TECHNOLOGICAL UNIVERSITY DUBLIN
KEVIN STREET CAMPUS

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

BSc. (Honours) Degree in Computer Science

Year 1

SEMESTER 2 EXAMINATIONS 2018/19

Data Exploration

Ms. Jane Ferris
Dr. Deirdre Lillis

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) Identify three descriptive statistics.

(5 marks)

b) Identify three inferential statistics.

(5 marks)

c) What is the purpose of inferential statistics?

(5 marks)

d) Identify the four iterative stages of exploratory data analysis.

(5 marks)

e) Describe and give an example of ordinal data.

(5 marks)

f) What is the difference between a qualitative and quantitative measurements?

(5 marks)

g) Describe the importance of measures of central tendency in data exploration.

(5 marks)

h) Describe the purpose of a histogram in data exploration.

(5 marks)

i) Briefly explain the purpose of data visualisation in the process of data exploration.

(5 marks)

j) In an experiment to assess the relationship between drug dosage on the concentration levels of children, four sets of drug dosage are administered to children:

D0 (a placebo dose with 0/Kg); D15 (a dosage of 15/Kg); D30 (a dosage of 30/Kg) and D45 (a dosage of 45/Kg).

Identify what the Null hypothesis and alternate hypothesis are for this experiment and identify the inferential statistical technique used to assess the hypothesis.

(5 marks)
(Total 50 marks)

Question Two

The following Employee table records the hours an employee spends working for a department each week 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 to add an employee into the Employee table.

(9 Marks)

b)   Identify three issues with the assigned data types declared in relation to the sample case provided below.
{H12345, John Kelly, 23, 30000}

(6 Marks)

c)   Identify the correct data types for storing of the variables in the table.
{H12345, John Kelly, 23, 30000}

(3 Marks)

d)   Provide the SQLite code to delete the table.

(4 Marks)

e)   Identify one issue with the original assigned data types declared in the table Employee in relation to the proposed wish of the company to prepare an average wages calculation.

(3 Marks)
(Total 25 marks)

Question Three

An experiment is performed to calculate the average weight of the 62 students in the Data Exploration class.
There is insufficient time to weigh every student so a random sampling technique is performed.
The average weight of 5 students {55, 63, 67, 50, 65} measured is calculated to be 60Kgs.
Answer the following sub-questions a) to d) in relation to the data set of class weights.

a)   Identify the number of people in the population.

(2 marks)

b)   Identify the number of people in the sample.

(3 marks)

c)   Assuming that the Central Limit Theory applies to this experiment, estimate the class average weight at a 95% confidence.

(10 marks)

d)   Explain the origins of standard error and calculate the standard error for the experiment.

(10 marks)
(Total 25 marks)

3

Question Four

a) Identify four characteristics of a good data visualisation.

(8 marks)

b) Briefly describe each characteristic identified in section a) of this question and give specific examples of visualisations.

(8 marks)

c) Compare and contrast the use of MS Excel and Tableau in relation to data analysis.

(5 marks)

d) Compare and contrast the use of MS Excel and Tableau in relation to data visualisation.

(4 marks)
(Total 25 marks)