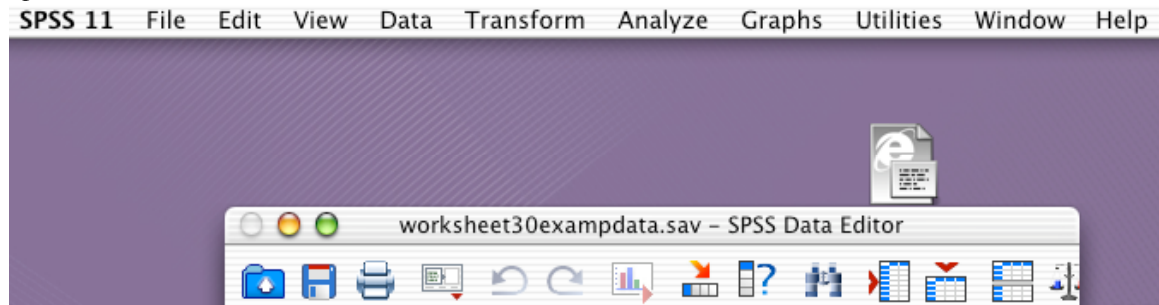


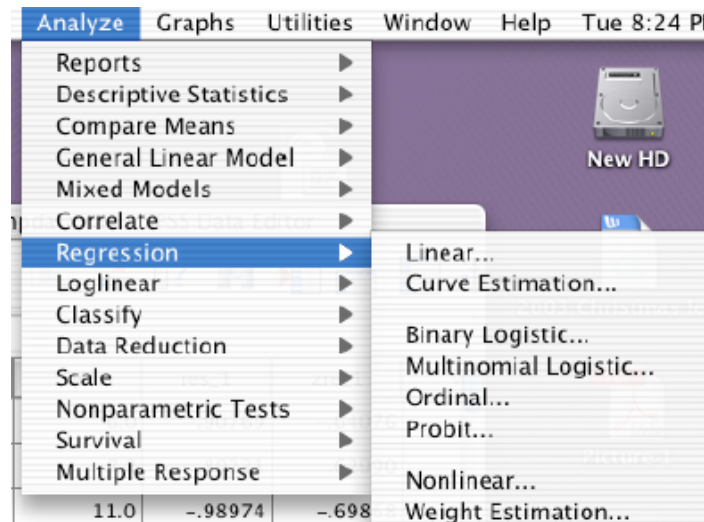
Doing Multiple Regression with SPSS

Multiple Regression for Data Already in Data Editor

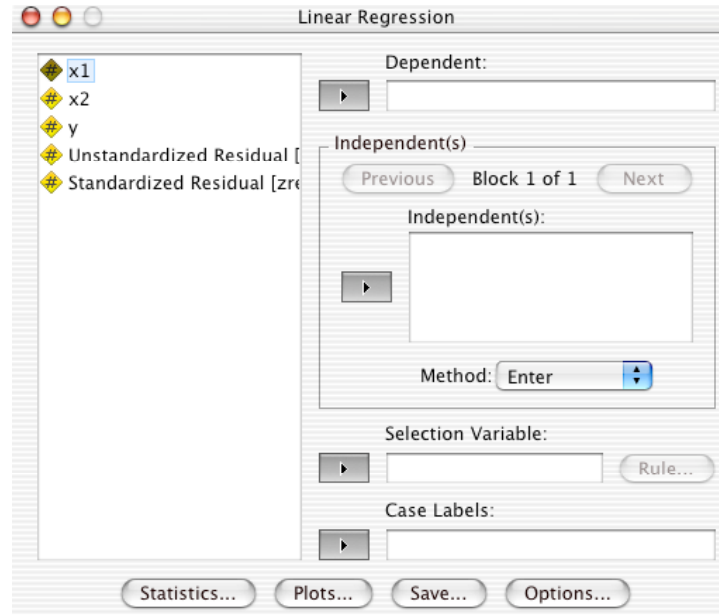
Next we want to specify a multiple regression analysis for these data. The menu bar for SPSS offers several options:



In this case, we are interested in the “Analyze” options so we choose that menu. It gives us a number of choices:



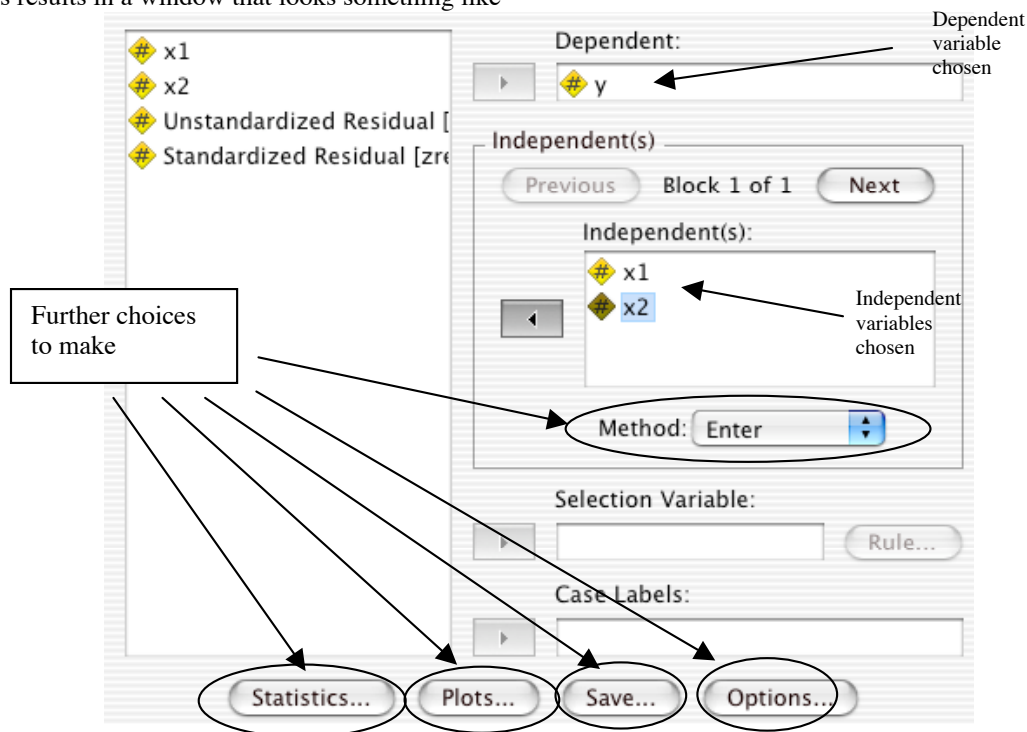
In this case we are interested in Regression and choosing that opens a sub-menu for the type of regression, which for us is “Linear” since that is all that we have studied, whether it be SLR or MR. This leads to the following window:



We use this window to choose the dependent variable and all of the independent variables. I choose x1 and x2 as independent variables and y as the dependent variable. The other two variables were actually generated in a previous analysis of this data and we will discuss them further.

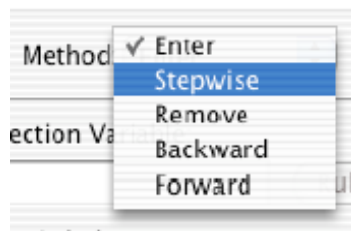
You enter a dependent variable by highlighting (selecting) it in the list on the left and then clicking on the arrow under “Dependent Variable” which will move it into that box.

You enter independent variables by highlighting (selecting) each (or all by using a shift-click) and clicking on the arrow for Independent Variables, which will move them into that box. SPSS also leaves a copy of the independent variables chosen in the box on the left so if you see them there that is not a mistake. This results in a window that looks something like



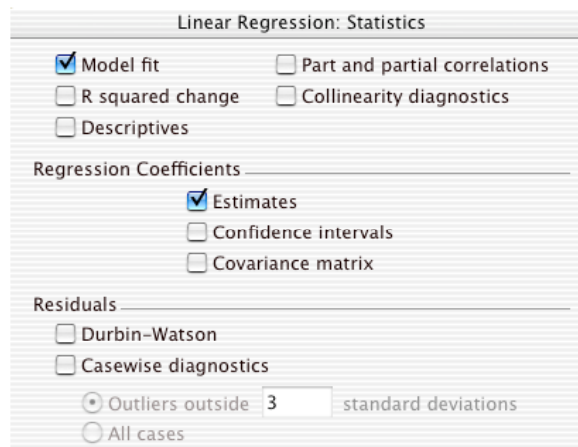
In addition to having the dependent and independent variables where you want them, there are several further choices to make as seen in the five buttons to which I have drawn arrows in the picture above. We will discuss each in turn.

The first choice is in the “Method” to use as indicated by the “Method” button. Pressing it gives something like the following:



We will discuss the other options like “Stepwise” later but for now we choose “Enter” which means “Enter all of the independent variables in the model at the same time. We click on “Enter”, the list goes away and we have “Method: Enter” showing in the window.

