

8 The pre-release material contains information on Pulse Rate and Body Mass Index (BMI). A student is investigating whether there is a relationship between pulse rate and BMI. A **section** of the available data is shown in the table.

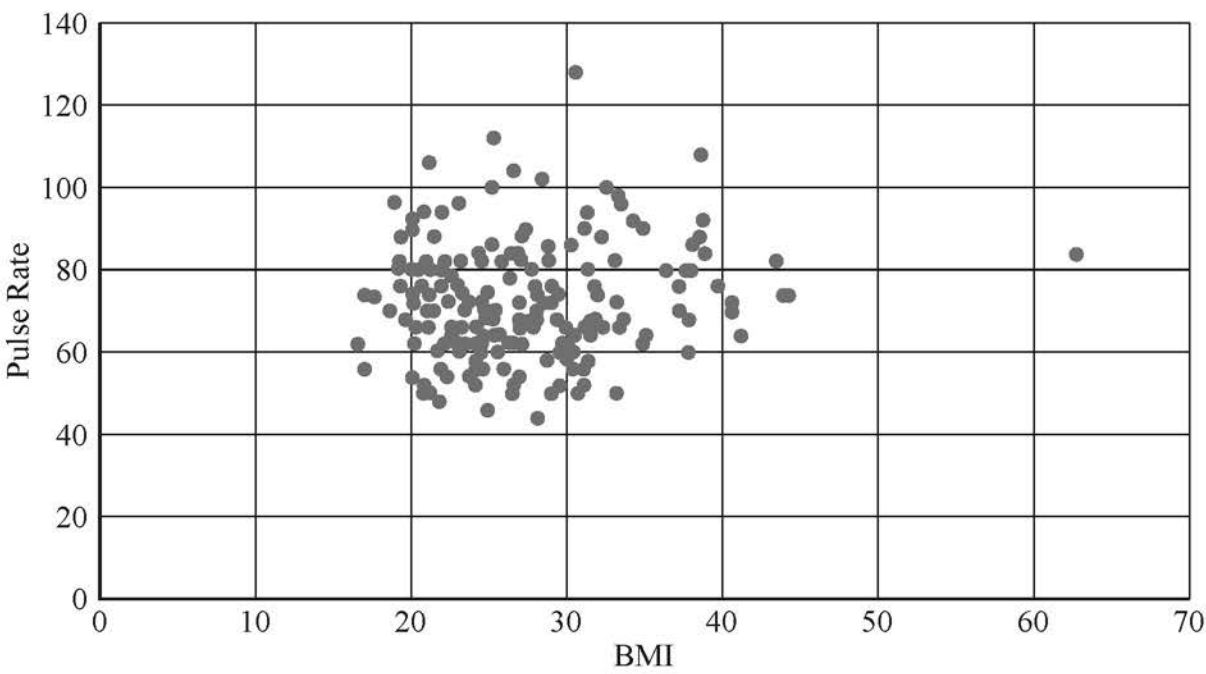
Sex	Age	BMI	Pulse
Male	62	29.54	60
Female	20	23.68	#N/A
Male	17	26.97	72
Male	35	24.7	64
Male	17	20.09	54
Male	85	23.86	54
Female	81	24.04	#N/A

The student decides to draw a scatter diagram.

(a) With reference to the table, explain which data should be cleaned before any analysis takes place. [1]

The student cleans the data for BMI and Pulse Rate in the pre-release material and draws a scatter diagram.

Scatter diagram of Pulse Rate against BMI



The student identifies **one** outlier.

(b) On the copy of the scatter diagram in the Printed Answer Booklet, circle this outlier. [1]

The student decides to remove this outlier from the data. They then use the LINEST function in the spreadsheet to obtain the following formula for the line of best fit.

$$P = 0.29Q + 64.2,$$

where P = Pulse Rate and Q = BMI.

They use this to estimate the Pulse Rate of a person with BMI 23.68.
They obtain a value of 71 correct to the nearest whole number.

- (c)** With reference to the scatter diagram, explain whether it is appropriate to use the formula for the line of best fit. **[1]**

It is suggested that all pairs of values where the pulse rate is above 100 should also be cleaned from the data, as they must be incorrect.

- (d)** Use your knowledge of the pre-release material to explain whether or not all pairs of values with a pulse rate of more than 100 should be cleaned from the data. **[1]**