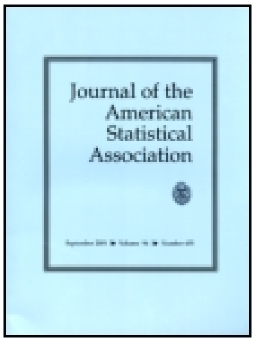

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

A Simplified Formula for Mean Difference

Tseng-Tung Cheng

A SIMPLIFIED FORMULA FOR MEAN DIFFERENCE

By Tseng-Tung Cheng
National Amoy University, China

Some statisticians[1] have advocated the use of the quantity called "mean difference" in place of the commonly used mean deviation or standard deviation as a measure of the tendency of the spreading of a set of statistical data. The formula for this quantity as given in statistical books is rather long and not amenable to routine computation. By using a simple formula from the calculus of finite differences, a new relation for the computation of the sum of the differences between any two values of a set of N items when the latter are grouped into class intervals having the same length is obtained. By definition, this sum divided by the number of pairs of items in the set is their mean difference.

Let n = number of class intervals into which the set of N items is grouped;

s = length of each class interval;

f_k = frequency in the kth interval;

x_k = mid-value of the kth interval;

$F_i = \sum_{k=1}^{i} f_k$ = sum of the number of items in the first i class-intervals;

N = total number of items in all the intervals $= F_n$;

m = mean difference;

$C_2^N = \frac{1}{2}N(N-1)$ = number of combination of N things taken two at a time.

Then by definition,

$$C_2^N \times m = \sum_{i=2}^{n} \sum_{j=1}^{i-1} (x_i - x_j) f_i f_j = \sum_{i=2}^{n} \sum_{j=1}^{i-1} (i - j) s f_j f_i. \qquad (1)$$

As is well-known in the calculus of finite differences,

$$\sum_{j=m}^{p} U_j V_j = U_{p+1} \sum_{j=m}^{p} V_j - \sum_{j=m}^{p} \Delta U_j \sum_{k=m}^{j} V_k, \qquad (2)$$

Editor's Note.—This and the following paper have come by devious routes to the Journal office. The Editor has found it very interesting to realize that, after seven long years of war, statisticians in China can still make worthy contributions to statistical methodology.

It is regretted that most details of the story are missing. The following quotation from a letter can but whet one's appetitite. "His University fled from the Coast at the approach of the Japanese, but only a couple hundred miles west to Chanting which, in common with most parts of China a few miles from railroads and waterways, has never been in Japanese hands—though it has been bombed by them."

[1] Bowley, *Elements of Statistics*, p. 114.

where ΔU_j denotes the first difference of U_j, i.e. $U_{j+1}-U_j$. Now taking $U_j=i-j$, $V_j=f_j$, we find $U_i=0$, $\Delta U_j=-1$, and

$$\sum_{j=1}^{i-1}(i-j)f_j = U_i\sum_{j=1}^{i-1}f_j - \sum_{j=1}^{i-1}\Delta U_j \sum_{k=1}^{j} f_k = \sum_{j=1}^{i-1}\sum_{k=1}^{j} f_k = \sum_{j=1}^{i-1} F_j. \tag{3}$$

By substitution, equation (1) becomes

$$C_2^N \times m = \sum_{i=2}^{n}\sum_{j=1}^{i-1} F_j f_i s. \tag{4}$$

In (4), by interchanging the order of summation and remembering that the starting value for j is 1, that the final value for i is n, while i is greater than j by 1 at the least, we obtain

$$\begin{aligned} C_2^N \times m &= \sum_{j=1}^{n-1}\sum_{i=j+1}^{n} f_i F_j s = \sum_{j=1}^{n-1}(F_n - F_j)F_j s \\ &= \left(N\sum_{j=1}^{n-1}F_j - \sum_{j=1}^{n-1}F_j^2\right)s, \end{aligned} \tag{5}$$

whence

$$\text{the mean difference, } m = \frac{\left[N\sum_{j=1}^{n-1}F_j - \sum_{j=1}^{n-1}F_j^2\right]s}{C_2^N}. \tag{6}$$

It should be pointed out that the quantity M^2 defined below as the mean of the squares of the differences between any two values of a set of N items bears a definite relation to the standard deviation of these N items as mentioned by Bowley.[2] This is readily shown algebraically as follows. Let

M^2 denote the mean square difference;
X_i and X_j, any two values in the set;
A, the arithmetic average; and
N, the total number of items in the set.

Then by definition

$$\begin{aligned} M^2 &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i - X_j)^2}{N(N-1)} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i - A + A - X_j)^2}{N(N-1)} \\ &= \frac{2N\sum_{i=1}^{N}(X_i - A)^2}{N(N-1)}. \end{aligned} \tag{7}$$

[2] *Loc. cit.*, p. 115.

Now the standard deviation σ is defined by $N\sigma^2=\sum_{i=1}^{N}(X_i-A)^2$, and so

$$M^2 = 2N\sigma^2/(N-1), \quad \text{or} \quad M = \sqrt{2N/(N-1)}\sigma. \tag{8}$$

On account of this relation it appears that whatever merit the quantity m called mean difference may possess is equally shared by the standard deviation σ or its analogue, the root mean square difference M.

In concluding the writer wishes to express his gratitude to President A.P.-T. Sah of the National Amoy University for valuable suggestions.