THE UNIVERSITY OF TEXAS AT AUSTIN
SCHOOL OF INFORMATION

MATHEMATICAL NOTES FOR LIS 397.1
INTRODUCTION TO RESEARCH IN
LIBRARY AND INFORMATION SCIENCE

Ronald E. Wyllys
Last revised: 2003 Jan 15

DEFINITIONAL vs. CALCULATIONAL FORMS
of the
STANDARD DEVIATION and VARIANCE FORMULAS

The formulas for the standard deviation and variance have several different-looking, but arithmetically equivalent, forms. By "arithmetically equivalent" I mean that the different-looking formulas nevertheless yield the same numerical results.

One set of formulas, which I like to call the "definitional" forms, display directly the definitions of the standard deviation and variance. These formulas show how the standard deviation and variation are constructed according to the mathematical theory of statistics. Another set of formulas, which you will also encounter and which I like to call the "calculational forms," are designed for ease of use with simple, four-function calculators in today's world (and were designed originally for use on the manually and electrically powered mechanical calculators available during the 1910s through the 1960s).

You should note, however, that moderately sophisticated electronic calculators today make it possible for you not even to have to use the calculational formulas. These calculators allow you to enter the observed values directly into the calculator, i.e., into the internally stored calculational formulas, by simply keying in each observed value from a set of observations and then using one or two additional keystrokes to cause the calculator to display the value of the standard deviation or variance. (It is calculators with at least this degree of sophistication that I recommend for use by students in LIS 397.1.)

Since many texts, including that of Stephens, merely assert the equivalence of the definitional and calculational formulas, this note shows how the two types of formulas are algebraically related. This note is intended only for those students interested in understanding the details of the algebraic relationship, and the numerical equivalence, of the two types. There is no need for students in LIS 397.1 to memorize any of the discussion below; the discussion is intended merely to satisfy the curiosity of those who may be wondering about the equivalence.

We illustrate the steps by working with the population variance. The first task is to show how formula (3.4) on page 42 of the Stephens text can be transformed into formula (3.8) on page 43.

We start with the equation

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

If we multiply both sides of this equation by N, the equation becomes

$$N\sigma^2 = \sum (x-\mu)^2$$

which we shall call Intermediate Step 1, or IS 1. For the next few steps we will work with only the right-hand side of the above equation. To make clearer what is going on, we also start using explicit indices with the summation sign. By expanding the square, i.e., the parenthesized expression, we have

$$\sum_{i=1}^{N} (x_i-\mu)^2 = \sum_{i=1}^{N} \left(x_i^2 - 2\mu x_i + \mu^2\right) = \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} 2\mu x_i + \sum_{i=1}^{N} \mu^2$$

which we shall call Intermediate Step 2. Now we concentrate briefly on the second term in the right-most expression. Examining it, we find

$$\sum_{i=1}^{N} 2\mu x_i = 2\mu x_1 + 2\mu x_2 + \cdots + 2\mu x_N = 2\mu\left(x_1 + x_2 + \cdots + x_N\right) = 2\mu\sum_{i=1}^{N} x_i$$

which we can call Intermediate Step 3.

Recall that the population mean is defined as

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

from which it follows that

$$\sum_{i=1}^{N} x_i = N\mu$$

We can call this Intermediate Step 4. From Intermediate Steps 3 and 4 it follows that

$$\sum_{i=1}^{N} 2\mu x_i = 2\mu\sum_{i=1}^{N} x_i = 2\mu N\mu = 2N\mu^2$$

We can note also that

$$\sum_{i=1}^{N} \mu^2 = N\mu^2$$

since the left-hand side just amounts to adding μ^2 to itself a total of N times.

Putting these last two expressions into the earlier expression, IS 2, for the expanded square, we see that

$$\sum_{i=1}^{N} \left(x_i - \mu\right)^2 = \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} 2\mu x_i + \sum_{i=1}^{N} \mu^2 = \sum_{i=1}^{N} x_i^2 - 2N\mu^2 + N\mu^2 = \sum_{i=1}^{N} x_i^2 - N\mu^2$$

which we shall call Intermediate Step 5.

Dividing both sides of IS 4 by N, we see that

$$\frac{\sum x}{N} = \mu$$

from which it follows that

$$\mu^2 = \left[\frac{\sum x}{N}\right]^2 = \frac{\left(\sum x\right)^2}{N^2}$$

Multiplying the left-hand and right-hand extremes of the above expression by *N*, we see that

$$N\mu^2 = \frac{\left(\sum x\right)^2}{N}$$

Putting this last step into IS 5 and combining the result with IS 1, we have

$$N\sigma^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{N}$$

which we shall call Intermediate Step 6.

It follows that

$$\sigma^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{N}}{N}$$

In the preceding discussion we have shown how the definitional form of the population variance, σ^2, the very first equation in this note, leads to the calculational form; i.e., we have shown how formula (3.4) on page 42 of the Stephens text can be transformed into formula (3.8) on page 43.

But what about the definitional and calculational forms of the sample variance? We start with the equation

$$s^2 = \frac{\sum \left(x - \overline{x}\right)^2}{n-1}$$

If we multiply both sides of this equation by *n-1*, the equation becomes

$$\left(n-1\right)\sigma^2 = \sum \left(x - \overline{x}\right)^2$$

If we worked through the analogs to Intermediate Steps 2 through 5, these would lead us to

$$\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2\overline{x}x_i + \sum_{i=1}^{n} \overline{x}^2 = \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

In working toward the analog of IS 6, we need to note that since

$$\overline{x} = \frac{\sum x}{n}$$

the analog of IS 6 is

$$\left(n-1\right)s^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

and the final result is

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}$$

Thus we have also shown how the definitional form of the sample variance, s^2, in formula (3.3) on page 42 of the Stephens text can be transformed into the calculational form, formula (3.7) on page 43.