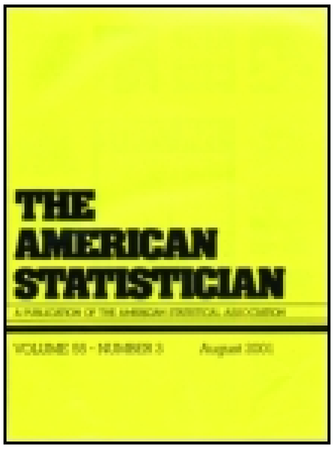

The American Statistician

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Is Kurtosis Really "Peakedness?"

Richard B. Darlington [a]

[a] Cornell University , USA

Published online: 12 Mar 2012.

[3,4] A similar editorial policy has been recommended before. Rosenthal (1966), concerned that researchers focus almost exclusively on outcomes, advocates a review procedure that excludes results:

> "What we need is a system for evaluating research based only on the procedures employed. If the procedures are judged appropriate, sensible, and sufficiently rigorous to permit conclusions from the results, the research cannot then be judged inconclusive on the basis of the results and rejected by the referees or editors. Whether the procedures were adequate would be judged independently of the outcome. To accomplish this might require that procedures only be submitted initially for editorial review or that only the result-less section be sent to a referee or, at least, that an evaluation of the procedures be set down before the referee or editor reads the results. This change in policy would serve to decrease the outcome-consciousness of editorial decisions, but it might lead to an increased demand for journal space. This practical problem could be met in part by an increased use of "brief reports" which summarize the research in the journal but promise the availability of full reports to interested workers."

REFERENCES

Bakan, David. "The test of significance in psychological research." *Psychological Bulletin*, 1966, *66* (6), 423–437.

Faia, Michael A. "A proposal for the standardization of Type II error in the replication of social research." *Sociology and Social Research*, 1966, *51*, 87–93.

Hudson, Jeffrey. *A Case of Need*. New York: The New American Library, Inc., 1968.

Kish, Leslie. "Some statistical problems in research design." *American Sociological Review*, 1959, *24*, (3), 328–338.

Lykken, David T. "Statistical significance in psychological research." *Psychological Bulletin*, 1968, *70*, (3), 151–159.

McNemar, Quinn. "At random: Sense and nonsense." *American Psychologist*, 1960, *15*, 295–300.

Meehl, P. E. "Theory testing in psychology and physics: A methodological paradox." *Philosophy of Science*, 1967, *34*, 103–115.

Newman, S. H. "Improving the evaluation of submitted manuscripts." *American Psychologist*, 1966, *21*, 980–981.

Rosenthal, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Croft, 1966.

Selvin, Hanan C. "A critique of tests of significance in survey research." *American Sociological Review*, 1957, *22*, (5), 519–527.

Selvin, Hanan C. and Stuart, A. "Data-dredging procedures in survey analysis." *American Statistician*, *20*, (3), 1966, 20–23.

Sterling, T. D. "Publication decisions and their possible effects on inferences drawn from tests of significance—or visa versa." *Journal of The American Statistical Association*, 1959, *54*, 30–34.

Tullock, Gordon. "Publication decisions and tests of significance—a comment." *Journal of the American Statistical Association*, 1959, *54*, 593.

Is Kurtosis Really "Peakedness?"[1]

RICHARD B. DARLINGTON
Cornell University

Abstract

Kurtosis is best described not as a measure of peakedness versus flatness, as in most texts, but as a measure of unimodality versus bimodality.

In a recent survey of the elementary statistical texts on his shelf, the present writer found 11 which attempted to explain the concept of kurtosis. Ten of these unambiguously used terms like "peaked" and "flat-topped" to describe high-kurtosis (leptokurtic) and low-kurtosis (platykurtic) distributions respectively. One pointed out that *platy* is the Greek word for flat. Only one specialized text [1, p. 68] suggested that the description "flat" does not adequately describe distributions with low kurtosis, and even there the details were unclear. The purpose of the present paper is to make somewhat clearer what kurtosis measures, what sorts of distributions have high and low kurtosis, and how kurtosis is altered when observations are added to an existing distribution.

