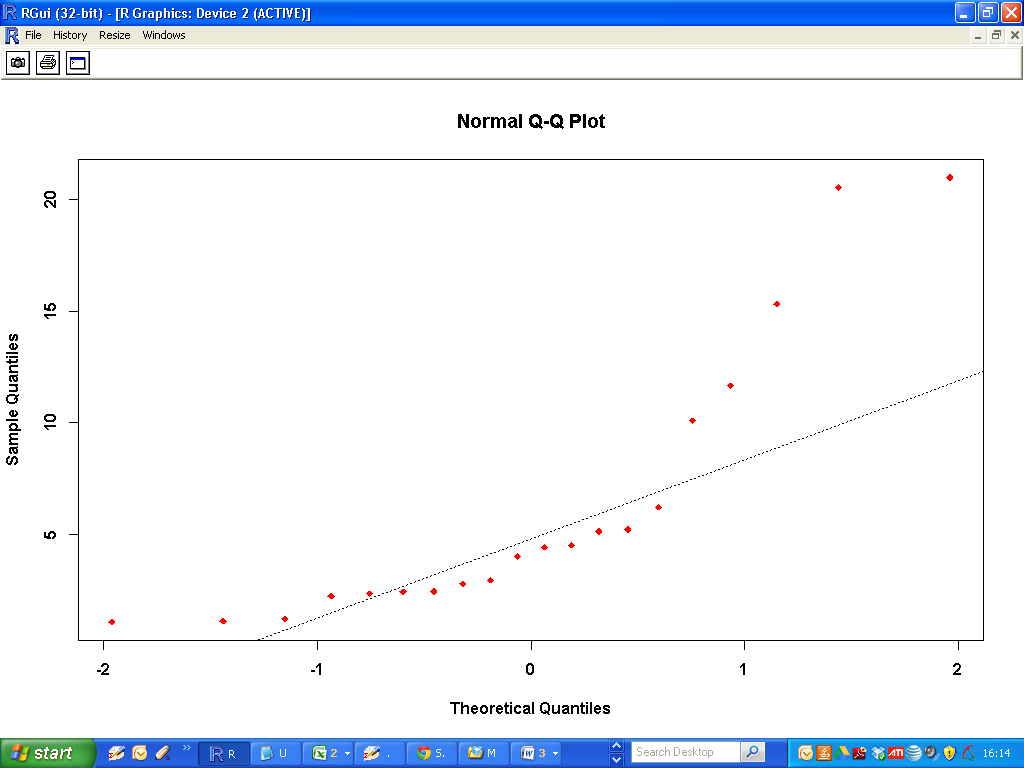
**Testing Normality - Question 1 (6 Marks)**

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

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | X | DAT94 |
| 2 | Y | DAT95 |

The following graphical procedure was carried out to determine whether or not the data set ***X*** is normally distributed. ****

1) Explain how this graphical procedure is used to make such a determination. [2 Marks]

2) What is your conclusion for this data set [1 Mark]

3) Perform a formal hypothesis test for testing the normality of data set ***Y***. State your null and alternative hypothesis.[1 Mark]

4) What is your conclusion for this procedure? Justify your answer with reference to the p-value.[2 Marks]

**Non-Parametric Inference Procedures - Question 2 (8 marks)**

*This question is a continuation of Question 1. You may assume that all the statements and conclusions made in Question 1 apply to Question 2.*

Consider the datasets ***X*** and ***Y*** from Question 1.

|  |  |  |
| --- | --- | --- |
|  | Variable | Data Set |
| 1 | X | DAT94 |
| 2 | Y | DAT95 |

1. For the data sets X and Y, give a brief explanation as to why non-parametric inference procedures should be used in preference to other types of testing procedures. [2 Mark]
2. Perform an appropriate hypothesis test to determine whether or not the data sets X and Y can be approximated by the same probability distribution.

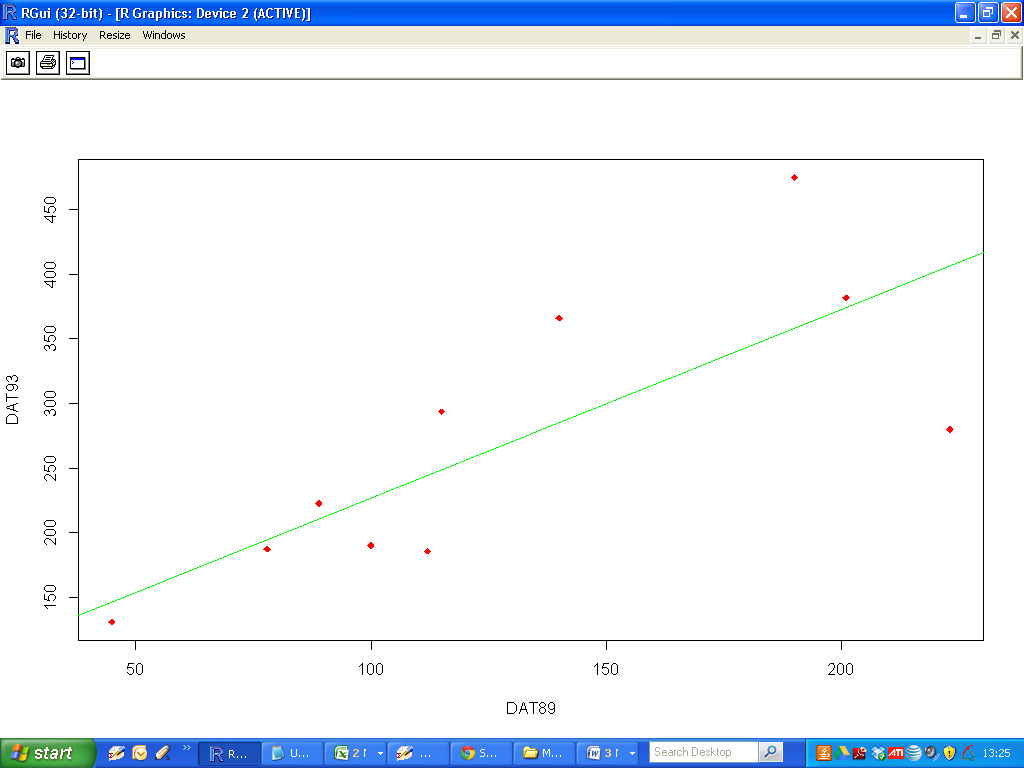
State your null and alternative hypotheses clearly [2 Marks]

1. What is your conclusion from this procedure? Justify your answer with reference to the p-value.[1 Mark]
2. Perform an appropriate hypothesis test to determine whether or not the data sets X and Y have the same median value. State your null and alternative hypotheses clearly [2 Marks]
3. What is your conclusion from this procedure? Justify your answer with reference to the p-value.[1 Mark]

**Diagnostic for Linear Models - Question 3 (6 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*** | DAT89 |
| 2 | ***Y*** | DAT93 |

****

1. Based on the scatter-plot, presented above, discuss whether or not a simple linear regression model is an appropriate model. Give one reason for your answer. [1 Marks]
2. Fit a linear model as specified. Construct and examine the diagnostic plots for this fitted model. Sketch the Residual plot ( i.e. *Residual values Versus Fitted Values )*. Comment on the pattern of covariates in this plot. [ 2 Marks]
3. Which two covariates have the highest Cook’s Distance value? [1 Mark]
4. 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]