KING EDWARD VI
HANDSWORTH GRAMMAR
SCHOOL FOR BOYS

KING EDWARD VI
ACADEMY TRUST
BIRMINGHAM

Year 12 and Year 13
Statistics
Large Data Set

Name: _____

Class: _____

The Large Data Set

Locations

5 UK weather stations

3 overseas

Time Periods

May – October 1987 (6 months)
May – October 2015 (6 months)

Seasons

May/June are the end of spring
July-Sept is summer
October is autumn

Perth (Australia) is in the southern hemisphere, so July-Sept is winter

UK Great Storm

The night of 15-16th October 1987
Gusts up to 100 knots recorded

Florida hurricanes

12 October 1987 Hurricane Floyd
1-2 October 2015 Hurricane Joaquin

Variables Recorded

Daily Mean Temperature
°C

Daily Total Rainfall
mm

Daily Mean Pressure
hPa (hectoPascal)
1 hPa = 100 Pa

Daily Mean Windspeed;
Daily Maximum Gust
knots (1kn = 1.15mph)
and Beaufort scale

Daily Mean Wind Direction;
Daily Maximum Gust Direction
bearing (°)
and cardinal direction

Daily Total Sunshine
hours

Daily Maximum Relative Humidity
%; mist and fog if > 95%

Daily Mean Total Cloud
oktas (eighths); 0 – 8

Daily Mean Visibility
Dm (decametres)
1 Dm =10 m

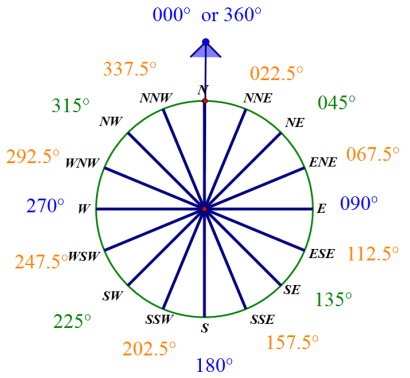
n/a
data not available

tr (trace)
0 < rainfall < 0.05mm

Beaufort Scale

Discrete, scale of 13 values:
0 (calm, < 1kn)
12 (hurricane, 64kn+)

Cardinal Directions



Oktas

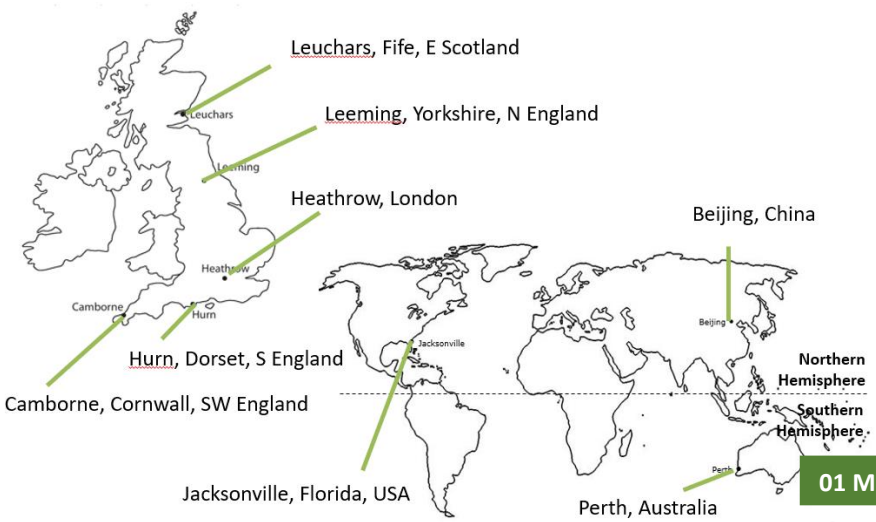
Eighths of the sky covered by cloud
Discrete, scale of 9 values:
0 (clear sky)
8 (completely overcast)

Variables in red are available for the international locations; orange are UK-only

Sources

Maps:	*Pearson*
Compass:	*mathsmutt.co.uk*

Edexcel Large Data Set – Factsheet



5 UK locations

Have a rough idea of UK geography (e.g. the more north a location, the more likely temperature will be lower, rainfall will be higher, etc …)

3 overseas locations

Again, have a rough idea of world geography (e.g., Florida is prone to hurricanes and Perth is the only location in the LDS that is in the southern hemisphere)

01 May to 31 October (6 months, 184 days)

UK – 14 variables

1. Daily Mean Temperature (°C) to 1 d.p.
2. Daily Total Rainfall (mm) to 1 d.p.
3. Daily Total Sunshine (hours) to 1 d.p.
4. Daily Maximum Relative Humidity (%)
5. Daily Mean Visibility (Dm): 1 Dm = 10 m
6. Daily Mean Total Cloud (oktas)
7. Daily Mean Pressure (hPa)
8. Daily Mean Windspeed (knots)
9. Daily Mean Windspeed (Beaufort conversion)
10. Daily Maximum Gust (knots)
11. Daily Mean Wind Direction (bearing)
12. Daily Mean Wind Direction (cardinal direction)
13. Daily Maximum Gust Direction (bearing)
14. Daily Maximum Gust Direction (cardinal direction)

7 wind related variables!

Majority of rainfall values in ALL locations is 0!

above 95% associated with mist and fog

note link with humidity

Qualitative data

October 1987

All UK locations experienced relatively higher rainfall in comparison to May – September due to the great storm/hurricane

October 1987 & 2015

Beijing significantly colder relative to May – September

Overseas – 5 variables

1. Daily Mean Temperature (°C) to 1 d.p.
2. Daily Total Rainfall (mm) to 1 d.p.
3. Daily Mean Pressure (hPa)
4. Daily Mean Windspeed (knots)
5. Daily Mean Windspeed (Beaufort conversion)

Direction wind is blowing FROM, e.g. direction of 270° (W) means wind is blowing from West to East

Did you know?

'Tr' and n/a does NOT appear overseas?

Trace 'tr'

Rainfall is the only variable with 'tr'.

Trace means values of rainfall less than 0.05mm.

If using trace in a calculation, mean & standard deviation for example, use 0.025 (cleaning data).

