

The geometric mean: Confidence limits and significance tests

EDWARD F. ALF and JOHN M. GROSSBERG
San Diego State University
San Diego, California 92182

The variability of magnitude estimations has been found to grow approximately in proportion to the magnitude, and to produce distributions that are roughly log normal. Consequently, averaging is done best by taking geometric means of the estimations. This method of averaging also has the advantage that, despite the different ranges of numbers used by different observers, no normalizing is needed prior to averaging. (Stevens, 1966, p. 531)

Ernst Heinrich Weber (1795-1878) is generally credited with being the first to observe that the increment in a stimulus that is just noticed is proportional to the size of the stimulus. According to Boring (1957), Weber performed many experiments verifying this, and reported them in his Latin monograph, *De tactu: annotationes anatomicae et physiologicae*, which appeared in 1834. Weber noted, for example, that, in lifting weights, we can "just perceive" the difference between two weights if they differ by about 1/40th, e.g., if the heavier is 1/40th larger than the lighter of the two weights. The heavier the weight, the greater must be the increment in order for it to be noticed.

Although just perceptible increments change as a function of stimulus size, the difference between the logarithms of the stimulus and the just noticeably larger stimulus will be constant. Thus, to the extent that Weber's law is true, the use of logarithms equalizes the just perceptible differences throughout the usual stimulus range. Furthermore, these logarithms tend to be normally distributed, and tend to have the same standard deviation for different stimuli throughout the stimulus range. As the opening quotation from Stevens (1966) indicates, it can be argued that in combining several independent estimates of a judged quantity, one should find the mean of the logarithms of the estimates, and then take the anti-logarithm of the result, which is the geometric mean of the observations.

Very possibly, the geometric mean of the observations provides a more stable estimate of the subjective evaluation of a stimulus than does the arithmetic mean. However, two important questions arise regarding the resulting statistic: (1) What is the formula for finding the approximate standard error of the geometric mean? (2) How can we calculate the exact confidence limits for the geometric mean? The present paper provides answers to these questions.

APPROXIMATING THE STANDARD ERROR

Derivation

Consider a sample of n values, $x_1, x_2, \dots, x_i, \dots, x_n$, from the random variable X , in which $x_i > 0$ for all population values. The sample geometric mean, G , is defined as:

$$G = (\prod x)^{1/n}. \quad (1)$$

From Equation 1, the arithmetic mean of the logarithms of the x values, M , may be written as:

$$M = (\sum \ln x)/n = \ln G. \quad (2)$$

That is, the logarithm of the geometric mean, $\ln G$, is equal to M ; the arithmetic mean of the logarithms of the sample values. It is sometimes more convenient to calculate G as the antilogarithm of the mean of the logarithms:

$$G = f(M) = e^M. \quad (3)$$

We may define the parameter μ as:

$$\mu = E(M). \quad (4)$$

Note that μ is the population arithmetic mean, or expected value, of the random variable $\ln X$.

We may define a second parameter, μ_e , as an explicit function of μ :

$$\mu_e = f(\mu) = e^\mu. \quad (5)$$

Note that μ_e is the geometric mean of the random variable X . And when the distribution of $\ln X$ is symmetrical, μ_e will also be the population median of the random variable X .

Any sample geometric mean, G , may be expressed in the form

$$G = \mu_e + \Delta G = f(\mu + \Delta M), \quad (6)$$

where ΔG represents the error in G (i.e., its deviation from the parameter μ_e), and ΔM represents the error in M (i.e., its deviation from the parameter μ).

Expanding the right side of Equation 6 by Taylor's theorem (Scarborough, 1962, p. 505) and discarding higher derivatives,

$$\mu_e + \Delta G \cong f(\mu) + \Delta M f'(\mu), \quad (7)$$

where $f'(\mu)$ indicates that the first derivative is to be evaluated at the value μ .

Subtracting Equation 5 from Equation 7,

$$\Delta G \cong \Delta M f'(\mu). \tag{8}$$

Notice in Equation 8 that the errors in G are approximately equal to the errors in M multiplied by a constant. Therefore,

$$\sigma_G^2 \cong [f'(\mu)]^2 \sigma_M^2. \tag{9}$$

From Equation 5,

$$f(\mu) = e^\mu = f'(\mu) = \mu_e. \tag{10}$$

Substituting Equation 10 in Equation 9,

$$\sigma_G^2 \cong \mu_e^2 \sigma_M^2; \tag{11}$$

and the standard error of the geometric mean will be approximately equal to the square root of Equation 11,

$$\sigma_G \cong \mu_e \sigma_M. \tag{12}$$

Equation 12 refers to parameter values. In an applied problem, one would replace these parameters by their sample estimates,

$$s_G \cong G s_M, \tag{13}$$

where G is the sample geometric mean as defined in Equation 1, and s_M is the sample estimate of the standard error of the mean of the logarithms of the sample values, defined as

$$s_M = \{[n \sum (\ln x)^2 - (\sum \ln x)^2] / [n^2(n-1)]\}^{1/2}. \tag{14}$$

Accuracy of the Approximation

The approximation in Equation 11 doesn't require the random variable X to be lognormally distributed, but its accuracy *does* depend on the degree to which the relation between G and M is approximately linear. And the degree of linearity between G and M is heavily dependent on the value of σ_M . Let us examine this relation for the lognormal case, using realistic values for σ_M .

If X is, in fact, lognormally distributed, then the variance of G is exactly

$$\sigma_G^2 = \mu_e^2 e^{\sigma_M^2} (e^{\sigma_M^2} - 1) \tag{15}$$

(see, e.g., Hastings & Peacock, 1975, p. 84).

