Topics from GCE AS and A Level Mathematics

|  |
| --- |
| * Interpret box and whisker plots and histograms for single-variable data.
* Interpret measures of central tendency and variation, extending to standard deviation.
* Recognise and interpret possible outliers in data sets and statistical diagrams.
* Select and critique data presentation techniques in the context of a statistical problem.
* Be able to clean data, including dealing with missing data, errors and outliers.
 |

**Investigation 2**

Compare the daily mean temperature and daily total rainfall for 1987 and 2015 at Heathrow for the months May – October (inclusive).

**Investigation 2 (a)**

Compare the daily mean temperature for the years 1987 and 2015 at Heathrow for the months May – October (inclusive).

**The data**

Open the Excel workbook **Pearson Edexcel GCE AS and AL Mathematics data set.xlsx.**

Select the **Information** worksheet. Read the information in cell **A12.**

1. Why is it important to know where the daily mean temperature is taken?
2. What type of variable is daily mean temperature?
3. Explain why the variable daily mean temperature is continuous whereas the data are not. (Although data are often referred to as continuous.)

Select the worksheet **Heathrow May-Oct 2015.**

*In Excel copy the data for Heathrow 2015 into a new workbook*

**

*Select the whole worksheet*

*Click on the small blue square in the left hand corner*

***Right click*** *then* ***Copy***

*Open a new workbook*

*Select the* ***Office button*** *then* ***New***

*Then double click on* ***Blank Workbook***

*Select* ***A1*** *right click* ***Paste***

*Rename the worksheet*

*Double click on the tab* ***Sheet 1*** *at the bottom of the worksheet*

*Type* ***Heathrow2015,*** *then click anywhere on the worksheet*

*It is difficult to analyse these data if the headers are not in row 1.*

*Delete rows 1 – 5*

*Select rows 1 – 5 right click* ***Delete***

*Save workbook as* ***Heathrow***

*Copy the data for Heathrow 1987 into the workbook Heathrow on a new worksheet*

*In the workbook* **Pearson Edexcel GCE AS and AL Mathematics data set.xlsx***select the worksheet* ***Heathrow May-Oct 1987***

*Copy the whole worksheet*

*In the workbook* ***Heathrow*** *select* ***Sheet 2***

*Paste in the data*

*Rename the worksheet* ***Heathrow 1987***

*Delete rows 1 – 5*

***Save your work***

*Copy the data for date, daily mean temperature and daily total rainfall for Heathrow 1987 and 2015 into a new worksheet*

*Copy and paste columns* ***A, B*** *and* ***C*** *for* ***Heathrow 1987*** *and* ***2015*** *into* ***Sheet 3***

*Rename the worksheet* ***TempRain***

*To help with labelling the box plots rename the columns* ***Temperature 1987, Rainfall 1987,******Temperature 2015*** *and* ***Rainfall 2015****.*

1. Are there any missing values or coded entries for daily total rainfall?

**Process**

The distributions of the variables can be compared using box plots or histograms. Whichever graphs are plotted it is important that the same scales are used.

Plot box plots to compare the daily mean temperatures in Heathrow for 1987 and 2015 May – October.

It is tricky to plot box plots in Excel. (See Appendix 1, ‘**Box plots in Excel’,** for instructions on how to do this.) Here the box plots are plotted using GeoGebra.

*Plot a box plot in GeoGebra*

*Open* ***GeoGebra*** *select* ***View*** *then* ***Spreadsheet***

*Copy and paste the data for daily mean temperature for Heathrow 1987 into Geogebra*

*In the workbook* ***Heathrow*** *select the worksheet* ***TempRain***

*Select column* ***B***

*Right click* ***Copy***

*In* ***GeoGebra*** *select* ***A1*** *and**right click* ***Paste***

*Copy and paste the data for daily mean temperature for Heathrow 2015 into Geogebra*

*In the workbook* ***Heathrow*** *select the work sheet* ***TempRain***

*Select column* ***E***

*Right click* ***Copy***

*In* ***GeoGebra*** *select* ***B1*** *and**right click* ***Paste***

*Plot box plots*

*In* ***GeoGebra***

*Select columns* ***A*** *and* ***B***

*Click on the bottom right hand arrow of the histogram icon*

*Select* ***Multiple Variable Analysis***

**

*Click on the  icon select* ***Use Header as Title*** *then* ***Analyse***

*(Note: if the  icon cannot be found delete the headers in A1 and B1.)*

*Right click on the graph and select* ***Copy to Graphics View***

*Close the original box plot window*

*Tidy up the graph and label*

*Right click on* ***a = 15.3*** *(this is on the left hand side of the main window)**click on* ***Show Label***

*a =15.3 will have been removed from the boxplot.*

*Repeat this for* ***b = 14.55***

**

*Right click on the graph area, not on the boxes.*

*Select* ***Graphics*** *(if this does not appear in the menu at first try right clicking in different places in the graph).*

*Select the* ***Basic*** *tab*

***x Min*** *enter* ***0***

***xMax*** *enter* ***30***

*Select the* ***xAxis*** *tab*

*In* ***Label*** *enter* ***Degrees C***

**

*Select the* ***yaxis*** *tab*

*Untick* ***Show yAxis***

*Close the* ***Graphics*** *window*

*Copy the graph into Word*

*Select the* ***Edit*** *tab then* ***Graphics View to Clipboard***

*In the Word document right click where the box plot is to be pasted* ***Paste***

*Add the title above or below the boxplot:*

***Daily mean temperature for Heathrow May to October***

**

*GeoGebra shows outliers as small crosses.*

1. Comment on any outliers.
2. Inspect the data to decide if the outlier should be included or excluded.
3. Identify the outlier.



1. In the worksheet Heathrow2015 look at the values for other variables on, just before and just after the day the outlier occurred to see if they are consistent with a hot day.
2. Look at the temperatures for other areas in the UK for that period. Comment on your findings giving examples.
3. Do a search on the internet to see if there is any mention of a heat wave in the UK during June/July in 2015.
4. Should the outlier be included in the data? Give a reason for your answer.
5. Comment on the shape of the distributions for the two variables.
6. Calculate and compare measures of location and spread for the two distributions.

To do this, use the functions in Excel to complete the summary statistics table.

|  |
| --- |
| Daily mean temperature oC |
|  | 1987 | 2015 |
| Mean |  |  |
| Standard deviation |  |  |
| Minimum |  |  |
| Lower quartile,Q1 |  |  |
| Median, Q2 |  |  |
| Upper quartile, Q3 |  |  |
| Maximum |  |  |
| Interquartile range, IQR |  |  |



Copy the above table and paste this into **H1:J10** in the worksheet **TempRain**. Increase the width of column **H** by double clicking on the edge of the column **H** .

*Calculate summary statistics in Excel*

*To enter a function in Excel you must start with* ***=***

*Calculate a mean*

*Select* ***I3***

*Type* ***=AVERAGE(B:B)*** *then* ***Enter***

*The function AVERAGE calculates the mean for an array of data.*

*Calculate a population standard deviation*

*Select* ***I4***

*Type* ***=STDEVP(B:B)***

*Calculate quartiles*

*The function* ***QUARTILE*** *calculates the three quartiles for an array of data.*

*As you enter the function a message appears telling you what the function will do. As you enter an open bracket a message will appear to let you know in what format to enter the information.*

*The function* ***Quartile*** *asks for an array and a quart.*

*The array is the column with the data.*

*The quart is to let Excel know which quartile to calculate.*

*Enter 1 for the lower quartile (Q1)*

 *2 for the median (Q2)*

 *3 for the upper quartile (Q3).*

*The interquartile range (IQR) is given by: Q3 – Q1.*

*Continue by entering the functions as shown in the table.*

**

*Reduce all the statistics in the table to one decimal place value*

*In the* ***Home*** *tab select the statistics* ***I3:J10*** *and click on the icon the appropriate number of times.*

***Save your work***

Complete the table

|  |
| --- |
| Daily mean temperature oC |
|   | 1987 | 2015 |
| Mean |  |  |
| Standard deviation |  |  |
| Minimum |  |  |
| Lower quartile,Q1 |  |  |
| Median, Q2 |  |  |
| Upper quartile, Q3 |  |  |
| Maximum |  |  |
| Interquartile range, IQR |  |  |

1. Explain why the median is almost equal to the mean for both distributions.

**Report**

1. Compare the measures of location and spread for daily mean temperature for Heathrow in 1987 and 2015.

**Investigation 2 (b)**

Compare the daily total rainfall for 1987 and 2015 at Heathrow for the months May – October (inclusive).

**The data**

Open the Excel workbook **Pearson Edexcel GCE AS and AL Mathematics data set.xlsx.**

Select the **Information** worksheet. Read the information in cell **A13.**

1. What rainfall unit measure is used?
2. What does ‘tr’ represent?
3. What does ‘n/a’ represent?
4. Are there any missing values for daily total rainfall?
5. How many coded entries are there for daily total rainfall in 1987 and 2015?

*Use the Excel COUNTIF function to count how many cells have the code tr*

*In the workbook* ***Heathrow*** *select the worksheet* ***TempRain*** *then select* ***G2*** *type* ***=count***

*A list of COUNT functions will appear.*

*Double click on* ***COUNTIF***

*This function counts the number cells within a range that meet a given condition.*

*Select column* ***C***

*Type* ***, “tr" (****i.e. a comma followed by* ***“tr”)***

*The COUNTIF function will count the number of* ***tr****s in column C. The quotation marks are needed as* ***tr*** *is text and not a number.*

*Repeat this for column F.*

There are ……… trs for the daily total rainfall in 1987 and ……… in 2015.

1. The rainfall for the code tr is between 0 and 0.05cm. What would be a sensible value to replace the code tr with?

*Use the Replace function in Excel to replace tr with 0.025*

*Select the* ***Home*** *tab*

*Select the column for* ***Rainfall 1987*** *(column* ***C****)*

*Select* ***Find & Select***

*Then* ***Replace***

**

*Enter* ***tr*** *in* ***Find what***

*Enter* ***0.025*** *in* ***Replace with***

***Replace All***

***Close***

*Repeat for the* ***Rainfall 2015***

***Save your work***

**Process**

Use the appropriate graphs to compare the distributions of daily total rainfall for Heathrow in 1987 and 2015, May – October.

1. When are the most extreme rainfalls? Do a search on the internet to see if these extreme rainfalls were reported.
2. Comment on the shapes of the distributions for the two variables.
3. Calculate and compare measures of location and spread for the two distributions.

|  |
| --- |
| Daily total rainfall (mm) |
|   | 1987 | 2015 |
| Mean |  |  |
| Standard deviation |  |  |
| Minimum |  |  |
| Lower quartile,Q1 |  |  |
| Median, Q2 |  |  |
| Upper quartile, Q3 |  |  |
| Maximum |  |  |
| Interquartile range, IQR |  |  |

1. Which measures of location and spread to use?

**Report**

**Investigation 2 (c)**

Compare the daily maximum gust for the years 1987 and 2015 at Heathrow for the months May - October. Plot the appropriate graphs and calculate summary statistics.

**Process**

Students will have to deal with missing values. Use all the data that are available. Geogebra will ignore missing values.

Investigate any outliers.

Calculate summary statistics

|  |  |
| --- | --- |
|  | Daily maximum gust (kn)With outlier included |
|   | 1987 | 2015 |
| Mean |  |  |
| Standard deviation |  |  |
| Minimum |  |  |
| Lower quartile,Q1 |  |  |
| Median, Q2 |  |  |
| Upper quartile, Q3 |  |  |
| Maximum |  |  |
| Interquartile range, IQR |  |  |

**Report**

**Investigation 2 (d)**

A couple wish to go on a holiday some time in June or July. They cannot decide whether to visit Perth in Australia, Jacksonville in America or Beijing in China. The weather will be a deciding factor.

Based on the weather data from 2015 compare the temperatures, rainfall and windspeeds for the three cities and report to the couple your findings.

You report should include appropriate graphs and measures of location and spread as evidence to support your advice.

*Appendix 1*

*Box plot in Excel*

**

*Open the workbook* ***Boxplots.xlsx*** *and worksheet* ***HeathrowTemp.***

*Complete the table in the Excel worksheet*

**

**

***Save your work***

*Plot a stacked bar chart*

*Select* ***E1:F1*** *hold down* ***Ctrl*** *and select* ***E7:F9***

*Select the* ***Insert*** *tab then* ***Bar*** *then* ***Stacked Bar***

**

*The plot is the wrong way around.*

*Right click on the chart then* ***Select Data*** *then* ***Switch Row/Column*** *and* ***OK***

**

*Delete the legend*

*Right click on the legend and* ***Delete***

*Delete grid lines*

*Right click on grid lines and* ***Delete***

**

*Add a title and vertical axis label*

*Click on the graph select the* ***Layout*** *tab then* ***Chart Title*** *then* ***Above Chart***

*Type***Daily mean temperature for Heathrow May to October**

*and* ***Enter***

*You may like to make the font size smaller.*

**

*Select* ***Axis Titles*** *then* ***Primary Horizontal Axis Title*** *then* ***Title Below Axis***

*Type* ***Temperature (degrees C)*** *and* ***Enter***

*Type* ***1987*** *in* ***E1*** *and* ***2015*** *in* ***F1***



*Add a chart border*

*Right click on the chart select* ***Format Plot Area*** *then* ***Border Color*** *select* ***Solid line*** *open up the* ***Color*** *arrow select* ***Black*** *and* ***Close***

***Save your work***

*Add lower whiskers*

*Click on the bottom box of the bar chart (blue)*

*Select the* ***Layout*** *tab then* ***Error Bars*** *then* ***More Error Bar Options***

**

*Select* ***Minus*** *and*

***Custom*** *and* ***Specify Value***

**

 *In* ***Custom Error Bars*** *click into* ***Negative Error Value*** *delete* ***={1}*** *select* ***E10:F10*** *then* ***OK*** *and* ***Close***

***Save your work***

*Change colour of bottom box*

*Right click on bottom box and select* ***Format Data Series*** *then* ***Fill*** *then* ***No fill*** *and* ***Close***

*Add upper whiskers*

*Click on the top box of bar chart (green)*

*Select the* ***Layout*** *tab then* ***Error Bars*** *then* ***More Error Bar Options***

*Select* ***Plus*** *then* ***Custom*** *and* ***Specify Value***

*In* ***Custom Error Bars*** *click into* ***Positive Error Value*** *delete* ***={1}*** *select* ***E11:F11*** *then* ***OK*** *and* ***Close***

*Change the centre box to one colour and to add borders*

*Right click on one of the red boxes select* ***Format Data Series*** *then* ***Fill*** *then* ***Solid fill*** *open the* ***Colour*** *choices and select* ***grey***

*Select* ***Border Colour*** *then* ***Solid line*** *select* ***black*** *from the colour choice*

*Click on a green box and do the same then* ***Close***

***Save your work***



Note: It is difficult to show outliers when plotting a box plot in Excel.