SPSS Chapter 2 Example 1 – Scatterplot

The math department of a large university wants to plan the number of sections and instructors necessary for its elementary courses. The head of the department hopes that the number of students that will enrol in these courses can be predicted from the number of first year students. The data for several years is presented below in the SPSS Data Editor window.

🖬 *Untitled1 [DataSet0] - SPSS Data Editor									
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	nalyze <u>G</u> raph	ns <u>U</u> tilities <u>V</u>	⊻indow <u>H</u> elp				
🖻 📙	🗁 🖶 📇 🐨 🐟 🚈 🕼 🛤 📲 🏥 🖽 🐺 🕸 🔕								
10 : enro	10 : enroll								
	firstyr	enroll	var	var	var	Var	var	~	
1	4595.00	7364.00							
2	4827.00	7547.00							
3	4427.00	7099.00						=	
4	4258.00	6894.00							
5	3995.00	6572.00							
6	4330.00	7156.00							
7	4265.00	7232.00							
8	4351.00	7450.00							
9									
10									
11									
12									
13									
14									
15									
16 ∎ ● \ Da	nta View λ ∀a	ariable View <i>J</i>	7		<			>	
	SPSS Processor is ready								

The explanatory variable (x) is the number of first year students (**firstyr**). The response variable (y) is the number of first year students enrolled in elementary math courses (**enrol**).

We will now use SPSS to create a scatterplot of the data.

The following are the steps needed to create a scatterplot.

- 1. Click <u>G</u>raphs on the main menu bar.
- 2. Click '**Chart Builder...**' in the <u>G</u>raphs menu, which will open a Chart Builder dialog box identical to that below.

🗖 Chart Builder	
Variables:	
<pre> firstyr [firstyr]</pre>	Drag a Gallery chart here to use it as your starting point OR Click on the Basic Elements tab to build a chart element by element
Categories: No variables selected	
(Chart preview uses example data
Gallery	Favorites
Optional Elements	Bar Line Area Pie/Polar Scatter/Dot
Elemen <u>t</u> Properties	
Options	
	OK <u>P</u> aste <u>R</u> eset Cancel Help

- 3. Select the '**Scatter/Dot**' option from the "Choose from" sub-dialog box, and then double-click on the '**Simple Scatter**' icon.
- 4. Step 3 will open a template of a simple scatterplot in the "Chart Preview" subdialog box.
- 5. Since the number of first year students is the independent variable, it should be placed on the X-axis of the scatterplot. To define the x-axis of 'firstyr,' click 'firstyr' in the box containing the variable list and drag it onto the 'X-axis' label in the "Chart Preview" sub-dialog box.
- 6. To define the Y-axis as the number of first year students enrolled in elementary math courses, select '**enroll**' in the box containing the variable list. Then drag it onto the '**Y-axis**' label in the "Chart Builder" sib-dialog box. The immediate visual effect of steps 5 and 6 is that 'firstyr' and 'enroll' have been entered into the X-axis box and Y-axis box respectively. The Chart Builder dialogue box should now resemble the one below.



7. Click the **OK** command pushbutton located at the bottom left corner of the dialogue box. The result of step 6 is the opening of a new SPSS Viewer window. The window consists of two panes: and outline pane and the results pane. The outline pane is located on the left and looks like a table of contents. The results pane is located on the right and contains the actual output (e.g., charts, graphs, etc.). If you click on one of the items in the outline pane it will appear in the results pane with a box around it, showing that it is the active object (i.e., and can be modified, etc.). You will not that the basic scatterplot of firstyr versus enroll is already displayed in the results window, similar to the one shown below.



8. In order to do some editing of the basic scatterplot (i.e., fitting the least-squares regression line, assigning a more appropriate title to the plot, etc.), the first step is to highlight the scatterplot of firstyr vs. enrol and then click '<u>Edit</u>' on the menu bar. Scroll to the bottom of the menu and select "SPSS Chart Object" and then select "<u>Open</u>." This series of commands will then open a SPSS Chart Editor window that contains a new icon bar and I new menu bar. Nevertheless, the original basic scatterplot is reproduced in full in the new window.



9. To the plot the least-squares regression line, **click once on a point** that is plotted on the graph. This should cause all the points to turn blue (as shown above). On the toolbar at the top of the page, click **Elements**, and select **<u>Fit</u> line at Total**.

10. A Properties box will appear. Select **Linear**. Then click the **Close** button.

Properties	×
Chart Size Lines Fit Line	
🗖 <u>D</u> isplay Spikes	
Fit Method	ור
C Mean of Y	
🖉 🖸 Linear 📈 C Cubic	
% of points to fit: 50	
Kernel: Epanechnikov 🔽	
Confidence Intervals	
⊙ <u>N</u> one	
C M <u>e</u> an	
C Individual	
% : 95	
Apply <u>C</u> lose <u>H</u> elp	

11. Your graph should have a line through the middle of al the points like the graph below. Close the Chart editor and return to the output file.



After viewing what the scatterplaot looks like, the next step will be to conduct a regression analysis on this sat of data. This procedure can be executed from within the SPSS Viewer window. Perform the following steps in order to produce output that is similar to that shown on page 118 of the course text:

- 1. Click **Analyze** on the main menu bar.
- 2. Click **Regression** on the <u>Statistics menu</u>.
- 3. Click <u>Linear...</u>in the <u>Regression</u> submenu. This will open a Linear Regression dialog box similar to the one shown below.

Linear Regression		
🖋 firstyr	Dependent:	OK <u>P</u> aste <u>R</u> eset Cancel Help
	Selection Variable: Rule Case Labels: WLS Weight: Statistics Plots Save Opti	ons

4. The next step is to specify which is the independent variable and which is the dependent variable. In this example, 'firstyr' is the independent variable and 'enroll' is the dependent. To specify 'firstyr' as the independent variable, simply click the arrow button to the left of the 'Independent[s]:' text box. 'Firstyr' is then entered into that text box. Finally, to specify 'enroll' as the dependent variable, go to the variable source list ad click 'enrol' and then click the arrow button to the left of the 'Dependent:' text box.

SPSS for Windows will not produce the Residuals Statistics shown in the course text by default. The following steps outline how to instruct SPSS for Windows to produce summary statistics for the residuals.

 Click the 'Save...' pushbutton located at the bottom of the Linear Regression dialog box. This will open a "Linear Regression: Save New Variables" dialog box like the one shown below:

Linear Regression: Save		
Predicted Values ✓ Unstandardized ✓ Standardized ✓ Adjusted ✓ S.E. of mean predictions Distances ✓ Mahalanobis ✓ Cook's ✓ Leverage values Prediction Intervals Mean Individual Confidence Interval: 95 %	Residuals ✓ Unstandardized ✓ Standardized ✓ Studentized ✓ Deleted ✓ Studentized deleted Influence Statistics DfBeta(s) Standardized DfBeta(s) DfFit Standardized DfFit Covariance ratio	Continue Cancel Help
Coefficient statistics Create coefficient statistics Create a new dataset Dataset name: Write a new data file File Export model information to XML file Include the covariance matrix	e Bro <u>w</u> se	

- 2. Click all the check boxes within all the following three boxes in the dialog box, namely: **Predicted Values; Distances;** and **Residuals.**
- 3. Click the **Continue** command pushbutton to register these choices and to return to the Linear Regression dialog box.

Your output of the residual statistics will be displayed in the **Data view**, as illustrated below,

🔛 *Unti	tled1 [DataS	iet0] - SPSS	Data Editor						
<u>File E</u> dit	<u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	nalyze <u>G</u> raphs <u>U</u> tilitie	es <u>W</u> indow <u>H</u> elp					
🖻 🔒	🗁 🖶 📴 🛧 🔶 🐜 🕴 🛤 📲 🏥 🖽 🎼 🥸 🔕								
12 : PRE_1									
	firstyr	enroll	PRE 1	RES_1	DRE_1				
1	4595.00	7364.00	7392.44298	-28.44298	-37.03622				
2	4827.00	7547.00	7639.82977	-92.82977	-226.33390				
3	4427.00	7099.00	7213.30083	-114.30083	-131.37195	Ξ			
4	4258.00	6894.00	7033.09235	-139.09235	-165.65625				
5	3995.00	6572.00	6752.64957	-180.64957	-342.91526				
6	4330.00	7156.00	7109.86756	46.13244	53.09160				
7	4265.00	7232.00	7040.55661	191.44339	226.94856				
8	4351.00	7450.00	7132.26033	317.73967	364.00602				
9									
10									
- 11									
12									
13									
14									
15									
			/			V			
	ata View <u>(</u> Va	ariable View _/				>			
SPSS Processor is ready									

The SPSS output for this example of the Linear Regression is the following:

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	firstyr(a)		Enter

a All requested variables entered.

b Dependent Variable: enroll

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.833(a)	.694	.643	188.94853

a Predictors: (Constant), firstyr

The correlation is r = .833. the R² statistic above indicated that 69.4% of the variance in y ("response variable") is explained by changes in x ("explanatory variable").

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	486552.22 9	1	486552.229	13.628	.010(a)
	Residual	214209.27 1	6	35701.545		
	Total	700761.50 0	7			

a Predictors: (Constant), firstyr

b Dependent Variable: enroll

We will discuss the ANOVA table in chapter 10.

Coefficients(a)

		Unstandardized Coefficients		Standardized Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1	(Constant)	2492.692	1267.199		1.967	.097
	firstyr	1.066	.289	.833	3.692	.010

a Dependent Variable: enroll

The least square equation for the model is found in the *Unstandardized Coefficient* table under the *B* column. It is y = 2492.692 + 1.066x.

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	Ν
Predicted Value	6752.6494	7639.8296	7164.2500	263.64268	8
Std. Predicted Value	-1.561	1.804	.000	1.000	8
Standard Error of Predicted Value	67.363	145.116	90.049	30.550	8
Adjusted Predicted Value	6914.9150	7773.3340	7196.6584	275.51490	8
Residual	۔ 180.64957	317.73969	.00000	174.93234	8
Std. Residual	956	1.682	.000	.926	8
Stud. Residual	-1.317	1.800	068	1.063	8
Deleted Residual	- 342.91525	364.00601	-32.40843	237.03902	8
Stud. Deleted Residual	-1.426	2.422	.010	1.241	8
Mahal. Distance	.015	3.254	.875	1.258	8
Cook's Distance	.004	.779	.207	.272	8
Centered Leverage Value	.002	.465	.125	.180	8

a Dependent Variable: enroll