Using PowerStats to Calculate the Variance Inflation Factor

When one or more independent variables in a regression are highly correlated, the resulting "multicollinearity" can make it difficult for statistical software to implement the calculations required to complete your analysis (see Tabachnick & Fidell, 1996, for a thorough discussion). As a result, it has become somewhat common practice to recommend analysts implement collinearity diagnostics to identify any potentially problematic variables (Keith, 2006; Kennedy, 2003).

One method to characterize multicollinearity is to calculate an independent variable's *variance inflation factor*. Kennedy (2003) notes:

> "The inverse of the correlation matrix is also used in detecting multicollinearity. The diagonal elements of this matrix are called variance inflation factors, VIF_i. They are given by $(1-R_i^2)^{-1}$ where R_i^2 is the R^2 from regressing the *i*th independent variable on all the other independent variables" (p. 213)

Using commercial spreadsheet software, you can quickly invert the correlation matrix output by PowerStats and calculate variance inflation factors. *An important note: VIF, like most statistics, was conceptualized within the "simple random sample" framework, not the complex design employed by NCES studies. As a result, while this statistic may give an analyst insight in to potentially problematic collinearity in a model, this test must be interpreted with caution.*

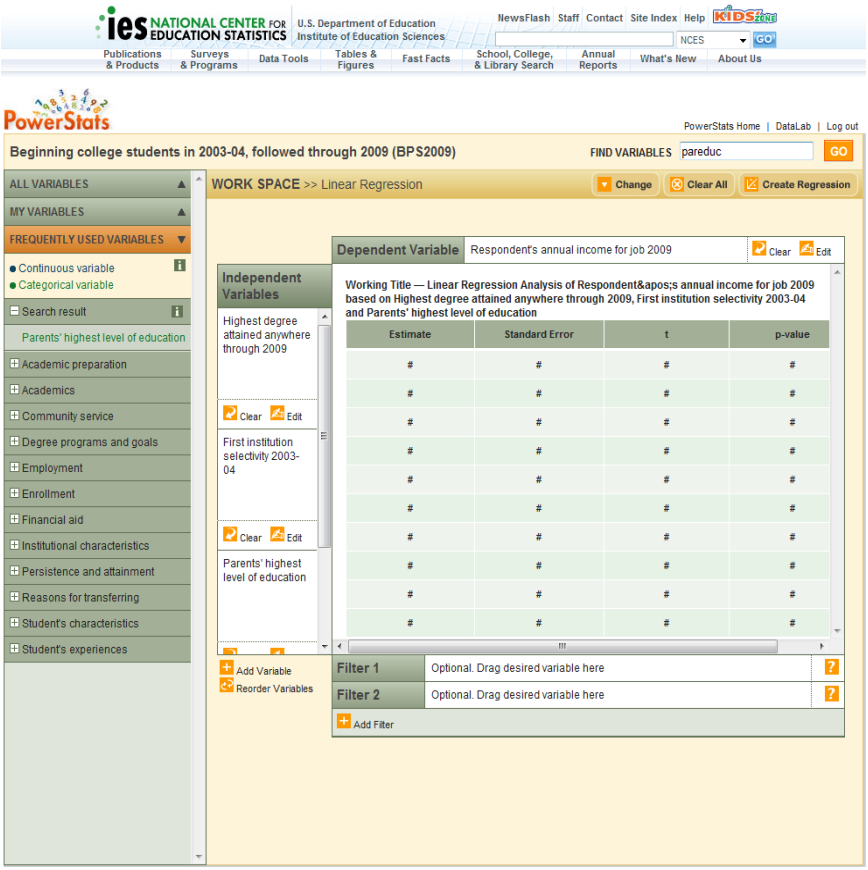
The following screenshots and accompanying instructions outline how to use Microsoft Excel to calculate the VIFs associated with independent variables used in PowerStats regressions.

References

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Pearson Education, Inc.

Kennedy, P. (2003). *A guide to econometrics.* Cambridge, MA: The MIT Press.