The formal definition of kurtosis

For simplicity of notation we shall confine the discussion to distributions with finite numbers of observations. The generalization to infinite distributions is straightforward.

If m and s denote the mean and standard deviation of a distribution (where the denominator in s^2 is N rather than $N - 1$), then the usual measure of kurtosis[2] is

$$k = \frac{N^{-1}\sum(X - m)^4}{s^4}.$$

k is unaffected by changes in the mean or standard deviation of the distribution. k can be simply expressed as a function of the z scores; it is easy to show that

$$k = N^{-1} \sum z^4.$$

What does kurtosis mean in intuitive terms?

Most elementary texts describe kurtosis as a measure of the "peakedness" of a distribution. We shall attempt to show that this term is misleading, and that a far better term for describing kurtosis is "bimodality," where the lower the kurtosis, the greater the bimodality. We shall make this point in three ways:

(a) by an algebraic analysis of the formula for k, (b) by examining distributions with low and high kurtosis, and (c) by examining the change in k resulting from modifying an existing distribution by adding observations at various points.

Our "algebraic analysis of the formula for k" can be completed in one short paragraph. In a distribution of z scores, it is true by definition that

$$\text{Mean}(z^2) = 1.$$

Considerable insight into the meaning of k can be obtained from the fact that

$$\text{Variance}(z^2) = k - 1.$$

The proof is simple:

$$\text{Var}(z^2) = \frac{\sum (z^2)^2}{N} - [\text{Mean}(z^2)]^2$$

$$= \frac{\sum z^4}{N} - 1 = k - 1.$$

Thus k can be interpreted as a measure of the degree to which the values of z^2 cluster around their mean of 1; the greater the clustering, the lower k. Since z equals $+1$ or -1 when $z^2 = 1$, k can also be interpreted as a measure of the degree to which a distribution's z-scores cluster around $+1$ and -1. The most descriptive one-word summary of such clustering is "bimodality."

Using the term "bimodality" as a guide, the reader should be able to guess what distributions have the highest and lowest kurtosis of all possible distributions. Clearly in ordinary language the "most bimodal" of all possible distributions is a symmetric two-point distribution, while the "least bimodal" (or most unimodal) distribution is concentrated entirely at one point. The reader can check that these distributions have respectively lowest and highest kurtosis from the fact that in a symmetric three-point distribution in which p is the density at the mean,

$$k = 1/(1 - p).$$

As p approaches 1 (i.e., as the distribution approaches being concentrated entirely at its mean), k approaches infinity. On the other hand, when $p = 0$ (i.e., when the distribution is a two-point, rather than a three-point, distribution) k achieves its lowest possible value of 1. k cannot be less than 1 since $k - 1 = \text{Var}(z^2)$, and $\text{Var}(z^2)$ can obviously not be less than 0. A few moments' reflection shows that $\text{Var}(z^2)$ is zero (i.e., all z's are concentrated at $+1$ and -1) for no distribution except a symmetric two-point distribution, so that this is the only distribution for which k is 1.

The same conclusion follows if one examines, instead of the family of all symmetric three-point distributions, the family of all two-point distributions. If the densities at the two points are p and q, then it can be shown that

$$k = 1/pq - 3.$$

Again the lowest value of k occurs when $p = q = \frac{1}{2}$, and k approaches infinity as either p or q approaches zero (i.e., as the distribution concentrates at one point or the other).

Another way of grasping the meaning of k is to ask how k changes as new observations are added to an existing distribution. Let Z be some particular point in a probability distribution of z scores. If we add an observation at Z, then the distribution's mean will be changed (except of course when Z is zero). This complicates the analysis considerably, so we shall imagine that observations are added at both $+Z$ and $-Z$, thereby leaving the mean unchanged. Let D denote the derivative of k in respect to the total change in the size of the distribution. D is a convenient measure of the effect of the new observations on k. To say that D is positive is to say that the new observations have raised k; to say that D is negative is to say that k has been lowered.

