SM-1402: Exercise 2 (Data Summaries)

1. Recall, from Exercise 1, the heights of 20 students in centimetres.

162	177	189	173	151	178	165	179	168	158
168	147	171	168	153	184	173	166	165	158

   (a) Calculate the mean $\bar{x}$ and standard deviation s of the data set.

   We will now compare the answers you obtained in part (a) with the grouped version of the mean and standard deviation

   (b) The following grouped data table was obtained from Exercise 1. Complete the remaining blank columns.

From	To	Freq (f)	Centre (m)	$m \times f$	$m^2 \times f$
140	150	1	145		
150	160	4	155		
160	170	7	165		
170	180	6	175		
180	190	2	185		

   (c) To compute the grouped mean, sum all the values in the $m \times f$ column, and divide by the total number of observations. I.e.,

$$\bar{x}_{grp} = \frac{\sum mf}{\sum f}$$

   (d) To compute the grouped standard deviation,

      i. Calculate the average of squares: Sum all the values in $m^2 \times f$ column, and divide by the total number of observations.
      ii. Calculate the grouped variance: Subtract the quantity in i. by the squared grouped mean.
      iii. Calculate the grouped SD: Take positive square root of quantity in ii.

   Mathematically,

$$s_{grp} = \sqrt{\frac{\sum m^2 f}{\sum f} - \bar{x}^2_{grp}}.$$

   (e) Compare with the answers in (a) and comment.

2. A student obtained the following marks (in percentage) for their assignments over the course of a year in their studies.

   (a) Find, for each subject, the range and interquartile range.

   (b) Which subject is the student most "consistent" in? Explain your answer.

   (c) What is the student's "best" subject? Explain your answer.

Geography	56	49	63	58	52	50	57	61	
English	61	70	53	60	57	52	48	79	65
Science	68	56	58	73	39	47	55	76	
Mathematics	45	46	42	48	40	45	44	41	47

3. It is possible to *estimate* the quartiles (median, Q1, Q3, etc.) of a grouped data set, in the absence of raw data. A factory producing batteries is interested in finding out the number of hours (t) their batteries lasted. The data from the experiment on the time batteries lasted is represented in the table below:

From	To	Freq.	Cumulative Freq.
0.0	0.5	0	
0.5	1.0	5	
1.0	1.5	12	
1.5	2.0	32	
2.0	2.5	40	
2.5	3.0	16	
3.0	3.5	5	
3.5	4.0	4	
4.0	4.5	4	
4.5	5.0	1	

   (a) Complete the column entitled "Cumulative Freq." by summing up consecutive values in the Freq. column. Start with 0, then add 5, then add 12, etc. The last entry should be the total frequency count.

   (b) On a graph, mark the values "To" on the x-axis and "Cumulative Freq." on the y-axis by points. Connect all the points using a straight line. This is called a *cumulative frequency graph*.

   (c) To estimate the median life hours of batteries, trace the value $n/2$ on the Cumulative Freq. axis horizontally until it reaches the graph, and trace the vertical point of intersection on the x-axis.

   (d) What is the estimate of the interquartile range?