Discrete variables (relevant in probability questions)

Cloud cover (0 – 8, 9 outcomes),

Cardinal directions (N, NNE, etc … 16 outcomes)

Not available 'n/a'

The only 4 variables to have n/a are: daily total sunshine, daily mean windspeed (and Beaufort conversion) and daily maximum gust.

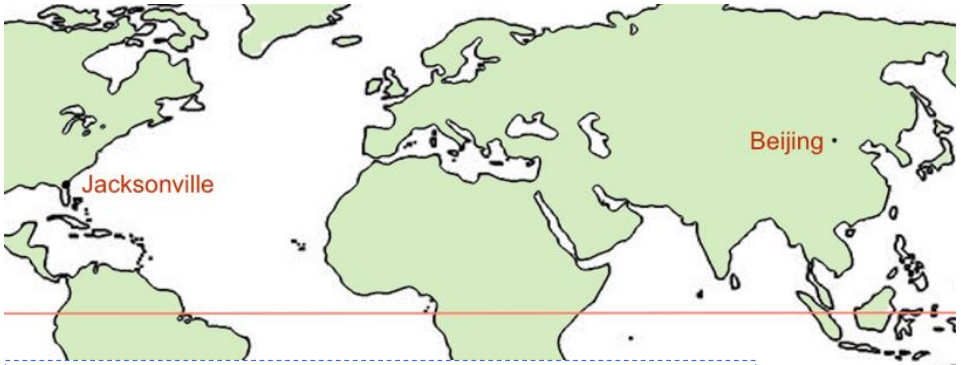
The four aforementioned variables are all n/a for the first two weeks of May 1987 – relevant in sampling questions.

UK

Heathrow is warmest.
Heathrow and **Hurn** are the driest.
Leuchars is the wettest, coldest, windiest, and furthest north.

The **GREAT STORM** happened in the UK in Oct 1987, with high wind speeds affecting the Southeast the most.
In 2015 May was the windiest month.
Lots of missing data in May 1987

Worldwide



Perth has both 0 rainfall and very high rainfall.
Jacksonville is hot and humid. **Beijing** is inland.
Florida hurricanes in Octobers of both years.

LARGE DATA SET

1987 and 2015

May to October

Temperature

1987 to 2015 **increasing** due to global warming.

Cleaning data

Delete **"n/a"**
Removing **anomalies** (mistakes)
Converting the **'tr'** values to 0

Types of data

"n/a" not available
"tr" trace (0 to 0.05mm)
Qualitative - Beaufort Scale
Fresh, light, moderate, strong
Quantitative
Discrete
Cloud coverage 0 - 8 Oktas
Continuous
The rest of the data
Outliers Unusual data
Anomalies Errors

Variables

Daily Mean Temp
Degrees C

Daily Total Rainfall
mm

Daily Mean Windspeed
Beaufort Scale
Gust
knots (1kn - 1.15mph)
Direction
Bearing

Daily Total Sunshine
Hours

Daily Max Rel. Humidity
% and fog >95%

Daily Mean. Total Cloud
oktas (eighths) - 0 to 8

Daily Mean Pressure
1 hPA - 100 Pa

Daily Mean Visibility
Dm (decametres)

THE LARGE DATA SET

KEY WORDS & DEFINITIONS

1. Daily Mean Temperature
The average of hourly temperature readings in a 24hour period, in Celsius.

2. Daily Total Rainfall
The depth of precipitation as a liquid. All precipitation is included, not just rainfall, but it is melted if necessary for the measurement. Heights less than 0.05mm are recorded as a "trace" or "tr".

3. Daily Total Sunshine
Recorded to the nearest 10th of an hour (6 minutes).

4. Daily Mean Wind Direction
Given as a bearing and/or in cardinal (compass) directions.

5. Daily Mean Windspeed
Averaged over 24 hours of a day (midnight to midnight), in Knots, nautical miles per hour where 1 Knot = 1.15mph. Can also be categorised by the Beaufort Scale.

6. Daily Maximum Gust
The highest instantaneous windspeed recorded, in Knots.

7. Daily Maximum Gust Direction
The direction of the maximum gust of wind recorded.

8. Daily Maximum Relative Humidity
A percentage of air saturation with water vapour. Relative humidities above 95% result in mist or fog.

9. Daily Mean Cloud Cover
Measured in eighths of the sky that is covered (Oktas).

10. Daily Mean Visibility
The greatest horizontal distance at which an object can be seen in daylight, measured in decametres (Dm).

11. Daily Mean Pressure
Measured in hectopascals (hPa)

WHAT DO I NEED TO KNOW?

1. What the Large Data Set is about
The Edexcel LDS has samples on weather data in different locations for certain time periods. The data is provided by the Met Office.
The LDS contains the weather data for 5 UK weather stations and 3 weather stations overseas.

2. How to clean the data
N/A should be removed before calculations
tr (trace) should be turned to 0

3. Locations
Learn maps and understand geographical significance of North, South, Coastal etc,

4. Dates
Remember the Large Data Set only has information from May–October 1987 and May–October 2015. Anything between November and April is outside the range of our data

5. Understand OKTAS
A measure of the fraction of the celestial dome covered by cloud, measured in
eighths. 0 oktas represents a clear sky,
while a value of 8 indicates complete overcast.

6. How to convert units
1 knot = 1.151 mph

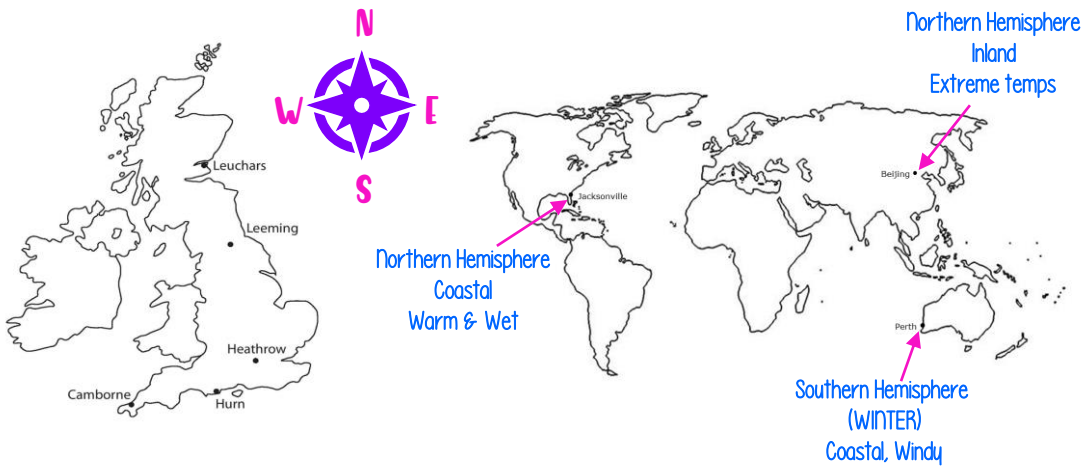
7. Limitations
These stations do not tell us about the whole UK

