ECOR 1010 – INTRODUCTION TO ENGINEERING

BRYCE MARINO KUHN

ASSIGNMENT #6

Assignment Title: Bivariate Data: Linear Least-Squares Regression

TO Marking TA:

Name: Michael Gutland

Email: Michael.Gutland@carleton.ca

Lab Section: L3

Room: 4301CB – Carleton University

FROM:

Email:brycemkuhn@cmail.carleton.ca

Student Number: 100999222

Number of Figures and Tables (Including handwritten ones): 10

Last Date and Time of Revision: @1:35PM

INTRODUCTION

The purpose of this lab was to use bivariate statistics to test relationships between two coupled variables. Using Microsoft Excel and knowledge of correlation coefficient, total sum of squares, sum of the squares, coefficient of determination and other equations taught in class, the purpose of the lab was to create a linear, third degree and sixth degree equation with the data and predict new values with these equations.

MATERIALS AND METHODS

The majority of calculations were done within the Microsoft Excel program itself. The coefficient of correlation and coefficient of determination were calculated separately as well as calculated in the table. The data was inserted into the program and from there the mean was calculated with =AVERAGE, the standard deviation was calculated with =STDEV and the sums were calculated with =SUM. Every column in the data could be derived from the data itself using various equations for example the equation for the first value of $(x-\bar{x})*(y-\bar{y})$ was [= ((B3)-(B12))*((C3)-(C12))].

RESULTS

The results show that the regression value for linear is significantly smaller than for the sixth and third degree polynomials. The 0.33 wt% values predicted were 387.8117 (linear), 380.5740238 (third degree polynomial) and 358.1164 (sixth degree polynomial). The 1.0 wt% values predicted were 603.21 (linear), 490.2 (third degree polynomial) and -172.9 (sixth degree polynomial).

