3 The Scalar Algebra of Variances, Covariances, and Correlations

In this chapter, we review the definitions of some key statistical concepts: *means, covariances,* and *correlations.* We show how the means, variances, covariances, and correlations of variables are related when the variables themselves are connected by one or more linear equations by developing the *linear combination and transformation rules.* These rules, which hold both for lists of numbers and *random variables*, form the algebraic foundations for regression analysis, factor analysis, and SEM.

3.1 Means, Variances, Covariances, and Correlations

In this section, we quickly review the basic definitions of the statistical quantities at the heart of structural equation modeling. Our approach is to keep the mathematical statistical content to a minimum while still conveying the important details

3.1.1 Means

The *mean* of a statistical population is simply the arithmetic average of all the numbers in the population. Usually, we model statistical populations as *random variables* with a particular statistical distribution. Since this is not a course in probability theory, we will not review the formal definition of a random variable as a function assigning numbers to outcomes, but rather use a very informal notion of such a variable as a process that generates numbers according to a specifiable system. The mean of the population can be represented as the *expected value* of the

random variable.

> **Definition (Expected Value of a Random Variable)** The expected value of a random variable X, denoted $\mathcal{E}(X)$, is the long run average of the numbers generated by the random variable X.

This definition will usually prove adequate for our purposes in this text, although the veteran of a course or two in statistics undoubtedly knows the formal definitions in terms of sums and integrals, and recognizes that the formal definitions are different for discrete and continuous random variables.

The arithmetic average of a sample of numbers taken from some statistical population is referred to as the *sample mean*. This was defined already in section 2.3.4.

3.1.2 Deviation Scores

The *deviation score* plays an important role in statistical formulas. The definitions are quite similar, for the sample and the population. In each case, the deviation score is simply the difference between the score and the mean.

Deviation scores in a statistical population are obtained by subtracting the population mean from every score in the population. If the population is modeled with a random variable, the random variable is converted to *deviation score form* by subtracting the population mean from each observation, as described in the following definition.

> **Definition (Deviation Score Random Variable)** The deviation score form of a random variable X is defined as $[dX] = X - \mathcal{E}(X)$.

Sample deviation scores were defined analogously in section 2.3.4.

3.1.3   Variances and Standard Deviations

The variance of a statistical sample or population is a measure of how far the numbers are spread out around the mean. Since the deviation score reflects distance from the mean, perhaps the two most obvious measures of spread are the average absolute deviation and the average squared deviation. For reasons of mathematical tractability, the latter is much preferred by inferential statisticians.

> **Definition (Population Variance)** The variance of a statistical population, denoted $\mathrm{Var}(X)$ or σ_X^2, is the average squared deviation score in that population. Hence,
> $$\sigma_X^2 = \mathcal{E}\left(X - \mathcal{E}(X)\right)^2.$$

An important relationship, which we give here without proof, is

> **Definition (Population Variance Computational Formula)**
> $$\begin{aligned} \sigma_X^2 &= \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 & (3.1)\\ &= \mathcal{E}\left(X^2\right) - \mu^2 & (3.2) \end{aligned}$$

The population standard deviation relates to the variance analogously to the corresponding sample quantities.

> **Definition (The Population Standard Deviation)** The population standard deviation is the square root of the population variance, i.e.,
> $$\sigma_X = \sqrt{\sigma_X^2}$$

The natural transfer of the notion of a population variance to a sample statistic would be to simply compute, for a sample of n numbers, the average squared (sample)

deviation score. Unfortunately, such an estimate tends to be, on average, lower than the population variance, i.e., *negatively biased.* To correct this bias, we divide the sum of squared deviations by $n - 1$, instead of n.

Definition (The Sample Variance) The sample variance for a set of n scores on the variable X is defined as

$$S_X^2 = 1/(n-1) \sum_{i=1}^{n} (X_i - \overline{X}_\bullet)^2$$

As an elementary exercise in summation algebra, it can be shown that the sample variance may also be computed via the following equivalent, but more convenient formula:

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

Definition (The Sample Standard Deviation) The sample standard deviation is the square root of the sample variance, i.e.,

$$S_X = \sqrt{S_X^2}$$

3.2   The Algebra of Means

The *linear transformation* is an important concept in statistics, because many elementary statistical formulas involve linear transformations.

Definition (Linear Transformation) Suppose we have two variables, X and Y, representing either random variables or two lists of numbers. Suppose furthermore that the Y scores can be expressed as linear functions of the X scores, that is, $Y_i = aX_i + b$ for some constants a and b. Then we say that Y

is a linear transformation (or "linear transform") of X. If the multiplicative constant a is positive, then the linear transformation is order-preserving.

3.2.1   The Linear Transform Rule

Both sample and population means behave in a predictable and convenient way if one variable is a linear transformation of the other.

Theorem 3.2.1 (The Sample Mean of a Linear Transform)   *Let Y and X represent two lists of n numbers. If, for each Y_i, we have $Y_i = aX_i + b$, then $\overline{Y}_\bullet = a\overline{X}_\bullet + b$.*

__Proof__ . For lists of numbers, we simply substitute definitions and employ elementary summation algebra. That is,

$$
\begin{aligned}
\overline{Y}_\bullet &= (1/n)\sum_{i=1}^{n} Y_i \\
&= (1/n)\sum_{i=1}^{n} (aX_i + b) \\
&= (1/n)\sum_{i=1}^{n} aX_i + (1/n)\sum_{i=1}^{n} b \\
&= a(1/n)\sum_{i=1}^{n} X_i + (1/n)nb \\
&= a\overline{X}_\bullet + b
\end{aligned}
$$

The second line above follows from the distributive rule of summation algebra, given on page 23. The third line uses the second constant rule of summation algebra given on page 22, for the left term, and the first constant rule (page 20), for the right.

