**Inference Procedures - Question 1 (6 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 12 randomly selected patients suffering from the disease in one area of the UK

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

(see DAT49)

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)

The variance of both data sets are equal.

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. 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]
4. Compute a 95% confidence interval for the difference of means of both data sets. [1 Mark ]

**Multiple Linear Regression - Question 2 (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 3 (8 Marks)**

The*DNase* dataset contains data obtained during development of an ELISA assay for the recombinant protein DNase in rat serum.

* The variable ***conc*** is a numeric vector giving the known concentration of the protein. (The data is contained in data set DAT19)
* The variable ***density*** is a numeric vector giving the measured optical density in the assay. (The data is contained in data set DAT03)

Suppose the clinical scientists constructed a simple linear regression model to predict ***density*** using ***conc*** as the independent variable.

**Questions**

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 two reasons for your answer. [2 Marks]
3. Compute an estimate for the correlation coefficient for ***conc*** and ***density.*** [ 1 Mark]
4. Determine a 95% confidence interval for this correlation estimate. [1 Mark].
5. Write down the regression equation for this fitted model. [1 Mark]
6. What are the p-values associated with each of the regression coefficients? [1 Mark]
7. For the slope estimate, interpret the corresponding p-value. [1 Mark]