International Journal of Aviation Research

Volume 11 | Issue 01 Peer-Reviewed Article

Does Slight Skewness Matter?

David Trafimow
New Mexico State University

Tonghui Wang
New Mexico State University

Cong Wang
New Mexico State University

Hunter A. Myüz
New Mexico State University

Researchers feel they have the green light to ignore low levels of skewness because the central limit theorem indicates that such levels do not seriously compromise the validity of significance tests (e.g., Rouaud, 2013). And yet, significance testing is not the only issue, and the present focus is on two other issues where the central limit theorem does not come to the rescue. First, there is the question of how well the sample location statistic estimates the population location parameter. Surprisingly, skewness increases the precision of the estimation, and this increase in precision is impressive even with very low levels of skewness. Thus, by ignoring low levels of skewness, researchers are throwing away an important advantage. Second, experimental manipulations can cause differences in means across conditions, when there is no difference in locations across conditions; so the experiment seems to have worked based on means, when it really has not worked based on locations. Thus, moderate effect sizes can be caused by slight changes in skewness. For both reasons, it is recommended that researchers always attend to skewness, even when it is slight; and consider locations whenever they consider means.

Recommended Citation:

Trafimow, D., Wang, T., Wang, C, & Myüz, H.A. (2019). Does Slight Skewness Matter? *International Journal of Aviation Research, 11*(01), 11-24.

Researchers who are serious about the proper interpretation of data are consistent about advising researchers to carefully peruse their data at descriptive levels (Tukey, 1977), including using visual displays (Valentine, Aloe, & Lau, 2015). An advantage of careful data perusal pertaining to skewness—the present issue of concern—is that researchers can assess whether the data are skewed, and by how much. There are several rules of thumb about acceptable levels of skewness, such as that skewness should be between -1 and 1, or between -.8 and .8.[1] Surely, if the absolute magnitude of the skewness is sufficiently small, such as less than 1 or .8, it is small enough to be ignored. If the data are close to normal, and the calculated skewness is sufficiently slight, the central limit theorem justifies that the researcher can treat the data as normally distributed, without negative implications for the validity of typical parametric null hypothesis significance tests.

Our main goal is to present arguments that counter the foregoing conclusion. Although it is granted that slight skewness usually is not an important problem for the validity of typical parametric null hypothesis significance tests, significance testing is not the only issue.[2] Our goal is to demonstrate that low levels of skewness have important effects on sampling precision and data interpretation. A consequence is that, contrary to current data analysis trends, researchers almost always should worry about skewness. There will be two main arguments. First, skewness increases precision, and much of the increase occurs at absolute levels much less than 1, or even .8 (Trafimow, Wang, & Wang, 2019). By ignoring skewness, and depending on sample means instead of sample locations, researchers are depriving themselves of a great deal of precision that they otherwise would gain. According to this argument, skewness is desirable—a friend rather than an enemy—and should be embraced rather than seen as a problem to be rectified by data transformation. Second, by ignoring skewness—even slight skewness—the differences in means that provide the main tests of hypotheses in typical experimental research, are subject to an alternative explanation that can be eliminated simply by considering locations too, as opposed to only considering means (Trafimow, Wang, & Wang, 2018). But a prerequisite for considering locations is that researchers must consider skewness, even slight skewness.

Skew-Normal Distributions

Both arguments to be presented, about precision and about an alternative explanation for differences between means, depend on a minimal level of understanding of the family of skew-normal distributions. The present section presents that preparatory information (see Azzalini, 2014 for a comprehensive review of skew-normal distributions). Subsequently, there is the presentation of arguments about precision and about an alternative explanation for differences between means.

Let us commence with the family of normal distributions that constitute a small subset of the much larger family of skew-normal distributions. Normal distributions have two parameters: these are the mean μ and the standard deviation σ. In contrast, skew-normal distributions have

[1] There are many such recommendations (e.g., see Bulmer, 1979 for a classic textbook).

