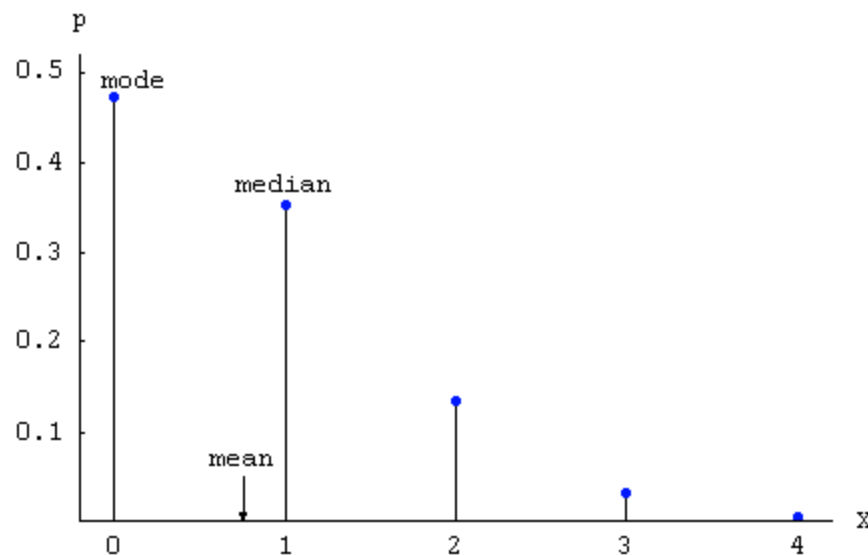
Figure 5. The Poisson distribution with $\mu = 0.75$. The skew is to the right, yet the mean is left of the median.

Continuous violations are rarer, but do exist. Multimodal continuous densities, for example, can easily break the rule. If the modes are narrow enough, a multimodal density approximates a discrete distribution, and we have already seen that discrete violations are commonplace. To construct a multimodal violation, simply take a discrete violation (e.g., Figure 2 or Figure 5) and add random normal "noise" to each value of X. The noise makes the distribution continuous, but if the noise variance is small there will be little change to the mean, median, mode, or skew. A density constructed in this way can be severely multimodal; such craggy densities are unusual, but not unheard of. The emission spectrum of hydrogen is severely multimodal (Dyson and Willhams 1997), and craggy densities approximate the small-N sampling distributions of many sample statistics (e.g., Cytel 2004). Extreme cragginess is not required to exchange the positions of median and mean; Figure 3, for example, is only mildly bimodal.

It is also worth noting that a multimodal density can put the mode simply anywhere in relation to the median and mean. To see this, in Figure 1 add a tall spike of density on the right, at say $X = 4$. If the spike is tall enough, it becomes the primary mode, but if the spike is narrow enough it leaves the mean, median and skew substantially unchanged. The result is a right-skewed density where the primary mode is right of the median and mean. This sounds artificial, but a similar method could be used to construct the empirical violation in Figure 3; start with a left-skewed density with a single mode at $X = 64$, then add a taller lump near $X = 52$. The result is a left-skewed density where the primary mode is left of the median and mean. Using a similar method, Dudewicz and Mishra (1988, p. 217) construct a right-skewed density where the primary mode is between the median and mean.

Unimodal continuous densities are more cooperative. Groeneveld and Meeden (1977) prove that the skew gives the relative positions of mean, median and mode for the F, beta, and gamma densities (the gamma includes the exponential and the chi-square). More generally, MacGillivray (1981) proves the relationship for a large class of continuous unimodal densities including the entire Pearson family.

Outside the Pearson family, however, the rule can fail. For example, Groeneveld (1986) points out

violations in the Weibull density with shape parameter β. Figure 6 plots the mean, median, mode, and skew of the Weibull density for β in the interval (3.20, 3.60). Although the skew is consistently positive, the mean can be on either side of the median, and the median or mean can be on either side of the mode. Figure 7 plots the Weibull density with $\beta = 3.44$; the skew is to the right, but the mean is left of the median, and the median is left of the mode. This violation is quite mild, however; the skew is nearly invisible, and the mean, median, and mode differ hardly at all.