THE BEAUFORT SCALE

Beaufort Scale	Description	Av. Wind Speed 10m above ground
0	Calm	< 1 Knot
1-3	Light	1 – 10 Knots
4	Moderate	11 – 16 Knots
5	Fresh	17 – 21 Knots

UK DATA

Location (N to S)	Temp Range (°C)	Wind Speed Range (kn)
Leuchars	4 – 9	3 – 23
Leeming	4 – 23	3 – 17
Heathrow	8 - 29	3 – 19
Hurn	6 - 24	2 – 19
Camborne	10 - 20	3 – 18



Large data sheet Summary

	Unit of measure	average	range	International	info to note
Date	dd-mm-yyyy	-	May to Oct	x	weather for Perth opposite season to UK
Daily Mean Temperature	°C	13	4 to 30	x	increase in ave from 1987 to 2015 (global warming?)
Daily Total Rainfall	mm (tr <0.05)	2	0 to 55	x	includes snow, drought in June 1987
Daily Total Sunshine	hrs (to 1dp)	5	0 to 16		missing data in May 1987
Daily Mean Windspeed	kn (1.15 mph)	8	1 to 20	x	never 0 kn
Daily Mean Windspeed	Beaufort Scale	light	light>moderate>fresh	x	never calm (0 kn)
Daily Maximum Gust	kn (1.15 mph)	20	7 to 70		Hurricane Oct 1987 affected south
Daily Maximum Relative Humidity	%	94	65 to 100		95%+ mist and fog
Daily Mean Total Cloud	oktas	5	0 to 8		negative correlation with sunshine
Daily Mean Visibility	dam (10 m)	2400	(0) 600-6300		one day in Camborne 19887 visibility zero (error?)
Daily Mean Pressure	hPa (1 mb)	1013	980 to 1040	x	990 low (bad weather) 1025 high (good weather)
Daily Mean Wind Direction	bearing °	220	0 to 360		999, 0 and N/A all appear in the data
Cardinal Direction	compass direction	SSW	3 figure compass points		the bearing and cardinal directions for Leuchars in 1987 don't match
Daily Max Gust Direction	bearing °	210	0 to 360		N/A's recorded in Camborne 1987
Cardinal Direction	compass direction	SSW	3 figure compass points		N/A's recorded in Camborne 1987

The large data set (Edexcel)

What is the large data set?

- Edexcel's LDS looks at weather data across 5 locations within the UK and 3 locations abroad.
- Edexcel will test knowledge and familiarity of the data set.
- You will **not** be required to take copies of the LDS into the exam and you will not be expected to have a detailed knowledge of the actual data within the data set.

The Edexcel specification states for questions that use the LDS:

- *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.*

In particular, it makes the following remarks about questions testing the LDS:

- Questions may assume familiarity with the terminology and contexts of the data and may not explain them. This is so that students that have not seen or studied the data set do not have the same opportunities to access marks as students that have seen and studied the data set.
- Questions may use summary statistics or selected data from, or statistical diagrams, based on the data set – these might be given in the question/task, or as stimulus material;
- Questions may be 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.
- Questions may require students to interpret data in ways that would be too demanding in an unfamiliar context.

Locations of the large data set

The LDS contains data for 5 UK weather stations and 3 weather stations overseas for May to October 1987 and May to October 2015. This is important that you know which months the LDS contains.

The UK weather stations are:

- Camborne
- Heathrow
- Hurn
- Leeming
- Leuchars

Advice: You will need to have a rough idea of where these locations are in terms of how north and south they are within the UK.

Leuchars is a lot further north than any other location and it is by the sea so we would expect it to be colder than other locations like Hurn and Heathrow.

The great storm of 1987: This happened between 15th October 1987 and 16th October 1987. This brought heavy rain, as well as heavy wind speeds and devastating damage. Check out the videos below for context:

- https://www.youtube.com/watch?v=NnxjZ-aFkjs
- https://www.youtube.com/watch?v=ciBox3QHYq4

The overseas locations are:

- Jacksonville (Florida, USA)
- Beijing (China)
- Perth (Australia)

Data set variables

The LDS features 11 variables:

- Daily Mean Temperature
- Daily Total Rainfall
- Daily Total Sunshine
- Daily Maximum Relative Humidity
- Daily Mean Windspeed
- Daily Maximum Gust
- Daily Mean Wind Direction
- Daily Maximum Gust Direction
- Cloud Cover
- Visibility
- Pressure

If a reading is not available, then it is listed as 'n/a'.

Daily Maximum Temperature

This is measured in Degrees Celsius. Data values are given to one decimal place. A negative value indicates a temperature less than 0℃.

Daily Total Rainfall

All totals are given in millimetres. **If the total amount of rainfall recorded is less than 0.05mm, then it is recorded as 'tr' (trace)**. **A "tr" value is treated as 0 if we have to do a calculation with it.**

Daily Total Sunshine

Values for this are given in hours and recorded to one decimal place. For example, an entry of '4.5' would indicate that there were 4.5 hours (e.g. four and a half hours) of sunshine on that particular day in that particular location.

Daily Maximum Relative Humidity

Relative humidity is a measure of the saturation of water vapour in the atmosphere. Higher relative humidities indicate that the air contains more water vapour. The values are given as a percentage. Values above 95% are associated with mist and fog. If the relative humidity is 100%, then the air is fully saturated and condensation can occur.

Daily Mean Windspeed

Daily mean windspeed is measured in knots in one column. **1 knot is 1.15 miles per hour.** The daily mean windspeed is also recorded using the Beaufort scale in another column. This is a non-numerical and empirical scale that maps windspeeds to a number and then a descriptive term. The descriptive terms are **light, moderate and fresh**. Light happens the most in most locations.

Daily Maximum Gust

This is the maximum instantaneous speed that occurred over a 24 hour period. It is measured in knots.

Daily Mean Wind Direction

The daily mean wind direction is averaged over the 24 hours of the day. This value is given in degrees relative to true North and is the direction the wind was blowing **from**. The corresponding cardinal direction is also given. Values are rounded to the nearest 10 degrees.

Daily Maximum Gust Direction

This is the direction the wind was blowing **from** in the hour that the corresponding daily maximum gust occurred. Values are given in degrees relative to true North. The corresponding cardinal direction is also given. The direction of the maximum gust is that direction from which the wind was blowing.

Cloud cover

This is a measure of the fraction of the celestial dome covered by cloud. It is measured in eighths. The technical unit in this case is the okta. A value of 0 oktas represents a clear sky, while a value of 8 indicates complete overcast **(discrete variable with 9 values)**.

Visibility

This is measured horizontally. Readings are given in decametres **(1dm=10m)**. Unavailable data is indicated by a dash. Usually between 1000dm and 7000dm. Less than 100dm is fog and less than 200dm is mist.

Pressure

- The measurement is taken at the location and converted into the equivalent amount for the pressure at sea level (to remove the effect of altitude at location).
- Pressure is measured in hectopascals (hPa) to the nearest whole number.
- This value is always around 1000 hPa.
- Higher pressure is associated with good weather
- Lower pressure is associated with rain, cold, cloud.
- The UK locations have pressures of 980 to 1040 hPa.
- The overseas locations have pressures of 1000 to 1040 hPa.

Analysis of the data

UK locations

For the UK locations in 2015, the LDS gives (rounded to the nearest whole):

UK location	Temperature range	Rainfall range	Windspeed range
Camborne	10-20	0-34	3-18
Heathrow	8-29	0-52	3-19
Hurn	6-24	0-26	2-19
Leeming	4-23	0-26	3-17
Leuchars	4-19	0-23	3-23

You can see that Heathrow has the highest recorded value of temperature. This is influenced by the airport and its black asphalt runways and airport buildings which naturally absorb more heat.

Heathrow has an extremely high rainfall range. The highest rainfall value of 51.6ml was recorded on 26th August 2015 and was down to the fact of extreme rainfall within the UK at the time. The next highest was 37.6ml and after this it goes down to 19ml. This suggests that we have two outliers (could be an exam question in the future).

The windspeed ranges are fairly similar amongst the UK locations with Leuchars having some higher values. We can suggest that this is because of its close proximity to the sea.

The locations of Leuchars and Leeming have the lowest temperatures due to their locations being further north than the others.

Overseas locations

The overseas locations have less variables. They only have daily average temperature, daily rainfall, daily mean pressure, and daily mean windspeed in knots and then in the **Beaufort scale (descriptive terms and not continuous data).**

For the overseas locations in 2015, the LDS gives (rounded to the nearest whole):

UK location	Temperature range	Rainfall range	Windspeed range
Jacksonville	15-31	0-80	1-12
Beijing	8-33	0-49	2-9
Perth	8-25	0-102	4-14

Jacksonville

This is located in Florida in the USA, close to the sea. It is in the Northern hemisphere. We have high temperatures here. There is more rainfall here than in the UK due to being nearer the equator and close to where hurricanes form. The UK's rainfall pattern is more spread out as Jacksonville's rainfall is more intense. Jacksonville had 102 days without rainfall in 2015 (May 1st - October 31st). There are moderate windspeeds at this location.

Beijing

This is located in China and is in the Northern hemisphere. High temperatures are expected here with values similar to Jacksonville, however, we see lower temperatures than Jacksonville due to the phenomenon of monsoon circulation. This effect is where cold currents from the North West prevail in the winter and then hot and humid currents from the tropics replace these in the summer. Monsoon circulation also brings **intense** rainfall at times during the summer months. Despite this, Beijing still has periods of time without rain and this is why it had 115 days without rain in 2015 (May 1st - October 31st).

Perth

This is located in Australia. A key difference is that Perth is in the Southern Hemisphere. The significance of this is that the seasons are reversed in the Southern Hemisphere, so between June-August, it is winter. It is located next to the sea, so we expect higher windspeeds here. Perth has two high rainfall values of 63ml and 102ml. The rainfall range would have been 0-38ml if these two values were excluded (probably outliers).

A Level Maths
Large Data Set

CM

Introduction

These notes will look at the Edexcel Large Data Set (LDS). We will explore the variables from the data set and note key points from it. We will also give examples of exam questions that require you to apply your knowledge of the LDS to illustrate what the expectation is.

What is the Large Data Set?

All exam boards have designed a large data set for the use in Statistics sections of A Level Maths exams. Schools are expected to dedicate some teaching time to exploring the large data set as some exam questions will test knowledge and familiarity of the data set. You will not be required to take copies of the LDS into the exam and you will also not be expected to have a detailed knowledge of the actual data within the data set.

The Edexcel specification states for questions that use the LDS,

> "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."

In particular, it makes the following remarks about questions testing the LDS:

- questions may assume familiarity with the terminology and contexts of the data and may not explain them. This is so that students that have not seen or studied the data set do not have the same opportunities to access marks as students that have seen and studied the data set;
- questions may use summary statistics or selected data from, or statistical diagrams based on, the data set – these might be given in the question/task, or as stimulus material;
- questions may be 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;
- questions may require students to interpret data in ways that would be too demanding in an unfamiliar context.

You can download a copy of the data set from the Edexcel website.

What is the Edexcel LDS about?

The Edexcel LDS contains data about weather in several locations and during certain time periods. The focus is therefore to study weather patterns in these locations, make comparisons and be able to explain any findings using basic meteorological knowledge that you will develop by working through the data set.

Locations

The LDS contains data for 5 UK weather stations and 3 weather stations overseas for May to October 1987 and May to October 2015. Notice that November, December, ..., April are not included.

The UK weather stations are:

- Camborne
- Heathrow
- Hurn
- Leeming
- Leuchars

Study Tip

You will need to have a rough idea of the geography of all the locations; for example, that Leuchars is more North than Heathrow. This information will be useful in explaining trends.