[2] Null hypothesis significance testing has recently come under strong attack (see Briggs, 2016; Hubbard, 2016; Ziliak & McCloskey, 2016 for recent reviews), has been banned from *Basic and Applied Social Psychology* (Trafimow & Marks, 2015), and was widely criticized at the American Statistical Association Symposium on Statistical Inference (2017). We strongly disfavor its use, though this argument is not necessary for present purposes.

three parameters. These are the location ξ, scale ω, and skewness or shape λ. When there is no skewness, and the distribution is normal, the location and mean are the same ($\xi = \mu$), and the scale and standard deviation are the same ($\omega = \sigma$). But when the distribution is skewed ($\lambda \neq 0$), the location is different from the mean ($\xi \neq \mu$), and the scale is different from the standard deviation ($\omega \neq \sigma$). For skew-normal distributions, because locations and scales are parameters and means and scales are not; locations are more informative than means, and scales are more informative than the standard deviations. Azzalani (2014) presents the probability density function and cumulative density function of skew-normal distributions, but this level of understanding is not necessary to comprehend the arguments to be made here.

The Precision Argument

The precision argument derives from a consideration of locations and scales in the context of work by Trafimow concerning *a priori* thinking (2017; also see Trafimow & MacDonald, 2017; Trafimow et al., 2019). Under the assumptions of random sampling from a normal distribution, Trafimow (2017) provided an accessible proof of Equation 1 below. The idea is to determine the sample size n for any desired level of precision f, expressed as a fraction of a standard deviation of the sample mean from the population mean, and the *z*-score Z_C corresponding to the desired level of confidence of having the sample mean be within f of the population mean. That is,

$$n = \left(\frac{Z_C}{f}\right)^2. \quad (1)$$

For example, suppose that a researcher tests a new type of aviation training and wishes to have a 95% probability of obtaining a sample mean that is within .1 of a standard deviation of the population mean. The sample size needed to meet these prescriptions is as follows: $n = \left(\frac{1.96}{.1}\right)^2 = 384.16$. Thus, rounding to the nearest upper whole number, the researcher needs to collect 385 participants to have a 95% probability of obtaining a sample mean within one tenth of a standard deviation of the corresponding population mean. Remembering that under the normal distribution, the location equals the mean and the scale equals the standard deviation, if the goal were to be 95% confident that the sample location is within .1 of a scale unit of the corresponding population location, the researcher again would need 385 participants to meet specifications.

In contrast, suppose the shape parameter is .8, or 1. What difference does this make for the required number of participants to meet the same specifications? The complex details of how to make the calculations are described by Trafimow et al. (2019).[3] What matters at present is that slight skewness renders dramatic change regarding the sample size necessary to meet specifications. To see this, suppose the researcher wishes to be 95% confident of being within .1 of a scale unit of the population location. When the shape parameter is .8, the sample size needed to meet specifications shrinks to 150. Moving to a shape parameter of 1, the sample size needed

[3] Unfortunately, although there is a simple closed form expression for the normal case (e.g., Equation 1), there is no closed form expression for the skew-normal case. Thus, solutions must be obtained numerically via computer. Trafimow et al. (2019) provide the mathematics and computer programming.

to meet specifications shrinks to 146. These are dramatic reductions from the necessary sample size of 385 when there is normality!

The reader might be puzzled about why seemingly trivial skews matter so much. The mathematical answer, and how to perform the calculations, is complex (see Trafimow et al., 2019 for details). But the explanation can be rendered qualitatively. Specifically, as skewness increases, the bulk of the distribution becomes narrower and taller. Because most data points randomly sampled from a distribution are from the bulk rather than the tail, the narrower the bulk, the greater the precision. Thus, it takes fewer data points to provide a good estimate of the population location.

Given pronouncements about levels of skewness that safely can be ignored, the surprising fact uncovered by a concern with precision is that most of the difference that skewness makes occurs at very low levels. To clarify, consider Figure 1, where the necessary sample size to meet specifications is represented along the vertical axis, as a function of skew along the horizontal axis. Each of the curves represents a different level of precision $(f = .1, .2, .3, \text{or } .4)$, using 95% as the confidence level. In all cases, the curves decrease monotonically, and reach asymptote quickly. As a dramatic example, even moving from a shape parameter of 0 to an extremely small skew when the shape parameter is .1, when $f = .1$, reduces the necessary sample size from 385 to 227. Clearly, from the point of view of precision, even an extremely small degree of skewness—perhaps too small to see easily in a histogram—matters.