The theorem actually includes more fundamental rules about the sample mean as special cases. First, by letting $b = 0$ and remembering that a can represent either multiplication or division by a constant, we see that multiplying or dividing a variable (or every number in a sample) by a constant multiplies or divides its mean by the same constant. Second, by letting $a = 1$, and recalling that b can represent either addition or subtraction, we see that adding or subtracting a constant from a variable adds or subtracts that constant from the mean of that variable.

After quickly mentioning two additional results, we shall construct an "$\mathcal{M}$-algebra" or *algebra of means*. Subsequently, we'll discover that this simple set of rules allows us to derive many important results.

> **Theorem 3.2.2 (Mean of a Constant)** *Suppose that, over a list of numbers X, X_i never varies. For example, suppose $X_i = b$ for all i. Then $\overline{X}_\bullet = b$.*

The proof of this result is left to an exercise.

> **Theorem 3.2.3 (Mean of a Sum)** *Suppose two lists X and Y of equal length are added together to produce a new list W. Then $\overline{W}_\bullet = \overline{X}_\bullet + \overline{Y}_\bullet$.*

Proof

$$
\begin{aligned}
\overline{W}_\bullet &= \frac{1}{n}\sum_{i=1}^{n} W_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i\right) \quad (by\ Equation\ 2.4) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} Y_i \\
&= \overline{X}_\bullet + \overline{Y}_\bullet
\end{aligned}
$$

3.2.2 Constructing an $\mathcal{M}$-Algebra

In the preceding section, we derived 3 key results:

(a) The mean of a linear transformation is the same transformation rule applied to the original mean, i.e., if $Y = aX + b$, then $\overline{Y}_\bullet = a\overline{X}_\bullet + b$.

(b) The mean of a variable that stays constant at b is equal to b.

(c) The mean of a sum of two variables is the sum of their means.

Now we'll simply restate these rules in a slightly different form. In what follows, define capital letters as variable names and lower-case letters as constants. Define the *mean operator* $\mathcal{M}$ to represent the sample mean of whatever quantity it is applied to. For example, $\mathcal{M}(X)$ represents the sample mean of a variable X representing a list of numbers X_i. Then the following rules represent the *algebra of sample means*.

(a) $\mathcal{M}(a) = a$

Note that the third rule also implies that $\mathcal{M}(X - Y) = \mathcal{M}(X) - \mathcal{M}(Y)$

(b) $\mathcal{M}(aX + b) = a\mathcal{M}(X) + b$

(c) $\mathcal{M}(X + Y) = \mathcal{M}(X) + \mathcal{M}(Y)$

We can use this algebra to prove many things about sample means more compactly than we can with summation algebra. The main hurdle is simply getting used to the notation.

> **Example 3.2.4 (Mean of a Set of Deviation Scores)**   A set of scores in the variable X can be converted to deviation scores via the formula $[dX] = X - \mathcal{M}(X)$. Now it is absolutely trivial to prove that the mean of a set of deviation scores is zero. Here we go:
>
> $$
> \begin{aligned}
> \mathcal{M}([dX]) &= \mathcal{M}(X - \mathcal{M}(X)) \\
> &= \mathcal{M}(X) - \mathcal{M}(\mathcal{M}(X)) \quad \text{(Third Rule of } \mathcal{M}\text{-algebra)} \\
> &= \mathcal{M}(X) - \mathcal{M}(X) \quad \text{(First Rule of } \mathcal{M}\text{-algebra)} \\
> &= 0
> \end{aligned}
> $$

The third line combines the first rule with the fact that $\mathcal{M}(X)$ is itself a constant.

3.3   The Algebra of Expected Values

The algebra of sample means we developed in the preceding section has a counterpart for random variables. Recall that the mean of a random variable X is called its "expected value" and is denoted $\mathcal{E}(X)$. For random variables X and Y, and constants a and b, the three primary laws of expected value algebra are:

(a) $\mathcal{E}(a) = a$

(b) $\mathcal{E}(aX + b) = a\mathcal{E}(X) + b$

(c) $\mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y)$

These rules for both discrete and continuous random variables are a straightforward consequence of the formulas for expected value in the discrete and continuous cases, and we omit them here. However, the astute reader should ask the following very serious question: would not *any* reasonable measure of "central tendency" have to follow these rules?

3.4   Linear Transform Rules for Variances and Standard Deviations

In this section, we examine the effect of a linear transformation on the variance and standard deviation. Since both the variance and standard deviation are measures on deviation scores, we begin by investigating the effect of a linear transformation of raw scores on the corresponding deviation scores. We find that the multiplicative constant "comes straight through" in the deviation scores, while an additive or subtractive constant has no effect on deviation scores.

Theorem 3.4.1 (Linear Transforms and Deviation Scores) *Suppose a variable X is is transformed linearly into a variable Y, i.e., $Y = aX + b$. Then the Y deviation scores must relate to the X deviation scores via the equation $[dY] = a[dX]$.*

Proof *First, consider the case of sets of n scores. If $Y_i = aX_i + b$, then*

$$
\begin{aligned}
[dY]_i &= Y_i - \overline{Y}_\bullet \\
&= (aX_i + b) - \overline{Y}_\bullet \\
&= (aX_i + b) - (a\overline{X}_\bullet + b) \text{ [Theorem 3.2.1]} \\
&= aX_i + b - a\overline{X}_\bullet - b \\
&= a(x_i - \overline{X}_\bullet) \\
&= a[dX]_i
\end{aligned}
$$

Next, consider the case of random variables. We have

$$
\begin{aligned}
[dY] &= Y - \mathcal{E}(Y) \\
&= aX + b - a\mathcal{E}(X) + b \text{ [Theorem 3.2.1]} \\
&= a(X - \mathcal{E}(X)) \\
&= a[dX]
\end{aligned}
$$

This completes the proof.

