Multiple R^2 and Partial Correlation/Regression Coefficients[©]

b_i is an unstandardized partial slope.

Consider the case where we have only two predictors, X_1 and X_2. Were we to predict Y from X_2 and predict X_1 from X_2 and then use the residuals from X_1, that is, $(X_1 - \hat{X}_{1 \bullet 2})$, to predict the residuals in Y, that is, $(Y - \hat{Y}_2)$, the slope of the resulting regression would be b_1. That is, b_1 is the number of units that Y changes per unit change in X_1, **after we have removed the effect of X_2 from both X_1 and Y**. Put another way, b_i is the average change in Y per unit change in X_i with all other predictor variables held constant.

β is a standardized partial slope.

Look at the formulas for a trivariate multiple regression.

$$\hat{Z}_y = \beta_1 Z_1 + \beta_2 Z_2 \qquad\qquad \beta_1 = \frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2} \qquad\qquad \beta_2 = \frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2}$$

β_1 represents the unique contribution of X_1 towards predicting Y in the context of X_2. We must remove the effect of X_2 upon both X_1 and Y to obtain this unique contribution. If you look at the formula for β_1 you will see how this is done: The larger r_{y1}, the larger the β_1. Also, the larger the r_{y2} and the r_{12}, the smaller the β_1 (due to greater redundancy between X_1 and X_2 with respect to their overlap with Y).

Were we to predict Z_Y from Z_2, and Z_1 from Z_2, and then use the residuals from Z_1, that is, $(Z_1 - \hat{Z}_{1 \bullet 2})$, to predict the residuals in Z_Y, that is, $(Z_Y - \hat{Z}_{Y \bullet 2})$, the <u>slope</u> of the resulting regression would be β_1. That is, β_1 is the number of standard deviations that Y changes per standard deviation change in X_1 a**fter we have removed the effect of X_2 from both X_1 and Y**. It should be clear that the value of β_i can be greatly affected by the correlations of other predictors with Y and with X_i. Removing from X_i its overlap with other predictors usually reduces its correlation with Y, but it can increase it (when the variance removed from X_i is variance that is not related to Y).

R^2 can be interpreted as a simple r^2, a proportion of variance explained.

$$R^2_{Y \bullet 12 \ldots i \ldots p} = r^2_{y\hat{y}} = \frac{\sigma^2_{\hat{y}}}{\sigma^2_y}$$

The variance in predicted Y, that is, $\sigma^2_{\hat{y}}$, represents differences in Y due to the linear "effect" upon Y of the optimally weighted combination of the X's. Thus, R^2 represents the proportion of the total variance in Y that is explained by the linear relationship between Y and the weighted combination of X's.

**R^2 can be obtained from beta weights and zero-order correlation coefficients.**

$$R^2_{Y \bullet 12...i...p} = \sum \beta_i r_{yi} = \sum \beta_i^2 + 2\sum \beta_i \beta_j r_{ij} \qquad (i \neq j)$$

The sum of the squared beta weights represents the sum of the unique contributions of the predictors while the rightmost term represents the redundancy among the predictors.

$$R^2_{y \bullet 12} = \frac{r^2_{Y1} + r^2_{Y2} - 2r_{Y1}r_{Y2}r_{12}}{1 - r^2_{12}} = \beta_1 r_{Y1} + \beta_2 r_{Y2}$$

Note that in determining R^2 we have added together the two bivariate coefficients of determination and then corrected (reduced) that sum for the redundancy of X_1 and X_2 in predicting Y.

Squared correlation coefficients represent proportions of variance explained.

$$a + b + c + d = 1$$

$$r^2_{Y1} = b + c \qquad sr^2_1 = \frac{b}{(a + b + c + d) = 1} = b$$

$$r^2_{Y2} = d + c$$
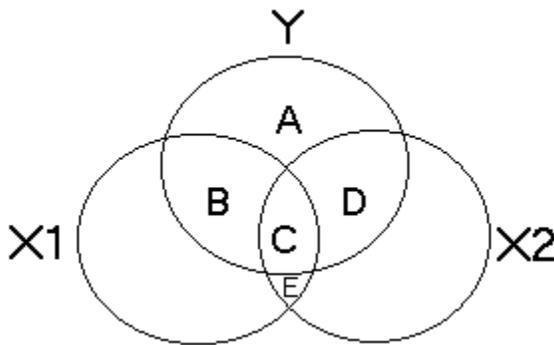
$$r^2_{12} = c + e$$

$$R^2_{Y \cdot 12} = b + c + d$$

c = redundancy (aka commonality)

A **squared semipartial correlation** represents the proportion of <u>all</u> the variance in Y that is associated with one predictor but not with any of the other predictors. That is, in terms of the Venn diagram,

$$sr^2_1 = \frac{b}{(a + b + c + d) = 1} = b.$$

The squared semipartial can also be viewed as the decrease in R^2 that results from removing a predictor from the model, that is,

$$sr^2_i = R^2_{Y \bullet 12...i...p} - R^2_{Y \bullet 12...(i)...p}$$

In terms of residuals, the **semipartial correlation** for X_i is the r between all of Y and X_i from which the effects of all other predictors have been removed. That is,

$$sr_1 = \text{corr between} Y \text{ and} (X_1 - \hat{X}_{1 \bullet 2})$$

A **squared partial correlation** represents a fully partialled proportion of the variance in Y: Of the variance in Y that is not associated with any other predictors, what proportion is associated with the variance in X_i. That is, in terms of the Venn diagram,

$$pr_1^2 = \frac{b}{a+b}$$

The squared partial can be obtained from the squared semipartial:

$$pr_i^2 = \frac{sr_i^2}{1 - R_{Y \bullet 12...(i)...p}^2} \qquad pr_i^2 > sr_i^2$$

The (i) in the subscript indicates that X_i is not included in the R^2.

In terms of residuals, the **partial correlation** for X_i is the r between Y from which all other predictors have been partialled and X_i from which all other predictors have been removed. That is,

$$pr_1 = \text{corr between}(Y - \hat{Y}_2) \text{ and}(X_1 - \hat{X}_{1 \bullet 2})$$

If the predictors are well correlated with one another, their partial and semipartial coefficients may be considerably less impressive than their zero-order coefficients. In this case it might be helpful to conduct what some call a commonality analysis. In such an analysis one can determine how much of the variance in Y is related to the predictors but not included in the predictors partial or semipartial coefficients. For the Venn diagram above, that is area c. For more details, please see my document Commonality Analysis.

A demonstration of the partial nature of multiple correlation and regression coefficients.

Run the program Partial.sas from my SAS programs page. The data are from an earlier edition of Howell (6th edition, page 496). Students at a large university completed a survey about their classes. Most of the questions had a five-point scale where "1" indicated that the course was lousy and "5" indicated that it was great. The variables in the data set are:
- Overall: the overall quality of the lectures in the class
- Teach: the teaching skills of the instructor
- Exam: the quality of the tests and exams
- Knowledge: how knowledgeable the instructor was
- Grade: the grade the student expected to receive (1 = F, …, 5 = A)
- Enroll: the number of students in the class.

The data represent a random sample from the population of classes. Each case is from one class.

Look at the data step. Here I create six Z scores, one for each of the variables in the model.

The **first invocation of Proc Reg** does a multiple regression predicting Overall from the five predictor variables. SCORR2 tells SAS I want squared semipartial correlation coefficients. PCORR2 requests squared partial correlation coefficients. TOL requests tolerances. STB tells SAS I want

Beta weights. If you look at the output, you will see that I have replicated the results reported in Howell. "Parameter Estimates" are unstandardized slopes (b), while "Standardized Estimates" are Beta weights.

The **second invocation of Proc Reg** conducts the same analysis on the standardized data (Z scores). Note that the parameter estimates here (page 2) are identical to the beta weights produced by the previous invocation of Proc Reg. That is, beta is the number of standard deviations that Y increases for every one standard deviation in X_i, partialled for the effects of all remaining predictors.

The **third invocation of Proc Reg** builds a model to predict Teach from all remaining predictors Note that Teach is pretty well correlated with the other predictors. If we subtract the R^2 here, .5818, from one, we get the **tolerance** of Teach, .4182. When the tolerance statistic gets very low, we say that we have problem with **multicollinearity**. In that case, the partial statistics would be unstable, that is, they would tend to vary wildly among samples drawn from the same population. The usual solution here is to drop variables from the model to eliminate the problem with multicollinearity. The Output statement here is used to create a new data set (Resids1) which includes all of the variables in the previous data set and one more variable, the residuals (Teach_Resid). For each observation, this residual is the difference between the actual Teach score and the Teach score that would be predicted given the observed values on the remaining predictor variables. Accordingly, **these residuals represent the part of the Teach variable that is not related to the other predictor variables**.

The **fourth invocation of Proc Reg** builds a model to predict Overall from all of the predictors except Teach (page 4). If you take the R^2 from the full model, .7554, and subtract the R^2 from this reduced model, .5722, you get .1832, the **squared semipartial** correlation coefficient for Teach, the portion of the variance in Overall that is related to Teach but not to the other predictors. The Output statement here is used to create another new data set (Resids2) with all of the earlier variables and one new one, Overall_Resid, the difference between actual Overall score and the Overall score predicted from all predictor variables except Teach. **These residuals represent the part of the Overall variable that is not related to the Exam, Knowledge, Grade, and Enroll predictors**.

The **fifth invocation of Proc Reg** builds a model to predict Overall_Resid from Teach_Resid -- that is, to relate the part of Overall that is not related to Exam, Knowledge, Grade and Enroll to the part of Teach that is not related to Exam, Knowledge, Grade and Enroll. Both Overall and Teach are adjusted to take out their overlap with Exam, Knowledge, Grade and Enroll. The squared correlation between these two residuals is .4284. This is the **squared partial** correlation between Overall and Teach. Of the variance in Overall that is not explained by the other predictors, 43% is explained by Teach. Note also that the slope here, .76324, is identical to that from the initial model. The slopes given in the output of a multiple regression analysis are partial slopes, representing the amount by which the criterion variable changes for each one point change in the predictor variable, with both criterion and predictor adjusted to remove any overlap with the other predictor variables.

Let me summarize:
- b_i is the partial slope for predicting Y from X_i – that is, the slope for predicting (Y from which we have removed the effects of all other predictors) from (X_i from which we have removed the effects of all other predictors)
- β_i is the standardized slope for predicting (Y from which we have removed the effects of all other predictors) from (X_i from which we have removed the effects of all other predictors)
- pr_i is the partial correlation between (Y from which we have removed the effects of all other predictors) and (X_i from which we have removed the effects of all other predictors)

You know that **β and *r* are the same quantity in a bivariate correlation** – the number of standard deviations that Y increases for each one standard deviation in X, which can be computed as $\beta = r = b\dfrac{s_i}{s_y}$. **Why are β and *pr* not the same quantity in a multiple regression?** It is a matter of which standard deviations are used. Look at the descriptive statistics I computed on Overall, Overall_Resid, Teach, and Teach_Resid. The unstandardized slope for Teach in the full model is .76324. We can standardize that as $\beta = b\dfrac{s_i}{s_y} = .76324\dfrac{.5321347}{.6135378} = .66197$. Note that the standard deviations are those of the original variables (Teach and Overall). If we use the standard deviations of the adjusted variables (Teach_Resid and Overall_Resid) look what we get:

$$pr = b\dfrac{s_{Adjusted\_i}}{s_{Adjusted\_y}} = .76324\dfrac{.3441182}{.4012950} = .65449.$$

Finally, Proc Corr is use to obtain the correlation between Teach_Resid and Overall, that is, the correlation between all of the variance in Overall and the variance in Teach which is not shared with the other predictors. Of course, this is a semipartial correlation coefficient. If you square it, $.42808^2$, you get the squared semipartial correlation coefficient obtained earlier, .18325.

The remaining pages of this document have the program and annotated output.

The Program

```
options pageno=min nodate formdlim='-';
title 'Multiple Regression, Data from Page 496 of Howell (6th ed.)'; run;
data Lotus;
input Overall Teach Exam Knowledge Grade Enroll @@;
Z_Overall = (Overall - 3.55) / .6135378;
Z_Teach = (Teach - 3.664) / .5321347;
Z_Exam = (Exam - 3.808) / .4931531;
Z_Knowledge = (Knowledge - 4.176) / .4078615;
Z_Grade = (Grade - 3.486) / .3510974;
Z_Enroll = (Enroll - 88) / 145.059453;
  cards;
3.4 3.8 3.8 4.5 3.5 21   2.9 2.8 3.2 3.8 3.2 50   2.6 2.2 1.9 3.9 2.8 800
3.8 3.5 3.5 4.1 3.3 221   3.0 3.2 2.8 3.5 3.2 7   2.5 2.7 3.8 4.2 3.2 108
3.9 4.1 3.8 4.5 3.6 54   4.3 4.2 4.1 4.7 4.0 99   3.8 3.7 3.6 4.1 3.0 51
3.4 3.7 3.6 4.1 3.1 47   2.8 3.3 3.5 3.9 3.0 73   2.9 3.3 3.3 3.9 3.3 25
4.1 4.1 3.6 4.0 3.2 37   2.7 3.1 3.8 4.1 3.4 83   3.9 2.9 3.8 4.5 3.7 70
4.1 4.5 4.2 4.5 3.8 16   4.2 4.3 4.1 4.5 3.8 14   3.1 3.7 4.0 4.5 3.7 12
4.1 4.2 4.3 4.7 4.2 20   3.6 4.0 4.2 4.0 3.8 18   4.3 3.7 4.0 4.5 3.3 260
4.0 4.0 4.1 4.6 3.2 100   2.1 2.9 2.7 3.7 3.1 118   3.8 4.0 4.4 4.1 3.9 35
2.7 3.3 4.4 3.6 4.3 32   4.4 4.4 4.3 4.4 2.9 25   3.1 3.4 3.6 3.3 3.2 55
3.6 3.8 4.1 3.8 3.5 28   3.9 3.7 4.2 4.2 3.3 28   2.9 3.1 3.6 3.8 3.2 27
3.7 3.8 4.4 4.0 4.1 25   2.8 3.2 3.4 3.1 3.5 50   3.3 3.5 3.2 4.4 3.6 76
3.7 3.8 3.7 4.3 3.7 28   4.2 4.4 4.3 5.0 3.3 85   2.9 3.7 4.1 4.2 3.6 75
3.9 4.0 3.7 4.5 3.5 90   3.5 3.4 4.0 4.5 3.4 94   3.8 3.2 3.6 4.7 3.0 65
4.0 3.8 4.0 4.3 3.4 100   3.1 3.7 3.7 4.0 3.7 105   4.2 4.3 4.2 4.2 3.8 70
3.0 3.4 4.2 3.8 3.7 49   4.8 4.0 4.1 4.9 3.7 64   3.0 3.1 3.2 3.7 3.3 700
4.4 4.5 4.5 4.6 4.0 27   4.4 4.8 4.3 4.3 3.6 15   3.4 3.4 3.6 3.5 3.3 40
4.0 4.2 4.0 4.4 4.1 18   3.5 3.4 3.9 4.4 3.3 90
;
* STEP 1; ********************************************************************;
Proc Reg; Model Overall = Teach -- Enroll / scorr2 pcorr2 tol stb;
Title2 'Analysis on Raw Data'; run;
* STEP 2; ********************************************************************;
Proc Reg; Model Z_Overall = Z_Teach -- Z_Enroll;
Title 'Analysis on Standardized Data'; run;
* STEP 1; ********************************************************************;
Proc Reg; Model Teach = Exam -- Enroll;
Output out = Resids1  r = Teach_Resid;
Title 'Create Residuals for Teach Predicted From All Remaining Predictors'; run;
* STEP 3; ********************************************************************;
Proc Reg; Model Overall = Exam -- Enroll;
Output out = Resids2  r = Overall_Resid;
Title 'Create Residuals for Overall Predicted From All Except Teach'; run;
* STEP 4; ********************************************************************;
Proc Reg; Model Overall_Resid = Teach_resid / stb;
Title 'Use the Part of Teach Not Related to the Other Predictors';
Title2 'To Predict the Part of Overall Not Related to those Other Predictors';
run;
proc means mean stdev; var Overall Overall_Resid Teach Teach_resid;
Title 'Descriptive Statistics on Overall, Teach, and Their Residuals'; run;
* STEP 5; ********************************************************************;
Proc Corr nosimple; Var Overall; With Teach_resid;
Title 'Correlation Between All of Overall and the Part of Teach Not Related';
Title2 'To the Other Predictors'; run;
```

Annotated Output

STEP 1: Multiple Regression, Data from Page 496 of Howell (6th ed.)   Analysis on Raw Data

The REG Procedure
Model: MODEL1
Dependent Variable: Overall

Number of Observations Read	50
Number of Observations Used	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13.93426	2.78685	27.18	<.0001
Error	44	4.51074	0.10252		
Corrected Total	49	18.44500			

Root MSE	0.32018	R-Square	0.7554
Dependent Mean	3.55000	Adj R-Sq	0.7277
Coeff Var	9.01923		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance
Intercept	1	-1.19483	0.63116	-1.89	0.0649	0	.	.	.
Teach	1	0.76324[E]	0.13292	5.74	<.0001	0.66197[F]	0.18325[B]	0.42836[C]	0.41819[A]
Exam	1	0.13198	0.16280	0.81	0.4219	0.10608	0.00365	0.01472	0.32457
Knowledge	1	0.48898	0.13654	3.58	0.0008	0.32506	0.07129	0.22570	0.67463
Grade	1	-0.18431	0.16550	-1.11	0.2715	-0.10547	0.00689	0.02742	0.61969
Enroll	1	0.00052549	0.00039008	1.35	0.1848	0.12424	0.01009	0.03961	0.65345

You might be misled by the small size of the unstandardized partial slope for Enroll, but notice that it has the third highest Beta weight. The unstandardized slope is the number of points the Overall rating increases for each one additional student enrolled in the class. One additional student is a very small change in enrollment. By contrast, the unstandardized slope for Grade is number of points the Overall rating increases for each one point change in anticipated grade in the class. A one point change in anticipated grade is an entire letter grade.

STEP 2: Analysis on Standardized Data

The REG Procedure
Model: MODEL1
Dependent Variable: Z_Overall

Number of Observations Read	50
Number of Observations Used	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	37.01700	7.40340	27.18	<.0001
Error	44	11.98299	0.27234		
Corrected Total	49	48.99999			

Root MSE	0.52186	R-Square	0.7554
Dependent Mean	2.22322E-16	Adj R-Sq	0.7277
Coeff Var	2.347327E17		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0	0.07380	0.00	1.0000
Z_Teach	1	0.66197	0.11528	5.74	<.0001
Z_Exam	1	0.10608	0.13086	0.81	0.4219
Z_Knowledge	1	0.32506	0.09077	3.58	0.0008
Z_Grade	1	-0.10547	0.09470	-1.11	0.2715
Z_Enroll	1	0.12424	0.09223	1.35	0.1848

Standardizing the variables did not affect the output excepting that the slopes are now standardized, that is, they are Beta weights.

STEP 3: Create Residuals for Teach Predicted From All Remaining Predictors

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	8.07275	2.01819	15.65	<.0001
Error	45	5.80245	0.12894		
Corrected Total	49	13.87520			

Root MSE	0.35909	R-Square	0.5818[A]
Dependent Mean	3.66400	Adj R-Sq	0.5446
Coeff Var	9.80040		

[A] The correlation between Teach and all of the other predictors is .5818. If this value were much higher, we would have a problem with multicollinearity. SAS reports the tolerance for each predictor, that is, 1 minus this R^2 between the predictor and the remaining predictors. In this case, tolerance = 1 - .5818 = **.4182**[A]. If you look back at Page 1 you will see that SAS does report a tolerance of .4182 for the Teaching variable.

STEP 4: Create Residuals for Overall Predicted From All Except Teach

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10.55416	2.63854	15.05	<.0001
Error	45	7.89084	0.17535		
Corrected Total	49	18.44500			

Root MSE	0.41875	R-Square	0.5722[B]
Dependent Mean	3.55000	Adj R-Sq	0.5342
Coeff Var	11.79579		

[B] The full model R^2 was .7554. When we removed the Teaching predictor from the model, the R^2 dropped to .5722. The sr^2 for the Teaching predictor is .7554 - .5722 = .1832[B]. If you look back at Page 1 you will see that SAS reported the sr^2 to be .1832.

STEP 5: Use the Part of Teach Not Related to the Other Predictors To Predict the Part of Overall Not Related to those Other Predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Overall_Resid Residual

Number of Observations Read	50
Number of Observations Used	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.38010	3.38010	35.97	<.0001
Error	48	4.51074	0.09397		
Corrected Total	49	7.89084			

Root MSE	0.30655	R-Square	0.4284^C
Dependent Mean	-1.0569E-15	Adj R-Sq	0.4164
Coeff Var	-2.90039E16		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	Intercept	1	-9.6881E-16	0.04335	-0.00	1.0000	0
Teach_Resid	Residual	1	0.76324^E	0.12726	6.00	<.0001	0.65449^D

^C The R^2 between the part of Teaching that is not related to the other predictors and the part of Overall that is not related to the other predictors, .4284, is the squared partial correlation for Teaching. If you look back at Page 1, you will see that SAS reported the value as .4284.

^D The unsquared partial correlation is SQRT(.4284) = .6545. Notice that this is the value which SAS reports as the standardized slope for predicting that part of Overall that is unrelated to the other predictors from that part of Teaching that is unrelated to the other predictors.

^E The unstandardized slope for predicting that part of Overall that is unrelated to the other predictors from that part of Teaching that is unrelated to the other predictors is .76324. If you look back at Page 1 you will see that is the value SAS reports for the partial slope for predicting Overall from Teaching.

Descriptive Statistics on Overall, Teach, and Their Residuals

The MEANS Procedure

Variable	Label	Mean	Std Dev
Overall		3.5500000	0.6135378
Overall_Resid	Residual	0	0.4012950
Teach		3.6640000	0.5321347
Teach_Resid	Residual	0	0.3441182

$\beta_i = \dfrac{b_i s_i}{s_y}$ For Overall and Teaching, **beta = (.76324)(.5321347/.6135378) = .66197[F]**, the beta weight for Teaching, the standardized slope for predicting that part of Overall that is unrelated to the other predictors from that part of Teaching that is unrelated to the other predictors. Notice that standardizers (standard deviations) are from the original variables, Overall and Teach.

If we standardize using the standard deviations of the residuals, we get **beta = (.76324)(3441182/.401295) = .65449[D],** which is the partial correlation coefficient for Teaching.

Correlation Between All of Overall and the Part of Teach Not Related To the Other Predictors

The CORR Procedure

1 With Variables:	Teach_Resid
1 Variables:	Overall

Pearson Correlation Coefficients, N = 50
Prob > |r| under H0: Rho=0

	Overall
Teach_Resid	**0.42808**
Residual	0.0019

This is the semi-partial correlation for Teaching, the correlation between all of Overall and that part of Teaching that is unrelated to the other predictors. If you square it, you get **.18325[B]**, the value reported for sr^2 on Page 1.

Return to Wuensch's Stats Lessons Page