DUBLIN INSTITUTE OF TECHNOLOGY

DT211C BSc. (Honours) Degree in Computer Science (Infrastructure)

DT228 BSc. (Honours) Degree in Computer Science

Year 1

SUMMER EXAMINATIONS 2017/18

DATA EXPLORATION [CMPU1039]

MS. JANE FERRIS
DR. DEIRDRE LILLIS

THURSDAY 17TH MAY 9.30 A.M. – 11.30 A.M.

DURATION
Two hours.

INSTRUCTIONS TO CANDIDATES

ANSWER *QUESTION ONE* & TWO OTHER QUESTIONS.

QUESTION ONE IS *COMPULSORY* & CARRIES 50 MARKS.

ALL OTHER QUESTIONS CARRY 25 MARKS.

Question 1.

a) Explain what a Primary Key is in the context of a Database?

(5 marks)

b) List three different kinds of constraints expressed in the schema for a database.

(5 marks)

c) Explain what a Null Hypothesis is in relation to the proposal that "Blueberries reverse the ageing process".

(5 marks)

d) Describe and give an example of nominal data.

(5 marks)

e) What is the difference between an interval and a ratio scale of measurement?

(5 marks)

f) Describe the importance of measures of dispersion in data exploration.

(5 marks)

g) Identify two measures of dispersion in data exploration.

(5 marks)

h) Identify the origination of standard errors in inferential statistics.

(5 marks)

i) In an experiment to assess the differences between Undergraduate and Postgraduate Computer Scientist students' attitudes to binge drinking. It is proposed that the Postgraduate students binge drink more than Undergraduate students.
Respondents were asked to detail their average weekly alcoholic drinks in a questionnaire. If a T-test is performed on the data and a value of .6 is calculated. Is the proposal accepted or rejected, explain your answer in relation to the p value and the Null Hypothesis.

(10 marks)
(Total 50 marks)

Question Two

The following *Employee* table records the hours an employee spends working for a Company department each week. The database table is stored and manipulated in SQLite:

```
CREATE TABLE Employee (
                        PPSN INTEGER not null PRIMARY KEY,
                        name TEXT(20) not null,
                        age INTEGER not null,
                        salary INTEGER not null,
                        );
```

a) Write a valid SQL statement that will query the *Employee* table to identify all staff that are over forty. The query will provide the employee's name and salary in the results.

(7 marks)

Question Two

b) Draw a Data Flow Diagram or Entity Relational Diagram for the Employee table.

(9 marks)

c) Write a valid SQL statement that will query the Employee table to count all the unique names of employees.

(9 marks)
(Total 25 marks)

Question Three

a) Find the mean, median, and mode of the following data set:
{5,10, 7, 19, 25, 12, 15, 7, 6, 8}

(9 marks)

b) Describe what an outlier is in relation to data sets.

(3 marks)

c) Explain the significance of an outlier on the measures of central tendency.

(4 marks)

d) Draw a histogram for the data presented in section a) of this question.

(9 marks)
(Total 25 marks)

Question Four

a) Identify three methods of contemporary data acquisition.

(9 Marks)

b) Briefly explain the importance of the process of data cleaning in the exploration of data.

(7 Marks)

c) Identify three ways in which data becomes 'bad' or 'dirty' in relation to user interaction with data.

(9 Marks)
(Total 25 marks)