Key Concepts in Descriptive Statistics (Version of 9/01/92)

James H. Steiger

This course deals, for the most part, with classical parametric statistical procedures. These are primarily concerned with inferences about means, variances, covariances, and correlations. Since, ultimately, we shall be so concerned with these quantities, we should try, at the outset, to develop a succinct description (and understanding) of their behavior, both in the population, and in the sample case.

We will begin with the sample statistics. Our goal will be to develop a (small) set of interlocking rules which describe and explain what is important about sample means, variances, covariances, and correlations.

Because the linear transform and the linear combination play such an important role in psychological statistics, we will also pay special attention to the behavior of statistics when sample scores are linearly transformed or linearly combined.

I assume that most students bring basic background in undergraduate behavioral statistics to this discussion. Those who are unfamiliar with the notion of a "summation operator" $\sum$ or subscript notation should review these concepts in an undergraduate text such as Glass and Hopkins.

We now develop our key basic results and definitions.

Result 1. ("First constant rule of summation algebra.") Let a be an algebraic expression which is constant with respect to i. Then

$$\sum_{i=x}^{y} a = (y - x + 1)a$$

(***Comment***. This is simply a restatement of the principle, taught in third grade arithmetic, that adding a number n times is the same as multiplying it by n. Introductory texts in undergraduate behavioral statistics often give a special case of this rule in which x is 1 and y is n. In that case, we have the more familiar result,

$$\sum_{i=1}^{n} a = na$$

Although, in general, we will find this latter result sufficient for our purposes in this course, it is important to remember that a summation need not start with 1 and end with *n*, and so we must be cautious in applying the first constant rule. For example, consider

$$\sum_{i=2}^{4} 2$$

This is not equal to 8, but, rather 6. (Do you see why?)

Result 1 says that, if we have a summation operator governing an expression which, with respect to that summation operator, is constant, we can eliminate that summation sign and replace it with a constant to be multiplied by that expression. This is an important result for simplifying complicated summation expressions. The key to the application of Result 1 is remembering that constants (with respect to a particular summation operator) are not necessarily simple expressions, or expressions without subscripts.

For example, consider the following expression.

$$\sum_{i=1}^{n} \left(Y_j + 4\right) = n\left(Y_j + 4\right)$$

The expression within parentheses is a constant with respect to *i*, even though it contains a subscripted term.

Result 2. ("Second constant rule of summation algebra.") Let a be any algebraic expression which is constant with respect to *i*. Then

$$\sum_{i=1}^{n} aX_i = a\sum_{i=1}^{n} X_i$$

Comments similar to those for Result 1 hold here as well, i.e., a need not be a simple expression.

Result 3. The summation operator may be distributed over addition and subtraction, but not over multiplication or division. For example,

$$\sum_{i=1}^{n}\left(aX_i + bY_i - c\right)$$

$$= \sum_{i=1}^{n} aX_i + \sum_{i=1}^{n} bY_i + \sum_{i=1}^{n} c$$

$$= a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} Y_i + nc$$

However, it is **not true** in general that

$$\sum_{i=1}^{n} \frac{X_i}{Y_i} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i}$$

We now introduce some basic definitions.

Definition 1. The sample mean $\overline{X}_\bullet$ for the *n* scores $X_i,\ i = 1,n$ is defined as

$$\overline{X}_\bullet = \left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i .$$

The sample mean is the simple "arithmetic average" we are all familiar with. The "deviation scores" corresponding to the X_i are defined as follows.

Definition 2. Suppose we have *n* raw scores $X_i,\ i = 1,n$ with sample mean $\overline{X}_\bullet$. The deviation score, denoted d_i (if more that one variable name is used, we will employ a notation like $\underline{dx_i}$) is defined as

$$d_i = X_i - \overline{X}_\bullet$$

Often important theoretical relationships in basic statistics are seen more easily when the statistics are expressed in terms of deviation scores rather than in terms of the raw scores (i.e., the original X_i). "Expressing scores in deviation score form" means to transform them into deviation scores.

We now establish an important basic result about deviation scores.

