

Answer each of the following questions in the space provided. Complete solutions and good form are expected for full marks.

Knowledge	Thinking	Application	Communication
/17	/2	/5	/4 (form /2)

Andrew is a business student at Durham College. He is interested in statistics so he took a co-op placement at a market research company to learn more about the field.

1. His first assignment is to conduct a survey on Durham College students. Several businesses are interested in targeting this group and would like some up-to-date information. This information will help with their advertising and promotions.

a) Identify the population to be surveyed and explain why Andrew should consider taking a sample. [2C]

✓ The population is the total students at Durham College
 ✓ Sample is less expensive and time-saving and hence more accurate and reliable

b) Explain how Andrew could choose a systematic sample of this population. [1C]

→ By first assigning a number to all the students, then he decides how large his sample size should be and then divide the population to his sample size

c) Explain how Andrew could choose a random stratified sample of this population. [1C]

Andrew should divide the whole population of students into homogeneous groups (strata) and then select a sample from each stratum.

2. His second assignment is to analyze some data collected by the market research company on behalf of a local charity. The charity holds 4 fund-raising events each year: an auction, a barbeque, a marathon, and a formal dinner and dance.

a) Classify each set of data as discrete or continuous and choose an appropriate type of graph.

i) the amount of money earned from each event [1K, 1T]

Discrete data

ii) the heights of the people who attended the formal dinner and dance [1K, 1T]

Continuous data

b) Below is the number of people who entered the marathon each year for the past 20 years.

118, 109, 122, 112, 143, 135, 129, 147, 156, 138, 124, 112, 125, 117, 148, 140, 131, 151, 159, 142

Ordered:

$$\text{Mean} = \frac{\sum_{i=1}^{20} x_i}{n} = \frac{118 + 109 + \dots + 142}{20} = \frac{2652}{20}$$

$$\text{Median} = \frac{1}{2} \left[\frac{1}{2} \text{observation} + \frac{1}{2} \text{next observation} \right]$$

$$\textcircled{1} \text{ Mean} = 132.9 \text{ ANS}$$

i) Determine the mean, median and mode of this data set. [5K]

$$\textcircled{11} M = \frac{1}{2} [10^{\text{th}} + 11^{\text{th}} \text{ obs}] = \frac{1}{2} [131 + 135] = 133 \text{ ANS}$$

Median = 133

$$\textcircled{111} \text{ Mode} = 112 \quad \text{Mode } 112 \text{ ANS}$$

ii) Calculate the standard deviation for the data. [7K]

$$S^2 = \frac{1}{n} \sum (x_i - \bar{x})^2, \quad \bar{x} = 132.9$$

Data	Mean - data	(mean - data) ²
118	132 14.9	222.01
109	+23.9	571.21
122	+10.9	118.81
112	+20.9	436.81
143	-10.1	102.01
135	132 -2.1	4.41
129	3.9	15.21
147	-14.1	198.81
156	-23.1	533.61
138	-5.1	26.01
124	8.9	79.21
112	20.9	436.81
125	7.9	62.41
117	15.9	252.81
148	-15.1	228.01
140	-7.1	50.41
131	1.9	3.61
151	-18.1	327.61
159	-26.1	681.21
142	-9.1	82.81
2658		47433.8

iii) Determine a 5 number summary and use this to display the data using a box-and-whisker plot [3K, 5A]

$$S^2 = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} = \sqrt{\frac{47433.8}{20}} = 14.8893 \approx 14.89$$

$$\text{Lower quartile} = \frac{1}{4} (n+1)^{\text{th}} \text{ observation} = \frac{1}{4} (20+1) = 5.25$$

$$5^{\text{th}} \text{ observation} + \frac{1}{4} (6-5) \text{ observations} = 118 + \frac{1}{4} (122-118) = 118 + 1 = 119$$

$$\text{Upper quartile} = \frac{3}{4} (20+1) = 15.75$$

$$15^{\text{th}} \text{ observation} + \frac{3}{4} (16-15)^{\text{th}} \text{ obs}$$

$$= 143 + \frac{3}{4} (147-143) = 143 + \frac{3}{4} (4) = 146$$

① Minimum = 109

IQR

② $Q_1 = 119$

③ $Q_3 = 146$

④ Median = 133

⑤ Max = 159

~~IQR~~ =

