Introduction to Excel Workshop

Empirical Reasoning Center

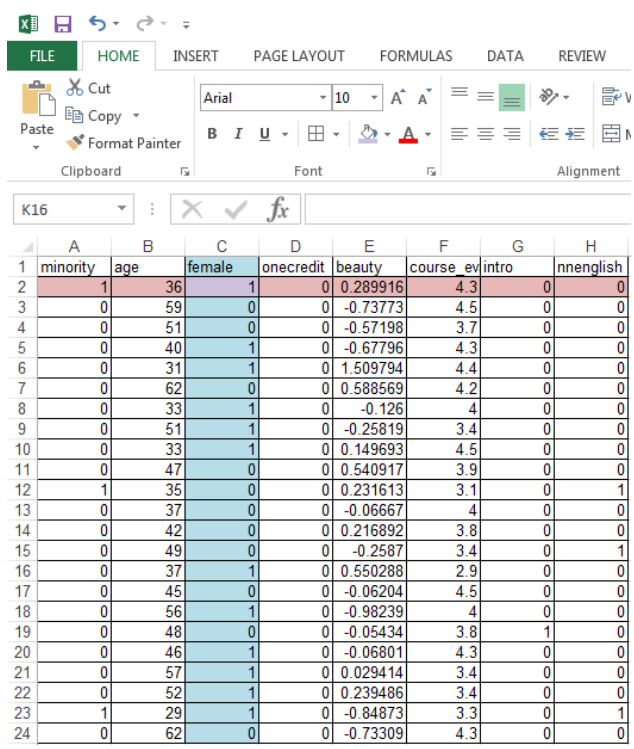
2017

1 Important Terminology

1. Rows are identified by numbers.

2. Columns are identified by letters.

3. Cells are identified by the row-column combination.



	A	B	C	D	E	F	G	H
1	minority	age	female	onecredit	beauty	course_ev	intro	nnenglish
2	1	36	1	0	0.289916	4.3	0	0
3	0	59	0	0	-0.73773	4.5	0	0
4	0	51	0	0	-0.57198	3.7	0	0
5	0	40	1	0	-0.67796	4.3	0	0
6	0	31	1	0	1.509794	4.4	0	0
7	0	62	0	0	0.588569	4.2	0	0
8	0	33	1	0	-0.126	4	0	0
9	0	51	1	0	-0.25819	3.4	0	0
10	0	33	1	0	0.149693	4.5	0	0
11	0	47	0	0	0.540917	3.9	0	0
12	1	35	0	0	0.231613	3.1	0	1
13	0	37	0	0	-0.06667	4	0	0
14	0	42	0	0	0.216892	3.8	0	0
15	0	49	0	0	-0.2587	3.4	0	1
16	0	37	1	0	0.550288	2.9	0	0
17	0	45	0	0	-0.06204	4.5	0	0
18	0	56	1	0	-0.98239	4	0	0
19	0	48	0	0	-0.05434	3.8	1	0
20	0	46	1	0	-0.06801	4.3	0	0
21	0	57	1	0	0.029414	3.4	0	0
22	0	52	1	0	0.239486	3.4	0	0
23	1	29	1	0	-0.84873	3.3	0	1
24	0	62	0	0	-0.73309	4.3	0	0

Figure 1: Reference Terminology

In the figure above, A2:H2 is highlighted in red; C1:C24 is highlighted in blue; and, C2 is highlighted in purple.

