DOCUMENT RESUME

ED 235 205 TM 830 618

AUTHOR	O'Brien, Francis J., Jr.
TITLE	A Derivation of the Sample Multiple Correlation Formula for Raw Scores.
INSTITUTION	National Opinion Research Center, New York, NY.
PUB DATE	24 Jun 83
NOTE	64p.; For related document, see ED 223 429.
PUB TYPE	Guides - Classroom Use - Materials (For Learner) (051)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Correlation; Higher Education; Instructional Materials; *Mathematical Formulas; *Scores; *Statistics; *Supplementary Reading Materials
IDENTIFIERS	Linear Models; *Multiple Correlation Formula

ABSTRACT

This paper, a derivation of the multiple correlation formula for unstandardized (raw) scores, is the fourth in a series of publications. The purpose of these papers is to provide supplementary reading for students of applied statistics. The intended audience is social science graduate and advanced undergraduate students familiar with applied statistics. The minimum background for most of the existing and forthcoming papers is knowledge of applied statistics through rudimentary analysis of variance, and multiple correlation and regression analysis. The unique feature of this set of papers is detailed proofs and derivations of important formulas and derivations which are not readily available in textbooks, journal articles, and other similar sources. Each proof or derivation is presented in a clear, detailed and consistent fashion. When necessary, a review of relevant algebra is provided. Calculus is not used or assumed. This series seeks to address the needs of students to see a full, comprehensible statement of a mathematical argument. (PN)

A Derivation of the Sample Multiple Correlation Formula for Raw Scores

Francis J. O'Brien, Jr., Ph.D.

Assistant Sampling Director
National Opinion Research Center

NORC
Sampling Department
902 Broadway
New York, NY 10010

June 24, 1983

ERRATA SHEET

"A derivation of the sample multiple correlation formula for raw scores" by Francis J. O'Brien, Jr., June 24, 1983

Page*	Now Reads	Correct to
10, footnote, 4 lines down	$n\bar{X}\,\bar{Y}$	$n\bar{X}_1\bar{Y}$
13	$var(b_2, x_2)$	$var(b_2 x_2)$
17, footnote	Multiple R	multiple R
29, equation $\sum x_p Y$	$b_p \sum x_p$	$b_p \sum x_p^2$
36, 2 lines from bottom of text	mathematical calculus	mathematical statistics
38, 2nd equation	S_1^2	S_Y^2
43, last line in text	$S_j^2 = 0$	$S_j^2 = 1$

*Page refers to original numbers (at top).

Table of Contents

List of Tables

A Derivation of the Sample Multiple Correlation Formula for Raw Scores

Francis J. O'Brien, Jr., Ph.D.
National Opinion Research Center, New York

Introduction

This paper is the fourth in a series of publications. The purpose of these papers is to provide supplementary reading for students of applied statistics. (See O'Brien, 1982a, 1982b, 1982c). My intended audience is social science graduate and advanced undergraduate students familiar with applied statistics. The minimum background for most of the existing and forthcoming papers is knowledge of applied statistics through rudimentary analysis of variance, and multiple correlation and regression analysis.

The unique feature of this set of papers is detailed proofs and derivations of important formulas and derivations which are not readily available in textbooks, journal articles, and other similar sources. Each proof or derivation is presented in a clear, detailed and consistent fashion. When necessary, a review of relevant algebra is provided. Calculus is not used or assumed.

As a former instructor of applied statistics on the graduate level, I know that many students are very capable of understanding the proofs and derivations presented in these papers. My experience has been that many students desire to see a full, comprehensible statement of a mathematical argument. This series seeks to address such needs.

The present paper is a companion work to an earlier paper (O'Brien, 1982c). Each is a derivation of the multiple correlation formula for the linear model. The first paper formulated a detailed derivation of the multiple correlation formula for standard (z) scores.[1] The present paper is a derivation of the multiple correlation formula for unstandardized (raw) scores. Readers should find each paper interesting and informative.

[1] Typographical errors appeared in this paper. For the readers' convenience, corrections are summarized in Appendix B of the present paper. The author would be grateful if other errors in that paper or the present paper were communicated to him.

The two papers taken together are meant to be preparatory reading for a related paper.[1]

Overview of Derivation

In this paper we will present a derivation of the linear multiple correlation formula for raw scores. The basic objective is to derive this formula for one raw score criterion (dependent variable) and any finite number of raw score predictors (independent variables).

Let us first state the formula we will derive and introduce the notation used. The linear multiple correlation between one criterion and p predictors can be expressed as:

$$R_{Y.x_1,x_2,\ldots,x_j,\ldots,x_p} = \frac{\sqrt{b_1 r_{y1} s_y s_1 + b_2 r_{y2} s_y s_2 + \ldots + b_j r_{yj} s_y s_j + \ldots + b_p r_{yp} s_y s_p}}{s_y}$$

Writing the right hand side in summation notation:

$$R_{Y.x_1,x_2,\ldots,x_j,\ldots,x_p} = \frac{\sqrt{\sum_{j=1}^{p} b_j r_{yj} s_y s_j}}{s_y}$$

where:

$R_{Y.x_1,x_2,\ldots,x_j,\ldots,x_p}$	=	multiple correlation of raw scores,
Y	=	the observed raw score criterion to be predicted,
$x_1,x_2,\ldots,x_j,\ldots,x_p$		raw score predictors of the criterion,

[1] Forthcoming with the expected title: "A Derivation of the Unbiased Sample Standard Error of Estimate: the General Case." It will appear in ERIC.

$b_1, b_2, \ldots, b_j, \ldots, b_p$ = slope coefficients or regression weights,

$r_{y1}, r_{y2}, \ldots, r_{yj}, \ldots, r_{yp}$ = product moment criterion/predictor correlations,

$S_1, S_2, \ldots, S_j, \ldots, S_p$ = standard deviations of the predictors,

S_y = the standard deviation of the criterion.

This is the formula that is derived in this paper. We will first present a derivation for the simplest multivariate case: one criterion and two predictors. A derivation is then presented for three predictors. The latter derivation is a useful exercise because it allows a review of the logic and procedures used in the derivation. In addition, it will motivate the use of summation when the algebra becomes complex. The derivation is then presented for the general case of p (finite) predictors. An integral part of this paper is Appendix A. In that appendix, a method is presented for finding the "normal equations" in regression analysis for raw score linear models.

Prior to starting the derivation for two predictors, let us outline the plan which will be followed in the derivations. The steps we will use are:

1. state the regression model
2. derive the normal equations (see Appendix A)
3. define the multiple correlation
4. apply rules of covariance and variance algebra to simplify the definitional form of the multiple correlation formula
5. substitute the normal equations into the multiple correlation formula
6. simplify.

We will refine these steps to suit a particular application.

Brief Overview of Regression Analysis and Derivation for Two Predictors

In this section we will review the basic concepts, logic and our notation for regression analysis. Introductory applied statistics textbooks can be consulted for more detailed information on regression analysis theory. See, for example, Lindeman , et al., 1982. The intention in this section is to review the rationale of regression analysis.

The primary use of statistical regression analysis is controlled prediction and explanation of quantatative data. The basic principle that lay behind regression analysis involves selecting a general mathematical function that best matches the underlying form of variables over which one desires to exercise predictability. Assume one is attempting to predict one raw score criterion by use of two raw score predictors. Assume further that the relationship between each predictor and the criterion is linear in form. The mathematical function most often selected to obtain the best linear "fit" for these conditions is provided by the following equation:

$$\hat{Y} = a + b_1x_1 + b_2x_2$$

where:

$\hat{Y}$	=	the predicted (not actual or observed) criterion,
a, b_1, b_2	=	constants to be selected by the "least squares" procedure; a = the slope intercept, and b_1 and b_2 = slope coefficient terms,
x_1, x_2	=	predictor variables in deviation score form.

It is conventional to express the predictor variables in deviation score form. That is, for each predictor, first find its mean and then subtract the mean from each predictor. For example,

$$x_1 = X_1 - \bar{X}_1$$
$$x_2 = X_2 - \bar{X}_2$$

Here, for either variable, "cap X" is the actual (or gross) raw score and $\bar{X}$ is its arithmetic mean. It is not necessary for any mathematical reason to re-express the predictors in deviation score form. This is done simply to force the algebra to be more tractable. As such, it is a matter of convenience.[1] Note that we do not re-express $\hat{Y}$ (or Y) as deviations. We could re-express each type of criterion. However, we have chosen not to do this since most authors follow this convention.

Using deviation scores for the predictors, we can now write the two predictor raw score model as follows:

$$\hat{Y} = a + b_1(X_1 - \bar{X}_1) + b_2(X_2 - \bar{X}_2)$$
$$= a + b_1x_1 + b_2x_2$$

As stated, we will use the second form in this paper.

The regression model stated above is an idealized mathematical model. If a variable set consisting of one criterion and two predictors can be assumed to be linear, then the model is a reasonable one to apply for prediction of actual or observed criterion scores. It is idealized in the sense that it assumes no error is made in the prediction of $\hat{Y}$. In practice, when an actual criterion score is compared to the criterion

---

[1]Readers of the 1982c paper may wonder why on page 2 thereof the raw score regression model was stated in terms of gross raw score (and not deviation score) predictors. As stated, it is not necessary mathematically to re-express. In any case, the major result we are seeking in this paper is unaffected by the initial form of the predictors. The derivation could be made without the translation of predictors into deviation score form, but the result would involve unnecessary and unwanted complexities. Practically speaking, this paper would have been very much longer if re-expression was not done.

score generated by model, some error is likely to occur--the "fit" is less than perfect. If we call the actual sample raw score criterion Y, we can state another model (an observed raw score model):

$$Y = \hat{Y} + e$$

where:

e = the amount of numerical error resulting from using the idealized mathematical model ($\hat{Y}$) to predict the actual criterion score (Y).

