

**A Level Mathematics B (MEI)**

**H640/02** MEI Pure Mathematics and Statistics

Large Data Set 1  
Statistics

**Question Set 6**

- 1 The probability that Chipping FC win a league football match is  $P(W) = 0.4$ .
- (a) Calculate the probability that Chipping FC fail to win each of their next two league football matches. [1]

The probability that Chipping FC lose a league football match is  $P(L) = 0.3$ .

- (b) Explain why  $P(W) + P(L) \neq 1$ . [1]
- 2 A survey of the number of cars per household in a certain village generated the data in Fig. 2.

Number of cars	0	1	2	3	4
Number of households	8	22	31	27	7

Fig. 2

- (a) Calculate the mean number of cars per household. [1]
- (b) Calculate the standard deviation of the number of cars per household. [1]

- 3 Every morning before breakfast Laura and Mike play a game of chess. The probability that Laura wins is 0.7. The outcome of any particular game is independent of the outcome of other games. Calculate the probability that, in the next 20 games,

- (a) Laura wins exactly 14 games, [2]
- (b) Laura wins at least 14 games. [2]

- 4 At the end of each school term at North End College all the science classes in year 10 are given a test. The marks out of 100 achieved by members of set 1 are shown in Fig. 4.

3		5
4		0 9
5		2 3 6
6		0 1 3 5 6
7		0 1 2 5 6 8 9 9
8		3 4 6 6 8 8 9
9		5 5 5 6 7

Key 5 | 2 represents a mark of 52

Fig. 4

- (a) Describe the shape of the distribution. [1]
- (b) The teacher for set 1 claimed that a typical student in his class achieved a mark of 95. How did he justify this statement? [1]
- (c) Another teacher said that the average mark in set 1 is 76. How did she justify this statement? [1]

Benson's mark in the test is 35. If the mark achieved by any student is an outlier in the lower tail of the distribution, the student is moved down to set 2.

- (d) Determine whether Benson is moved down to set 2. [2]

- 5 The screenshot in Fig. 5 shows the probability distribution for the continuous random variable  $X$ , where  $X \sim N(\mu, \sigma^2)$ .

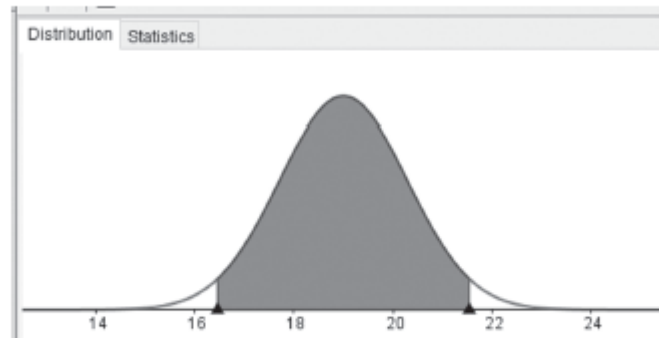


Fig. 5

The area of each of the unshaded regions under the curve is 0.025. The lower boundary of the shaded region is at 16.452 and the upper boundary of the shaded region is at 21.548.

- (a) Calculate the value of  $\mu$ . [1]
- (b) Calculate the value of  $\sigma^2$ . [3]
- (c)  $Y$  is the random variable given by  $Y = 4X + 5$ .
- (i) Write down the distribution of  $Y$ . [3]
- (ii) Find  $P(Y > 90)$ . [1]

- 6 The discrete random variable  $X$  takes the values 0, 1, 2, 3, 4 and 5 with probabilities given by the formula

$$P(X=x) = k(x+1)(6-x).$$

- (a) Find the value of  $k$ . [2]

In one half-term Ben attends school on 40 days. The probability distribution above is used to model  $X$ , the number of lessons per day in which Ben receives a gold star for excellent work.