From the previous discussion we would expect that D would be negative (that is, adding new observations would lower k) when the new observations are added in the regions around $z = 1$ and $z = -1$ (that is, when Z^2 is in the neighborhood of 1), and positive otherwise. This turns out to be the case. It can be shown that

$$D = (Z^2 - k)^2 - (k^2 - k).$$

From this formula it can be shown that D is positive for both low and high values of Z^2, and negative for intermediate values. Specifically, D is negative if Z^2 is within the interval between $k - (k^2 - k)^{1/2}$ and $k + (k^2 - k)^{1/2}$, zero if Z^2 is at either end of the interval, and positive if Z^2 is outside the interval. For a normal curve, for example, (for which $k = 3$) the interval runs from $|Z| = .74$ to $|Z| = 2.33$. Symmetrically-added observations within this region lower k, those outside raise it. If $k = 1$, the interval has zero width, being concentrated entirely at 1. As k approaches infinity, the lower end of the interval for Z^2 approaches $\frac{1}{2}$ (without ever reaching it), and the upper end approaches infinity. Thus unity is the only value of Z^2 which, regardless of the value of k, never makes D positive. This is of course consistent with our interpretation of k as a measure of the degree to which values of z^2 are dispersed around unity; adding new observations at $z^2 = 1$ never raises k, and in fact lowers k unless k has already reached its minimum value of 1.

[1] For helpful criticisms the author is indebted to Mrs. Jean Rom. Reprints are available free from the author at: Department of Psychology, Morrill Hall, Cornell Univ., Ithaca, N.Y. 14850.

[2] Often some simple function of k is used, like $(k - 3)$. We shall use k, but our discussion can be generalized readily to such other measures.

REFERENCE

1. Horst, P. *Psychological Measurement and Prediction*. Belmont, Calif.: Wadsworth, 1966.

Some Proofs on Kurtosis

Notation

m = mean of the distribution
s = standard deviation of the distribution
X = raw scores
z = standard scores

Theorem 1.

$$k = N^{-1} \sum z^4$$

Proof.

$$k = N^{-1}\frac{\sum(X-m)^4}{s^4} = N^{-1}\sum\left(\frac{X-m}{s}\right)^4$$
$$= N^{-1}\sum z^4$$

Theorem 2. In a two-point probability distribution with frequencies p, q, $k = 1/pq - 3$.

Proof. Without loss of generality, assume the two points of the distribution are at 0 and 1, with p being the frequency at 1. Then $m = p$, $s = (pq)^{1/2}$

$$z_1 = \frac{1-m}{s} = \frac{q}{(pq)^{1/2}} = \left(\frac{q}{p}\right)^{1/2}$$

$$z_0 = \frac{0-m}{s} = \frac{-p}{(pq)^{1/2}} = -\left(\frac{p}{q}\right)^{1/2}$$

$$k = N^{-1}\sum z^4 = (pz_1^4 + qz_0^4) = \left(\frac{q^2}{p} + \frac{p^2}{q}\right)$$

The subsequent reduction to the form $k = 1/pq - 3$ is then a straightforward algebraic process. A less straightforward shortcut is based on the fact that $a^2 - b^2 = (a+b)(a-b)$. Recalling always that $p + q = 1$, we have from the last line above,

$$k = \frac{q^2-1}{p} + \frac{1}{p} + \frac{p^2-1}{q} + \frac{1}{q}$$

$$= \frac{(q-1)(q+1)}{p} + \frac{(p-1)(p+1)}{q} + \frac{1}{p} + \frac{1}{q}$$

$$= -(q+1) - (p+1) + \frac{1}{p} + \frac{1}{q} = \frac{1}{p} + \frac{1}{q} - 3$$

$$= \frac{1}{pq} - 3$$

Theorem 3. In a symmetric three-point distribution in which p is the density at the mean, $k = 1/(1-p)$.