One may also prove the result for samples of size n using the $\mathcal{M}$-algebra. The proof is essentially the same as the proof for random variables, with $\mathcal{M}$ substituted for $\mathcal{E}$ throughouth.

A simple numerical example will help "concretize" these results. Consider the original set of scores shown in Table 3.1. Their deviation score equivalents are -1, 0, and $+1$, as shown. Suppose we generate a new set of Y scores using the equation $Y = 2X + 5$. The new deviation scores are twice the old set of deviation scores. The multiplicative constant 2 has an effect on the deviation scores, but the additive constant 5 does not.

From the preceding results, it is easy to deduce the effect of a linear transformation on the variance and standard deviation. Since additive constants have no effect on deviation scores, they cannot affect either the variance or the standard deviation. On the other hand, multiplicative constants constants "come straight through" in the deviation scores, and this is reflected in the standard deviation and the variance.

$[dX]$	X	$Y = 2X + 5$	$[dY]$
$+1$	3	11	$+2$
0	2	9	0
-1	1	7	-2

Table 3.1: Effect of a Linear Transform on Deviation Scores

Theorem 3.4.2 (Effect of a Linear Transform on the Variance and SD)
Suppose a variable X is transformed into Y via the linear transform $Y = aX + b$. Then, for random variables, the following results hold.

$$\sigma_Y = |a|\sigma_X$$

$$\sigma_Y^2 = a^2\sigma_X^2$$

Similar results hold for samples of n numbers, i.e.,

$$S_Y = |a|S_X$$

$$S_Y^2 = a^2 S_X^2$$

Proof *Consider the case of the variance of sample scores. By definition,*

$$
\begin{aligned}
S_Y^2 &= 1/(n-1)\sum_{i=1}^{n}[dY]_i^2 \\
&= 1/(n-1)\sum_{i=1}^{N}(a[dY]_i)^2 \quad \text{(from Theorem 3.4.1)} \\
&= 1/(n-1)\sum_{i=1}^{n}a^2[dX]_i^2 \\
&= a^2\left(1/(n-1)\sum_{i=1}^{n}[dX]_i^2\right) \\
&= a^2 S_X^2
\end{aligned}
$$

For random variables, we have

$$
\begin{aligned}
\sigma_Y^2 &= \mathcal{E}(Y - \mathcal{E}(Y))^2 \\
&= \mathcal{E}(aX + b - \mathcal{E}(aX + b))^2 \\
&= \mathcal{E}(aX + b - (a\mathcal{E}(X) + b))^2 \quad \text{(from Theorem 3.2.1)} \\
&= \mathcal{E}(aX + b - a\mathcal{E}(X) - b)^2 \\
&= \mathcal{E}(aX - a\mathcal{E}(X))^2 \\
&= \mathcal{E}(a^2(X - \mathcal{E}(X))^2) \\
&= a^2\mathcal{E}(X - \mathcal{E}(X))^2 \quad \text{(from 2nd rule, Section 3.3)} \\
&= a^2\sigma_X^2
\end{aligned}
$$

3.5   Standard Scores

Having deduced the effect of a linear transformation on the mean and variance of a random variable or a set of sample scores, we are now better equipped to understand the notion of "standard scores." Standard scores are scores that have a particular desired mean and standard deviation. So long as a set of scores are not all equal, it

is always possible to transform them, by a linear transformation, so that they have a particular desired mean and standard deviation.

Theorem 3.5.1 (Linear Standardization Rules) *Suppose we restrict our-selves to the class of order-preserving linear transformations, of the form* $Y = aX + b$, *with* $a > 0$. *If a set of scores currently has mean* $\overline{X}_\bullet$ *and standard deviation* S_X, *and we wish the scores to have a mean* $\overline{Y}_\bullet$ *and standard deviation* S_Y, *we can achieve the desired result with a linear transformation with*

$$a = \frac{S_Y}{S_X}$$

and

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

Similar rules hold for random variables, i.e., a random variable X *with mean* μ_X *and standard deviation* σ_X *may be linearly transformed into a random variable* Y *with mean* μ_Y *and standard deviation* σ_Y *by using*

$$a = \frac{\sigma_Y}{\sigma_X}$$

and

$$b = \mu_Y - a\mu_X$$

Example 3.5.2 (Rescaling Grades)   Result 3.5.1 is used frequently to rescale grades in university courses. Suppose, for example, the original grades have a mean of 70 and a standard deviation of 8, but the typical scale for grades is a mean of 68 and a standard deviation of 12. In that case, we have

$$a = \frac{S_Y}{S_X} = \frac{12}{8} = 1.5$$

$$b = \overline{Y}_\bullet - a\overline{X}_\bullet = 68 - (1.5)(70) = -37$$

Suppose that there is a particular "desired metric" (i.e., desired mean and standard deviation) that scores are to be reported in, and that you were seeking a formula for transforming any set of data into this desired metric. For example, you seek a formula that will transform any set of X_i into Y_i with a mean of 500 and a standard deviation of 100. Result 3.5.1 might make it seem at first glance that such a formula does not exist, i.e., that the formula would change with each new data set. This is, of course, true in one sense. However, in another sense, we can write a single formula that will do the job on any data set. We begin by defining Z-scores.

