

Year 12 Induction Day 2nd July 2019

The Large Data Set

One part of the Statistics portion of your A Level in Maths is to become familiar with a large data set. For Edexcel, this data set is a collection of data about the weather at eight different locations from 1987 and 2015. In order to prepare for your A Level examination, you will need to work through this booklet over the summer and bring it with you to your first lesson in September to show your teacher.

Getting Started

When you join the school in September, you will be given a google classroom login and will be invited to a Maths page in which you can access the large data set. In the meantime, you can find the large data set at <https://qualifications.pearson.com/en/qualifications/edexcel-a-levels/mathematics-2017.coursematerials.html>. Since you will need to work on and interact with the spreadsheet, you will need to download a copy and save it on your personal drive/cloud/desktop.

Spend 10 minutes or so looking at the different tabs on the large data set. Are there any questions you want to ask?

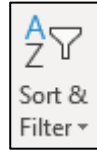


Beat the Teacher! Can you think of a question related to the large data set that the teacher cannot answer?

Some Useful Techniques in Excel

1. Sorting and filtering data

- Highlight the row containing the table headers (Row 6 on this large data set)
- Click on the “Sort & Filter” icon found in the “home” toolbar and select “filter”.



A small arrow in a box should appear in each cell in row 6.

- To filter out data from just one month, say, click on the small arrow at the top of the Date column, and select a single month.
- To order data, for example from the hottest day to the coolest, click on the small arrow at the top of the “daily mean temperature” column and select “sort largest to smallest”.

Go to the spreadsheet for Leuchars May – Oct 1987.

Order the data according to daily mean windspeed.

Write down the date of the windiest day by this measure and the daily mean windspeed for this day.

Select only data from June.

Write down the date of the windiest day in June and the daily mean windspeed for this day.

2. Calculating the average (mean)

- The quickest way to do this is to highlight the cells containing the data you want to find the average of, then simply look in the bottom right hand corner of the screen – “Average” refers to the mean average.
- Alternatively, to find the mean of the data in column B, say, type the following into an empty cell: `=AVERAGE(B7:B190)` and push enter.

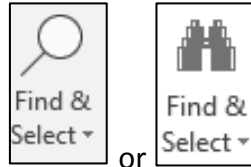
Go to the spreadsheet for Heathrow May-Oct 2015.

Use Excel to calculate the mean daily pressure.

3. Find and replace

You may notice that some rainfall readings are recorded as “tr”. The first spreadsheet will inform you of what this means, but when it comes to analysing the data, we would need to allocate a numerical number to “tr” values. We can do this quickly by following these steps:

- Click on the “Find & Select” icon in the “home” toolbar, then click on “replace”



- In the replace tab, where it says “find what” type “tr”
- Where it says “replace with”, type in a suitable numerical value (see first spreadsheet for a suitable numerical value)
- Click “replace all”

Go to the spreadsheet for Leeming May-Oct 1987.
Replace “tr” values with the figure 0.05.
Calculate the mean daily rainfall for May-Oct in Leeming, 1987. Round your answer to 2 decimal places.

Challenge

As you know from GCSE, it is sometimes more useful to work with a sample of data rather than the entire data set.

(a) With reference to this large data set in particular, suggest why a sample may be better to use than the entire data set.

(b) How could we generate a random sample on excel?

Large Data Set Quiz.

Answer the following questions from the large data set:

1. Who has provided the weather data in the large data set?

2. Does the data cover weather readings all year round? If not, what data is included?

3. List the eight locations included in the large data set, starting with the furthest North and ending with the furthest South.
 -
 -
 -
 -
 -
 -
 -
 -

4. How is snow and hail measured?

5. Wind direction (measured in degrees) is measured from?

(Like a, seen previously in Maths).

6. By considering daily total sunshine, which location reported the sunniest day in 1987?

7. Daily total sunshine is measured in

8. What is meant by “tr”?

9. Pressure can be measure in two units – name them.

10. Which of the units of pressure are used in the large data set and how can this be converted to the other unit mentioned in Q9?

11. The direction of maximum gust is the direction the wind blows from/to. (Delete as appropriate)

12. Cloud cover is measured in? Describe what this measurement tells you.

13. Windspeed is measured in, and 1 of these is equivalent to?

14. What is meant by cardinal wind direction?

15. “NGR” appears on all tabs which include data. What does “NGR” mean?

16. In which month was the wettest day in the UK recorded in 2015?

17. By considering daily maximum wind gust in the location closest to WHSG, on average which month was windiest in 1987?

18. Is the same month, in the same location, windier on average in 2015?

19. What famous weather phenomenon occurred in the UK in 1987?

20. Decide whether each of the eight locations are coastal or inland? What might you expect regarding the weather of coastal towns compared to inland towns?

Coastal

- 1.
- 2.
- 3.
- 4.
- 5.

Inland

- 1.
- 2.
- 3.
- 4.
- 5.

21. Which of the eight locations are closest to the equator? What might you expect the weather to be like closer to the equator?

Exam Style Question

Alex is collecting information about weather patterns in different regions around the world.

He uses the large data set to collect some of this information.

He gains information about weather in the UK by looking at the data for Leuchars in 2015.

As part of his investigation, Alex calculates the average daily mean temperature for each month in Leuchars. His calculations are shown in the table below to one decimal place. The data are not in month order.

Month	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
Average daily mean temperature (°C)	18.0	13.8	14.0	17.3	16.3	19.1

Month *B* had the lowest daily mean temperature.

(a) Using your knowledge of the large data set, suggest, giving a reason, the month corresponding to *B*. (1)

(b) Suggest why giving Alex's calculations to one decimal place in the table is suitable. (1)

To gather information on cloud cover in the UK, he takes a simple random sample of 38 data points from Leuchars in 2015 and calculates the mean.

(c) Describe how Alex can use simple random sampling to obtain a sample of 38 data points from the large data set. (2)

The data collected by Alex is summarised in the table below.

Cloud cover	0	1	2	3	4	5	6	7	8
Frequency	2	0	3	1	5	8	5	4	10

(d) Calculate the mean of Alex's data. (1)

(e) Calculate the standard deviation of these data and state the units. (2)

(f) State **one** limitation of how Alex has used the large data set for his study. (1)

(g) Explain why Alex should not only use the large data set for the purpose of his study. (1)

(9 marks)

You will learn more about standard deviation in Year 12, but essentially it is a measure of how spread out the data is. We can calculate the standard deviation by using the following formula:

$$\sigma = \sqrt{\frac{\sum fx}{\sum f} - \left(\frac{\sum f^2 x^2}{\sum f}\right)^2}$$