Proof. Assume without loss of generality that the three points of the distribution are at -1, 0, 1. Then

$$k = \frac{\left(\frac{1-p}{2}\right)(-1)^4 + p(0)^4 + \left(\frac{1-p}{2}\right)(1)^4}{s^4} = \frac{(1-p)}{s^4}$$

$$s^2 = \left(\frac{1-p}{2}\right)(-1)^2 + p(0)^2 + \left(\frac{1-p}{2}\right)(1)^2 = 1-p$$

$$\therefore k = \frac{1-p}{(1-p)^2} = \frac{1}{1-p}$$

Theorem 4. If two densities in a probability distribution, equidistant from the mean and at opposite sides of it, are each increased, with the total increase in both denoted as Δp, then

$$\frac{dk}{dp} = (Z^2-k)^2 - (k^2-k)$$

where Z^2 denotes the squared standard scores of the two points.

Proof. Without loss of generality, consider a discrete distribution of x scores, with zero mean and density p_i at the ith point in the distribution. Then the second and fourth central moments are respectively $\sum p_i x^2_i/\sum p_i$ and $\sum p_i x^4_i/\sum p_i$. Thus

$$k = \frac{\sum p_i x^4_i/\sum p_i}{(\sum p_i x^2_i/\sum p_i)^2} = \frac{(\sum p_i x^4_i)(\sum p_i)}{(\sum p_i x^2_i)^2} \quad (1)$$

Suppose two values of p_i, representing frequencies equidistant from the mean but at opposite sides therefrom, are each increased by an amount $\Delta p/2$. Then the mean will remain at zero, so Formula 1 will still be an accurate formula for k. The total size of the distribution will be increased by Δp. We find dk/dp_j through ordinary elementary rules of differentiation.

If we redefine p_i as the sum of the two probabilities at x_i and $-x_i$, and let the summations in Formula 1 be only over nonnegative values of x, then Formula 1 still holds. Differentiating the numerator and denominator of k, we have

$$\frac{d(\sum p_i x^4_i)(\sum p_i)}{dp_j} = (\sum p_i x^4_i) + (\sum p_i)x^4_j \quad (2)$$

$$\frac{d(\sum p_i x^2_i)^2}{dp_j} = 2(\sum p_i x^2_i)x^2_j \quad (3)$$

Using the familiar rule for the derivative of a ratio then gives

$$\frac{dk}{dp_j} = \frac{\sum(p_i x^4_i)}{(\sum p_i x^2_i)^2} + \frac{(\sum p_i)x^4_j}{(\sum p_i x^2_i)^2} - 2(\sum p_i x_i^2)x^2_j\frac{(\sum p_i x^4_i)(\sum p_i)}{(\sum p_i x^2_i)^4} \quad (4)$$

If $\sum p_i = 1$, then use of (1) gives

$$\frac{dk}{dp_j} = k + \frac{x^4_j}{(\sum p_i x^2_i)^2} - 2\left(\frac{x^2_j}{\sum p_i x^2_i}\right)k \quad (5)$$

Since $s^2 = \sum p_i x^2_i$, we have $z^2_{x_j} = x^2_j/(\sum p_i x^2_i)$. Denoting $z^2_{x_j}$ by Z^2, we have from (5)

$$\frac{dk}{dp_j} = k + (Z^2)^2 - 2\,Z^2k \quad (6)$$

which can be rewritten as

$$\frac{dk}{dp_j} = (Z^2-k)^2 - (k^2-k). \quad (7)$$

Theorem 5. If $D = (Z^2-k)^2 - (k^2-k)$, then D is negative if Z^2 is within the interval between $k - (k^2-k)^{1/2}$ and $k + (k^2-k)^{1/2}$, zero if Z^2 is at either end of the interval, and positive if Z^2 is outside the interval.

Proof. Setting $D = 0$ and solving for Z^2 gives, from the quadratic formula $Z^2 = k \pm (k^2 - k)^{1/2}$. Both values for Z^2 are real for any possible value of k. Since the formula for D describes an upright parabola in respect to Z^2, D is negative when Z^2 is between these limits, and positive when outside.

Theorem 6. D is positive if $Z^2 \leq \frac{1}{2}$. If $Z^2 > \frac{1}{2}$, D is negative for sufficiently large k.

Proof.

$$\begin{aligned} D &= (Z^2 - k)^2 - (k^2 - k) \\ &= (Z^2)^2 - 2Z^2k + k \\ &= (Z^2)^2 + (1 - 2Z^2)k. \end{aligned}$$

If $Z^2 \leq \frac{1}{2}$, then $(1 - 2Z^2) \geq 0$. Since $k > 0$ and $(Z^2)^2 \geq 0$, it follows that $D > 0$. On the other hand, if $Z^2 > \frac{1}{2}$, then $(1 - 2Z^2) < 0$, so that D is always negative for a sufficiently large value of k.

On Calculating the Gamma Function of Non-Integral Arguments

MORTON S. RAFF
National Heart and Lung Institute

For real numbers greater than 1 the gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \qquad (1)$$

may be regarded as a generalization of the factorial $(x - 1)!$ since it is equal to the factorial at integer values of x and takes other values at all points in between in a well-behaved continuous fashion. Applications of the gamma function for non-integer values of x are well known, as in the density functions for t, F, and χ^2.

The gamma function is tabled [2, 7]. Because of the recurrence relation

$$\Gamma(x + 1) = x\Gamma(x) \qquad (2)$$

the function for non-integer values is usually tabulated in detail only for a range of width one. For example, reference [2] gives $\Gamma(x)$ for $x = 1(.005)2$. The gamma function of a larger number can then be computed by repeated applications of (2), e.g.,

$$\Gamma(6.38) = (5.38)(4.38)(3.38)(2.38)(1.38)\Gamma(1.38).$$

This procedure is clumsy for large values of x, making it worth while to look for a more direct approach. The approximate formula given below avoids the repeated multiplications, has good accuracy even for small values of x, and dispenses with the gamma-function table, requiring only the factorials of integers.

Let the positive number x be partitioned into n plus p, where n is the integer part and p is the fractional part (perhaps 0, if x is an integer). Then for $n \geq 2$ the following approximation to $\Gamma(x)$ is never in error by more than 1 part in 13,000 and gets better as n increases:

$$\Gamma_F(n + p) = n![(n + \tfrac{1}{2}p)^2 + p(2 - p)/12]^{(p-1)/2}.$$
$$n = 2, 3, \cdots; 0 \leq p \leq 1 \qquad (3)$$

For values of x between 1 and 2 it is suggested that $\Gamma(x)$ be approximated by dividing $\Gamma_F(2 + p)$ by $(1 + p)$.

Formula (3) is a special case of an asymptotic relation given by J. S. Frame [5] and is denoted by the subscript F for that reason.[1] It will be noted that the right-hand expression gives the gamma function without error when $p = -1$, 0, 1, 2, or 3. Consequently the approximation $\Gamma_F(n + p)$ has error only for fractional values of p and differs in this respect from the well-known Stirling approximation. With any fixed number of terms the latter is never exact but becomes steadily better (ratio-wise) as x increases.

Table 1 shows how the accuracy of $\Gamma_F(n + p)$ varies with p for a fixed n and Table 2 shows how the maxi-

TABLE 1

Relative Error in Γ_F Approximation for

p	$\Gamma(2 + p)$	$\Gamma(10 + p)$
0	0	0
.1	-39×10^{-6}	-8×10^{-8}
.2	−62	−12
.3	−73	−16
.4	−74	−18
.5	−68	−18
.6	−57	−15
.7	−44	−13
.8	−29	−8
.9	−14	−4
1	0	0

[1] Frame's formula (7) reads
$\Gamma[N + \frac{1}{2}(1 + u)]/\Gamma[N + \frac{1}{2}(1 - u)] \sim [N^2 + (1 - u^2)/12]^{u/2}$.
Formula (3) above is obtained from this by letting $N = n + \frac{1}{2}p$ and $u = p - 1$.