That is, an actual criterion consists of a predicted quantity plus an error component.

The error made in predicting the observed criterion score by the idealized mathematical model is:

$$e = Y - \hat{Y}$$

This is the quantity we want to be as small as possible in order to minimize the error in prediction. It can be seen that if e=0, the actual criterion is perfectly predicted by the idealized model ($Y=\hat{Y}$).

The technique most often used in the social sciences to accomplish this goal is the "least squares" procedure. Essentially, this procedure seeks to maximize predictability by minimizing prediction error. The least squares criterion or goal is summarized in the following expression:[1]

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \text{a minimum}$$

If we substitute the quantity for $\hat{Y}$ previously defined, we can rewrite the least squares criterion as:

---

[1] If it is understood that the summation limits range from the first observation (i=1) to the last (i=n) then we can drop the summation limits; n refers to the total number of observations for the criterion and predictors. This sample size is the same regardless of the number of predictors in the regression model. Later in the paper when the algebra becomes more complex, we use summation limits extensively.

$$\sum\left[Y - (a + b_1x_1 + b_2x_2)\right]^2 = \sum(Y - a - b_1x_1 - b_2x_2)^2 = \sum e^2 = \text{minimum}$$

(As an aside, "least squares" means we determine values for a, b_1 and b_2 in $\hat{Y}$ such that the squared error term results in the least possible value).

Normal Equations

Having stated the multiple regression model for two predictors, we now derive the so-called "normal equations". A discussion of the procedures and results we will need is presented in Appendix A. The reader may wish to read Appendix A at this point (or take the next step on faith).

The normal equations are derived from the least squares criterion using calculus. The basic idea that lay behind the technique for two predictors is to generate an equation for each of the constants in the regression model (a, b_1 and b_2). For the two predictor model, the normal equations for a, b_1 and b_2, respectively are found to be:

$$\sum Y = na + b_1\sum x_1 + b_2\sum x_2$$

$$\sum x_1Y = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2$$

$$\sum x_2Y = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2$$

In the first normal equation (for a), n is the sample size.

These normal equations can be simplified by substituting various descriptive statistics into terms of the equations. Other terms will cancel in the process. For the readers convenience in following the substitutions, some basic formulas for sample descriptive statistics are presented in Table 1.

Table 1

Descriptive Sample Statistics

Statistic	Raw Score Form	Deviation Score Form	
Mean	$\bar{X}_1 = \frac{\sum X_1}{n}$	same	
Variance	$S_1^2 = \frac{\sum (X_1 - \bar{X}_1)^2}{n-1}$	$S_1^2 = \frac{\sum x_1^2}{n-1}$	
Standard Deviation	$S_1 = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2}{n-1}}$	$S_1 = \sqrt{\frac{\sum x_1^2}{n-1}}$	
Correlation of Y and X_1	$r_{y1} = \frac{\sum (X_1 - \bar{X}_1)(Y - \bar{Y})}{(n-1)S_1 S_2}$	$\frac{\sum x_1 y}{(n-1)S_1 S_2}$	(where $y = Y - \bar{Y}$)
	$= \frac{\sum (X_1 - \bar{X}_1)Y}{(n-1)S_1 S_2}$	$\frac{\sum x_1 Y}{(n-1)S_1 S_2}$	
	$= \frac{\sum X_1 (Y - \bar{Y})}{(n-1)S_1 S_2}$	$\frac{\sum X_1 y}{(n-1)S_1 S_2}$	

Note: For "mean" it is understood that the summation extends across all n values of X_1 (and Y for "correlation"). This applies equally to other statistics defined in the table.

1. In the first normal equation, we recognize that, on the right hand side:

$$\sum x_1 = \sum (X_1 - \bar{X}_1) = 0$$

$$\sum x_2 = \sum (X_2 - \bar{X}_2) = 0$$

2. In the second normal equation, we can see that:

$$\sum x_1 = 0$$

$\sum x_1^2 = \sum (X_1 - \bar{X}_1)^2$, but the sample variance, S_1^2, is:

$$\frac{\sum (X_1 - \bar{X}_1)^2}{n-1} \quad \text{or} \quad (n-1)S_1^2 = \sum (X_1 - \bar{X}_1)^2$$

This may be substituted for $\sum x_1^2$.

As for $\sum x_1 x_2$, we can use the definition of the sample correlation between x_1 and x_2 to simplify this term. By definition, for samples:

$$r_{12} = \frac{\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{(n-1)S_1 S_2} = \frac{\sum x_1 x_2}{(n-1)S_1 S_2} \quad \text{or:}$$

$(n-1) r_{12} S_1 S_2 = \sum x_1 x_2$. This may be substituted.

Finally, $\sum x_1Y$ may be simplified as follows:

$$\sum x_1Y = \sum (X_1-\overline{X}_1)Y \text{ . Now,}$$

$\sum x_1Y$ is identical to[1] $\sum (X_1-\overline{X}_1)(Y-\overline{Y})$ or

$\sum x_1Y = \sum x_1y$ (where $y = Y-\overline{Y}$). This is recognized to be the numerator of the correlation between x_1 and Y (r_{y1} or r_{1y}). Hence,

$$r_{y1} = \frac{\sum x_1Y}{(n-1)S_1S_2} \quad \text{or} \quad \sum x_1Y = (n-1)r_{y1}S_yS_1.$$

This may be substituted into the second normal equation.

---

[1]PROOF:

$$\sum(X_1-\overline{X}_1)Y = \sum(X_1Y - \overline{X}_1Y) = \sum X_1Y - \overline{X}_1\sum Y = \sum X_1Y - n\overline{X}_1\overline{Y}$$

Now, $\sum(X_1-\overline{X}_1)(Y-\overline{Y}) = \sum(X_1Y - X_1\overline{Y} - \overline{X}_1Y + \overline{X}_1\overline{Y})$

$$= \sum X_1Y - \sum X_1\overline{Y} - \sum \overline{X}_1Y + \sum \overline{X}\,\overline{Y}$$

$$= \sum X_1Y - \overline{Y}(n\overline{X}_1) - \overline{X}_1(n\overline{Y}) + n\overline{X}_1\overline{Y} = \sum X_1Y - n\overline{X}\,\overline{Y}$$

Therefore, $\sum(X_1-\overline{X}_1)Y = \sum(X_1-\overline{X}_1)(Y-\overline{Y})$ End of proof.

3. For the equation $\sum x_2Y$ we can write down immediately the following simplifications:

$\sum x_2Y$ translates into $\sum x_2Y = (n-1)r_{y2}S_yS_2$. In addition,

$$\sum x_2 = 0$$

$$\sum x_1x_2 = (n-1)r_{12}S_1S_2$$

$$\sum x_2^2 = (n-1)S_2^2$$

Making all these substitutions, we arrive at a simplified set of the originally stated normal equations.

$$\sum Y = na + b_1(0) + b_2(0)$$

$$(n-1)r_{y1}S_yS_1 = a(0) + b_1(n-1)S_1^2 + b_2(n-1)r_{12}S_1S_2$$

$$(n-1)r_{y2}S_yS_2 = a(0) + b_1(n-1)r_{12}S_1S_2 + b_2(n-1)S_2^2$$

To further simplify, eliminate zero terms, and for the last two normal equations, divide each term by (n-1). This gives us:

$$\sum Y = na$$

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2$$

As a final simplification, we can divide through the first equation by n:

$$\bar{Y} = a$$

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2$$

These are the normal equations we want to work with in the derivation for two predictors. For the reader's convenience in working through the derivation, we will restate them prior to the derivation.

Multiple Correlation

We are now ready to define the multiple correlation for one criterion and two predictors. By definition:[1]

$$R_{Y.x_1,x_2} = \text{corr}(Y,\hat{Y}) = \text{corr}(Y, a + b_1x_1 + b_2x_2)$$

$$= \frac{\text{cov}(Y,\hat{Y})}{\sqrt{\text{var}(Y)\ \text{var}(\hat{Y})}}$$

$$= \frac{\text{cov}(Y, a + b_1x_1 + b_2x_2)}{\sqrt{\text{var}(Y)}\ \sqrt{\text{var}(a + b_1x_1 + b_2x_2)}}$$

where:

corr means correlation ,
cov means covariance and,
var means variance.

---

[1] Alternative notation systems use $R_{Y.x_1+x_2}$ or $R_{Y.x_1x_2}$, among others.

It is important to remember that a, b_1 and b_2 function as constants.

Elementary covariance and variance operations performed on the above correlation formula yield in the first step:

$$R_{Y.x_1,x_2} = \frac{\mathrm{cov}(Y,a) + \mathrm{cov}(Y,b_1x_1) + \mathrm{cov}(Y,b_2x_2)}{\sqrt{\mathrm{var}(Y)}\sqrt{\begin{array}{c}\mathrm{var}(a) + \mathrm{var}(b_1x_1) + \mathrm{var}(b_2,x_2) + \\ 2\mathrm{cov}(a,b_1x_1) + 2\mathrm{cov}(a,b_2x_2) \quad 2\mathrm{cov}(b_1x_1,b_2x_2)\end{array}}}$$

Applying rules of covariance and variance for variables and constants, we can achieve further simplification.[1] This is done on the next page.