Next we click on the “Statistics” button and get something like:



Linear Regression: Statistics

☒ Model fit ☐ Part and partial correlations
☐ R squared change ☐ Collinearity diagnostics
☐ Descriptives

Regression Coefficients

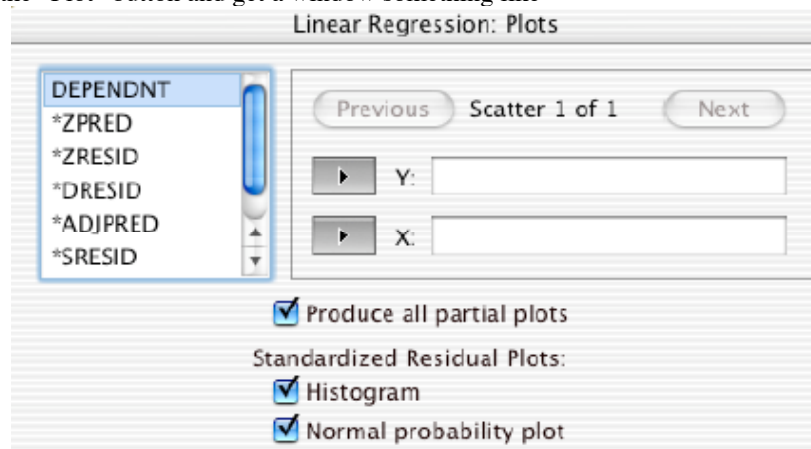
☒ Estimates
☐ Confidence intervals
☐ Covariance matrix

Residuals

☐ Durbin-Watson
☐ Casewise diagnostics
☒ Outliers outside 3 standard deviations
☐ All cases

Obviously there a lot of choices. For now we choose only “Model fit” which tells SPSS to fit an MR model using the specified independent and dependent variables. We also choose “Estimates” under “Regression Coefficients” since that will tell SPSS to give us the $\hat{\beta}_i$ ’s. We don’t specify anything under “Residuals” right now since we will do the simple things we need to do in later steps with SPSS. However, once you have the hang of doing an MR with SPSS, you may want to explore some of these other options. After having made our choices we press continue.

We next press the “Plot” button and get a window something like



Linear Regression: Plots

DEPENDNT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID

Previous Scatter 1 of 1 Next

Y:

X:

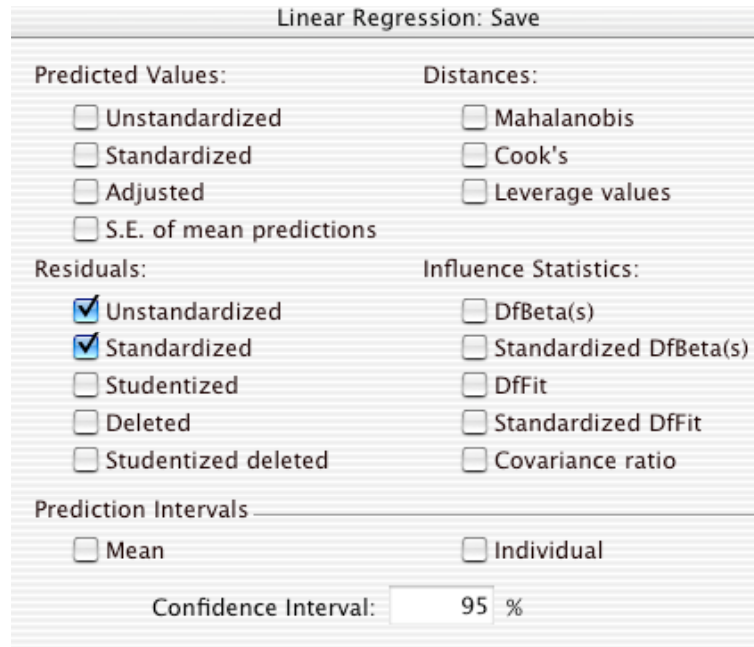
☒ Produce all partial plots

Standardized Residual Plots:

☒ Histogram
☒ Normal probability plot

For now, I choose “Produce all partial plots”, and for the “Standardized Residual Plots” I choose “Histogram” and “Normal probability plot”. Above in the set of windows labeled X and Y you can choose variables from the list at left to produce as many scatter plots as you wish. We will return to this later. Having made our choices we indicate OK and continue to the Save button.

