Q1. Statistics – The Basics (2nd Year Statistics – Unit 5)

a) State what type of data each of the following are using the terms numerical, categorical, discrete, continuous, ordinal, nominal:
   i) The number of cars sold by a garage last month
   ii) The blood types of the pupils in your class
   iii) The length of time taken to complete a crossword
   iv) The sizes of T-shirts worn by a group if they are S, M, L, XL
   v) Patients' body temperature
   vi) The number of animals in a zoo
   vii) The country of birth of a student

b) i) List two different methods of obtaining primary data.
   ii) Describe an advantage and a disadvantage of each of the methods mentioned in part (i)

c) Explain, using an example, what is meant by a "biased sample".

d) State three tips that should be observed when designing a questionnaire.

e) Describe briefly one way you could obtain a simple random sample of students from your school.

f) Write down the mean, mode and median of the following sets of data.
   i) 2.4, 3.1, 2.5, 3.1, 1.8, 3.4, 2.7   ii) 23, 45, 10, 52, 24, 13, 52, 4

g) For the set of data 9, 11, 11, 15, 17, 18, 94, which two measures of centre (i.e. mean, mode or median) would you choose to best describe the numbers? Explain why you would use the measures of centre chosen.

h) Explain what is meant by an "outlier" in a set of data.

i) Here are the times, in minutes, for a bus journey:
   15, 7, 9, 12, 9, 19, 6, 11, 9, 16, 8

   i) Find the range of these times.
   ii) Find the lower quartile, the upper quartile and the interquartile range

j) A 3rd year Maths class get the following results in a class test: 25, 36, 76, 55, 76, 83, 34, 50, 73, 41, 52, 47, 71, 76, 65, 44, 53, 82, 41
   i) Draw a Stem and Leaf diagram to represent the data.
   ii) What is the interquartile range of the data set?
   iii) What percentage of the class got a B grade or higher?

Q2. Statistics (2nd Year Statistics – Unit 5)

a) A class was surveyed and asked how many people they had in their families. The results are shown below.

No. Of People in Family	3	4	5	6	7
No. Of Students	6	8	9	4	3

i) What type of data has been collected in this survey?
ii) Draw a Pie Chart to represent the data above.
iii) What is the modal number of people per family?
iv) What is the median number of people per family?
v) What is the mean number of people per family?

b) The amount of time spent studying during the week, in minutes, was measured for a particular 3rd year group. The results are shown below:

Time Studying in Mins	0 - 60	60 - 120	120 - 180	180 - 240	240 - 300
No. Of Students	9	12	18	15	6

i) Draw a histogram of the data shown above.
ii) What was the modal range of times?
iii) In which interval does the median lie?
iv) Using mid-interval values, calculate an estimate of the mean.
v) What is the greatest number of people that could have spent more than 3.5 hours a week?

Q3. Extra Challenge and Problem Solving

a) The following set of data has 2 modes: 2, 3, 4, 4, 5, 6, 7, 8, a + 1, 4, 3, 5. What are the possible values of a?

b) The mean of 4 numbers is 9. Three of the numbers are 6, 11 and 14. What is the fourth number?

c) The mean price of five packets of biscuits is €1.80. The prices of two of the packets are €2.10 and €1.71. The other three packets are all the same price. Find this price.

d) Write down five numbers so that the mode is 4, the mean is 6 and the median is 5.

e) The numbers 4, 8, 12, 17, x are arranged in order of size. If the mean of the numbers is equal to the median, find x.

f) On four tests, each marked out of 100, my mean score was 85. What is the lowest mark I could have scored on any one test?

g) The following five numbers have a median of 6 and a range of 9. They are given in increasing order: 2, 2, x, 7, y   Find the values of x and y.

h) The following six numbers have a median of 15, a mean of 18 and a range of 30. They are given in increasing order: a, 8, 14, b, 26, c
Find the values of a, b and c.

i) The mean of 3, 7, 8, 10 and x is 8. Find x.

j) A sample of 10 measurements has a mean of 15.7 and a sample of the other 20 measurements has a mean of 14.3. Find the mean of all 30 measurements. Give your answer correct to one decimal place.

k) The range for eight numbers is 40 and seven of the numbers are shown below. Find two possible values for the missing number. 27, 5, 33, 42, 11, 13, 19.

l)   Amy says that "whenever the median of a list of numbers is 12, then at least one of the numbers in the list must be 12". Investigate, using examples, if her statement is true or not.

m)   All the students in a class took Spelling Test 1. A week later they all took Spelling Test 2. The scores are shown in the tables below:

Spelling Test 1			
21	27	13	9
18	35	41	39
13	26	7	48
15	36	33	21

Spelling Test 2			
19	26	37	11
35	42	47	31
25	38	5	30
23	30	28	24

i) Draw a back-to-back stem and leaf plot to display the students' scores.

ii) Both tests were out of the same mark. Estimate from the results what the total marks for the test were. Explain how you know.

iii) Compare the scores on the two spelling tests. Refer to **at least** one measure of central tendency (mean, mode, median) and **at least** one measure of variability (spread) in your answer.

iv) Jamie says that every student in the class must have done better on Spelling Test 2 than on Spelling Test 1. Is Jamie correct? Explain your answer.

