**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.

A study has obtained the following test scores for 15 randomly selected patients suffering from the disease in Edinburgh, Scotland.

***113 115 120 109 105 103 103 99 128 96 125 107 115 131 119***

(see DAT88)

A similar study has obtained the following test scores for 14 randomly selected patients suffering from the disease in Dublin, Ireland.

***120 140 112 109 114 116 99 108 109 111 109 131 117 101***

(see DAT79)

You may assume that both data sets are normally distributed.

The clinical research organization wishes to determine if there is a significant difference between the two groups of patients.

1. Perform a appropriate formal test to determine whether or not both data sets have equal variance. State both the null hypothesis and alternative hypothesis. [2 Marks]
2. From the R code output, what is your conclusion of this test, using the resultant p-value? [2 Marks]
3. Compute a 95% confidence interval for the variance ratio of both data sets. [1 Marks]

**Inference Procedures - Question 2 (6 marks)**

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

The clinical research organization wishes to determine if there is a significant difference between the two groups of patients.

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 for this procedure. [2 Mark]
2. From the R code output, what is the resultant p-value? [1 Mark]
3. What is the 95% confidence interval for the difference in means? [1 Mark]
4. What is your conclusion for the hypothesis test, based on the p-value only? [2 Marks]

**Multiple Linear Regression - Question 3 (6 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 sets of independent variables.

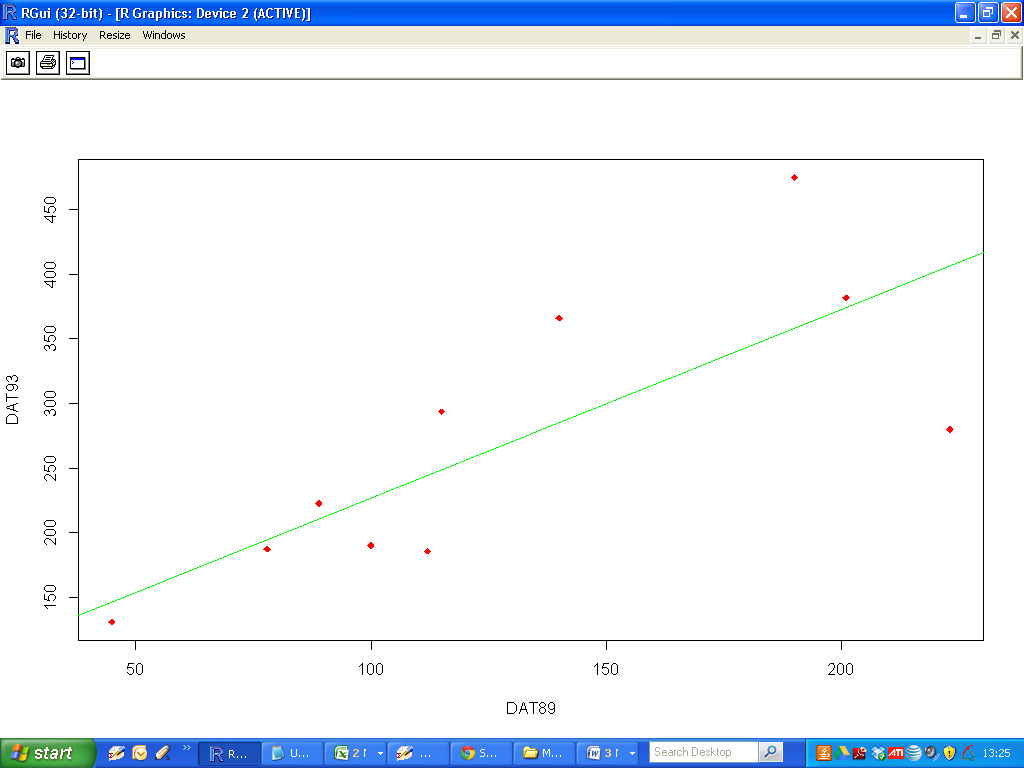
|  |  |
| --- | --- |
| Model | Variables |
| A | Oleic (DAT72), Linoleic (DAT10) , Stearic (DAT60) |
| B | Palmitic (DAT69), Linolenic (DAT78) |

1. Write down the regression equations for both fitted models A and B. [2 marks ]
2. Write down the *Akaike Information Criterion* values for both linear models. [2 Marks]
3. Which of the two models better fits the data? Explain your answer. [2 Marks]

**Linear Models - Question 4 (3 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. Compute an estimate for the correlation coefficient for ***X*** and ***Y.*** [ 1 Mark]
3. Determine a 95% confidence interval for this correlation estimate. [1 Mark].