---

[1]To briefly review: the variance of any constant is zero; the variance of a product term containing a constant yields the squared constant times the variance of the variables--for example ;

$$\mathrm{var}(b_1x_1) = b_1^2\ \mathrm{var}(x_1)$$

When a covariance term contains constants, factor the constants outside the covariance operator (sometimes this reduces the covariance to zero)--for example,

$$\mathrm{cov}(a,b_1x_1) = ab_1\mathrm{cov}(1,x_1) = 0$$

but

$$\mathrm{cov}(b_1x_1,b_2x_2) = b_1b_2\mathrm{cov}(x_1,x_2)$$

By definition, the covariance is related to the simple correlation--for example,

$$\mathrm{cov}(x_1,x_2) = r_{12}S_1S_2$$

This should appear correct since, by definition,

$$r_{12} = \frac{\mathrm{cov}(x_1,x_2)}{\sqrt{\mathrm{var}(x_1)\ \mathrm{var}(x_2)}}$$

$$R_{Y.x_1,x_2} = \frac{0 + b_1\text{cov}(Y,x_1) + b_2\text{cov}(Y,x_2)}{S_y\sqrt{\begin{array}{l}0 + b_1^2\text{var}(x_1) + b_2^2\text{var}(x_2) + \\ 0 + 0 + 2b_1b_2\text{cov}(x_1,x_2)\end{array}}}$$

As mentioned, by definition:

$$\text{cov}(Y,x_1) = r_{y1}S_yS_1$$
$$\text{cov}(Y,x_2) = r_{y2}S_yS_2$$
$$\text{cov}(x_1,x_2) = r_{12}S_1S_2$$

One further observation should be made with respect to the variance of the predictors. For example, the variance of x_1 is:

$$\text{var}(x_1) = \text{var}(X_1 - \overline{X}_1)$$

By definition, the variance of this difference is:

$$\text{var}(X_1) + \text{var}(\overline{X}_1) - 2\text{cov}(X_1,\overline{X}_1)$$

Since $\overline{X}_1$ is a constant,

$$\text{var}(x_1) = \text{var}(X_1) + 0 - 0$$
$$= S_1^2$$

Similar results obtain for $\text{var}(x_2)$. Therefore, when all substitutions are made:

$$R_{Y.x_1,x_2} = \frac{b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}}$$

This is the form of the multiple R we will use in the derivation. It will be restated for the readers convenience.

Derivation

The following formula for one criterion and two predictors appears in many applied statistics textbooks:

$$R_{Y \cdot x_1, x_2} = \frac{\sqrt{b_1 r_{y1} S_y S_1 + b_2 r_{y2} S_y S_2}}{S_y}$$

We are now able to show its derivation.

For the readers convenience, a restatement of the simplified set of normal equations and the multiple R formula is given in Table 2.

The derivation involves two steps: a)substitute the normal equations into the numerator of the multiple R formula and b)simplify algebraically.

See the page following Table 2.

Table 2

Normal Equations and Multiple Correlation Formula for Two Raw Score Predictors

Normal Equations[1]

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2$$

Multiple Correlation

$$R_{Y.x_1,x_2} = \frac{b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}}$$

---

[1] The $a = \bar{Y}$ term is omitted because it plays no role in the derivation (other than zero).

NOTE: Proof involves the substitution of the normal equations into the numerator of the multiple R formula and simplifying. See text for details.

Notice that the numerator of the multiple R formula contains the terms $r_{y1}S_yS_1$ and $r_{y2}S_yS_2$. These terms are functionally related to the normal equations. If we substitute normal equations for each term into R and rearrange terms, we obtain the following results:

$$R_{Y.x_1,x_2} = \frac{b_1(b_1S_1^2 + b_2r_{12}S_1S_2) + b_2(b_1r_{12}S_1S_2 + b_2S_2^2)}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}}$$

$$= \frac{b_1^2S_1^2 + b_1b_2r_{12}S_1S_2 + b_1b_2r_{12}S_1S_2 + b_2^2S_2^2}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}}$$

$$= S_y\left[\frac{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}{\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2}}\right]$$

(Hence, $b_1r_{y1}S_yS_2 + b_2r_{y2}S_yS_2 = b_1^2S_1^2 + b_2^2S_2^2 + 2b_1b_2r_{12}S_1S_2$)

Now, the bracketed term of the denominator can be simplified algebraically if we remember radicals and laws of exponents.[1]

---

[1] Let the denominator (inside the brackets) be called A. Thus, the structure of the Multiple R is:

$$R = S_y\left[\frac{A}{\sqrt{A}}\right]$$

Recall the following permissible operation (rationalizing the denominator):

$$R = S_y\left[\frac{A\sqrt{A}}{\sqrt{A}\sqrt{A}}\right] = Sy\left[\frac{A\sqrt{A}}{\sqrt{A^2}}\right] = S_y\left[\frac{A\sqrt{A}}{A}\right] = \frac{\sqrt{A}}{S_y}$$

Simplifying:

$$R_{Y.x_1,x_2} = \frac{\sqrt{b_1^2 S_1^2 + b_2^2 S_2^2 + 2b_1 b_2 r_{12} S_1 S_2}}{S_y}$$

Therefore,

$$R_{Y.x_1,x_2} = \frac{\sqrt{b_1 r_{y1} S_y S_1 + b_2 r_{y2} S_y S_2}}{S_y}$$

END OF PROOF[1]

---

[1]For readers familiar with the 1982c paper, it is possible to obtain a "cheap" proof in the analogous standard score regression model. If variables are in standard score form, then the standard deviations become unity;

$S_{z_y} = S_{z_1} = S_{z_2} = 1$. Thus, in the notation of

the 1982c paper,

$$R_{z_y.z_1,z_2} = \sqrt{B_1 r_{y1} + B_2 r_{y2}}$$

Derivation for Three Predictors

Let us now work out the derivation for a three predictor raw score linear regression model. This will allow us to review the logic and procedures of the derivation. We will also introduce the use of summation which becomes necessary for the general case of p predictors.

The first step is to state the regression model. For three predictors:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + b_3x_3$$

We have simply added an independent variable to our prediction (idealized) mathematical model to form a four dimensional model ($\hat{Y}$ and three predictors with their associated slope terms).

As in the two predictor model, we make use of the least squares criterion to establish our goal of minimizing the prediction error:

$$\sum(Y - \hat{Y})^2 = \sum(Y - a - b_1x_1 - b_2x_2 - b_3x_3)^2 = \sum e^2 = \text{a minimum}$$

The next step is the application of partial differentiation to find derivatives of each of the terms in the prediction model (a, b_1, b_2 and b_3). This procedure produces the set of normal equations. Appendix A shows the procedures involved. Omitting the cumbersome algebra involved in simplifying the original set of normal equations, we can state the final and simplified set of normal equations as follows:

$$\bar{Y} = a$$

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2 + b_3r_{13}S_1S_3$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2 + b_3r_{23}S_2S_3$$

$$r_{y3}S_yS_3 = b_1r_{13}S_1S_3 + b_2r_{23}S_2S_3 + b_3S_3^2$$

Recall that the value of a is determined in practice but it plays no role in the derivation since it "drops out" in covariance and variance operations of the multiple R derivation.

The above normal equations are the ones we will make use of in the derivation of the multiple R formula for three predictors. A restatement of them is presented in Table 3 for easy reference.

The third step is to define the multiple correlation of one criterion and three raw score predictors. Rules of covariance and variance algebra will allow us to simplify the definitional form of R.

The multiple R is defined on the following page.[1]

[1] The term a is omitted. For justification, the reader may want to include it in the definition of R and ascertain the result.

$$R_{Y.x_1,x_2,x_3} = \text{corr}(Y,\hat{Y}) = \text{corr}(Y, b_1x_1 + b_2x_2 + b_3x_3) = \frac{\text{cov}(Y,\hat{Y})}{\sqrt{\text{var}(Y)\ \text{var}(\hat{Y})}}$$

$$= \frac{\text{cov}(Y, b_1x_1 + b_2x_2 + b_3x_3)}{\sqrt{\text{var}(Y)}\sqrt{\text{var}(b_1x_1 + b_2x_2 + b_3x_3)}}$$

All of the above forms state equivalent ways to define the multiple R. The last is amenable to operations of covariance and variance. Applying rules of covariance and variance algebra:

$$R_{Y.x_1,x_2,x_3} = \frac{\text{Cov}(Y,b_1x_1) + \text{cov}(Y,b_2x_2) + \text{cov}(Y,b_3x_3)}{S_y\sqrt{\text{var}(b_1x_1) + \text{var}(b_2x_2) + \text{var}(b_3x_3) + 2\text{cov}(b_1x_1,b_2x_2) + 2\text{cov}(b_1x_1,b_3x_3) + 2\text{cov}(b_2x_2,b_3x_3)}}$$

$$= \frac{b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2 + b_3r_{y3}S_yS_3}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + b_3^2S_3^2 + 2b_1b_2r_{12}S_1S_2 + 2b_1b_3r_{13}S_1S_3 + 2b_2b_3r_{23}S_2S_3}}$$

This is as far as we can simplify the multiple R at this point. We will retain this for eas[illegible]erence. See Table 3.