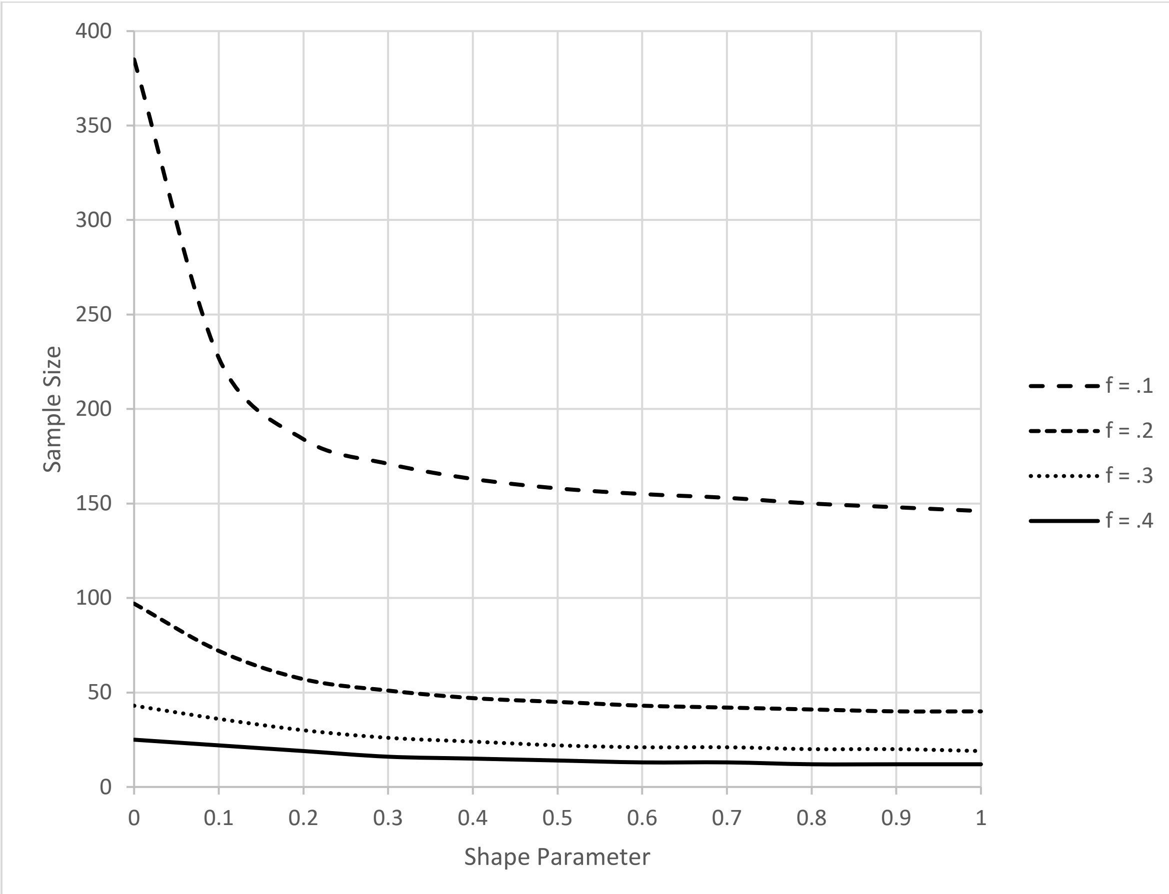

Figure 1. The sample size required to meet specifications for 95% confidence and precision levels of *f* = .1, .2, .3, or .4; as a function of the shape parameter.

An Alternative Explanation for Differences Between Means

To understand a potential alternative explanation for a difference in means, consider the relation between location and mean, and between scale and standard deviation (e.g., Trafimow, Wang, & Wang, 2018). These are presented below in Equation 2.
The mean and variance are

$$\mathrm{E}(X) = \xi + \sqrt{\frac{2}{\pi}}\delta\omega \quad \text{and} \quad \mathrm{V}(X) = \omega^2\left(1 - \frac{2}{\pi}\delta^2\right), \tag{2}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.[4] An implication of Equation 2 is that as λ increases, so that δ increases too, the mean is increasingly displaced from the location, in the direction of the skewness; and the standard deviation decreases relative to the scale.

Imagine a typical aviation training experiment where a researcher randomly assigns participants to a new training procedure or the old one; these are the experimental and control

[4] Delta δ is commonly used for at least two reasons. First, it shortens equations such as Equation 2. Second, it has the nice property of ranging from 0 to 1 as a function of the shape parameter.

conditions, respectively. Let us suppose that mean performance (by whatever measure) is significantly higher in the experimental condition than in the control condition; how much does this finding support the hypothesis that the new training procedure is superior to the old one? Further, to avoid issues pertaining to statistical significance, let us imagine full access to population parameters, so all values used for illustration in this section pertain to population parameters.

