



American Society for Quality

Note on a Method for Calculating Corrected Sums of Squares and Products

Author(s): B. P. Welford

Source: *Technometrics*, Vol. 4, No. 3 (Aug., 1962), pp. 419-420

Published by: American Statistical Association and American Society for Quality

Stable URL: <http://www.jstor.org/stable/1266577>

Accessed: 17/03/2010 12:50

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=astata>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association and American Society for Quality are collaborating with JSTOR to digitize, preserve and extend access to *Technometrics*.

<http://www.jstor.org>

Notes

Note on a Method for Calculating Corrected Sums of Squares and Products

B. P. Welford

*Imperial Chemical Industries Limited, Pharmaceuticals Division.
Alderley Park, Macclesfield, Cheshire, England.*

In many problems the "corrected sum of squares" of a set of values must be calculated i.e. the sum of squares of the deviations of the values about their mean. The most usual way is to calculate the sum of squares of the values (the "crude" sum of squares) and then to subtract a correction factor (which is the product of the total of the values and the mean of the values). This subtraction results in a loss of significant figures and if a large set of values is being handled by a computer, this can result in a corrected sum of squares which has many fewer, accurate significant figures than the computer uses in calculations.

Various alternative schemes are available to combat this. One method is to scale the values to an arbitrary origin which is approximately equal to the mean: if successful, this will reduce the loss in significant figures. An alternative method is to first calculate the mean and then sum the powers of the deviations from the mean. This involves each value being considered twice: first in evaluating the mean and then when calculating its deviation from the mean. If the set of values is large and is being handled by a computer this can involve either storing the data in a slow speed store or reading the same data into the computer twice. A third method which is less cumbersome than either of these is outlined below.

The basis of the method is an iteration formula for deriving the corrected sum of squares for n values from the corrected sum of squares for the first $(n - 1)$ of these. We are given a set of x_i 's ($i = 1, \dots, k$), for which we require the corrected sum of squares.

$$S = \sum_{i=1}^k (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \sum_{i=1}^k x_i/k$$

We define

$$m_n = 1 \sum_{i=1}^n x_i/n \quad n = 1, \dots, k$$

and

$$S_n = \sum_{i=1}^n (x_i - m_n)^2 \quad n = 1, \dots, k$$

Thus

$$S_k = S$$

The following identities hold:—

$$m_n = \frac{n-1}{n} m_{(n-1)} + \frac{1}{n} x_n \quad (1)$$

For

$$i < n, \quad x_i - m_n = x_i - m_{(n-1)} - \frac{1}{n} (x_n - m_{(n-1)}) \quad (2)$$

$$x_n - m_n = \frac{n-1}{n} (x_n - m_{(n-1)}) \quad (3)$$

for $n = 1, 2, \dots, k$

$$\begin{aligned}
 \therefore S_n &= \sum_{i=1}^n (x_i - m_n)^2 \\
 &= \sum_{i=1}^{n-1} \left[(x_i - m_{(n-1)}) - \frac{1}{n} (x_n - m_{(n-1)}) \right]^2 \\
 &\quad + \left(\frac{n-1}{n} \right)^2 (x_n - m_{(n-1)})^2 \\
 &= \sum_{i=1}^{n-1} (x_i - m_{(n-1)})^2 + \left[\frac{n-1}{n^2} + \frac{(n-1)^2}{n^2} \right] (x_n - m_{(n-1)})^2 \\
 &= \underline{S_{(n-1)} + \left(\frac{n-1}{n} \right) (x_n - m_{(n-1)})^2} \quad I
 \end{aligned}$$

Using this formula, the corrected sum of squares is computed using for each value its deviation from the mean of all the previous values. At no stage are significant figures lost and each value is only used once and need not be stored. Formula I is similar in form to that derived by Box and Hunter (1959) for computing the change in residual sum of squares after each cycle of an 'evolutionary operation' design.

A similar formula can be derived for calculating corrected sums of products, viz.

$$\begin{aligned}
 S_n &= \sum_{i=1}^n (x_i - m_n)(y_i - m'_n) \\
 &= \underline{S_{(n-1)} + \left(\frac{n-1}{n} \right) (x_n - m_{(n-1)})(y_n - m'_{(n-1)})} \quad II
 \end{aligned}$$

where

$$m'_n = \sum_{i=1}^n y_i/n$$

Similar formulae to I can also be derived for the iterative calculation of corrected sums of higher powers of values although these are a little more complex since they involve the corrected sums of the lower powers. However, since these corrected sums would usually be required for calculating moments about means and one would generally require all moments up to a certain order, this would not be a drawback.

Define

$$S_n^{(r)} = \sum_{i=1}^n (x_i - m_n)^r \quad n = 1, \dots, k,$$

and $S_k^{(r)} = S^{(r)}$, the required corrected sum of r th powers of deviations about the mean.

Then, using identities (2) and (3), it can be shown that

$$\underline{S_n^{(r)} = \sum_{i=0}^r {}^r C_i (1/n)^{r-i} (m_{(n-1)} - x_n)^{r-i} S_{(n-1)}^{(i)} + \left(\frac{n-1}{n} \right)^r (x_n - m_{(n-1)})^r} \quad III$$

Remembering that

$$S_n^{(0)} = n \quad \text{for } n = 1, \dots, k$$

$$S_n^{(1)} = 0 \quad \text{for } n = 1, \dots, k$$

it is easily verified that III reduces to I when $r = 2$.

REFERENCE

G. E. P. BOX AND J. S. HUNTER, "Condensed calculations for Evolutionary Operation Programs," *Technometrics* Vol. 1, No. 1, February 1959, pp. 77-95.