14 The pre-release material includes data on unemployment rates in different countries. A sample from this material has been taken. All the countries in the sample are in Europe. The data have been grouped and are shown in Fig 14.1.

Unemployment rate	0–	5–	10–	15–	20–	35–50
Frequency	15	21	5	5	2	2

Fig. 14.1

A cumulative frequency curve has been generated for the sample data using a spreadsheet. This is shown in Fig. 14.2.

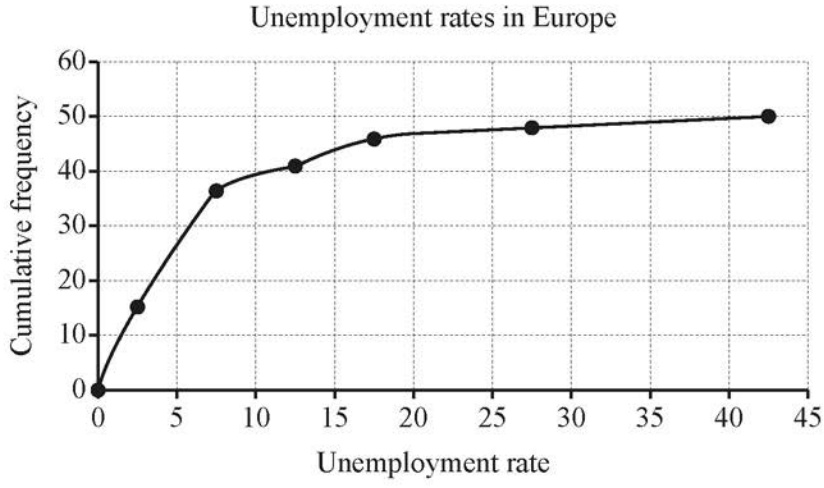

Fig. 14.2

Hodge used Fig. 14.2 to estimate the median unemployment rate in Europe. He obtained the answer 5.0. The correct value for this sample is 6.9.

(i) (*A*) There is a systematic error in the diagram.

- Identify this error.
- State how this error affects Hodge's estimate. **[2]**

(*B*) There is another factor which has affected Hodge's estimate.

- Identify this factor.
- State how this factor affects Hodge's estimate. **[2]**

(ii) Use your knowledge of the pre-release material to give another reason why any estimation of the median unemployment rate in Europe may be unreliable. **[1]**

(iii) Use your knowledge of the pre-release material to explain why it is very unlikely that the sample has been randomly selected from the pre-release material. **[1]**

The scatter diagram shown in Fig. 14.3 shows the unemployment rate and life expectancy at birth for the 47 countries in the sample for which this information is available.

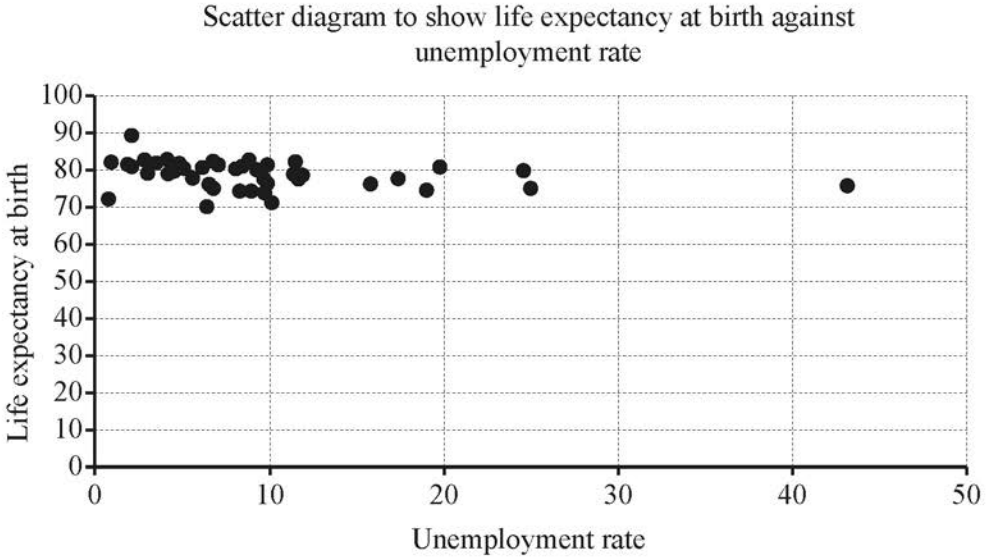

Fig. 14.3

The product moment correlation coefficient for the 47 items in the sample is – 0.2607.
The p-value associated with $r = -0.2607$ and $n = 47$ is 0.0383.

(iv) Does this information suggest that there is an association between unemployment rate and life expectancy at birth in countries in Europe? **[2]**

Hodge uses the spreadsheet tools to obtain the equation of a line of best fit for this data.

(v) The unemployment rate in Kosovo is 35.3, but there is no data available on life expectancy. Is it reasonable to use Hodge's line of best fit to estimate life expectancy at birth in Kosovo? **[1]**