**MS4024 Numerical Computation - R component.**

Mr Kevin O’Brien

27th April 2011

* This assessment is worth 20% of the overall grade for this module.
* The duration of the examination is 50 minutes.
* Write your answer to all questions on this document.
* Once you have submitted your script, you must leave the room.

**Part 1 : Exploratory Data Analysis (3 Marks)**

Analyse the data set ***SetN*** and answer the following questions:

1. What is the mean of the values of ***SetN***? [1 Mark]
2. What is the standard deviation of the values of ***SetN***? [1 Mark]
3. What is the median of dataset ***SetN?*** [1 Mark]

**Part 2 : Inference Procedures (5 Marks)**

The ***Anderson-Darling test*** and the ***Shapiro-Wilk test*** are two inference procedures that provide similar tests (i.e. one is an alternative to the other).

Perform the Shapiro-Wilk test on the data set ***SetG*.**

1. What is the null hypothesis for this test? [1 Mark]
2. What is the p-value of this test? [1 Mark]
3. Interpret the outcome of this test [2 Marks]

Write down a brief description of the R command ***ks.test()*** [1 Mark]

**Part 3 : Graphical Procedures (3 Marks)**

Run the following piece of R code on dataset ***SetH***, and answer the questions below.

|  |
| --- |
| >qqnorm(SetH)  >qqline(SetH) |

1. Describe how this graphical procedure works. [2 Marks]
2. Interpret the outcome of this graphical procedure for data set ***SetH***. [1 Mark]

**Part 4a : Bivariate Data ( 3 Marks)**

Data was collected on the heights( in inches) and weights ( in pounds ) of a team of six football players.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cyclist | A | B | C | D | E | F |
| Height | 68 | 64 | 62 | 65 | 66 | 61 |
| Weight | 132 | 108 | 102 | 115 | 128 | 106 |

(See R code provided : ***SetW*** and ***SetP***)

Fit a linear model to this data and answer the following questions:

1. Use the regression coefficients to write down the regression model equation.
2. What is the p-value for the slope estimate?
3. Interpret the p-value for the slope estimate, commenting on the relevance for the overall model. Justify your conclusions.

**Part 4b : Bivariate Data ( 3 Marks)**

Perform an inference procedure to test the correlation coefficient for height and weight in part 4a.

1) What is the correlation coefficient for height and weight. [1 Mark]

2) What is the confidence interval for the correlation coefficient estimate? [1 Mark]

3) Interpret the outcome of this inference procedure. Justify your conclusions. [1 Mark]

**Part 5 : Paired t –test (3 Marks)**

The weights of one group of Irish students were recorded both at the beginning of year 1 of their studies and at the end of year 4. The results (in kg) are given below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Year 1 | 72 | 58 | 68 | 81 | 65 | 69 | 75 | 84 |
| Year 4 | 74 | 61 | 69 | 83 | 69 | 74 | 76 | 82 |

( See R code provided : ***SetM*** and ***SetU***)

Carry out an inference procedure to test the diet does not on average change an individual’s weight (i.e. null: true difference in weight is zero) .

1. What is the p-value for this test?
2. What is the confidence interval for the change of an individual’s weight?
3. Interpret the outcome of this inference procedure. Justify your conclusions.