7. TUTORIAL FOR NEURAL NET FORECASTING

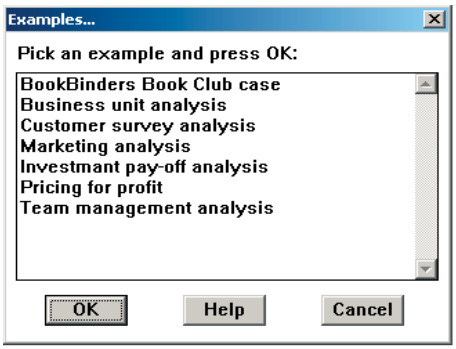
CASE: BOOKBINDERS BOOK CLUB, P. 185

A neural network is a general response model that can help firms to understand complex relationships between the dependent and independent variables of a marketing system. The network consists of a number of interconnected nodes (neurons), each of which contributes in a specific way to the overall relationship between the variables. Unlike regression models, neural network models do not require the users to prespecify the nature of the relationship between the dependent and independent variables.

The neural network software is supplied by Cognos Corporation and is called 4Thought. It uses a feedforward network with up to two hidden layers and a sigmoidal activation function. It fits the model using back propagation of errors, and it minimizes the possibility of overfitting by continuously testing the model on a hold-out sample.

***Note**: 4Thought is a comprehensive software program that includes an online help file. Here we describe only the features that you will need for the BookBinders Book Club (BBBC) case.*

For this exercise, you will build a neural net model to explain response to a direct mail offer from Bookbinders Book Club. From the **Model** menu, select **Neural Net Forecasting**. You will then see the following screen:



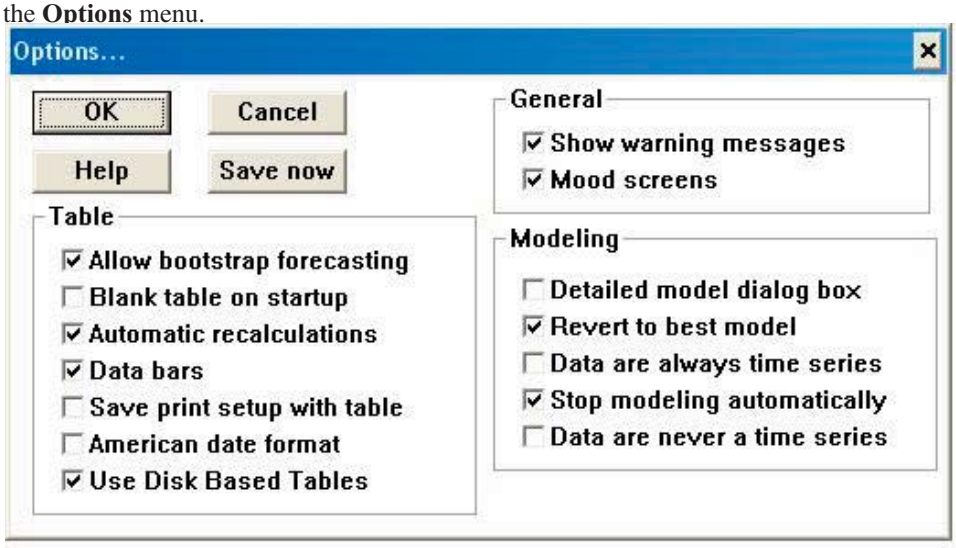
Choose the **Bookbinders Book Club Case** and hit **OK**. Then you will go through a series of informational dialog boxes, accept them, until you see a spreadsheet listing the variables for this exercise. The first 400 observations are from customers who responded to the direct mail offer (Choice = 1), and the remaining 1200 observations are from customers who did not (Choice = 0). The file also contains an additional 2300 observations that you can use for verification or forecasting.

BBBCNN.4TH

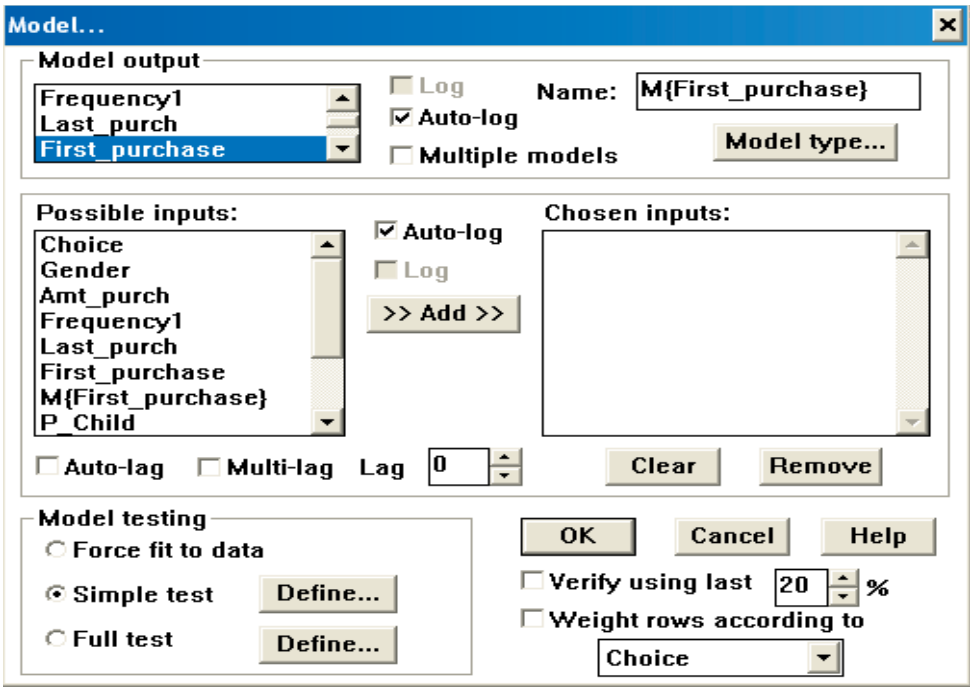
Undo | Redo | Recall

14:46	123	123	123	123	123	123	123	123	123	123
1Rx11C	Choice	Gender	Amt_purch	Last_purc	First_purc	Frequency	P_Child	P_Youth	P_Cook	P_DIY
1	1.00	1.00	113.00	8.00	1.00	8.00	0.00	1.00	0.00	0.00
2	1.00	1.00	418.00	6.00	11.00	66.00	0.00	2.00	3.00	2.00
3	1.00	1.00	336.00	18.00	6.00	32.00	2.00	0.00	1.00	1.00
4	1.00	1.00	180.00	16.00	5.00	42.00	2.00	0.00	0.00	1.00
5	1.00	0.00	320.00	2.00	3.00	18.00	0.00	0.00	0.00	1.00
6	1.00	1.00	268.00	4.00	1.00	4.00	0.00	0.00	0.00	0.00
7	1.00	1.00	198.00	2.00	12.00	62.00	2.00	3.00	2.00	1.00
8	1.00	0.00	280.00	6.00	2.00	12.00	0.00	2.00	0.00	0.00
9	1.00	1.00	393.00	12.00	11.00	50.00	3.00	0.00	3.00	0.00
10	1.00	1.00	138.00	10.00	7.00	38.00	2.00	3.00	0.00	0.00
11	1.00	0.00	142.00	14.00	2.00	20.00	0.00	1.00	0.00	0.00
12	1.00	0.00	251.00	10.00	2.00	18.00	0.00	0.00	2.00	0.00
13	1.00	0.00	130.00	14.00	2.00	18.00	1.00	0.00	0.00	0.00
14	1.00	1.00	308.00	6.00	1.00	6.00	0.00	0.00	0.00	0.00
15	1.00	1.00	251.00	10.00	2.00	18.00	1.00	0.00	0.00	0.00

First, go to the **Options** menu and choose **Options**, and then select the appropriate settings and press **OK**. In particular, check the **Detailed model dialog box**. Until you become familiar with the software, you may also want to choose the **Simple menus** from the **Options** menu.



Next, specify a model for analysis. From the **Specify** menu, select **Model**. You will see



the following screen:

DEFINING THE MODEL

Specifying the dependent variable: First select a dependent variable (here Choice) for **Model output**. The dependent variable can be any table column containing numerical data. If you wish to build models of several columns all at once, all using the same inputs, check the **Multiple models** box and select all the columns you wish to model. If you wish to model the logarithm of the column's contents instead of the raw data, check **Log** before adding the column to the input list. If you want 4Thought to automatically decide whether the distribution of the data in the column merits taking a log of the raw data, then select the **Auto-log** check box before adding the column to the input list.