Table 3

Normal Equations and Multiple Correlation Formula for Three Raw Score Predictors

Normal Equations[1]

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2 + b_3r_{13}S_1S_3$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2 + b_3r_{23}S_2S_3$$

$$r_{y3}S_yS_3 = b_1r_{13}S_1S_3 + b_2r_{23}S_2S_3 + b_3^2$$

Multiple Correlation

$$R_{Y.x_1,x_2,x_3} = \frac{b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2 + b_3r_{y3}S_yS_3}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + b_3^2S_3^2 + 2b_1b_2r_{12}S_1S_2 + 2b_1b_3r_{13}S_1S_3 + 2b_2b_3r_{23}S_2S_3}}$$

[1] Again, we note that the term a $(=\bar{Y})$ is omitted from normal equations and the multiple R.

NOTE: Derivation involves substituting the normal equations into the multiple R and simplifying. See the text for details.

We have stated the multiple regression model and least squares criterion, and presented the normal equations and the multiple R formula. The fourth step is to substitute the normal equations into the multiple R.

If we substitute each of the normal equations for appropriate terms in the numerator of R we obtain (see Table 3):

$$\mathrm{cov}(Y,\hat{Y}) = b_1 r_{y1} S_y S_1 + b_2 r_{y2} S_y S_2 + b_3 r_{y3} S_y S_3$$

$$= b_1(b_1 S_1^2 + b_2 r_{12} S_1 S_2 + b_3 r_{13} S_1 S_3) + b_2(b_1 r_{12} S_1 S_2 + b_2 S_2^2 + b_3 r_{23} S_2 S_3)$$
$$+ \; b_3(b_1 b_3 r_{13} S_1 S_3 + b_2 b_3 r_{23} S_2 S_3 + b_3 S_3^2)$$

$$= (b_1^2 S_1^2 + b_1 b_2 r_{12} S_1 S_2 + b_1 b_3 r_{13} S_1 S_3) + (b_1 b_2 r_{12} S_1 S_2 + b_2^2 S_2^2 + b_2 b_3 r_{23} S_2 S_3)$$
$$+ \; (b_1 b_3 r_{13} S_1 S_3 + b_2 b_3 r_{23} S_2 S_3 + b_3^2 S_3^2)$$

Now, let us write each parenthesised term on a separate line to form a covariance matrix:

$$\begin{aligned}\mathrm{cov}(Y,\hat{Y}) = \; & b_1^2 S_1^2 && + \; b_1 b_2 r_{12} S_1 S_2 && + \; b_1 b_3 r_{13} S_1 S_3 \\ & b_1 b_2 r_{12} S_1 S_2 && + \; b_2^2 S_2^2 && + \; b_2 b_3 r_{23} S_2 S_3 \\ & b_1 b_3 r_{13} S_1 S_3 && + \; b_2 b_3 r_{23} S_2 S_3 && + \; b_3^2 S_3^2\end{aligned}$$

At this point we will introduce summation to simplify the algebra. Consider the three squared terms along the northwest to southeast diagonal of the covariance matrix. It is clear that we might express these terms in summation as follows:

$$b_1^2S_1^2 + b_2^2S_2^2 + b_3^2S_3^2 = \sum_{j=1}^{3} b_j^2S_j^2$$

The remaining six terms in the matrix consist of three <u>pairs</u> of quantities:

$$2b_1b_2r_{12}S_1S_2 + 2b_1b_3r_{13}S_1S_3 + 2b_2b_3r_{23}S_2S_3$$

One common way to express this in summation is as follows:[1]

$$2(b_1b_2r_{12}S_1S_2 + b_1b_3r_{13}S_1S_3 + b_2b_3r_{23}S_2S_3) = 2\sum_{j=2}^{3} \sum_{i=1}^{2} b_ib_jr_{ij}S_iS_j$$

---

[1] One of several forms often seen in multivariate statistics textbooks is as follows:

$$2\sum_{i<j} b_ib_jr_{ij}S_iS_j \quad ; \quad i \neq j$$

The total number of terms to be summed is determined by multiplying the upper limits (3x2=6). In the double summation operation, the inside summation operator is set to 1; then increment the outer operator (j=2,3) giving i j=12 + 13. Now increment i to 2 and complete the limits of j (with the side condition that i≠j -- e.g., i j=22 is not permitted). The subscripts that result from all of the summation operations are: 12 + 13 + 23. Each value ,of course, is taken twice.

Thus, the nine covariance terms of the multiple R numerator can be written in all of the following ways:

$$\text{cov}(Y,\hat{Y}) = b_1^2 S_1^2 + b_2^2 S_2^2 + b_3^2 S_3^2 + 2b_1 b_2 r_{12} S_1 S_2 + 2b_1 b_3 r_{13} S_1 S_3 + 2b_2 b_3 r_{23} S_2 S_3$$

$$= \sum_{j=1}^{3} b_j^2 S_j^2 + 2\sum_{j=2}^{3}\sum_{i=1}^{2} b_i b_j r_{ij} S_i S_j$$

$$\equiv b_1 r_{y1} S_y S_1 + b_2 r_{y2} S_y S_2 + b_3 r_{y3} S_y S_3 \equiv \sum_{j=1}^{3} b_j r_{yj} S_y S_j$$

This last equation is simply a restatement of multiple R numerator from Table 3. The second equation was just derived from the first equation.

Turning to the denominator of the multiple R in Table 3, it is readily apparent that it is similar to the covariance term above. That is:

$$\sqrt{\text{var}(Y)}\sqrt{\text{var}(\hat{Y})} = S_y\sqrt{b_1^2 S_1^2 + b_2^2 S_2^2 + b_3^2 S_3^2 + 2b_1 b_2 r_{12} S_1 S_2 + 2b_1 b_3 r_{13} S_1 S_3 + 2b_2 b_3 r_{23} S_2 S_3}$$

$$= S_y\sqrt{\sum_{j=1}^{3} b_j^2 S_j^2 + 2\sum_{j=2}^{3}\sum_{i=1}^{2} b_b b_j r_{ij} S_y S_j}$$

If we now form the ratio of covariance and variance terms for the multiple R, we can complete the derivation for three predictors:

$$R_{Y.x_1,x_2,x_3} = \frac{\sum_{j=1}^{3} b_j^2 S_j^2 + 2\sum_{j=2}^{3}\sum_{i=1}^{2} b_i b_j r_{ij} S_i S_j}{S_y\sqrt{\sum_{j=1}^{3} b_j^2 S_j^2 + 2\sum_{j=2}^{3}\sum_{i=1}^{2} b_i b_j r_{ij} S_i S_j}}$$

Notice that the numerator and denominator (under the radical) are identical in form. If we make the same algebraic simplification we made for the two predictor derivation, we obtain:

$$R_{Y.x_1,x_2,x_3} = \frac{\sqrt{\sum_{j=1}^{3} b_j^2 S_j^2 + 2\sum_{j=2}^{3}\sum_{i=1}^{2} b_i b_j r_{ij} S_i S_j}}{S_y}$$

$$= \frac{\sqrt{\sum_{j=1}^{3} b_j r_{yj} S_y S_j}}{S_y} \qquad \text{END OF PROOF}$$

This completes the derivation for three predictors. We now derive the multiple R for any [possi]ible (finite) number of predictors in the linear regression model.

Derivation for p Predictors

The derivation of the multiple correlation formula for any number of predictors will be presented,as a generalization of the two and three predictor cases. A rigorous mathematical proof that the generalization holds for p predictors could be provided by "mathematical induction". Our approach in this section is a straightforward multivariate generalization.

For reference, the following is a listing of the general steps for the p predictor variable case:

1. state the regression model for p predictors
2. derive the normal equations (see Appendix A)
3. define the multiple R
4. substitute normal equations into numerator of R
5. express the covariance term in summation
6. express the variance term in summation
7. simplify

The linear regression model is:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + \ldots + b_jx_j + \ldots + b_px_p$$

The least squares criterion is:

$$\sum(Y - \hat{Y})^2 = \sum e^2 = \text{a minimum}$$

Substituting for $\hat{Y}$:

$$\sum(Y - a - b_1x_1 - b_2x_2 - \ldots - b_jx_j - \ldots - b_px_p)^2 = \sum e^2 = \text{a minimum}$$

Next we derive the normal equations. In unsimplified form we have:[1]

---

[1]Note that the normal equations for terms $\sum x_1Y$, $\sum x_2Y$ etc. are written such that the first subscript is always less that the second one. Since these products are symmetric ($\sum x_1Y = \sum Yx_1$ etc.) this method simplifies the algebra. See Appendix A for more detail.

$$\Sigma Y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2 + b_3 \Sigma x_3 + \ldots + b_j \Sigma x_j + \ldots + b_p \Sigma x_p$$

$$\Sigma x_1 Y = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + b_3 \Sigma x_1 x_3 + \ldots + b_j \Sigma x_1 x_j + \ldots + b_p \Sigma x_1 x_p$$

$$\Sigma x_2 Y = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 + b_3 \Sigma x_2 x_3 + \ldots + b_j \Sigma x_2 x_j + \ldots + b_p \Sigma x_2 x_p$$

$$\Sigma x_3 Y = a \Sigma x_3 + b_1 \Sigma x_1 x_3 + b_2 \Sigma x_2 x_3 + b_3 \Sigma x_3^2 + \ldots + b_j \Sigma x_3 x_j + \ldots + b_p \Sigma x_3 x_p$$

$$\vdots$$

$$\Sigma x_p Y = a \Sigma x_p + b_1 \Sigma x_1 x_p + b_2 \Sigma x_2 x_p + b_3 \Sigma x_3 x_p + \ldots + b_j \Sigma x_j x_p + \ldots + b_p \Sigma x_p$$

If we apply the same logic and make the same substitutions we made for 2 and 3 predictors, we obtain a simplified set of normal equations:

$$\bar{Y} = a$$

$$r_{y1} S_y S_1 = b_1 S_1^2 + b_2 r_{12} S_1 S_2 + b_3 r_{13} S_1 S_3 + \ldots + b_j r_{1j} S_1 S_j + \ldots + b_p r_{1p} S_1 S_p$$

$$r_{y2} S_y S_2 = b_1 r_{12} S_1 S_2 + b_2 S_2^2 + b_3 r_{23} S_2 S_3 + \ldots + b_j r_{2j} S_2 S_j + \ldots + b_p r_{2p} S_2 S_p$$

$$r_{y3} S_y S_3 = b_1 r_{13} S_1 S_3 + b_2 r_{23} S_2 S_3 + b_3 S_3^2 + \ldots + b_j r_{3j} S_3 S_j + \ldots + b_p r_{3p} S_3 S_p$$

$$\vdots$$

$$r_{yp} S_y S_p = b_1 r_{1p} S_1 S_p + b_2 r_{2p} S_2 S_p + b_3 r_{3p} S_3 S_p + \ldots + b_j r_{jp} S_j S_p + \ldots + b_p S_p^2$$

restatement of the normal equations is given in Table 4.

Multiple Correlation for p Predictors and Derivation

We are now ready to derive the multiple correlation formula for p predictors. See Table 4 for a statement of the definition of the multiple R.

The covariance term is:[1]

$$\mathrm{cov}(Y, a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_jx_j + \ldots + b_px_p)$$

$$= b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2 + b_3r_{y3}S_yS_3 + \ldots + b_jr_{yj}S_yS_j + \ldots + b_pr_{yp}S_yS_p$$

Now, substitute the normal equations (line for line--see Table 4):

$$\mathrm{cov}(Y,\hat{Y}) = b_1(b_1S_1^2 + b_2r_{12}S_1S_2 + \ldots + b_pr_{1p}S_1S_p) + b_2(b_1r_{12}S_1S_2 + b_2S_2^2 + \ldots + b_pr_{2p}S_2S_p) + \ldots + b_p(b_1r_{1p}S_1S_p + b_2r_{2p}S_2S_p + \ldots + b_pS_p^2)$$

Multiply each of the b_j terms inside the parentheses and write each parenthesized sum on a separate line:

$$\begin{aligned}
\mathrm{cov}(Y,\hat{Y}) &= b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2 + b_3r_{y3}S_yS_3 + \ldots + b_pr_{yp}S_yS_p \\
&= b_1^2S_1^2 + b_1b_2r_{12}S_1S_2 + b_1b_3r_{13}S_1S_3 + \ldots + b_1b_pr_{1p}S_1S_p \; + \\
&\quad b_1b_2r_{12}S_1S_2 + b_2^2S_2^2 + b_2b_3r_{23}S_2S_3 + \ldots + b_2b_jr_{2j}S_2S_j \; + \\
&\quad b_1b_3r_{13}S_1S_3 + b_2b_3r_{23}S_2S_3 + b_3^2S_3^2 + \ldots + b_3r_{3p}S_3S_p \; + \\
&\quad \vdots \\
&\quad b_1b_pr_{1p}S_1S_p + b_2b_pr_{2p}S_2S_p + b_3r_{3p}S_3S_p + \ldots + b_jb_pr_{jp}S_jS_p + \ldots + b_p^2S_p^2
\end{aligned}$$

[1] For reasons presented earlier, the term a is omitted in the derivation.

Table 4

Normal Equations and Multiple Correlation Formula for p Raw Score Predictors

Normal Equations[1]

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2 + b_3r_{13}S_1S_3 + \ldots + b_jr_{1j}S_1S_j + \ldots + b_pr_{1p}S_1S_p$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2 + b_3r_{23}S_2S_3 + \ldots + b_jr_{2j}S_1S_j + \ldots + b_pr_{2p}S_2S_p$$

$$r_{y3}S_yS_3 = b_1r_{13}S_1S_3 + b_2r_{23}S_2S_3 + b_3S_3^2 + \ldots + b_jr_{3j}S_3S_j + \ldots + b_pr_{3p}S_3S_p$$

$$\vdots$$

$$r_{yp}S_yS_p = b_1r_{1p}S_1S_p + b_2r_{2p}S_2S_p + b_3r_{3p}S_3S_p + \ldots + b_jr_{jp}S_jS_p + \ldots + b_pS_p^2$$

Multiple Correlation

$$R_{Y.x_1,x_2,x_3,\ldots,x_j,\ldots,x_p} = \text{corr}(Y,\hat{Y}) = \text{corr}(Y,\ b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_jx_j + \ldots + b_px_p)$$

$$= \frac{\text{cov}(Y,\hat{Y})}{\sqrt{\text{var}(Y)\ \text{var}(\hat{Y})}} = \frac{\text{cov}(Y, b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_jx_j + \ldots + b_px_p)}{\sqrt{\text{var}(Y)}\sqrt{\text{var}(b_1x_1 + b_2x_2 + \ldots + b_jx_j + \ldots + b_px_p)}}$$

$$= \frac{b_1r_{y1}S_yS_1 + b_2r_{y2}S_yS_2 + \ldots + b_jr_{yj}S_yS_j + \ldots + b_pr_{yp}S_yS_p}{S_y\sqrt{b_1^2S_1^2 + b_2^2S_2^2 + \ldots + b_p^2S_p^2 + 2b_1b_2r_{12}S_1S_2 + 2b_1b_3r_{13}S_1S_3 + \ldots + 2b_ib_jr_{ij}S_iS_j + \ldots + 2b_{p-1}b_pr_{p-1,p}S_{p-1}S_p}}$$

[1] The $a=\bar{Y}$ term is omitted from the normal equations and multiple R.

NOTE: Derivation consists of substituting each $r_{yj}S_yS_j$ normal equation term into the covariance term of the multiple R. See text for details.

To facilitate working with such a complex matrix, we will introduce summation at this point. As the first step, we count the total number of terms to be summed. An inspection of the covariance matrix (page 30) above makes it evident that each row consists of p terms. Since there is a total of p such rows, the entire covariance matrix consists of $p \times p = p^2$ terms. For example, in the derivation for three predictors, we worked with three rows, each of which contained three terms or a total of $3 \times 3 = 3^2 = 9$ terms. In the p predictor model, the covariance matrix consists of two kinds of terms: diagonal terms ($b_1^2 S_1^2$ to $b_p^2 S_p^2$) and off diagonal terms. It is evident that there are p such diagonal terms. A little algebra will tell us how many off diagonal terms are in the covariance matrix. Let X represent the total number of off diagonal terms. Then:

$$\text{TOTAL MATRIX} = p^2 = p + X$$

or

$$X = p^2 - p$$

$$X = p(p-1)$$

Thus, the entire covariance matrix consists of p diagonal terms and $p(p-1)$ off diagonal terms for a total of p^2 terms.

We can view the structure of the covariance matrix in another way. This view is the "trick" in understanding the expression of the matrix in summation notation. Notice that the off diagonal terms exhibit a pattern (as we saw in the two and three predictor cases). Each $b_i b_j r_{ij} S_i S_j$ corresponds to one other term in the matrix that is identical to it. For example, the first off diagonal term in row one is $b_1 b_2 r_{12} S_1 S_2$, and the first term in row two is identical to it. In general, any off diagonal term in row i, column j is identical to the term in row j, column i (e.g., row 2, column 5 = row 5, column 2). Thus, the off diagonal terms consist of a number of identical pairs of terms. There are $p(p-1)$ such pairs of off diagonal terms. Suppose we halve the total p^2 matrix and consider the upper half only that makes a right triangle. In this halved matrix, we are considering the p $b_j^2 S_j^2$ diagonal terms and $p(p-1)/2$ off diagonal terms. That is, the upper triangle consists of $p + \frac{p(p-1)}{2}$ terms. To represent the entire covariance matrix (p^2 terms), simply double the number of off diagonal terms in the half matrix: $p^2 = p + 2\left[\frac{p(p-1)}{2}\right] = p + p(p-1)$ total terms.

Examine the matrix of covariance terms for the three predictor case for further clarification.

As explained, the $\text{cov}(Y,\hat{Y})$ matrix consists of $p \times p = p^2$ terms; there are p $b_j^2 S_j^2$ and $p(p-1) b_i b_j r_{ij} S_i S_j$ $\left\{\textbf{or}\ 2\left[p(p-1)/2\right]\right\}$ terms in the total matrix.

Expressing the total number of diagonal terms in summation notation:

$$b_1^2 S_1^2 + b_2^2 S_2^2 + \ldots + b_j^2 S_j^2 + \ldots + b_p^2 S_p^2 = \sum_{j=1}^{p} b_j^2 S_j^2$$

The off diagonal terms can be expressed in summation notation as follows:[1]

$$2(b_1 b_2 r_{12} S_1 S_2 + \ldots + b_j b_p r_{jp} S_j S_p) = 2 \sum_{j=2}^{p} \sum_{i=1}^{p-1} b_i b_j r_{ij} S_i S_j$$

---

[1] For those readers familiar with combinatorics, the following may assist in clarifying the logic.

There is a total of p $b_j^2 S_j^2$ which are combined with all such terms one at a time. In combinatorial notation, this means that p $b_j^2 S_j^2$ terms are combined one at a time--that is:

$$\text{total number of } b_j^2 S_j^2 \text{ terms} = \binom{p}{1} = \frac{p!}{1!(p-1)!} = \frac{p(p-1)(p-2)\ldots 1}{1(p-1)(p-2)\ldots 1} = p$$

For the off diagonal terms, we construct $2\binom{p}{2}$ terms (pairs of identical terms, each combined with all other like terms two at a time). Thus:

$$\text{Total number of } b_i b_j r_{ij} S_i S_j \text{ terms} = 2\binom{p}{2} = 2\left[\frac{p!}{2!(p-2)!}\right] = 2\left[\frac{p(p-1)(p-2)(p-3)\ldots 1}{2\ (p-2)(p-3)\ldots 1}\right] = p(p-1)$$

Hence, the entire covariance matrix consists of:

$$\binom{p}{1} + 2\binom{p}{2} = p + p(p-1) = p^2 \text{ terms}$$

For example, in the three predictor model, the first off diagonal term was seen to be $b_1 b_2 r_{12} S_1 S_2$, and the last was seen to be $b_2 b_3 r_{23} S_2 S_3$; In the case of a 10 predictor model, first and last terms, respectively, would be: $b_1 b_2 r_{12} S_1 S_2$ and $b_9 b_{10} r_{9,10} S_9 S_{10}$.

We can now express the full covariance matrix in summation notation as:

$$\mathrm{cov}(Y,\hat{Y}) = \sum_{j=1}^{p} b_j^2 S_j^2 + 2 \sum_{j=2}^{p} \sum_{i=1}^{p-1} b_i b_j r_{ij} S_i S_j$$

Equivalently,

$$\mathrm{cov}(Y,\hat{Y}) = b_1 r_{y1} S_y S_1 + b_2 r_{y2} S_y S_2 = \ldots + b_j r_{yj} S_y S_j$$

$$= \sum_{j=1}^{p} b_i r_{yj} S_y S_j$$

Thus,

$$\mathrm{cov}(Y,\hat{Y}) = \sum_{j=1}^{p} b_j^2 S_j^2 + \sum_{j=2}^{p} \sum_{i=1}^{p-1} b_i b_j r_{ij} S_i S_j = \sum_{j=1}^{p} b_i r_{yj} S_y S_j$$

The latter equation is very important in the final steps.

If the variance terms of the multiple R are examined, we see that $\sqrt{\mathrm{var}(Y)}$ is simply S_y by definition. The term, $\sqrt{\mathrm{var}(\hat{Y})}$ can be manipulated by covariance and variance rules to produce the following (see Table 4):

$$\sqrt{\mathrm{var}(\hat{Y})} = \sqrt{\mathrm{var}(b_1x_1 + b_2x_2 + \ldots + b_jx_j + \ldots + b_px_p)}$$

$$= \sqrt{b_1^2S_1^2 + b_2^2S_2^2 + \ldots + b_j^2S_j^2 + \ldots + b_p^2S_p^2 + 2b_1b_2r_{12}S_1S_2 + 2b_1b_3r_{13}S_1S_3 + 2b_1b_4r_{14}S_1S_4 + \ldots + 2b_ib_jr_{ij}S_iS_j + \ldots + 2b_pb_{p-1}r_{p,p-1}S_pS_{p-1}}$$

In summation notation:

$$\sqrt{\mathrm{var}(\hat{Y})} = \sqrt{\sum_{j=1}^{p} b_j^2S_j^2 + 2\sum_{j=2}^{p}\sum_{i=1}^{p-1} b_ib_jr_{ij}S_iS_j}$$

THEREFORE, AFTER MUCH LABOR, WE CAN STATE THE MULTIPLE CORRELATION :

$$R_{Y.x_1,x_2,x_3,\ldots,x_j,\ldots,x_p} = \frac{\sum_{j=1}^{p} b_j^2S_j^2 + 2\sum_{j=2}^{p}\sum_{i=1}^{p-1} b_ib_jr_{ij}S_iS_j}{S_y\sqrt{\sum_{j=1}^{p} b_j^2S_j^2 + 2\sum_{j=2}^{p}\sum_{i=1}^{p-1} b_ib_jr_{ij}S_iS_j}}$$

$$= \frac{\sqrt{\sum_{j=1}^{p} b_j^2S_j^2 + 2\sum_{j=2}^{p}\sum_{i=1}^{p-1} b_ib_jr_{ij}S_iS_j}}{S_y}$$

$$= \frac{\sqrt{\sum_{j=1}^{p} b_jr_{yj}S_yS_j}}{S_y}$$

END OF PROOF FOR p PREDICTORS

Appendix A

Normal Equations in Regression Analysis

Introduction

In this appendix, we outline a set of procedures to apply in regression analysis for finding normal equations. The procedures are appropriate when:

a) the regression model is linear, and
b) the measures are in raw score.

If variables are transformed to a nonlinear form prior to regression analysis procedures, the procedures described in this appendix would not apply. Examples of nonlinear transformations include logarithmic, exponential and square root re-expression, or, in general, whenever the exponents of the variables in the regression model are not equal to unity. For example,

$$\hat{Y} = a + b_1 x_1 + b_2 x_2^2$$

This is a nonlinear mathematical model since the exponent of x_2 is not equal to 1.

To derive normal equations for a given regression model requires knowledge of elementary differential calculus which makes use of partial differentiation.[1] Students who are familiar with calculus may read any textbook of mathematical calculus for the details (for example, Hoel, Port and Stone, 1971).

[1] For students who need to review this procedure, or who know some calculus and want to learn the technique, see Goodman, 1977, for a good introduction.

To render a conceptual understanding of normal equations as they are employed in the least squares procedure, let us take an example of a two predictor model. The mathematical model applied to a distribution assumed linear in each predictor is the one given in the text, namely:

$$\hat{Y} = a + b_1x_1 + b_2x_2$$

The raw score model includes an error component, and the error made in prediction of the criterion (Y) with the above model may be negative, zero or positive. The raw score model is:

$$Y = \hat{Y} + e$$

Solving for e, we obtain:

$$Y - \hat{Y} = e$$

This represents the amount of numerical error made on a score-by-score basis when we predict Y with the idealized model, $\hat{Y}$. To obtain an overall indication of the amount of prediction error for the entire raw score distribution, we might be tempted to define:

$$\Sigma(Y-\hat{Y}) = \Sigma e \quad \text{(over all n observations)}$$

The problem with this approach is that the resulting sum on the left side turns out to be exactly zero[1]; $\Sigma(Y-\hat{Y}) = \Sigma e = 0$. That is, positive errors cancel out negative errors leaving zero as the overall sum. This is obviously problematical because no matter how good or bad a particular mathematical model (linear or nonlinear) is for empirical score prediction, we would have no way of determining its utility (using the sensible criterion of minimizing prediction error).

---

[1]Proof. For two predictors:

$$\Sigma(Y-a-b_1x_1-b_2x_2) = \Sigma(Y-\bar{Y}- b_1x_1-b_2x_2)$$
$$= \Sigma(Y-\bar{Y}) - b_1\Sigma x_1 - b_2\Sigma x_2 = 0$$

The generalization of this for p predictors is obvious.

For these reasons, the most widely used and accepted procedure for finding normal equations is based on the least squares criterion; i.e.,

$$\Sigma(Y - \hat{Y})^2 = \Sigma(Y - a - b_1 x_1 - b_2 x_2)^2 = \Sigma e^2 = \text{minimum}$$

(The summation ranges from i=1 to i=n or over the entire set of observations). In words, least squares states: find numerical values for a, b_1 and b_2 which will make the prediction error the smallest possible numerical amount upon substitution.

The reader is already aware of one least squares type of result from elementary statistics. A kind of least squares criterion (and procedure) is used in defining the sample variance of a distribution; i.e.,

$$S_1^2 = \frac{1}{n-1} \Sigma(Y-\overline{Y})^2$$

The arithmetic mean, $\overline{Y}$, is used in variance formulas (instead of medians or other numbers) because the resulting variance is the smallest possible value when the mean is used rather than any other number (or combination of numbers) in that given distribution.[1] This is derived through the same calculus procedure used in deriving normal equations, and is based on the same principle: optimization or minimization.

---

[1] Take an example:

Y	$(Y-2)^2$	$(Y-4)^2$	$(Y-8)^2$	$(Y-10)^2$	$(Y-11)^2$	$(Y-7)^2$
2						
4						
8						
10						
11						

$\overline{Y} = 7$

Find each squared sum and compare it against $(Y-\overline{Y})^2$ (The n-1 can be ignored since it is a constant and has no material bearing on the result). It will be seen that only

$(Y-7)^2$ gives the smallest squared deviation sum.

Our task in regression analysis is to find numerical values corresponding to terms in the model to satisfy the least squares criterion of minimum error of prediction. The resulting values, when substituted into the regression equation, satisfies the criterion of minimization. In essence, we solve p+1 equations (p= the number of predictors, and 1 corresponds to the slope intercept term), or one equation for one term in the model. Each equation is then solved simultaneously to determine computing formulas to obtain the numerical values for the p+1 terms in the model. Finally, each predictor (and the slope intercept term) is passed through the resulting prediction equation to find a unique predicted criterion for each observation in the data set. The rest is statist cal theory (see Lindeman, et al. for an excellent discussion of regression theory).

