|  |  |
| --- | --- |
| ***MA4605 Chemometrics*** | ***November 2nd 2012*** |
| ***Duration 45 minutes*** | ***20 Marks*** |
| ***Group 1*** |  |
| ***Answer All Questions*** |  |

**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 needs to test whether the mean score of sufferers from a particular disease differs from the mean score of the general population on this test.

A study has obtained the following test scores for 12 randomly selected patients suffering from the disease.

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

(see DAT49)

1. Compute a 99% confidence interval for the mean of patient’s blood factor scores.[2 Marks]
2. We wish to determine whether or not the mean score for patients is significantly different from the general population. Using the confidence interval, perform a hypothesis test. State your null and alternative hypothesis clearly. [3 Marks]
3. Perform a hypothesis test for this procedure, using a significance level of 5%. What is the p-value? [1 Mark]

**Linear Models - Question 2 (6 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. Write down the regression equation for this fitted model. [1 Mark]
2. Determine 95% confidence intervals for regression estimates. [2 Marks]
3. What are the p-values associated with each of the regression coefficients? [1 Mark]
4. For the slope estimate, interpret the corresponding p-value.[ 1 Mark]
5. Compute the Akaike Information Criterion (AIC) for this fitted Model [ 1 Mark]

**Multiple Linear Regression and ANOVA - Question 3 (8 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 | Palmitic (DAT69), Linoleic (DAT10), Stearic (DAT60) |
| B | Oleic(DAT72), Stearic (DAT10) |

1. Write down the regression equations for both fitted models A and B. [2 marks ]
2. Write down the adjusted R square values for both linear models. [2 Marks]
3. Which of the two models better fits the data? Explain your answer. [2 Marks]
4. Complete the *Analysis of Variance* table for fitted model B. [ 2 Marks ]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) | (asterisks) |
| DAT72 |  | 483222 |  |  | < 2.2e-16 |  |
| DAT10 | 1 |  | 29173 |  |  |  |
| Residuals |  | 57157 |  |  |  |  |