JOURNAL OF COMPUTATIONAL PHYSICS 77, 537–539 (1988)

Note

A Note on the Progressive Calculation of the Mean and Variance

Direct calculations of means and variances from the definitions

$$m = \sum_{r=1}^{i} x_r / i$$

$$sx = \sum_{r=1}^{i} (x_r - m)^2 / (i-1)$$

are usually introduced in statistics texts and are followed by the suggestion that

$$sx = \left[\sum_{r=1}^{i} x_r^2 - \left(\sum_{r=1}^{i} x_r \right)^2 \Big/ i \right] \Big/ (i-1)$$

is more economical both for hand and for machine calculation. This is because only a single pass through the data is required for its evaluation as against the two passes required for the original form. It is often not explained that the latter formula may be dangerous when the two quantities in [ ] are large and nearly equal. In this case most micro-computers, working to single precision, may give totally incorrect values for sx_i.

Recently modified formulae have appeared [1] which retain the advantages of single pass calculation and have the added advantage of giving a running update of m and sx if required. These are

$$m_{i+1} = (i \cdot m_i + x)/(i+1)$$

$$sx_{i+1} = sx_i + i \cdot (x - m_i)^2/(i+1)$$

with $m_0 = sx_0 = 0$.

For grouped data, variants of the "classical" equations given at the start of this note, are well known [2]. However, we have been unable to find grouped data variants of the above iterations in any of the literature available to us and were thus constrained to develop the following expressions which may be of interest to others. The necessary algebra is fairly elementary and will not be reproduced here. If j is the group size of variable x, the result is

$$sx_{i+j} = sx_i + i \cdot j \cdot (x - m_i)^2/(i+j)$$

$$m_{i+j} = (i \cdot m_i + j \cdot x)/(i+j).$$

0021-9991/88 $3.00

It may be of interest to note that the cross product expression

$$sxy = \sum_{r=1}^{i} (x_r - m_x)(y_r - m_y)$$

has the progressive form

$$sxy_{i+1} = s_i + i \cdot (x - mx_i)(y - my_i)/(i+1)$$

and that, for grouped data (which is probably rarely useful), we have shown that

$$sxy_{i+j} = p_i + i \cdot j \cdot (x - mx_i)(y - my_i)/(i+j).$$

The following programme fragment, in MICROSOFT (T.M.) BASIC illustrates the above results:

PROGRAMME FOR PROGRESSIVE EVALUATION OF MEANS AND PRODUCT SUMS

```
 10 REM EACH LINE OF FILE F$ IS ASSUMED TO CONTAIN X Y J
 20 INPUT "ENTER DRIVE:FILENAME";F$:OPEN "I", 1, F$
 30 MX = 0:MY = 0:SX = 0:SY = 0:SXY = 0:I = 0
 40 WHILE NOT EOF(1)
 50 INPUT#1, X, Y, J
 60 SX = SX + I*J*(X - MX)*(X - MX)/(I + J):SY = SY + I*J*(Y - MY)*
    (Y - MY)/(I + J)
 70 SXY = SXY + I*J*(X - MX)*(Y - MY)/(I + J)
 80 MX = (I*MX + J*X)/(I + J):MY = (I*MY + J*Y)/(I + J)
 90 I = I + J
100 WEND:CLOSE(1)
110 PRINT "MEAN X = ";MX, "ST · DEVIATION (N - 1) FORM = ";
    SQR(SX/(I - 1)):PRINT
120 PRINT "MEAN Y = ";MY, "ST · DEVIATION (N - 1) FORM = ";
    SQR(SY/(I - 1)):PRINT
130 PRINT "CORRELATION COEFFICIENT = ";SXY/SQR(SX*SY)
140 END
```

BASIC PROGRAMME TO CALCULATE MEANS, VARIANCES, AND CORRELATION

REFERENCES

1. D. COOKE, A. H. CRAVEN, AND G. M. CLARKE, *Basic Statistical Computing* (Edward Arnold, London, 1984), pp. 53–56.

2. D. Cooke, A. H. Craven, and G. M. Clarke, *Basic Statistical Computing* (Edward Arnold, London, 1984), p. 63, Ex. 8.
3. D. Cooke, A. H. Craven, and G. M. Clarke, *Basic Statistical Computing* (Edward Arnold, London, 1984), p. 5.

Received: February 13, 1987

Andrew D. Booth

The Institute of Ocean Sciences,
9860 West Saanich Road,
Sidney, B. C., Canada V8L4B2

Ian J. M. Booth[1]

Department of Physics,
Simon Fraser University, Burnaby, B.C., Canada

[1] Current address: University of British Columbia, Vancouver, B.C., Canada.