Department of Mathematics-University of Ruhuna
Bachelor of Science (Special) Degree (Level II)-Semester I
Mathematics

MST 4053 Statistics Laboratory

Assignment No: 02 05/02/2010

1. Data corresponding to the life span of a particular electronic component and the heat to which it is exposed are given by following table.

Heat (C)	50	100	150	200	250	300
Life span (hours)	875	884	762	424	365	128

   a) Construct scatter plot between two variables.
   b) What can you say about relationship between Heat and Life span?
   c) Find the least squares regression equation.
   d) Find the residuals and Calculate SSE.
   e) Plot the regression line on same scatter plot.
   f) Use **identify** command to identify observations which are seems to be outliers.
   g) Find *Multiple R-Squared* value and what conclusions can you make about model?
   h) Find the predicted life span for a component exposed to 180.
   i) Find Analysis of variance table.
   j) Draw box plot, histogram and normal probability plot of residuals and make conclusion about model.

2. A study of the association between reading ability and IQ scores was conducted by a reading coordinator in a lager public school system. A random sample of 14 eighth-grade students was given a reading achievement test and an IQ test. The score are recorded in the table:

Reading Score	42	35	61	28	48	46	59	21	47	29	65	37	35	53
IQ score	105	110	122	92	112	100	120	85	125	96	130	90	107	120

   a) Organize these data in a scatter plot.

b) What is meant by bivariate outliers and are there any bivariate outliers in the data?

c) Dose it appear that a straight line would fit the data reasonably well?

d) On your scatter plot, sketch in a straight line that fits the data.

---

3. In World War II, the actual number of German submarines sunk each month by the U.S. did not exactly match the number of reported by the Navy. Use the data set *Submarin*.

a) Construct a scatter plot with actual count on the vertical axis and the reported count on the horizontal axis.

b) Is there a reasonable linear pattern to the data? Is there a positive or negative trend?

c) Are there any bivariate outliers in the data?

d) Do you see two separate clusters in the scatterplot?

e) Describe the variability of the actual count when the reported count was 5 or less and the variability of the actual count when the reported count was greater than 5.

---

4. Violent crimes are the offenses of murder, forcible rape, robbery, and aggravated assault. Following are the rates per 100,000 inhabitants for each of the 50 states and the District of Columbia for 1983 and 1993.Use data set *Crime.*

a) Construct a scatterplot of the 1993 crime rate versus the 1983 crime rate.

b) Are there any outliers in the data?

c) Would the removal of outliers change the relationship between the 1983 crime rate and the 1993 crime rate?

d) Would the removal of outliers have an effect on the correlation coefficient?

e) Calculate the correlation both with and without the outlier and compare the results. Would you classify the outlier as an influential observation?

---

5. The *Iceberg* data give the number of icebergs sighted each month south of Newfoundland and south of the Grand Banks in 1920. Assess the relationship between the two. Do you think that a linear regression equation can be used to predict the number of icebergs in Newfoundland based on the number in the Grand Banks? If so, obtain the equation and use it to predict the number in Newfoundland when there are 15 in the Grand Banks.