From Equations 11 and 15, the percent error in Equation 11 may be expressed as

$$\% \text{ error} = 100\{[\mu_e^2 \sigma_M^2] / [\mu_e^2 e^{\sigma_M^2} (e^{\sigma_M^2} - 1)]\}. \tag{16}$$

Finally, we may express σ_M^2 in the form

$$\sigma_M^2 = \sigma^2 / n, \tag{17}$$

where σ_M^2 is the variance of the mean of the logarithms, and σ^2 is the variance of the parent population of logarithms from which samples of size n are drawn.

Reasonable estimates of σ^2 may be taken from Luce and Mo (1965, Figure 6, p. 170), which shows the distribution of 2,000 log x values for each of two subjects. From these we would estimate σ to fall in the range $.15 \leq \sigma \leq .35$. From these estimates and Equation 17, we would estimate σ_M^2 for samples of size n = 3 to fall in the range $.0075 \leq \sigma_M^2 \leq .0408$. Substituting these values in Equation 16, we would estimate $-.37 \leq \% \text{ error} \leq -5.94$.

In summary, using reasonable estimates of σ^2 , in samples of size n = 3 we would expect the approximation in Equation 11 to lead to an underestimation of σ_G^2 of between about .5% to about 6%, when the lognormal assumption is met.

The Lognormal Assumption

The data of Luce and Mo (1965) suggest that the distribution of magnitude estimates may not be lognormal. This may well be correct. But, as Galton (1879) suggested in one of the earliest papers on this topic, the distribution of M will approach normality as n increases, for all parent distributions to which the central limit theorem applies. Thus, the distribution of G will approach the lognormal form, even though the parent distribution of X may not be lognormal.

The approach to normality is usually quite rapid. For example, Edwards (1968, p. 78) illustrates how the sampling distribution of the mean of even a completely rectangular distribution approaches normality closely in samples as small as n = 4. For unimodal leptokurtic distributions like those in Figure 6 of Luce and Mo (1965), the approach to normality should be even more rapid. To check on this assumption, we constructed hypothetical distributions similar to those in Figure 6 of Luce and Mo, and the actual standard errors of the geometric means in samples of size n = 3 were determined using Monte Carlo methods. For both subjects, the approximation in Equation 11 was closer to the true value than the approximation using the square root of Equation 15. Equation 11 overestimated σ_G by 3.03% and 4.9% for subjects 11 and 8, respectively; while the square root of Equation 15 overestimated σ_G by 4.65% and 6.9%, respectively. The slightly greater accuracy of Equation 11 may be due to its fewer assumptions regarding the parent distribution; so it may be preferable when the lognormal assumption is questionable. In any case, both approximations become more accurate as sample size increases.

Computational Procedures

Let x be a subject's judgment, and assume we have n values of x for which we wish to find the geometric

mean and its standard error. The computational procedures are as follows: (1) Find the natural logarithm, $\ln x$, for each x value. (2) Find the mean of the logarithms, which is simply the sum of the $\ln x$ values divided by n , the number of such values. (3) Find the antilogarithm of the mean of the logarithms. This is the geometric mean. (4) Find the standard error of the mean of the logarithms. This is given by Equation 14. (5) Multiply the standard error of the mean of the logarithms by the geometric mean. The resulting product is the standard error of the geometric mean.

EXACT METHODS

The methods of the preceding sections are approximate, but they should be serviceable enough for most purposes. However, when the assumptions of Stevens (1966) are met, and the responses follow a log normal distribution, then exact significance tests and exact confidence limits can be obtained for the logarithms of the responses. Because of the exact monotonic relation between the mean of the logarithms and the geometric mean of the responses, it is also possible, under these assumptions, to make exact significance tests on the geometric mean.

Exact Confidence Limits

Continuing the terminology of the previous sections, if the logarithms of the observations are normally distributed, then

$$(M - \mu)/(s_M) = t \quad (18)$$

will follow Student's t distribution.

Exact confidence limits can then be found as follows. For the $(1 - \alpha)\%$ confidence limits, the upper limit is

$$U = M + (s_M)t_{(\alpha/2; df)} \quad (19)$$

and the lower limit is

$$L = M - (s_M)t_{(\alpha/2; df)} \quad (20)$$

The limits in Equations 19 and 20 are for the expected value of the mean of the logarithms. Since the geometric mean is a monotonic function of the mean of the logarithms, the upper and lower confidence limits for the geometric mean are

$$\text{Upper Limit} = e^U, \quad (21)$$

$$\text{Lower Limit} = e^L. \quad (22)$$

Significance Tests

For hypotheses regarding the difference between geometric means, it is possible to perform conventional t tests on the logarithms of the observations. Since the geometric mean is a monotonic function of the mean of the logarithms, significance of the difference between the means of the logarithms implies significance of the difference between geometric means.

REFERENCES

- BORING, E. G. *A history of experimental psychology*. New York: Appleton-Century-Crofts, 1957.
- EDWARDS, A. L. *Experimental design in psychological research* (3rd ed.). New York: Holt, Rinehart and Winston, 1968.
- GALTON, F. The geometric mean in vital and social statistics. *Proceedings of the Royal Society of London*, 1879, **29**, 365-367.
- HASTINGS, N. A. J., & PEACOCK, J. B. *Statistical distributions*. New York: Wiley, 1975.
- LUCE, R. D., & MO, S. S. Magnitude estimation of heaviness and loudness by individual subjects: A test of a probabilistic response theory. *British Journal of Mathematical and Statistical Psychology*, 1965, **18**, 159-174.
- SCARBOROUGH, J. B. *Numerical mathematical analysis* (5th ed.). Baltimore: Johns Hopkins Press, 1962.
- STEVENS, S. S. A metric for the social consensus. *Science*, 1966, **151**, 530-541.

(Received for publication January 30, 1979;
revision accepted September 26, 1979.)