

## Chapter 5: Multiple Regression

### 5.1 Partial and semi-partial correlation

Before starting on multiple regression *per se*, we need to consider the concepts of **partial** and **semi-partial correlation**. Both of these can be computed when you have 3 (or more) variables. As Glass and Hopkins (1984) argue, “These concepts are of interest in themselves”, and “they are also useful for understanding the subsequent topics of multiple regression and multiple correlation.”

#### 5.1.1 Partial correlation

Perhaps the best way to understand *partial correlation* is to consider a simple example. I will borrow the one given by Howell (1997). It concerns a researcher who wished to know if there is a relationship between success in university and annual income. To his great delight, the correlation between these variables was highly significant. Note however, that both variables are almost certainly correlated positively with IQ. That is, people with high IQs tend to do well in university, and also earn a lot of money. So as Howell says, perhaps “the correlation between income and college success is an artifact of this relationship.”

The partial correlation between college success and income, with IQ partialled out of both variables, can be very useful in a situation like this. Conceptually, it is the correlation between **that portion of college success that cannot be explained by IQ** and **that portion of annual income that cannot be explained by IQ**.

What exactly is “that portion of college success that cannot be explained by IQ”? For just a moment, imagine that college success and IQ are the only variables we have. Let  $Y$  = college success, and  $X$  = IQ. For each subject, we could compute a deviation score as follows:

$$\text{Deviation score} = (Y_i - \bar{Y}) \quad (5.1)$$

If we performed the regression of  $Y$  on  $X$  (college success on IQ), we could partition each deviation score into two components as follows:

$$(Y_i - \bar{Y}) = (Y' - \bar{Y}) + (Y_i - Y') \quad (5.2)$$

This equation ought to look familiar, because you saw it in section 4.10 of the previous set of notes. You may recall that the difference between  $Y'$  and the mean of  $Y$  is the part of the deviation from the mean that can be explained by the linear relationship between  $X$  and  $Y$ ; and  $(Y_i - Y')$  is the part that **cannot** be explained by the linear relationship between  $X$  and  $Y$ . Translating this into the terms of the present example, we have:

$(Y' - \bar{Y})$  = the portion of college success that **can** be explained by IQ;  
and  
 $(Y_i - Y')$  = error in prediction (or **residual**)  
= the portion of college success that **cannot** be explained by IQ.

To find the portion of **annual income** that cannot be explained by IQ, we could let  $Y$  = annual income and  $X$  = IQ, and carry out the regression of  $Y$  on  $X$ . The **residuals** (i.e., errors in prediction) from this regression would correspond to the **portion of annual income that cannot be explained by IQ**.

And so, after carrying out these two regressions, we are left with two residuals (i.e.,  $Y_i - Y'$  scores) for each subject. As Howell (1997) suggests, you could think of these residuals as purified measures--"purified of the influence of IQ."

Finally then, **the partial correlation between college success and income, with the influence of IQ partialled out of both variables** is equal to **the correlation between these two sets of residuals**. As you might have suspected, this is a conceptual description of partial correlation. When it comes to calculation of the partial correlation, you would not go through all of the aforementioned steps. Rather you would use a **computational formula**, such as:

$$r_{01.2} = \frac{r_{01} - r_{02}r_{12}}{\sqrt{(1 - r_{02}^2)(1 - r_{12}^2)}} \quad (5.3)$$

The subscripts on  $r_{01.2}$  indicate that it is the **correlation between variables 0 and 1 with variable 2 partialled out of both**. If we let  $X_0$  = annual income,  $X_1$  = college success, and  $X_2$  = IQ, then  $r_{01.2}$  is the correlation between annual income and college success with IQ partialled out of both.

### 5.1.2 Semi-partial correlation

**Semi-partial correlation** (also known as **part correlation**) is closely related to partial correlation. It differs in that *the influence of a third variable is removed from, or partialled out of only one of the two variables you wish to correlate*. Continuing with the preceding example, we might wish to compute the semi-partial correlation between annual income and college success with IQ partialled out of college success. Conceptually, this would be the correlation between annual income and the residual scores (i.e., errors in prediction) from the regression of college success on IQ. Putting this another way, it is the correlation between **income** (including the part that can be explained by IQ) and **that portion of college success that cannot be explained by IQ**.

