Getting real with statistics

A-level MATHS

MathsConf 31

Resource booklet

Published: Spring 2023

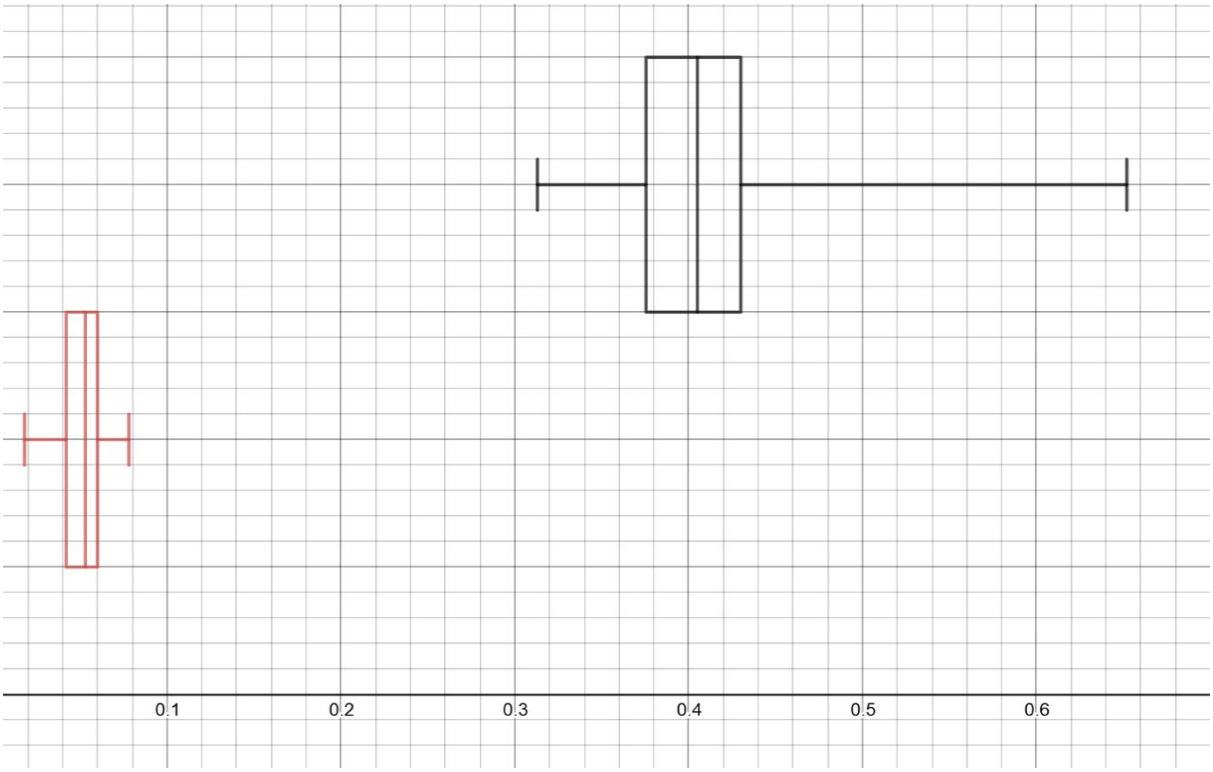
The content of this training course contains no reference to future exam content as far as we know at the time of production.

Contents	Page

Introductory activity

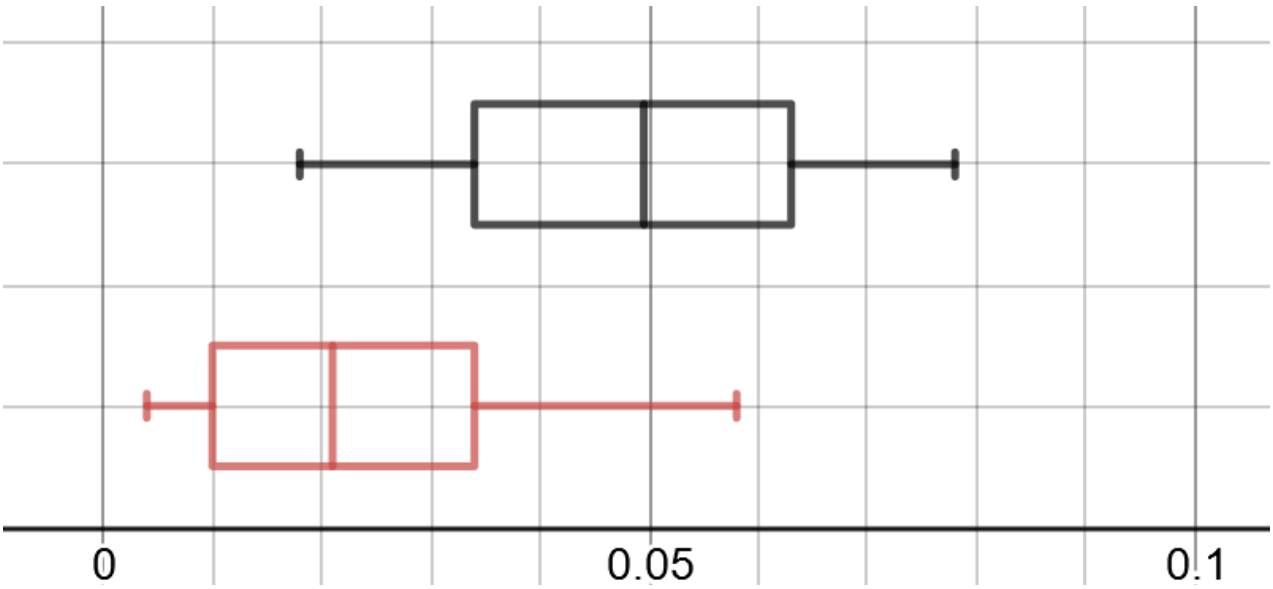
NOX emissions for diesel cars registered in NW in 2002 (above) and in 2016 (below)

- An outlier is defined as a value more than 1.5 times the IQR below Q1 or above Q3
- Determine whether either of the samples represented in these boxplots contains any outliers.
- Using information from the boxplots find an appropriate value for the percentage reduction in NOX emissions between 2002 and 2016.
- Rose claims that these graphs show there was a very large reduction in the NOX emissions of cars in the UK.
- What assumptions would Rose have to make to support her claim, based on these graphs?



NOX emissions for cars registered in NW in 2016 – diesel (above) and petrol (below)

- Compare the distributions of NOX emissions for diesel and petrol cars.
- Julian claims that, based on this evidence, petrol engines are much better for the environment than diesel engines. He says that the proposed ban on diesel engines is justified.
- What assumptions must Julian make, based on this evidence, to support his claim?



Activity 1

The following worksheets and activities are extracts from AQA's LDS Teaching Guidance.

Worksheet LDS1: Exploring the LDS

1 How many cars are included in the LDS?

2 What makes of car are included and why?

3 How many electric powered cars are there?

4 How many cars have owners in the North West region?

5 How many cars have female registered keepers?

6 How many are company cars?

7 What is the largest engine size?

8 How many cars were registered in 2012?

9 What makes are the cars with the lowest carbon dioxide emissions?

10 What is the highest level of carbon monoxide emissions?

11 How many vehicles have nitrogen oxide emissions less than 0.01 g/km?

12 What is the highest level of particulate emissions?

13 What is the actual mass of the car with the largest mass?

14 What make is the car with the highest hydrocarbon emissions?

15 Jamie drives a BMW to work and home each day. Each journey is approximately 40 miles. Estimate the total CO2 emissions from Jamie's car per annum by finding the average CO2 emissions for all BMWs in the LDS.

Activity 2

Worksheet LDS 2: Sampling from the Cars LDS

1 Take a sample of 50 diesel powered cars.
   a. For how many of these cars is there data on their particulate emissions.
   b. Use this data to make an estimate of the median particulate emissions of diesel cars in the LDS.

2 Take a sample of 100 petrol powered Fords.
   a. Find the mean and standard deviation of the carbon monoxide emissions of these cars.
   b. Use your values to make estimates of the mean and variance of the carbon monoxide emissions of all petrol-powered Fords in the LDS.

3 In the LDS 26% of cars are registered in London, 30% in the North West and 44% in the South West. Petra wants to take a stratified random sample of 375 cars from the LDS with her strata being the three regions.

How many cars should she select from each region?

4 The table shows the number of cars in the LDS broken down by make and year of registration.

Make	2002	2016
BMW	141	410
Ford	407	682
Toyota	154	121
Vauxhall	260	809
Volkswagen	337	506

Charley is taking a quota sample of cars from the LDS. She will take a sample of 50 for each year of registration, with 10 of each make.

   a. Explain whether Charley's sample is representative of the population.

Guillermo takes a sample of 50 petrol-powered Toyotas from the LDS. Otis takes a sample of 100 petrol-powered Toyotas.

Guillermo calculates the mean CO_2 emissions to be 163 g/km; Otis calculates a value of 156 g/km.

   b. Using these values give an estimate of the mean CO_2 emissions of all petrol-powered Toyotas in the LDS.

The known mean of all 174 petrol-powered Toyotas in the LDS is 159 g/km.

   c. Give, with a reason, an estimate of the mean CO_2 emissions of all Toyotas registered in England.

Activity 3

Investigating hypotheses and interpreting results

Take a hypothesis and investigate it.

Example

The mean CO2 emissions of cars registered in 2016 is lower than the mean CO2 emissions of cars registered in 2002.

Using the filter arrows select the cars registered in 2002. Select column J. The mean CO2 emissions is displayed in the status bar at the bottom of the sheet as the Average.
Repeat this for cars registered in 2016.

For 2002 the mean is 171, for 2016 the mean is 120.

Make a comment to interpret this result.

You will expect students initially to make imprecise comments such as:
It's lower.
The average has gone down.

You want them to make more precise statements such as:
The mean CO2 emissions for cars registered in 2016 is 120 g/km, which is much lower than the mean CO2 emissions for cars registered in 2002 which is 171 g/km.

You are training students to use precise language to reach correct conclusions – echoing the words used in the hypothesis. You do not expect students to do a formal hypothesis test. However, making a judgement of the significance of the change would be worthwhile. (Percentage change would be useful.)

Whilst this LDS was chosen to highlight topical environmental issues, we are teaching students to work as mathematicians/statisticians who are trying to be objective about the conclusions they reach. We don't need emotional value judgements in answers to exam questions, but they are worth considering in the classroom.

Repeat for other hypotheses.

The LDS as a sample

The LDS is itself a sample of all data in the Department of Transport's stock vehicle database (about 40 million vehicles in the database when we requested the data – that is a large data set!) As such, the population of all vehicles in England would represent the "population." However, it is helpful to consider the LDS as a population in its own right, so that samples from it can be used to compare with the whole LDS.

(Note that technical definitions of population and sample can refer to the actual data as the population and sample rather than the objects to which the data are assigned. The carbon dioxide emissions of the vehicles in the LDS would constitute a population, from which we might take a sample. We are happy to consider this technical definition or the less formal one in which the population and sample are cars as both being correct.)

Making informal inferences about the population

Using the sample statistics produced by the LDS with Tools, simple inferences would be such things as:

The mean of CO2 emissions in my sample is 144 g/km, so I estimate that the mean CO2 emissions of all cars in the LDS is also 144 g/km.

The highest band of CO2 emissions for vehicle tax is more than 255 g/km. In my sample of 50 cars only 1 has a value in this range, so I estimate that 2% of all cars will be in this highest band.

You can do the same for any sample statistic.

We know that we can find all population parameters for the LDS and we are using samples in this context in order to learn about both sampling and the LDS. Students should recognise that we don't consider the reliability of an estimate based on how close it is to a known population parameter.

Activity One

Select a few different simple random samples selected using different criteria and write down the mean for each one. Discuss which would give the best estimate of the population mean.
E.g. a sample of 50 cars registered in 2002; a sample of 25 cars with no filters set; a sample of 100 Toyotas. Find the mean CO emissions for each sample.

Note that when selecting a sample of size n from the LDS, the tool selects n cars. If there is no data for the relevant variable for k of these cars, the actual sample size is $n - k$. This actual sample size is shown on the Sample statistics sheet.

The LDS tool and LDS worksheets are on AQA All About Maths

Notes

Contact us

Our friendly team will be happy to support you between 8am and 5pm, Monday to Friday.

Tel: 0161 957 3852
Email: maths@aqa.org.uk
Twitter: @AQA

aqa.org.uk