Specifying the model type: You can specify the type of neural network model that you wish to explore by clicking on **Model type**. Unless you are very familiar with neural network models, we recommend that you retain the **Chosen by 4Thought** option.

Specifying the independent variables: Define the independent variables (Possible inputs) in the lower section of the dialog box. You can add items by clicking on the **Add** button. You can remove an item by clicking on the **Remove** button; you can remove all items from the input list by clicking on the **Clear** button. To select multiple columns, keep the Ctrl key pressed as you select each variable.

If you select a column that is not numeric and that has not already been categorized, the **Categories** dialog box opens. You can indicate whether data for each input is to be logged or auto-logged as we did with the independent variable. If you are using time-series data, you can also choose to lag an input. This means that instead of using data from the input column's corresponding row number, you can use data from X number of rows forward

(down) or backwards (up) in the table. Alternatively, you can select the **Auto-lag** option to tell the system to designate a suitable lag period. In that case, the system will try out different lags, measuring the correlation between the data and the lagged data. It stops when the correlation begins to drop off. Note that if you have only one column of data and are building a model of it based upon itself, you have to use lags.

If you wish to add a time and/or seasonality input to the model and you have not previously specified these columns, do so at this point by clicking **Time**.

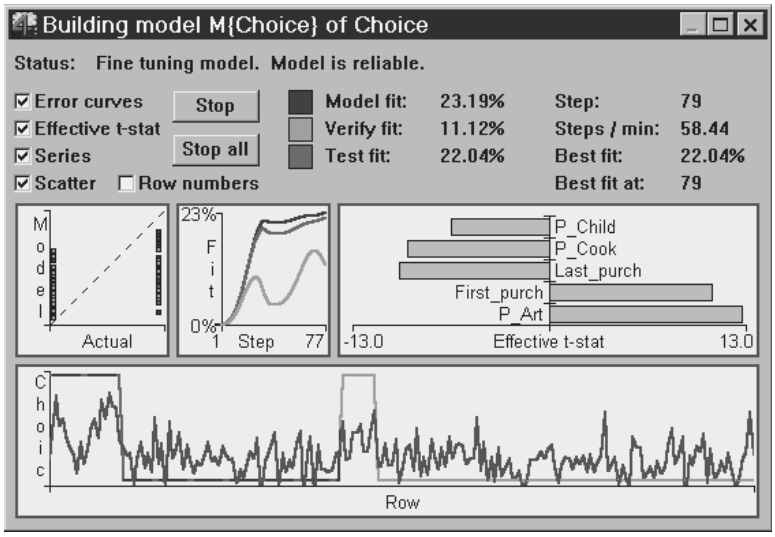
Specifying model tests: You can choose one of three methods to test the quality of the model you are building using the **Model testing** section of the dialog box. 4Thought uses the test data to decide when to stop the modeling process.

1. ***Force fit to data***: In this case, the system does not perform any holdout prediction tests. This option is useful if you wish to compare the results from the neural net with that of a traditional statistical model.
2. ***Simple test***: This option separates out a specific region of the data for testing during the model development process. We recommend this option. By clicking on the Define button, you can exercise more control on how specifically the test data is selected. For cross-sectional data, we recommend that you use the **Evenly spread** option. For time-series data, you can select **Contiguous groups** or **At end of data set**. In the case of the contiguous groups, the system selects a number of representative contiguous portions of data for testing. We recommend that you use two contiguous groups.
3. ***Full test***: Under this option, the system builds many models with different test points in order to establish an optimum point at which to stop modeling. The system then builds a final model using all the data. We recommend this option if you have a small data set.

If you wish, you can set aside additional data points for verification. This is a wholly independent test, not used in model development or holdout prediction. Data selected for verification should be marked as V under the column titled RU. In that same column, the system marks the data it has selected for model development as M and the data used for holdout prediction as T.

***Note**: For the BBBCNN.4TH data, you have to mark rows 1,601 to 3,900 with a V before you run the model. This will ensure that these data are not used for model development or testing. However, you will then obtain model predictions for the dependent variable (Choice) for these rows. The simplest way to do this is by checking **Verify using last** in the **Model** dialog box and setting this value to 59 (=2300/3900) percent.*

After you complete the model specifications, click **OK**. You will see the following screen, which displays the neural net modeling process. (Setting up the screen may take a while, and running the model may take a long time if you have a large data set.) If the **Model fit** (shown in blue) and **Test fit** (red) are roughly similar, the model should be reliable. In addition, if you set aside data for verification, the display will include **Verify fit**. If its value is roughly the same as the Model fit, you have further evidence that the model is reliable.



