FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS AND STATISTICS

MID-TERM ASSESSMENT EXAMINATION

MODULE CODE: MA4605 SEMESTER: Autumn 2016

MODULE TITLE: Chemometrics DURATION OF EXAM: 60 minutes

LECTURER: Mr. Kevin O'Brien GRADING SCHEME: 30 Marks

INSTRUCTIONS TO CANDIDATES

- **IMPORTANT: THIS PAPER MUST BE RETURNED**
- There are 7 parts in this exam. You must attempt 6 parts.
- Each question will be worth either 5 Marks.
- The exam will be marked out of 30 Marks.
- The exam is worth 20% of the overall grade
- This exam is optional. You may revert to the original grading structure, should you wish.
- **IMPORTANT** You must attempt parts A,B and C.
- **IMPORTANT for LENS Student:** Specifically approved LENS students must answer any 5 of the 7 parts, but must attempt 2 parts from parts A,B and C.

COVER PAGE

Attempt ALL questions

Part A (5 Marks)

The mercury level of several tests of sea-water from costal areas was determined by atomic-absorption spectrometry. The analysis of the relationship between concentration and absorbance is carried out with R and presented below.

```
> x = seq(0,100,by=10)
> y = c(0.321, 0.834, 1.254, 1.773, 2.237, 2.741, 3.196, 3.678,
4.217, 4.774, 5.261)

>myModel = lm(y~x)
>summary(myModel)

Call:
lm(formula = y ~ x)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.2933636  0.0234754   12.50 5.45e-07
x           0.0491982  0.0003968  123.98 7.34e-16
---

Residual standard error: 0.04162 on 9 degrees of freedom
Multiple R-squared: 0.9994,     Adjusted R-squared: 0.9993
F-statistic: 1.537e+04 on 1 and 9 DF,  p-value: 7.337e-16

>confint(myModel)
                 2.5 %     97.5 %
(Intercept) 0.24025851 0.34646876
x           0.04830054 0.05009582
```

(i) (2 Marks) State the Regression Equation for the fitted model.

(ii) (2 marks) State the 95% confidence interval for the slope and the intercept coefficients. Interpret this intervals with respect to any relevant hypothesis tests.

(*Please Turn Over*)

(iii) (1 Mark) The following piece of `R` code gives us a statistical metric. What is this metric? What is it used for? How should it be interpreted.

```
> AIC(model)
[1] -34.93389
```

Part B (5 Marks)

Suppose we have a regression model, described by the following equation

$$\hat{y} = 28.81 + 6.45x_1 + 7.82x_2$$

We are given the following pieces of information.

- The standard deviation of the response variance y is 10 units.
- There are 53 observations.
- The *Coefficient of Determination* (also known as the *Multiple R-Squared*) is 0.75.

Complete the *Analysis of Variance* Table for a linear regression model. The required values are indicated by question marks.

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Regression	?	?	?	?	$< 2.2e^{-16}$
Error	?	?	?		
Total	?	?			

Part C (5 Marks)

Consider the following inference procedure performed on data set Z.

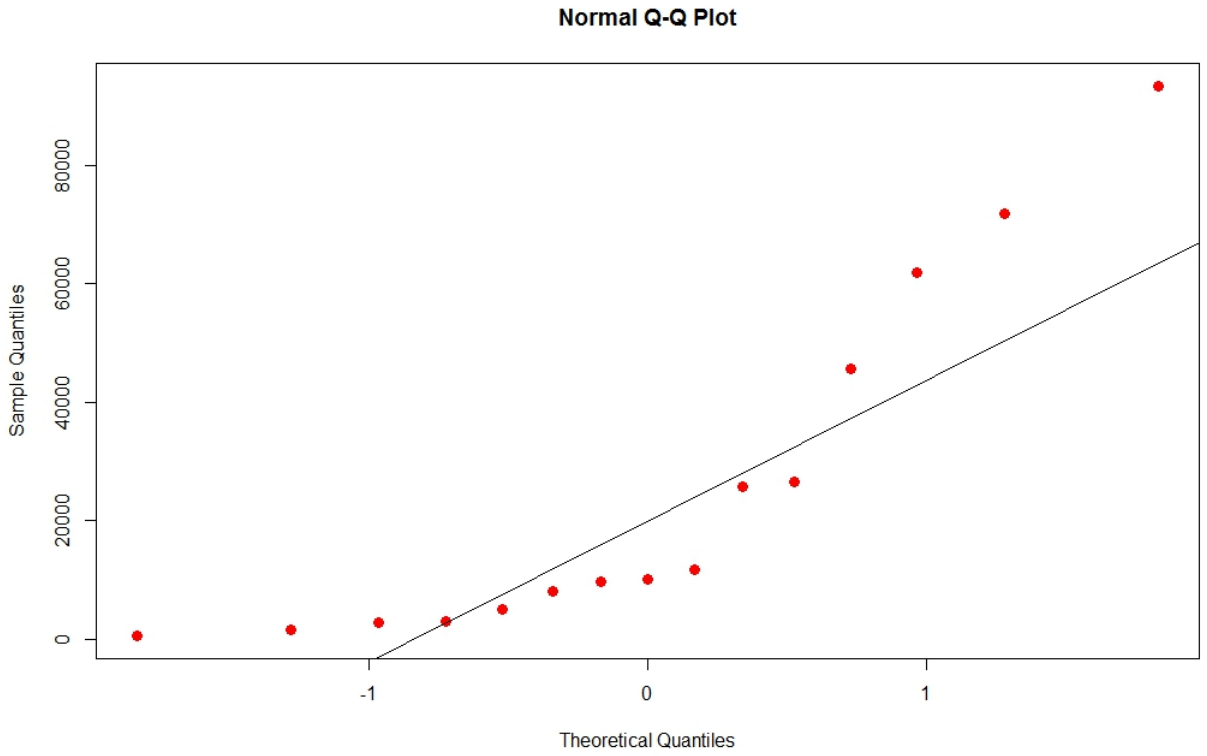
```
> shapiro.test(X)

Shapiro-Wilk normality test

data:  X
W = 0.8914, p-value = 0.007047
```

(i.) (1 Mark) Describe what is the purpose of this procedure.

(ii.) (1 Mark) What is the null and alternative hypothesis?

(iii.) (1 Mark) Write the conclusion that follows from it.

A graphical procedure was carried out to assess whether or not this assumption of normality is valid for data set `Z`. Consider the figure below.



(iv.) (1 Mark) Provide a brief description on how to interpret this plot.

(v.) (1 Mark) What is your conclusion for this procedure? Justify your answer.

Part D (5 Marks)

In certain circumstances, Robust Regression may be used in preference to Ordinary Least Squares Regression. Answer the following questions relating to Robust Regression.

(i.) (2 Marks) Explain the process of Huber Weighting for Residuals, stating the algorithm used to compute weightings.

(ii.) (3 Marks) Suppose that Huber Weighting, with a tuning constanct of $k = 13.45$, was applied to the observations tabulated below. What would be the outcome of the procedure for each case.

Observation (i)	Residual (e_i)
18	-8.011
23	16.54
25	-15.11
32	18.91

Part E (5 Marks)

(i.) (3 Marks) Provide a brief description for three tests from the family of Grubb's Outliers Tests. Include in your description a statement of the null and alternative hypothesis for each test

(ii.) (2 Marks) Describe any required assumptions for tests, and the limitations of these tests.

Part F (5 Marks)

Numeric Transformations, such as logarithmic transformation, are often used in statistical analysis as an approach for dealing with non-normal data.

(i) (1 Marks) Discuss the importance of numeric transformations, such as logarithmic transformation, in Statistics.

(ii.) (3 Marks) Give two examples of a transformation for various types of skewed data (i.e. an example for both types of skewness).

(iii.) (1 Mark) Discuss the limitations of numeric transformations.

Part G (5 Marks)

```
Int=c(2.1,5.0,9.0,12.6,17.3,21.0,24.7)
Conc=c(0,2,4,6,8,10,12)
cor.test(Int,Conc)

Pearson's product-moment correlation

data:  Int and Conc
t = 47.197, df = 5, p-value = 8.066e-08

alternative hypothesis: ..............................

95 percent confidence interval:
0.9920730 0.9998421

sample estimates:
cor
0.9988796
```

(i) (2 Marks) Describe the Statistical Procedure that you are carrying out. State the Null and Alternative Hypothesis

(ii) (1 Mark) By reference to the p-value, interpret the output of this analysis.

(iii) (2 Marks) By reference to the 95% confidence interval , interpret the output of this analysis. Explain how you came to this conlcusion

(END OF EXAM)