

## **Section 3: Bivariate data**

## **Exercise level 3**

- 1. The four box plots below are for the following four variables (but not in that order) for a random sample of adult women.
  - Height in cm.
  - Systolic blood pressure in mm Hg.
  - Waist measurement in cm.
  - Weight in kg.



Each of the two scatter diagrams below shows the relationship between two of the four variables.



## Scatter diagram A

Scatter diagram B

- (i) Decide which box plot shows each of the variables. Explain your reasoning.
- (ii) Decide which variable is on which axis for each of the two scatter diagrams. Explain your reasoning.
- (iii) Interpret what each scatter diagram says about association between specific measurements for adult women.



## **Edexcel Collecting and interpreting data**

2. The scatter diagram below shows the relationship between engine size and  $CO_2$  emissions for a random sample of cars.



(i) A correlation coefficient is a measure of correlation.
It is a number between -1 and 1, with -1 representing perfect negative correlation, 0 representing no correlation, and 1 representing perfect positive correlation.

Match the correlation coefficients from the following list to the data set below. In each case, justify your choice.

Correlation coefficients: 0.611, 0.788, 0.894

- (A) Petrol cars data set.
- (B) Diesel cars data set.
- (C) All cars data set (i.e. petrol and diesel combined).
- (ii) For which of the three data sets would a straight line of best fit provide the most appropriate model of the relationship between engine size and  $CO_2$  emissions? Justify your answer.
- (iii) Does knowing whether a car is fuelled by petrol or diesel help to predict its CO<sub>2</sub> emissions? Justify your answer.