

# 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

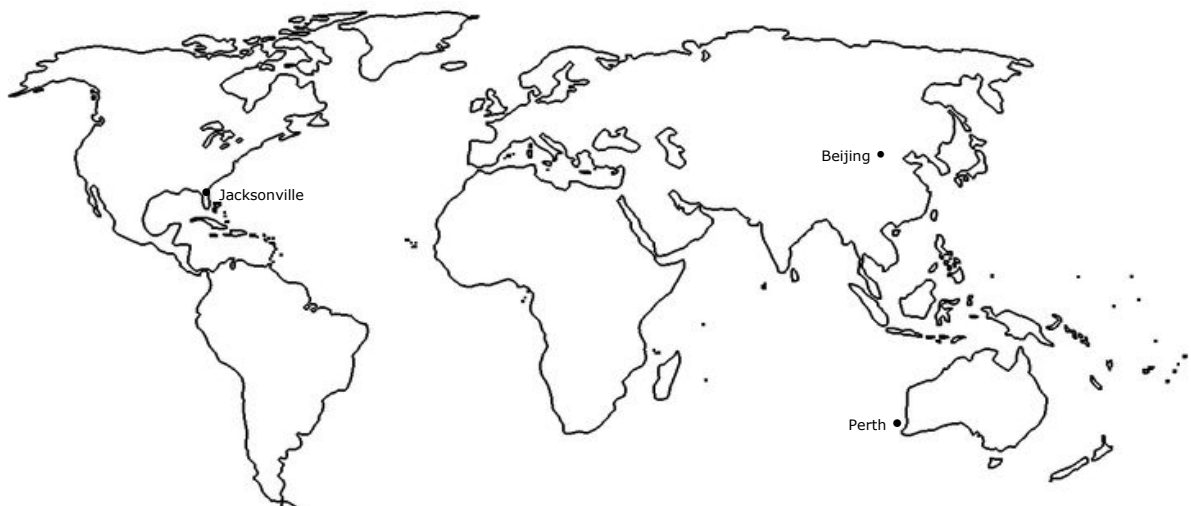
### Study Tip

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 millimetres. 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 meteorology 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  $\bar{t} = 27.2$  and  $\bar{w} = 3.5$ .

(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.

2019 © crashMATHS Limited