The degree of support the finding provides for the hypothesis depends, in large part, on alternative explanations. And there is one alternative explanation that is generally a potential problem. Suppose that the new training procedure causes the distribution to become more positively skewed relative to the control condition, without influencing the location or scale. Well, then, increasing skewness will (a) cause the mean to increase and (b) cause the standard deviation to decrease. Thus, although the lack of change in location, in our example, clearly indicates that the experiment did not work, the changes in mean and standard deviation will nevertheless render the appearance that it did work.

To see this, consider a specific example where the distribution is normal in the control condition and skew-normal in the experimental condition, with the shape parameter equal to .8, which, by conventional thinking, indicates minimal skewness and safely can be ignored. Let us also assume that the location is 10 in both conditions, and the scale is 4 in both conditions. In the control condition, where the location and mean are the same, and the scale and standard deviation are the same, it should be clear that the mean is 10 and the standard deviation is 4. But matters change for the experimental condition, again keeping the location at 10 and scale at 4. The bullet listed calculations make use of Equation 2 to exemplify the changes in mean and standard deviation caused by inducing the allegedly minimal degree of skewness in the experimental condition.

- $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} = \frac{.8}{\sqrt{1+.8^2}} = .6247;$
- $\mathrm{E}(X) = \xi + \sqrt{\frac{2}{\pi}}\delta\omega = 10 + \left(\sqrt{\frac{2}{\pi}} \cdot .6247 \cdot 4\right) = 11.5908;$
- $\mathrm{V}(X) = \omega^2\left(1 - \frac{2}{\pi}\delta^2\right) = 4^2\left(1 - \frac{2}{\pi} \cdot .6247^2\right) = 12.0250,$
    - $\text{Standard deviation} = 3.4677.$

Thus, although the fact that the locations are the same indicates that the experiment did not work, there is a moderate difference in means, from 10 in the control condition to 11.5908 in the experimental condition. Or in terms of the effect size, pooling the standard deviations in the two conditions, as is standard practice for obtaining the denominator of Cohen's *D* $\left(\text{Cohen's } D = \frac{Difference\ in\ means}{Pooled\ standard\ deviation}\right)$, the effect size is .4260, which most reviewers and editors would consider sufficiently strong to justify publication. Remember, though, that this is the effect size using the difference in means in the numerator of Cohen's *D* (Cohen, 1988). If the difference in locations is used, the effect size appropriately is 0.

And matters become worse if one uses 1, rather than .8, as the shape parameter in the experimental condition, again which, by conventional thinking, is slight and can be ignored:

- $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} = \frac{1}{\sqrt{1+1^2}} = .7071;$
- $\mathrm{E}(X) = \xi + \sqrt{\frac{2}{\pi}}\delta\omega = 10 + \left(\sqrt{\frac{2}{\pi}} \cdot .7071 \cdot 4\right) = 11.8006;$
- $\mathrm{V}(X) = \omega^2\left(1 - \frac{2}{\pi}\delta^2\right) = 4^2\left(1 - \frac{2}{\pi} \cdot .7071^2\right) = 10.9070,$
    - $\text{Standard deviation} = 3.3026.$

The difference in means is from 10 in the control condition to 11.8006 in the experimental condition. Or in terms of the effect size, Cohen's *D* is .4931.

The example effect sizes when the shape parameter in the experimental condition increases to .8 or 1.0 may seem surprisingly large in the context of no differences in locations or scales and beg some explanation. It is important to understand that two factors contribute to effect sizes when there is no change in location or scale but there is a change in shape. First, of course, the mean moves in the direction of the skew. This is illustrated in Figure 2. Second, as the shape parameter increases, the standard deviation decreases, as is illustrated in Figure 3. Thus, the numerator of Cohen's *D* increases; and the denominator decreases, due to pooling the decreased standard deviation in the experimental condition with the standard deviation in the control condition. As Figure 4 illustrates, both effects combine to render moderate values for Cohen's *D*. Note that the seeming effects demonstrated in Figure 2, Figure 3, and especially Figure 4, occur at allegedly unimportant levels of shape parameters that are less than or equal to 1.

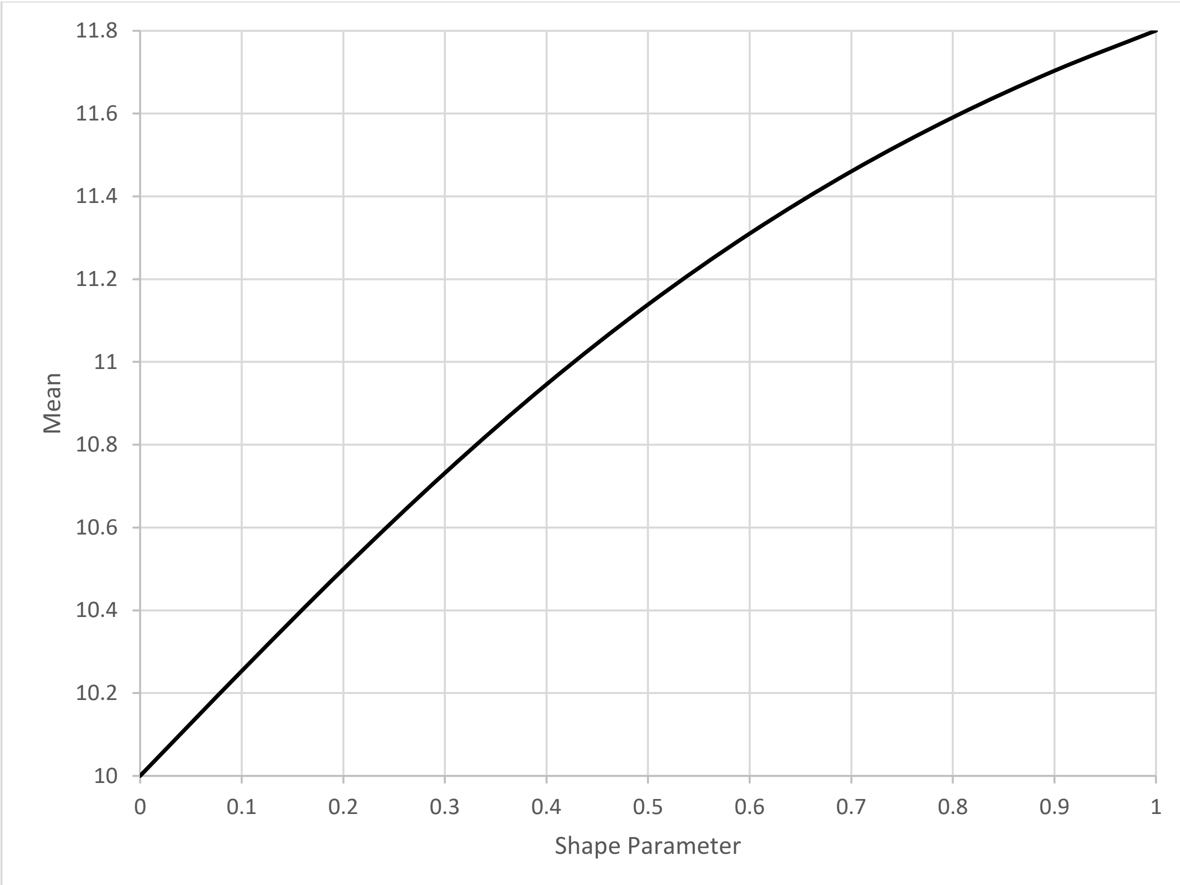

Figure 2. The mean of the experimental condition as a function of the shape or skewness parameter, with location = 10 and scale = 4.