Revision Sheet 8 Solutions:

Q1.

a) i) Discrete Numerical   ii) Nominal Categorical   iii) Continuous Numerical

iv) Ordinal Categorical v) Continuous Numerical   vi) Discrete Numerical

vii) Nominal Categorical

b) i) Face-to-face Interview/Telephone Interview/Online Questionnaire e.t.c.

ii) Advs: can explain the questions/can explain the questions/people more honest when not face-to-face…….

Disadvs: people might not be honest face-to-face/can be expensive/questions can't be explained…….

c) A biased sample is one that doesn't properly represent the population being surveyed. For example, if surveying people on attitudes to GAA by asking people coming out of a GAA game, the sample would be biased.

d) Keep questions simple at the start/Have no leading questions/Avoid personal questions e.t.c.

e) Assign every student in the school a number. Draw the numbers out of a hat or use the random number generator on your calculator to pick the numbers randomly.

f) i) Mean = 2.71, Median = 2.7, Mode = 3.1   ii) Mean = 27.88, Median = 23.5, Mode = 52

g) Mode or Median. There is a mode and neither are affected by the outlier 94.

h) A data value that is way off the other values in the data set.

i) i) Range = 13   ii) LQ = 8, UQ = 15, IQ Range = 7

j) i)

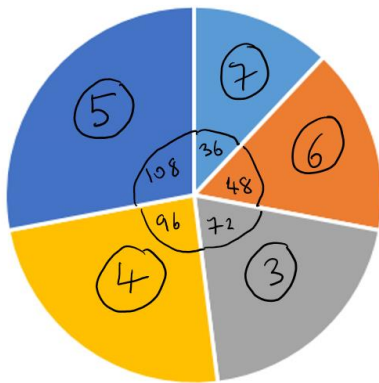| 2	5
3	4, 6
4	1, 1, 4, 7
5	0, 2, 3, 5
6	5
7	1, 3, 6, 6, 6
8	2, 3

Key: 4|1 = 41%

ii) IQ Range = 35   iii) 36.8%

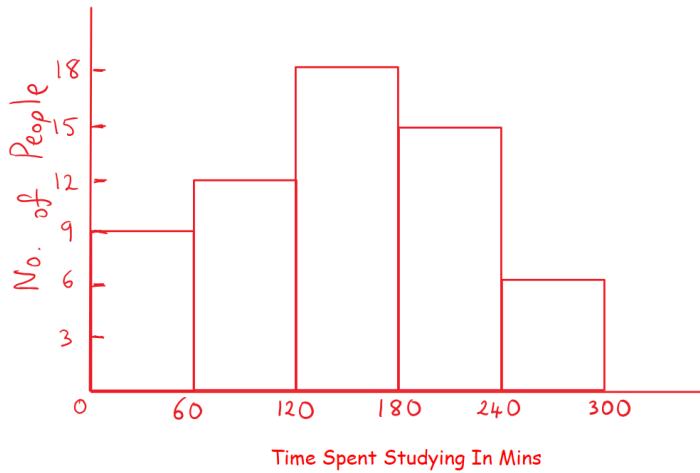## Q2.

a) i) Discrete Numerical

ii)

No of People in Family

iii) Mode = 5

iv) Median = 5

v) Mean = 4.67

b) i)

Time Spent Studying In Mins

ii) Modal Range = 120 - 180

iii) Median lies in 120 – 180 range

iv) Mean = 147mins

v) 21 people

Q3.

a) 2 or 4   b) 5   c) €1.73   d) One possibility 4, 4, 5, 8, 9   e) 19

f) 40   g) x = 6, y = 11   h) a = 7, b = 16, c = 37   i) x = 12

j) Mean = 14.77   k) 2 or 45

l) This is not true as the two middle numbers could be 11 and 13, and this would still give a median of 12. E.g. 5, 6, 7, 11, 13, 14, 16, 17

m) i)

Spelling Test 1		Spelling Test 2
9, 7	0	5
8, 5, 3, 3	1	1, 9
7, 6, 1, 1	2	3, 4, 5, 6, 8
9, 6, 5, 3	3	0, 0, 1, 5, 7, 8
8, 1	4	2, 7

Key: 1|4 = 41 marks   Key: 3|0 = 30 marks

ii) Total marks for the test were probably 50 as the two highest results were 47 and 48, which must be close to what the test was out of, and only a few students got those scores.

iii) Overall, the students did slightly better in test 2 as evidenced by the mean and the median. The median in test 1 is 21 whereas the median in test 2 is slightly higher at 28. The mean also shows this as the mean for test 1 is 25.13, while it is 28.19 for test 2. The data in test 1 is a bit more spread out than the data in test 2. We can see this as the interquartile range for test 1 is 23 while the interquartile range for test 2 is smaller at 14.5. The range for both tests is similar, which doesn't tell us anything about the spread in this case.

iv) He can't say every student definitely did better without finding out what each student got but we can see from the data that generally students did better in test 2 than test 1 as more students got above 23 on test 2 than test 1.