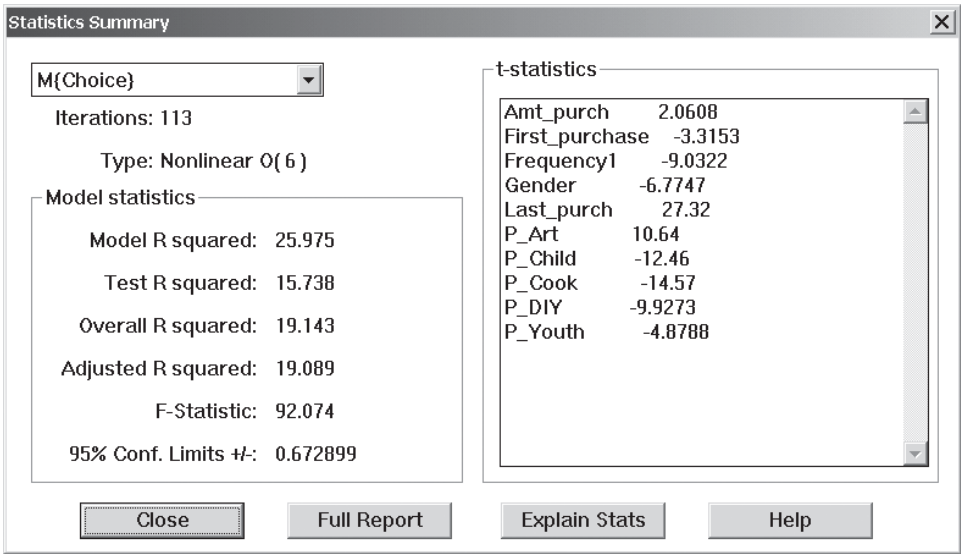
Once the program selects a final model, the spreadsheet will include two new columns. The column titled "RU" indicates how a particular row of data was used in the modeling process. M indicates use in model development, T in model testing, and V in model verification. The column titled "Model" (in red) indicates the predicted value of the dependent variable.

