STAT E-150 - Statistical Methods
Introduction to SPSS

How to access SPSS from Harvard Computer Services

Before you begin to install SPSS from Harvard Computer Services:

- You must have a Harvard ID (HUID) and PIN to activate an FAS account or to access to most online applications. You can find your HUID on your Extension School registration confirmation statement and/or bill. (Note: this is *not* the same as your DCEID which begins with "@")
- If you do not have a PIN, you can request one online at www.pin.harvard.edu. Your PIN will be emailed to the email address you provided when you registered with the Extension School.
- You must email HUIT Support at help@fas.harvard.edu to request an Authorization Code. The Code will be sent to you upon resolution of your support request, and then you can follow the instructions below.

Once you have your HUID, PIN, and authorization code, go to: www.fas.harvard.edu/computing

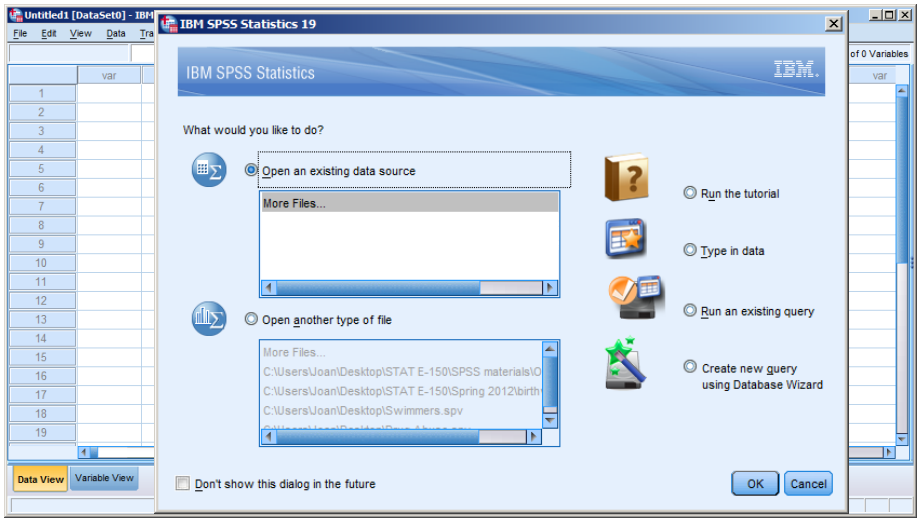
1. Click on "Other Students" under "New to FAS?"
You will use your Harvard ID and PIN in order to create an email account.

2. Once you have an account, click on Software Downloads at the top of the page.
This will lead you to a page where you login using the ID and PIN discussed above.

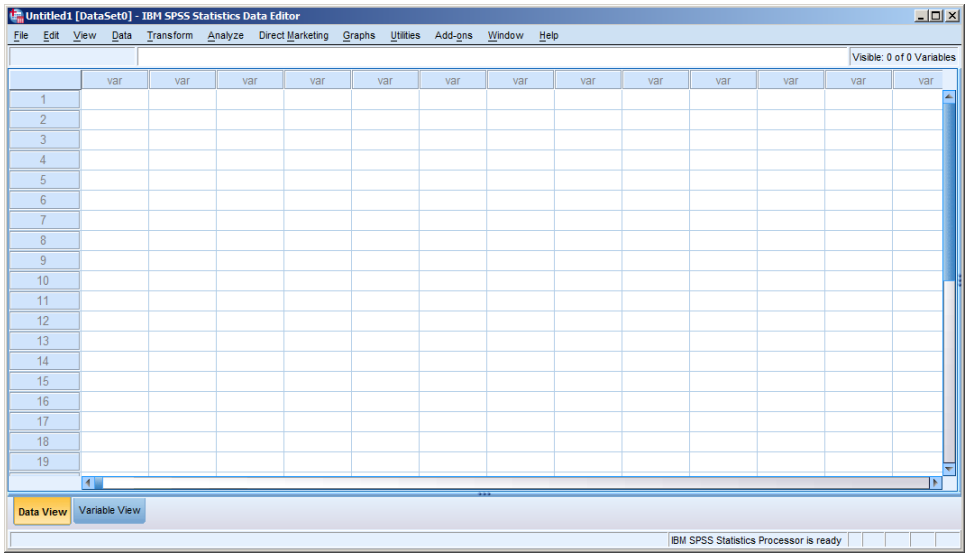
3. After you log in, you will be directed to a list of software downloads and you can download SPSS.

Entering and Saving Data

When you open SPSS your screen should look like this, with a menu dialog box in the front and the Data Editor behind it. The Data Editor looks like a spreadsheet, and that is where you can define variables and enter data.



Click on **Cancel** to access the Data Editor window. Each column represents a variable, and each row represents an observation.

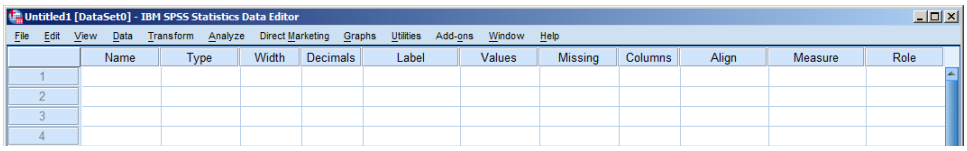


Data can be entered by opening a data file, but we will start by typing data into the Data Editor. First click on the Variable View tab at the bottom of the window.

Here is the data we have been discussing in class:

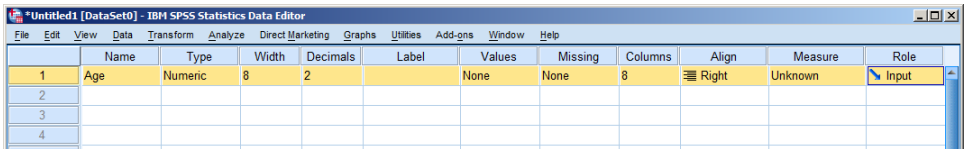
Observation	1	2	3	4	5	6	7	8	9	10
Maternal Age	15	17	18	15	16	19	17	16	18	19
Birthweight (g)	2289	3393	3271	2648	2897	3327	2970	2535	3138	3573

Define the variables by entering a name, the type of data, the number of decimal places, etc., for each variable. You can see the column headings at the top:

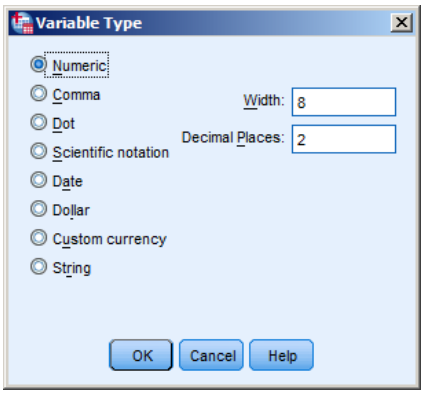


1. Go to the first empty cell in Row 1, under Name, and type in the first variable name. We will use the variable name **Age**.

2. Press **Enter** or **Tab**.

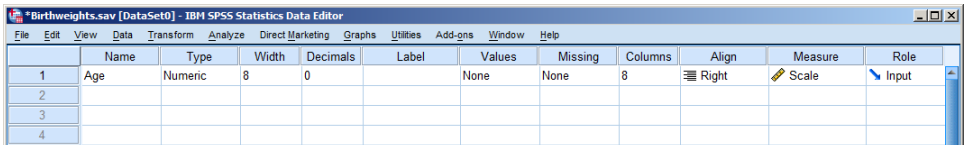


3. Click in the **Type** column, and a small button with three dots will appear. Click on this button to see this dialog box:

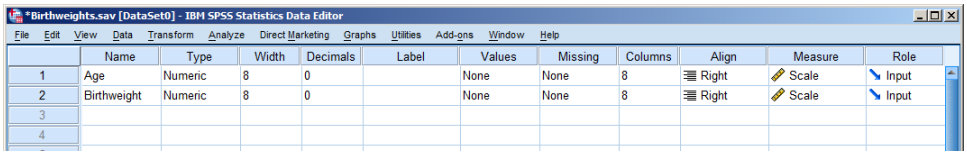


Note that the default variable type is Numeric, which is appropriate for the data we are using. Set the number of decimal places to 0.

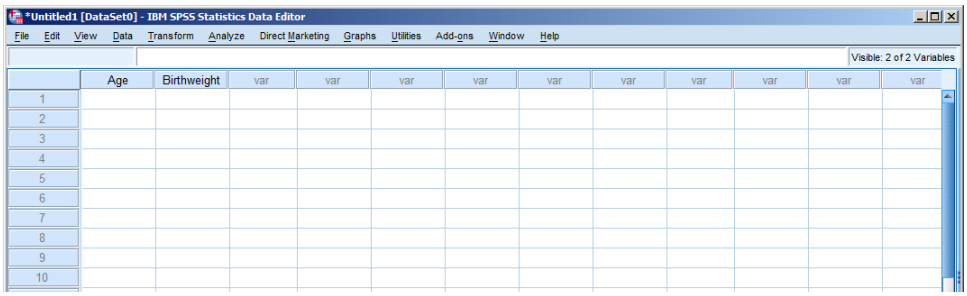
4. Set the Measurement Level of these two variables to **Scale.**



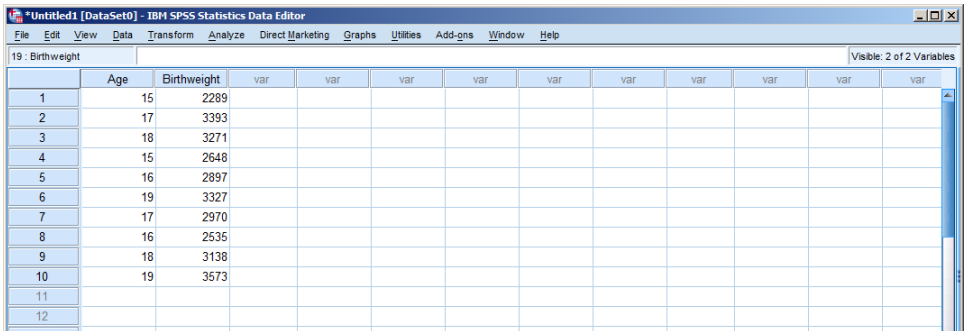
5. Repeat these steps to create the **Birthweight** variable.



6. Switch to the Data View to enter the data:



Your window should look like this:



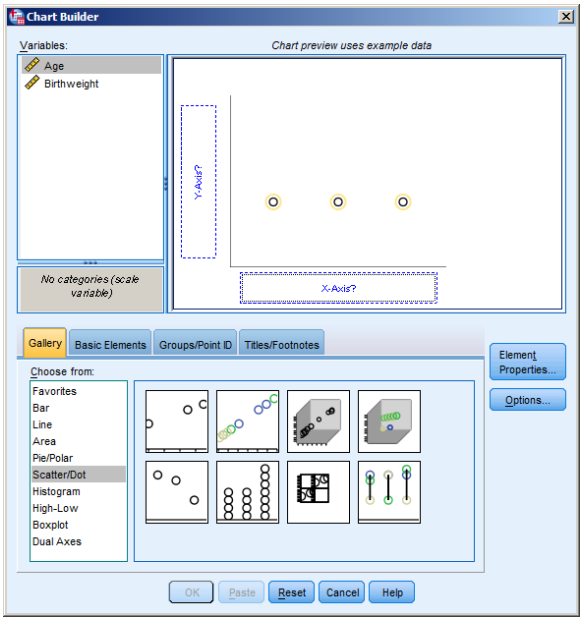
7. To save your file, go to **File > Save As…** and indicate where you would like to save your work. Be sure to use an appropriate filename. Click **Save** when you have entered all of the necessary information.

You will see an Output Viewer window that confirms that your file has been saved.

Creating a Scatterplot

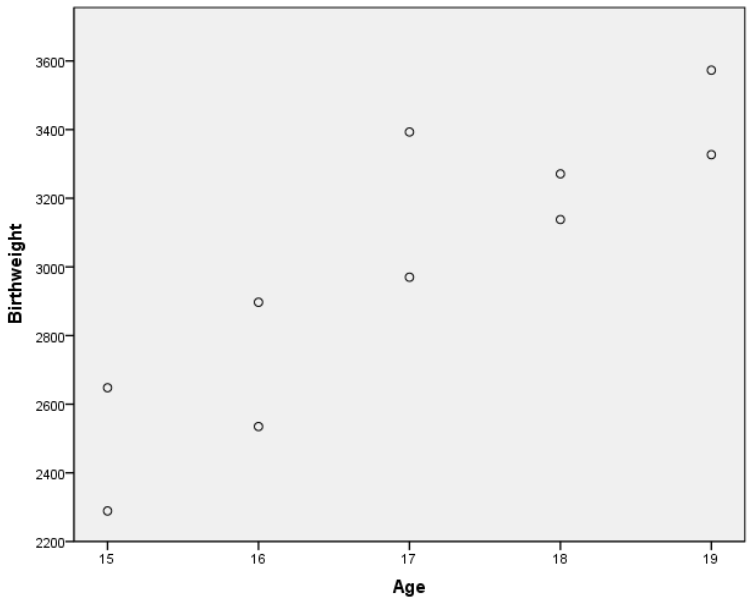
1. Click on **Graphs > Chart Builder**.

   Select **Scatter/Dot** from the list of charts.
   Drag the first Scatterplot option to the window.



   Drag the explanatory variable, Age, to the x-axis, and drag the response variable, Birthweight, to the y-axis. You will see a sample scatterplot; this is not a display of the data that was entered.

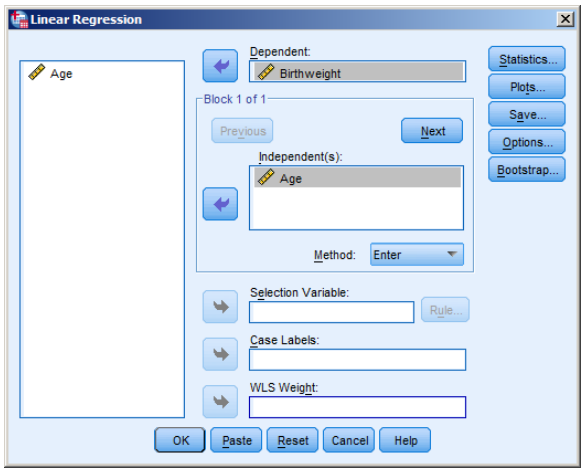
   Click on OK. The scatterplot of the data will appear in the **Output Viewer**.



Finding the Linear Regression Equation

1. Click on **Analyze > Regression > Linear**

   Drag the variables to the appropriate locations and then click on **OK**



To see the output, go to the **Output Viewer** which you can access by using the **Window** menu.

You will see several tables; these are the three we are interested in at this point:

The **Variables** table lists the independent, or explanatory variables in the model:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Birthweight

The **Model Summary** table reports the value of the correlation coefficient and the coefficient of determination:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.884[a]	.781	.754	205.308

a. Predictors: (Constant), Age

The **Table of Coefficients** table reports the coefficients in the regression model, as well as t-statistics and significance levels for the regression inference:

Coefficients[a]

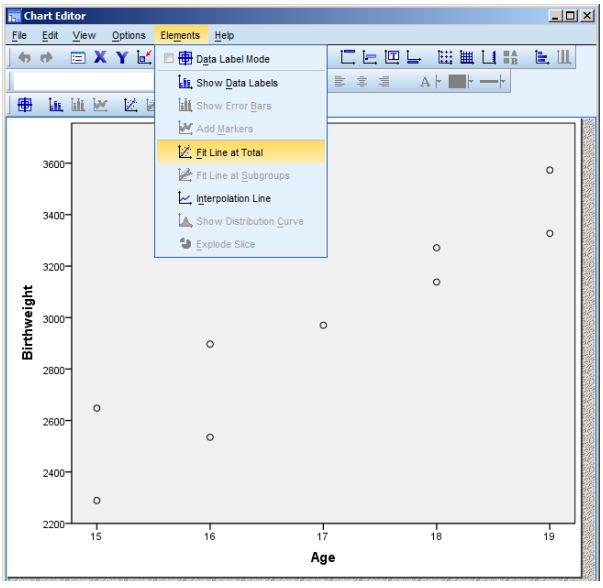
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-1163.450	783.138		-1.486	.176
	Age	245.150	45.908	.884	5.340	.001

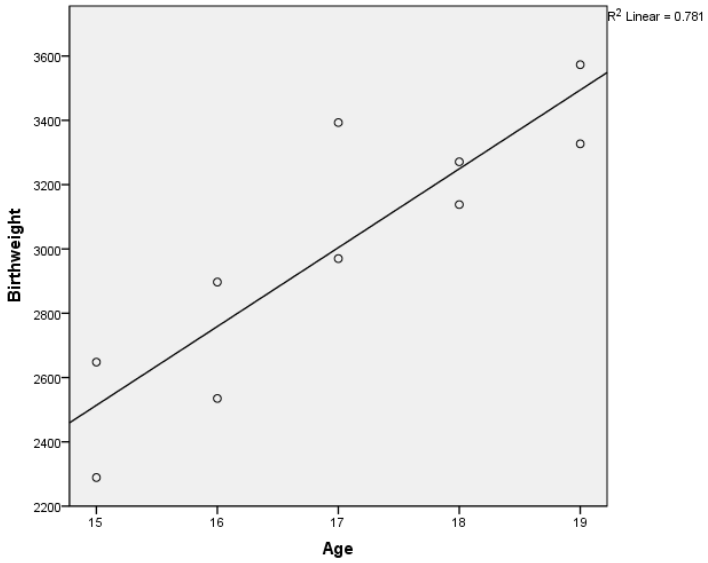
a. Dependent Variable: Birthweight

Displaying the Regression Line

1. Go to the Output viewer and double-click on the scatterplot you created. This will open the Chart Editor.

In the Chart Editor, click on > **Elements** > **Fit Line at Total** and then close the Chart Editor. In the upper right corner, you will see value of R^2



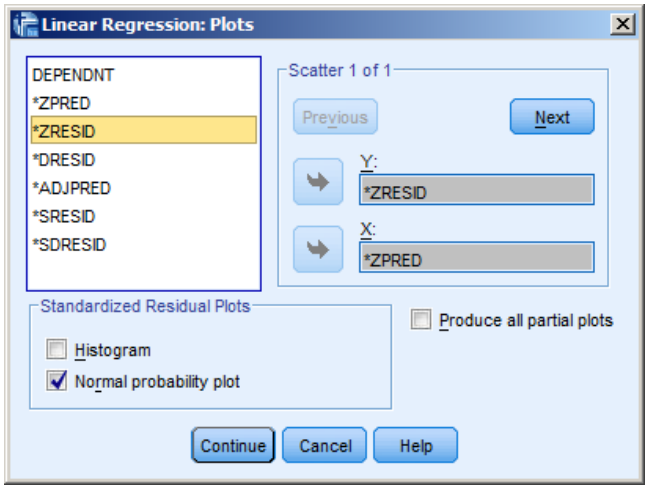


Examining Residuals

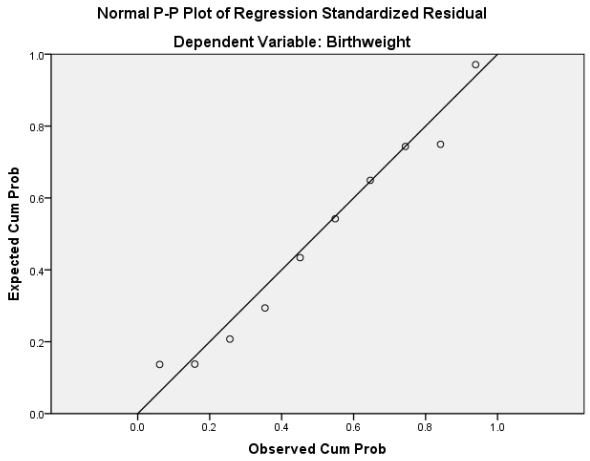
1. Create a scatterplot with the regression line displayed.

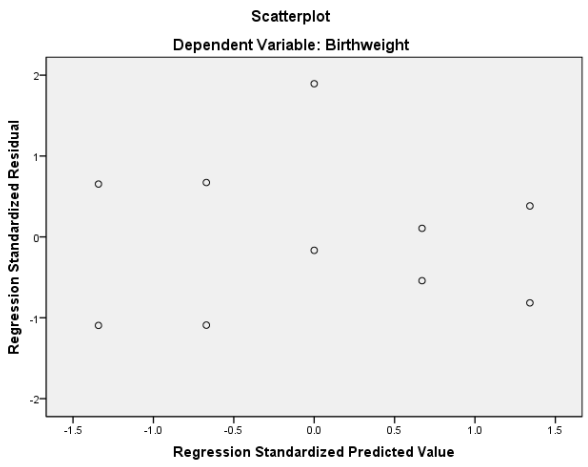
2. Click on **Analyze > Regression > Linear**

3. As before, choose the explanatory and response variables, and click on **Statistics**. This will open a new dialog box which will allow you to choose the output you need.



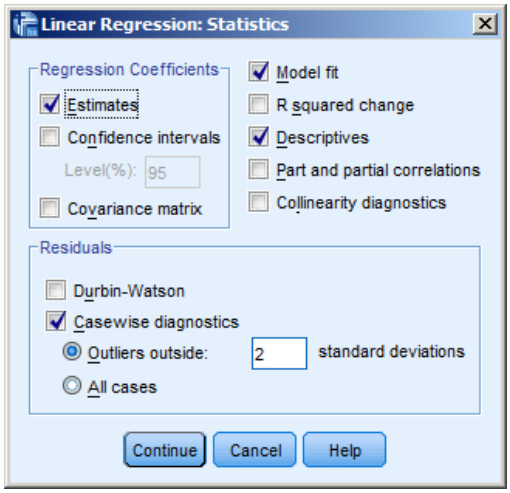
4. In the next dialog box, click on **Plots**, and make the selections shown above in the **Linear Regression Plots** dialog box. This will produce a Normal Probability Plot of the residuals and a scatterplot of the standardized predicted values (*ZPRED) vs. the standardized residuals (*ZRESID).





Getting confidence intervals:

Make the selections shown in the **Linear Regression Statistics** box, and then click on **Continue**:



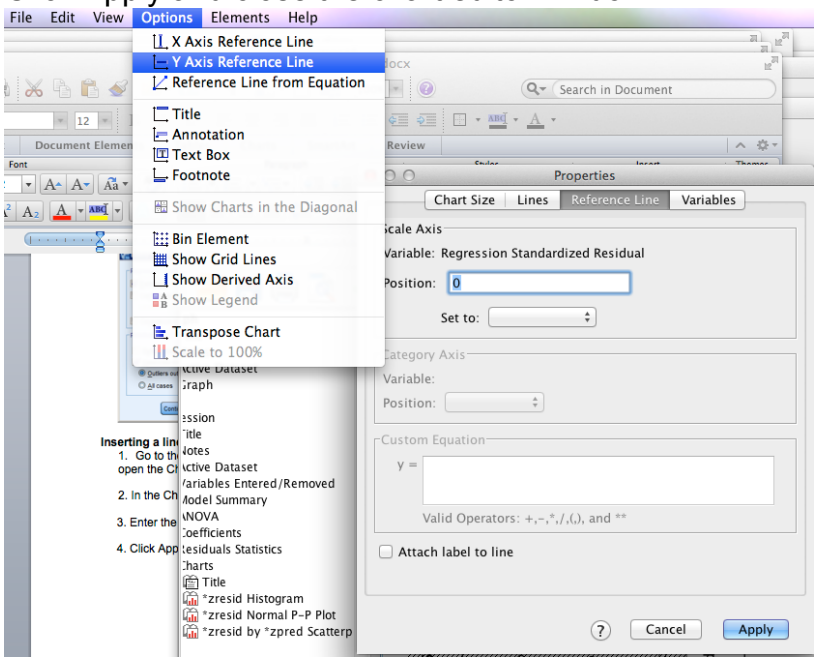
Inserting a line with a constant value, like at 0 for the residual graph.

1. Go to the Output viewer and double-click on the scatterplot you created. This will open the Chart Editor.

2. In the Chart Editor, click on > **Options** > **Y axis reference line**

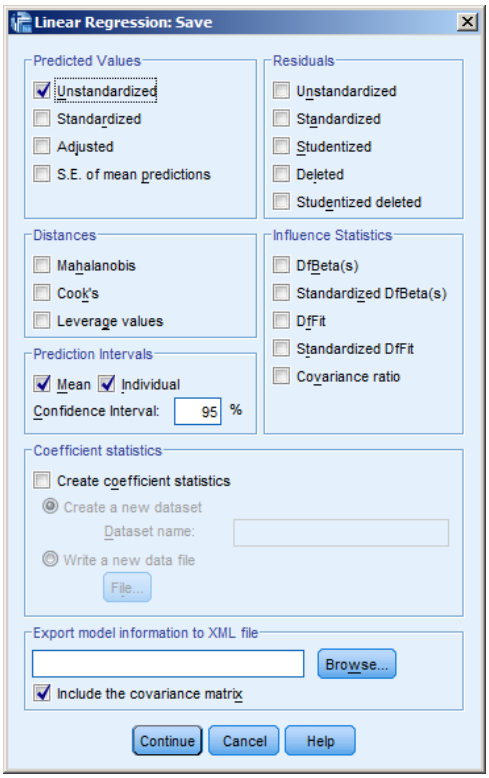
3. Enter the value you want the line at in the Position: box

4. Click Apply and close the chart editor window.

5. 

Making Predictions

Return to your data, and again select **> Analyze > Regression > Linear**
Click on **Save…** and make the selections shown below.

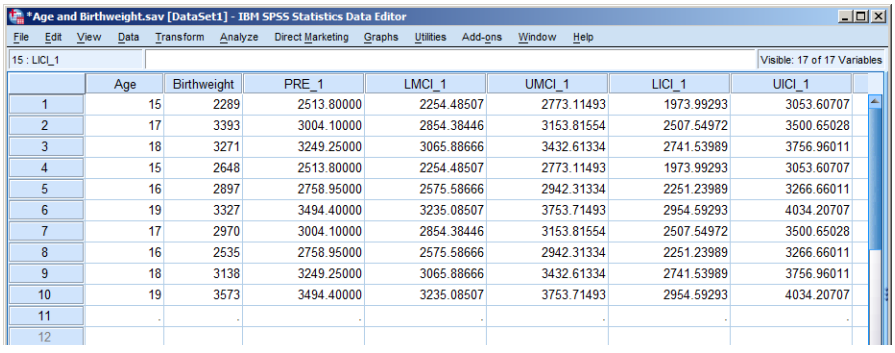


2. Click on **> Window** and select your dataset.
   You will see the table shown below.

   The **PRE_1** column shows the values predicted by the regression equation.

   The **LMCI_1 and UMCI_1** columns show the upper and lower bounds of the 95% *confidence* interval for the mean values of y.

   The **LICI_1 and UICI_1** columns show the upper and lower bounds of the 95% *prediction* interval for the individual values of y.



	Age	Birthweight	PRE_1	LMCI_1	UMCI_1	LICI_1	UICI_1
1	15	2289	2513.80000	2254.48507	2773.11493	1973.99293	3053.60707
2	17	3393	3004.10000	2854.38446	3153.81554	2507.54972	3500.65028
3	18	3271	3249.25000	3065.88666	3432.61334	2741.53989	3756.96011
4	15	2648	2513.80000	2254.48507	2773.11493	1973.99293	3053.60707
5	16	2897	2758.95000	2575.58666	2942.31334	2251.23989	3266.66011
6	19	3327	3494.40000	3235.08507	3753.71493	2954.59293	4034.20707
7	17	2970	3004.10000	2854.38446	3153.81554	2507.54972	3500.65028
8	16	2535	2758.95000	2575.58666	2942.31334	2251.23989	3266.66011
9	18	3138	3249.25000	3065.88666	3432.61334	2741.53989	3756.96011
10	19	3573	3494.40000	3235.08507	3753.71493	2954.59293	4034.20707
11	.	.	.	.	.	.	.
12							

To make predictions for a value of the explanatory variable which is not in your dataset:

1. Click on **> Window** and select your dataset.

2. Enter the new value in the column for the explanatory variable. Here we will make predictions for a mother who is 20 years of age.

3. Click on **> Analyze > Regression > Linear**
   then click on **Save**

   You will see the same dialog box you used earlier, and will generate the same results for the new variable you entered.

   Click on **Continue**

4. In the next dialog box, click on **OK**

5. Click on **Window** and select your dataset.
   At the left side of the spreadsheet you will see the value you entered.

6. Scroll to the right; in the row where you entered the value of the explanatory variable you will see the predicted value of the response variable as well as the confidence interval and the prediction interval for this value.