2 Calculations

1. Functions always begin with '='
2. You can write the function yourself, or
3. You could refer to a preset function

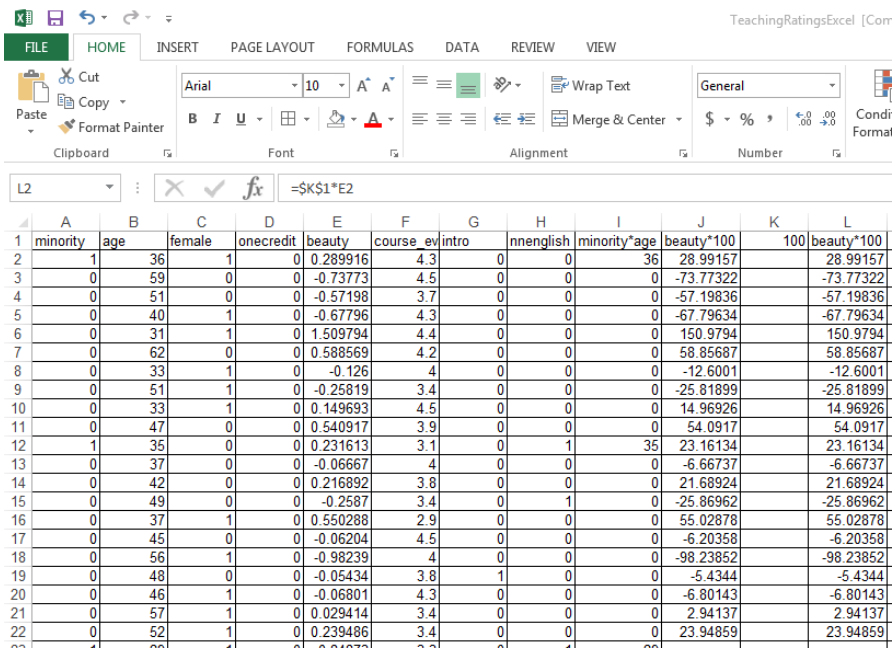
2.1 Example I: A×B

Multiply age by minority. Start by labeling the column where this product will go. In cell I1 type 'minority*age'. Then in the next cell you will type the function. In cell I2 type '=A2*B2' and press enter. Now you want this function to apply to the whole column. To do this, select cell I2 then place your cursor over the bottom right corner of cell I2; you will see the cursor become a small black cross. Click and drag down the whole column (until the data ends). If you select any cell in that column, you should see the formula but the row numbers should refer to that row's data.

2.2 Example II: A×100

Multiply beauty by 100. There are 2 methods: the first uses relative referencing, and the second uses absolute referencing. For the first method, you are going to label column J 'beauty*100' then use the formula '=100*E2' in cell J2 and apply it to the whole column. This is using relative referencing because the formula updates the cell it refers to as you go row by row.

The second method uses absolute referencing. In cell K1, type 100, and title the L column 'beauty*100'. For this method we want to refer to the number 100 as cell K1 for every row. To do this we are going to change how we reference that cell. In cell L2 type the formula '=$K$1*E2'. When you apply this formula to the rest of the column, the dollar signs tell Excel only to refer to cell K1 and not to update the row as you move down the column.



L2 =$K$1*E2

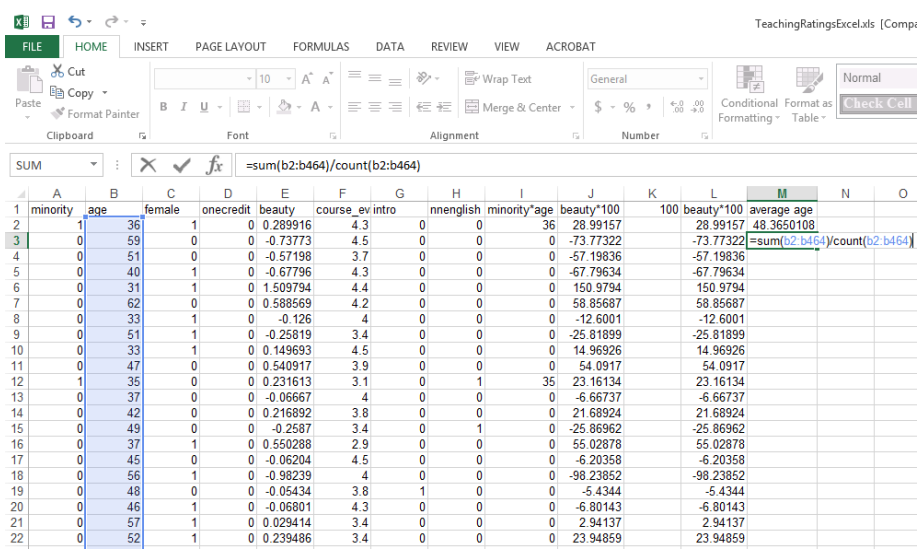
	A	B	C	D	E	F	G	H	I	J	K	L
1	minority	age	female	onecredit	beauty	course_ev	intro	nnenglish	minority*age	beauty*100	100	beauty*100
2	1	36	1	0	0.289916	4.3	0	0	36	28.99157		28.99157
3	0	59	0	0	-0.73773	4.5	0	0	0	-73.77322		-73.77322
4	0	51	0	0	-0.57198	3.7	0	0	0	-57.19836		-57.19836
5	0	40	1	0	-0.67796	4.3	0	0	0	-67.79634		-67.79634
6	0	31	1	0	1.509794	4.4	0	0	0	150.9794		150.9794
7	0	62	0	0	0.588569	4.2	0	0	0	58.85687		58.85687
8	0	33	1	0	-0.126	4	0	0	0	-12.6001		-12.6001
9	0	51	1	0	-0.25819	3.4	0	0	0	-25.81899		-25.81899
10	0	33	1	0	0.149693	4.5	0	0	0	14.96926		14.96926
11	0	47	0	0	0.540917	3.9	0	0	0	54.0917		54.0917
12	1	35	0	0	0.231613	3.1	0	1	35	23.16134		23.16134
13	0	37	0	0	-0.06667	4	0	0	0	-6.66737		-6.66737
14	0	42	0	0	0.216892	3.8	0	0	0	21.68924		21.68924
15	0	49	0	0	-0.2587	3.4	0	1	0	-25.86962		-25.86962
16	0	37	1	0	0.550288	2.9	0	0	0	55.02878		55.02878
17	0	45	0	0	-0.06204	4.5	0	0	0	-6.20358		-6.20358
18	0	56	1	0	-0.98239	4	0	0	0	-98.23852		-98.23852
19	0	48	0	0	-0.05434	3.8	1	0	0	-5.4344		-5.4344
20	0	46	1	0	-0.06801	4.3	0	0	0	-6.80143		-6.80143
21	0	57	1	0	0.029414	3.4	0	0	0	2.94137		2.94137
22	0	52	1	0	0.239486	3.4	0	0	0	23.94859		23.94859

Figure 2: Ex II) Beauty × 100

3 Functions

Calculate the average age. To do this you use the average formula that Excel has preset. In cell M1 type the label 'average age'. In cell M2 you will calculate the average age by entering the formula '=average(B2:B464)'.

You can also use more than one function in a single cell. Calculate the average age using the sum and count functions. In cell M3 enter the formula '=sum(B2:B464)/count(B2:B464)'.



SUM =sum(b2:b464)/count(b2:b464)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	minority	age	female	onecredit	beauty	course_ev	intro	nnenglish	minority*age	beauty*100	100	beauty*100	average age		
2	1	36	1	0	0.289916	4.3	0	0	36	28.99157		28.99157	48.3650108		
3	0	59	0	0	-0.73773	4.5	0	0	0	-73.77322		-73.77322	=sum(b2:b464)/count(b2:b464)		
4	0	51	0	0	-0.57198	3.7	0	0	0	-57.19836		-57.19836			
5	0	40	1	0	-0.67796	4.3	0	0	0	-67.79634		-67.79634			
6	0	31	1	0	1.509794	4.4	0	0	0	150.9794		150.9794			
7	0	62	0	0	0.588569	4.2	0	0	0	58.85687		58.85687			
8	0	33	1	0	-0.126	4	0	0	0	-12.6001		-12.6001			
9	0	51	1	0	-0.25819	3.4	0	0	0	-25.81899		-25.81899			
10	0	33	1	0	0.149693	4.5	0	0	0	14.96926		14.96926			
11	0	47	0	0	0.540917	3.9	0	0	0	54.0917		54.0917			
12	1	35	0	0	0.231613	3.1	0	1	35	23.16134		23.16134			
13	0	37	0	0	-0.06667	4	0	0	0	-6.66737		-6.66737			
14	0	42	0	0	0.216892	3.8	0	0	0	21.68924		21.68924			
15	0	49	0	0	-0.2587	3.4	0	1	0	-25.86962		-25.86962			
16	0	37	1	0	0.550288	2.9	0	0	0	55.02878		55.02878			
17	0	45	0	0	-0.06204	4.5	0	0	0	-6.20358		-6.20358			
18	0	56	1	0	-0.98239	4	0	0	0	-98.23852		-98.23852			
19	0	48	0	0	-0.05434	3.8	1	0	0	-5.4344		-5.4344			
20	0	46	1	0	-0.06801	4.3	0	0	0	-6.80143		-6.80143			
21	0	57	1	0	0.029414	3.4	0	0	0	2.94137		2.94137			
22	0	52	1	0	0.239486	3.4	0	0	0	23.94859		23.94859			

Figure 3: Calculate Average Age

If you think Excel may have the function you want to use you can go to the 'Formulas' tab and select

'Insert Function'. The functions are organized categorically in the function library (to the right of the 'Insert Function' box).

4 Inserting and Editing a Scatter Plot

Scatter plots are useful for visualizing the relationship between two variables. Plot the relationship between beauty and course evaluations. Begin by selecting the data (columns E and F). Then go to the 'Insert' tab, within the 'Charts' group select the 'Scatter' drop-down box, and select the first option (a series of unconnected points).

The initial chart should look like Figure 4.

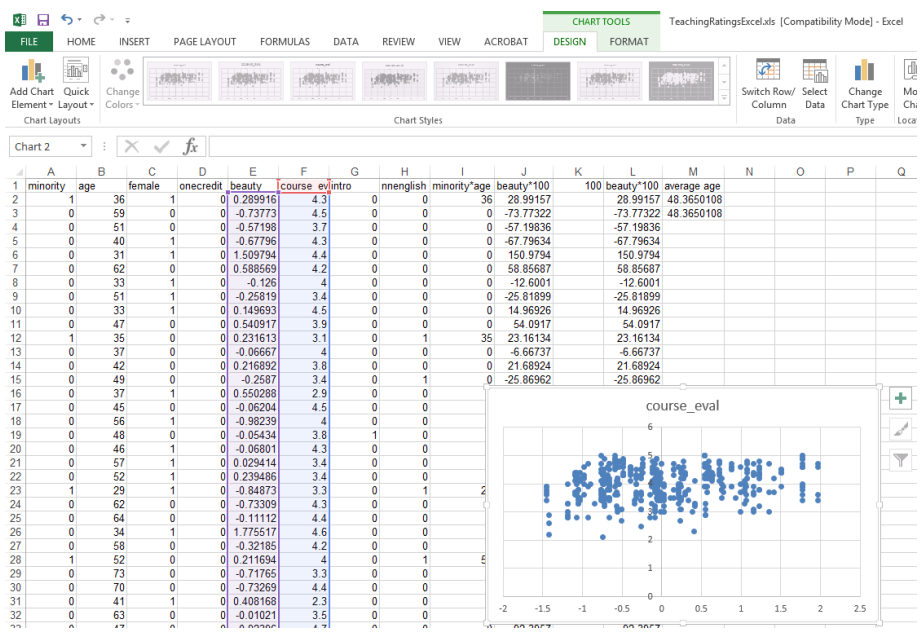


Figure 4: Initial Scatter Plot

Edit the series by right clicking on the chart, choosing 'Select data', and selecting 'Edit'.

For a chart to be effective, the information needs to be very clear. Formatting the chart will make a significant difference. Important elements include:

- Chart and Axes Titles

To add chart and axes titles, select the chart and then the green plus sign in the top right corner of the chart. Select 'Chart Title' and 'Axes Titles'. You can edit the title in the text boxes that appear. Change the chart title to 'Beauty & Course Evaluations'. Label the X-axis 'Beauty' and the Y-axis 'Course Evaluations'. To delete a title, select a text box and press Backspace.

- Axis limits

Select the green plus sign, select 'Axes', select the arrow to the right of 'Axes', and select 'More Options' to edit formatting. To change the numbering on the axes, select 'Axis Options', then select the bar chart icon, and select the 'Axis Options' drop down. You will have to edit each axis. To edit the X-axis select the X-axis on the chart. Make the Y-axis to go from zero to ten instead of zero to six. Also change the major unit to two.

- Gridlines

To insert or delete gridlines, select the green plus, select or unselect 'Gridlines', and select 'More Options' to edit the formatting. Delete the gridlines.

- Chart size

Select the chart and move your mouse over one of the square boxes on each corner and in the center of the borders. When the cursor looks like a double arrow then click and drag the borders of the chart to the preferred size. Solid lines will show the new size. Make the chart larger.

- Legend

To insert or delete a legend, select the green plus, select or unselect 'legend'. To change the placement of the legend, select the arrow to the right of 'legend' and choose among the options. Delete the legend.

- Data Markers

To format the data markers, select the chart, right click on a data marker, and select 'Format Data Series'. To change the color of the markers, select the paint can icon, select the 'Marker' option, and edit the 'Fill' and 'Border' options. Select solid fill and a new color. Change the fill of all of the markers to green with a darker green border.

4.1 Trendlines

Linear trendlines are useful for showing linear relationships between variables. To add a trendline to the chart select the green plus sign, select 'Trendline', select the arrow to the right, and select 'Linear'. To format the trendline and display the equation for the trendline and the R-squared value, select 'More Options', and select the bar chart icon. At the bottom of the box are three check boxes; select the bottom two. You can move the equation on the chart by clicking and dragging the text box. Display the equation and R^2 in the bottom right hand corner.

To format the trendline, select the paint can icon. Change the color to black and the 'Dash type' to a solid line.

Your scatter plot should look like Figure 5.

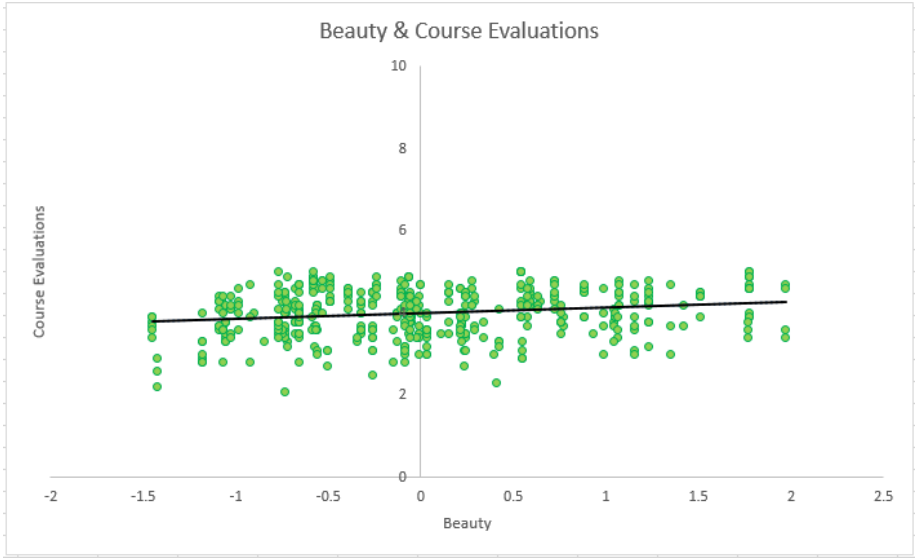


Figure 5: Final Scatter Plot

4.2 Error Bars

Error bars are useful for showing variation in the data. Select the scatter plot. Select the green plus sign, select 'Error bars,' select the arrow to the right of 'Error bars,' and select 'More options' to format the error bars. To set or change the calculation select the bar chart icon. Under error amount you can choose between a fixed value, fixed percentage, the standard deviation, standard error, or use your own calculations (custom). Be cautious about Excel's calculation of the standard error. It may not be the calculation you need.

You'll need to format the horizontal and vertical error bars separately. To change from editing the horizontal bars to the vertical bars (and vice versa), select the 'Error Bar Options' drop down. At the bottom of that menu you should see 2 options, one for the series' 'X Error Bars' and one for the 'Y Error Bars.' Clicking between those two options changes which error bars are edited.

Add error bars with the standard error in both the horizontal and vertical directions.

5 Significant Figures

Making sure your data represents the appropriate significant figures is important. You can format the numbers in your spreadsheet under the 'Home' tab in the ribbon, in the Number section. The icons look like this:

The icon on the left increases the number of decimal places (of selected cells) by 1 every time you click it. The icon on the right decreases the number of decimal places (of selected cells) by 1 every time you click it. To change an entire column, you first have to highlight the entire column. Change the significant figures for the beauty variable to 1 decimal place. Then change it to 4 decimal places.

Please be aware that when you change the significant figures in your data, this can affect your chart. You can fix the number of significant figures in your chart by selecting the green plus sign, select 'Axes', select the arrow to the right of 'Axes', and select 'More Options' to edit formatting. To change the number formatting on the axis, select 'Axis Options', then select the bar chart icon. Then towards the bottom of the list, select the 'Number' drop down. First you will want to make sure the 'Category' field says 'Number.' Then you can change the number of decimal places. You will have to edit each axis. The Options window will look like this:

Change the significant figure on both the X- and Y-axes to 2 decimal places.

6 Copying & Pasting into Word

Select the chart; copy (CTRL + C) & paste it (CTRL + V) into a Word document. The formatting will change, and the chart will remain editable. To prevent this, paste the chart as a picture. Right click on the Word document and under 'Paste Options' select 'Picture'.

7 Data Analysis ToolPak I

The Data Analysis ToolPak is used for statistical analyses. It is an Add-In that you will need to install. To install it, select 'File', then 'Options', then select the 'Add-Ins' tab. Under 'Inactive Application Add-ins' select 'Analysis ToolPak'. For the 'Manage' dropdown menu select 'Excel Add-ins' and select 'Go'. Select 'Analysis ToolPak' and select 'OK'.

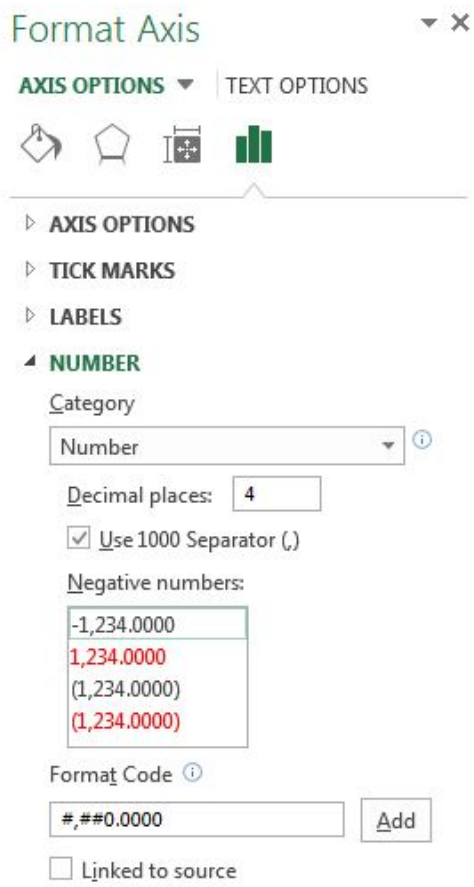


Figure 6: Changing the Significant Figures

If you select the 'Data' tab, the 'Analysis' group will now show a 'Data Analysis' feature.

Older versions of Excel do not have the Data Analysis Toolpak. However, all campus computers can install the Data Analysis Toolpak. If you have to use your own computer, use the Linest function to run regressions. The Linest function is described in the last section.

7.1 Regressions

Linear regression analysis uses the "least squares" method to fit a line to the set of observations. You can analyze how a dependent variable is affected by one or more independent variables.

To run a univariate regression, select the 'Data Analysis' feature, select 'Regression', and then select 'OK'. For 'Input Y Range' select the dependent variable. For this example select the course eval column (F1:F424). For 'Input X Range' select the independent variable. For this example select the beauty column

(E1:E424). Select 'Labels' and 'New Worksheet Ply'. Label the new worksheet 'Univariate'. Then select 'Line Fit Plots' then 'OK'. The Univariate Worksheet will contain the output and a chart.

Notable values in the output are

- Slope and standard error
- Intercept and standard error
- R-squared
- F-statistic
- Regression sum of squares
- Degrees of freedom
- Residual sum of squares.

8 Linest Function (Optional)

Linest is used to estimate the parameters of a line of best fit (the trendline). It also provides the standard errors for the parameter of the line (which the graph does not).

Linest is an array function, so it acts a bit differently than the other functions we covered earlier.

1. Select 10 cells in a 5 by 2 formation (5 rows, 2 columns)
2. Start the function with '='
3. Type Linest and follow the prompts in Excel
    - Select your known y's
    - Select your known x's
    - For [const] type 1 (because you do NOT want to force the intercept to be 0)
    - For [stats] also type 1 (because you want the additional statistics)
4. Press Ctrl + Shift + Enter (NOT just Enter!!)

After pressing Ctrl + Shift + Enter all 10 cells should have numbers in them. The output is labeled here:

	A	B	C	D	E
1	slope	0.133001	3.998272	intercept	
2	error	0.032178	0.025349	error	
3	R-squared	0.035736	0.545452	standard deviation of y	
4	F-statistic	17.08473	461	degrees of freedom	
5	Regressior	5.083008	137.1556	residual sum of squares	

Figure 7: Linest Output