BBBCNN.4TH

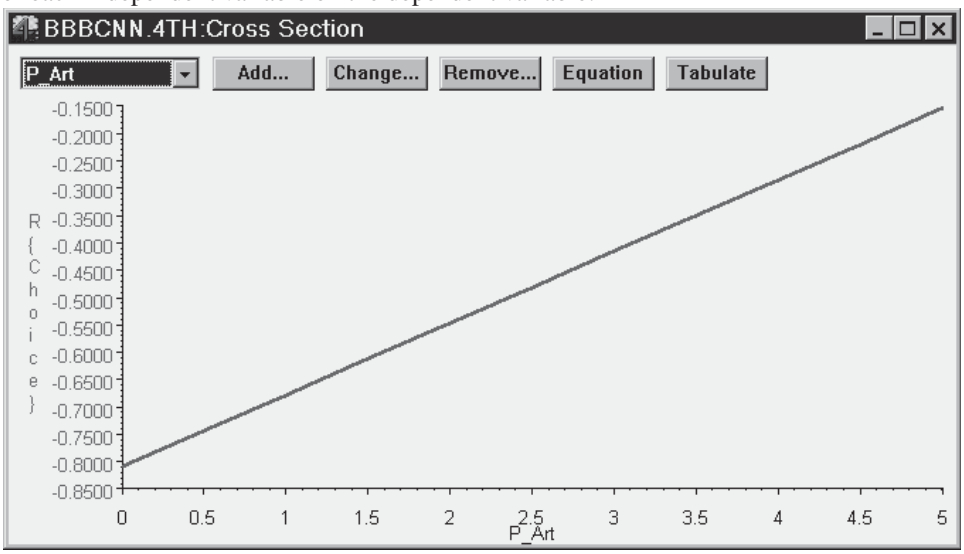
Undo | Redo | Recall | Model of Choice

15:07	123	123	Model	123	123	123	123	123	123
M{Choice}	Actual	Use	Model	0	0	0	0	0	0
1Rx1C	Choice	RU	M{Choice}	Gender	Amt_purch	Last_purc	First_purc	Frequency	P_Child
1	1.00	RU	0.19	1.00	113.00	8.00	1.00	8.00	0.00
2	1.00	T	0.71	1.00	418.00	6.00	11.00	66.00	0.00
3	1.00	M	0.39	1.00	336.00	18.00	6.00	32.00	2.00
4	1.00	M	0.28	1.00	180.00	16.00	5.00	42.00	2.00
5	1.00	M	0.81	0.00	320.00	2.00	3.00	18.00	0.00
6	1.00	T	0.36	1.00	268.00	4.00	1.00	4.00	0.00
7	1.00	M	0.60	1.00	198.00	2.00	12.00	62.00	2.00
8	1.00	M	0.39	0.00	280.00	6.00	2.00	12.00	0.00
9	1.00	M	0.60	1.00	393.00	12.00	11.00	50.00	3.00
10	1.00	M	0.39	1.00	138.00	10.00	7.00	38.00	2.00
11	1.00	T	0.47	0.00	142.00	14.00	2.00	20.00	0.00
12	1.00	M	0.21	0.00	251.00	10.00	2.00	18.00	0.00
13	1.00	M	0.30	0.00	130.00	14.00	2.00	18.00	1.00
14	1.00	M	0.48	1.00	308.00	6.00	1.00	6.00	0.00
15	1.00	M	0.38	1.00	251.00	10.00	2.00	18.00	1.00

To interpret the model, use the **Analysis** menu. Select **Statistics** to see details of model fit. From the **Statistics Summary** dialog box, click **Full Report** to get additional details on the results.



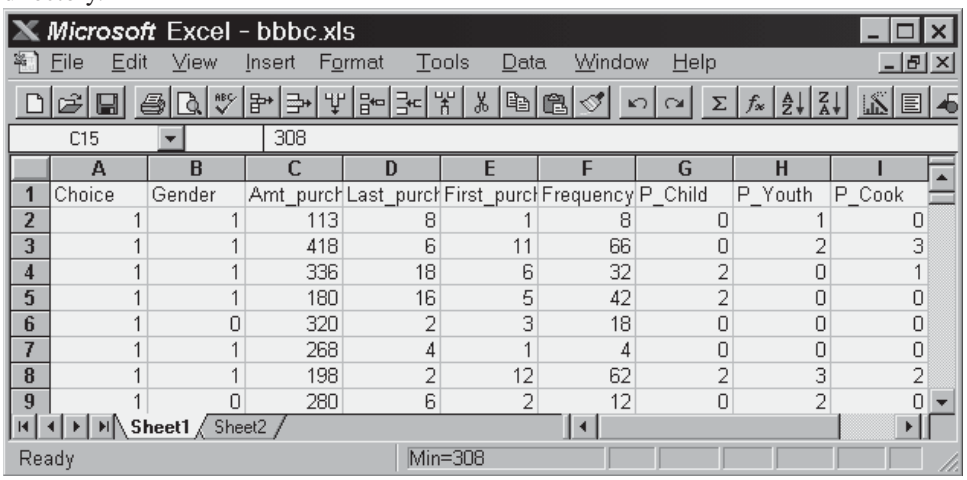
From the **Analysis** menu, select **Cross-section** to view graphs that depict the impact of each independent variable on the dependent variable.



Use **Chart** and **Scenario (**also under **Analysis** menu) to get further details on model results.

