ON IMPROVING EQUATION REFERENCES IN MATHEMATICAL WRITING

by

August, 1977

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

In mathematical writing, simplification of a formula usually requires reference back to equations developed earlier. For example, suppose that in presenting sample moments we have already observed that for any constant θ

$$\sum_{i=1}^{n} \theta = n\theta \ , \tag{1}$$

and have also defined

$$\bar{x} = \sum_{i=1}^{n} x_i/n \tag{2}$$

and now wish to develop the result $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. We could proceed as follows:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \ . \tag{3a}$$

From (1) and (2), equation (3a) becomes

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \ . \tag{4a}$$

BU-625-M (Revised)

In this development the simplification of (3a) by using (1) and (2) is noted by temporarily halting the development at (3a) and using a new sentence. After this interruption the algebra restarts by writing down the left-hand side of (3a) again and then continuing. An alternative development is

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \Sigma(x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \qquad (3b)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2, \quad \text{from using (1) and (2),} \qquad (4b)$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 .$$

Here there is no interruption, but after the algebra in (4b) there is now the tagged-on phrase "from using (1) and (2)". This achieves the same thing as the interruptive sentence used between (3a) and (4a), but logically it is in a most unsatisfying position. Since it indicates how the steps from (3b) to (4b) are made, it should come before and not after the algebra in (4b).

Both of these ways of referring back to already-stated equations are used in a variety of ways in mathematical writing. And yet both of them have deficiencies insofar as easy readability is concerned. The first, the interruptive sentence, interrupts the flow of the algebra and the second, the tagged-on phrase, is logically unsatisfying. Furthermore, both methods require the reader to look back in his reading to where the referred-to equations are located. These deficiencies are greatly exacerbated for complicated algebra, particularly when references back are numerous and when some are several pages, or maybe chapters, earlier in the work concerned. (Readers must sometimes wish they had more than five fingers on their left hands to keep some books open at the many pages referred to.)

To avoid these difficulties, a different style of referencing is now suggested: parenthetical sentences, adapted from the convention of parenthetical numbers for equation labels. For example

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \qquad (3c)$$

[Use (2): $\sum_{i=1}^{n} x_i = n\bar{x}$.]

[Use (1): $\sum_{i=1}^{n} \bar{x}^2 = n\bar{x}^2$.]

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \; . \qquad (4c)$$

The parenthetical sentences between (3c) and (4c) are logically in the correct position, and although they create a break in the spacing of the main development it is only a spatial interruption, not a linguistic one, with no need for repeating the left-hand side of the main development, as in (4a). Furthermore, such sentences can give not only the label of the equation being referred to but also the equation itself or some "obvious" consequence of it. Not only does this reduce the reader's need for a many-fingered left hand but it can also show the particular use of an earlier result that is now pertinent. Ofttimes this may entail repetition – but that is no disadvantage, it aids the learning process.