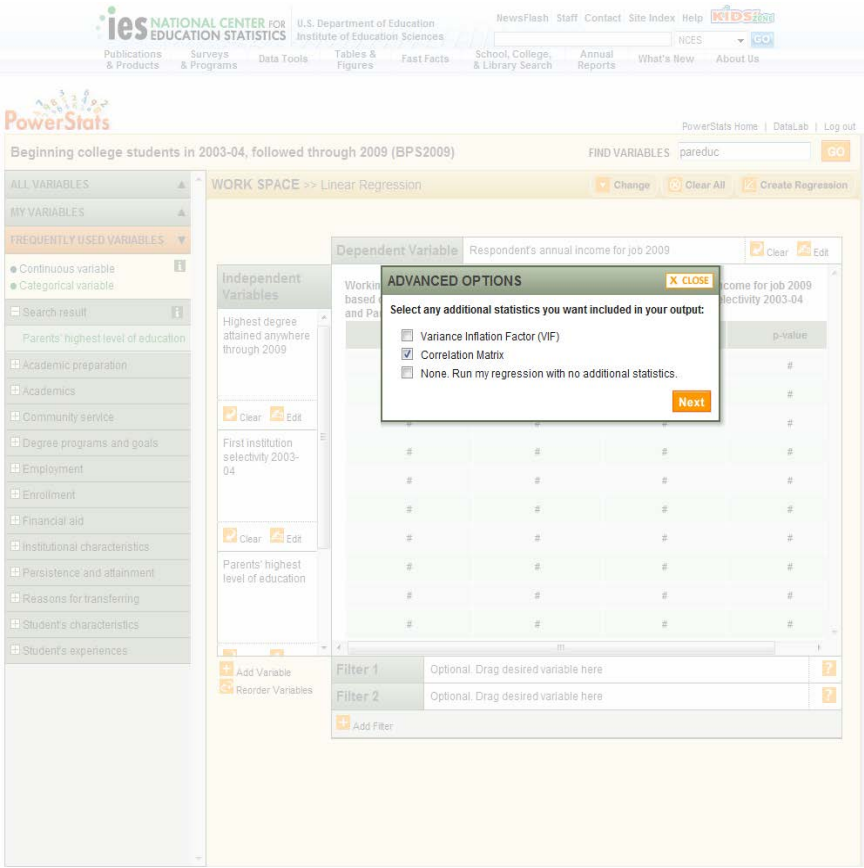
Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York, NY: HarperCollins College Publishers.



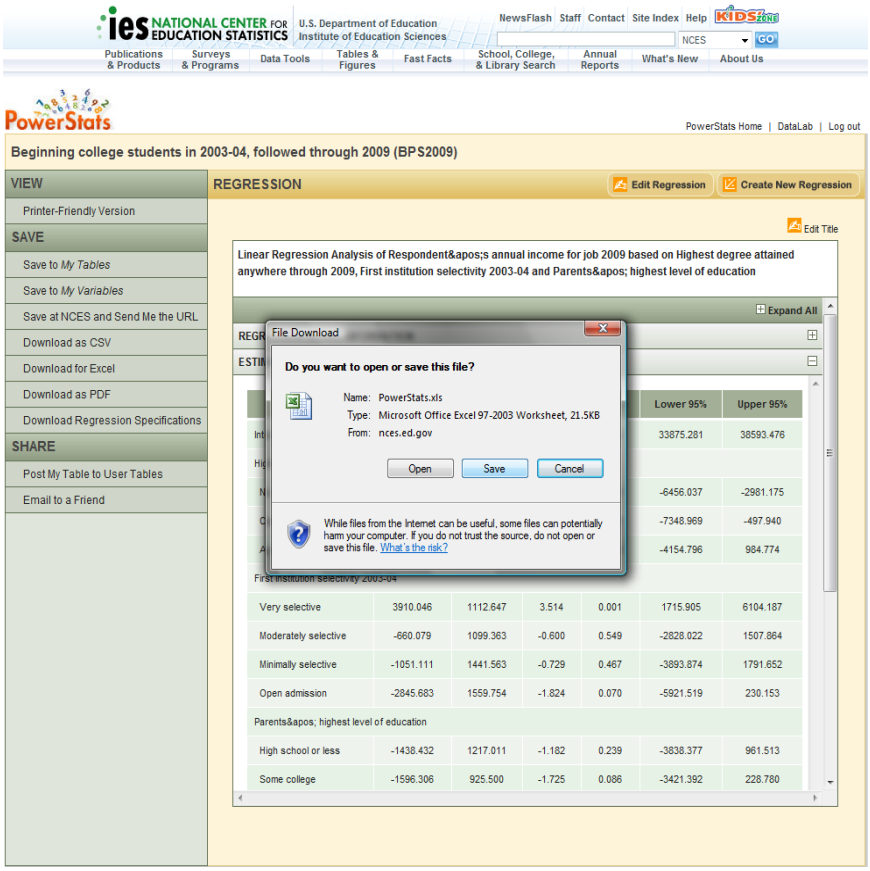
Step 1.