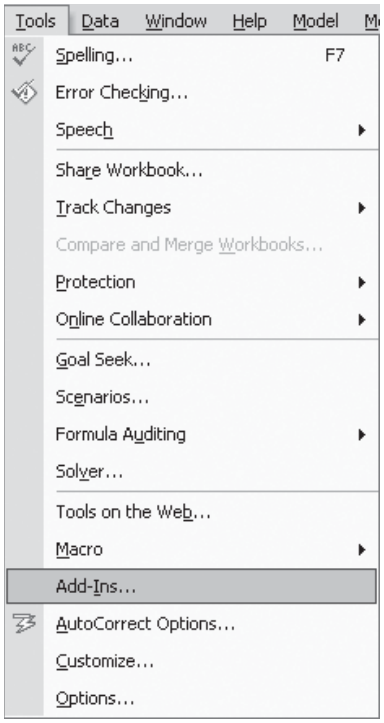
REGRESSION IN EXCEL

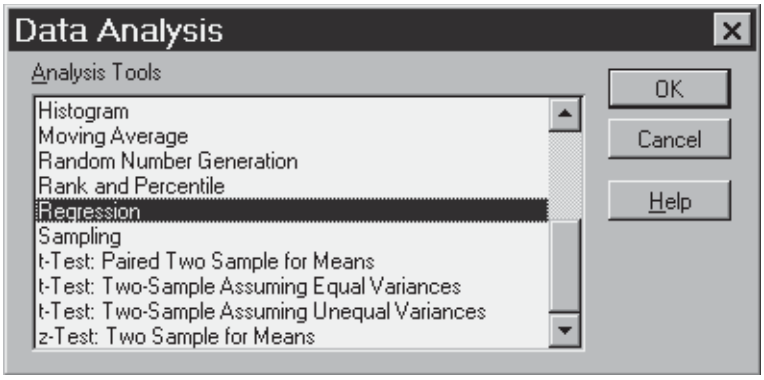
As a point of comparison for the Neural Network model, run the BBBC data using ordinary least squares regression. Open BBBC.XLS file, typically located in the .../mktgeng/data directory.



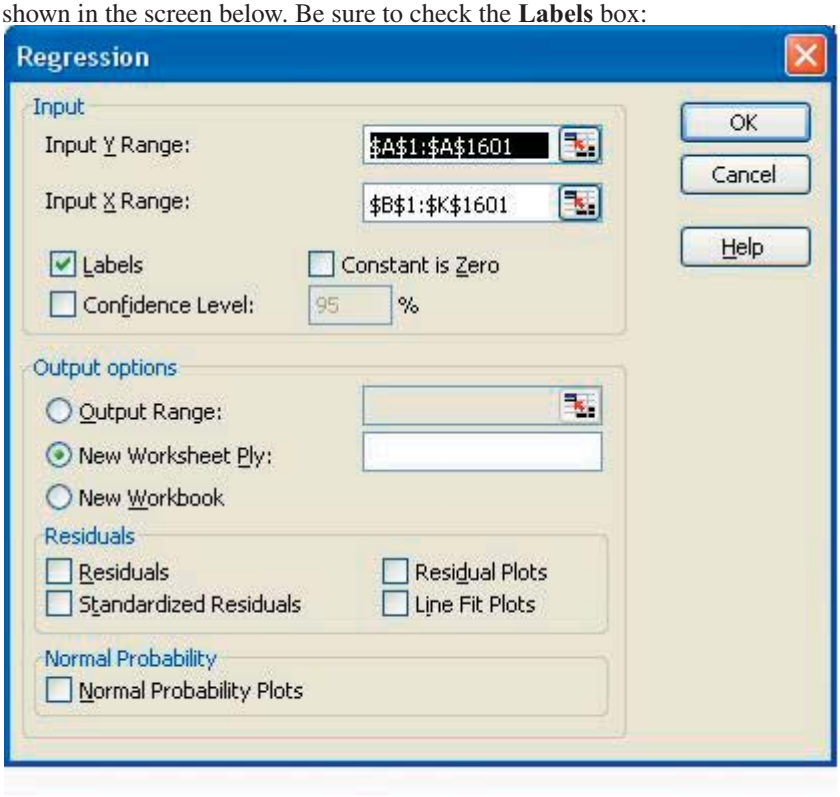
	A	B	C	D	E	F	G	H	I
1	Choice	Gender	Amt_purch	Last_purch	First_purch	Frequency	P_Child	P_Youth	P_Cook
2	1	1	113	8	1	8	0	1	0
3	1	1	418	6	11	66	0	2	3
4	1	1	336	18	6	32	2	0	1
5	1	1	180	16	5	42	2	0	0
6	1	0	320	2	3	18	0	0	0
7	1	1	268	4	1	4	0	0	0
8	1	1	198	2	12	62	2	3	2
9	1	0	280	6	2	12	0	2	0

To start the regression-analysis tool, open the **Tools** menu, select **Data Analysis**, and then **Regression**.





You are now all set to conduct regression analysis. Specify the regression model, as shown in the screen below. Be sure to check the **Labels** box:



After the model runs, you should get regression results shown in the screen below: