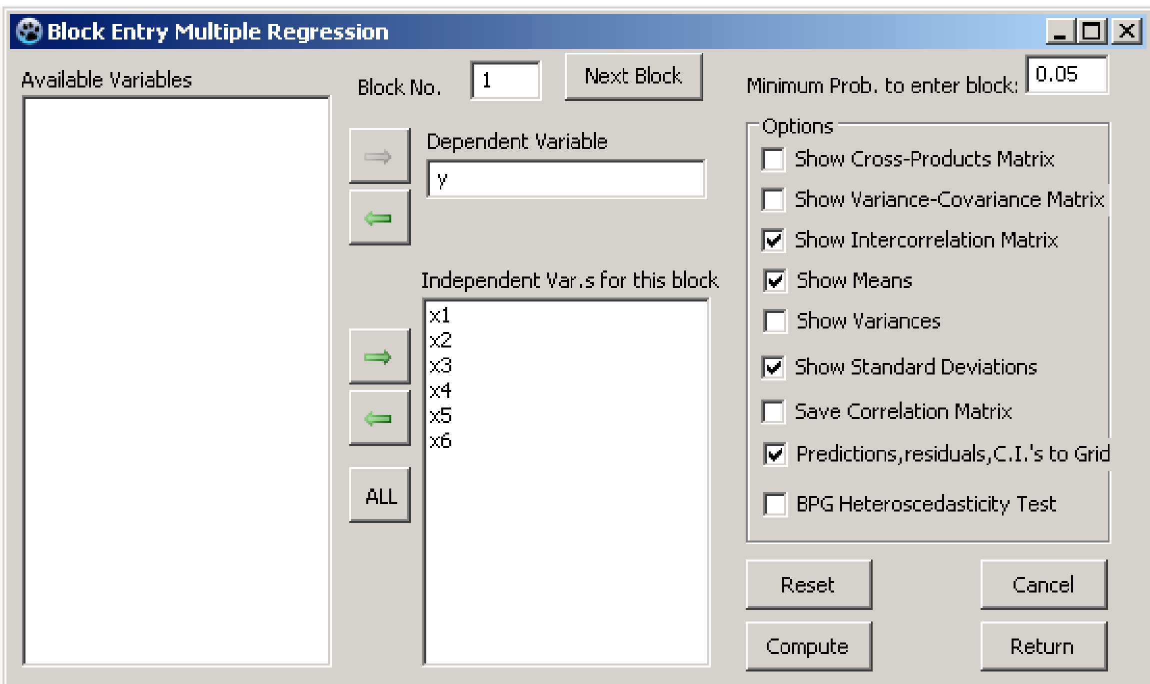
BLOCK ENTRY MULTIPLE REGRESSION

When conducting research to explore the relationship between a variable of interest (dependent variable) and other variables, it is often the case that previous research suggests some of these other (independent variables) may have a higher relationship to the dependent variable than others do. In addition, the cost of gathering the additional variables for the purpose of predicting future samples may vary. For these reasons, it is useful to be able to select sub-sets of independent variables to enter a regression equation. In the Block Entry method, the researcher can enter a block of one or more variables, obtain the results and then enter additional blocks of variables. The procedure can be limited by tests for the significance of variable entry as well as elimination should the contribution of a variable be lessened by entry of other variables.

We will demonstrate the block entry using the Longley.LAZ file. This particular set of data was created by the NIST government agency as a test of a computer program to handle somewhat “strange” data. If you load this file you will note that several variables are extremely large. This can cause “overflow” problems with some programs. In addition, several variables are very highly correlated – much more than would typically be expected if random sampling was done and the variables were truly independent. Many procedures will report that the inverse of the matrix of correlations among the variables is zero. Here then is the dialog and the results of our analysis. We have only entered a single block of variables (all of the independent variables.)




```
Block Entry Multiple Regression by Bill Miller

----------------- Trial Block 1 Variables Added ------------------

Product-Moment Correlations Matrix with   16 cases.


Variables
```

	x1	x2	x3	x4	x5
x1	1.000	0.992	0.621	0.465	0.979
x2	0.992	1.000	0.604	0.446	0.991
x3	0.621	0.604	1.000	-0.177	0.687
x4	0.465	0.446	-0.177	1.000	0.364
x5	0.979	0.991	0.687	0.364	1.000
x6	0.991	0.995	0.668	0.417	0.994

y	0.971	0.984	0.502	0.457	0.960

Variables

	x6	y
x1	0.991	0.971
x2	0.995	0.984
x3	0.668	0.502
x4	0.417	0.457
x5	0.994	0.960
x6	1.000	0.971
y	0.971	1.000

Means with 16 valid cases.

Variables	x1	x2	x3	x4	x5
	101.681	387698.438	3193.313	2606.688	117424.000

Variables	x6	y
	1954.500	65317.000

Standard Deviations with 16 valid cases.

