Statistical Control using Partial and Semi-partial (Part) Correlations

A study of performance in graduate school produced the following correlations.

Correlations

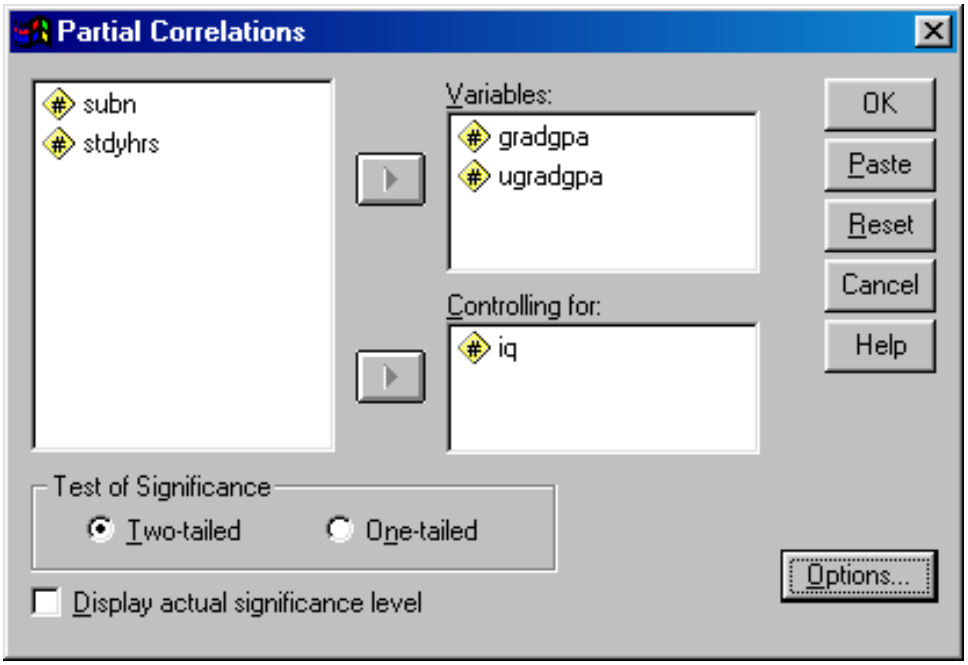
		1st year graduate gpa -- criterion variable	undergraduate gpa	full scale IQ score	number weekly study hours
1st year graduate gpa -- criterion variable	Pearson Correlation	1	.642	.612	.268
	Sig. (2-tailed)	.	.000	.000	.003
	N	122	122	122	122
undergraduate gpa	Pearson Correlation	.642	1	.464	.532
	Sig. (2-tailed)	.000	.	.000	.000
	N	122	122	122	122
full scale IQ score	Pearson Correlation	.612	.464	1	.465
	Sig. (2-tailed)	.000	.000	.	.000
	N	122	122	122	122
number weekly study hours	Pearson Correlation	.268	.532	.465	1
	Sig. (2-tailed)	.003	.000	.000	.
	N	122	122	122	122

Not surprisingly graduate grades were positively correlated with all three of these likely predictors -- undergraduate grades, IQ and number of hours studied each week. However the researcher wanted to explore these relationships further.

Partial Correlations

The first question was whether there was a relationship between graduate and undergraduate grades after controlling both for IQ. Since there is no way to randomly assign folks to and manipulate their IQ, statistical control must be applied. This partial correlation is easily obtained using SPSS.

Analyze → Correlation → Partial



Move the "core variables" into the "Variables" window and the control variable into the "Controlling for" window.

SPSS Syntax

```
PARTIAL CORR
  /VARIABLES=GRADGPA UGRADGPA BY iq.
```

← variables being correlated **BY** control(s)

We get …

Correlations

Contr			GRADGPA	UGRADGPA
IQ	GRADGPA	Correlation	1.000	.511
		Significance (2-tailed)	.	.000
		df	0	119
	UGRADGPA	Correlation	.511	1.000
		Significance (2-tailed)	.000	.
		df	119	0

There is a substantial, significant positive correlation between the two types of grades, even after controlling both for IQ.

Notice that the partial correlation is somewhat smaller than the simple correlation. This suggests that part of the simple correlation is due to each of the variables being related to IQ. When IQ is "removed" from the grades, their relationship is somewhat weaker. This is a common result when partialing.

A second question was whether there was a relationship between graduate grades and study time after controlling both for IQ. The results were…

Correlations

Contr			GRADGPA	UGRADGPA
IQ	GRADGPA	Correlation	1.000	-.024
		Significance (2-tailed)	.	.797
		df	0	119
	UGRADGPA	Correlation	-.024	1.000
		Significance (2-tailed)	.797	.
		df	119	0

While there was a significant positive simple correlation between graduate grades and study time, the correlation "disappears" when both variables are controlled for IQ.

Both grade variables are substantially correlated with IQ, so "removing" IQ from both takes away most of the apparent relationship between them.

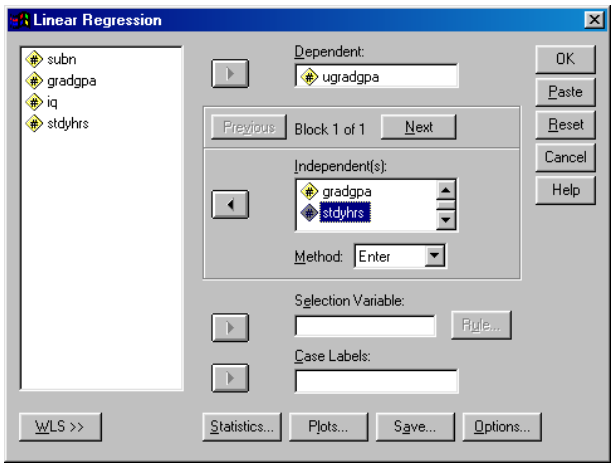
A third question involved the multiple partial correlation of the two grade variables controlling both for IQ and study time. However, after careful consideration the researcher decided that this wouldn't make sense. The study time variable referred to study time during graduate school. So, it didn't make sense to control undergraduate grades for graduate study time! Always carefully consider whether your intended partialing makes sense -- especially consider the temporal aspects of the data collection.

Semi-partial (Part) Correlations

A modification of the third question occurred to the researcher. What is the correlation between the two types of grades controlling graduate grades for graduate study time? This is a semi-partial (part) correlation question and makes sense than the partial version mentioned earlier.

We can't get a matrix of semi-partial (or multiple semi-partial) correlations from SPSS, but we can obtain these correlations via multiple regression analyses.

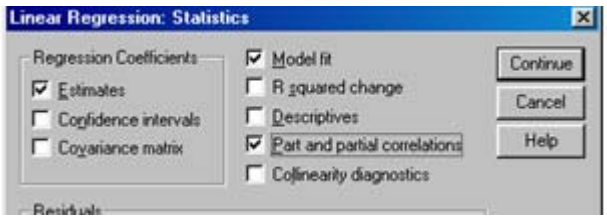
Analyze → Regression →Linear



The semi-partial we are after is …

$r_{ugradgpa(gradgpa.stdyhrs)}$

Be sure to put the control variable and the variable being controlled into the "Independent(s)" window and the other variable into the "Dependent" window.



Click "Statistics" …

and be sure that "Part and partial correlations" is checked.

SPSS Syntax

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA ZPP
  /DEPENDENT ugradgpa
  /METHOD=ENTER gradgpa stdyhrs
```

← ZPP gets the partial and semi-partial (part) correlations
← put the variable not being controlled as the criterion
← put both the correlate and the control in as predictor

Here are the pertinent parts of the resulting output…

The number we need to answer the research question is .519 -- the semi-partial (part) correlation between undergraduate gpa and graduate gpa, controlling graduate gpa for study time. The t-test of the related regression weight (B) provides a significance test for this semi-partial correlation.

So we can see that there is a substantial positive part correlation between these two types of grades -- folks with better undergraduate grades tend to also have better graduate grades, even after controlling the latter for number of study hours.

Coefficients[a]

Model		Unstandardized Coefficients	Standardized Coefficients	t	Sig.	Correlations		
		B	Beta			Zero-order	Partial	Part
1	(Constant)	-2.964		-8.269	.000			
	1st year graduate gpa -- criterion variable	.902	.538	8.455	.000	.642	.613	.519
	number weekly study hours	.388	.388	6.085	.000	.532	.487	.373

a. Dependent Variable: undergraduate gpa

Multiple Semi-partial (Part) Correlations

The last research question was whether graduate grades and study time were correlated after controlling the later for IQ and undergraduate grades. This is a multiple semi-partial (part) question. We'd put graduate grades in the "Dependent" window and study hours, IQ and undergraduate grades in the "Independent(s)" window, and get…

SPSS Syntax

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA ZPP
  /DEPENDENT ugradgpa                          ← put the variable not being controlled as the criterion
/METHOD=ENTER stdyhrs iq gradgpa               ← put both correlate and controls in as predictors
```

Coefficients[a]

Model		Unstandardized Coefficients	Standardized Coefficients			Correlations		
		B	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	3.299		92.0	.000			
	number weekly study hours	-.147	-.245	-3.32	.001	.268	-.292	-.198
	full scale IQ score	.281	.469	6.638	.000	.612	.521	.397
	undergraduate gpa	.332	.555	7.523	.000	.642	.569	.450

a. Dependent Variable: 1st year graduate gpa -- criterion variable

The number we want is -.198 -- the multiple semi-partial (part) correlation between graduate grades and study time, controlling the latter for IQ and undergraduate grades. The t-test tells us this correlation is significant.

What's interesting here is that this multiple semi-partial correlation is negative, whereas the simple correlation between graduate grades and study time is positive (r=.268). So, study time is positively correlated with graduate grades, but the part of study time that is independent of IQ and undergraduate grades is negatively correlated with graduate grades. A substantive interpretation of this might be that the graduate students with lower IQ and poorer graduate grades do more studying to maintain their graduate grades.

You should notice that the correlations involved in this last question are in the .2-.25 range -- if we had had a smaller sample (say, by 1/2) we'd have "missed" the whole issue!!

Customized Control

There are occasions when we want to control the X and the Y variables for different "other variables". While simple to do, it is important to have a solid *a priori* good reason to do this.

Probably the most meaningful situation is when Y is controlled for a "psychometric confound" and X is controlled for a "design confound".

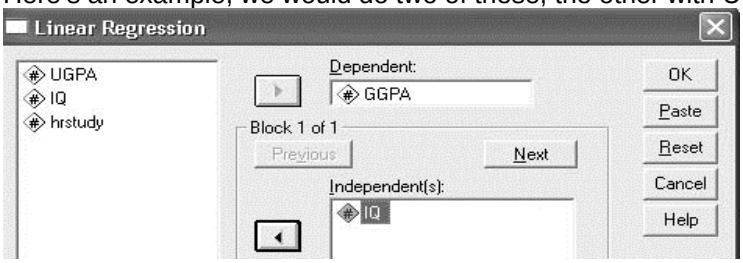
Using these variables (pretending for the moment that the # hours refers to undergraduate), imagine that we want to examine the causal relationship between the causal variable "how much is learned in undergraduate school" and the effect variable "how much is learned in graduate school". We have UGPA and GGPA, however we suspect that the bivariate correlation between the two (r = .642) doesn't tell the whole story. We can't randomly assign and manipulated UGPA. Based on this, we might want to control it for # hours of study (certainly not the only confound involved, but remember, the game is "more complex models are more accurate, on average"). Also, it is known that GGPA (perhaps more than other levels of scholastic performance) is related to intelligence. Given this, we might decide that we want to control GGPA for IQ.

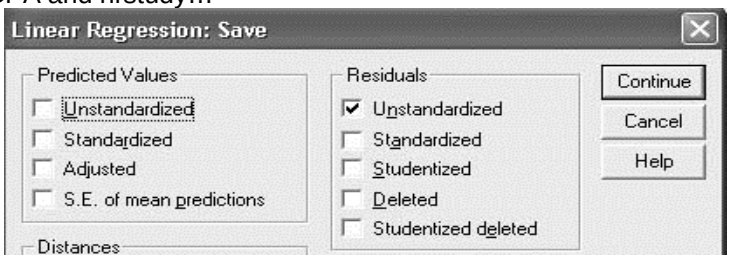
Thus, we want the correlation between "the part of GGPA that is not IQ" and "the part of UGPA that is not # study hours". Getting these residualizations is simple with SPSS. We'll just…

1. Get a model with IQ as the predictor and GGPA as the criterion → keeping each participant's residual
2. Get a model with #hrs as the predictor and UGPA as the criterion → keeping each participant's residual
3. Correlate the residuals (remember that df is really N-3, so the p-value of this correlation is bit small)

Analyze → Regression → Linear

Here's an example, we would do two of these, the other with UGPA and hrstudy…





Then we get the correlation between the two residuals…

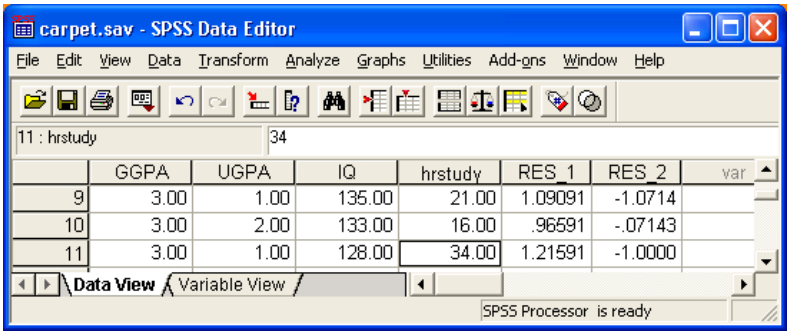
SPSS Syntax

```
REGRESSION
  /DEPENDENT gradgpa
  /METHOD=ENTER iq
  /SAVE RESID.
```

```
REGRESSION
  /DEPENDENT ugradgpa
  /METHOD=ENTER stdyhrs
  /SAVE RESID.

CORR VARIABLES = RES_1 RES_2.
```

Here's the resulting dataset and correlations…



Correlations

		Unstandardized Residual	Unstandardized Residual
Unstandardized Residual	Pearson Correlation	1	.426
	Sig. (2-tailed)	.	.000
	N	.	120
Unstandardized Residual	Pearson Correlation	.426	1
	Sig. (2-tailed)	.000	.
	N	120	.

We'd conclude the part of grad GPA that isn't IQ and the part of undergrad GPA that isn't study hours are related!