ME 303 Lab #0
Exercise #2: MS Excel Regression Analysis Exercise

OBJECTIVE

This computer exercise is intended to show you how to use the MS Excel spreadsheet software to perform statistical analysis of two variables, i.e., regression analysis.

KEY LABORATORY LEARNINGS

- How to perform professional statistical analysis using MS Excel
- How to create suitable statistical plots as would be put in a professional scientific report or presentation

EQUIPMENT

- Computer with MS Excel

SUMMARY

Experimental data is often too large and cumbersome to be analyzed by hand. Therefore, computer software can assist the experimentalist in analyzing data. Furthermore, computer software can assist in creating professional looking plots suitable for placement in scientific reports or presentations.

In this lab session a dataset will be analyzed statistically and suitable professional plots created. Despite the fact that this data set is “small” and can be analyzed by hand, students are required to use Excel functions to create the appropriate data cells. This will prepare students for future labs in which much larger datasets (where it is unreasonable to analyze by hand) will be obtained.

KEYWORDS

Population, sample, random, probability, probability density, array, class interval, class boundary, class mark, class frequency, frequency distribution, relative frequency distribution, histogram, central tendency, (arithmetic) mean, median, mode, variance, standard deviation, accuracy, precision, confidence interval, accuracy, bias, normal (Gaussian) distribution, variance of the means, regression.

PROCEDURE

Part 1: Linear Regression Analysis

1. Open MS Excel and enter the data shown below in the text into two columns in the spreadsheet. This data is from a static load cell in which an inputted load in Newtons results in an outputted change in the voltage of the load cell.

Load (N)	Load Cell Reading (V)
1.0	2.4
2.0	3.3
3.0	6.5
4.0	8.2
5.0	9.5

2. Be sure that the Analysis ToolPak is available. In order to do this, left click on the "Tools" heading. Choose the "Add-Ins" option. Check the box next to Analysis ToolPak and click OK.

3. Perform a linear (i.e., a first order polynomial) regression analysis on the data.

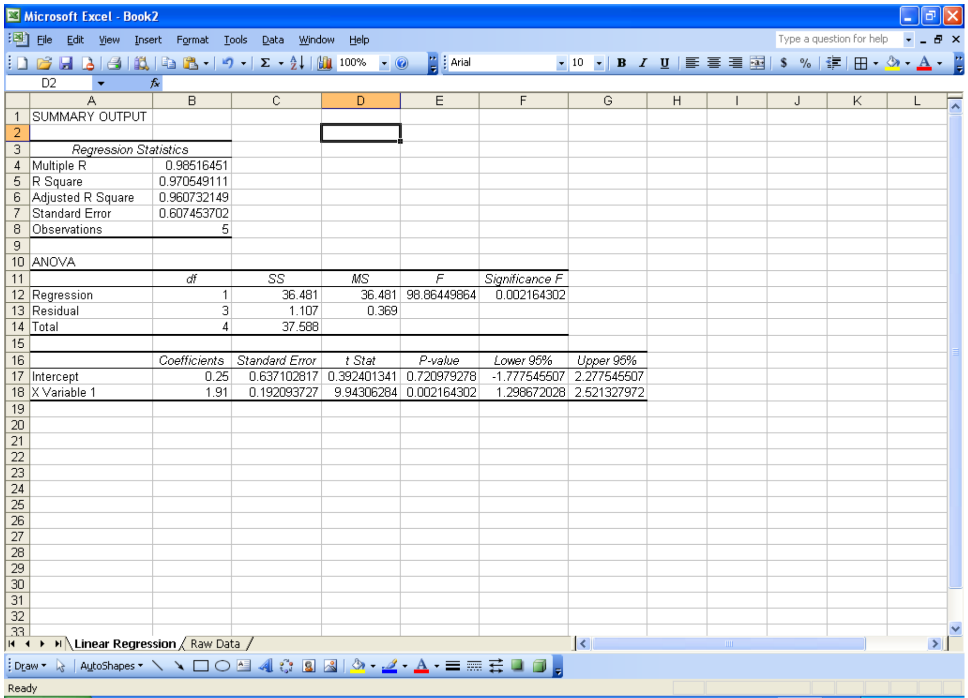
    a. Select "Tools" > "Data Analysis" > "Regression" from the top menu and click OK.

    b. For the dependent variable or output ("Input Y range:") select the load cell reading data.

