Massey University

228.371 Assignment 1 **ID 00001**

Name: ____________________

Student ID: ____________________

Signature: ____________________

You need to complete this assignment using a word processor and statistical software. The choice of word processor is up to you, but you must use R for doing the statistical work required. You may use any other software you like for calculations and formatting of your work.

Once you have completed your assignment, you will need to create a pdf file for uploading to the class Stream site. A drop box will be used to receive the assignments. Please ensure you have fully submitted your assignment. Many students make the mistake of leaving this file in a draft form.

Please ensure that the front page of your assignment has nothing but the paper details, your full name and ID number.

Solutions will be made available via the drop box on Stream once the due date has passed.

1. Suppose that a standard gas turbine has, on average, a heat rate of 9750kJ/kWh. Perform a t-test to see if the mean heat rate for the turbines in your data file exceeds 9750kJ/kWh. What do you conclude about your set of gas turbines compared to the average gas turbine?

   My answer for this question is:

   The mark for this question is:

2. The temperature is measured at two points for each turbine. Compare these two measurements and make a statement (backed with evidence) about the mean reduction in temperature from the inlet to the exhaust points of the turbines.

   My answer for this question is:

   The mark for this question is:

3. Do turbines with the advanced engine type have higher power than those with traditional engines? What assumptions are required if your findings are to be generalised to all advanced and traditional turbines?

   My answer for this question is:

   The mark for this question is:

4. Determine if the reduction in temperature from the inlet to the exhaust is different for the three types of turbine. (This should include a graphical representation as well as a model.)

   My answer for this question is:

   The mark for this question is:

5. Conduct a hypothesis test to ascertain if there is potentially a linear relationship between air flow and power.

   My answer for this question is:

   The mark for this question is:

6. Use both a graphical method and a formal hypothesis test to make a judgement about the normality of the power of the turbines in your sample.

   My answer for this question is:

   The mark for this question is:

7. Using the heat rate as the response variable, construct suitable graphs to identify which of the numeric variables might prove useful as predictors in a simple linear regression of the form $E(y) = \beta_0 + \beta_1 x$.

   My answer for this question is:

   The mark for this question is:

8. Use speed as a candidate for the predictor in a simple linear regression model to explain the heat rate of the turbines.

   My answer for this question is:

   The mark for this question is:

9. Choose another variable as the predictor in a simple linear regression model. Is this model better or worse than using speed as the sole predictor of heat rate?

   My answer for this question is:

   The mark for this question is:

10. Now fit a model that has both speed and the other variable you chose. Determine if the interaction of the two variables is required. You must therefore choose between the models $E(y) = \beta_0 + \beta_1 Speed + \beta_2 x_2$ and $E(y) = \beta_0 + \beta_1 Speed + \beta_2 x_2 + \beta_3 Speed \times x_2$.

    My answer for this question is:

    The mark for this question is:

11. Create a multiple regression model that uses all numeric variables in your data set as predictors of the heat rate. Do not include any interactions or polynomial terms.

    My answer for this question is:

    The mark for this question is:

12. Give a practical interpretation of your estimates of the β's from this model.

    My answer for this question is:

    The mark for this question is:

13. Consider the standard deviation of the residuals from the models created thus far. Use this as a means of describing the value of the last model compared to the model that included only two predictors.

    My answer for this question is:

    The mark for this question is:

14. Interpret the R^2 value from the last model.

    My answer for this question is:

    The mark for this question is:

15. Is this model useful for predicting heat rate? Justify your answer using a hypothesis test for the utility of the entire model at a significance level of $\alpha = 0.01$.

    My answer for this question is:

    The mark for this question is:

16. Investigate the leverages of your multiple regression model. Are there any points that are having undue influence on the model?

    My answer for this question is:

    The mark for this question is:

17. Now create a reduced model by appropriately removing terms from your multiple regression model.

    My answer for this question is:

    The mark for this question is:

18. Compare this reduced model with the complete multiple regression model that included all predictors, using a single hypothesis test.

    My answer for this question is:

    The mark for this question is:

19. Generate the residual analysis for your reduced model. Is there anything to be concerned about?

    My answer for this question is:

    The mark for this question is:

20. Obtain the Cook's distances and variance inflation factors for this model. Is there anything to worry about?

    My answer for this question is:

    The mark for this question is:

21. For any single predictor, demonstrate your knowledge about how to build a polynomial regression model to explain heat rate. That is, a model of the form $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 \ldots$ etc.

    My answer for this question is:

    The mark for this question is:

22. Consider a model for heat rate of a gas turbine that is a function of cycle speed and cycle pressure ratio. Fit a second-order model using just these two variables. Construct a graph that compares the cycle pressure ratio with the heat rate that is predicted under this model for a cycle speed of 4900 rpm.

    My answer for this question is:

    The mark for this question is:

23. Up to this point we have ignored the engine type variable in your data. Create a graph that plots the residuals from the last model against the Speed, which indicates the differences between the three engine types.

    My answer for this question is:

    The mark for this question is:

24. Fit the model that adds the engine type to the second-order model using only cycle pressure ratio and cycle speed. Does allowing for differences in the mean response for each engine type improve the ability of the model to predict heat rate?

    My answer for this question is:

    The mark for this question is:

25. The second-order model using cycle pressure ratio and cycle speed is called a response surface model. Do you believe there is a different response surface for each engine type?

    My answer for this question is:

    The mark for this question is:

Your total mark is: ______________________