**Linear Models - Question 1 (5 Marks)**

Consider the variables ***X*** and ***Y***. We wish to fit a linear model to the data, where ***X*** is the independent variable and ***Y***  is the dependent variable.

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | ***X*** | X090 |
| 2 | ***Y*** | X119 |

1. Fit a linear model as instructed above. State the Regression Equation [1 Mark ]
2. Construct and examine the diagnostic plots for this fitted model.

Sketch the “*Residual Versus Fitted Values*” Plot(i.e. Diagnostic Plot 1 ).

Comment on the pattern of covariates in this plot. [2 Marks]

Sketch the *Normal Probability Plot for Residuals* ( i.e. Diagnostic Plot 2).

Comment on the pattern of covariates in this plot. [1 Marks]

1. For any points specifically identified in either of the diagnostic plots – state the Cook’s Distance value. [1 Mark]

Hint: use “ X <-round(X,5)” to make the answers easier to read.

**Two Sample Tests for Centrality - Question 2 (6 Marks)**

Consider the Samples ***J*** and ***K***.

We wish to know if they can be assumed to have the same (or at least similar) centrality (i.e. mean and median)

|  |  |  |
| --- | --- | --- |
|  | Sample | Data Set |
| 1 | ***J*** | X107 |
| 2 | ***K*** | X108 |

Use a 5% significance level in all of these procedures.

1. Carry out a test for equality of population means, given that equality of population means is **assumed. [ 2 Marks]**
2. Carry out a test for equality of population means, given that equality of population means is NOT **assumed. [ 2 Marks]**
3. Carry out a test for equal population medians (**N.B. Ignore the output warnings**) **[ 2 Marks ]**

**Analysis of Variance - Question 3 (7 Marks)**

Suppose you want to determine whether the brand of laundry detergent used and the temperature affects the amount of dirt removed from your laundry. You are also interested in determining if there is an interaction between the two variables.

To this end, you buy two different brand of detergent (“ ***Super***” and “***Best***”) and choose three different temperature levels (“***cold***”, “***warm***”, and “***hot***”).



|  |
| --- |
| * The Full Data is contained in X075. * The grouping variables for Factor A and Factor B are contained in X055 and X060. |

1. Write out the ANOVA table. *You are not required to add the “Totals” Row.*

[4 Marks]

1. State whether or not each of the effects are statistically significant. [3 Marks]

**Testing Normality - Question 4 (2 Marks)**

The Puromycin data set contains data on the velocity of an enzymatic reaction were obtained by Treloar (1974).

The number of counts per minute of radioactive product from the reaction was measured as a function of substrate concentration in parts per million (ppm) and from these counts the initial rate (or velocity) of the reaction was calculated (counts/min/min).

The experiment was conducted once with the enzyme treated with Puromycin, and once with the enzyme untreated.

We are interested in determining if the data for the ***conc*** variable is normally distributed.

The values for this data set can be found here:

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | ***conc*** | X087 |
|  |  |  |

Answer the following questions.

1. Use a graphical procedure was carried out to determine whether or not the data set ***conc***  is normally distributed. Sketch your output. [1 Mark]
2. Perform a formal hypothesis test for testing the normality of data set ***conc***. State your null and alternative hypothesis. [1 Mark]