Definition (Z-Scores) Z-scores are scores that have a mean of 0 and a standard deviation of 1. Any set of n scores x_i that are not all equal can be converted into Z-scores by the following transformation rule

$$z_i = \frac{x_i - \overline{x}_\bullet}{S_X}$$

We say that a random variable is "in Z-score form" if it has a mean of 0 and a standard deviation of 1. Any random variable X having expected value μ and standard deviation σ may be converted to Z-score form via the transformation

$$Z = \frac{X - \mu}{\sigma}$$

That the Z-score transformation rule accomplishes its desired purpose is derived easily from the linear transformation rules. Specifically, when we subtract the current mean from a random variable (or a list of numbers), we do not affect the standard deviation, while we reduce the mean to zero. At that point, we have a list of numbers with a mean of 0 and the original, unchanged standard deviation. When we next divide by the standard deviation, the mean remains at zero, while the standard deviation changes to 1.

Once the scores are in Z-score form, the linear transformation rules reveal that it is very easy to transform them into any other "desired metric." Suppose, for example, we wished to transform the scores to have a mean of 500 and a standard deviation of 100. Since the mean is 0 and the standard deviation 1, multiplying by 100 will change the standard deviation to 100 while leaving the mean at zero. If we then

add 500 to every value, the resulting scores will have a mean of 500 and a standard deviation of 100.

Consequently, the following transformation rule will change any set of X scores into Y scores with a mean of 500 and a standard deviation of 100, so long as the original scores are not all equal.

$$y_i = 100 \left(\frac{x_i - \bar{x}_\bullet}{S_X} \right) + 500 \tag{3.3}$$

The part of Equation 3.3 in parentheses transforms the scores into Z-score form. Multiplying by 100 moves the standard deviation to 100 while leaving the mean at 0. Adding 500 then moves the mean to 500 while leaving the standard deviation at 100.

Z-scores also have an important theoretical property, i.e., their invariance under linear transformation of the raw scores.

> **Theorem 3.5.3 (Invariance Property of Z-Scores)** *Suppose you have two variables X and Y, and Y is an order-preserving linear transformation of X. Consider X and Y transformed to Z-score variables $[zX]$ and $[zY]$. Then $[zX] = [zY]$. The result holds for sets of sample scores, or for random variables.*

Proof *Suppose $Y = aX + b$, and assume $a > 0$. Then, by Theorems ?? and ??, $\mathcal{E}(Y) = a\mathcal{E}(X) + b$, and $\sigma_Y = a\sigma_X$. We then have*

$$
\begin{aligned}
[zY] &= \frac{Y - \mathcal{E}(Y)}{\sigma_Y} \\
&= \frac{aX + b - (aE(X) + b)}{a\sigma_X} \\
&= \frac{\cancel{a}(X - \mathcal{E}(X)) + (b - b)}{\cancel{a}\sigma_X} \\
&= \frac{X - \mathcal{E}(X)}{\sigma X} \\
&= [zX]
\end{aligned}
$$

To construct a simple example that dramatizes this result, we will employ a well-known fundamental result in statistics.

Example 3.5.4 (Properties of 3 Evenly Spaced Numbers) Consider a set of 3 evenly spaced numbers, with middle value M and "spacing" D, the difference between adjacent numbers. For example, the numbers 3,6,9 have $M = 6$ and $D = 3$. Prove that, for any such set of numbers, the sample mean $\overline{X}_\bullet$ is equal to the middle value M, and the sample standard deviation S_X is equal to D, the spacing.

Solution The proof is straightforward if you set up the notation properly. We may represent the ordered list of 3 numbers as $X_1 = M - D$, $X_2 = M$, and $X_3 = M + D$. We then have

$$
\begin{aligned}
\overline{X}_\bullet &= \frac{1}{3}\sum_{i=1}^{3} X_i \\
&= \frac{1}{3}(X_1 + X_2 + X_3) \\
&= \frac{1}{3}(M - D + M + M + D) \\
&= \frac{1}{3}(3M) \\
&= M
\end{aligned}
$$

We then have $[dX]_1 = -D$, $[dX]_2 = 0$, and $[dX]_3 = D$, whence

$$
\begin{aligned}
S_X^2 &= \frac{1}{3-1}\sum_{i=1}^{3}[dX]_i^2 \\
&= \frac{1}{2}([dX]_1^2 + [dX]_2^2 + [dX]_3^2) \\
&= \frac{1}{2}(D^2 + 0^2 + D^2) \\
&= \frac{1}{2}(2D^2) \\
&= D^2
\end{aligned}
$$

$[zX]$	X	$Y = 2X - 5$	$[zY]$
+1	115	225	+1
0	100	195	0
−1	85	165	−1

Table 3.2: Invariance of Z-Scores Under Linear Transformation

and so $S_X = D$.

This result makes it easy to construct simple data sets with a known mean and standard deviation. It also makes it easy to compute the mean and variance of 3 evenly spaced numbers by inspection.

Example 3.5.5 (Invariance of Z-Scores Under Linear Transformation)
Consider the data in Table 3.2. The X raw scores are evenly spaced, so it is easy to see they have a mean of 100 and a standard deviation of 15. The $[zX]$ scores are +1, 0, and −1 respectively. Now, suppose we linearly transform the X raw scores into Y scores with the equation $Y = 2X - 5$. The new scores, as can be predicted from the linear transformation rules, will have a mean given by

$$\overline{Y}_\bullet = 2\overline{x}_\bullet - 5 = 2(100) - 5 = 195,$$

and a standard deviation of

$$S_Y = 2S_X = 2(15) = 30$$

Since the Y raw scores were obtained from the X scores by linear transformation, we can tell, without computing them, that the $[zY]$ scores must also be +1, 0, and −1. This is easily verified by inspection.

