

Second Year Assignment 15

This assignment is slightly different. It is all to do with the “Large Data Set”
You should download the Large Data Set and explore it. (Google “Edexcel Large Data Set”)

Here is an extract from the A level Specification which explains how important this is.

Use of data in statistics

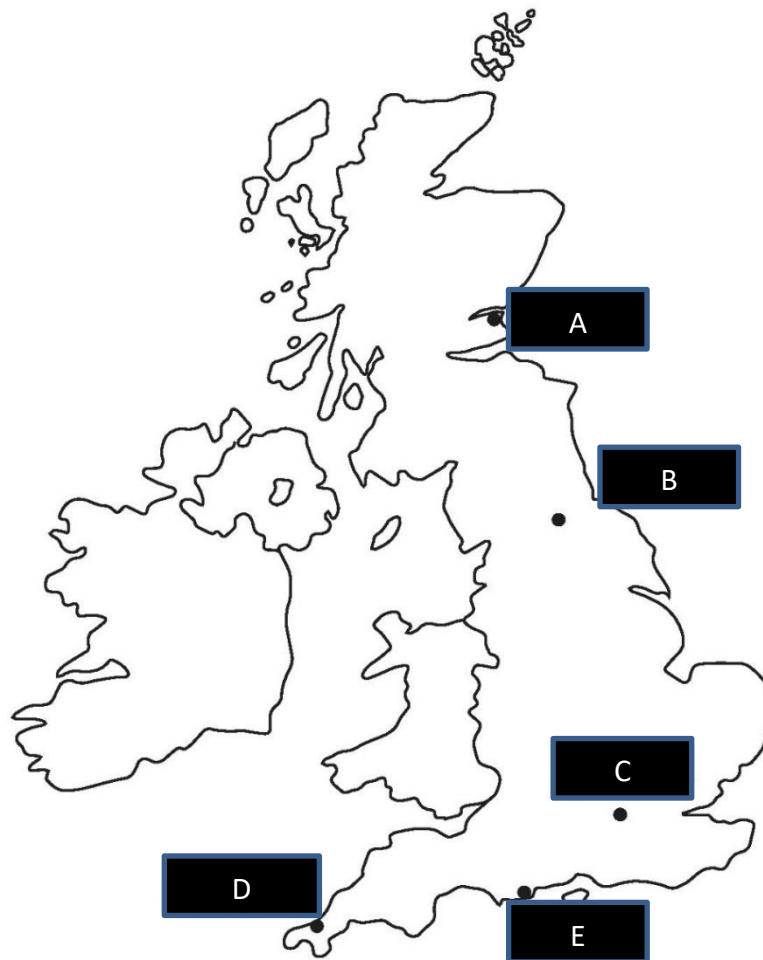
EDEXCEL has provided a large data set, which will support the assessment of Statistics in Paper 3: Statistics and Mechanics. Students are required to become familiar with the data set in advance of the final exam.

Exams will be designed in such a way that questions assume knowledge and understanding of the data set. The expectation is that these questions should be likely to give a material advantage to students who have studied and are familiar with the data set. They might include questions/tasks that:

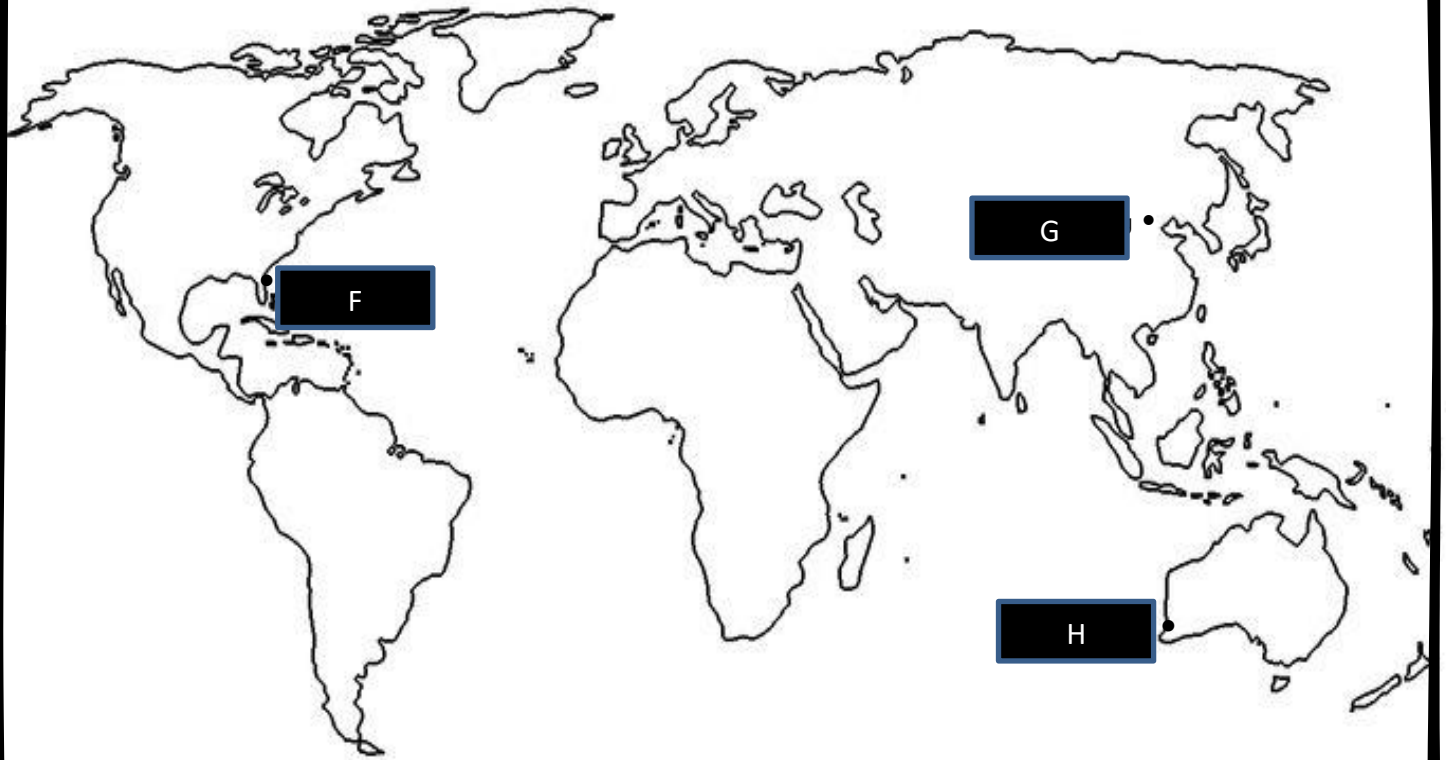
- assume familiarity with the terminology and contexts of the data, and do not explain them in a way that gives students who have not studied the data set the same opportunities to access marks as students who have studied them
- use summary statistics or selected data from, or statistical diagrams based on, the data set – these might be provided in the question or task, or as stimulus materials
- are based on samples related to the contexts in the data set, where students’ work with the data set will help them understand the background context and/or
- require students to interpret data in ways that would be too demanding in an unfamiliar context.

Students will not be required to have copies of the data set in the examination, nor will they be required to have detailed knowledge of the actual data within the data set.

SECTION A



1. i) Name the towns at A, B, C, D and E



- ii) a) Name the cities at F, G, and H
b) In which countries are cities F, G, and H ?