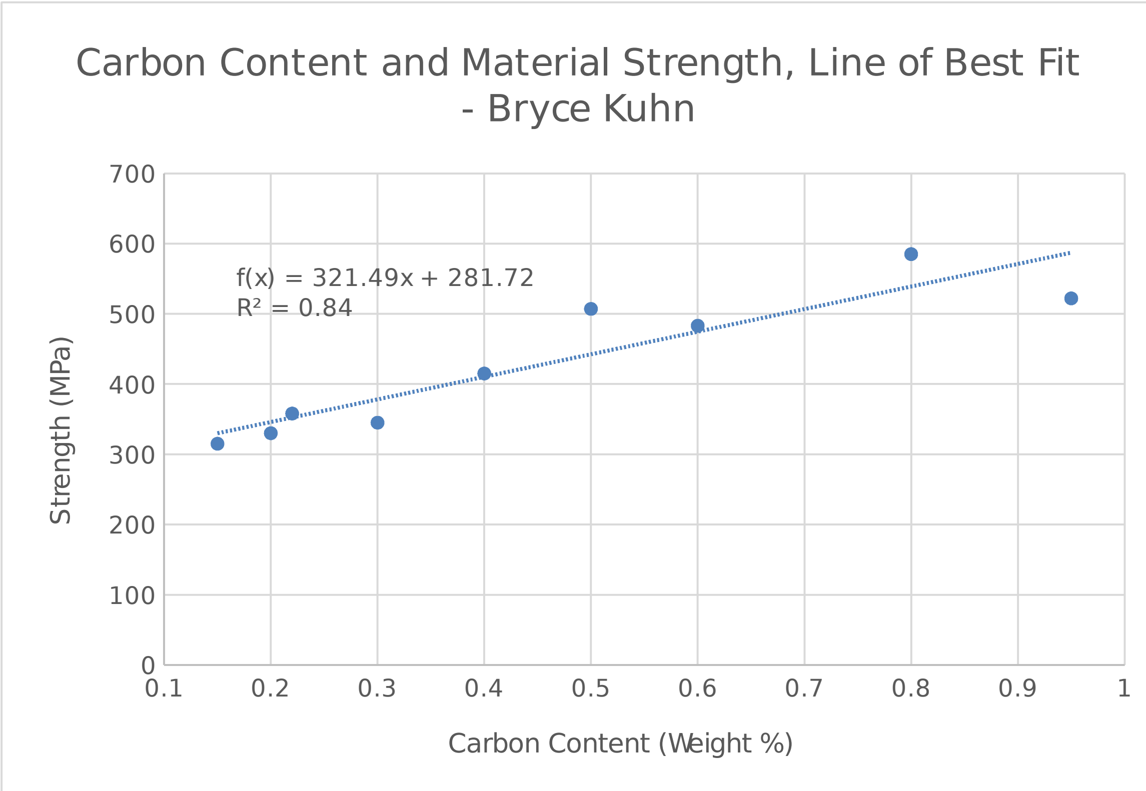
DISCUSSION

Both correlation coefficients calculated were basically the same value. The one calculated manually was more precise, giving more decimal values than the one calculated by Microsoft Excel. The results that were determined from the data given showed that the most acceptable graph would either be the linear or the third degree polynomial. The reason the sixth degree polynomial cannot be considered useful is because when it was used to predict the 1.0 wt% it gave a negative value. Between the linear and the third degree polynomial it makes more sense that that the linear is more accurate as it predicts a stronger 1.0 wt% than the third degree polynomial. It also follows the trend that as the wt% increases the strength does as well.

CONCLUSIONS

Overall it appears that the safest way to approach the strength estimation of 0.33 wt% and a 1.0 wt% carbon content would be to use the linear equation that was solved for. The sixth degree polynomial does predict the closest values to the ones already determined, however it gives a negative value from the 1.0 wt%. This is because the third and sixth degree polynomials are on the descending side of a global maximum. The linear or the third degree polynomial could both be used to predict the values, the only difference is that at 1.0 wt% the linear predicts a higher values while the third degree polynomial predicts a lower value. Due to the fact that the real value is unknown it is difficult to make a prediction. Due to the fact that the sixth degree polynomial predicts a negative value 1.0 wt% it is most wise to use the linear for anything greater than 0.8 wt% and to use the sixth degree for anything smaller than 0.8 wt%.

APPENDICES- FIGURES AND TABLES



Steel Grade	Carbon Content (Weight %)	Strength (MPa)
Observation	x	y
1015	0.15	315
1020	0.2	330
1022	0.22	358
1030	0.3	345
1040	0.4	415
1050	0.5	507
1060	0.6	483
1080	0.8	585
1095	0.95	522

All graphs were derived from this set of data

Kuhn, Bryce

Steel Grade	Carbon Content (Weight %)	Strength (MPa)	(x-x̄)	(y-ȳ)	(x-x̄)*(y-ȳ)	(x-x̄)^2	(y-ȳ)^2	y = 321.49x + 281.72	(ŷ-ȳ)^2	(y-ŷ)^2
Observation	x	y								
1015	0.15	315	-0.307778	-113.888889	35.05	0.0947272	12970.67901	329.9435	9790.189982	223.3081923
1020	0.2	330	-0.257778	-98.8888889	25.49	0.0664494	9779.012346	346.018	6867.584225	256.576324
1022	0.22	358	-0.237778	-70.8888889	16.86	0.0565383	5025.234568	352.4478	5843.240071	30.82692484
1030	0.3	345	-0.157778	-83.8888889	13.24	0.0248938	7037.345679	378.167	2572.710012	1100.049889
1040	0.4	415	-0.057778	-13.8888889	0.80	0.0033383	192.9012346	410.316	344.9522017	21.939856
1050	0.5	507	0.042222	78.11111111	3.30	0.0017827	6101.345679	442.465	184.3107929	4164.766225
1060	0.6	483	0.142222	54.11111111	7.70	0.0202272	2928.012346	474.614	2090.785786	70.324996
1080	0.8	585	0.342222	156.1111111	53.42	0.117116	24370.67901	538.912	12105.08498	2124.103744
1095	0.95	522	0.492222	93.11111111	45.83	0.2422827	8669.679012	587.1355	25041.98993	4242.63336
Mean of	0.457777778	428.8888889								
Standard deviation of	0.28003472	98.1547814								
Sum of			0	-2.2737E-13	201.69	0.6273556	77074.88889	3860.0188	64840.84798	12234.52951
$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y}$	0.917204661									

Number of data	9	Slope	321.49
$\sum_{i=1}^{n}(x_i-\bar{x})^2$	0.627356	Intercept	281.72
$\sum_{i=1}^{n}(y_i-\bar{y})^2$	77074.88889	Correlation Coefficient	0.917204661
$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$	201.69	Coefficient of Determination	0.841271142
$\bar{x}$	0.457777778		
$\bar{y}$	428.8888889		
s_x	0.28003472		
s_y	98.1547814		
TSS	77074.8889		
SSR	64840.84798		
SSE	12234.52951		

Correlation Coefficient

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 201.687778$

$n = 9$

$S_x = 0.280034$

$S_y = 98.1547814$

$$r_{xy} = \frac{201.68778}{(9-1)(0.280034)(98.1547814)}$$

$r_{xy} = 0.91720703$

Sum of Squares from regression

$SSR = \sum(\hat{y} - \bar{y})^2$

$= 64840.84798$

Total sum of squares

$TSS = \sum(y - \bar{y})^2$

$= 77074.8889$

Coefficient of Determination.

$= \frac{SSR}{TSS}$

$\frac{64840.84798}{77074.8889} = 0.84127142$

	Coefficient of Determination	Correlation Coefficient
Calculations	0.84	1.031855255
Excel	0.841271142	1.031855244
CORREL(B2:B3,C2:C3)=	-1	
RSQ(B2:B3,C2:C3)=	1	

SUMMARY OUTPUT

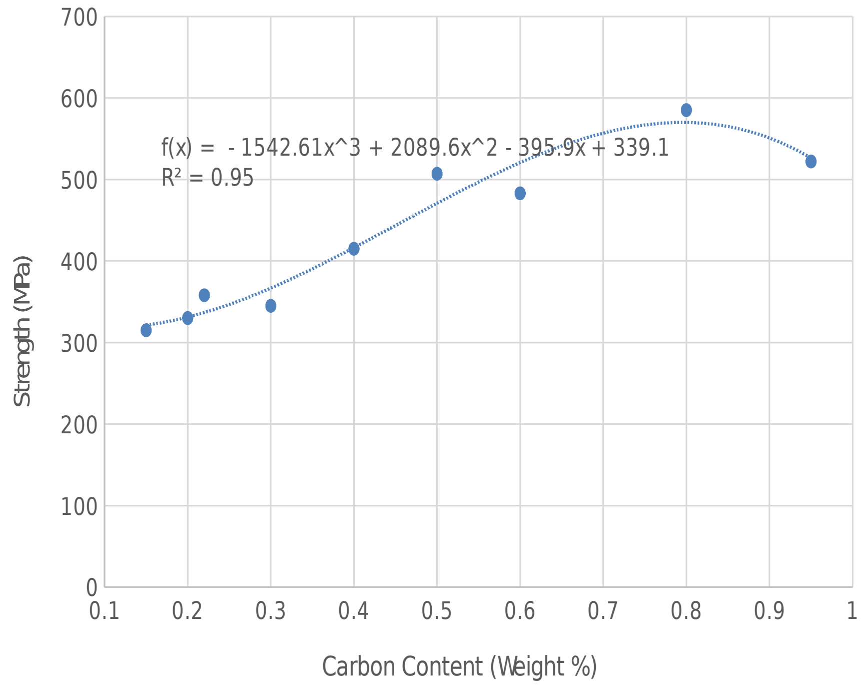
Regression Statistics	
Multiple R	0.917204661
R Square	0.84126439
Adjusted R Square	0.818587875
Standard Error	41.80657753
Observations	9

