

Chapter 8: Regression

Self-test answers



SELF-TEST Residuals are used to compute which of the three sums of squares?

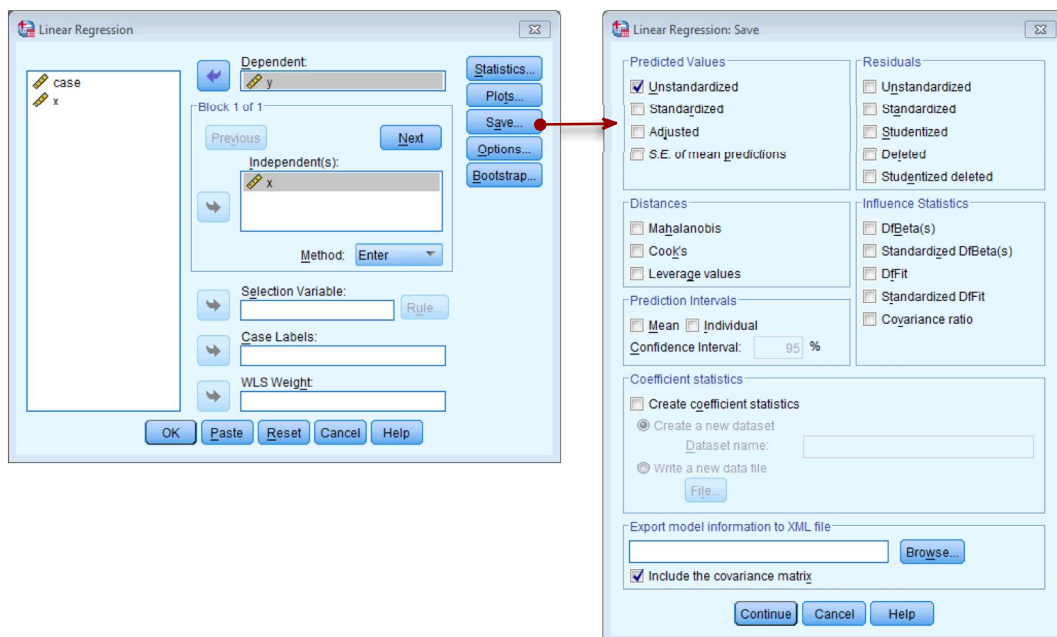
The residuals are used to calculate the residual sum of squares (SS_R). This value is calculated by first squaring the residuals and then adding them up.



SELF-TEST Once you have read Section 8.4, run a regression first with all the cases included and then with case 30 deleted.

To run the analysis on all 30 cases, you need to access the main dialog box by selecting **Analyze > Regression > Linear...**. The figure below shows the resulting dialog box.

There is a space labelled *Dependent* in which you should place the outcome variable (in this example *y*). There is another space labelled *Independent(s)* in which any predictor variable should be placed. In simple regression we use only one predictor (in this example, *x*). Click on **Save...** and tick *Unstandardized* predicted values (see figure below), and then click on **Continue** to return to the main dialog box. Click on **OK** to run the analysis.



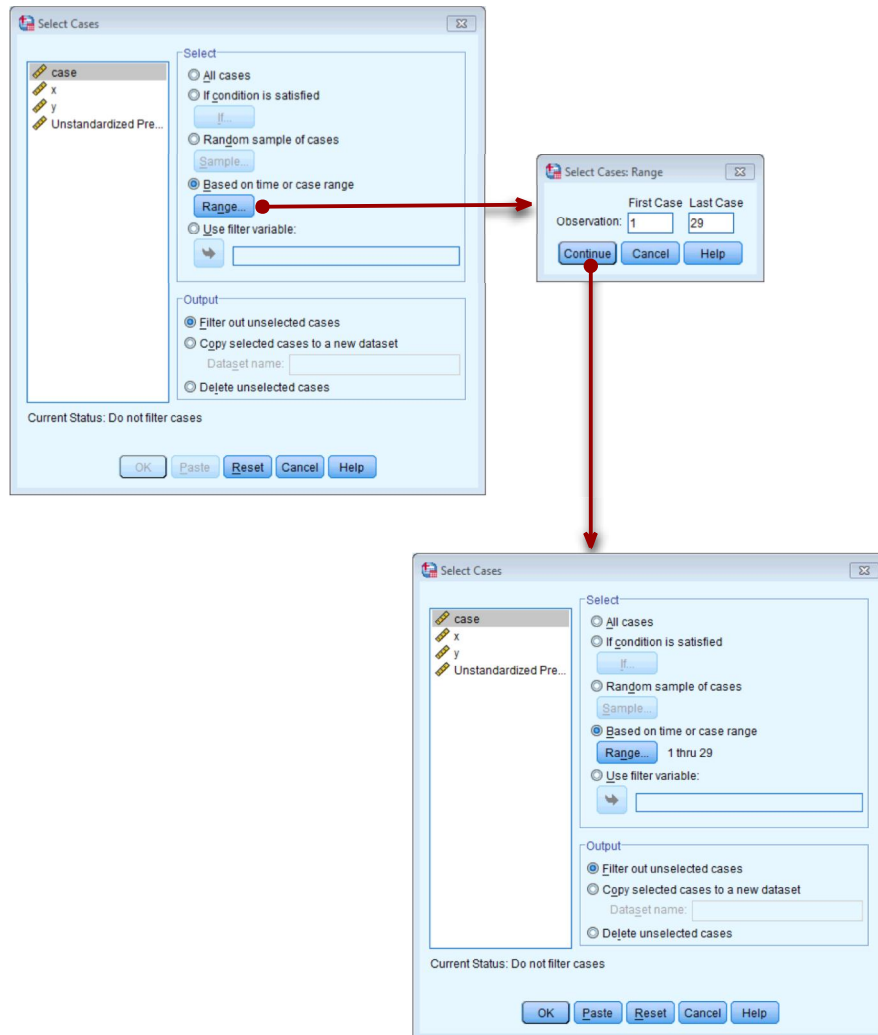
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	29.000	.992		29.237	.000
	x	-.903	.056	-.950	-16.166	.000

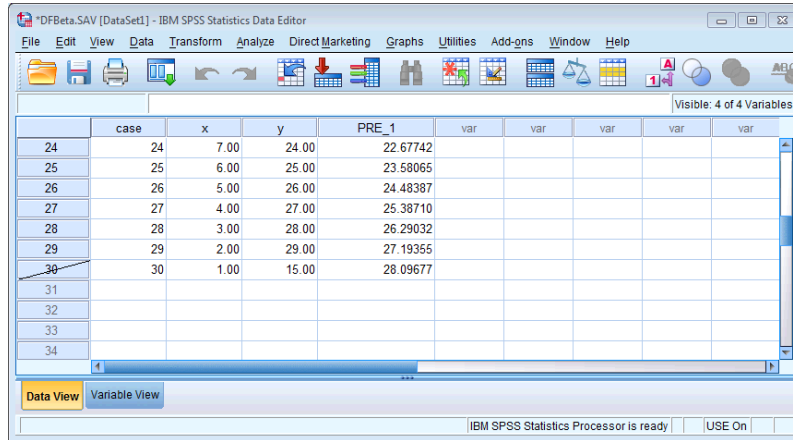
a. Dependent Variable: y

After running the analysis you should get the output above. See the book chapter for an explanation of these results.

To run the analysis with case 30 deleted, go to [Data >> Select Cases...](#) to open the dialog box in the figure below. Once this dialog box is open select *Based on time or case range* and then click on *Range*. We want to set the range to be from case 1 to case 29, so type these numbers in the relevant boxes (see figure below). Click on [Continue](#) to return to the main dialog box and then click on [OK](#).



Once you have done this, your data should look like mine below. You will see that case 30 now has a diagonal strike through it to indicate that this case will be excluded from any further analyses.



Now we can run the regression in the same way as we did before by selecting **Analyze** **Regression** **Linear...**. As before, we need to place the outcome variable (in this example **y**) in the space labelled *Dependent* and place the predictor variable (in this example, **x**) in the space labelled *Independent(s)*. Next, click on **Save...** and tick *Unstandardized* predicted values, and then click on **Continue** to return to the main dialog box. Click on **OK** to run the analysis. You should get the same output as mine below (see the book chapter for an explanation of the results).

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	31.000	.000		.	.
	x	-1.000	.000	-1.000	.	.


a. Dependent Variable: y

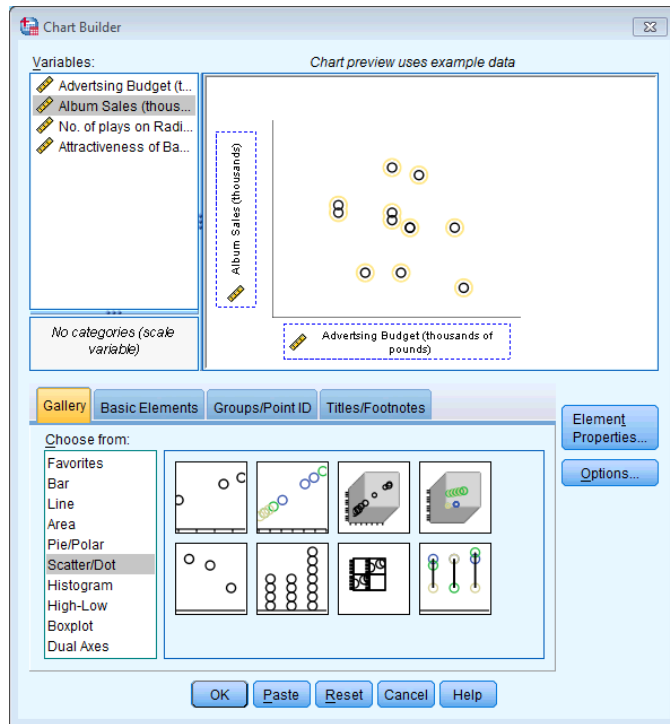
	case	x	y	PRE_1	PRE_2	var	var	var
	24	7.00	24.00	22.67742	24.00000			
	25	6.00	25.00	23.58065	25.00000			
	26	5.00	26.00	24.48387	26.00000			
	27	4.00	27.00	25.38710	27.00000			
	28	3.00	28.00	26.29032	28.00000			
	29	2.00	29.00	27.19355	29.00000			
	30	1.00	15.00	28.09677	.			
	31							
	32							
	33							
	34							



Once you have run both regressions, your data view should look like mine above. You can see two new columns **PRE_1** and **PRE_2** which are the saved unstandardized predicted values that we requested.

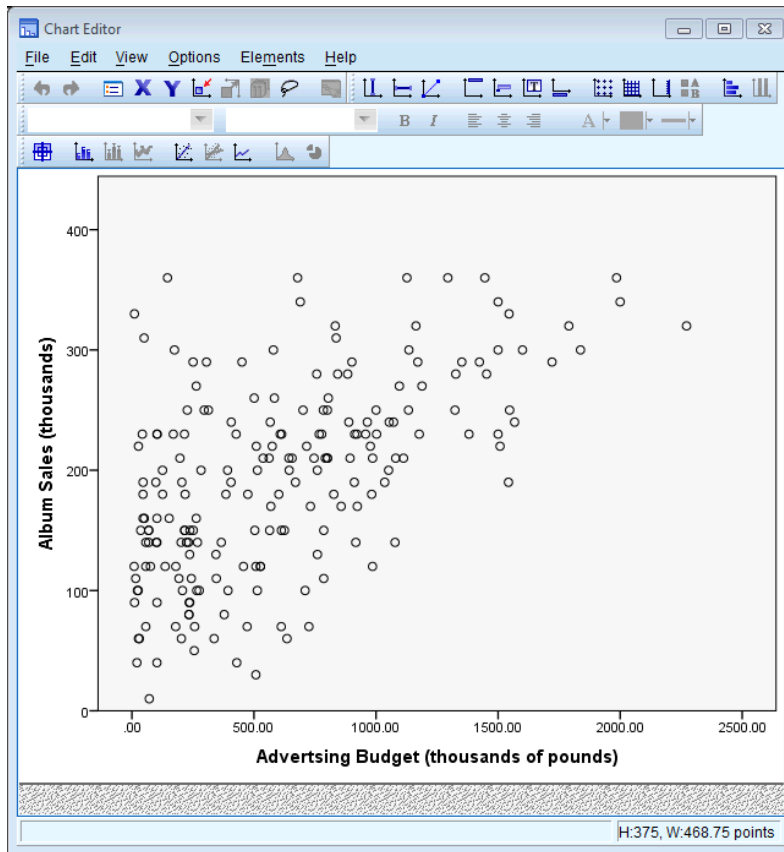


SELF-TEST Produce a scatterplot of advertising budget (*x*-axis) against sales (*y*-axis). Include the regression line.

To do the scatterplot, first access the chart builder, complete the dialog box as I have done below and then click on .



To add a regression line, double-click on the scatterplot to open the chart editor (see below), click on  which will open the *Properties* dialog box, select  Linear and then click on



The "Properties" dialog box is open, showing the "Fit Line" tab. It includes the following settings:

- Display Spikes
- Suppress intercept
- Fit Method:**
 - Mean of Y
 - Linear
 - Loess
 - Quadratic
 - Cubic
- % of points to fit:
- Kernel:
- Confidence Intervals:**
 - None
 - Mean
 - Individual
- Confidence Interval %:

Buttons at the bottom: Apply, Close, Help.

See the book chapter for the completed scatterplot.



SELF-TEST How is the t in Output 8.3 calculated? Use the values in the table to see if you can get the same value as SPSS.

It is calculated using this equation:

$$\frac{134.14 - 17.79}{0.096} = \frac{116.35}{0.096} = 1211.875$$

Using the values from SPSS Output 8.3 to calculate t for the constant ($t = 134.140/7.537 = 17.79$), for the advertising budget, we get: $0.096/0.01 = 9.6$. This value is different than the one in the output ($t = 9.979$) because SPSS rounds values in the output to three decimal places, but calculates t using unrounded values (usually this doesn't make too much difference, but in this case it does!). In this case the rounding has had quite an effect on the standard error (its value is 0.009632 but it has been rounded to 0.01). To obtain the unrounded values, double-click the table in the SPSS output and then double-click the value that you wish to see in full. You should find that $t = 0.096124/0.009632 = 9.979$.




SELF-TEST How many records would be sold if we spent £666,000 on advertising the latest CD by black metal band Abgott?

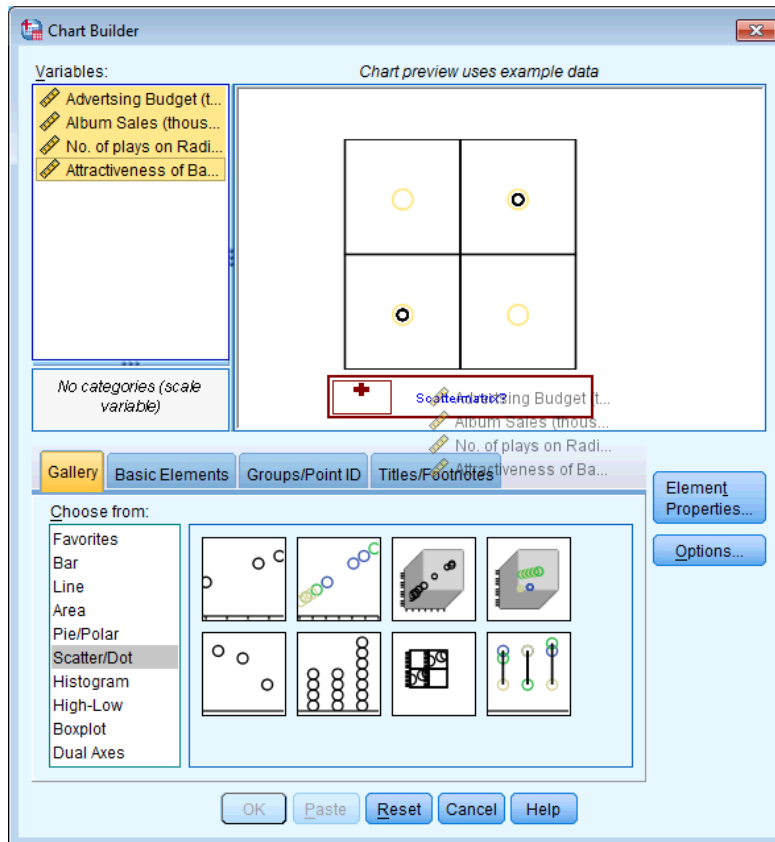
He would sell 198,080 CDs:



$$\text{album sales} = 134.14 + 0.096 \times \text{advertising budget} \\ = 134.14 + 0.096 \times 666 = 198.08$$

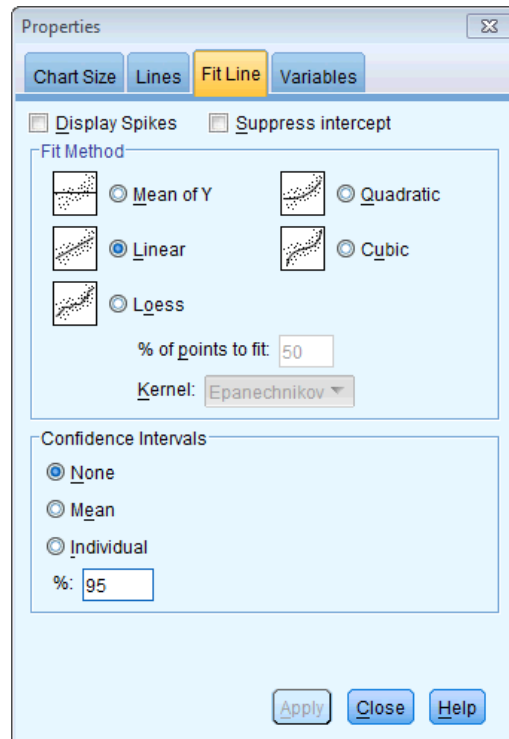


SELF-TEST Produce a matrix scatterplot of **Sales, Adverts, Airplay** and **Attract** including the regression line.

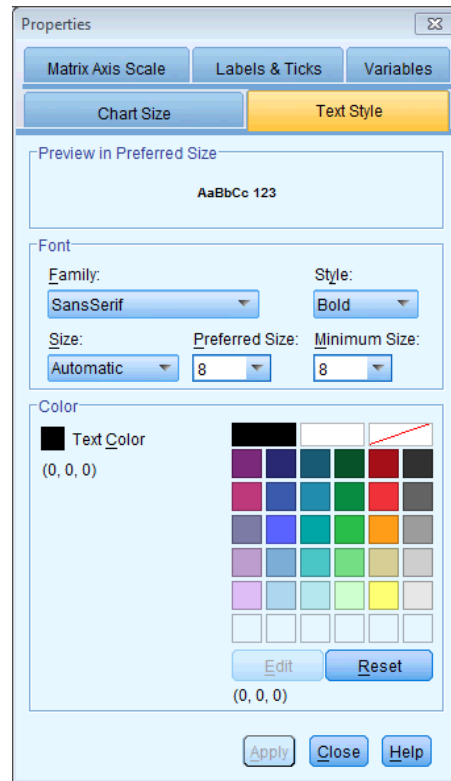
First you need to access the chart builder and select the matrix scatterplot. Next select all the variables in the *Variables list* on the left, move them into the box on the x-axis of the matrix (see my completed dialog box below), and then click on .



To add regression lines to your matrix, double-click on the resulting matrix scatterplot in your output to open the chart editor (see below) and click on , which will open the *Properties* dialog box. Select  **Linear** and then click on **Apply**.



You will probably find that because the variable labels are all quite long, SPSS will not display them all on the axes of the graph. To rectify this you can edit the font size in the chart editor window. To do this, double click on one of the variables along the x-axis to open the *Text style* tab in the *Properties* dialog box (see below). Here you can change the values of *Preferred Size* and *Minimum Size* to a smaller number (I chose 8), then click on **Apply** to apply this change to your matrix. You then need to repeat this for the y-axis.



You can then close the chart editor to reveal your completed matrix scatterplot. See the book chapter for my completed matrix.