- (b) Find the probability that Ben receives no gold stars on each of the first 3 days of the half-term and two gold stars on each of the next 2 days. [2]
- (c) Find the expected number of days in the half-term on which Ben receives no gold stars. [2]

- 7 In the summer of 2017 in England a large number of candidates sat GCSE examinations in both mathematics and English. 56% of these candidates achieved at least level 4 in mathematics and 80% of these candidates achieved at least level 4 in English. 14% of these candidates did not achieve at least level 4 in either mathematics or English.

Determine whether achieving level 4 or above in English and achieving level 4 or above in mathematics were independent events. [5]

- 8 Each weekday Keira drives to work with her son Kaito. She always sets off at 8.00 a.m. She models her journey time,  $x$  minutes, by the distribution  $X \sim N(15, 4)$ .

Over a long period of time she notes that her journey takes less than 14 minutes on 7% of the journeys, and takes more than 18 minutes on 31% of the journeys.

- (a) Investigate whether Keira's model is a good fit for the data. [3]
- Kaito believes that Keira's value for the variance is correct, but realises that the mean is not correct.
- (b) Find, correct to two significant figures, the value of the mean that Keira should use in a refined model which does fit the data. [2]

Keira buys a new car. After driving to work in it each day for several weeks, she randomly selects the journey times for  $n$  of these days. Her mean journey time for these  $n$  days is 16 minutes. Using the refined model she conducts a hypothesis test to see if her mean journey time has changed, and finds that the result is significant at the 5% level.

- (c) Determine the smallest possible value of  $n$ . [5]

9

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 9.1.

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

Fig. 9.1

A cumulative frequency curve has been generated for the sample data using a spreadsheet. This is shown in Fig. 9.2.



Fig. 9.2

Hodge used Fig. 9.2 to estimate the median unemployment rate in Europe. He obtained the answer 5.0. The correct value for this sample is 6.9.

- (a) (i) There is a systematic error in the diagram.
- Identify this error.
  - State how this error affects Hodge's estimate. [2]
- (ii) There is another factor which has affected Hodge's estimate.
- Identify this factor.
  - State how this factor affects Hodge's estimate. [2]
- (b) 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]
- (c) 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. 9.3 shows the unemployment rate and life expectancy at birth for the 47 countries in the sample for which this information is available.

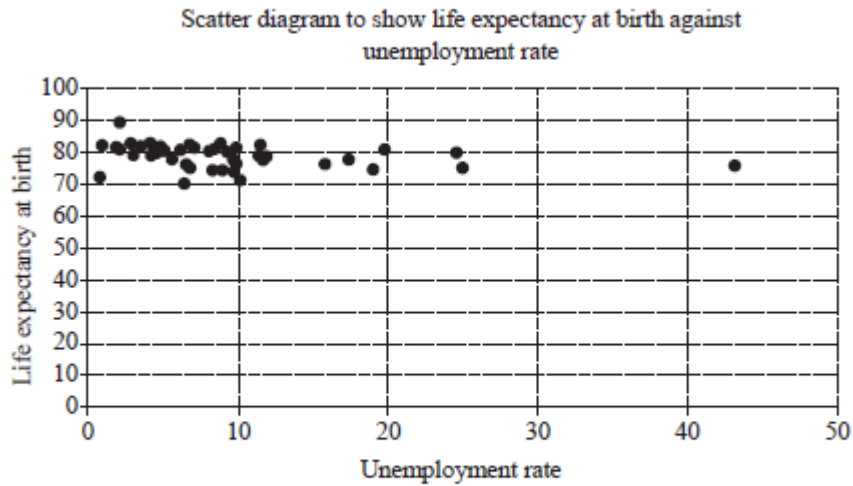


Fig. 9.3

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  $0.0383$ .

- (d) 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.

