**Inference Procedures - Question 1 (5 marks)**

A test of a specific blood factor has been devised so that, for adults in the UK and Ireland, the test score is normally distributed with mean 100 and standard deviation 10.

A clinical research organization is carrying out research on the blood factor levels for sufferers of a particular disease. Specifically they are carrying out research on the effects of experimental medication has on the blood factor level.

For a group of 14 volunteer patients the following test scores for were obtained both prior to the medication, and after the medication.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
| Before | 120 | 140 | 112 | 109 | 114 | 116 | 99 | 108 | 109 | 111 | 109 | 131 | 117 | 101 |
| After | 104 | 112 | 110 | 107 | 101 | 103 | 101 | 102 | 103 | 102 | 101 | 120 | 112 | 103 |

(see DAT79 and DAT81)

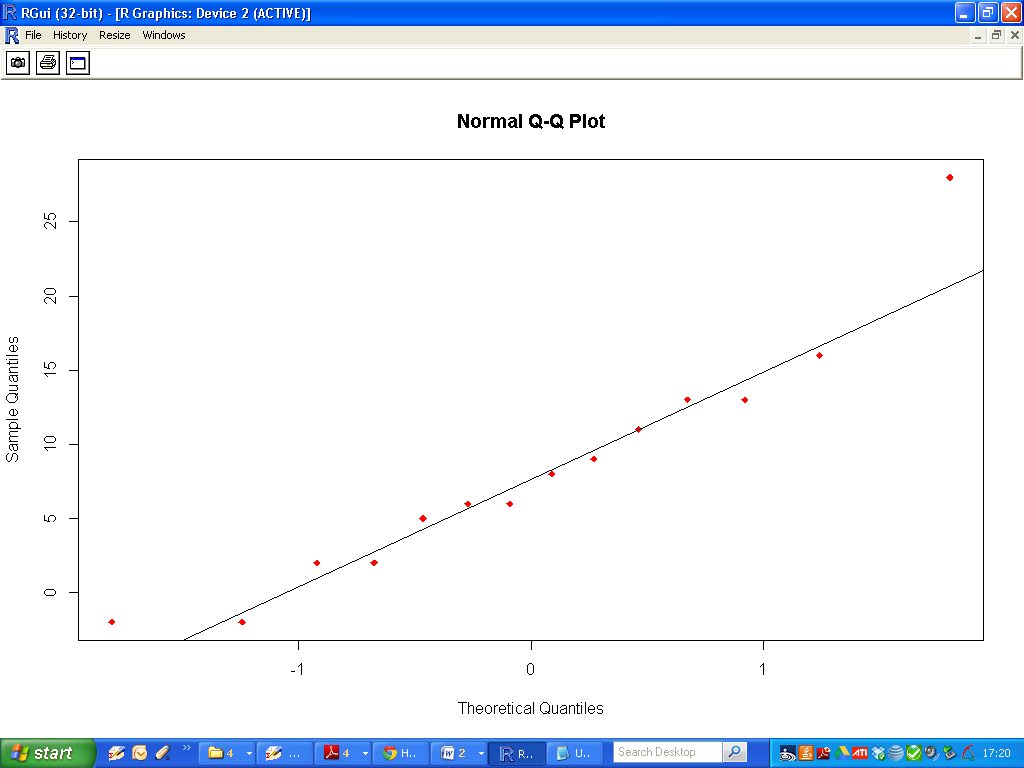
For now , you may assume that the case-wise differences are normally distributed.

The clinical research organization wishes to determine if there is a significant improvement (lessening of the blood factor level) due to the medication.

Perform an appropriate hypothesis test for this hypothesis test, using a significance level of 5%.

1. Formally state both the null hypothesis and alternative hypothesis. [2 Mark]
2. From the R code output, what is the resultant p-value? [1 Mark]
3. What is your conclusion for the hypothesis test, based on this p-value? [2 Marks]

**Testing Normality - Question 2 (6 Marks)**

The following graphical procedure was carried out to determine whether or not the case-wise differences from the previous question are normally distributed. ****1) Explain how this graphical procedure is used to make a determination. [2 Marks]

2) What is your conclusion for the set of case-wise differences from the study described in Question 1. [1 Mark]

3) Perform a formal hypothesis test for testing the normality the case-wise differences (See DAT82). 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]

**Multiple Linear Regression - Question 3 (5 Marks)**

Olive oil is composed, in part, of triacylglycerols (triglycerides or fats).

The major fatty acids in olive oil triacylglycerols are:

* Oleic Acid - a monounsaturated omega-9 fatty acid (DAT72)
* Linoleic Acid -a polyunsaturated omega-6 fatty acid (DAT10)
* Palmitic Acid - a saturated fatty acid (DAT69)
* Stearic Acid - a saturated fatty acid (DAT60)
* Linolenic Acid - a polyunsaturated omega-3 fatty acid (DAT78)

A sample of 40 brands of olive were sampled and given a quality rating (DAT47).

Using the quality rating as a dependent variable, fit a linear model to predict the quality of olive oil using the following set of independent variables.

|  |
| --- |
| Oleic (DAT72), Linoleic (DAT10) , Stearic (DAT60), Linolenic (DAT78) |

1. Write down the regression equations for the fitted model . [2 marks ]
2. Write down the *adjusted R squared value* for this fitted model. [1 Marks]
3. The model would possibly be improved by removing one of the independent variables. Which independent variable, if any, would you remove from the model? Justify your answer. [2 Marks]

**Linear Models - Question 4 (4 Marks)**

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).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Variable | Description | Data Set |
| 1 | Conc | Substrate concentration (ppm) | DAT57 |
| 2 | Rate | Reaction velocity (counts/min/min) | DAT51 |

Fit a linear model to the data, where ***Conc*** is the independent variable and ***Rate*** is the dependent variable

1. Sketch the scatter-plot for this data. [1 Mark]
2. Based on this scatter-plot, discuss whether or not a simple linear regression model is an appropriate model. Give 1 reason for your answer. [1 Marks]
3. Write down the regression equation for this fitted model. [1 Mark]
4. What are the p-values associated with each of the regression coefficients? [1 Mark]