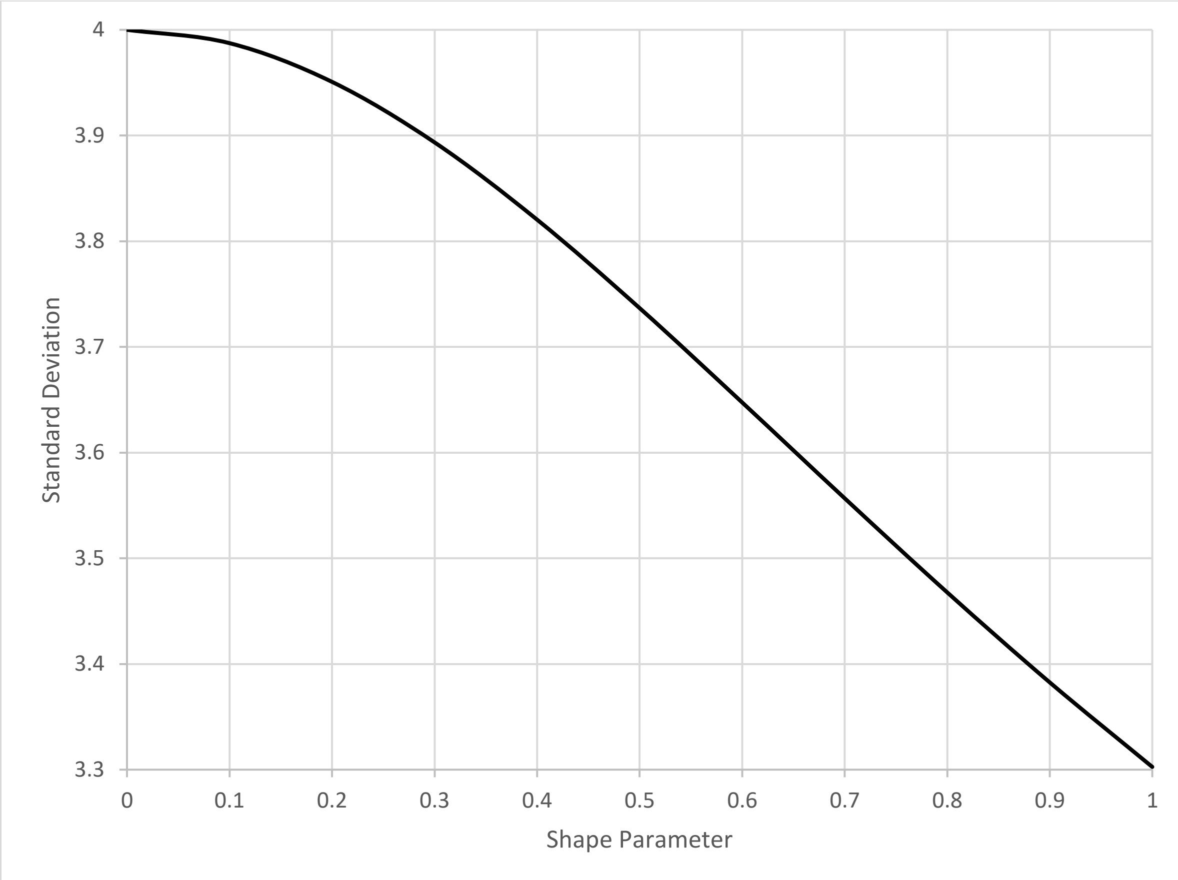

Figure 3. The standard deviation of the experimental condition as a function of the shape or skewness parameter, with location = 10 and scale = 4.