An important implication of Result 3.5.3 is that any statistic that is solely a function of Z-scores must be invariant under linear transformations of the observed variables.

3.6 Covariances and Correlations

Multivariate analysis deals with the study of sets of variables, observed on one or more populations, and the way these variables *covary*, or behave jointly. A natural measure of how variables vary together is the covariance, based on the following simple idea. Suppose there is a dependency between two variables X and Y, reflected in the fact that high values of X tend to be associated with high values of Y, and low values of X tend to be associated with low values of Y. We might refer to this as a direct (positive) relationship between X and Y. If such a relationship exists, and we convert X and Y to deviation score form, the cross-product of the deviation scores should tend to be positive, because when an X score is above its mean, a Y score will tend to be above its mean (in which case both deviation scores will be positive), and when X is below its mean, Y will tend to be below its mean (in which case both deviation scores will be negative, and their product positive). Covariance is simply the average cross-product of deviation scores.

> **Definition (Population Covariance)** The covariance of two random variables, X and Y, denoted σ_{XY}, is the average cross-product of deviation scores, computed as
>
> $$
> \begin{aligned}
> \sigma_{XY} &= \mathcal{E}\left[(X - \mathcal{E}(X))(Y - \mathcal{E}(Y))\right] & (3.4) \\
> &= \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) & (3.5)
> \end{aligned}
> $$

Proving the equality of Equations 3.4 and 3.5 is an elementary exercise in expected value algebra that is given in many mathematical statistics texts. Equation 3.5 is generally more convenient in simple proofs.

When defining a sample covariance, we average by $n - 1$ instead of n, to maintain consistency with the formulas for the sample variance and standard deviation.

Definition (Sample Covariance) The sample covariance of X and Y is defined as

$$S_{XY} = 1/(n-1)\sum_{i=1}^{n}[dX]_i[dY]_i \tag{3.6}$$

$$= 1/(n-1)\sum_{i=1}^{n}(X_i - \overline{X}_\bullet)(Y_i - \overline{Y}_\bullet) \tag{3.7}$$

$$= 1/(n-1)\left(\sum_{i=1}^{n}Y_iY_i - \frac{\sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{n}\right) \tag{3.8}$$

Note that the variance of a variable may be thought of as its covariance with itself. Hence, in some mathematical discussions, the notation σ_{XX} is used (i.e., the covariance of X with itself) is used to signify the variance of X.

The covariance, being a measure on deviation scores, is invariant under addition or subtraction of a constant. On the other hand, multiplication or division of a constant affects the covariance, as described in the following result.

Theorem 3.6.1 (Effect of a Linear Transform on Covariance) *Suppose the variables X and Y are transformed into the variables W and M via the linear transformations $W = aX + b$ and $M = cY + d$. Then the covariance between the new variables W and M relates to the original covariance in the following way. For random variables, we have*

$$\sigma_{WM} = ac\sigma_{XY}$$

For samples of scores, we have

$$S_{WM} = acS_{XY}$$

Proof By Theorem *3.4.1*, we know that $[dW] = a[dX]$ and $[dM] = c[dY]$. So, for sample data, we have

$$
\begin{aligned}
S_{WM} &= \frac{1}{n-1}\sum_{i=1}^{n}[dW]_i[dM]_i \\
&= \frac{1}{n-1}\sum_{i=1}^{n}a[dX]_i c[dY]_i \\
&= ac\frac{1}{n-1}\sum_{i=1}^{n}[dX]_i[dY]_i \\
&= acS_{XY}
\end{aligned}
$$

For random variables, we have

$$
\begin{aligned}
\sigma_{WM} &= \mathcal{E}[dW][dM] \\
&= \mathcal{E}(a[dX]c[dY]) \\
&= ac\mathcal{E}([dX][dY]) \\
&= ac\sigma_{XY}
\end{aligned}
$$

Covariances are important in statistical theory. (If they were not, why would we be studying "the analysis of covariance structures"?) However, the covariance between two variables does not have a stable interpretation if the variables change scale. For example, the covariance between height and weight changes if you measure height in inches as opposed to centimeters. The value will be 2.54 times as large if you measure height in centimeters. So, for example, if someone says "The covariance between height and weight is 65," it is impossible to tell what this implies without knowing the measurement scales for height and weight.

3.6.1 The Population Correlation

As we noted in the preceding section, the covariance is not a scale-free measure of covariation. To stabilize the covariance across changes in scale, one need only divide it by the product of the standard deviations of the two variables. The resulting statistic, ρ_{xy}, the "Pearson Product Moment Correlation Coefficient," is defined as

the average cross-product of Z scores.

From Result 3.5.3, we can deduce without effort one important characteristic of the Pearson correlation, i.e., it is invariant under changes of scale of the variables. So, for example, the correlation between height and weight will be the same whether you measure height in inches or centimeters, weight in pounds or kilograms.