In the screenshot above, we have specified a linear regression model. The dependent variable ("respondent's annual income for job 2009") is displayed at the top of the workspace, and the independent variables (e.g., "highest degree attained anywhere through 2009") are displayed along the side of the workspace.

Click on *Create Regression*

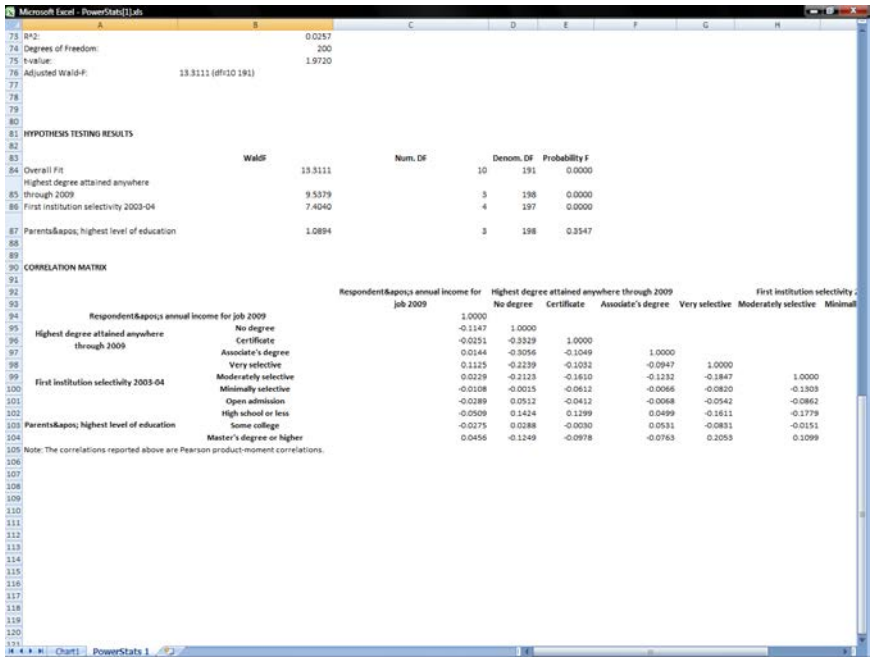


Step 2.
After clicking *Create regression*, the "Advanced Options" pane will open. Select *Correlation Matrix* and click "Next"



Step 3.

After the regression routine has completed, click on "*Download for Excel*" from the "SAVE" option pane, and save the resulting Excel file to your desktop.