2. Copy and complete this table

Attribute	Units	Discrete or Continuous?	Min	Max	Qualitative or Quantitative	Possible values
<i>Daily Mean Temperature</i>						
<i>Daily Total Rainfall</i>						
<i>Daily Total Sunshine</i>						
<i>Daily Maximum Relative Humidity</i>						
<i>Daily Mean Windspeed</i>						
<i>Daily Maximum Gust</i>						
<i>Daily Mean Wind Direction</i>						
<i>Cardinal Wind Direction</i>						
<i>Daily Maximum Gust Direction</i>						
<i>Cardinal Gust Direction</i>						
<i>Cloud Cover</i>						
<i>Visibility</i>						
<i>Daily Mean Pressure</i>						

3. i) Convert these wind speeds to miles per hour

- a) 1 knot b) 10 knots c) $\frac{20}{23}$ knots d) x knots

ii) Convert these speeds to wind speeds

- a) 1 m.p.h. b) 10 m.p.h. c) $\frac{23}{20}$ m.p.h. d) x m.p.h.

4. The pressure unit used in meteorology was previously the millibar. However, this has been replaced by the SI unit of pressure — the pascal (Pa).

New units

1 Hectopascal (hPa) = 100 Pascals (Pa)

Old units

1 Bars = 1000 Millibar (mb)

Conversion between new and old units

1 Bar = 100 000 Pascals (Pa)

Find x and y

1 mb = x Pa

1 mb = y hPa

5. Look at the columns for “Wind Direction” and “Cardinal Direction” for Camborne in 2015. On July 7th the Wind Direction was 90 and the Cardinal Direction was East. On August 5th the Wind Direction was 180 and the Cardinal Direction was South.

a) What is the name for the system used to measure Wind Direction?

b) Convert the following Wind Directions to Cardinal Directions

i) 270

ii) 135

iii) 225

iv) 202.5

v) 310

vi) 20

c) Find the smallest bearing that would give a Cardinal Direction of

i) East

ii) South East

iii) North North West

iv) East South East

6. a) Copy and complete this table

Windspeed (Beaufort conversion)	Beaufort number	Wind speed (knots)
Calm	0	< 1
Light air	1	1 – 3 knots
Light breeze	2	
Gentle breeze	3	
Moderate breeze	4	
Fresh breeze	5	
Strong breeze	6	
High wind, moderate gale	7	
Gale	8	
Strong gale	9	
Storm	10	
Violent storm	11	
Hurricane	12	

b) In the Large Data Set, numbers 1,2 and 3 have been combined. What description has been given to this combination?

c) On which dates in 2015 was there a “Strong” wind in Leuchars?

SECTION B

1. James kept a diary during his stay in Cornwall in the Summer of 2015. These are extracts from his diary on the first day of every month between May and October. Use the large data set to find the month of each entry

A: After such a sunny day yesterday, it was disappointing that I only saw the sun briefly today.

B: It was the fourth dry day in a row. Beautiful sunny day although a bit on the cold side.

C: I was looking forward to today because I fancied going on a cliff top walk to look at the views. The weather let me down. It was cloudy all day and fairly windy.

D: It was a warm humid day.

E: Woke up late. It was pouring with rain outside and it was cold and windy.

F: Sunshine and showers today. I went for a walk on the cliff tops and the views were amazing. I could see for miles, or so it seemed.

2. What is humidity?

3. Investigate the correlation between humidity and cloud cover.

4. Investigate the correlation between humidity and temperature

5. It's a bright cold clear calm dry sunny day in Leuchars.

Use estimation to complete this table

Month	
Daily Mean Temperature (0900-0900) (°C)	
Daily Total Rainfall (0900-0900) (mm)	
Daily Total Sunshine (0000-2400) (hrs)	
Daily Mean Windspeed (0000-2400) (kn)	
Daily Mean Total Cloud (oktas)	
Daily Maximum Relative Humidity %	

6. It's a hot humid windy rainy cloudy day in Perth.

Use estimation to complete this table

Month	
Daily Mean Temperature (0900-0900) (°C)	
Daily Total Rainfall (0900-0900) (mm)	
Daily Total Sunshine (0000-2400) (hrs)	
Daily Mean Windspeed (0000-2400) (kn)	
Daily Mean Total Cloud (oktas)	
Daily Maximum Relative Humidity %	

SECTION C

1. Sara is investigating the variation in the daily maximum gust, t kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

(a) State the sampling technique used by Sara.