    c. For the independent variable or input ("Input X range:") select the load data.

    d. Output the results into a new worksheet ply titled "Linear Regression Analysis"

    e. When completed, click OK, and the following regression analysis should result:



	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0.98516451					
5	R Square	0.970549111					
6	Adjusted R Square	0.960732149					
7	Standard Error	0.607453702					
8	Observations	5					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	1	36.481	36.481	98.86449864	0.002164302	
13	Residual	3	1.107	0.369			
14	Total	4	37.588				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	0.25	0.637102817	0.392401341	0.720979278	-1.777545507	2.277545507
18	X Variable 1	1.91	0.192093727	9.94306284	0.002164302	1.298672028	2.521327972

4. In a new worksheet, create a dataset of the linear regression results. This can be done be creating a load column at even intervals across the range of the data. For example, you may create a load column with values 0.0, 0.1, 0.2, …, 5.8, 5.9, 6.0. Then, using the results from the linear regression analysis, create a predicted voltage output using the following formula:

$$V = C_0 + C_1 L$$

   where V is the predicted voltage output, C_0 is the coefficient for the intercept, C_1 is the coefficient of the first variable (i.e., the load), and L is the load from the column that you created.

5. In addition to the predictions outputted by the regression analysis, please make note of the key statistical parameters such as the upper and lower 95% confidence bounds of the predictions, the R^2 value, and the adjusted R^2 value (note that the adjusted R^2 value includes a confidence "penalty" for the number of parameters that are input into a given regression analysis). The technical definition of these R^2 values is that, for an R^2 value of 0.95, 95% of the variation in the data is explained through the given model.

Part 2: Polynomial Regression Analysis

1. In your raw data table, create additional columns for $(\text{Load})^2$, $(\text{Load})^3$, $(\text{Load})^4$, $(\text{Load})^5$, and $(\text{Load})^6$.

Load (N)	Load^2	Load^3	Load^4	Load^5	Load^6	Load Cell Reading (V)
1.0	1.0	1.0	1.0	1.0	1.0	2.4
2.0	4.0	8.0	16.0	32.0	64.0	3.3
3.0	9.0	27.0	81.0	243.0	729.0	6.5
4.0	16.0	64.0	256.0	1024.0	4096.0	8.2
5.0	25.0	125.0	625.0	3125.0	15625.0	9.5

2. Perform a sixth order polynomial regression analysis on the data.

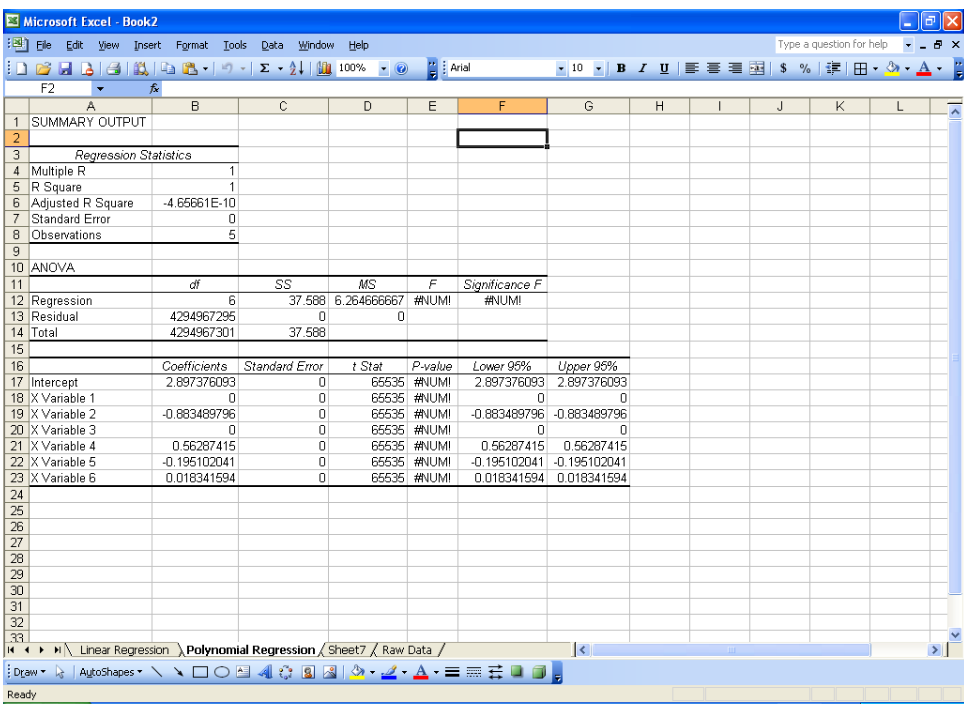
    a. Select "Tools" > "Data Analysis" > "Regression" from the top menu and click OK.

    b. For the dependent variable or output ("Input Y range:") select the load cell reading data.

    c. For the independent variable or input ("Input X range:") select 2D data set for all of the load columns (i.e., select the square with diagonal corners of 1.0 and 15625.0).

    d. Output the results into a new worksheet ply titled "Polynomial Regression Analysis"

    e. When completed, click OK, and the following regression analysis should result:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	-4.65661E-10
Standard Error	0
Observations	5

ANOVA

	df	SS	MS	F	Significance F
Regression	6	37.588	6.264666667	#NUM!	#NUM!
Residual	4294967295	0	0		
Total	4294967301	37.588			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2.897376093	0	65535	#NUM!	2.897376093	2.897376093
X Variable 1	0	0	65535	#NUM!	0	0
X Variable 2	-0.883489796	0	65535	#NUM!	-0.883489796	-0.883489796
X Variable 3	0	0	65535	#NUM!	0	0
X Variable 4	0.56287415	0	65535	#NUM!	0.56287415	0.56287415
X Variable 5	-0.195102041	0	65535	#NUM!	-0.195102041	-0.195102041
X Variable 6	0.018341594	0	65535	#NUM!	0.018341594	0.018341594

f. Note the confidence intervals, R^2 values, and adjusted R^2 values provided here.

3. In the same worksheet created in Part 1, Step 4, use the results from the sixth order polynomial regression analysis to create a predicted voltage output using the following formula:

$$V = C_0 + C_1L + C_2L^2 + C_3L^3 + C_4L^4 + C_5L^5 + C_6L^6$$

where V is the predicted voltage output, C_0 is the coefficient for the intercept, C_i is the coefficient of the i-th variable, and L is the load from the column that you created.

Part 3: Creating the plot

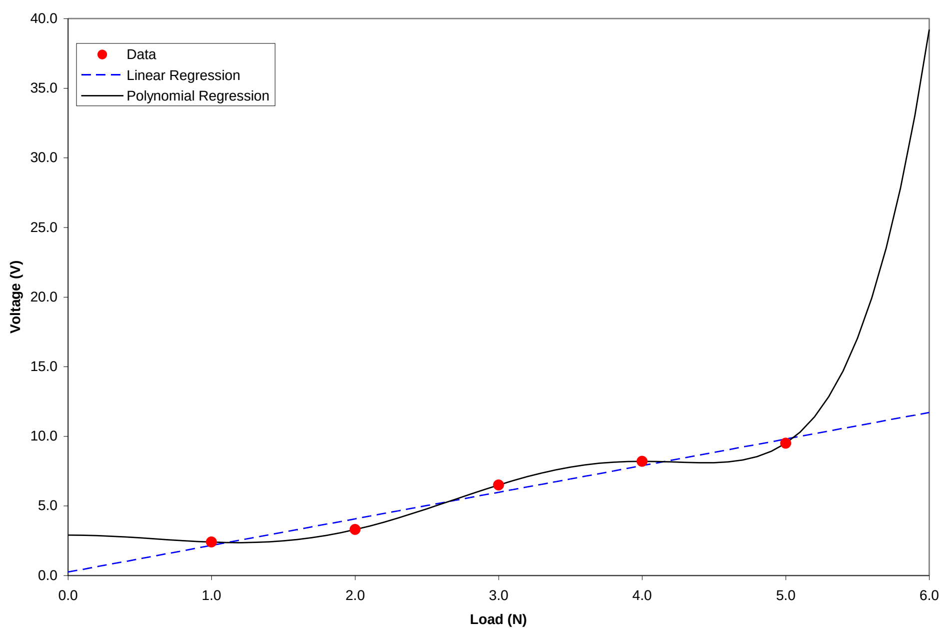
1. Create a scatter plot containing the raw data as well as the two models.

2. Format your plot professionally.

    a. Choose appropriate colors and sizes for the data that are easily distinguishable

    b. Raw data should be presented as individual points

    c. Predictions or models should be presented as lines without individual points



3. Once you are completed with this exercise, save the file.