As with partial correlation, there is a **computational formula** for semi-partial correlation:

$$r_{0(1.2)} = \frac{r_{01} - r_{02}r_{12}}{\sqrt{1 - r_{12}^2}} \quad (5.4)$$

### 5.1.3 Higher order partial correlation

Hitherto, our discussion of partial correlation has been confined to partialling out the effects of one variable from two other variables we wish to correlate. This sort of partial correlation is called **first-order** partial correlation. Note that partial (and semi-partial) correlation techniques can also be used to remove the influence of **2 or more variables**. A partial correlation that removes the effects of 2 variables is called **second-order partial correlation**. The partial correlation between  $X_0$  and  $X_1$  with  $X_2$  and  $X_3$  partialled out of both can be computed as follows:

$$r_{01.23} = \frac{r_{01.2} - r_{03.2}r_{13.2}}{\sqrt{(1 - r_{03.2}^2)(1 - r_{13.2}^2)}} \quad (5.5)$$

A higher order partial correlation between  $X_0$  and  $X_1$  with the effects of predictors 2 through  $p$  removed would be symbolized as  $r_{01.23\dots p}$ . The computational formulae for higher order partial correlations are quite complex, and so they are typically computed using computerized statistics packages.

## 5.2 Multiple linear regression

Simple linear regression is concerned with using the relationship between two variables for purposes of prediction. Multiple regression extends this by increasing the number of predictors. In this case, the regression equation looks like:

$$Y' = b_1X_1 + b_2X_2 + \dots + b_pX_p + a \quad (5.6)$$

Following Howell (1997), I will use  $p$  to represent the number of predictor variables. The  $b$ 's are **regression coefficients** (like the slope constant in simple linear regression), and there is one for each predictor. As in simple regression,  $a$  is the intercept. Note however that in multiple regression, it is not uncommon to see  **$b_0$  used to represent the intercept**. And so the regression equation often appears as follows:

$$Y' = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p \quad (5.7)$$

Note as well that  $X_0$  is sometimes used (in place of  $Y'$ ) to represent the predicted score.

The least squares criterion that we discussed in relation to simple linear regression also applies to multiple regression. That is:

$$\Sigma (Y - Y')^2 = \mathbf{a\ minimum} \quad (5.8)$$

### 5.3 Standardized regression equation

If one were to standardize all variables (i.e., convert all raw scores to z-scores) before carrying out multiple regression, the result would be the standardized regression equation:

$$Z_{Y'} = \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p \quad (5.9)$$

You may recall that when simple linear regression is carried out on standard scores rather than raw scores, the regression line passes through the point (0,0). Thus the intercept ( $a$ ) is equal to zero, and can be left off of the equation. Similarly in the standardized multiple regression equation, the intercept ( $b_0$ ) is equal to zero, and so can be left off.

Note as well that the regression coefficients in this equation (i.e.,  $\beta_1, \beta_2, \dots, \beta_p$ ) are commonly referred to as **beta weights**. You may also come across other names for these coefficients, including **standardized regression coefficients**, or **standard partial regression coefficients**. As Glass and Hopkins (1984) explain, “standard to denote z-scale, and partial because the effects of the other variables are ‘held constant’ [or partialled out]. In other words,  $\beta_1$  is short for  $\beta_{1.23\dots m}$ .” (Note that Glass & Hopkins use  $m$  rather than  $p$  to stand for the number of predictors.)

### 5.4 Computational formulae with 2 predictor variables

When you have only 2 predictor variables, the formulae involved in computing multiple regression are not too complicated. Glass and Hopkins (1984) provide an example in which predictor 1 ( $X_1$ ) is a reading readiness score (RR), predictor 2 ( $X_2$ ) is IQ, and the criterion variable ( $Y$ ) is a reading score. The means, standard deviations, and pairwise correlations for these 3 variables are summarized in Table 5.1:

**Table 5.1** Intercorrelations between reading performance ( $Y$ ), Reading Readiness ( $X_1$ ), and IQ ( $X_2$ ) for 157 First-grade pupils

Variable	$X_2$	$Y$	Mean	s.d.
RR ( $X_1$ )	.466	.612	49.0	10.3
IQ ( $X_2$ )		.541	102.8	14.7
Reading Score ( $Y$ )			26.0	8.1

When you have only 2 predictor variables, the standardized regression equation can be written as:

$$Z_{Y'} = \beta_1 Z_1 + \beta_2 Z_2 \quad (5.10)$$

And the least squares estimates of the standardized regression coefficients (when  $p = 2$ ) are given by:

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} \quad \beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} \quad (5.11)$$

Substituting values from Table 5.1 yields the following:

$$\beta_1 = \frac{(0.612) - (0.541)(0.466)}{1 - (0.466)^2} = 0.460 \quad (5.12)$$

$$\beta_2 = \frac{(0.541) - (0.612)(0.466)}{1 - (0.466)^2} = 0.327 \quad (5.13)$$

And so, the standardized regression equation is:

$$Z_{Y'} = 0.460Z_1 + 0.327Z_2 \quad (5.14)$$

The standardized regression equation can be converted to a raw-score equation quite easily. The raw-score regression coefficient for predictor  $p$  can be computed as follows:

$$b_p = \beta_p \frac{s_Y}{s_p} \quad (5.15)$$

And as in simple linear regression,  $b_0$  (the intercept) can be obtained by rearranging and solving the regression equation after substituting in the mean value for each of the  $p$  predictors:

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 - \dots - b_p\bar{X}_p \quad (5.16)$$

For this example, the raw-score equation works out to:

$$Y' = -10.24 + (0.362)X_1 + (0.180)X_2 \quad (5.17)$$

## 5.5 Multiple correlation

Multiple correlation is the **simple correlation between  $Y$  and  $Y'$**  when there are 2 or more predictors:

$$R_{Y.12\dots p} = r_{YY'} \quad (5.18)$$

In simple linear regression, we saw that  $r^2$  gives the proportion of  $SS_Y$  that can be accounted for by the linear relationship between  $X$  and  $Y$ . In the case of multiple regression and correlation,  $R^2$  tells us the proportion of  $SS_Y$  that is accounted for by the  $p$  predictors.

When there are only 2 predictors,  $R^2$  can be calculated as follows:

$$R_{Y.12}^2 = \frac{r_{Y1}^2 + r_{Y2}^2 - 2r_{Y1}r_{Y2}r_{12}}{1 - r_{12}^2} \quad (5.19)$$

Using the data in Table 5.1, we can see that,

$$R_{Y.12}^2 = \frac{(0.612)^2 + (0.541)^2 - 2(0.612)(0.541)(0.466)}{1 - (0.466)^2} = 0.458 \quad (5.20)$$

Therefore, 45.8% of the variance in reading scores ( $Y$ ) can be accounted for by reading readiness (RR) and IQ.

Returning to Table 5.1, note that if you were to use only one predictor variable, RR would be a better predictor than IQ, because  $r^2_{Y1} = (.612)^2 = .375$ , whereas  $r^2_{Y2} = (.541)^2 = .293$ . In other words, when they are used as single predictors, RR accounts for 37.5% of the variance in reading scores, and IQ accounts for 29.3% of the variance in reading scores.

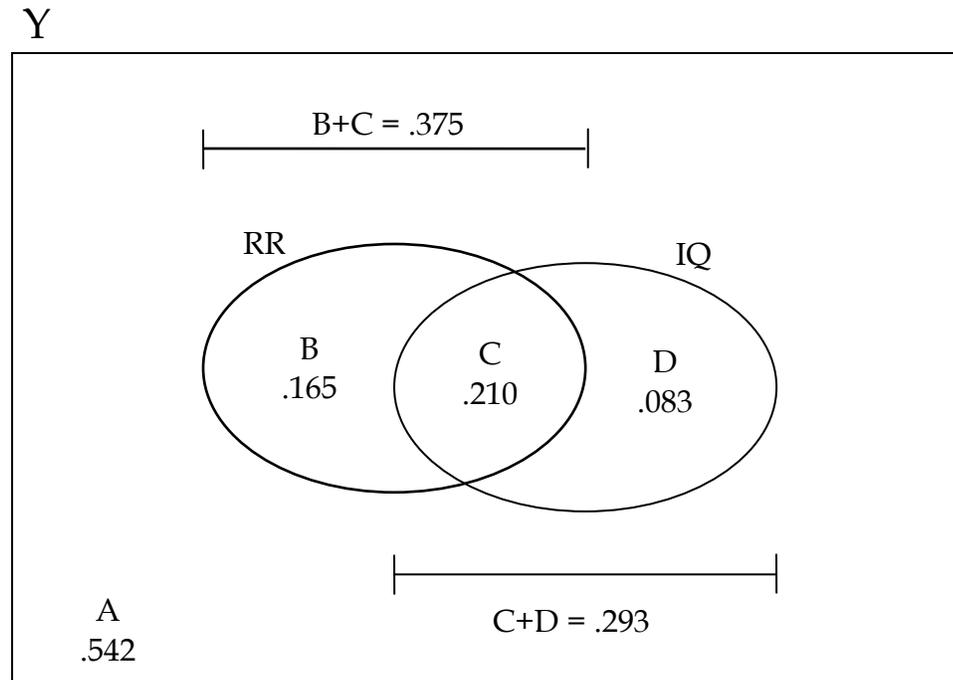
Note as well that the amount of variability in reading scores accounted for by **both** predictors is not substantially greater than the amount accounted for by RR alone. That is, 45.8% (for both predictors) is only 8.3% more than 37.5%. How can this be? If IQ accounts for 29.3% of the variance, why is there a mere 8.3% increase in variance accounted for when it is added to RR?

The answer to this conundrum is that RR and IQ are **not independent**. (If they were independent, the correlation between them would be equal to zero.) Putting this another way, some of the  $Y$  variability accounted for by RR is also accounted for by IQ. Variance in  $Y$  that is explained by two (or more) variables is often called **redundant variance**, whereas variance that is explained by only one predictor variable is called **unique variance**.

Figure 5.1 illustrates the concepts of unique and redundant variance. In the figure, the total variability of the  $Y$ -scores is represented by the area inside of the rectangle. For ease of discussion, this amount is said to be equal to 1. (We could have just as easily set it equal to 100%, but chose 1 because we want to talk about squared correlation coefficients which give us **proportions** rather than percentages of variance accounted for.) The circles (which came out looking more like ellipses!) represent the amounts of  $Y$  variability accounted for by RR and IQ. Area A is the area inside of the rectangle, but outside all of the circles. It represents the residual variance, or the amount of  $Y$  variability that **cannot** be explained by RR and IQ. It can be obtained as follows:  $A = 1 - R^2_{Y.12} = 1 - .458 = .542$ .

The area in (B+C+D) represents the amount of  $Y$  variability that is explained when RR and IQ are both used as predictors, which is  $R^2_{Y.12}$ , or .458. The area in B+C represents the amount of  $Y$  variability that can be explained by RR as a single predictor, which is .375; and the area in C+D represents the proportion that can be explained by IQ as a single predictor, or .293.

Area D represents the proportion of  $Y$  variability that can be explained by IQ *over and above that which is already explained by RR*. As we saw earlier, this is equal to  $.458 - .375 = .083$ . Area B represents the proportion of  $Y$  variability that is explained by RR *over and above that which is explained by IQ*. This is equal to  $.458 - .293 = .165$ . Note that Areas B and D both represent proportions of the variability of  $Y$  that are explained by only one variable. As such, they each represent **unique variance**.



**Figure 5.1** Redundant and unique variance among two predictors and one criterion, based on data in Table 5.1

Area C, on the other hand, represents the proportion of  $Y$  variability that is explained by **both** RR and IQ. This is the **redundant variance**, or common variance. It is equal to  $(B+C+D) - B - D = .458 - .165 - .083 = .210$ .

### 5.6 Partial and semi-partial correlation revisited

Those of you who are still awake may be thinking, “This talk of unique and redundant variance reminds me of something I read about earlier. Hmm, what could it be?” What it *should* be reminding you of is partial and semi-partial correlation.

Looking at Figure 5.1 again, note that the proportion of  $Y$  variability accounted for by RR is  $.165 + .210 = .375$ . This is also the value of  $r^2$  for RR and  $Y$ . Thus, after using RR for prediction, the **amount of  $Y$  variability that is not explained by RR** is  $1 - .375 = .625$ . The amount of  $Y$  variance that is explained uniquely by IQ is represented by area D, and is equal to  $.083$ . And so  $.083/.625$ , or  $.133$ , represents the **additional amount of  $Y$  that IQ explains, relative to the amount that remains to be explained**. As such, it is **the squared partial correlation between  $Y$  and IQ with RR partialled out of both**. I won’t hold my breath while I’m waiting, but some of you may wish to verify this by calculating the partial correlation with equation 5.3, and then squaring it. Your answer should be around  $.133$ . (If you do try this note that equation 5.3 is set up for partialling variable 2 out of variable 1 and the criterion. Since you want to partial RR out of IQ and  $Y$ , it may be easier if you call RR  $X_2$  and IQ  $X_1$ , which is opposite to what we did earlier.)

On the other hand, if you took area D and divided it by the total variability of  $Y$ , which is equal to 1, you would get  $.083/1 = .083$ . This is the **additional amount** of  $Y$  that IQ explains. Note that it is **relative to the entire amount of  $Y$  variability**, not to the amount that remains to be explained. This is the same thing as the **squared semi-partial correlation** (with RR partialled out of IQ, but not out of  $Y$ ). Again, you might wish to check this by computing the semi-partial correlation (with equation 5.4) and squaring it. (Again note that things might be easier if you call RR  $X_2$  and IQ  $X_1$ .)

#### Summary of Figure 5.1

A	$= 1 - R^2_{Y.12}$	= the residual (unexplained) variation in $Y$
B+C+D	$= R^2_{Y.12}$	= the proportion of $Y$ explained by both predictors
B+C	$= r^2_{Y1}$	= squared correlation between RR and $Y$
C+D	$= r^2_{Y2}$	= squared correlation between IQ and $Y$
B	$= r^2_{Y(1.2)}$	= the squared semi-partial correlation
D	$= r^2_{Y(2.1)}$	= the other squared semi-partial correlation
B/(A+B)	$= r^2_{Y1.2}$	= the squared partial correlation
D/(A+D)	$= r^2_{Y2.1}$	= the other squared partial correlation

### 5.7 Multiple correlation revisited

We are now in a position to say a bit more about multiple correlation. Howell (1997) says that “we can think of  $R$  as being based on as much information as possible from one variable, any *additional, nonredundant* information from a second, and so on.” If we think about  $R^2$  rather than  $R$ , this can be translated into the following:

$$R^2_{0.123\dots p} = r^2_{01} + r^2_{0(2.1)} + r^2_{0(3.12)} + \dots + r^2_{0(p.123\dots p-1)} \quad (5.21)$$

The first term on the right side of this equation is the square of the simple correlation between  $Y$  and predictor 1. The second term is the square of the semi-partial correlation between  $Y$  and predictor 2 with predictor 1 partialled out of predictor 2. The next term is the square of the (second order) semi-partial correlation between  $Y$  and predictor 3 with 1 and 2 both partialled out of 3. And so on.

Finally, it should be noted that if all of the predictors are mutually independent, then equation 5.21 reduces to:

$$R^2_{0.123\dots p} = r^2_{01} + r^2_{02} + r^2_{03} + \dots + r^2_{0p} \quad (5.22)$$

The reason for this is that if the variables are all independent of each other, there is no redundant (or common) variance to partial out.

### 5.8 Testing the significance of $R^2$

As we saw in section 4.11 (in the previous set of notes), the significance of  $R^2$  can be tested with an  $F$ -ratio:

$$F_{(p, N-p-1)} = MS_{Y'} / MS_{error} \quad (5.23)$$

Note that  $MS_{Y'}$  is often called  $MS_{regression}$ , and that there are  $p$  degrees of freedom associated with it (where  $p$  = the number of predictors). The number of degrees of freedom associated with  $MS_{error}$ , or  $MS_{residual}$ , is  $N-p-1$ . The null hypothesis for this  $F$ -test states that none of the variability in the criterion variable can be accounted for by the  $p$  predictor variables.

Finally, note that the **standard error of estimate** in multiple regression is equal to the square root of  $MS_{residual}$ .

### 5.9 Constructing a regression equation: Stepwise method

In multiple regression, the usual goal is to come up with a **subset** of the  $p$  predictors that explains a substantial proportion of the overall variability in  $Y$ . Imagine, for example, that you have 10 predictor variables. It could be that a **reduced model** that includes only the 3 best predictors can account for virtually all of the variance accounted for by the **full model** (with all 10 predictors). If so, the reduced model is preferred, because it is more parsimonious. There are various ways to go about coming up with this subset of predictors (see Howell 1997, 1992, for discussion of these). One method, which I will illustrate here, is called **stepwise regression**.

Before going any further, let be perfectly clear about one thing: According to Howell (1997) and many other authors, **stepwise regression is not a very good technique to use in actual practice**--in fact, Henderson and Denison (1989) suggested that it should be called "**unwise regression**". The reasons for this are explained by Howell. I also urge readers to see Rich Ulrich's Stats FAQ, which has a page that lists several objections to stepwise selection (<http://www.pitt.edu/~wpilib/statfaq/regfaq.html>).

Despite the very serious objections to stepwise regression, I will introduce it here for two reasons:

1. It has been so commonly used in the past that you are sure to read about it in journal articles. Therefore, it behooves you to have some understanding of it.
2. I can use it to illustrate how regression models are compared to test hypotheses about variables. This is done even when one uses better approaches to model building such as hierarchical regression (where the *researcher*, not a computer algorithm, decides which variables are to be included in the model, and in which order).

The variables shown in Table 5.1 are only two of five variables that were examined for the purpose of predicting how successful grade 1 students would be at reading. The complete set of variables was as follows:

$X_1$  = reading readiness (RR) as measured with a standardized test

$X_2$  = intelligence quotient (IQ)

$X_3$  = age of the child (AGE)

$X_4$  = socioeconomic status (SES)

$X_5$  = sex of the child (SEX); Male = 1, Female = 2

$Y$  = score on a standardized reading test at the end of grade 1

**Table 5.2 Intercorrelations between reading performance ( $Y$ ) and five predictor variables for 157 First-grade pupils**

	$X_2$	$X_3$	$X_4$	$X_5$	$Y$	Mean	s.d.
$X_1$	.466	.252	.160	.055	.612	49.0	10.3
$X_2$		-.119	.265	.072	.541	102.8	14.7
$X_3$			-.053	-.081	.050	76.8	4.11
$X_4$				-.040	.209	32.3	21.6
$X_5$					.197	1.48	.501
$Y$						26.0	8.10

It is clear from the correlation matrix in the shaded part of Table 5.2 that the best single predictor of reading success is  $X_1$ , reading readiness. (RR has the highest simple correlation with  $Y$ .) So the first step in stepwise regression is to carry out simple linear regression with  $X_1$  as the predictor and  $Y$  as the criterion.

#### Step 1

- simple regression of  $Y$  on  $X_1$
- $Y' = .481X_1 + 2.41$
- $r^2_{Y1} = .375$
- $s_{Y.1} = 6.428$

As a single predictor, RR accounts for 37.5% of the variance in  $Y$ . That leaves 62.5% of the variance to be explained. We will select the next variable to enter into our regression equation on the basis of **which one explains the greatest amount of that remaining variance**. That is, we will look at the **partial correlations** of the 4 remaining predictor variables:

$$r_{Y2.1} = .366$$

$$r_{Y3.1} = -.135$$

$$r_{Y4.1} = .142$$

$$r_{Y5.1} = .206$$

Clearly,  $X_2$  (IQ) explains the greatest amount of residual variance. Therefore, in Step 2, we will carry out multiple regression with  $X_1$  and  $X_2$  as predictors.

*Step 2*

- multiple regression with  $X_1$  and  $X_2$  as predictors
- $Y' = .362X_1 + .180X_2 - 10.24$
- $R^2_{Y.12} = .458$
- $s_{Y.12} = 6.003$

Note that as a result of including  $X_2$  as a predictor,  $R^2$  has increased from .375 to .458. Note as well that the standard error of estimate has decreased from 6.428 to 6.003. Thus, it appears that prediction is more accurate with RR and IQ as predictors than it is with RR alone.

But is this apparent improvement statistically significant? There are a few ways to address this question, but the most common is to test the increment in  $R^2$ . Let me remind you that  $R^2$  tells you the proportion of  $Y$  variance accounted for by a particular regression model. You may also recall that the ratio of two variances is equal to  $F$ . Thus it should not surprise you to discover that a test on the increment in  $R^2$  can be done with an  $F$ -test:

$$F_{(f-r, N-f-1)} = \frac{(N-f-1)(R_f^2 - R_r^2)}{(f-r)(1-R_f^2)} \quad (5.24)$$

where

$N$  = the number of subjects

$f$  = the number of predictors in the “fuller” of the 2 models

$r$  = the number of predictors in the “reduced” model

Plugging in the appropriate numbers from our example, we find:

$$F_{(1,154)} = \frac{(154)(.458 - .375)}{(1)(1 - .458)} = \frac{12.782}{.542} = 23.583 \quad (5.25)$$

This value of  $F$  is significant at the .001 level, and so we conclude that the inclusion of  $X_2$  as a predictor does improve prediction significantly.

Before going on to Step 3, we need to look at the partial correlations of the remaining 3 predictors, with predictors 1 and 2 partialled out:

$$r_{Y3.12} = -.038$$

$$r_{Y4.12} = .069$$

$$r_{Y5.12} = .201$$

From these figures, it is clear that  $X_5$ , the sex of the child, is the predictor that explains the largest amount of the remaining, unexplained variance.

**Step 3**

- multiple regression with  $X_1$ ,  $X_2$ , and  $X_5$  as predictors
- $Y' = .358X_1 + .175X_2 + 2.40X_5 - 13.16$
- $R^2_{Y.125} = .480$
- $s_{Y.125} = 5.900$

Note again that as the number of predictors is increased from 2 to 3, the value of  $R^2$  jumps from .458 to .480; and the standard error of estimate falls from 6.003 to 5.900. Again therefore, prediction appears to be more accurate with 3 predictors than with 2. However, as before, we need to test the significance of the increment in  $R^2$ . This time,  $f = 3$ , and  $r = 2$ : That is, there are 3 predictors in the fuller of the two models being compared, and 2 predictors in the reduced model:

$$F_{(1,153)} = \frac{(153)(.480 - .458)}{(1)(1 - .480)} = \frac{3.366}{.520} = 6.473 \quad (5.26)$$

This increment in  $R^2$  is also significant at the .025 level. And so, assuming that we had set  $\alpha = .05$  ahead of time, we would carry on to the next step.

The partial correlations for the remaining two predictors, with variables 1, 2, and 5 partialled out are:

$$\begin{aligned} r_{Y3.125} &= -.021 \\ r_{Y4.125} &= .083 \end{aligned}$$

Because the absolute value of the partial correlation for  $X_4$  (SES) is higher, we will include it as a predictor in Step 4.

**Step 4**

- multiple regression with  $X_1$ ,  $X_2$ ,  $X_5$ , and  $X_4$  as predictors
- $Y' = .356X_1 + .167X_2 + 2.46X_5 + .023X_4 - 13.01$
- $R^2_{Y.1254} = .483$
- $s_{Y.1254} = 5.899$

As in the previous steps,  $R^2$  has increased (from .480 to .483) and the standard error of estimate has decreased (from 5.900 to 5.899) as a result of adding another predictor. Note however that the changes are now quite a bit smaller than before. The  $F$ -test for the increment in  $R^2$  yields:

$$F_{(1,152)} = \frac{(152)(.483 - .480)}{(1)(1 - .483)} = \frac{.456}{.517} = .882 \quad (5.27)$$

Clearly, this value of  $F$  is not significant. In other words, prediction does not improve significantly with the addition of SES to our model. Therefore, we can stop, and revert to the model with 3 predictors:

$$Y' = .358X_1 + .175X_2 + 2.40X_5 - 13.16 \quad (5.28)$$

This model accounts for 48% of the variance in the  $Y$  scores, and the standard error of estimate is 5.900.

#### *Forward selection*

Before moving on, let me draw your attention to this comment on stepwise regression from Howell (1997):

At each step in the process, before we add a new variable we first ask whether a variable that was added on an earlier step should now be removed on the grounds that it is no longer making a significant contribution. If the test on a variable falls below “ $F$  to remove” (or “ $p$  to remove”), that variable is removed before another variable is added. Procedures that do not include this step are often referred to as **forward selection** procedures.

In the example we have just worked through, we did not eliminate any variables before going on to the next step. In other words, we used a forward selection procedure.

### 5.10 Other selection methods

Let me emphasize once more that stepwise regression is just one method for building a regression equation, and that according to many authors, it is not a very good one. The best approach, from a statistical point of view, is for the researcher to determine (guided by theory or previous literature, perhaps) which variables should be included in the model, and the order in they should be entered. This form of model building is often described in textbooks as *hierarchical* model building.

### 5.11 $R$ as an inferential statistic: Correcting for bias

A multiple regression equation “does its best to account for every bump and wiggle (including sampling error) in the data” (Howell, 1997, p. 543). Because of this,  $R^2_{Y.12...p}$  overestimates the correlation between  $Y$  and  $Y'$  in the population,  $\rho^2_{Y.12...p}$ . The value of  $R^2$  can be adjusted to remove this bias as follows:

$$adj R^2 = 1 - \frac{(1 - R^2)(N - 1)}{(N - p - 1)} \quad (5.29)$$

The adjusted  $R^2$ , which is displayed by SPSS and other programs, should always be used when your interest is in estimating the correlation in the population.

---

## Review Questions

1. Imagine looking at a large number of cities with varying populations, and recording for each city the number of prostitutes ( $X_1$ ) and the number of clergymen ( $Y$ ). If you actually did this, you would almost certainly find a strong positive correlation--no offense intended to any clergymen who might be reading this. What is the likely explanation for this correlation? How could you verify that your explanation is correct?
  2. Explain the difference between partial and semi-partial correlation. (It may be helpful to use an example.)
  3. Why is there no intercept constant in a standardized regression equation?
  4. What are redundant and unique variance? (Use an example if it helps.)
  5. When all  $p$  predictors are not mutually independent, then:  

$$R^2_{0.123\dots p} = r^2_{01} + r^2_{0(2.1)} + r^2_{0(3.12)} + \dots + r^2_{0(p.123\dots p-1)}$$
 But when all predictors are mutually independent:  

$$R^2_{0.123\dots p} = r^2_{01} + r^2_{02} + \dots + r^2_{0p}.$$
 Explain why this is so.
  6. How does one test the significance of  $R^2$ ? What is the null hypothesis?
  7. How does one test the significance of the change in  $R^2$ ? What is the null hypothesis?
  8. Is stepwise regression a good method for selecting the variables to enter into a regression equation? Why or why not? List 2 other methods, and provide a brief explanation of each.
  9. What is forward selection?
-