Surname	
Other Names	
Candidate Signature	

Centre Number						Candidate Number				

Examiner Comments	

Total Marks

MATHEMATICS

LARGE DATA SET PRACTICE

CM

[Insert series] (Edexcel Version)

Time allowed: [Insert time]

Instructions to candidates:

• In the boxes above, write your centre number, candidate number, your surname, other names and signature.

• Answer ALL of the questions.

• You must write your answer for each question in the spaces provided.

• You may use a calculator.

Information to candidates:

• Full marks may only be obtained for answers to ALL of the questions.

• The marks for individual questions and parts of the questions are shown in round brackets.

• There are 6 questions in this question paper. The total mark for this paper is X.

Advice to candidates:

• You should ensure your answers to parts of the question are clearly labelled.

• You should show sufficient working to make your workings clear to the Examiner.

• Answers without working may not gain full credit.

1 0 3 3 2 2 X X 8 L D S

1 Ellie wants to investigate rainfall in the UK in 2015.

She takes a random sample of 14 days from July 2015 for Heathrow from the large data set.

The data she collected is summarised in the table below.

Amount of rainfall (r mm)	Frequency
trace	7
$1 < r \leq 2$	4
$2 < r \leq 4$	3
$r > 4$	0

(a) Work out an estimate for the mean and standard deviation of Ellie's data.

(b) Interpret the value of your standard deviation in part (a).

(c) (i) Comment on the suitability of Ellie's sampling method for her investigation.

(ii) Suggest how Ellie could make better use of the large data set for her study.

TOTAL X MARKS

2 (a) Give an example of a discrete variable in the large data set.

James is studying the total amount of sunshine in Leeming in 2015.

He wants a sample of 15 data points from the large data set.

(b) Explain how James can use simple random sampling to obtain a sample of size 15 from the large data set for his study.

James works out that the mean of his 15 data points is 5.6 hours.

He concludes that in 2015, Leeming had an average of 5.6 hours of sunshine each day.

(c) Comment on the reliability of James' conclusion with reference to his sample size.

(d) State **one** limitation of James using the large data set for his study.

TOTAL X MARKS

3 Michael is using the large data set to investigate the relationship between the time of the year and the maximum daily temperature, T °C, in the UK.

He looks at the daily temperatures in Leuchars to do this.

Starting with 01/05/2015, he labels each of the days in the large data set with a number x. The day 01/05/2015 is given the number 1, the day 02/05/2015 is given the number 2 and so on.

He then plots a scatter diagram of T against x.

Michael expects there to be 184 values for x.

(a) Using your knowledge of the large data set, explain why

(i) he expects there to be 184 values for x,

(ii) there may be less than 184 values of x.

Michael calculates the regression line for T on x.

His regression line has the equation

$$T = 16.551 - 0.0027x$$

(b) Interpret the gradient of Michael's regression line.

(c) Estimate the temperature in Leuchars on 03/05/2015.

(d) Use your knowledge of the large data set to explain why it is unreliable to use Michael's regression line to estimate the temperature on days in Leuchars.

Question 3 continued

TOTAL X MARKS

4 Paul is doing an investigation about Beijing and wants to take a random sample of 12 data points from the large data set.

He looks at the daily mean wind speed for Beijing in 2015. The list of data is enumerated from 1 to 184 and 12 three digit random numbers are generated.

His list of random numbers are

192
138
289
986
004
103
736
420
075
602
387
138

(a) State the name of the sampling method used by Paul.

(b) Give one advantage of Paul using a sample of the data points from the large data set.

(c) How many distinct data points do Paul's random numbers correspond to?

Paul wants to investigate the relationship between wind speed and cloud cover in Beijing in 2015.

(d) Explain why Paul cannot use the large data set alone for this investigation.