Microsoft Excel - PowerStats[1].xls

R^2:		0.0257
Degrees of Freedom:		200
t-value:		1.9720
Adjusted Wald-F:	13.3111 (df=10 191)	

HYPOTHESIS TESTING RESULTS

	WaldF	Num. DF	Denom. DF	Probability F
Overall Fit	13.3111	10	191	0.0000
Highest degree attained anywhere through 2009	9.5379	3	198	0.0000
First institution selectivity 2003-04	7.4040	4	197	0.0000
Parents' highest level of education	1.0894	3	198	0.3547

CORRELATION MATRIX

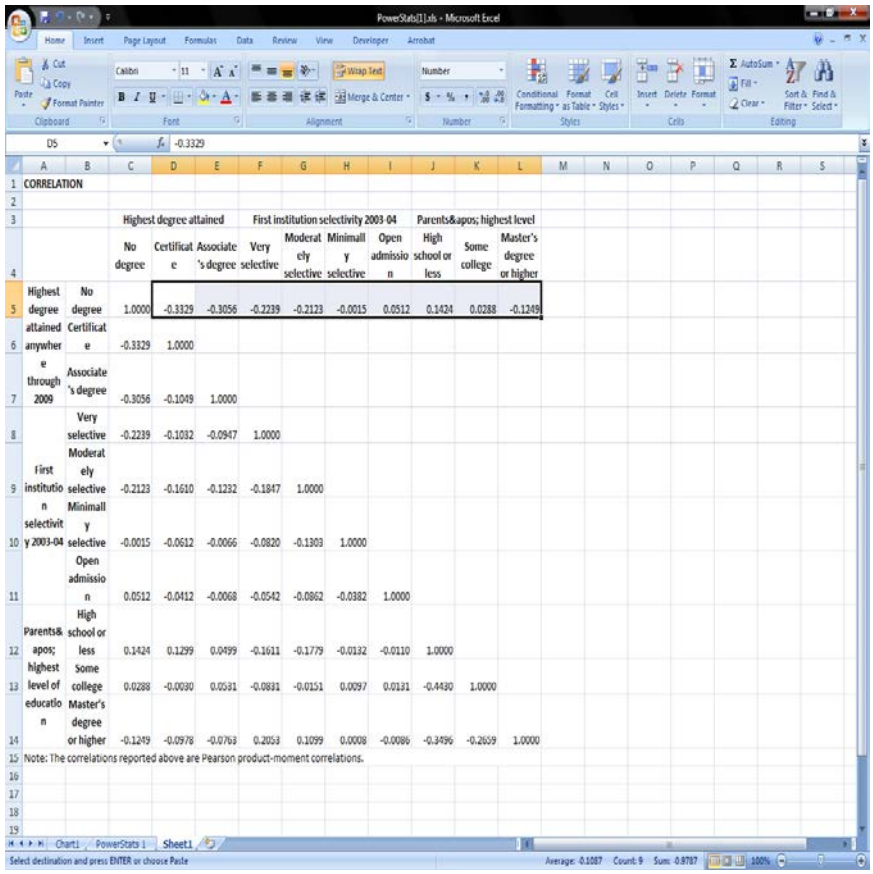
		Respondent's annual income for job 2009	Highest degree attained anywhere through 2009				First institution selectivity	
			No degree	Certificate	Associate's degree	Very selective	Moderately selective	Minimall
Respondent's annual income for job 2009		1.0000						
Highest degree attained anywhere through 2009	No degree	-0.1147	1.0000					
	Certificate	-0.0251	-0.3329	1.0000				
	Associate's degree	0.0144	-0.3056	-0.1049	1.0000			
	Very selective	0.1125	-0.2239	-0.1032	-0.0947	1.0000		
First institution selectivity 2003-04	Moderately selective	0.0229	-0.2123	-0.1610	-0.1232	-0.1847	1.0000	
	Minimally selective	-0.0108	-0.0015	-0.0612	-0.0066	-0.0820	-0.1303	
	Open admission	-0.0289	0.0512	-0.0412	-0.0068	-0.0542	-0.0862	
	High school or less	-0.0509	0.1424	0.1299	0.0499	-0.1611	-0.1779	
Parents' highest level of education	Some college	-0.0275	0.0288	-0.0030	0.0531	-0.0831	-0.0151	
	Master's degree or higher	0.0456	-0.1249	-0.0978	-0.0763	0.2053	0.1099	

Note: The correlations reported above are Pearson product-moment correlations.

Chart1 PowerStats 1

Step 4.