Result 4. Deviation scores always sum to zero, i.e.,

$$\sum_{i=1}^{n} d_i = 0 .$$

We shall prove this Result using the summation algebra (i.e., Results 1 - 3). This proof will be characteristic of most of the proofs in this course. It will require very little mathematical insight, and will rely instead on a methodical application of the following general approach.

(1) Define the problem and write one side of the equality which is to be shown.

$$\sum_{i=1}^{n} d_i$$

(2) Substitute known definitions, and expand where possible, and where expansion is not obviously counterproductive. (In the present problem we simply substitute Definition 2.)

$$= \sum_{i=1}^{n} \left(X_i - \overline{X}_\bullet \right)$$

(3) Distribute summation signs where possible.

$$= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X}_\bullet$$

(4) Apply the two "constant rules" (i.e., Results 1 and 2).

$$= \sum_{i=1}^{n} X_i - n\overline{X}_\bullet$$

(5) Examine the expression for obvious identities and fruitful substitutions. By examining Definition 1, we can see easily that

$n\overline{X}_\bullet = \sum_{i=1}^{n} X_i$, and that

$$\sum_{i=1}^{n} X_i - n\overline{X}_\bullet = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0$$

so the proof is complete.

Result 4 will turn out to be quite useful is some subsequent proofs. We note in passing the obvious fact that deviation scores always have a zero mean.

We recall, from our discussion of basic issues in measurement theory, that interval measurement is preserved under positive linear transformations, which are defined below.

Definition 3. Given scores X_i, a linear transformation of these scores into a new set of scores Y_i is any transformation of the form

$$Y_i = aX_i + b$$

A linear transformation is a positive linear transform if $a > 0$.

Throughout the course, I will be using positive linear transforms, and I will refer to them (for simplicity) as linear transforms.

We now examine the effect of a linear transformation of a sample mean. We prove the following very important result.

Result 5. Given *n* scores $X_i, i = 1, n$, with sample mean $\overline{X}_\bullet$. These scores are all transformed linearly as $Y_i = aX_i + b$. The sample mean $\overline{Y}_\bullet$ of the transformed scores must satisfy the relationship

$$\overline{Y}_\bullet = a\overline{X}_\bullet + b$$

Proof. Previous results and definitions used to derive each step are given in parenthesis.

$$\overline{Y}_{\bullet} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} Y_i \qquad \text{(Definition 1)}$$

$$= \left(\frac{1}{n}\right)\sum_{i=1}^{n}(aX_i + b) \qquad \text{(Definition 3)}$$

$$= \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} aX_i + \sum_{i=1}^{n} b\right) \qquad \text{(Result 3)}$$

$$= \left(\frac{1}{n}\right)\left(a\sum_{i=1}^{n} X_i + nb\right) \qquad \text{(Results 1 and 2)}$$

$$= a\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i + \left(\frac{1}{n}\right)nb$$

$$= a\overline{X}_{\bullet} + b \qquad \text{(Definition 1)}$$

This result is a very useful one, as we shall see later. It implies that we may calculate the sample mean of a set of transformed scores without actually calculating the scores themselves, provided we know the mean of the untransformed scores. Moreover, by considering $\overline{X}_{\bullet}$ to be the "current" mean of a set of scores, and $\overline{Y}_{\bullet}$ to be the "desired" mean, we can use Result 5 to derive a rule (given as a later Result) for transforming scores into a desired metric.

This Result subsumes some well known basic relations as special cases. For example, if $a = 1$ in the linear transform, then we are simply adding a constant to every X_i. Result 5 thus says that adding a constant to every score in a group increases the mean of that group by the additive constant.

By setting $b = 0$ in the linear transformation, we can examine the effect of multiplying every score in a group by a constant. Result 5 says that, in that case, the sample mean will simply be multiplied by the multiplicative constant *a*.

Some well known relationships in statistics can be derived much more compactly (that they often are in textbooks) by utilizing the following basic Result. This Result describes the effect of a linear transformation on deviation scores.

Result 6. Given the same setup as in Result 5 above, the deviation scores calculated on the transformed scores Y_i will have the following simple relationship to those calculated on the X_i.

$$\underline{dy}_i = a\underline{dx}_i \, .$$

Comment. This result says that the multiplicative constant (a) in the linear transform is carried directly through in the deviation scores, while the additive constant b has no effect. (Does this result surprise you? If it does, reflect a bit on the meaning of deviation scores.)

We now define the sample variance, and examine the effect of a linear transform on the sample variance.

Definition 4. The sample variance s_x^2 is

$$s_x^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \underline{dx}_i^2$$

(*Comment.* The variance of a set of scores is, essentially, their average squared deviation.)

The standard deviation s_x is the square root of the variance.

Using Result 6, we can easily establish the effect of a linear transformation on the variance of a set of scores.

Result 7. Given the same setup as in Result 5, the variance of the Y_i will be related to the variance of the X_i in the following way.

$$s_y^2 = a^2 s_x^2$$

Proof.

$$(n-1)s_y^2 = \sum_{i=1}^{n} \underline{dy}_i^2 \qquad \text{(Rearrange Definition 4)}$$

$$= \sum_{i=1}^{n} (a\underline{dx}_i)^2 \qquad \text{(Result 6)}$$

$$= \sum_{i=1}^{n} (a^2 \underline{dx}_i^2)$$

$$= a^2 \sum_{i=1}^{n} \underline{dx}_i^2 \qquad \text{(Result 2)}$$

This implies that

$$s_y^2 = a^2 \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} \underline{dx}_i^2$$

$$= a^2 s_x^2 \qquad \text{(Definition 4)}$$

and the proof is complete.

In Results 6 and 7, we see that deviation scores, while invariant under additive transformations, are not invariant under multiplicative transformations. Result 7 suggests a way of modifying deviation scores so that they will be invariant under both additive and multiplicative transformations. Specifically,

Definition 5. The *z*-score transformation of a score X_i (denoted $\underline{zx}_i$, or simply z_i if the discussion is concerned with only a single variable) is defined as

$$\underline{zx}_i = \frac{\underline{dx}_i}{s_x}$$

z-scores have some interesting properties. Since they are a multiplicative transform of deviation scores (which, by Result 4, have a mean zero), we see immediately from Result 5 that they must also have a mean of zero. Moreover, since deviation scores have a variance of s_x^2, it follows immediately from Result 7 and Definition 5 that *z*-scores always have a variance of 1. You should prove these results to your own satisfaction as an exercise.

A most important property of *z*-scores is the following:

Result 8. *z*-scores are invariant under linear transformations of raw scores.

Proof. Suppose the *n* scores X_i are converted to Y_i by the linear transformation $Y_i = aX_i + b$. Then, by Result 6,

$$\underline{dy}_i = a\underline{dx}_i$$

But, by Definition 5,

$$\underline{zy}_i = \frac{\underline{dy}_i}{s_y} = \frac{a\underline{dx}_i}{s_y}$$

However, Result 7 implies that $s_y = as_x$, and so

$$\underline{zy}_i = \frac{a\underline{dx}_i}{s_y} = \frac{a\underline{dx}_i}{as_x} = \frac{\underline{dx}_i}{sx} = \underline{zx}_i$$

Hence, we see that the *z*-score has not changed, and the proof is complete.

Reflection. We see that *z*-scores, are, in effect, a "metric-free" method for expressing interval data, in the sense that no matter which arbitrary linear rescaling of a set of data we imply, the *z*-scores will not change. Here is a question which I would like you to consider for a few moments. We could define the term "*z*-scores" to refer to any set of scores which have a mean of zero and a variance of 1. Suppose that "*t*-scores" had been defined as scores with a mean of 100 and a standard deviation of 16. Would "*t*-scores" still be "metric-free" in the same sense that "z-scores" are?

Application. Results 5 and 7 may be combined to provide a simple formula for determining the constants *a* and *b* for linearly transforming a set of scores X_i, with known mean $\overline{X}\bullet$ and standard deviation s_x, into new scores Y_i with desired mean $\overline{Y}\bullet$ and desired standard deviation s_y. Such a transformation can be useful in, for example, rescaling course grades into a desired metric.

Since, from Result 7,

$$s_y = as_x,$$

it follows immediately that the desired multiplicative constant is

$$a = \frac{s_y}{s_x}.$$

From Result 5, we have that

$$b = \overline{Y}_\bullet - a\overline{X}_\bullet = \overline{Y}_\bullet - \left(\frac{s_y}{s_x}\right)\overline{X}_\bullet$$

Hence, our desired transform is

$$Y_i = a\overline{X}_\bullet + b = \left(\frac{s_y}{s_x}\right)X_i + \overline{Y}_\bullet - \left(\frac{s_y}{s_x}\right)\overline{X}_\bullet,$$

which may be rewritten as

$$Y_i = s_y\left(\frac{X_i - \overline{X}_\bullet}{s_x}\right) + \overline{Y}_\bullet.$$

Since the term within parentheses is simply $\underline{zx}_i$, we may write

$$Y_i = s_y \underline{zx}_i + \overline{Y}_\bullet.$$

Hence, in transforming scores to a desired metric, we may view the transformation process as proceeding in two stages, the first of which carries the raw score into their *z*-score equivalents, while the second transforms the *z*-scores into the final, desired, metric. Since the *z*-scores have a mean of zero and a variance of 1, they may (obviously as a consequence of Results 5 and 7) be converted into scores with a desired mean and standard deviation by multiplying by the desired standard deviation and adding the desired mean.

So far, our discussion has concentrated on transformations of scores on a single variable. However, in practice, we will often be interested in situations involving scores on several variables for each subject. When a group of subjects is measured on several variables (i.e., the experiment involves “repeated measurements”), we will be interested in how the various measurements are related, or, more precisely, how they “vary together”, or “covary” across subjects. An elementary statistical quantity which expresses how variables covary is their “covariance,” defined as follows:

Definition 6. Given n pairs of observations X_i and Y_i, the "covariance of X and Y", denoted s_{xy}, is

$$s_{xy} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \underline{dx}_i \underline{dy}_i .$$

The covariance is roughly the average cross-product of deviation scores. Since the covariance is a measure on deviation scores, it is invariant under purely additive transformations, but it is not invariant under multiplicative transformations. The covariance is thus, except for its sign, an arbitrary quantity for interval data.

It is worthwhile remembering that the variance of a single variable is simply its covariance with itself. Variance is thus a special case of covariance. The practical implication of this for students of statistics is that variance and covariance can both be calculated from the same formula.

We also recall the well-known "computational formulas" for covariance and variance, which you should prove, using summation algebra Results 1-3, as an exercise.

Result 9. The covariance s_{xy} may be computed as

$$s_{xy} = \left(\frac{1}{n-1}\right)\left(\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right)$$

which implies in turn that the variance of a variable may be computed as

$$s_x^2 = \left(\frac{1}{n-1}\right)\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)$$

Proof. (To be supplied by the student as an exercise.)

The covariance s_{xy}, as I mentioned above, is not a scale-free measure of covariation. However, we can render the covariance scale-free by redefining it in terms of z-scores, rather than deviation scores. If we do so, we obtain the well known Pearson product-moment correlation coefficient.

***Definition** 7*. The Pearson product moment correlation coefficient, r_{xy}, is defined as

$$r_{xy} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \underline{zx}_i \underline{zy}_i$$

Comment. This formula may be unfamiliar to you if you have only been exposed to a very basic course in undergraduate behavioral statistics. However, it is conceptually very revealing, in that it allows one to see immediately that r_{xy} is invariant under linear transformations of either the X_i or the Y_i. (***Why? What are some implications of this fact?)***

Alternative computational formulas for r_{xy} are derived as follows. First, note

Result 10. Definition 7 may be rewritten as

$$\mathrm{r_{xy}} = \frac{\mathrm{s_{xy}}}{\mathrm{s_x s_y}}$$

Proof.

$$r_{xy} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \frac{\underline{dx}_i}{s_x}\frac{\underline{dy}_i}{s_y} \qquad \text{(Definition 5)}$$

$$= \left(\frac{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \underline{dx}_i \underline{dy}_i}{s_x s_y}\right)$$

$$= \frac{s_{xy}}{s_x s_y} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}}$$

Note that by substituting Result 9 into Result 10, and then multiplying numerator and denominator by *n*, one may obtain the following well-known computational formula for the sample correlation.

$$r_{xy} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

The student should attempt to derive this formula as an exercise.

We now introduce a concept, the linear combination (or weighted sum), which will prove to be very useful in the remainder of the course.

Definition 8. Given *J* variables $X_1, X_2, \ldots X_J$. A variable *K* is a linear combination of the X_i variables if it may be written in the form

$$K = \sum_{j=1}^{n} c_j X_j$$

where the c_j are constants called *linear weights*.

Many familiar quantities in statistics are linear combinations. For example, the sample mean $\overline{X}_\bullet$ is a linear combination of the *n* scores on which it is based, since it may be written

$$\overline{X}_\bullet = \sum_{i=1}^{n} \left(\frac{1}{n}\right) X_i .$$

In this case, all the linear weights are equal to $\left(\frac{1}{n}\right)$.

We are familiar with many other such examples. Grade point averages at most universities are linear combinations of the individual course grades, in which the linear weights are the number of credit hours for each course, divided by the total number of credit hours.

Many interesting statistical hypotheses can be expressed as hypotheses about linear combinations. For example, consider the hypothesis that two population means are equal, that $\mu_1 = \mu_2$. This may be expressed in the form

$$\mu_1 - \mu_2 = 0$$

The left side of the expression is a linear combination in which the linear weights are 1 and –1.

Most of your grade in this course will be based on a linear combination of your z-scores on a number of tests and exercises.

The linear combination is a pervasive, and important concept in basic behavioral statistics. We will find that our study of statistics can be streamlined considerably by developing and applying an understanding of the statistical behavior of linear combinations.

We now develop some results on linear combinations. We will need to employ double subscript notation to prove these results. Suppose we have scores for n subjects on J variables, and these scores are symbolized as X_{ij}. The scores for the jth variable have sample mean $\overline{X}_{\bullet j}$ and sample variance s_j^2. For example, these scores could represent the exam grades of my n Psychology 100 students on the J exams given this year.

Definition 9. Suppose we linearly recombine the scores on the J variables to obtain a new variable K. The scores for the n subjects on K will be

$$K_i = \sum_{j=1}^{J} c_j X_{ij}$$

Under the above setup, we find that the sample mean of the linear combination scores K_i is a linear combination of the sample means on the X variables, i.e.,

Result 11. The sample mean $\overline{K}_{\bullet}$ of the linear combination scores is

$$\overline{K}_{\bullet} = \sum_{j=1}^{J} c_j \overline{X}_{\bullet j}$$

Proof.

$$\overline{K}_{\bullet} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} K_i \qquad \text{(Definition 1)}$$

$$= \left(\frac{1}{n}\right)\sum_{i=1}^{n}\sum_{j=1}^{J} c_j X_{ij} \qquad \text{(Definition 9)}$$

$$= \left(\frac{1}{n}\right)\sum_{j=1}^{J}\sum_{i=1}^{n} c_j X_{ij} \qquad \text{(Why is this step correct?)}$$

$$= \sum_{j=1}^{J} c_j \left(\frac{1}{n}\right)\sum_{i=1}^{n} X_{ij} \qquad \text{(Result 2)}$$

$$= \sum_{j=1}^{J} c_j \overline{X}_{\bullet j}$$

and the proof is complete.

We now examine the behavior of deviation scores under linear combination of the raw scores.

Result 12. Given a linear combination K on J variables as defined in Definition 9, the deviation scores $\underline{dk}_i$ are related to the deviation scores of the original variables by the formula

$$\underline{dk}_i = \sum_{j=1}^{J} c_j \underline{dx}_{ij}$$

Proof. (follows rather easily from the definition and Result 11, and we leave it to the student as an exercise.)

Result 13.

$$\left(\sum_{j=1}^{J} x_j\right)^2 = \sum_{j=1}^{J} x_j^2 + 2\sum_{j=2}^{J}\sum_{k=1}^{J-1} x_j x_k$$

Proof. Consider the special case $J = 3$. Here the desired result is $(x_1 + x_2 + x_3)^2$. This result may be expressed as the sum of all the elements in the (3x3) square matrix whose entries are the cross products of marginal terms x_1, x_2, x_3. Specifically, the square matrix is

	x_1	x_2	x_3
x_1	x_1^2	$x_1 x_2$	$x_1 x_3$
x_2	$x_1 x_2$	x_2^2	$x_2 x_3$
x_3	$x_1 x_3$	$x_2 x_3$	x_3^2

Now, the terms on the diagonal are, when summed up, in the form given by the first term on the right of Result 13. Note that each off-diagonal term is actually repeated twice. Hence, the sum of the remaining non-diagonal terms is of the form given by the second term on the right of Result 13. That the result may be extended to any number of terms is obvious.

Result 14. Given a linear combination *K* on *J* variables, the variance of the linear combination is given by the formula

$$s_K^2 = \sum_{j=1}^{J} c_j^2 s_j^2 + 2\sum_{j=2}^{J} \sum_{k=1}^{J-1} c_j c_k s_{jk}$$

Proof. Using Result 12, we may write

$$s_K^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n} \underline{dk}_i^2$$

$$= \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(\sum_{j=1}^{J} c_j \underline{dx}_{ij}\right)^2$$

Now, consider the rightmost term. The term is of the form $\left(\sum x_j\right)^2$, and so Result 13 applies. Hence

$$s_K^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(\left[\sum_{j=1}^{J} c_j^2 \underline{dx}_{ij}^2\right] + 2\left[\sum_{j=2}^{J}\sum_{k=1}^{J-1} c_j c_k \underline{dx}_{ij}\underline{dx}_{ik}\right]\right)$$

The expression

$$\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}$$

may be distributed to each of the right hand terms in the above expression, up to the point where the deviation scores occur in each term (since the order of the summation signs is not important in this expression). The result then follows immediately.

Comment: The above result implies some well-known results often given in introductory undergraduate statistics texts. For example, if we have two variables, *X* and *Y*, and we recombine the X_i and Y_i scores by the formula $K_i = X_i + Y_i$, the variance of *K* follows the rule $s_K^2 = s_x^2 + s_y^2 + 2s_{xy}$.

More generally, if we study the above proof carefully, we can discover a "heuristic rule" which produces the same answer as Result 14. This heuristic rule always works and is easy to remember, although computer programmers would find the Result as stated above more convenient to apply in practice. The heuristic rule is in 3 steps as follows. Suppose you have an expression which characterizes a linear combination as the sum of several terms.

(1) Write the expression. For example, $K = 2X + Y$.

(2) Algebraically square the term.

(obtaining, in this case, $4X^2 + 4XY + Y^2$)

(3) Apply the following conversion rule: Wherever a squared variable occurs, write, in place of the squared variable, the variance of that variable. Wherever the product of two variables occurs, replace it with the covariance of the same two variables. Leave constants unchanged. Eliminate all terms involving variables which are not either squared, or multiplied by another variable.

(obtaining, here, $4s_x^2 + 4s_{xy} + s_y^2$)

This heuristic rule can be extended to handle the case in which you take two linear combinations on the same variables, and compute the covariance between the two linear combinations. In this case, one simply computes the algebraic cross product of the expressions for the two linear combinations, and then applies the conversion rule to the result. For example, if W = X + Y, and L = X – Y, then $s_{wl} = s_x^2 - s_y^2$.

Corollary 14a. In the special case in which the variables are uncorrelated (and hence have zero covariance), the variance of the linear combination simplifies to the form

$$s_k^2 = \sum_{j=1}^{J} c_j^2 \, s_j^2$$

This completes the basic results for descriptive statistics. As we move to inferential statistics, and deal with the concept of the random variable, we will discover that random variables, like sets of scores, have means, variances, covariances, correlations, and linear combinations. We will also (to our relief and benefit) discover that the rules we have derived for the behavior of sample statistics will generalize, unchanged, to the corresponding quantities for random variables.