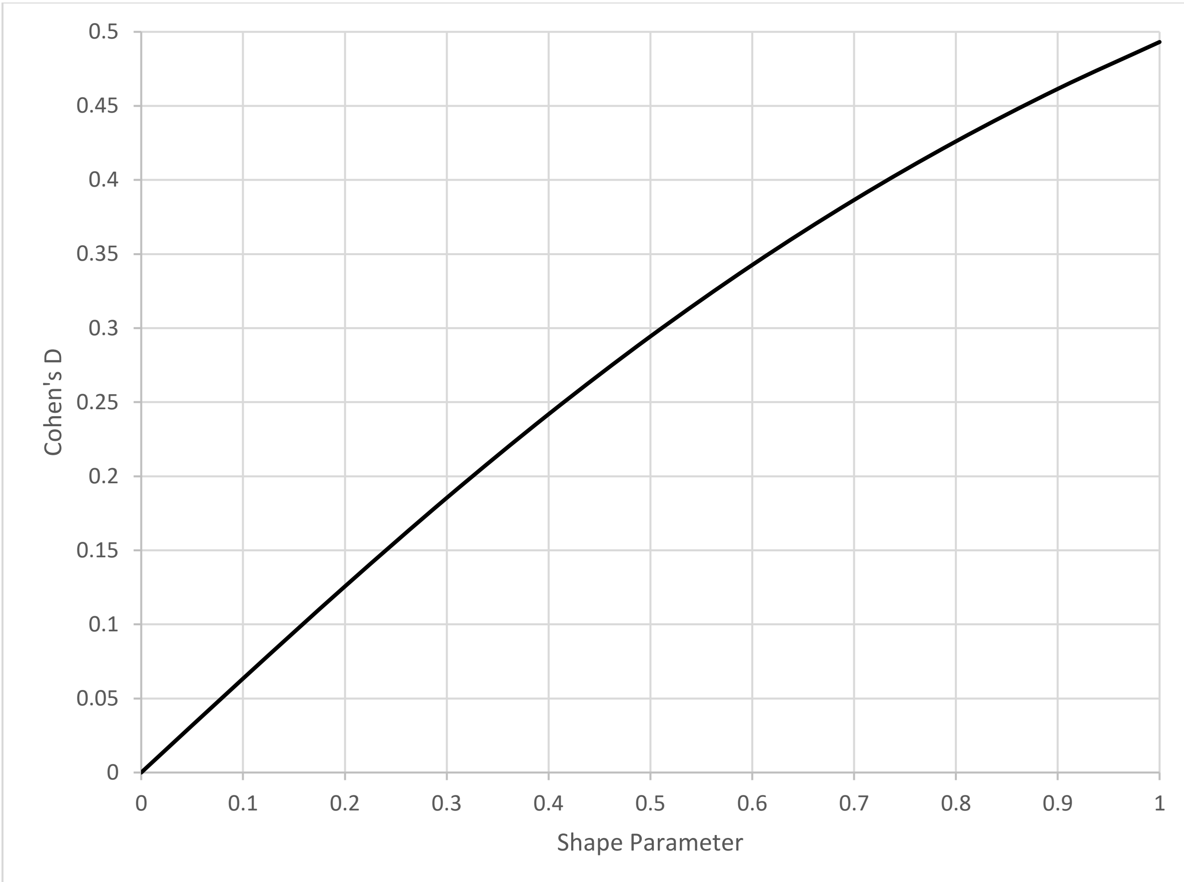

Figure 4. The effect size, expressed as Cohen's *D*, as a function of the shape or skewness parameter, with location = 10 and scale = 4 for both the experimental and control conditions.

Conclusion

Despite the received wisdom that small deviations from normality, such as shape parameters with an absolute value of 1 or .8, safely can be ignored, two important factors contradict the received wisdom. First, skewness renders fewer participants necessary for sample locations to be good approximations of population locations. Or if the sample size remains constant, skewness increases the precision with which the sample location estimates the population location. Moreover, the effect is a large one, even at allegedly slight or even miniscule levels of skewness. Second, slight changes in skewness, from the control condition to the experimental condition, can cause moderate differences in means and effect sizes, even when there is no difference whatsoever in the locations and scales of the two distributions. In general, then, it is a poor idea to ignore skewness; to fail to calculate sample location, scale, and skewness statistics; and to fail to compare locations across conditions, as well as means.

Nor are the calculations difficult. At the population level, Equation 2 shows how to obtain means and standard deviation from locations, scales, and shapes. After suitable algebraic manipulation, it is easy to write equations that allow the calculation of location and scale parameters from means, standard deviations and shapes, such as Equation 3.

$$\xi = \sqrt{\frac{2}{\pi}}\delta\omega - \mu \qquad \text{and} \qquad \omega^2 = \frac{\sigma^2}{(1-\frac{2}{\pi}\delta^2)}, \tag{3}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

At the sample level, it is necessary to use estimates. Fortunately, most computer programs will calculate a sample mean $\bar{X}$, standard deviation s, and skewness $\hat{\gamma}_1$ for each condition. In turn, these can be used to estimate population shape, location, and scale parameters for each condition, using the method of moments. There are alternative methods not covered here (Azzalani & Capitanio, 1999; Zhu, Ma, Wang, & Teetranont, 2017).

First, it is useful to estimate the shape parameter, which necessitates using Equation 4 to obtain an estimate of the absolute value of delta $|\hat{\delta}|$:

$$|\hat{\delta}| = \sqrt{\frac{\pi}{2}\frac{\hat{\gamma}_1^{\frac{2}{3}}}{\hat{\gamma}_1^{\frac{2}{3}}+\left(\frac{(4-\pi)}{2}\right)^{\frac{2}{3}}}}, \tag{4}$$

where the sign of $\hat{\delta}$ is the same as the sign of $\hat{\gamma}_1$.