The overseas weather stations are:

- Beijing
- Jacksonville
- Perth

Data Set Variables

The LDS features 11 variables:

- Daily Mean Temperature
- Daily Total Rainfall
- Daily Total Sunshine
- Daily Maximum Relative Humidity
- Daily Mean Windspeed
- Daily Maximum Gust
- Daily Mean Wind Direction
- Daily Maximum Gust Direction
- Cloud Cover
- Visibility
- Pressure

Now we will briefly give some more information on each of these variables.

Daily Maximum Temperature

This is measured in Degrees Celsius. Data values are given to one decimal place. A negative value indicates a temperature less than 0°C. If a reading is not available, it is listed as 'n/a'.

Daily Total Rainfall

All totals are given in millmetres. If the total amount of rainfall recorded is less than 0.05mm, then it is recorded as 'tr'. This stands for a trace (tr) of rain. If a reading is not available, it is recorded as 'n/a'.

Daily Total Sunshine

Values for this are given in hours and recorded to one decimal place. For example, an entry of '4.5' would indicate that there were 4.5 hours (eg four and a half hours) of sunshine on that particular day in that particular location. If a reading is not available, it is recorded as 'n/a'.

Daily Maximum Relative Humidity

Relative humidity is a measure of the saturation of water vapour in the atmosphere. Higher relative humidities indicates that the air contains more water vapour. The values are given as a percentage. Values above 95% are associated with mist and fog. If the relative humidity is 100%, then the air is fully saturated and condensation can occur. If a reading is not available, it is recorded as 'n/a'.

Daily Mean Windspeed

Daily mean windspeed is measured in knots. 1 knot is 1.15 miles per hour. If a reading is not available, it is recorded as 'n/a'.

The daily mean windspeed is also recorded using the Beaufort scale. This is a non-numerical and empirical scale that maps windspeeds to a number. Find more here.

Daily Maximum Gust

This is the maximum instantaneous speed that occurred over a 24 hour period. It is measured in knots. If a value is not available, it is listed as 'n/a'.

Daily Mean Wind Direction

This value is given in degrees relative to true North. The corresponding cardinal direction is also given. Values are rounded to the nearest 10 degrees.

Daily Maximum Gust Direction

This is the direction the wind was blowing in the hour the corresponding daily maximum gust occurred. Values are given in degrees relative to true North. The corresponding cardinal direction is also given.

Terminology
When we talk about the direction of a wind, we talk about the direction from which it is blowing. For example, a 270° wind is blowing *from* the West.

Cloud cover

This is a measure of the fraction of the celestial dome covered by cloud. It is measured in eighths. The technical unit in this case is the okta. A value of 0 oktas represents a clear sky, while a value of 8 indicates complete overcast.

Visibility

This is measured horizontally. Readings are given in metres. Unavailable data is indicated by a dash.

Pressure

This is recorded in hectopascals (hPa). Note Pascals is the unit and hecto is a unit prefix.

Another common unit used to measure pressure in meterology is the bar. 1 bar is 1000 millibars and 1000 millibars is 1 hectopascals.

Example Analysis of the Data

UK Locations

For the UK locations in 2015, the large data set gives:

UK Location	Temperature Range (°C)	Windspeed Range (kn)
Camborne	10-20	3-18
Heathrow	8-29	3-19
Hurn	6-24	2-19
Leeming	4-23	3-17
Leuchars	4-19	3-23

You can see that Heathrow has the highest recorded value of temperature; this is most likely influenced by the fact that the biggest UK airport is in Heathrow.

The windspeed ranges are fairly similar amongst the UK locations with Leuchars having some higher values. We can suggest that this is because of its close proximity to the sea. The location of Leuchars could also suggest that it is likely to be cool (due to the effect of the sea breeze).

The max and min temperature in Leuchars is lower than the max and min temperature of the other locations and this supports our suggestion. However, we would need to rule out the fact that this range is not skewed by extreme values and you could perhaps work out some averages and see if this further supports the suggestion or not.

This is one way to approach your studies of the data set. You should look at the locations and make predictions based on their geography. You can then analyse the data to see if your predictions are supported or not; if not, suggest reasons why.

Overseas locations

Jacksonville

This is located in Florida in the USA, close to the sea. It is in the Northern hemisphere. We expect high temperatures here. More rainfall here than in the UK and moderate windspeeds.

Beijing

This is located in China and is in the Northern hemisphere. High temperatures are expected here with values similar to Jacksonville.

However, we expect a lower average temperature than Jacksonville and much more variation in the temperature we see (**Task**: Does the data support this?).

Perth

This is located in Australia. A key difference is that Perth is in the Southern Hemisphere. The significance of this is that the seasons are reversed in the Southern Hemisphere, so between June-August, it is winter. It is located next to the sea, so we expect high windspeeds here.

Note
The overseas locations do not have information on all of the variables. For these locations, there is only data on daily mean temperature, daily total rainfall, daily mean windspeed and daily mean pressure.

Example Questions

Example 1

Source: Edexcel Sample Assessment Materials

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.

Comments

This question tests your knowledge of the large data set in part (b). You need to appreciate that the large data set has gaps to score the mark here. While this may initially seem like a 'standard' answer that you can rote-learn, not all the data sheets in the LDS have gaps so it is not a good idea to simply try to rote-learn answers to the 'obvious' questions. **You will need to study the data for each location carefully.**

Example 2

Source: Edexcel Sample Assessment Materials

[Part of question omitted.]

Using the same 9 days a location from the large data set gave $\overline{t} = 27.2$ and $\overline{w} = 3.5$.

(d) Using your knowledge of the large data set, suggest, giving a reason, the location that gave rise to these statistics.

Comments

This part of the question gives the mean daily mean temperature and mean daily mean windspeed for a location in the large data set. You are expected to be able to figure out what this location is. It should be clear that the given mean daily mean temperature is high, too high for the UK. This means the location is overseas and we can rule out Perth due to the high temperature. The mean daily mean windspeed is on the low end, so Beijing is an appropriate suggestion. **You will need to know how the values for each variable compare between the locations.**

Summary

This document is only intended to give a brief introduction to the data set and a few examples of what is required. Our main advice is that you treat familiarisation with the data set as topic in its own right.

Some key points to remember in your study:

- Conventions in data entry in the data set, eg unavailable data entered as 'n/a', low rainfall values up to 0.05 mm as 'tr', windspeeds are in knots and so on.
- The locations in the large data set and their geographical significance, eg Perth being in the Southern hemisphere and close to the sea and the significance of this in relation to the expected climate. Such knowledge can help you with the 'suggest' questions.
- Understanding how the data for each variable can be presented and what investigations can be made. For example, it would not be appropriate to represent cloud cover using a histogram, but we can potentially investigate how cloud cover is distributed (i.e. does it follow a binomial model?). This approach can help prepare you for potential exam questions.
- Find outliers within each data set. For example, if you investigate rainfall in the UK, you will discover that most values of rainfall are quite low/close to 0; higher values are rare. How high does a value need to be to be considered an outlier?

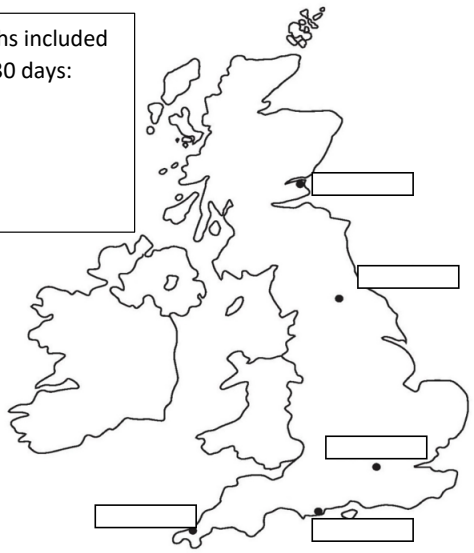
Of course, this is not an exhaustive list but it can hopefully get you started/thinking. To help you further, you can find some practice questions on the LDS on our website. Feel free to tweet us any further questions/queries or send us a message.

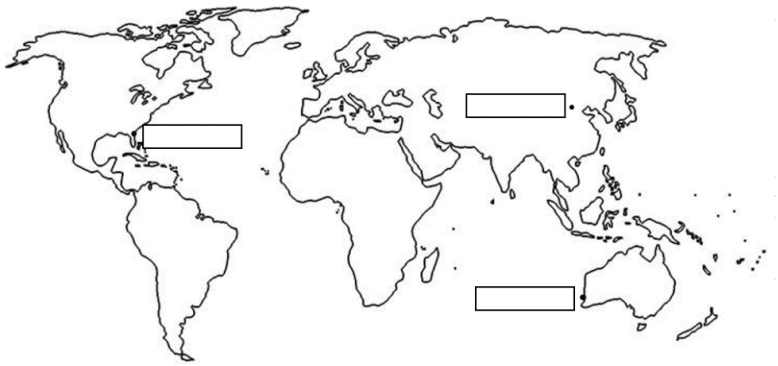
Large Data Set

Months included: _____ to _____

Years included: _____ & _____

Months included with 30 days:





Cities included

1. Beijing
2. Camborne
3. Heathrow
4. Hurn
5. Jacksonville
6. Leeming
7. Leuchars
8. Perth

Data recorded

Daily Mean Temperature	Units and accuracy	n/a means	Range of times covered	Range of temperatures
Daily Total Rainfall	Units and accuracy What does tr mean?	Max rainfall UK/worldwide	Range of times covered	
Daily Total Sunshine	Units and accuracy (how is it measured)		Max/min	
Daily Maximum Relative Humidity	What is it?		How is it measured?	
Wind	What is measured and how (4 types)?	How are they measured?	Max/min UK/worldwide	
Visibility	What is it?		How is it measured?	
Pressure	How is it measured?		Max/min	
Daily mean total cloud	What is it?		What units are used to measure it?	

Daily Mean Temperature

Which months have the highest average temperature in the UK?	Which locations are warmest in the UK?	Which locations are coldest in the UK?
Which locations are warmer worldwide?	Which locations are colder worldwide?	Are there any cities that have very different weather to the others?
What range of temperatures do you get in the UK? Worldwide?	Is there a difference between 1987 average temperatures and 2015?	

Daily Total Rainfall

Which months are wettest for UK cities?	Is there a difference in rainfall between 1987 and 2015	Do all cities follow a similar pattern for rainfall? Are there any with very different patterns?
Which of the cities is the wettest in the UK? What about worldwide?	Which of the cities is the driest in the UK? What about worldwide?	

Windspeed

Typical average windspeed	Windiest months UK? Any difference with worldwide cities?	Any differences between 1987 and 2015?

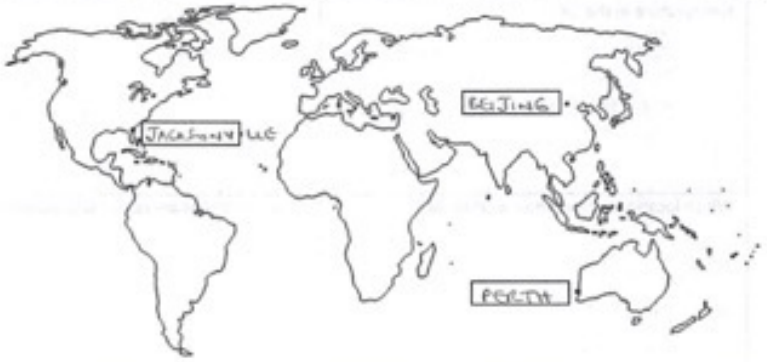
t.jones@shhs.gdst.net

Large Data Set

Months included: MAY to OCTOBER

Years included: 2015 & 1987

Months included with 30 days: JUNE, SEPTEMBER



Cities included

1. Beijing
2. Camborne
3. Heathrow
4. Hurn
5. Jacksonville
6. Leeming
7. Leuchars
8. Perth

Data recorded

Daily Mean Temperature	Units and accuracy: °C, $\frac{1}{10}$ °C	n/a means: Not available	Range of times covered: 24 hours 0900 – 0900	Range of temperatures e.g. Heathrow: 8 – 28.7°C
Daily Total Rainfall	Units and accuracy: mm, $\frac{1}{10}$ mm What does tr mean? Trace; less than 0.05mm	Max rainfall UK/worldwide: 51.6mm Heathrow 26/8/15 102 mm Perth 17/5/15	Range of times covered: 24 hours 0900 – 0900	
Daily Total Sunshine	Units and accuracy (how is it measured): Hours, $\frac{1}{10}$ Hour	Max/min Heathrow 2015: 15 hrs, 0 hrs.		
Daily Maximum Relative Humidity	What is it? Amount of water vapour in the air.	How is it measured? As a percentage of maximum water vapour that could be in the air		
Wind	What is measured and how (4 types)? Daily mean windspd; " " " direction; " max gust; " " " direction	How are they measured? Knots / Beaufort. Degrees, NSEW. Knots. Degrees, NSEW	Max/min UK/worldwide 2015. MAX 23 kns Leuchars 6/6/15. MIN 2 kns Hurn 3/10/15. MAX 14.1 kns Perth. MIN 0.9 kns Jacksonville	
Visibility	What is it? Greatest horizontal distance at which an object can be seen in daylight	How is it measured? Dm Decametres (= 10 metres)		
Pressure	How is it measured? hPa = hectopascals	Max/ming Heathrow 2015: 1036 hPa, 989 hPa		
Daily mean total cloud	What is it? How many eighths of the sky is covered	What units are used to measure it? OKtas		

CONSIDERING 2015

KEY
C – Camborne
H – Heathrow
Hu – Hurn
Lee – Leeming
Leu – Leuchars
B Beijing
J Jackson
P Perth.

Daily Mean Temperature

Which months have the highest average temperature in the UK?

	M	J	J	A	S	O
C	11.2	13.8	(15.7)	15.4	13.6	12.2
H	13.2	16.8	(18.8)	18.1	14.4	12.5
Hu	11.9	15.1	(16.8)	16.2	13.3	11.3
Lee	10.4	13.3	15.1	(15.6)	11.9	10.9
Leu	9.5	12.8	14.9	(15.1)	12.1	9.2

Which locations are warmest in the UK?

Based on the mean temperature

Heathrow	15.6°C
Hurn	14.1°C
Camborne	13.6°C
Leeming	12.7°C
Leuchars	12.2°C

Which locations are coldest in the UK?

LEUCHARS 12.2°C

Which locations are warmer worldwide?

Based on the mean temperature

J 24.8°C
B 22.6°C
P 15.2°C

Which locations are colder worldwide?

P 15.2°C outside UK
Leu 12.2°C including UK

Are there any cities that have very different weather to the others?

max temp Beijing 32.5°C
Cf max Leuchars 19.2°C.

What range of temperatures do you get in the UK? Worldwide?

UK All locations
3.8 – 28.7°C

Is there a difference between 1987 average temperatures and 2015?

	2015	1987	1987→2015
H	15.6	14.6	+1°C
Hu	14.1	13.7	+0.4°C
C	13.6	13.3	+0.3°C
Lee	12.7	12.5	+0.2°C
Leu	12.2	11.5	+0.7°C

Daily Total Rainfall

Which months are wettest for UK cities?

	M	J	J	A	S	O
C	2.1	1.3	3.2	(5.6)	2.3	1.9
H	1.4	0.4	2.3	(3.8)	1.7	1.3
Hu	2.2	1.0	1.9	(4.5)	2.5	2.2
Lee	2.5	0.7	2.6	2.2	1.3	(2.8)
Leu	2.5	1.2	(3.6)	1.7	1.3	1.6

Is there a difference in rainfall between 1987 and 2015 considering mean rainfall in UK

	2015	1987	1987→2015
C	2.8	2.7	Higher
H	1.8	2.6	Lower
Hu	2.4	2.0	Higher
Lee	2.0	2.1	Lower
Leu	2.0	2.3	"

Do all cities follow a similar pattern for rainfall? Are there any with very different patterns?

UK locations have a similar pattern; varied, peaking at similar times. Worldwide peaks occur at different times.

(See graphs on Excel)

Which of the cities is the wettest in the UK? What about worldwide? Based on mean rainfall

Camborne, UK 2.8mm
Jackson, Worldwide 5.9mm

Which of the cities is the driest in the UK? What about worldwide?

Heathrow, UK 1.8mm
Heathrow, Worldwide 1.8mm
(Just B, J & P then Beijing at 2.1mm)

Windspeed

Typical average windspeed

	Kn
C	9.4
H	8.0
Hu	7.7
Lee	7.4
Leu	8.7

Windiest months UK? Any difference with worldwide cities? MAY

	M	J	J	A	S	O
C	(10.8)	9.5	10.1	8.6	9.0	8.3
H	(9.8)	8.5	9.1	7.3	6.8	6.4
Hu	(9.1)	8.1	8.3	6.9	6.9	6.8
Lee	(9.4)	8.1	7.4	7.7	6.0	6.1
Leu	(11.5)	9.3	8.8	8.7	6.8	7.1
B	(5.5)	4.6	4.0	3.8	3.6	4.0
J	6.0	5.1	5.2	4.6	4.8	(6.1)
P	7.9	8.3	7.4	7.8	(8.8)	8.7

All may except Jacksonville & Perth; Oct & Sept respectively

Any differences between 1987 and 2015?

2015
Average windspeed
Heathrow 8.0 kns
1987 6.0 kns
Windspeed greater for Heathrow in 2015.

Statistics – Starter for 10

- Which is further North: Hurn or Cambourne?
- Which location in the UK is highest?
- What scales are used for measuring windspeed?
- How are snow or hail measured?
- Why can't we test the hypothesis "It's sunnier in the UK than in Perth" ?
- Why is it difficult to compare the temperature in Heathrow and the temperature in Beijing?
- Find some discrete data
- Find some categorical data
- Why might the Daily Mean Wind Direction not be a reliable indicator?
- What does a wind direction of 10 mean?
- What does "tr" mean?
- What is "normal air pressure" (not in the data set)

Statistics – Starter for 10 answers

- Which is further North: Hurn or Cambourne?
- *Hurn*

- Which location in the UK is highest?
- *Cambourne 87m*

- What scales are used for measuring windspeed?
- *Beaufort conversion, knots*

- How are snow or hail measured?
- *Melted first and then measured as rain*

- Why can't we test the hypothesis "It's sunnier in the UK than in Perth" ?
- *No sunshine data for overseas stations*

- Why is it difficult to compare the temperature in Heathrow and the temperature in Beijing? UK:
- *Heathrow: Highest Daily Temp, Beijing: Daily Average Temp*

- Find some discrete data:
- *Cloud Cover, Wind Direction of 20*

- Find some categorical data:
- *Wind Cardinal Direction (NNW)*

- Why might the Daily Mean Wind Direction not be a reliable indicator?
- *Doesn't give an indication of how variable the wind has been.*

- What does a wind direction of NE mean?
- *Wind is blowing <u>FROM</u> the North East and not towards the North East*

- What does "tr" mean?
- *A trace of rain but not enough to measure*

- What is "normal air pressure" (not in the data set)
- *1013 hPa*

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.

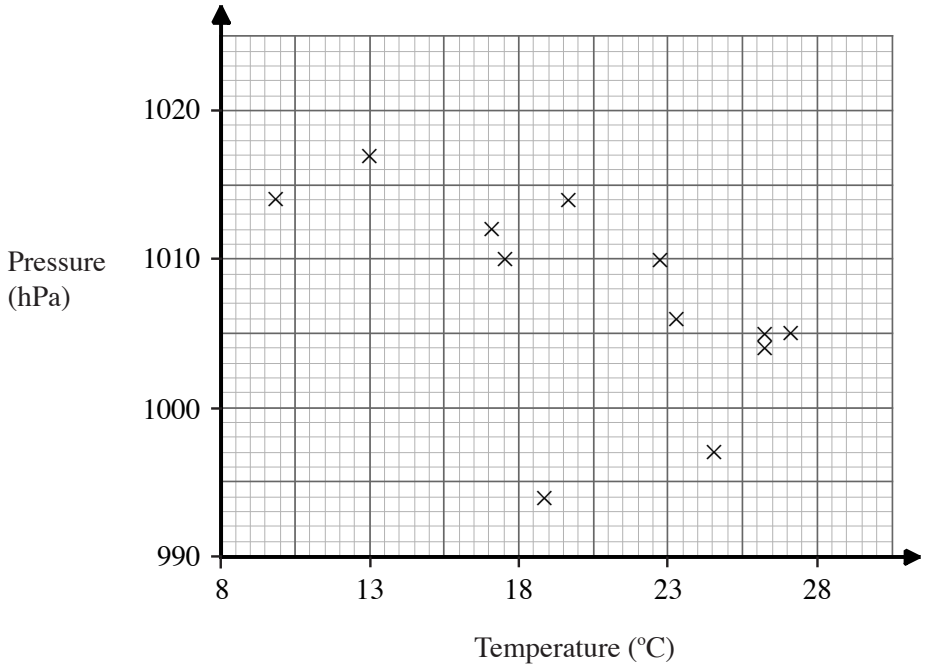
He takes a random sample of 12 days from May 2015 for Beijing from the large data set.

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

A Level
Large Data Set

Practice Questions

Question	Scheme
1	
(a)	Mean rainfall = $\frac{0.025(7)+1.5(4)+3(3)}{7+4+3}=\underline{1.083...}$ mm
	Standard deviation = $\sqrt{\left(\frac{0.025^2(7)+1.5^2(4)+3^2(3)}{7+4+3}\right)-(1.083...)^2}=\underline{1.182...}$ mm
(b)	Rainfalls between (about) $0 \le r \le 2.26$ mm are within one standard deviation of the mean
(c/i)	Not suitable because her sample only consisted of 14 days from one location and from one month
(c/ii)	e.g. Use more data from more **UK** locations and months **Must reference UK locations**

Question	Scheme
2	
(a)	e.g. Cloud cover **Accept 'Daily mean windspeed on the Beaufort scale'**
(b)	• Generate (some) two digit random numbers • Enumerate the data points. For each random number chosen, select the corresponding data point on the enumerated list. • If the random number does not correspond to a data point (due to gaps or being out of range), ignore it and choose another. If a random number is repeated, ignore it and choose another • Continue in this way until 15 data points are chosen
(c)	Not reliable since he only used 15 data points, which is unlikely to be a good representation of the weather in Lemming in 2015
(d)	The large data set only contains data for the months May-October and not the whole year

Question	Scheme
3	
(a/i)	The large data set contains data for the months May-October and there are 184 days between (1^{st}) May and (31^{st}) October.
(a/ii)	e.g. The large data set contains gaps
(b)	(Starting from 1^{st} May), each day the total amount of rainfall in Leuchars in 2015 decreases by 0.0027 mm
(c)	$x = 3 \Rightarrow T = 16.551 - 0.0027(3) = \underline{16.5429}$
(d)	*Idea that* The daily mean rainfall in Leuchars (in 2015) does not decrease at a steady rate, but fluctuates **IGNORE** references to 'extrapolation' – the question asks for discussion about the unreliability for any day in Leuchars in 2015, not just those outside of the data range

Question	Scheme
4	
(a)	Simple random sampling
(b)	e.g. easier/quicker/etc. to process / analyse/etc. the data since the large data set has a lot of data points
(c)	4
(d)	*Idea that* The large data set does not contain information on cloud cover for Beijing

Question	Scheme
5	
(a)	Temperature outliers are $T < 5$ and $T > 37.8$ Pressure outliers are $p < 993$ and $p > 1025$ But all values of T are between 9.7 and 27.2, so there are no temperature outliers and all values of p are between 994 and 1017, so there are no pressure outliers
(b)	Negative **Ignore quantifiers e.g. 'weak, strong'**
(c)	For every 1 °C increase in the temperature (in 2015 in Beijing), the pressure decreases by 0.71
(d)	$p = -0.71(8.5) + 1022 = \underline{1015.965}$ hPa
(e)	Unreliable because the large data set only contains data for May-October, and so December is outside of the data range (used to produce the regression line) **Accept 'extrapolation' but they must make reference to the fact that the LDS only contains data for May-October**
(f)	Use a greater number/sample size of days from Beijing **No marks for reference to 'Asia'/'other places', since Beijing is the only place in Asia the LDS has data for**
(g)	*Idea that* he should consider data from other places in Asia, but Beijing is the only place in Asia that the LDS has data for

Question	Scheme
6	
	$8.45 \times 1.15 = \dfrac{194.35}{n}$ $\Rightarrow n = \dfrac{194.35}{8.45 \times 1.15} = 20$ so the size of Zain's sample is 20 **[Here it has been used that 1knot = 1.15 mph]**