The Save button is probably not quite what you think it might be. It opens a window something like the following:

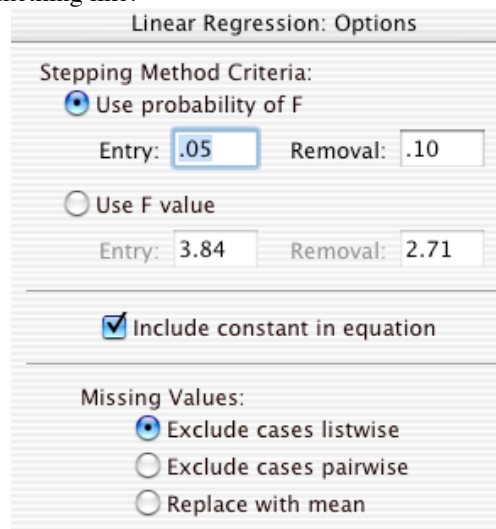


The 'Linear Regression: Save' dialog box contains several sections for saving regression results:

- Predicted Values:**
 - ☐ Unstandardized
 - ☐ Standardized
 - ☐ Adjusted
 - ☐ S.E. of mean predictions
- Distances:**
 - ☐ Mahalanobis
 - ☐ Cook's
 - ☐ Leverage values
- Residuals:**
 - ☒ Unstandardized
 - ☒ Standardized
 - ☐ Studentized
 - ☐ Deleted
 - ☐ Studentized deleted
- Influence Statistics:**
 - ☐ DfBeta(s)
 - ☐ Standardized DfBeta(s)
 - ☐ DfFit
 - ☐ Standardized DfFit
 - ☐ Covariance ratio
- Prediction Intervals:**
 - ☐ Mean
 - ☐ Individual
- Confidence Interval:** 95 %

Clearly it represents a large variety of possibilities. Mostly these are variables, statistics and calculations that you would want SPSS to make for each case and save for each case. For now, we will restrict ourselves to two simple ones: unstandardized and standardized residuals. This is where the two variables in the list I showed in the window earlier came from. I had specified that they be created and saved on a previous SPSS MR run. This window gives you all of the possibilities for things that SPSS can create and save for each observation.

Having made my “Save” choices, I indicate that we should continue and turn to the “Options” button, which produces a window something like:



The 'Linear Regression: Options' dialog box contains the following sections:

- Stepping Method Criteria:**
 - ☒ Use probability of F
 - Entry: .05
 - Removal: .10
 - ☐ Use F value
 - Entry: 3.84
 - Removal: 2.71
- ☒ Include constant in equation
- Missing Values:**
 - ☒ Exclude cases listwise
 - ☐ Exclude cases pairwise
 - ☐ Replace with mean

We are not using “Stepwise” as our Method yet so we ignore the “Stepping Method Criteria” section. We will discuss this later when we consider stepwise regression. Hopefully you won’t be dealing with missing values either but if you do, you should go with the default option of “Exclude cases listwise.” The only important panel on this page is to choose the default “Include constant in equation.” This means that the MR model fitted will include a β_0 . In advanced courses one studies options in which you might not want a β_0 , for example, situations in which you know y should equal 0 when all of the x -variables equal 0. For now we won’t worry about it.

Carrying Out the MR Analysis and Looking at the Results

At this point, you indicate “Done” or “OK” or “Continue” for the whole window specifying the MR analysis to be run. You will have to wait from 5 to 500 seconds depending on how powerful the PC is that you are using to do your regression. Nothing appears to be happening. Do not be discouraged. Don’t panic. OK, panic if you want to, but don’t do anything to the computer. When the analysis is done an output window will appear that looks something like this:

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	X2, X1 ^a	.	Enter

a. All requested variables entered.
b. Dependent Variable: Y

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980 ^a	.960	.948	1.4166

a. Predictors: (Constant), X2, X1
b. Dependent Variable: Y

This is the output window and it will continue to be updated with any further analyses you make until the window is closed. The left-hand column lists the parts of each analysis (see oval) and will list any later analyses that are made. It is a sort of “Table of Contents” of outputs to allow you to go to any of them quickly. At this point it might be well use “Save As” to save you output. ***Please note that you save the Data (in the Data Editor window and the output in the Output window separately. Saving one will not guarantee that the other is saved.*** I would use Save As on both once and then continue to use Save periodically as you make changes in either the data or the analysis output.

Now that we have the output, let’s consider each major part in turn.

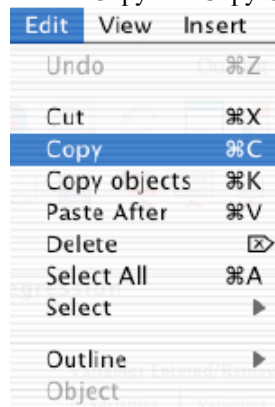
When you click on a piece of the output, it is boxed and a red arrow indicates that this is what you are looking at, as for example,

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980 ^a	.960	.948	1.4166

a. Predictors: (Constant), X2, X1
b. Dependent Variable: Y

This is the second piece of the output and gives you the value of the coefficient of determination, “R Square” and of $s = \sqrt{MSE}$, the “standard error of the estimate. If you wish to copy and past this into a Word or other document, you have two choices: Copy and Copy Object. That is:



Use Copy Object if you want just to past a (usually faint) picture of the output into Word. Use Copy if you want to past in output that can be manipulated. For example, Copy on this selected output produced the following when I pasted it into this document., This is in the table form of Word and can be manipulated like any other table.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980	.960	.948	1.4166
a. Predictors: (Constant), X2, X1				
b. Dependent Variable: Y				

The next output is the ANOVA table for the MR model and this to can be copied an pasted into a Word or PowerPoint document. Here I have just pasted in a picture of the output.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	332.933	2	166.466	82.954	.000 ^a
	Residual	14.047	7	2.007		
	Total	346.980	9			
a. Predictors: (Constant), X2, X1						
b. Dependent Variable: Y						

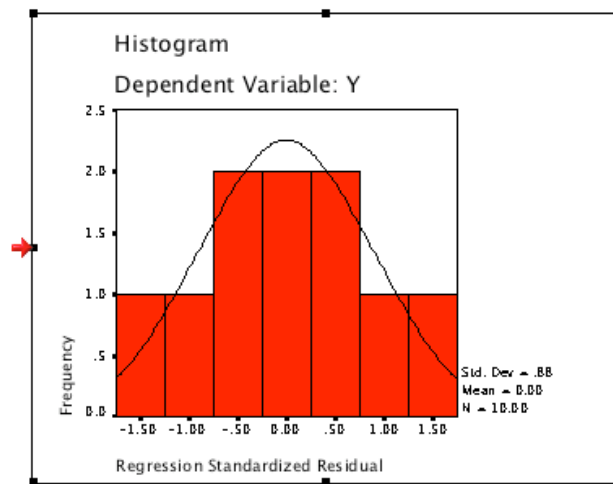
Notice that this comes complete with the Sum Squares column, the df column, the MS column, the F column and the p -value or observed significance of the F. From the df column, we see $k = 2$, the two predictor variables that we have. We see that $n - 1 = 9$ so $n = 10$, our sample size. We see that the MSE is 2.007 so this means that $s^2 = 2.007$. We see that $F = 82.954$ which looks as if it would be significant. Under the "Sig." Column for $F = 82.954$, we see .000. This is enough to tell us that the p -value or significance of the F is $p < .001$. Since this is the smallest value at which we can reject the hypothesis that the model doesn't help, we can reject at .05, .01, or even .001, in fact, anything above .00044 (which rounds to .000). This piece of the output is discussed in detail in Worksheets 30 and 31.

The next part of the output is the coefficients section:

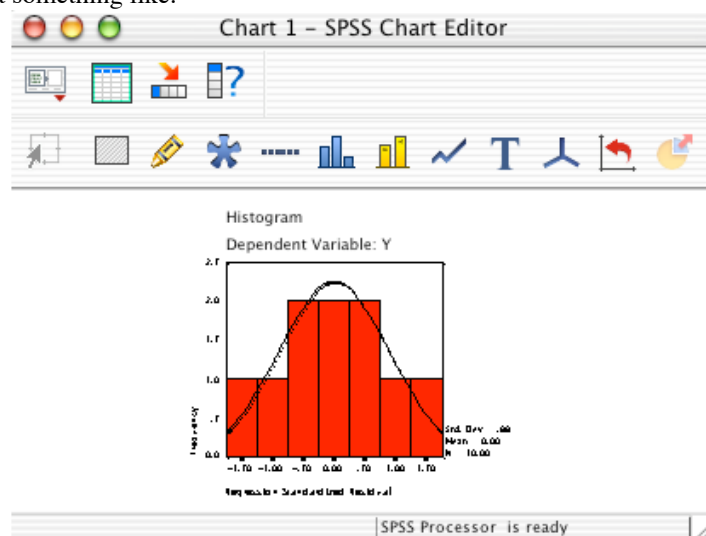
Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	*****	1.426		.989
	X1	3.236	2.158	.563	.177
	X2	1.846	1.644	.422	.298
a. Dependent Variable: Y					

This contains the $\hat{\beta}_i$'s. The row of asterisks for the constant, $\hat{\beta}_0$, turns into a number when you cut and past a larger version of this table. As before this can be copied and pasted as a formattable table into Word by using the Copy command rather than the Copy Object command. This piece of the output is discussed in considerable detail in Worksheet 30.