To take the two predictor example once again,

$$\Sigma(Y - a - b_1x_1 - b_2x_2)^2 = \Sigma e^2 = \text{minimum.}$$

We are not interested in finding a computational formula for a, b_1 and b_2. Our goal is to stop one step short of doing that. We are interested in finding the normal equations, and simplifying them to substitute into the multiple R.

Plan

We will now set down a plan for finding the normal equations. A four phase plan is used throughout this appendix for finding normal equations. This will help structure the presentation.

A. state the regression model, $\hat{Y}$

B. state the mathematical function of the least squares criterion, $\Sigma(Y-\hat{Y})^2$

C. derive the normal equations for each of terms in the model

D. summarize the normal equations

Finding Normal Equations for the Two Predictor Model

Let us apply the four phase plan first to the two predictor case.

A. the regression function is

$$\hat{Y} = a + b_1x_1 + b_2x_2$$

B. the least squares criterion is

$$\Sigma(Y - \hat{Y})^2 = \Sigma(Y - a - b_1x_1 - b_2x_2)^2 = \Sigma e^2$$

C. the procedures for deriving the normal equations are:

1. For the slope intercept term, a, we need to:

a) drop the exponent 2 and set **function equal** to 0
b) distribute the summation operator
c) apply rules of summation for constants
d) solve in terms of the criterion variable, Y
e) substitute descriptive statistics and simplify

Applying each step in a) through e) produces:

a) $\Sigma(Y - a - b_1x_2 - b_2x_2) = 0$

b) $\Sigma Y - \Sigma a - \Sigma b_1x_1 - \Sigma b_2x_2 = 0$

c) $\Sigma Y - na - b_1\Sigma x_1 - b_2\Sigma x_2 = 0$

d) $\Sigma Y = na + b_1\Sigma x_1 + b_2\Sigma x_2$

e) $\Sigma Y = na + b_1(0) + b_2(0)$

Recall that $\Sigma x_1 = \Sigma x_2 = 0$. Dividing through by n gives us the normal equation for a (in simplified form):

$$a=\overline{Y}$$

2. The procedures for finding the normal equation for b_1 are:

a) drop the exponent and set function equal to 0.
b) multiply the function by x_1
c) distribute the x_1 term
d) distribute the summation operator
e) apply rules of summation for constants
f) solve in terms of the criterion variable, Y
g) substitute descriptive statistics and simplify

Applying each step in turn produces:

a) $\Sigma(Y - a - b_1x_1 - b_2x_2) = 0$

b) $\Sigma(Y - a - b_1x_1 - b_2x_2)x_1$

c) $\Sigma(Yx_1 - ax_1 - b_1x_1^2 - b_2x_1x_2)$

d) $\Sigma Yx_1 - \Sigma ax_1 - \Sigma b_1x_1^2 - \Sigma b_2x_1x_2$

e) $\Sigma Yx_1 - a\,\Sigma x_1 - b_1\,\Sigma x_1^2 - b_2\,\Sigma x_1x_2$

f) $\Sigma Yx_1 = a\,\Sigma x_1 + b_1\,\Sigma x_1^2 + b_2\,\Sigma x_1x_2$

g) since $\Sigma Yx_1 = (n-1)r_{y1}S_yS_1$ and $\Sigma x_1^2 = (n-1)S_1^2$ and $\Sigma x_1x_2 = (n-1)r_{12}S_1S_2$,

we can substitute these quantities, and obtain:

$$(n-1)r_{y1}S_yS_1 = 0 + b_1(n-1)S_1^2 + b_2(n-1)r_{12}S_1S_2$$

(recall that $\Sigma x_1 = 0$).

If we divide the last equation by (n-1), we obtain:

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2$$

This is the normal equation in simplified form we used in the derivation (see Table 2).

3. The steps for finding the normal equation for b_2 parallel those for b_1:

   a) drop the exponent 2 and set function equal to 0
   b) multiply the function by x_2
   c) distribute the x_2 term
   d) distribute the summation operator
   e) apply rules of summation for constants
   f) solve in terms of the criterion variable, Y
   g) substitute descriptive statistics and simplify

Applying each step in order:

a) $\Sigma(Y - a - b_1x_1 - b_2x_2) = 0$

b) $\Sigma(Y - a - b_1x_1 - b_2x_2)x_2$

c) $\Sigma(Yx_2 - ax_2 - b_1x_1x_2 - b_2x_2^2)$

d) $\Sigma Yx_2 - \Sigma ax_2 - \Sigma b_1x_1x_2 - \Sigma b_2x_2^2$

e) $\Sigma Yx_2 - a\Sigma x_2 - b_1\Sigma x_1x_2 - b_2\Sigma x_2^2$

f) $\Sigma Yx_2 = a\Sigma x_2 - b_1\Sigma x_1x_2 - b_2\Sigma x_2^2$

g) since $\Sigma Yx_2 = (n-1)r_{y2}S_yS_2$ and

   $\Sigma x_1x_2 = r_{12}S_1S_2$ and $\Sigma x_2^2 = (n-1)S_1^2$

   we can substitute these quantities and obtain:

   $(n-1)r_{y2}S_yS_2 = 0 + b_1(n-1)r_{12}S_1S_2 + b_2(n-1)S_2^2$

If we divide through by $(n-1)$ we have:

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2$$

This was the simplified form of the normal equation for b_2 that was used in the derivation (see Table 2).

D. We now recapitulate. As noted, a normal equation is derived at the point when we solve in terms of the criterion variable, Y. Subsequent steps are used to simplify. The normal equations for a, b_1 and b_2 were:

For a:	$\Sigma Y = na$	+	$b_1\Sigma x_1$	+	$b_2\Sigma x_2$
For b_1:	$\Sigma Yx_1 = a\Sigma x_1$	+	$b_1\Sigma x_1^2$	+	$b_2\Sigma x_1x_2$
For b_2:	$\Sigma Yx_2 = a\Sigma x_2$	+	$b_1\Sigma x_1x_2$	+	$b_2\Sigma x_2^2$

When we simplified the normal equations, we obtained the following set used in the derivation for two predictors.[1]

$$\bar{Y} = a$$

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{12}S_1S_2$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2$$

Readers of the 1982c paper should recognize the remarkable similarity between raw score and standard score normal equations. If the above variables were standardized, each term $S_j^2 = 0$ and a=0 making each normal equation set equal.

[1]We actually disregarded the term a in the derivation because it was seen to "drop out" when it was included in the algebra. It is included here because the slope intercept term is included in the regression equation for criterion score calculation. The formula used is:

$$\hat{Y} = \bar{Y} + b_1x_1 + b_2x_2$$

See Lindeman, et al. for additional methods of writing this equation.

Finding Normal Equations for p Predictors

The rules and method for deriving a set of normal equations when the number of predictors is greater than two are generalizations for the two (or one) predictor case.[1] We will show two methods for the general case. The first method will use the four phase plan. The second is a short-cut technique. But the shorter method depends on first showing the longer one.

---

[1]What are the normal equations for the one predictor model? The reader may find it instructive to derive the normal equations for this linear model. This can be done using the above procedures as guidelines.

ANSWER:

$$\bar{Y} = a$$

$$r_{y1} S_y S_1 = b_1 S_1^2$$

The "multiple" R in this case is the simple Pearson product correlation, r_{y1} which is equal to $b_1 \frac{S_1}{S_y}$. This is obtained from the second equation.

Thus, the regression (prediction) equation upon substitution is:

$$\hat{Y} = a + b_1 x_1$$

$$= \bar{Y} + r_{y1} \frac{S_y}{S_1} x_1$$

Applying the four phase plan gives the following results for the general case.

A. The regression model is:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + \ldots + b_jx_j + \ldots + b_px_p$$

B. The function to be minimized accoring to the least squares criterion is:

$$\Sigma(Y - a - b_1x_1 - b_2x_2 - \ldots - b_jx_j - \ldots - b_px_p)^2$$

C. The procedures for finding the normal equations for a and <u>any</u> b_j term are as follows:

1. In deriving the normal equation for a, regardless of the number of predictors, the result is always the same-- $a = \bar{Y}$.