(b) From your knowledge of the large data set explain why this process may not generate a sample of size 20.

The data Sara collected are summarised as follows $n = 20, \sum t = 374, \sum t^2 = 7600$

(c) Calculate the standard deviation. Give your answer to 3 s.f.

2. Sara was studying the relationship between rainfall, r mm, and humidity, $h\%$, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set. She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1, Q_2 = 0.9, Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

(a) Show that $r = 20.6$ is an outlier.

(b) Give a reason why Sara might:

(i) include this day's reading

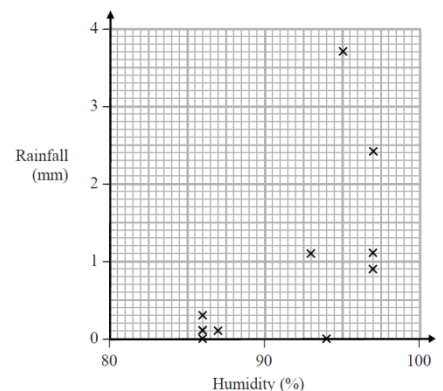
(ii) exclude this day's reading.

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h .

c) Give an interpretation of the correlation between rainfall and humidity.

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$

(d) Give an interpretation of the gradient of this regression line.



(e) (i) Comment on the suitability of Sara's sampling method for this study.

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

3. Lauren wants to find the average daily mean windspeed in Hurn in 1987. She only has access to the large data set. She uses it to obtain a random sample of the daily mean windspeeds, t knots, on n days in Hurn in 1987.

The data collected by Lauren are summarised as follows

$$\sum (t - 5) = 55, \quad \bar{t} = 10$$

(a) Find n .

Lauren uses the same sampling method to estimate that the average daily mean windspeed in Hurn in 2015 was 11 mph.

(b) Convert 11 mph into knots.

(c) Hence, compare the average daily mean windspeed in Hurn in 1987 and 2015.

(d) With reference to the large data set, state **one** limitation of your conclusion in part (c).

(e) Explain how Lauren can

(i) improve her data collection method

(ii) improve her data processing

to allow for a more reliable comparison in part (c).

4. The table shows the mean daily temperatures at each of the eight weather stations for August 2015

	Camborne	Heathrow	Hurn	Leeming	Leuchars	Beijing	Jacksonville	Perth
Mean daily mean temp °C	15.4	18.1	16.2	15.6	14.7	26.6	26.4	13.6

a) Give a geographical reason why the temperature in August might be lower in Perth than in Jacksonville

b) Comment on whether this data supports the conclusion that coastal locations experience lower average temperatures than inland locations.

5. Harriet believes that the random sample S , representing total daily hours of sunshine from the large data set, can be modelled by a discrete uniform distribution, once S has been rounded to the nearest integer.

a) Write down the probability distribution of S

b) Using this model, find the probability that the total daily hours of sunshine is less than 10

c) State what makes Harriet's assumption very unlikely

d) Suggest a refinement to Harriet's model.

6) Joshua compares the amount of rain in 2015 between Heathrow and the city X on the continent of Asia using the Large Data Set.

(a) Write down the name of the city X that Joshua compares with Heathrow.

At random, he selects 8 data points about the daily total rainfall, in mm, in May 2015 for the two cities. These 8 data points are shown below.

Heathrow: 7.0 0.2 1.2 tr 0.8 6.8 0.2 4.2

City X: 6.0 0.0 20.7 9.0 14.3 0.5 0.0 0.4

(b) Explain what is meant by the reading 'tr'.

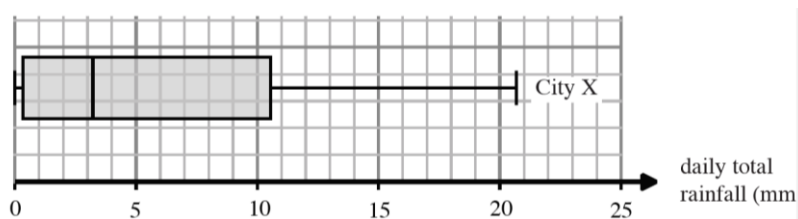
(c) State **one**

(i) advantage

(ii) disadvantage

of Joshua using 8 data points from the large data set for his comparisons.

The diagram below shows a box-plot for the data collected by Joshua on the rainfall in the city X in May 2015.



Draw another box-plot to represent the data collected by Joshua for Heathrow. In your data processing, take 'tr' to mean 0.0 mm of rainfall and ignore outliers.

(e) Compare the amount of rainfall in May 2015 between Heathrow and the city X.

Answers

SECTION A

1. i) A = Leucahrs, B = Leeming, C = Heathrow, D = Camborne, E = Hurn
 ii) a) E = Jacksonville, F = Beijing, G= Perth
 b) E = USA, F= China, G= Australia

2.

Attribute	Units	Discrete or Continuous?	Min	Max	Qualitative or Quantitative	Possible values
<i>Daily Mean Temperature</i>	°C	Discuss with your teacher	n/a	n/a	Quantitative	Integers + tenths
<i>Daily Total Rainfall</i>	millimetres	Discuss with your teacher	0	n/a	Quantitative	Integers + tenths
<i>Daily Total Sunshine</i>	Hours	Discuss with your teacher	0	24*	Quantitative	Integers + tenths
<i>Daily Maximum Relative Humidity</i>	Percentage	Discrete	0	100	Quantitative	Integers
<i>Daily Mean Windspeed</i>	Knots	Discuss with your teacher	0	n/a	Quantitative	Integers
<i>Daily Maximum Gust</i>	Knots	Discrete	0	n/a	Quantitative	Integers
<i>Daily Mean Wind Direction</i>	Degrees	Discrete	0	360	Quantitative	Multiples of 10
<i>Cardinal Wind Direction</i>	Points of compass	Discrete	n/a	n/a	Qualitative	N, NNE, NE, ENE, E, ESE, SE SSE, S, SSW, SW, WSW, W, WNW, NW NNW
<i>Daily Maximum Gust Direction</i>	Degrees	Discrete	0	360	Quantitative	Multiples of 10
<i>Cardinal Gust Direction</i>	Points of compass	Discrete	n/a	n/a	Qualitative	N, NNE, NE, ENE, E, ESE, SE SSE, S, SSW, SW, WSW, W, WNW, NW NNW
<i>Cloud Cover</i>	Oktas	Discrete	0	8	Quantitative	Integers
<i>Visibility</i>	Decametres	Discrete	0	n/a	Quantitative	Multiples of 100
<i>Daily Mean Pressure</i>	Pascal	Discrete	n/a	n/a	Quantitative	Integer

*You may not agree with this but bear in mind that the midnight sun occurs in places North of the Arctic Circle and South of the Antarctic Circle in the summer months. However, none of these places fall into this category. For more detail look at the Wikipedia entry for "Midnight Sun"

3.i) a) 1.15 b) 11.5 c) 1 d) $1.15x$

ii) a) $\frac{20}{23}$ b) $\frac{200}{23}$ c) 1 d) $\frac{20x}{23}$

4. $x = 100, y = 1$

5. a) Bearings

b) i) West ii) South East iii) South West

iv) South South West v) North West vi) North North East

c) i) 78.75 ii) 123.75 iii) 348.75 iv) 101.25

6.

Windspeed (Beaufort conversion)	Beaufort number	Wind speed (knots)
Calm	0	< 1
Light air	1	1 – 3 knots
Light breeze	2	4 – 6 knots
Gentle breeze	3	7 – 10 knots
Moderate breeze	4	11 – 16 knots
Fresh breeze	5	17 – 21 knots
Strong breeze	6	22 – 27 knots
High wind, moderate gale	7	28 – 33 knots
Gale	8	34 – 40 knots
Strong gale	9	41 – 47 knots
Storm	10	48 – 55 knots
Violent storm	11	56 – 63 knots
Hurricane	12	> 63 knots

b) Light,

c) June 2nd, June 6th, October 22nd

SECTION B

- 1) A: August 1st
- B: October 1st
- C: May 1st
- D: July 1st
- E: June 1st
- F: September 1st

2. **Humidity** is a term used to describe the amount of water vapour present in air.

5. These are not “the correct answers” but merely estimates

Month	October
Daily Mean Temperature (0900-0900) (°C)	8
Daily Total Rainfall (0900-0900) (mm)	0
Daily Total Sunshine (0000-2400) (hrs)	10
Daily Mean Windspeed (0000-2400) (kn)	7
Daily Mean Total Cloud (oktas)	1
Daily Maximum Relative Humidity %	70

6. These are not “the correct answers” but merely estimates

Month	January
Daily Mean Temperature (0900-0900) (°C)	30
Daily Total Rainfall (0900-0900) (mm)	10
Daily Total Sunshine (0000-2400) (hrs)	0
Daily Mean Windspeed (0000-2400) (kn)	15
Daily Mean Total Cloud (oktas)	7
Daily Maximum Relative Humidity %	90

SECTION C

1. a) Systematic
b) In the Large Data Set some days have gaps because the data was not recorded.
c) 5.51

2. b) i) e.g. It is a piece of data and we should consider all the data
ii) e.g. It is an extreme value and could unduly influence the analysis **or**
It could be a mistake.
c) e.g. "as humidity increases rainfall increases"
d) e.g. a 10% increase in humidity gives rise to a 1.5 mm increase in rainfall **or**
represents 0.15mm of rainfall per percentage of humidity
e) i) Not a good method since only uses 11 days from one location in one month
ii) e.g. She should use data from more of the UK locations and more of the months **or**
using a spreadsheet or computer package she could use all of the available UK data

3. a) 11
b) 9.56 knots
c) Hurn had a higher average daily mean windspeed in 1987 than in 2015
d) The Large Data Set only covers the months May to October **or**
The Large Data Set does not cover the whole year
e) i) use a larger data set so it is more representative
ii) consider standard deviation/variation, so she can compare/take into account the spread of the data **or**
consider another average/the median/the mode, so she can compare with respect to other averages **or**
consider/exclude outliers, to avoid the average being influenced by extreme values

4. a) Perth is in the Southern hemisphere so August is a winter month
b) The lowest temperatures in the U.K. are at coastal locations (Camborne and Leuchars).
The highest temperature is at an inland location (Beijing). There is some evidence to support this location although a greater sample size of different locations would lead to a more accurate conclusion.

5. a)

s	0	1	2	3	4	5	6	7	8	9	10	11	12
P(S=s)	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$

s	13	14	15	16	17	18	19	20	21	22	23	24	
P(S=s)	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	

b) $\frac{2}{5}$

c) None of these cities are in the Arctic Circle so there will always be some hours when the sun has set – so the maximum number will never be 24

d) The number of hours of sunshine will vary from month to month and place to place
So e.g. use a non-uniform distribution.

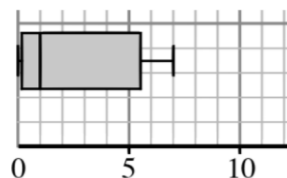
6. a) Beijing

b) tr means a rainfall of less than 0.05 mm

c) i) large data set has many more than 8 data points, so using 8 points is easier to process, it is quicker to process and requires less analysis

ii) large data set has many more than 8 data points, so using 8 points may not very representative and may lead to inaccurate and unreliable conclusions

d)



e) Heathrow had less rainfall on average than City X as the median is lower; Heathrow had less variation in the amount of rainfall it received than City X, as the Interquartile range and the range is smaller