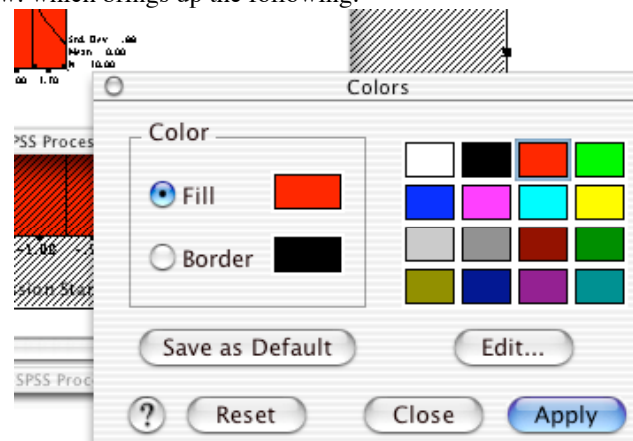
One of the next useful pieces of the output is the histogram of the residuals:



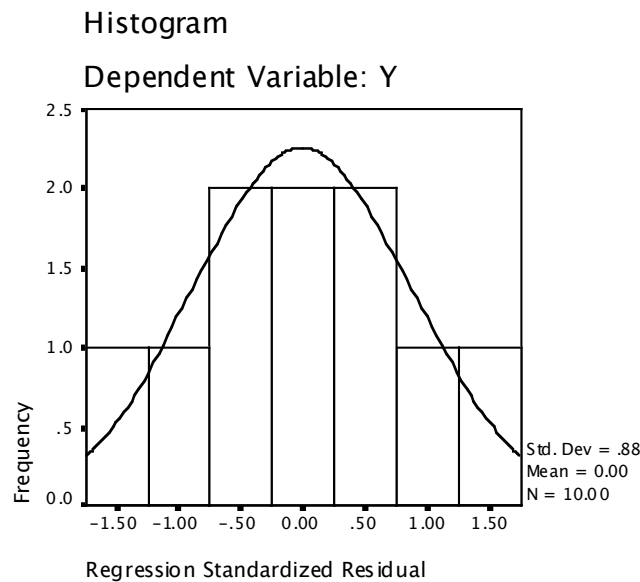
If this is not quite the way you like it, you can go to the Chart Editor by double clicking on the chart. If you do that here, you get something like:



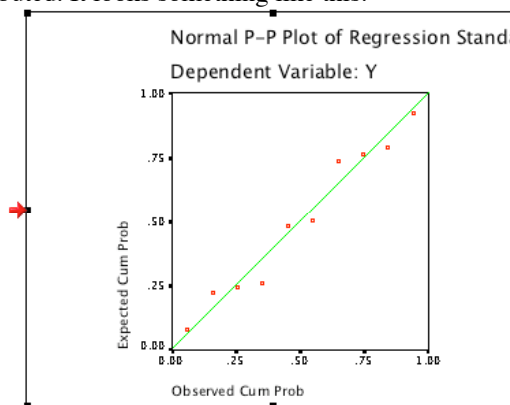
This has lots of options for changing the charts. I will make only one change here but will show others later. I don't like the red color in the bars of the histogram since I don't think it will photocopy well (look at the version in Worksheet 31), so I would like to change it to white. I click on the color icon (a crayon) in the Chart Editor window, which brings up the following:



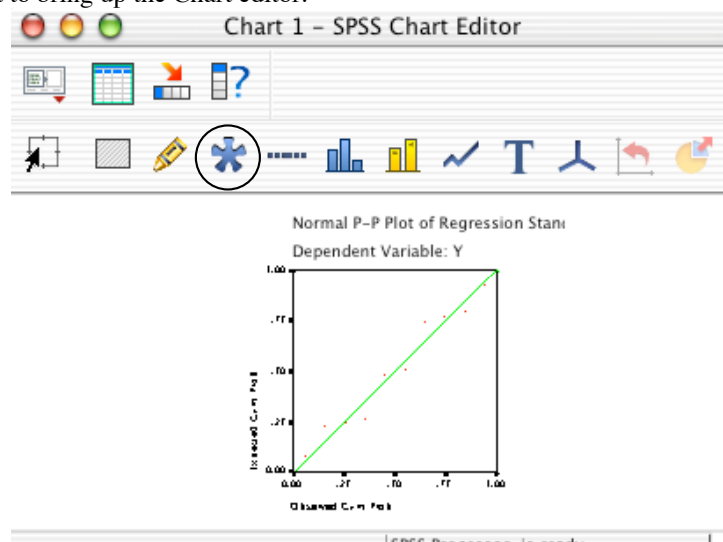
On this photocopy you may not be able to see the different colors in the upper right corner but of these the upper left is white. I choose it as the Fill color while leaving the Border color black. I click the Apply and then close this window and the chart editor window. The result is the picture that I have copied and pasted below:



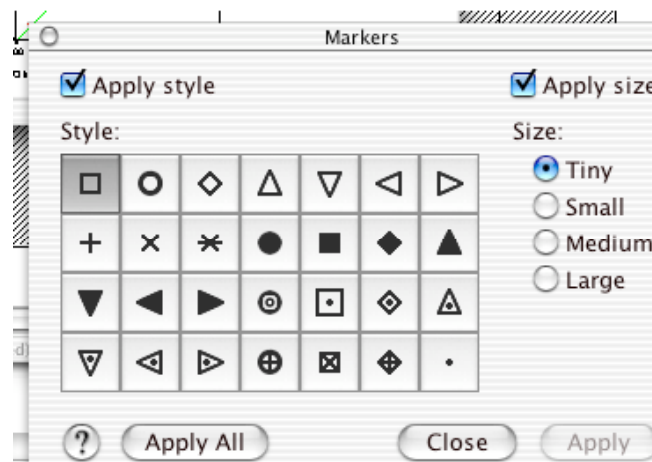
The next interesting piece of the output is the P-P plot (see Worksheet 31) to check on whether the residuals are normally distributed. It looks something like this:



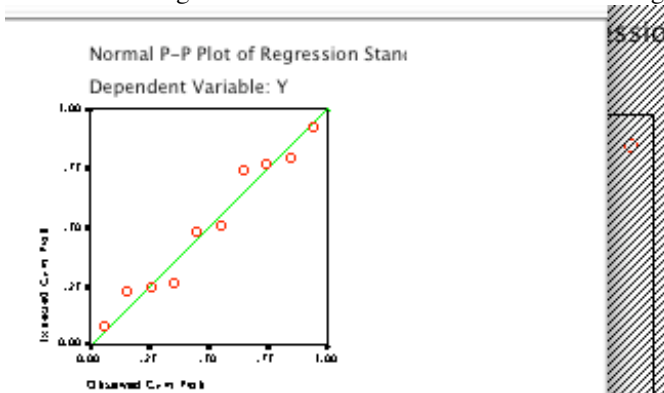
There are several things I don't like about this plot. You may not be able to tell but the dots are small red squares, the line is in green and the title, "Normal P-P Plot of...", runs off the picture on the right, I double click on this output to bring up the Chart editor:



The first thing I would like to do is change the shape and size of the dots. I click on the "asterisk" looking symbol (circled) above the graph and it opens the window:

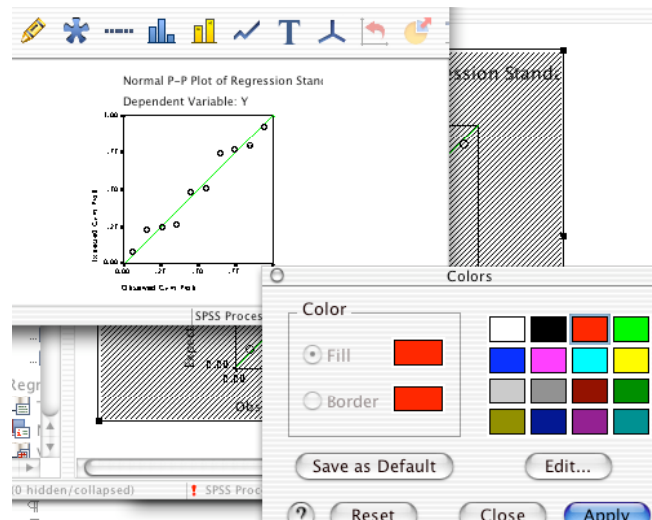


I choose the small circle by clicking on it and I choose the size “Small” instead of “Tiny”. I then click on Apply All, wait a moment for the change to be made and then close this window. I get something like:



The change to small circles is reflected both on the picture in the Chart Editor window and on the underlying main picture below it (shadowed).

I still don't like the color of the dots so while they are highlighted I click on the crayon to change color. This gives me



I click on the black and then click Apply. Although this changes both the fill and border to black, these symbols are unfilled circles so I get the black circles I wish. I still don't like the title being truncated so I double click on it and get:

Titles

Title 1:

Title 2:

Title Justification:

Subtitle:

Subtitle Justification:

I select “Normal P-P Plot” and get rid of the rest of the title. I also select “Center” instead of “Left” for the title justification. This gives me:

Titles

Title 1:

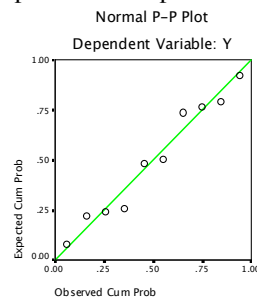
Title 2:

Title Justification:

Subtitle:

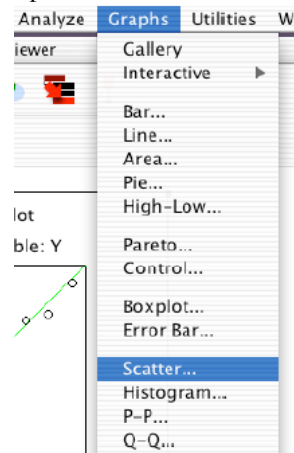
Subtitle Justification:

and when I close it I get the following graph to cut and paste into this document:

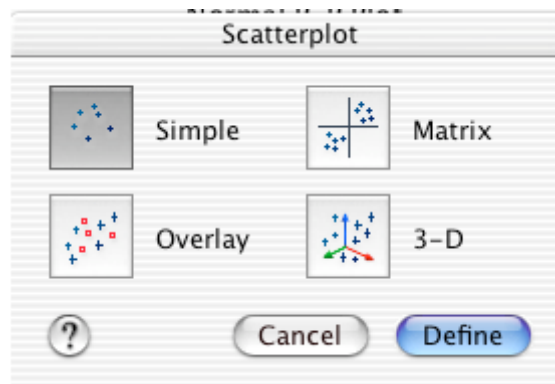


This has taken us through the main output an SPSS MR run. However, there are still some scatter plots we would like to have to check assumptions (see Worksheet 31).

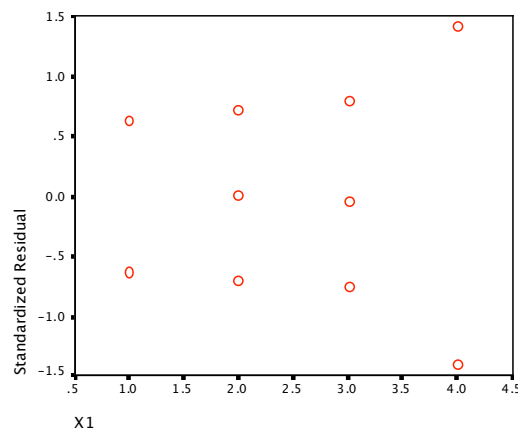
We can get any of them by choosing “Graph” in SPSS:



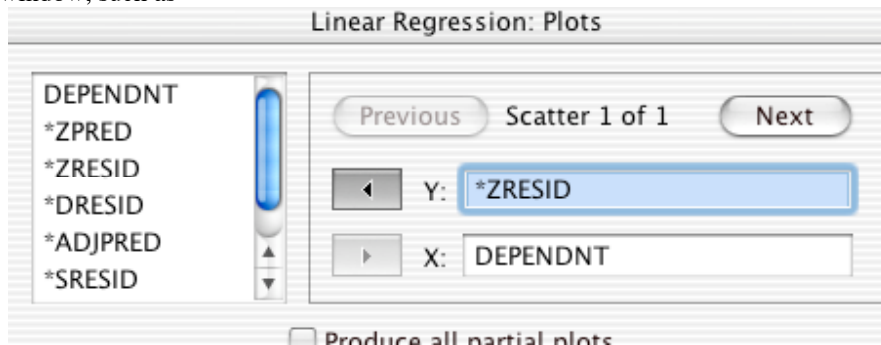
This gives us several options for graphs and we choose “Scatter...” which gives us



We choose “Simple” to make a simple scatterplot. By the way this is where you choose “3-D” to make three-dimensional scatterplots that you can spin, etc. An example of one is seen in Worksheets 30 and 31. After we have changed the symbols, etc., to suit our taste we get a scatterplot like the follow (a version of which has been seen in Worksheet 31):

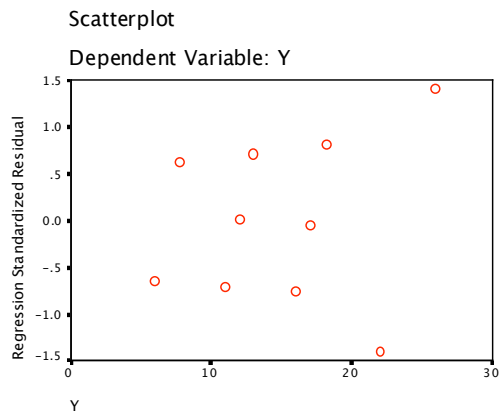


We can also specify some plots as a part of running the MR in the first place by using the Plot option on the Regression window, such as



I selected ZRESID for the y-axis of the scatterplot and the dependent variable (DEPENDNT) for the x-axis. By using the Next button I can go on to specify several scatterplots to be calculated as part of the regression run.

The choice above gives the following plot which I have cut and pasted after changing the symbols to what I wanted:



So you can make residual plots by using the scatterplot option under graph after having made the MR run in SPSS or you can specify some as a part of the run. You should choose any variables you are going to want such as “Unstandardized Residuals” as part of Save in specifying the MR run. Then SPSS will create and save these variables so that you have them later to make scatterplots, P-P plots (an option under graph), etc. This is all the basics about SPSS you will need for the Final and for the group projects.