THE LIMITS OF A MEASURE OF SKEWNESS

By Harold Hotelling and Leonard M. Solomons, *Columbia University*

The measure of skewness

$$s = \frac{\text{mean} - \text{median}}{\text{standard deviation}}$$

is sometimes recommended because of its simplicity. Obviously neither this nor any other statistic can be of much value until something at least is known of its distribution in samples from populations of some plausible form. For populations near the normal form the inefficiency of the median as a statistic of location suggests that the standard error of s may be considerably greater than that of μ_3/σ^3. We know of no investigation of the sampling distribution of s. Apparently even the range is unknown. The object of the present note is to show that s necessarily lies between -1 and 1.

The proof consists of three successive transformations of the sample, each increasing s, which nevertheless in the end remains less than unity.

1. Without loss of generality let us suppose that the median is zero and that the mean $\bar{x}$ is positive. Taking

$$\sigma^2 = \frac{\Sigma(x-\bar{X})^2}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2,$$

n being the number of observations, which we suppose odd, we have

$$s = \bar{x}/\sigma.$$

If a negative observation $-a$ be replaced by zero, the mean is increased by a/n. In the second of the expressions above for σ^2, the mean of the squares is diminished by a^2/n, while on account of the change in the mean, a further subtraction is made. Thus σ diminishes. Hence s increases if we alter the distribution by replacing all the negative observations by zero. The median remains unchanged at zero.

2. Let us further transform this altered distribution by replacing all the positive observations by the mean of these positive quantities. The general mean is left unchanged by this transformation, but the standard deviation is diminished. For, denoting by z the deviation of a positive observation from the general mean, Σz^2 is, for a fixed value of Σz, a minimum when all the z's are equal.

3. Thus the value of s is increased when we replace all the negative observations by the median value 0 and all the positive observations by a fixed quantity, which we may take as unity. Let there be h 0's and k 1's in this distribution. Then $h+k=n$. Moreover, since the median is at 0, $h>k$. The mean is k/n. while

$$n\sigma^2 = h\cdot 0^2 + k\cdot 1^2 - k^2/n = hk/n$$

Hence $s = (k/h)^{\frac{1}{2}}$

in this case in which s is a maximum. Since as just remarked, $h > k$. this is always less than unity, approaching unity when the observations are divided as nearly as is possible equally between the two values.

To go further in a study of the sampling distribution of s is possible only on the basis of special assumptions. The same is true of the somewhat more familiar but less definite measure of skewness.

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

It is clear that there is no limit of the range of this last quantity.

Harold Hotelling
Leonard M^c Solomons