2. Finding the normal equation for any b_j term can be done in seven steps:

   a) drop the exponent 2 and set the function equal to 0

   b) multiply the function by x_j

   c) distribute the x_j term

   d) distribute the summation operator

   e) apply rules of summation for constants

   f) solve in terms of the criterion variable, Y

   g) substitute descriptive statistics and simplify

Applying these steps in turn produces:

a) $\Sigma(Y - a - b_1x_1 - b_2x_2 - \ldots - b_jx_j - \ldots - b_px_p) = 0$

b) $\Sigma(Y - a - b_1x_1 - b_2x_2 - \ldots - b_jx_j - \ldots - b_px_p)x_j = 0$

c) $\Sigma(Yx_j - ax_j - b_1x_1x_j - b_2x_2x_j - \ldots - b_jx_j^2 - \ldots - b_px_jx_p) = 0$

d) $\Sigma Yx_j - \Sigma ax_j - \Sigma b_1x_1x_j - \Sigma b_2x_2x_j - \ldots - \Sigma b_jx_j^2 - \ldots - \Sigma b_px_jx_p = 0$

e) $\Sigma Yx_j - a\Sigma x_j - b_1\Sigma x_1x_j - b_2\Sigma x_2x_j - \ldots - b_j\Sigma x_j^2 - \ldots - b_p\Sigma x_jx_p = 0$

f) $\Sigma Yx_j = a\Sigma x_j + b_1\Sigma x_1x_j + b_2\Sigma x_2x_j + \ldots + b_j\Sigma x_j^2 + \ldots + b_p\Sigma x_jx_p$

g) $(n-1)r_{yj}S_yS_j = 0 + b_1(n-1)r_{1j}S_1S_j + b_2(n-1)r_{2j}S_2S_j + \ldots + b_j(n-1)S_j^2 + \ldots + (n-1)r_{jp}S_jS_p$

Dividing through by (n-1)

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + \ldots + b_jS_j^2 + \ldots + b_pr_{jp}S_jS_p$$

Thus, the normal equations for any number of predictors in the regression model consist of $a=\bar{Y}$ and p normal equations of the general form $r_{yj}S_yS_j$ defined above.

Alternate Procedure

The above normal equation for any b_j term $(r_{yj}S_yS_j)$ is a general result. Now a much simpler procedure which makes use of this fact will be presented.

Recall that the simple correlation of any variable with itself is equal to 1. That is, $r_{11}=r_{22}=\ldots=r_{jj}=\ldots=r_{pp}=1$. Also recall that the covariance of any variable with itself is equal to the variance of that variable; that is $cov(x_1,x_1)=S_1^2$; $cov(x_2,x_2)=S_2^2$ or, in general, $cov(x_j,x_j)=S_j^2$. Another way to denote $cov(x_1,x_1)$ is S_{11} ; in general we can write : $cov(x_j,x_j) = S_j^2$ or $S_{jj} = S_j^2$.

From these facts, it is possible to write down an entire set of normal equations for any number of predictors. If $r_{yj}S_yS_j$ holds for any b_j term, then it holds for j=1, j=2, j=3,..., j=p. For example, assume j=2 predictors. We know that the set of normal equations will consist of $j \times j = 2 \times 2 = 2^2 = 4$ terms. Thus, first write out the general result for $r_{yj}S_yS_j$ twice as follows:

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j$$

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j$$

Now, substitute the appropriate j value: j=1 for line 1, and j=2 for line 2:

$$r_{y1}S_yS_1 = b_1 r_{11} S_{11} + b_2 r_{21} S_2 S_1 \qquad\qquad r_{y1}S_yS_1 = b_1 S_1^2 + b_2 r_{21} S_2 S_1$$

$$r_{y2}S_yS_2 = b_1 r_{12} S_1 S_2 + b_2 r_{22} S_{22} \quad \text{OR} \quad r_{y2}S_yS_2 = b_1 r_{12} S_1 S_2 + b_2 S_2^2$$

OR

$$r_{y1}S_yS_1 = b_1 S_1^2 + b_2 r_{12} S_1 S_2$$

$$r_{y2}S_yS_2 = b_2 r_{12} S_1 S_2 + b_2 S_2^2$$

The last set shows the subscripts of the correlations between predictors and criterion, and the predictor standard deviations written so that the first subscript is less than the second subscript. As mentioned in the text, this convention makes it easier to read the matrix (and see the symmetry of off diagonal terms).

Example for Five Predictors

To exemplify the procedures for p predictors, we will work through the solution of normal equations for five predictors. We will show the solution by the short-cut method. The long method could be used by applying the steps listed above for any b_j term, but since the shorter method gives identical results, we will not work through the longer method.

We begin by writing out the $5^2 = 25$ terms for the general $r_{yj}S_yS_j$ normal equation. That is, write out $r_{yj}S_yS_j$ on five separate lines.

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + b_3r_{3j}S_3S_j + b_4r_{4j}S_4S_j + b_5r_{5j}S_5S_j$$

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + b_3r_{3j}S_3S_j + b_4r_{4j}S_4S_j + b_5r_{5j}S_5S_j$$

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + b_3r_{3j}S_3S_j + b_4r_{4j}S_4S_j + b_5r_{5j}S_5S_j$$

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + b_3r_{3j}S_3S_j + b_4r_{4j}S_4S_j + b_5r_{5j}S_5S_j$$

$$r_{yj}S_yS_j = b_1r_{1j}S_1S_j + b_2r_{2j}S_2S_j + b_3r_{3j}S_3S_j + b_4r_{4j}S_4S_j + b_5r_{5j}S_5S_j$$

Substitute the appropriate j value (j=1 for line 1, j=2 for line 2, etc.); set $r_{11}=1, r_{22}=2$, etc. and set $S_{11}=S_1^2$, $S_{22} = S_2^2$ etc.

$$r_{y1}S_yS_1 = b_1S_1^2 + b_2r_{21}S_2S_1 + b_3r_{31}S_3S_1 + b_4r_{41}S_4S_1 + b_5r_{51}S_5S_1$$

$$r_{y2}S_yS_2 = b_1r_{12}S_1S_2 + b_2S_2^2 + b_3r_{32}S_3S_2 + b_4r_{42}S_4S_2 + b_5r_{52}S_5S_2$$

$$r_{y3}S_yS_3 = b_1r_{13}S_1S_3 + b_2r_{23}S_2S_3 + b_3S_3^2 + b_4r_{43}S_4S_3 + b_5r_{53}S_5S_3$$

$$r_{y4}S_yS_4 = b_1r_{14}S_1S_4 + b_2r_{24}S_2S_4 + b_3r_{34}S_3S_4 + b_4S_4^2 + b_5r_{54}S_5S_4$$

$$r_{y5}S_yS_5 = b_1r_{15}S_1S_5 + b_2r_{25}S_2S_5 + b_3r_{35}S_3S_5 + b_4r_{45}S_4S_5 + b_5S_5^2$$

If one desires, the subscripts may be reversed for variables in the upper right hand triangle to render the first less than the second. The result is the same set of normal equations that would be obtained if the longer method were used to derive the normal equations.[1]

---

[1] The author would be pleased to receive comment and reactions by readers of this paper and others that appear in this series. My intention is to prepare a textbook of proofs and derivations for social science students. I have long felt the need to bridge the gap between the standard applied statistics (and psychometrics) textbooks currently on the market and mathematical statistics. The mathematical sophistication of students entering college and university is rising steadily, and a textbook such I am contemplating would make a contribution, I feel. While it is true that a "real" understanding of statistical (and probability) **theory** requires substantial mathematical coursework, it is nonetheless true that more in the way of explanation and justification of results in probability and statistics is possible. It is my belief that a textbook showing detailed presentations of proofs/derivations would be a welcome addition to the market.

I would like to hear from readers (students, professors and others) regarding these papers. For example, are they clear? Are there proofs that you would like to see (statistics or psychometrics) in this format? Please remember, at this time I am limiting my selections to those which can be presented with algebra.

I welcome comments on any level from readers of these papers. My mailing address is:

Francis J. O'Brien, Jr.
106 Morningside Drive, Apartment # 5
New York, New York 10027

Appendix B

ERRATA for " A derivation of the sample multiple Correlation formula for standard scores" ED 223 429

Page	Now Reads	Correct to*
2, Derivation for Two Predictors	Let us review some concepts, notation ...	Let us review some of the concepts, notation ...
3, footnote 1	[1] If it is understood that the all summations range from i=1 to i=n, then we can drop the summation limits.all together;	[1] If it is understood that all the summations range from i=1 to i=n, then we can drop the summation limits altogether;
21, first formula	$\text{cov}(Z_Y, B_1Z_1, B_2Z_2, B_3Z_3, \ldots, B_jZ_j, \ldots, B_pZ_p,$	$\text{cov}(Z_Y, B_1Z_1 + B_2Z_2 + B_3Z_3 + \ldots + B_jZ_j + \ldots + B_pZ_p)$
22	$\text{corr}(Z_Y, B_1Z_1, B_2Z_2, B_3Z_3, \ldots, B_jZ_j, \ldots B_pZ_p)$	$\text{corr}(Z_Y, B_1Z_1 + B_2Z_2 + B_3Z_3 + \ldots + B_jZ_j + \ldots + B_pZ_p)$.
27, statement under Plan	D	add period, D.
27, two lines under previous erratum	demomstrate	demonstrate
31, line 2	consisdered	considered
33, 4 sentences from bottom	first first	first

*Corrected prose is underlined, but formulas are rewritten with applied corrections only.

NOTE: "Page" refers to original numbers in upper right hand corner.

REFERENCES

Goodman, A.W. Calculus for the Social Sciences Philadelhpia: W.B. Saunders Company, 1977.

Hoel, Paul G., Sidney C. Port and Charles J. Stone. Introduction to Statistical Theory. Boston: Houghton Mifflin, 1971.

Lindeman, Richard H., Ruth Gold and Peter Merenda. Bivariate and Multivariate Analysis . Chicago: Scott, Foresman and Co., 1982.

O'Brien, Francis J. Jr. A proof that t^2 and F are identical: the general case, 1982a. ERIC ED 215 894.

_______. Proof that the sample bivariate correlation coefficient has limits $\pm$ 1, 1982b. ERIC ED 216 874

_______. A derivation of the sample multiple correlation formula for standard scores, 1982c. EDIC ED 223 429.