> **Definition (The Population Correlation ρ_{xy})** For two random variables X and Y the "Pearson Product Moment Correlation Coefficient," is defined as
>
> $$\rho_{XY} = \mathcal{E}([zX][zY]) \tag{3.9}$$
>
> Alternatively, one may define the correlation as
>
> $$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{3.10}$$

3.6.2   The Sample Correlation

The sample correlation coefficient is defined analogously to the population quantity. We have

> **Definition (The Sample Correlation r_{xy})** For n pairs of scores X_i and Y_i the "Pearson Product Moment Correlation Coefficient," is defined as
>
> $$r_{XY} = 1/(n-1) \sum_{i=1}^{n} [zx]_i [zy]_i \tag{3.11}$$
>
> There are a number of alternative computational formulas, such as
>
> $$r_{XY} = \frac{S_{XY}}{S_X S_Y}, \tag{3.12}$$

or its fully explicit expansion

$$r_{XY} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left(n \sum X_i^2 - (\sum X_i)^2\right)\left(n \sum Y_i^2 - (\sum Y_i)^2\right)}} \tag{3.13}$$

Since all summations in Equation 3.13 run from 1 to n, I have simplified the notation by eliminating them.

With modern statistical software, actual hand computation of the correlation coefficient is a rare event.

3.7   Linear Combination Rules

The *linear combination* (or LC) is a key idea in statistics. While studying factor analysis, structural equation modeling, and related methods, we will encounter linear combinations repeatedly. Understanding how they behave is important to the study of SEM. In this section, we define linear combinations, and describe their statistical behavior.

3.7.1   Definition of a Linear Combination

Definition (Linear Combination) A *linear combination* (or *linear composite*) L of J random variables X_j is any weighted sum of those variables, i.e..,

$$L = \sum_{j=1}^{J} c_j X_j \tag{3.14}$$

For sample data, in a data matrix $\mathbf{X}$ with n rows and J columns, representing n scores on each of J variables, a linear combination score for the ith set of observations is any expression of the form

$$L_i = \sum_{j=1}^{J} c_j X_{ij} \tag{3.15}$$

The c_j in Equations 3.14 and 3.15 are called *linear weights*.

It is important to be able to examine a statistical formula, decide whether or not it contains or represents a LC, and identify the linear weights, with their correct sign. A simple example is given below.

> **Example 3.7.1 (Course Grades as a Linear Combination)** Suppose an instructor announces at the beginning of a course that the course grade will be produced by taking the mid-term exam grade, adding twice the final exam grade, and dividing by 3. If the final grade is G, the mid-term is X, and the final exam Y, then the grades are computed using the formula
>
> $$G = \frac{X + 2Y}{3} = \frac{1}{3}X + \frac{2}{3}Y \tag{3.16}$$
>
> An example of such a situation is shown in Table 3.3. In this case, the course grade *is* a linear combination, and the linear weights are $+1/3$ and $+2/3$.
>
> *Note:* Table 3.3 gives the means and standard deviations for the mid-term exam (X), the final exam (Y), and the course grade (G). In subsequent sections of this chapter, we show how the mean and standard deviation for G can be calculated from a knowledge of the mean and standard deviation of X and Y, the linear weights that produced G, and the covariance between X and Y.

Student	Mid-term X	Final Y	Grade G
Judi	70	78	75.33
Fred	60	86	77.33
Albert	50	82	71.33
Mean	60	82	74.67
SD	10	4	3.06

Table 3.3: Course Grades as a Linear Combination

3.7.2   Mean Of A Linear Combination

When several variables are linearly combined, the resulting LC has a mean that can be expressed as a linear function of the means of original variables. This result, given here without proof, is the foundation of a number of important statistical methods.

> **Theorem 3.7.2 (Mean of a Linear Combination)**   *Given J random variables X_j having means μ_j, the mean of a linear combination*
>
> $$L = \sum_{j=1}^{J} c_j X_j$$

is

$$\mu_L = \sum_{j=1}^{J} c_j \mu_j$$

For a set of N sample scores on J variables, with linear combination scores computed as

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

the mean of the linear combination scores is given by

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

This simple, but very useful result allows one to compute the mean of a LC without actually calculating LC scores.

Example 3.7.3 (Computing the Mean of a Linear Combination) Consider again the data in Table 3.3. One may *calculate* the mean of the final grades by calculating the individual grades, summing them, and dividing by n. An alternative way is to compute the mean from the means of X and Y. So, for example, the mean of the final grades must be

$$
\begin{aligned}
\bar{G}_\bullet &= \frac{1}{3}\bar{X}_\bullet + \frac{2}{3}\bar{Y}_\bullet \\
&= \left(\frac{1}{3}\right)60 + \left(\frac{2}{3}\right)82 \\
&= \frac{224}{3} \\
&= 74.67
\end{aligned}
$$

3.7.3   Variance and Covariance Of Linear Combinations

The variance of a LC also plays a very important role in statistical theory.

Theorem 3.7.4 (Variance of a Linear Combination)   *Given J random variables X_j having variances $\sigma_j^2 = \sigma_{jj}$ and covariances σ_{jk} (between variables X_j and X_k), the variance of a linear combination*

$$L = \sum_{j=1}^{J} c_j X_j$$

is

$$\sigma_L^2 \;=\; \sum_{j=1}^{J}\sum_{k=1}^{J} c_j c_k \sigma_{jk} \tag{3.17}$$

$$=\; \sum_{j=1}^{J} c_j^2 \sigma_j^2 + 2\sum_{j=2}^{J}\sum_{k=1}^{J-1} c_j c_k \sigma_{jk} \tag{3.18}$$

For a set of n sample scores on J variables, with linear combination scores computed as

$$L_i = \sum_{j=1}^{J} c_j x_{ij},$$

the variance of the linear combination scores is given by

$$S_L^2 \;=\; \sum_{j=1}^{J}\sum_{k=1}^{J} c_j c_k S_{jk} \tag{3.19}$$

$$=\; \sum_{j=1}^{J} c_j^2 S_j^2 + 2\sum_{j=2}^{J}\sum_{k=1}^{J-1} c_j c_k S_{jk} \tag{3.20}$$

It is, of course, possible to define more than one LC on the same set of variables. For example, consider a personality questionnaire that obtains responses from individuals on a large number of items. Several different personality "scales" might be calculated as linear combinations of the items. (Items not included in a scale have a linear weight of zero.) In that case, one may be concerned with the covariance or correlation between the two LCs. The following theorem describes how the

covariance of two LCs may be calculated.

Theorem 3.7.5 (Covariance of Two Linear Combinations) *Consider two linear combinations L and M of a set of J random variables X_j, having variances and covariances σ_{ij}. (If $i = j$, a variance of variable i is being referred to, otherwise a covariance between variables i and j.) The two LCs are,*

$$L = \sum_{j=1}^{J} c_j X_j,$$

and

$$M = \sum_{j=1}^{J} d_j X_j,$$

The covariance σ_{LM} between the two variables L and M is given by

$$\sigma_{LM} = \sum_{j=1}^{J} \sum_{k=1}^{J} c_j d_k \sigma_{jk} \qquad (3.21)$$

The theorem generalizes to linear combinations of sample scores in the obvious way. For a set of n sample scores on J variables, with linear combination scores computed as

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

and

$$M_i = \sum_{j=1}^{J} d_j X_{ij}$$

the covariance of the linear combination scores is given by

$$S_{LM} = \sum_{j=1}^{J} \sum_{k=1}^{J} c_j d_k S_{jk} \qquad (3.22)$$

Some comments are in order. First, note that Theorem 3.7.5 includes the result of Theorem 3.7.4 as a special case (simply let $L = M$, recalling that the variance of a variable is its covariance with itself). Second, the results of both theorems can be expressed as a *heuristic rule* that is easy to apply in simple cases.

3.7.4   A General Heuristic Rule for Linear Transoforms and Linear Combinations

We can combine our previous rules for linear transformations and combinations into a single heuristic. To compute a variance of a linear combination or transformation, for example

$$L = X + Y,$$

perform the following steps:

(a) Express the linear combination or transformation in algebraic form;

$$X + Y$$

(b) Square the expression;

$$(X + Y)^2 = X^2 + Y^2 + 2XY$$

(c) Transform the result with the following *conversion rules*: (a) Wherever a squared variable appears, replace it with the variance of the variable. (b) Wherever the product of two variables appears, replace it with the covariance of the two variables. (c) Any expression that does not contain either the square of a variable or the product of two variables is eliminated.

$$X^2 + Y^2 + 2XY \rightarrow \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

To calculate the covariance of two linear combinations, replace the first two steps in the heuristic rules as follows:

(a) Write the two linear combinations side-by-side, for example;

$$(X + Y) \quad (X - 2Y)$$

(b) Compute their algebraic product;

$$X^2 - XY - 2Y^2$$

The conversion rule described in step 3 above is then applied, yielding, with the present example

$$\sigma_X^2 - 2\sigma_{XY} - 2\sigma_Y^2$$

The above result applies identically to random variables or sets of sample scores. It provides a convenient method for deriving a number of classic results in statistics. A well known example is the difference between two variables, or, in the sample case, the difference between two columns of numbers.

> **Example 3.7.6 (Variance of Difference Scores)** Suppose you gather pairs of scores X_i and Y_i for n people, and compute, for each person, a *difference score* as
>
> $$D_i = X_i - Y_i$$
>
> Express the sample variance S_D^2 of the difference scores as a function of S_X^2, S_Y^2, and S_{XY}.
>   *Solution.* First apply the heuristic rule. Squaring $X - Y$, we obtain $X^2 + Y^2 - 2XY$. Apply the conversion rules, and the result is
>
> $$S_D^2 = S_X^2 + S_Y^2 - 2S_{XY}$$

> **Example 3.7.7 (Variance of Course Grades)** The data in Table 3.3 give the variances of the X and Y tests. Since the deviation scores are easily cal-culated, we can calculate the covariance $S_{XY} = 1/(n-1)\sum[dX][dY]$ between the two exams as -20. Since the final grades are a linear combination of the exam grades, we can generate a theoretical formula for the variance of the final

grades. Since $G = 1/3X + 2/3Y$, application of the heuristic rule shows that

$$
\begin{aligned}
S_G^2 &= 1/9S_X^2 + 4/9S_Y^2 + 4/9S_{XY} \\
&= 1/9(100) + 4/9(16) + 4/9(-20) \\
&= 84/9 \\
&= 9.33
\end{aligned}
$$

The standard deviation of the final grades is thus $\sqrt{9.33} = 3.06$.

One aspect of the data in Table 3.3 is unusual, i.e., the scores on the two exams are negatively correlated. In practice, of course, this is unlikely to happen unless the class size is very small and the exams unusual. Normally, one observes a substantially positive correlation between grades on several exams in the same course.

In the following example, we use LC theory to explain a phenomenon, which we might call *variance shrinkage*, that has caught many a teacher by surprise when computing standardized course grades.

Example 3.7.8 (Variance Shrinkage in Course Grades) Suppose an instructor is teaching a course in which the grades are computed simply by averaging the marks from two exams. The instructor is told by the university administration that, in this course, final grades should be standardized to have a mean of 70 and a standard deviation of 10.

The instructor attempts to comply with this request. She standardizes the grades on each exam to have a mean of 70 and a standard deviation of 10. She then computes the final grades by averaging the standardized marks on the two exams, which, incidentally, had a correlation of $+.50$. She sends the grades on to the administration, but receives a memo several days later to the effect that, although her final grades had a mean of 70, their standard deviation was incorrect. It was supposed to be 10, but it was only 8.66. What went wrong?

Solution We can explain this phenomenon using linear combination theory. Since the final grades are obtained averaging the grades on the two mid-terms, the grades are computed as a linear combination, i.e.,

$$
G = (1/2)X + (1/2)Y
$$

Using the heuristic rule, we establish that the variance of the grades must follow the formula

$$S_G^2 = (1/4)S_X^2 + (1/4)S_Y^2 + (1/2)S_{XY}$$

However, since $S_X^2 = S_Y^2$, and $S_{XY} = S_X S_Y r_{XY} = S_X^2 r_{xy}$, the formula reduces to

$$S_G^2 = S_X^2 \frac{1 + r_{xy}}{2}$$

If $r_{xy} = .5$, then $S_G^2 = (3/4)S_X^2$. We have shown that, if the two tests correlate .5, and each test is standardized to yield a mean of 70 and a standard deviation of 10, it is an inevitable consequence of the laws of LC theory that the final grades, calculated as an average of the two tests, will have a variance that is 75, and a standard deviation equal to $\sqrt{75} = 8.66$. More generally, we have demonstrated that in similar situations, if the individual tests are standardized to a particular mean and variance, the average of the two tests will have a variance that "shrinks" by a multiplicative factor of $(1 + 2r_{xy})/2$. On the other hand, it is easy to verify that the mean of the final grades will be the desired value of 70.

Example 3.7.9 (The "Rich Get Richer" Phenomenon)   Ultimately, the hypothetical instructor in Example 3.7.8 will have to do something to offset the variance shrinkage. What should she do?

Solution   The answer should be apparent from our study of linear transformation theory. There are several ways to approach the problem, some more formal than others. The problem is that the variance of the scores is too low, so they need to be "stretched out" along the number line. Intuitively, we should realize that if we subtract 70 from each student's average grade, we will have a set of scores with a mean of 0 and a standard deviation of 8.66. If we then multiply each score by 10/8.66, we will have scores with a mean of 0 and a standard deviation of 10. Adding back 70 to the numbers will then restore the proper mean of 70, while leaving the standard deviation of 10. This transformation

can be written literally as

$$M_i = (G_i - 70)(10/8.66) + 70 \qquad (3.23)$$

There are, of course, alternative approaches. One is to transform each student's mark to a Z-score, then multiply by 10 and add 70. This transformation can be written

$$M_i = 10\frac{G_i - 70}{8.66} + 70$$

With a small amount of manipulation, you can see that these two transformation rules are equivalent.

Study Equation 3.23 briefly, and see how it operates. Notice that it "expands"the individual's deviation score (his/her difference from the mean of 70), making it larger than it was before. In this case, the multiplying factor is 10/8.66, so, for example, a grade of 78.66 would be transformed into a grade of 80. It is easy to see that, if the individual exam grades and the final mark are scaled to have the same standard deviation, any student with an exam average above the mean will receive a final grade higher than their exam average. On the other hand, any student with an exam grade below average will receive a mark lower than their exam average. For example, a person with a grade of 61.34, and a deviation score of -8.66, would have the deviation score transformed to -10, and the grade changed to 60. In my undergraduate courses, I refer to this phenomenon as "The Rich Get Richer, The Poor Get Poorer." This phenomenon can be the source of considerable consternation for a student whose grades have been below the mean, as the student's final grade may be lower than *any* of his/her grades on the exams. Indeed, the lower the student's mark, the more it is reduced! On the other hand, students with consistently above-average performance may get a pleasant surprise if they are unaware of this effect.

Linear combination theory can be used to prove a number of surprising results in statistics. The following example shows how to take two columns of numbers and compute two linear combinations on them that are guaranteed to have a zero correlation.

Example 3.7.10 (Constructing Sets of Uncorrelated Scores) Statistics instructors sometimes construct sample data sets with certain characteristics. Here is a simple way to construct two columns of numbers that have an *exactly zero correlation.* I will describe the method first, then we will use linear combination theory to explain why the scores must have a zero correlation.

The method is as follows: First, simply write down a list of N different numbers in a column. Call these numbers X. Next to the X column, write the exact same numbers, but in a different order. Call these Y. Now, create two additional columns of numbers, W and M. W is the sum of X and Y, M the difference between X and Y. W and M will have a zero correlation.

An example of the method is shown in Table 3.7.4. Using any general purpose statistical program, you can verify that the numbers indeed have a precisely zero correlation. Can you explain why?

X	Y	$W = X + Y$	$M = X - Y$
1	2	3	−1
2	3	5	−1
3	5	8	−2
4	1	5	3
5	4	9	1

Table 3.4: Constructing Numbers with Zero Correlation

Solution We can develop a general expression for the correlation between two columns of numbers, using our heuristic rules for linear combinations. Let's begin by computing the covariance between W and M. First, we take the algebraic product of $X + Y$ and $X - Y$. We obtain

$$(X + Y)(X - Y) = X^2 - Y^2$$

Applying the conversion rule, we find that

$$S_{X+Y,X-Y} = S_X^2 - S_Y^2$$

Note that the covariance between W and M is unrelated to the covariance between the original columns of numbers, X and Y! Moreover, if X and Y have the same variance, then W and M will have a covariance of exactly zero. Since the correlation is the covariance divided by the product of the standard deviations, it immediately follows that W and M will have a correlation of zero if and only if X and Y have the same variance. If X and Y contain the same numbers (in permuted order), then they have the same variance, and W and M will have zero correlation.