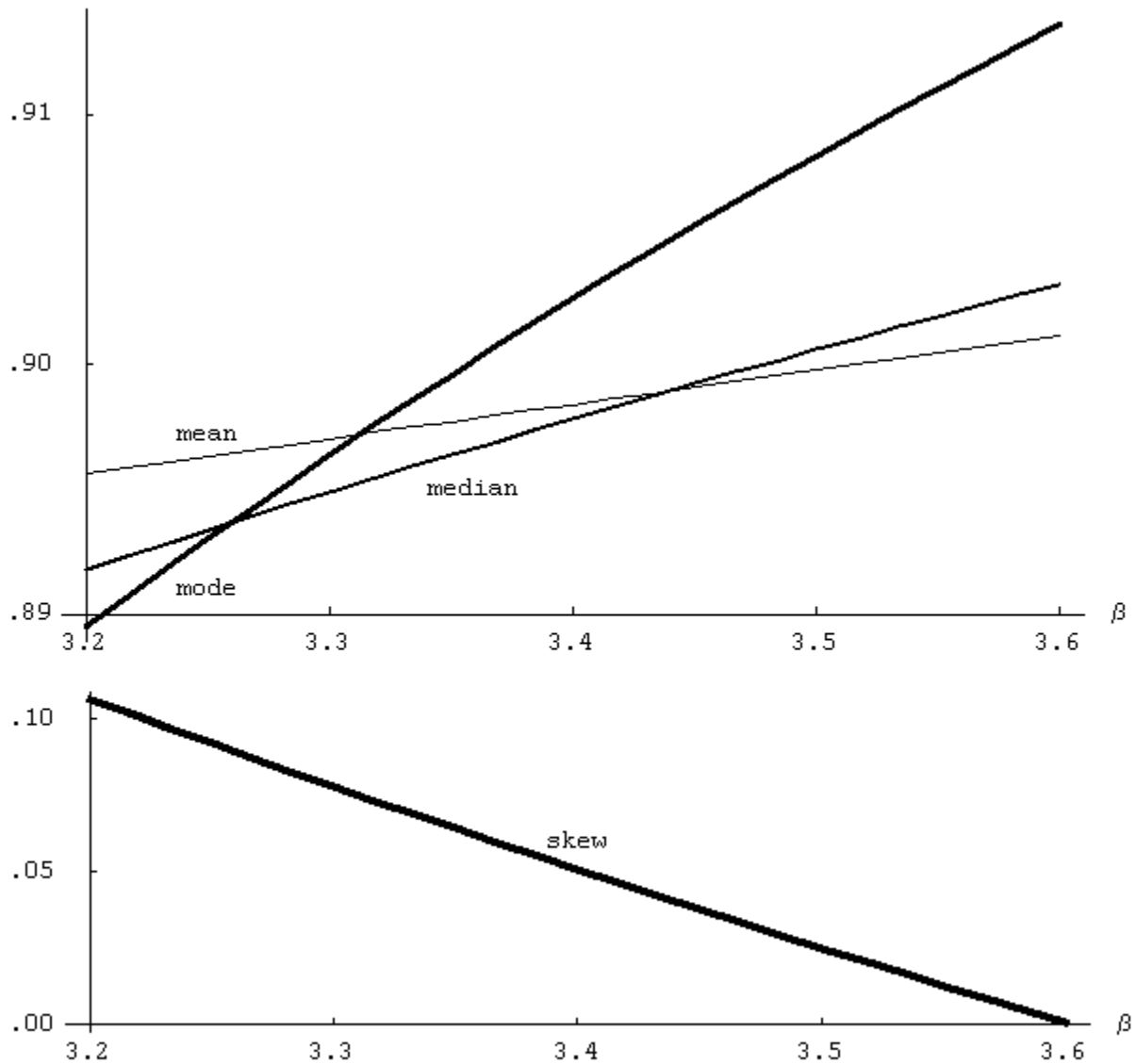
Figure 6. The mean, median, mode, and skew of a Weibull distribution with shape parameter β. For β < 3.60 the skew is positive, yet for β > 3.26 the median is less than the mode, for β > 3.31 the mean is less than the mode, and for β > 3.44 the mean is less than the median. (Adapted from Groeneveld 1986.)
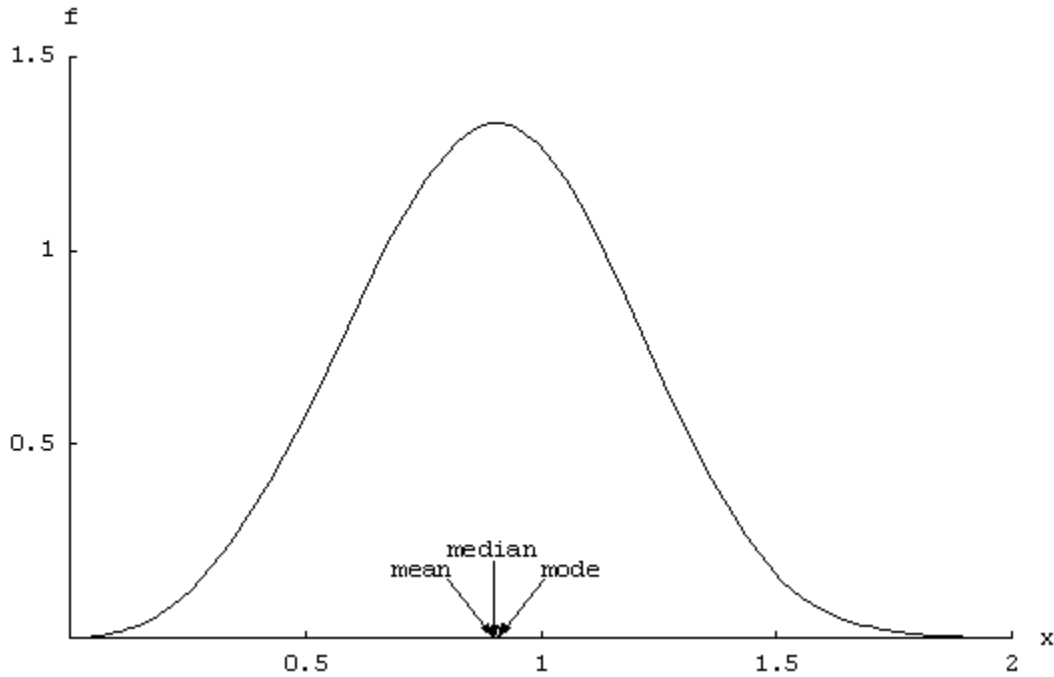
Figure 7. A Weibull density with shape parameter $\beta = 3.44$. The skew is slightly to the right (0.04), but the mean is just left of the median, and the median is just left of the mode.

A stronger though more contrived violation arises from juxtaposing the triangular and exponential densities. Generalizing from examples in <u>Dharmadhikari and Joag-Dev (1988)</u>, let f be a continuous density function that is triangular to the left of the origin and exponential to the right:

$$f(X) = \begin{cases} 0 & \text{if } X < \dfrac{-2p}{1-p} \\[2ex] (1-p)\left(1 + \dfrac{1-p}{2p}X\right) & \text{if } \dfrac{-2p}{1-p} \leq X \leq 0 \\[2ex] (1-p)e^{-X} & \text{if } X \geq 0 \end{cases}$$

The parameter p in the interval (0, 1) determines what proportion of the area is in the triangular region. <u>Figure 8</u> plots the mean, median, mode, and skew as functions of p. For $p < 0.755$, the skew is positive, yet the mean can be on either side of the median, and the mean or median can be on either side of the mode. <u>Figure 9</u> plots this density with $p = 0.75$; the skew is to the right, yet the mean is left of the median, and the median is left of the mode.
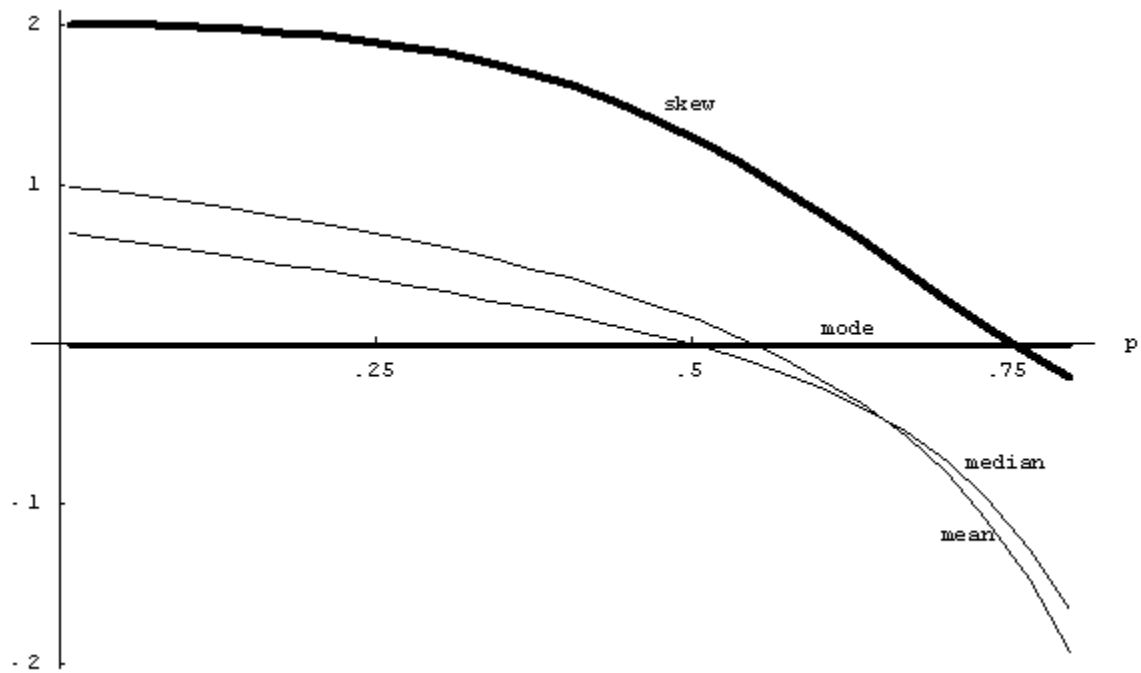
Figure 8. The mean, median, mode, and skew of a left-triangular, right-exponential continuous density with its mode at the origin. The parameter p determines what proportion of the area is in the triangular region. For $p < 0.755$ the skew is positive, yet for $p > 0.5$ the median is less than the mode, for $p > 0.55$ the mean is less than the mode, and for $p > 0.61$ the mean is less than the median.
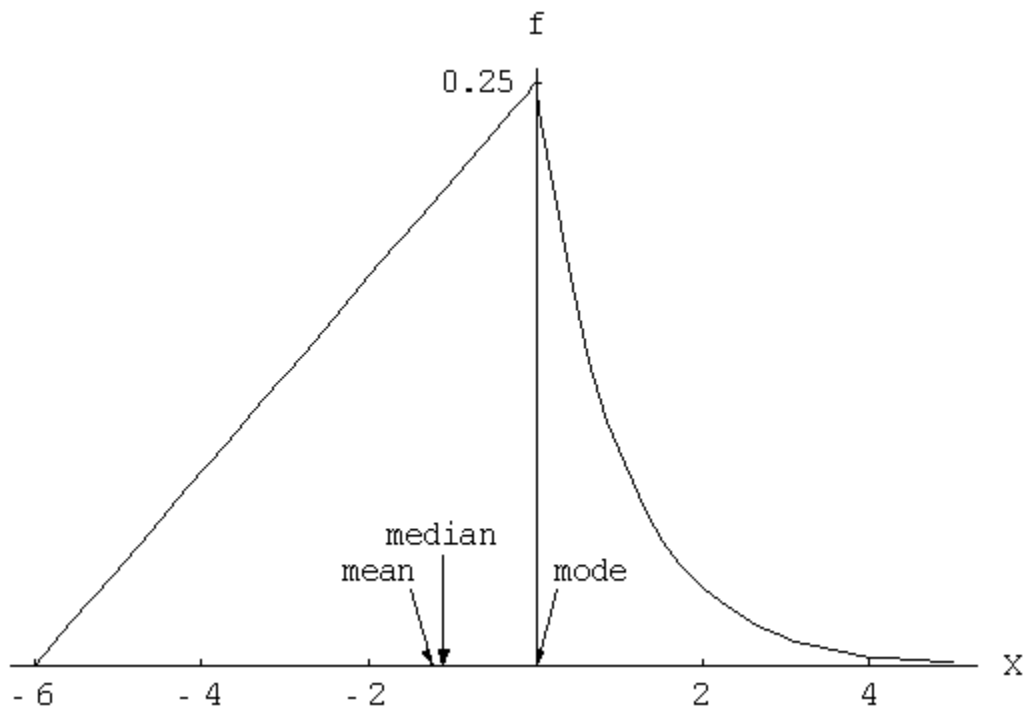

Figure 9. A 75% triangular, 25% exponential density. The skew is slightly to the right (0.023), but the mean is left of the median, and the median is left of the mode.

Figure 9 follows the pattern of Figure 2, Figure 3 and Figure 7, with greater area to one side of the mode, but greater length to the other. In Figure 9, the left side of the mode has greater area, but the right side is infinitely long.

Again, the question arises whether the rule could be guaranteed by an alternative definition of skew. The answer is yes; a small theoretical literature has developed a suitable definition based on comparing the weights of the left and and right tails at all possible distances from the median (Ageel 2000; Dharmadhikari and Joag-Dev 1983; Zwet 1979). This definition, however, does not always square with our intuitive sense of skew; it implies, for example, than none of the counterexamples in this paper has skew at all.

3. What to Teach?

We have shown that a widely taught rule of thumb has a surprising number of exceptions. In a skewed distribution, it is quite possible for the median to be further out in the long tail than the mean. This configuration is common for discrete variables, especially when the areas to the left and right of the median are not equal. Exceptions are rarer for continuous variables, but can still occur if the density is bimodal or multimodal, or if one tail is long but the other is heavy.

Notwithstanding these exceptions, the relationship between skew, median, and mean conveys useful intuition. It seems desirable to preserve or enhance this intuition, without giving students an inaccurate picture.

In a data analysis course, it is certainly possible to continue teaching the relationship between skew, median, and mean. The treatment, however, should be more qualified than it is in current textbooks.

1. First, the relationship should be introduced using clearly continuous, clearly unimodal densities. While most textbooks already begin with such densities, those that don't should be revised (e.g., Thorne and Giessen 2000, Figure 9-5; Freund 2004, Figure 2.4).
2. Next, it should be pointed out that the rule is imperfect, and that the most common exceptions occur when the variable is discrete.

Discrete violations provide a nice opportunity to refine students' interpretation of the median. Most textbooks teach that half the area falls on each side of the median, but this is far from true in Figure 2 and Figure 5. In discrete distributions, significant area can coincide with the median, so that the areas to each side can be unequal and substantially less than one-half. Continuous densities lack this possibility, so their violations tend to be rarer and milder.

The distinction between discrete and continuous variables is useful here, but it can be hard to draw in practice. An inherently continuous variable can be made discrete if the recorded values are rounded. Conversely, a Poisson distribution with (say) $\mu = 10.75$ is "nearly continuous," yet despite mild right skew the mean is left of the median (see Figure 4). The convergence between discrete and continuous distributions is well worth discussing in an introductory course.

A similar approach could be taken in a mathematical statistics course. Because the relationship between skew and center is just a rule of thumb, it can be taught rather informally. Teachers with an affection for the topic may ask students to demonstrate the rule using, say, the F density, or demonstrate its violation using the Poisson or Weibull.

The basic idea is that extreme values influence all distributional moments; a few large values increase the first moment (mean), the second moment (variance), and the third moment (skew) ([Groeneveld 1991](#)). The third moment will be most affected since the extreme values are cubed. From this perspective, the relationship between skew and mean comes from a shared sensitivity to influential points. A focus on influential points connects naturally to sensitive and robust statistics, and paves the way for a discussion of influence in bivariate and multivariate settings.

Acknowledgements

References

Theoretical literature

Ageel, M.I. (2000), "The Mean-Median-Mode Inequality for Discrete Unimodal Probability Measure," *Far East Journal of Mathematical Sciences*, 2, 187-192.

Cytel Software Corporation (2004), "StatXact Example 3: FDA Animal Toxicology Data Yields Sky-Scraper Distribution for Stratified Trend Test," Accessed 18 February 2004. Available at www.cytel.com/StatXact/example_03.asp

Dharmadhikari, S.W., and Joag-dev, K. (1983), "Mean, Median, Mode III," *Statistica Neerlandica*, 33, 165-168.

_____ (1988), *Unimodality, Convexity, and Applications*, Boston: Academic Press.

Dudewicz, E.J., and Mishra, S.N. (1988), *Modern Mathematical Statistics*, New York: Wiley.

Dyson, J.E., and Williams, D.A. (1997), *The Physics of the Interstellar Medium*, Bristol, UK: Institute of Physics.

Groeneveld, R.A. (1986), "Skewness for the Weibull Family," *Statistica Neerlandica*, 40, 135-140.

Groeneveld, R.A. (1991), "An Influence Function Approach to Describing the Skewness of a Distribution," *The American Statistician*, 45, 97-102

Groeneveld, R.A., and Meeden, G. (1977), "The Mode, Median, and Mean Inequality," *The American Statistician* 31(3), 120-121.

MacGillivray, H.L. (1981), "The Mean, Median, Mode Inequality and Skewness for a Class of

Densities," *Australian Journal of Statistics*, 23(2), 247-250.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Zwet, W.R. van (1979), "Mean, Median, Mode II," *Statistica Neerlandica*, 33, 1-5.

Textbooks for introductory data analysis

13 textbooks that teach the relationship between skew, median, and mean

Agresti, A., and Finlay, B. (1997), *Statistical Methods for the Social Sciences*, 3rd ed., Upper Saddle River, NJ: Prentice Hall.

Bartz, A.E. (1999), *Basic Statistical Concepts*, 4th ed., Upper Saddle River, NJ: Prentice Hall.

Frankfort-Nachmias, C., and Leon-Guerrrero. (2002), *Social Statistics for a Diverse Society*, 3rd ed., Thousand Oaks, CA: Pine Forge.

Freund, J.E. (2004), *Modern Elementary Statistics*, Upper Saddle River, NJ: Pearson Prentice Hall.

Gravetter, F.J., and Wallnau, L.B. (2000), *Statistics for the Behavioral Sciences*, 5th ed., Belmont, CA: Wadsworth.

Kendrick, J.R. (2005), *Social Statistics: An Introduction Using SPSS for Windows*, Boston: Pearson.

Knoke, D., Bohrnstedt, G.W., and Mee, A.P. (2002), *Statistics for Social Data Analysis*, 4th ed., Itasca, IL: Peacock.

Levin, J., and Fox, J.A. (2003), *Elementary Statistics in Social Research*, 9th ed., Boston: Allyn and Bacon.

Levin, J., and Fox, J.A. (2004), *Elementary Statistics in Social Research: The Essentials*, Boston: Allyn and Bacon.

Maxwell, N. (2004), *Data Matters: Conceptual Statistics for a Random World*, Emeryville, CA: Key College Publishing.

Moore, D.S. (2000), *The Basic Practice of Statistics*, 2nd ed., New York: Freeman.

Moore, D.S., and McCabe, G.P. (2003), *Introduction to the Practice of Statistics*, 4th ed., New York: Freeman.

Ritchey, F. (2000), *The Statistical Imagination: Elementary Statistics for the Social Sciences*, Boston: McGraw-Hill.

Thorne, B.M., and Giessen, J.M. (2000), *Statistics for the Behavioral Sciences*, 3rd ed., Mountain View, CA: Mayfield.

4 textbooks that do not teach the relationship

Aron, A., and Aron, E.N. (2002), *Statistics for the Behavioral and Social Sciences: A Brief Course*, 2nd ed., Upper Saddle River, NJ: Prentice Hall.

Berry, D.A. (1996), *Statistics: a Bayesian Perspective*, Belmont, CA: Wadsworth.

Sweet, S.A., and Grace-Martin, K. (19xx), *Data Analysis with SPSS: A First Course in Applied Statistics*, 2nd ed., Boston: Allyn and Bacon.

Watkins, A.E., Scheaffer, R.L., Cobb, G.W. (2004), *Statistics in Action: Understanding a World of Data*, Emeryville, CA: Key College Publishing.

Addendum

Volume 13, Number 3, of the *Journal of Statistics Education* contains a [Letter to the Editor](#) concerning this article.

Paul T. von Hippel
Department of Sociology and Initiative in Population Research
300 Bricker Hall
190 N. Oval Mall
The Ohio State University
Columbus, OH 43210
USA
von-hippel.1@osu.edu