THE RECURRENCE FORMULAE FOR MEANS AND VARIANCES

BU-764-M by December, 1981

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

Abstract

Derivation and illustration is shown of the simple and well-known formulae relating the mean (and variance) of n observations to that of those n plus one more.

Introduction

Walter (1981) illustrates how averaging successive averages with a new observation leads to weighted averages rather than to successive values of the arithmetic mean. An easy and useful alternative is the well-known and simple recurrence formula for calculating from the mean of n observations the mean of those n plus one more observations. A corresponding recurrence formula is also available for the sum of squares (corrected for the mean), and hence for the unbiased estimator of variance available from the sum of squares.

Formula for the mean

Let

$$M_n = (x_1 + x_2 + \dots + x_n)/n$$

be the arithmetic mean of the n observations $x_1, x_2, \dots, x_n$. Then for those n observations plus one more, x_{n+1}, the mean is

---

Paper No. BU-764-M in the Biometrics Unit.

$$M_{n+1} = (x_1 + x_2 + \ldots + x_n + x_{n+1})/(n + 1)$$
$$= (nM_n + x_{n+1})/(n + 1)$$
$$= [(n + 1)M_n + x_{n+1} - M_n]/(n + 1)$$

Hence

$$M_{n+1} = M_n + (x_{n+1} - M_n)/(n + 1) \quad (1)$$

By this formula the mean of the n + 1 observations is obtained from that of the n observations by the simple expedient of adding the difference (divided by n + 1) between the (n + 1)'th observation and the mean of the n observations - a difference that can be called that between the new observation and the old mean.

Example Suppose a student gets scores of 66, 72, 63, 83 and 66 on five successive tests. Table 1 shows these scores, their cumulative totals, and successive means after each test calculated directly and by formula (1).

TABLE 1 CALCULATING SUCCESSIVE MEANS

n + 1	x_{n+1}	Cumulative total of x_{n+1}	Mean M_{n+1}	Calculation of (1) $M_{n+1} = M_n + (x_{n+1} - M_n)/(n + 1)$
1	66	66	66	66 = 0 + (66 - 0)/1
2	72	138	69	69 = 66 + (72 - 66)/2
3	63	201	67	67 = 69 + (63 - 69)/3
4	83	284	71	71 = 67 + (83 - 67)/4
5	66	350	70	70 = 71 + (66 - 71)/5

Formula for estimated variance

The sum of squares (corrected for the mean) of the n observations is

$$S_n^2 = (x_1 - M_n)^2 + (x_2 - M_n)^2 + \ldots + (x_n - M_n)^2$$

which also has the well-known equivalent

$$S_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2 - nM_n^2 .$$

Hence for n + 1 observations

$$S^2_{n+1} = x_1^2 + x_2^2 + \ldots + x_n^2 + x_{n+1}^2 - (n+1)M^2_{n+1} \tag{2}$$

$$= x_1^2 + x_2^2 + \ldots + x_n^2 - nM_n^2 + nM_n^2 + x^2_{n+1} - (n+1)M^2_{n+1}$$

$$= S_n^2 + nM_n^2 + x^2_{n+1} - (n+1)M^2_{n+1} \ . \tag{3}$$

But, from (1),

$$x_{n+1} = (n+1)M_{n+1} - nM_n \ ,$$

so that in (3)

$$S^2_{n+1} = S_n^2 + nM_n^2 + [(n+1)M_{n+1} - nM_n]^2 - (n+1)M^2_{n+1}$$

which, after a little simplification, reduces to

$$S^2_{n+1} = S_n^2 + n(n+1)(M_{n+1} - M_n)^2 \ . \tag{4}$$

Thus the new sum of squares, S^2_{n+1} , is simply the old sum of squares, S_n^2, plus $n(n + 1)$ times the square of the difference between the new and old means. This represents a very simple calculation, once (1) has been made, compared to using (2) de nouveau. Notice, too, that (4) shows that S^2_{n+1} always exceeds S^2_n except that the two are equal only when $M_{n+1} = M_n$, i.e., when $x_{n+1} = M_n$.

Example (continued) For the same example as used in Table 1, Table 2 shows calculation of S^2_{n+1} from (2) and from the recurrence formula (4).

TABLE 2 CALCULATING SUCCESSIVE SUMS OF SQUARES

$n + 1$	x^2_{n+1}	Cumulative total of x^2_{n+1} (a)	$(n+1)M^2_{n+1}$ (b)	Equation (2) for S^2_{n+1} (a)-(b)	Calculation of (4): $S^2_{n+1} = S^2_n + n(n+1)(M_{n+1} - M_n)^2$
1	4356	4356	4356	0	$0 = 0 + 0(1)(66-0)^2$
2	5184	9540	9522	18	$18 = 0 + 1(2)(69-66)^2$
3	3969	13509	13467	42	$42 = 18 + 2(3)(67-69)^2$
4	6889	20398	20167	234	$234 = 42 + 3(4)(71-67)^2$
5	4356	24754	24500	254	$254 = 234 + 4(5)(70-71)^2$

Formula for variance estimator

When $x_1, x_2, \ldots x_n$ represent a simple random sample

$$\hat{\sigma}_n^2 = S_n^2/(n - 1)$$

is an unbiased estimator of variance. Applied to (4) this gives

$$\hat{\sigma}_{n+1}^2 = S_{n+1}^2/n \quad (5)$$

$$= S_n^2/n + (n + 1)(M_{n+1} - M_n)^2$$

$$= (1 - 1/n)\hat{\sigma}_n^2 + (n + 1)(M_{n+1} - M_n)^2 \; . \quad (6)$$

This is a recurrence formula for calculating successive values of $\hat{\sigma}^2$ as additional observations become available, one by one. Notice that whereas from (4), $S_{n+1}^2 \geq S_n^2$, formula (6) indicates that

$$\hat{\sigma}_{n+1}^2 > \sigma_n^2 \text{ only when } n(n + 1)(M_{n+1} - M_n)^2 > \hat{\sigma}_n^2 \; .$$

Example (continued)

Table 3 shows calculation of $\hat{\sigma}_{n+1}^2$ (for $n + 1 \geq 2$ because $\hat{\sigma}_1{}^2 = 0$) from both its definition, (5), and from its recurrence formula, (6).

TABLE 3 CALCULATING SUCCESSIVE VALUES OF $\hat{\sigma}^2$

n + 1	$\hat{\sigma}_{n+1}^2 = S_{n+1}^2/n$	Calculation of (6): $\hat{\sigma}_{n+1}^2 = (1 - 1/n)\hat{\sigma}_n^2 + (n + 1)(M_{n+1} - M_n)^2$
2	18/1 = 18	18 = (1-1/1) 0 + 2(69-66)^2
3	42/2 = 21	21 = (1-1/2)18 + 3(67-69)^2
4	234/3 = 78	78 = (1-1/3)21 + 4(71-67)^2
5	254/4 = 63½	63½ = (1-1/4)78 + 5(70-71)^2

Summary

When one new observation is to be included among n observations already at hand, calculation of the mean, sum of squares and unbiased estimate of variance for the n + 1 observations can be made from the corresponding values for the n observations by using the recurrence formulae (1), (4) and (6), without any need for direct recalculation of the sum or sum of squares.

Reference

Walter, M. (1981) How to get a higher grade: beware of averages. Teaching Statistics, 3, 77-79.