ANOVA

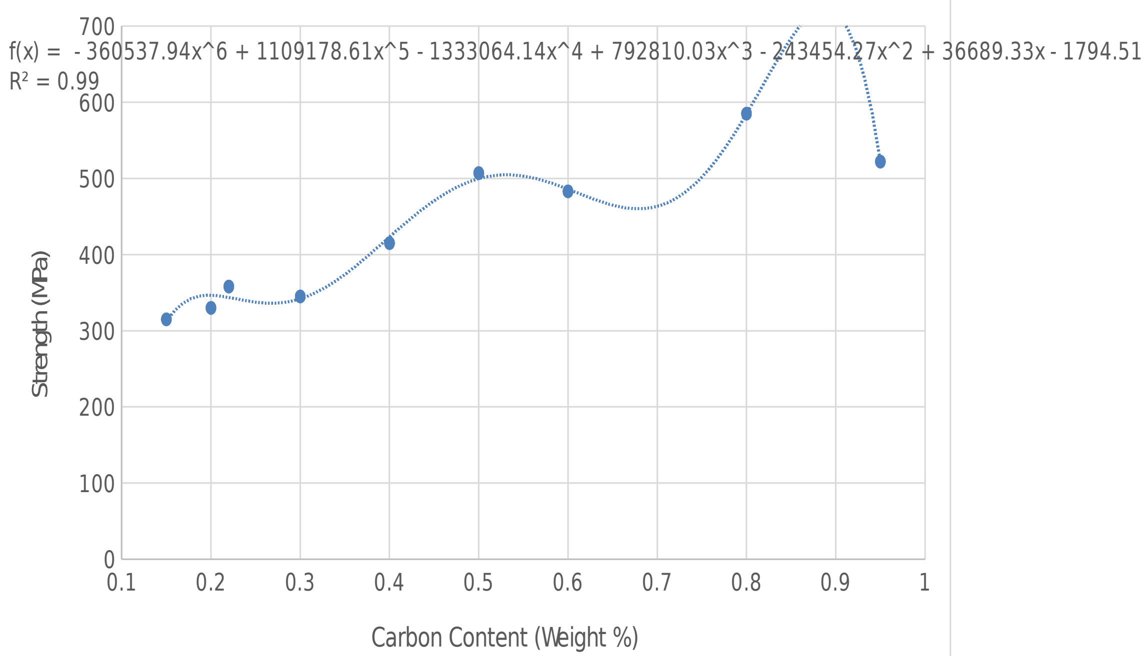
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	64840.35942	64840.36	37.09849	0.000495531
Residual	7	12234.52947	1747.79		
Total	8	77074.88889			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	281.7184655	27.89313265	10.09992	2E-05	215.7616876	347.67524
X Variable 1	321.4887889	52.78223082	6.090853	0.000496	196.6786459	446.29893

Carbon Content and Material Strength, Third-Order Polynomial - Bryce Kuhn



Carbon Content and Material Strength, Sixth-Order Polynomial - Bryce Kuhn



Carbon Content (Weight %)	Strength (MPa)	y = 321.49x + 281.72	$y = -1542.6x^3 + 2089.6x^2 - 395.9x + 339.1$	y = -360,537.95x6 + 1,109,178.62x5 - 1,333,064.16x4 + 792,810.04x3 - 243,454.27x2 + 36,689.33x - 1,794.51
0.15	315	329.9435	321.524725	312.1600779
0.20	330	346.018	331.1632	346.6255936
0.22	358	352.4478	336.7130352	343.7665441
0.30	345	378.167	366.7438	341.9279651
0.40	415	410.316	416.3496	423.1644896
0.50	507	442.465	470.725	499.7589063
0.60	483	474.614	520.6144	485.8752
0.80	585	538.912	569.9128	584.6715808
0.95	522	587.1355	526.272325	522.04203
R^2 value for the regression		0.84	0.9488	0.99
Interpolation: 0.33		387.8117	380.5740238	358.1163936
Extrapolation: 1.00		603.21	490.2	-172.9