

Edexcel

A Level

A Level Mathematics

Understanding the Large
Dataset (Answers)

Name:

M M E

Mathsmadeeasy.co.uk

Total Marks:

Understanding the Large Dataset- Answers

Edexcel

- 1) i) Draw a boxplot for the Daily Mean Temperatures in 1987 from all UK sites combined and compare it to a similar boxplot for 2015. [6]

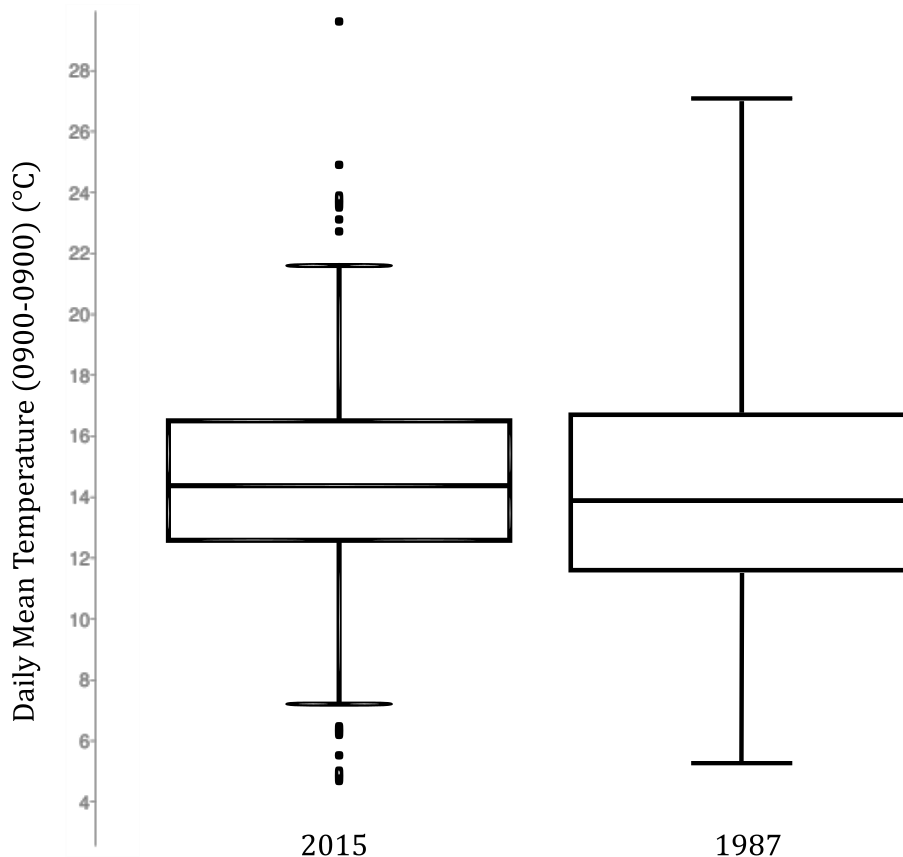
Here, the key is understanding that the Daily Mean Temperatures can be treated as observations that we can calculate a median and range of. Combining all of the 1987, 2015 data gives a UK wide sample of temperatures. [2]

[1 mark for each graph box drawn correctly- 2 max]

[1 mark for graph whiskers drawn correctly- 2 max]

[1 mark for correct axes titles (both)- 1 max]

[1 mark for outliers identified]



- ii) Comment on the distributions of the Daily Mean Temperatures.

[1 mark for comment on median/interquartile range]

The median temperature is higher for 2015 by 0.25 °C than in 1987.

[1 mark for comment on range]

The maximum and minimum values for 2015 are higher by 5 °C and lower 0.2 °C than in 1987, respectively. The extremes of temperatures are more likely to be outliers in 2015.

- 2) Summarise the number of observations, mean, median and range of Daily Total Rainfall for each of the UK sites in 1987. [15]

This is a data cleaning and summary statistics task. Those entries where rainfall was recorded as *tr* need removing as they cannot be used in this instance.

[1 mark for each location for correct number of observations - 5 max]

[1 mark for each location for correct mean- 5 max]

[1 mark for each location for correct median and range- 5 max]

	Daily Total Rainfall (0900-0900) (mm)				
	Camborne	Heathrow	Hurn	Leeming	Leuchars
No. of Observations	156	153	151	155	154
Mean	3.13	3.07	2.44	2.44	2.71
Median	0.5	0.6	0.2	0.6	0.55
Range	36.4	53.1	28.5	31.7	31.8

- 3) i) Find the variable most correlated to Daily Maximum Gust (0000-2400) (kn) in the Heathrow 1987 data.

[2 marks- 1 mark for correctly removing n/a data and 1 mark for correlation]

Daily Mean Windspeed (0000-2400) (kn) is correlated to Daily Maximum Gust with a correlation coefficient of 0.852391112.

- ii) Use this as the independent variable in a linear regression model, predicting the Daily Maximum Gust. State this model.

[2 marks- 1 mark for correct intercept, 1 mark for correct gradient]

$$Gust = 2.59 + 5.06 \times Windspeed$$

- iii) Work out the mean absolute error between the predicted and actual Daily Maximum Gust.

[1 mark]

$$Mean\ Absolute\ Error = \frac{\sum |Actual - Predicted|}{n}$$

where n is the number of observations.

$$Mean\ Absolute\ Error = 2.50\ (to\ 2dp)$$

- iv) Use the same model to predict the Daily Maximum Gust for Heathrow 2015. Then work out the mean absolute error between the predicted and actual Daily Maximum Gust.

[2 marks]

$$Mean\ Absolute\ Error = 4.820003563$$

- v) What is the difference in error between 1987 and 2015. What might be the reason for this?

[1 mark]

The difference in error is 2.32. The reason for this is that in 2015 the relationship between maximum gust and average windspeed has a higher correlation (much more linearly related).

- 4) On the following sketch of the UK draw an arrow showing mean Daily Mean Wind Direction (o) for each airport in 2015.

[5 marks- 1 for each correctly drawn arrow]

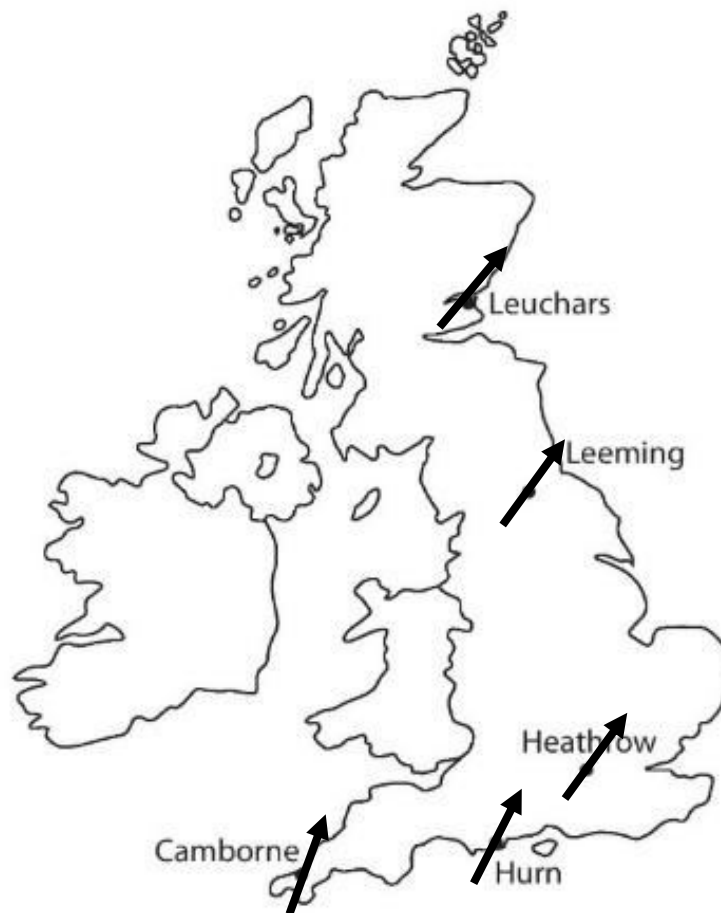
As it is written in the key, wind direction is given in the direction the wind has come from and as a result the arrow should be facing the opposite way (-180°).

The mean Daily Mean Wind Direction for each is shown below.

Camborne	Heathrow	Hurn	Leeming	Leuchars
200.3804	206.7935	207.3913	216.6848	220.3261

You should see, and note, that the average wind direction for all sites is south-westerly.

We know that the prevailing wind in the UK, contributing to warmer weather than expected at this latitude and frequent rain, is from the south-west



5) Plot and comment on the 7-day moving average Rainfall (24 hour total) (line) and Daily Mean Air Temperature (bar) on the same graph for Beijing 1987.

- [1 mark for each axis labelled correctly – 3 max]
- [2 marks for line graph of rainfall (as includes moving average)]
- [2 marks for line graph of rainfall (as includes moving average)]
- [1 mark for key (named legend in Excel)]
- [2 marks for comment, 1 for rain, 1 for temperature]

The graph shows that the temperature peaks just below 30 degrees in the middle of July. The temperature falls away into October at around 10 degrees. The rainfall is inconsistent, most of the rain occurs in the summer months, peaking in August. There is no rain in October.