In turn, it is easy to obtain an estimate of the shape or skewness parameter $\hat{\lambda}$ using Equation 5:

$$\hat{\lambda} = \frac{\hat{\delta}}{\sqrt{1-\hat{\delta}^2}}. \tag{5}$$

Finally, the estimate of delta can be used to estimate the location and scale parameters. Equation 6 below rewrites Equation 3, but with estimators obtained from samples, rather than using population parameters.

$$\hat{\xi} = \sqrt{\frac{2}{\pi}}\hat{\delta}\hat{\omega} - \bar{X} \qquad \text{and} \qquad \hat{\omega}^2 = \frac{s^2}{(1-\frac{2}{\pi}\hat{\delta}^2)}. \tag{6}$$

To recapitulate, there is no disagreement that a low level of skewness often is unimportant from the point of view of performing typical null hypothesis significance tests; but the present goal is to emphasize that this is not the only issue. The figures and examples illustrate that an allegedly unimportant level of skewness nevertheless has important effects on both precision and on effect sizes. Our main goals were to educate aviation researchers to appreciate the importance of these points and to encourage careful consideration of skewness in their own research. Embracing skewness by computing location statistics, rather than settling only for means, will increase precision. And because precision is intimately connected to replicability, increasing precision also should increase replicability (Trafimow, 2018). In addition, computing location statistics can be an invaluable aid for evaluating whether experimental manipulations really do shift distributions, whether they influence distribution shapes, both, or neither.

Moreover, it is easy to estimate the parameters of skew-normal distributions; such as location, scale, and skewness; given that most computer programs provide statistics pertaining to sample means, sample standard deviations and sample skewness. Because most distributions are skewed (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013; Ho & Yu, 2015; Micceri, 1989); and even slight levels of skewness have important consequences for precision and effect sizes; it is urged that the estimation of location, scale, and skewness parameters become obligatory practices in aviation research.

References

Azzalini, A. (2014). *The skew-normal and related families.* Cambridge University Press: New York.

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, *61*(3), 579–602. doi: 10.1111/1467-9868.00194

Blanca, M.J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(2), 78-84. doi: 10.1027/1614-2241/a000057

Briggs, W. (2016). *Uncertainty: The soul of modeling, probability and statistics*. New York: Springer.

Bulmer, M. G. (1979). *Principles of statistics*. Dover Publications, Inc.: New York.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Erlbaum: Hillsdale, NJ.

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Ed. Psych. Meas. 75*, 365-388.

Hubbard, R. (2016). *Corrupt research: The case for reconceptualizing empirical management and social science*. Los Angeles, California: Sage Publications.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psych. Bull*. *105*, 156-166.

Rouaud, M. (2013). *Probability, statistics, and estimation: Propagation of uncertainties in experimental measurement*. Creative Commons.

Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from *a posteriori* to *a priori* inferential statistics. *Educational and Psychological Measurement*, *77*(5), 831 – 854. https://doi.org/10.1177/0013164416667977

Trafimow, D. (2018). An *a priori* solution to the replication crisis. *Philosophical Psychology, 31*(8), 1188-1214. doi: 10.1080/09515089.2018.1490707

Trafimow, D., & MacDonald, J. A. (2017). Performing inferential statistics prior to data collection. *Educational and Psychological Measurement*, *77*(2), 204-219. doi: 10.1177/0013164416659745

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*(1), 1-2. http://www.tandfonline.com/doi/full/10.1080/01973533.2015.1012991

Trafimow, D., Wang, T., & Wang, C. (2018). Means and standard deviations, or locations and scales? That is the question! *New Ideas in Psychology*, *50*, 34-37. https://doi.org/10.1016/j.newideapsych.2018.03.001

Trafimow, D., Wang, T., Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! *Educational and Psychological Measurement*, *79*(1), 129-150. doi: 10.1177/0013164418764801

Tukey, J. (1977). *Exploratory Data Analysis*. Reading, M. A.: Addison-Wesley.

Valentine, J. C., Aloe, A. M., & Lau, T. S. (2015). Life after NHST: How to describe your data without "p-ing" everywhere. *Basic and Applied Social Psychology*, *37*(5), 260-273. http://dx.doi.org/10.1080/01973533.2015.1060240

Zhu X., Ma Z., Wang T., Teetranont T. (2017). Plausibility regions on the skewness parameter of skew normal distributions based on inferential models. In Kreinovich V., Sriboonchitta S., Huynh VN. (eds), *Robustness in Econometrics*. *Studies in Computational Intelligence*, vol 692. Springer, Cham. https://doi.org/10.1007/978-3-319-50742-2_16

Ziliak, S. T., & McCloskey, D. N. (2016). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, Michigan: The University of Michigan Press.