Question 4 continued

TOTAL X MARKS

5 Luke is investigating the relationship between air temperature, T °C, and pressure, p hPa, in Asia in 2015.

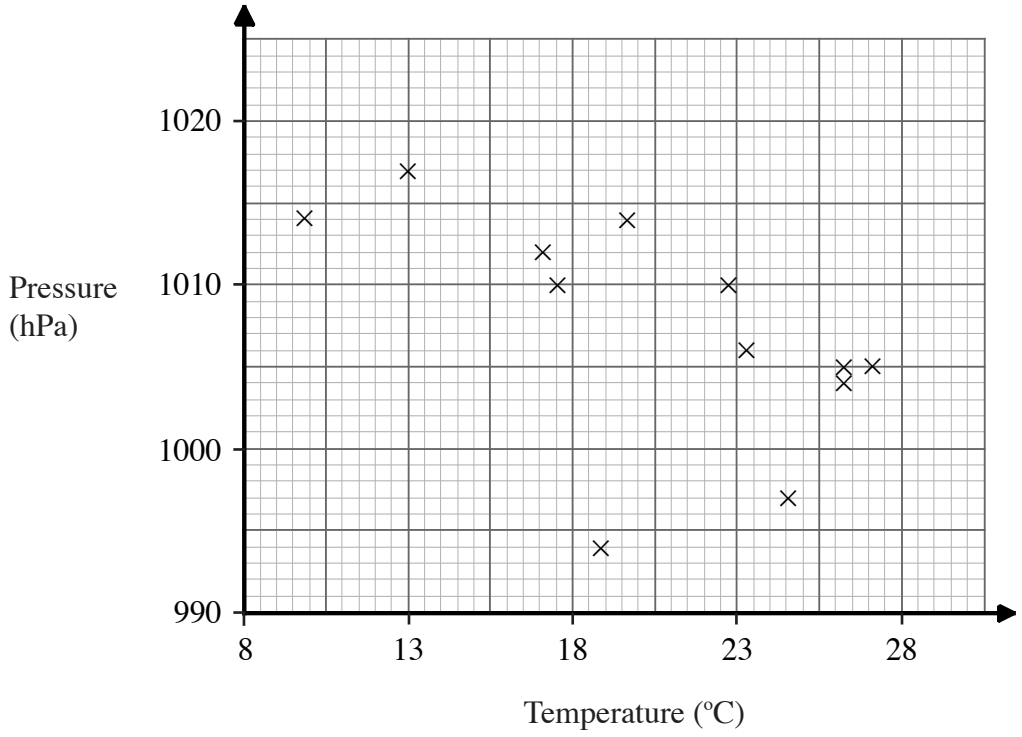
He takes a random sample of 12 days from May 2015 for Beijing from the large data set.

He obtained the following data.

T	17.5	13.0	24.6	23.3	19.6	26.3	22.8	17.1	9.7	18.9	26.3	27.2
p	1010	1017	997	1006	1014	1004	1010	1012	1014	994	1005	1005

Luke drew the following scatter graph for T and p and calculated the quartiles.

	Q_1	Q_2	Q_3
T	17.3	21.2	25.5
p	1005	1008	1013



An outlier is a value which is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(a) Show that this data has no outliers.

(b) Comment on the correlation between the daily mean temperature and the pressure in this sample.

Luke finds that the regression line for p on T is

$$p = -0.71T + 1022$$

(c) Give an interpretation to the figure –0.71 in this regression line.

Luke finds that the average temperature on one day in Beijing in December 2015 is 8.5 ºC.

(d) Estimate the pressure on that day using the regression line.

(e) Using your knowledge of the large data set, comment on the reliability of your estimate in part (e).

(f) Suggest how Luke could make better use of the large data set for his study.

(g) Explain why Luke should not use only the large data set for his study.

TOTAL X MARKS

6 Zain wants to calculate the average daily mean windspeed in Hurn in 2015.

To do this, he takes a simple random sample of the daily mean windspeeds, v knots, on n days in Hurn in 2015 using the large data set.

He converts his values for v into miles per hour. He calls the resulting values w.

Given that

$$\sum w = 194.35 \qquad \overline{v} = 8.45$$

find the size of Zain's sample.

Question 6 continued

END OF PAPER **TOTAL X MARKS**

TOTAL FOR PAPER IS X MARKS