Variables	x1	x2	x3	x4	x5
	10.792	99394.938	934.464	695.920	6956.102

Variables	x6	y
	4.761	3511.968

SOURCE	DF	SS	MS	F	Prob.>F
Regression	6	184172401.944	30695400.324	330.285	0.000
Residual	9	836424.056	92936.006		
Total	15	185008826.000			

Dependent Variable: y

R	R2	F	Prob.>F	DF1	DF2
0.998	0.995	330.285	0.000	6	9

Adjusted R Squared = 0.992

Std. Error of Estimate = 304.854

Variable	Beta	B	Std.Error	t	Prob.>t	VIF	TOL
x1	0.046	15.062	84.915	0.177	0.863	135.532	0.007
x2	-1.014	-0.036	0.033	-1.070	0.313	1788.513	0.001
x3	-0.538	-2.020	0.488	-4.136	0.003	33.619	0.030
x4	-0.205	-1.033	0.214	-4.822	0.001	3.589	0.279
x5	-0.101	-0.051	0.226	-0.226	0.826	399.151	0.003
x6	2.480	1829.151	455.478	4.016	0.003	758.981	0.001

Constant = -3482258.635
Increase in R Squared = 0.995
F = 330.285 with probability = 0.000
Block 1 met entry requirements

SOURCE	DF	SS	MS	F	Prob.>F
Regression	6	184172401.944	30695400.324	330.285	0.000
Residual	9	836424.056	92936.006		
Total	15	185008826.000			

Dependent Variable: y

R	R2	F	Prob.>F	DF1	DF2
0.998	0.995	330.285	0.000	6	9

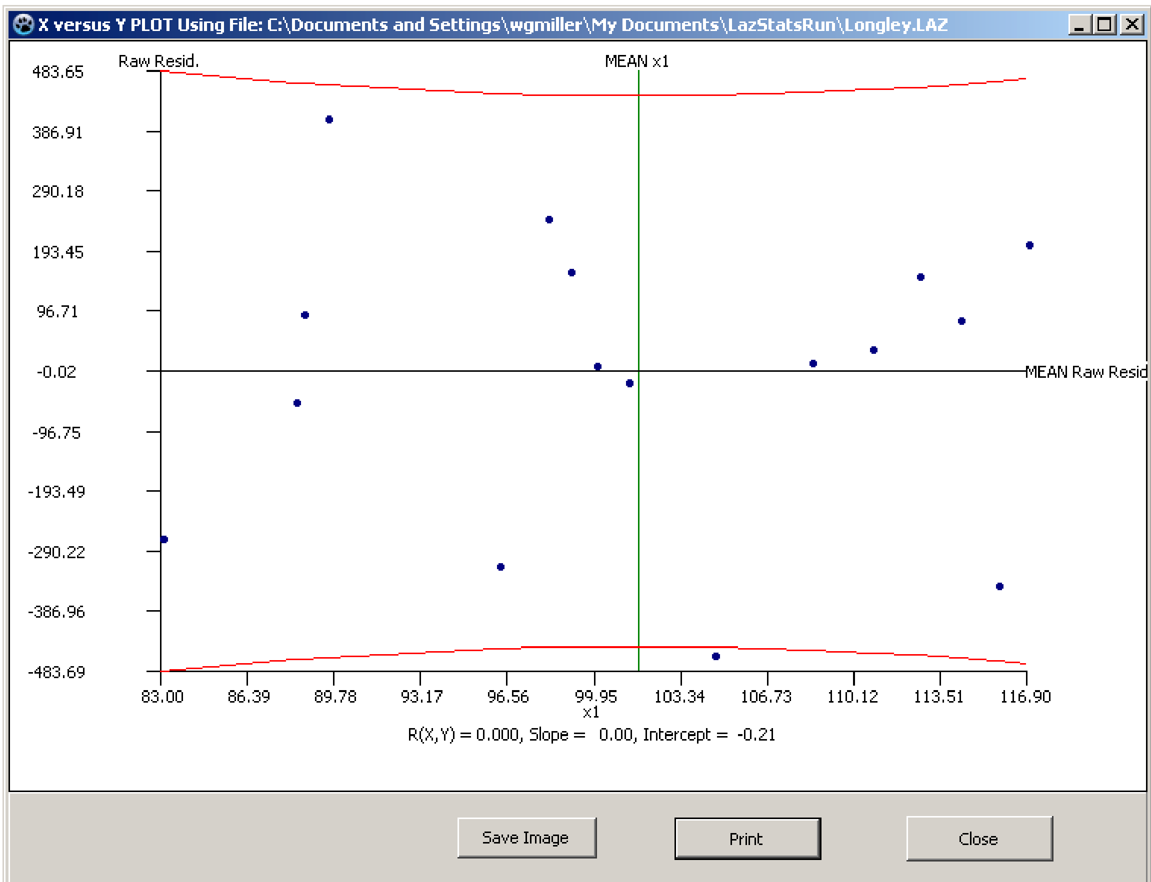
Adjusted R Squared = 0.992

Std. Error of Estimate = 304.854

Variable	Beta	B	Std.Error	t	Prob.>t	VIF	TOL
x1	0.046	15.062	84.915	0.177	0.863	135.532	0.007
x2	-1.014	-0.036	0.033	-1.070	0.313	1788.513	0.001
x3	-0.538	-2.020	0.488	-4.136	0.003	33.619	0.030
x4	-0.205	-1.033	0.214	-4.822	0.001	3.589	0.279
x5	-0.101	-0.051	0.226	-0.226	0.826	399.151	0.003
x6	2.480	1829.151	455.478	4.016	0.003	758.981	0.001

Constant = -3482258.635

It should be noted that the results above agree with those published by NIST. Since we also requested the residuals from prediction, you can also plot the residuals as found in the article "An Appraisal of Least Squares Programs for the Electronic Computer", Journal of the American Statistical Association, 62, pp. 819-841, 1967 by J.W. Longley. For example:



Note: Residuals can be obtained as original minus the predicted or predicted minus the original. The plot above is, in fact, a vertical reflection of that in the article mentioned above.