Open the Excel file, and copy the Correlation Matrix located at the bottom of the output. Paste it to a new sheet inside your Excel workbook.



CORRELATION

		Highest degree attained			First institution selectivity 2003-04				Parents' highest level		
		No degree	Certificate	Associate's degree	Very selective	Moderately selective	Minimally selective	Open admission	High school or less	Some college	Master's degree or higher
Highest degree attained anywhere through 2009	No degree	1.0000	-0.3329	-0.3056	-0.2239	-0.2123	-0.0015	0.0512	0.1424	0.0288	-0.1249
	Certificate	-0.3329	1.0000								
	Associate's degree	-0.3056	-0.1049	1.0000							
First institution selectivity 2003-04	Very selective	-0.2239	-0.1032	-0.0947	1.0000						
	Moderately selective	-0.2123	-0.1610	-0.1232	-0.1847	1.0000					
	Minimally selective	-0.0015	-0.0612	-0.0066	-0.0820	-0.1303	1.0000				
	Open admission	0.0512	-0.0412	-0.0068	-0.0542	-0.0862	-0.0382	1.0000			
Parents' highest level of education	High school or less	0.1424	0.1299	0.0499	-0.1611	-0.1779	-0.0132	-0.0110	1.0000		
	Some college	0.0288	-0.0030	0.0581	-0.0831	-0.0151	0.0097	0.0131	-0.4430	1.0000	
	Master's degree or higher	-0.1249	-0.0978	-0.0763	0.2053	0.1099	0.0008	-0.0086	-0.3496	-0.2659	1.0000

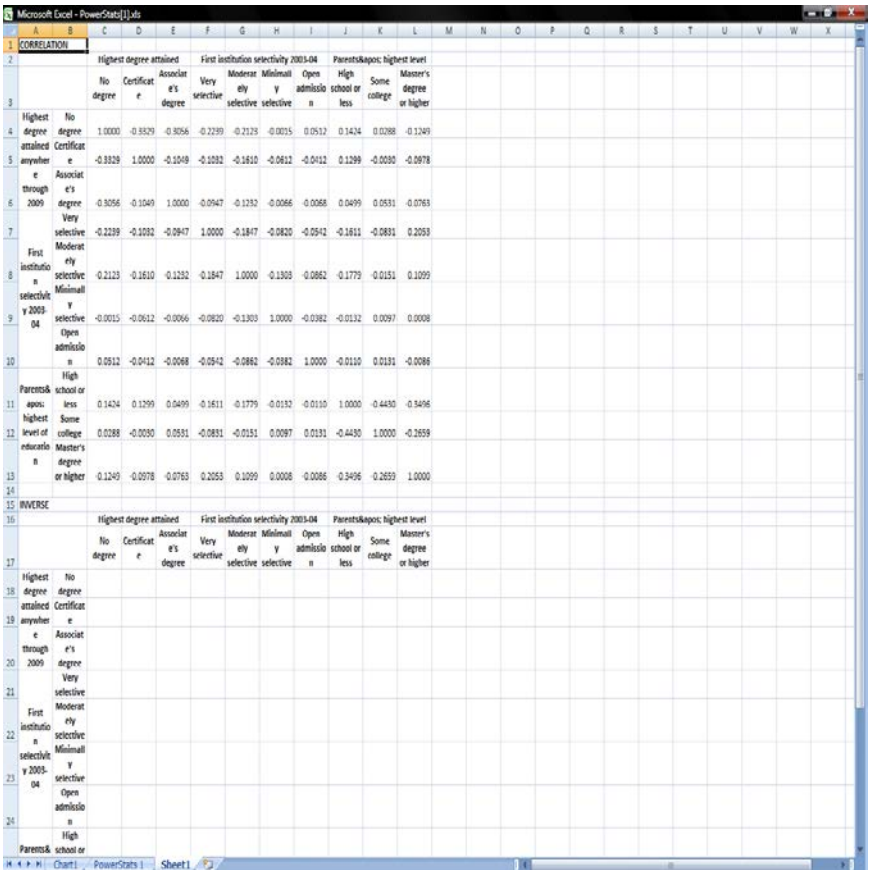
Note: The correlations reported above are Pearson product-moment correlations.

Step 5.

After removing the correlations associated with your *dependent variable* (i.e., delete the first row and column of correlations in your table), convert the *lower-half* matrix to a *symmetric* matrix. You will transpose the columns of variables in the lower half of the matrix to the rows in the upper half. (Note that the first <u>column</u> begins with the following values: 1.0, -0.33, -0.31, and so forth. Your goal is the make those same values also appear in the first <u>row</u>, and so forth.)

To do so:

(1) Select all values in the first <u>column</u> below the "1.0". Copy them by holding down the CTRL key and hitting the C key after the values are selected.
(2) Right-click in the first empty cell after the "1.0" in the first <u>row</u>.
(3) Select the "Paste Special" option, and make sure there is a check next to "Transpose." Click on "OK."
(4) Continue through all columns until all rows are filled.



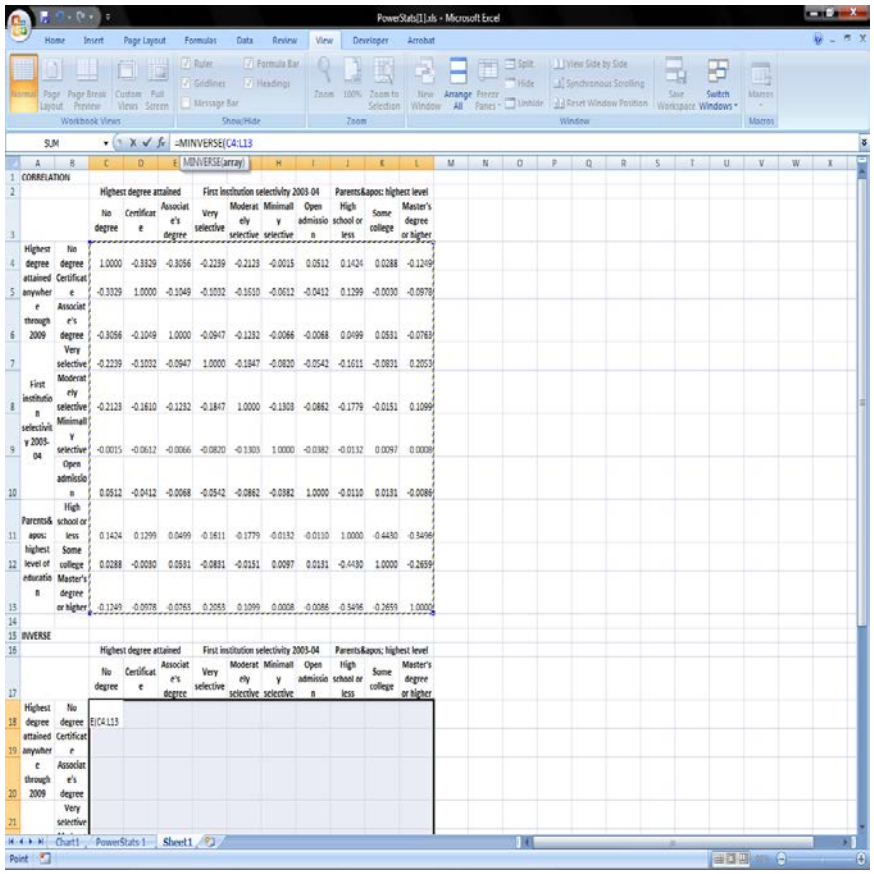
Microsoft Excel - PowerStats[1].xls

			C	D	E	F	G	H	I	J	K	L
1	CORRELATION											
2			Highest degree attained			First institution selectivity 2003-04				Parents' highest level		
3			No degree	Certificate	Associate's degree	Very selective	Moderately selective	Minimally selective	Open admission	High school or less	Some college	Master's degree or higher
4	Highest degree attained anywhere through 2009	No degree	1.0000	-0.3329	-0.3056	-0.2239	-0.2123	-0.0015	0.0512	0.1424	0.0288	-0.1249
5		Certificate	-0.3329	1.0000	-0.1049	-0.1032	-0.1610	-0.0612	-0.0412	0.1299	-0.0030	-0.0978
6		Associate's degree	-0.3056	-0.1049	1.0000	-0.0947	-0.1232	-0.0066	-0.0068	0.0499	0.0531	-0.0763
7	First institution selectivity 2003-04	Very selective	-0.2239	-0.1032	-0.0947	1.0000	-0.1847	-0.0820	-0.0542	-0.1611	-0.0831	0.2053
8		Moderately selective	-0.2123	-0.1610	-0.1232	-0.1847	1.0000	-0.1303	-0.0862	-0.1779	-0.0151	0.1099
9		Minimally selective	-0.0015	-0.0612	-0.0066	-0.0820	-0.1303	1.0000	-0.0382	-0.0132	0.0097	0.0008
10		Open admission	0.0512	-0.0412	-0.0068	-0.0542	-0.0862	-0.0382	1.0000	-0.0110	0.0131	-0.0086
11	Parents' highest level of education	High school or less	0.1424	0.1299	0.0499	-0.1611	-0.1779	-0.0132	-0.0110	1.0000	-0.4430	-0.5496
12		Some college	0.0288	-0.0030	0.0531	-0.0831	-0.0151	0.0097	0.0131	-0.4430	1.0000	-0.2659
13		Master's degree or higher	0.1249	-0.0978	-0.0763	0.2053	0.1099	0.0008	-0.0086	-0.5496	-0.2659	1.0000
14												
15	INVERSE											
16			Highest degree attained			First institution selectivity 2003-04				Parents' highest level		
17			No degree	Certificate	Associate's degree	Very selective	Moderately selective	Minimally selective	Open admission	High school or less	Some college	Master's degree or higher
18	Highest degree attained anywhere through 2009	No degree										
19		Certificate										
20		Associate's degree										
21	First institution selectivity 2003-04	Very selective										
22		Moderately selective										
23		Minimally selective										
24		Open admission										
	Parents'	High school or										

Chart1 | PowerStats1 | Sheet1

Step 6.

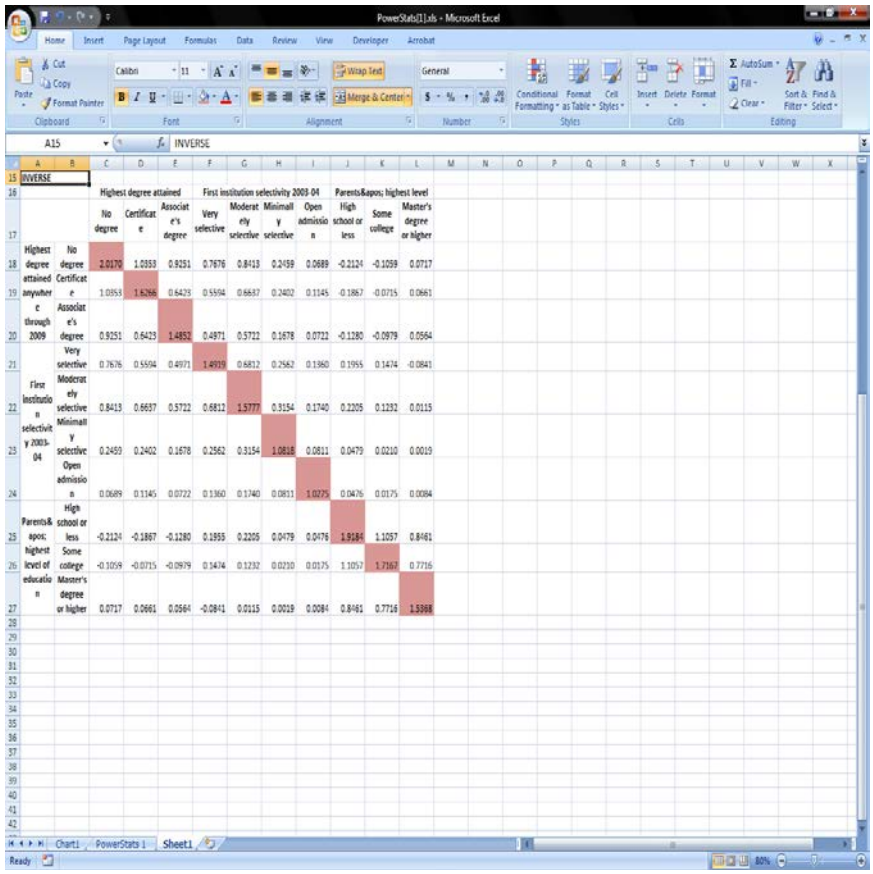
Leaving your now-symmetric correlation matrix intact, copy the entire matrix to empty space beneath it. Then, delete all the correlations in the pasted matrix so you have a full correlation matrix on top and an empty duplicate of it on the bottom. (In the picture above, the matrix extends below the bottom of the screen.)



Step 7.
Highlight all empty cells in the bottom matrix by clicking your mouse in the first empty cell and, while holding the mouse down, dragging down to its bottom right-hand corner.

In the formula bar at the top of the screen, enter the formula =MINVERSE( ). The "array" inside the parentheses that follow MINVERSE should correspond to the cells in the upper matrix that hold your correlations. Either enter them manually by entering the cell location of the upper left-hand corner of the correlation matrix, followed by a colon ':', followed by the cell location of the lower right hand corner of the matrix: =MINVERSE(C5:G9). Alternately, you can automatically enter the cells by clicking in the first cell in the upper left-hand corner of the correlation matrix and dragging down to the last cell in the lower right-hand corner.

Once the cells are highlighted (or manually entered), hit CTRL-SHIFT-ENTER.



		Highest degree attained			First institution selectivity 2003-04				Parents' highest level		
		No degree	Certificate	Associate's degree	Very selective	Moderately selective	Minimally selective	Open admission	High school or less	Some college	Master's degree or higher
Highest degree attained anywhere through 2009	No degree	2.0170	1.0353	0.9251	0.7676	0.8413	0.2459	0.0689	-0.2124	-0.1059	0.0717
	Certificate	1.0353	1.6266	0.6423	0.5594	0.6637	0.2402	0.1145	-0.1867	-0.0715	0.0661
	Associate's degree	0.9251	0.6423	1.4852	0.4971	0.5722	0.1678	0.0722	-0.1280	-0.0979	0.0564
First institution selectivity 2003-04	Very selective	0.7676	0.5594	0.4971	1.4979	0.6812	0.2562	0.1360	0.1955	0.1474	-0.0841
	Moderately selective	0.8413	0.6637	0.5722	0.6812	1.5777	0.3154	0.1740	0.2205	0.1232	0.0115
	Minimally selective	0.2459	0.2402	0.1678	0.2562	0.3154	1.0818	0.0811	0.0479	0.0210	0.0019
	Open admission	0.0689	0.1145	0.0722	0.1360	0.1740	0.0811	1.0275	0.0476	0.0175	0.0084
Parents' highest level of education	High school or less	-0.2124	-0.1867	-0.1280	0.1955	0.2205	0.0479	0.0476	1.9184	1.1057	0.8461
	Some college	-0.1059	-0.0715	-0.0979	0.1474	0.1232	0.0210	0.0175	1.1057	1.7167	0.7716
	Master's degree or higher	0.0717	0.0661	0.0564	-0.0841	0.0115	0.0019	0.0084	0.8461	0.7716	1.5368

Step 8.

The lower matrix will now be automatically populated with the inverse of the correlation matrix output by PowerStats. The values on the diagonal (highlighted above) are the VIFs associated with each predictor variable.

Kennedy (2003) and others have suggested a VIF threshold of 10 to indicate significant multicollinearity. As a reminder, analysts should take all VIFs calculated via PowerStats as advisory only: VIF was developed for use on simple random samples, not the complex sampling designs employed by NCES.