COMM 581 - Assignment #4
Multiple Linear Regression – Quadratic Transformation

Name: _____ **Total: 20 marks**
Due date: Wednesday Oct. 7, 2015 (11pm)

Background:
You are trying to determine the temperature at which to conduct a manufacturing process that generates the maximum yield. You have measured the yield at different temperatures.

The quadratic model equation for the population will be: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$
where x_1 is temperature and x_2 is temperature squared.

Include all graphs discussed in your assignment. Show calculations for test statistics, confidence intervals and measures of goodness of fit based on sums of squares.

1. Graph the relationship between yield and temperature. Does the relationship look linear? Discuss the form and strength of the relationship, and identify any outliers. (**1 mark**)

2. Fit a quadratic model (including two explanatory variables: temperature and the square of temperature) and create the residual plot for the model. Using the residual plot, assess the assumptions of linearity and equal variance. Divide the residual plot into 3-4 segments to assess these assumptions. State any concerns you have and their consequences. (**2 marks**)

3. Check the normality assumption using a histogram, normality plot, and normality tests. Adjust the bin width for the histogram to be informative. State any concerns you have and their consequences. (**2 marks**)

4. Discuss whether or not you think the assumption of independence is met. What could result in a violation of this assumption? State any concerns you have and their consequences. (**2 marks**)

5. Write the ANOVA table for this model showing each component of variation (variable 1, variable 2, error, total), its associated df, sum of squares and mean squares. (**3 marks**)

6. Test the significance of the regression. Show your calculations using the values from the ANOVA table. (**1.5 marks**)

7. Can you simplify the model? Test if the co-efficient associated with temperature is equal to 0 (and could therefore be removed from the model). Use a t-test for this co-efficient. What does it mean if this co-efficient is not equal to 0? (**1.5 marks**)

8. State the estimates of the co-efficients (b_0, b_1, b_2) and calculate their confidence intervals using the standard errors from the outputs. Do the confidence intervals for b_1 and b_2 overlap zero? What does it mean if they do overlap zero? (**2.5 marks**)

9. Calculate the R^2 value (co-efficient of multiple determination) and standard error of the estimate (root MSE) from the sums of squares. Show your calculations. Check if these are in agreement with the outputs from R. (**1 mark**)

10. What temperature gives the maximum yield based on the model? (**1 mark**)

11. What is the predicted yield at the optimal temperature? Calculate the prediction interval associated with this yield. (**1.5 marks**)

12. Create a plot of the original data (yield vs. temperature) including the line of best fit and lines for the prediction intervals. (**1 mark**)