**Testing Normality - Question 1 (4 Marks)**

Consider the following datasets ***A*** and ***B.*** The values for each data set are found in the specified data sets.

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | A | X101 |
| 2 | B | X102 |

Answer the following questions.

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

**Testing Equality of Variance - Question 2 (3 Marks)**

Consider the three following data sets ***X, Y*** and ***Z***.

|  |  |
| --- | --- |
| Sample | Data Set |
| X | X110 |
| Y | X114 |
| Z | X115 |

For each of the three possible pairings, state if the assumption of equality of variance is valid. Justify your conclusion.

**ANOVA Procedures - Question 3 (6 marks)**

Specimens of milk from dairies in three different districts are assayed for their concentrations of the radioactive isotope Strontium-90.

The results, in picocuries per litre, are as shown in the table below.

|  |  |  |
| --- | --- | --- |
| District | Observations | Dataset |
| A | 6.7 6.1 6.8 8.0 7.5 | X111 |
| B | 7.5 10.3 11.6 10.9 6.9 9.2 8.8 | X112 |
| C | 10.3 9.8 12.9 11.9 | X113 |
|  |  |  |
| Combined |  | X100 |
|  | Grouping Variable | X109 |

1. Write out the ANOVA table. *You are not required to add the “Totals” Row.* [4 Marks]
2. Carry out an analysis of variance of these data, conducting your significance test at the 5% level. [2 Marks]

**Diagnostic for Linear Models - Question 4 (7 Marks)**

Consider the variables ***X1***, ***X2*** and ***Y***. We wish to fit a linear model to the data, where ***Y*** is the dependent variable and ***X1*** and ***X2*** are the independent variables.

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | ***Y*** | X078 |
| 2 | ***X1*** | X077 |
| 3 | ***X2*** | X079 |

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. [1 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.

1. State the *Akaike Information Criterion* value for this model. [1 Marks]
2. Compute a robust linear regression model for this data set. *(Hint: Remember to load the MASS package. Also you may use the default weighting method.)*

Write out the regression equation for this fitted model. [2 Marks]