- (e) 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]

**Total Marks for Question Set 6: 51**

# Resource Materials

Question Set No: 6

Fig. 2

Number of cars	0	1	2	3	4
Number of households	8	22	31	27	7

Fig. 4

3	5
4	0 9
5	2 3 6
6	0 1 3 5 6
7	0 1 2 5 6 8 9 9
8	3 4 6 6 8 8 9
9	5 5 5 6 7

Key 5 | 2 represents a mark of 52

Fig. 5

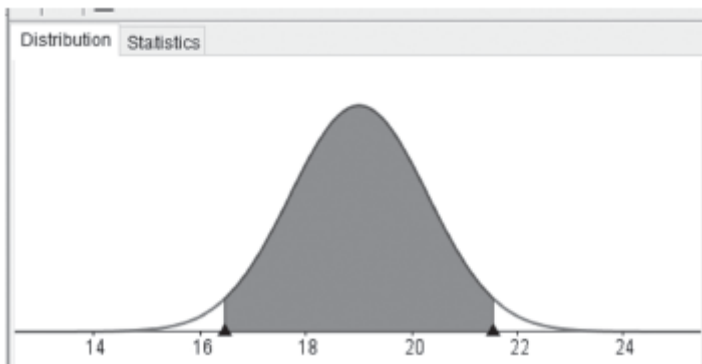


Fig. 9.1

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

Fig. 9.2

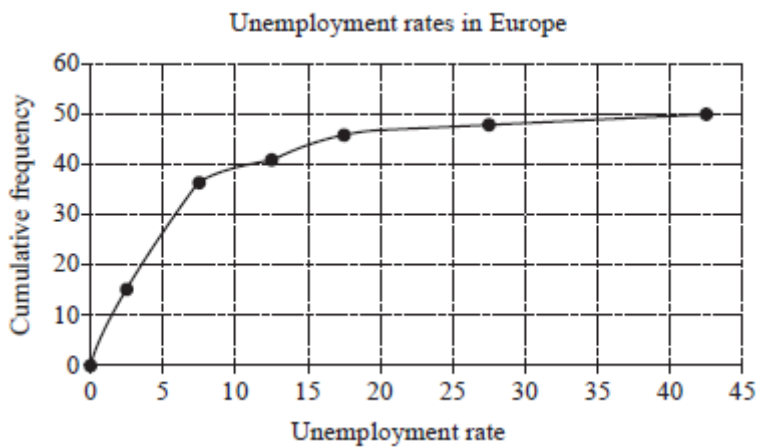
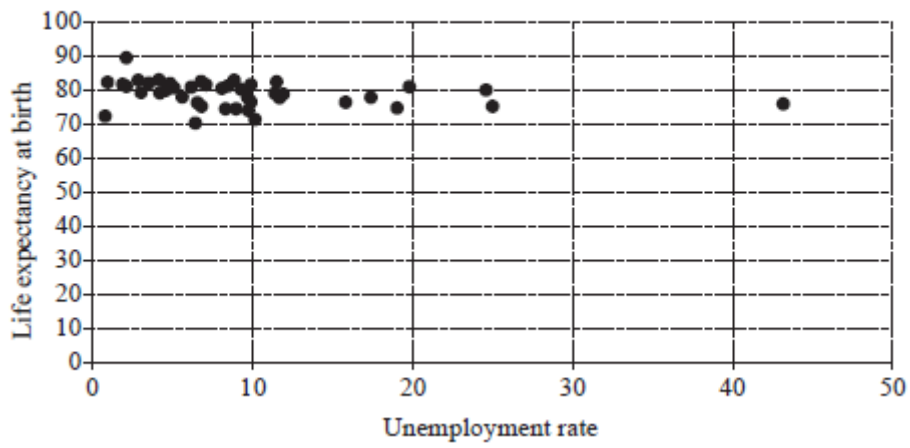


Fig. 9.3

Scatter diagram to show life expectancy at birth against unemployment rate



---

# OCR

Oxford Cambridge and RSA

## **Copyright Information**

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website ([www.ocr.org.uk](http://www.ocr.org.uk)) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge