Topics from GCE AS and A Level Mathematics

* Interpret scatter diagrams and regression lines for bivariate data, including recognition of scatter diagrams which include distinct sections of the population
* Understand informal interpretation of correlation.

**Investigation 4 (a)**

Investigate if there is a correlation between daily mean total cloud cover and daily mean pressure.

**The data**

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

Select the **Information** worksheet.

1. Read the information in cell **A17.**

Explain how cloud cover is calculated.

Cloud cover is the fraction of the celestial dome covered by cloud and is expressed as eighths.

1. Read the information in cell **A19.** What units are used to measure daily mean pressure?

The mean sea level pressure has been calculated from a measurement made at station level and measured in hectopascal (hPa).

1. What type of variable is daily mean pressure?

It is a continuous variable.

1. What type of variable is daily mean total cloud cover?

The data are expressed as eighths (oktas). The numerators have been converted to eighths and then expressed as whole numbers. So the variable is continuous but the data has been corrected to whole numbers.

The location Leuchars 2015 was randomly selected from:

Camborne 2015 Camborne 1987

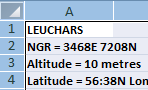
Hurn 2015 Hurn 1987

Leuchars 2015 Leuchars 1987

Leeming 2015 Leeming 1987

Heathrow 2015 Heathrow 1987.

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

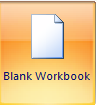
*It is difficult to analyse these data as it is presented in the dataset. The headers need to be in row 1.*

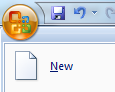
*Copy the data into a new workbook*

*Select the whole worksheet*

*Click on the small blue square in the left hand corner, this will select the whole worksheet.*

*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* ***Leuchars 2015,*** *then click anywhere on the worksheet*

*Delete rows 1 – 5*

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

*Save workbook as* ***Leuchars2015***

1. What are bivariate data?

Bivariate data are data that have two variables i.e. where each member of the sample requires the values of two variables.

In bivariate data if one of the variables is controlled (or explains the other variable), it is known as the independent (or explanatory) variable.

A dependent (or response) variable is a variable whose value depends on the value of another variable.

The dependent variable is usually plotted on the vertical axis. Note: if you are using a regression model to predict a value, the variable for the value you wish to predict should be the *Y* variable and plotted on the vertical axis of a scatter diagram.

For bivariate data if all the points in a scatter diagram seem to lie near a straight line, there is a linear correlation between the two variables.

Pearson’s product moment correlation coefficient provides a standardised measure of linear correlation. Its value lies within −1 and +1.

**Process**

Plot a scatter diagram to investigate if there is a correlation between daily mean total cloud cover and daily mean pressure. In this case there is no controlled variable. Later in this investigation a prediction for daily mean total cloud cover will be calculated based on the regression equation for daily mean total cloud cover against daily mean pressure.

1. Which variable should be plotted on the vertical axis?

Daily mean total cloud cover.

*Plot a scatter diagram in Excel*

*Open* ***Leuchars2015.xlsx’****.*

*When plotting a scatter diagram, Excel plots the variable in left hand column on the x-axis and the variable in the right hand column on the y-axis.*

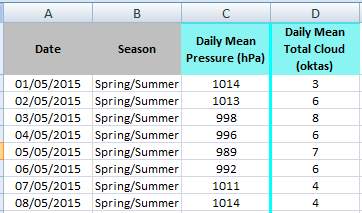
*Copy columns* ***A****,* ***I*** *and* ***K*** *into a new worksheet*

*In the worksheet* ***Leuchars2015*** *select column* ***A*** *right click* ***Copy***

*In worksheet* ***Sheet 2*** *select* ***A1*** *right click* ***Paste***

*In the worksheet* ***Leuchars2015*** *select column* ***K*** *right click* ***Copy***

*In worksheet* ***Sheet 2*** *select* ***C1*** *right click* ***Paste***

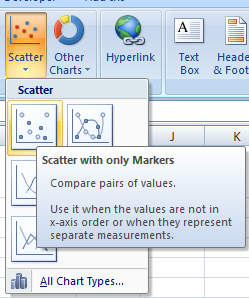
*In the worksheet* ***Leuchars2015*** *select column* ***I*** *right click* ***Copy***

*In worksheet* ***Sheet 2*** *select* ***D1*** *right click* ***Paste***

*Rename* ***Sheet 2 CloudPress***

*Select* ***B1*** *type* ***Season***

*In column* ***B*** *code the data for May, June, July and August as* ***Spring/Summer*** *and code the values for September and October* ***Autumn****. These codes will be used later.*

*Plot a scatter diagram*

*In the worksheet* ***CloudPress*** *select columns* ***C*** *and* ***D***

*Select the* ***Insert*** *tab then* ***Scatter****.*

*Select* ***Scatter with only Markers.***

*Format the diagram*

**

*Delete the legend*

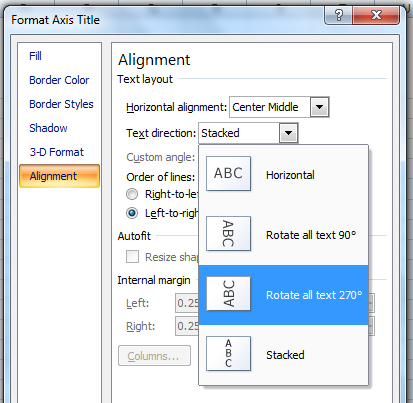
*Click on the* ***legend*** *and* ***delete***

*Add a title*

*Click on the title and type* ***Daily mean total cloud cover vs Daily mean pressure Leuchars 2015*** *then* ***Enter***

*Add a vertical axis title*

*Click on the chart then select the* ***Layout*** *tab select* ***Axis Titles*** *then* ***Primary Vertical Axis Title*** *then* ***Vertical Title*** *and type* ***Daily mean total cloud cover (oktas)*** *and* ***Enter***

**

*Change the alignment of the text in the vertical title*

*Right click on the* ***vertical title*** *select* ***Format Axis Title*** *then* ***Alignment*** *then click on the arrow next to* ***Text direction*** *click* ***Rotate all text 270o*** *and* ***Close***

*Add a horizontal axis title*

*In the* ***Layout*** *tab select* ***Axis Titles*** *then* ***Primary Horizontal Axis Title*** *then* ***Title Below Axis*** *and type* ***Daily mean pressure (hPa)*** *and* ***Enter.***

*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***



Figure 1

1. Comment on the correlation between the two variables.

The scatter diagram suggests there is a slight negative correlation between daily mean total cloud cover and daily mean pressure for Leuchars in 2015 between May and October.

1. Give an interpretation of the correlation between the two variables.

As daily mean pressure increases (rises) daily mean total cloud cover decreases (falls) in Leuchars May to October in 2015.

Or

As daily mean pressure decreases (falls) daily mean total cloud cover increases (rises) in Leuchars May to October in 2015.

1. Comment on the scatter.

The higher the daily mean pressure the more the scatter.

*Find the product moment correlation coefficient in Excel*

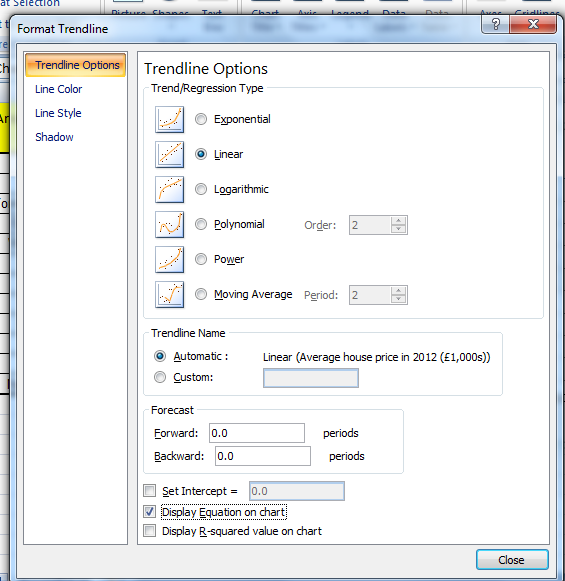
*Select* ***E2*** *type =corr*

**

**

*Select CORREL then column* ***C****, column* ***D Enter***

Correlation coefficient = – 0.3388

*To add a line of regression in Excel (This is called a trend line in Excel)*

*Right click on any* ***data point*** *select* ***Add Trendline***

*Select* ***Linear*** *then* ***Display Equation on Chart*** *and* ***Close***

*Drag the equation of regression (trend line) so it is clearly visible on the chart, reduce the number of decimal place values to two and format the text using the formatting in the* ***Home*** *tab.*



1. Write the line of regression using the names of the variables.

Daily mean total cloud cover = – 0.07 × Daily mean pressure + 77.53

1. Interpret the gradient of the line of regression for daily mean total cloud cover against daily mean pressure.

The regression model suggests for every hPa increase in daily mean pressure the daily mean total cloud cover decreases by 0.07 of one oktas.

Or

The regression model suggests for every 10 hPa increase in daily mean pressure the daily mean total cloud cover decreases by 0.7 of one eighth.

1. Does this regression model seem to fit the data?

Although there seems to be a negative relationship there is a lot of scatter for larger values of daily mean pressure.

1. Use the regression model (equation of regression) to predict the daily mean total cloud cover for a daily mean pressure of 1030 hPa.

Daily mean total cloud cover = − 0.07 × 1030 + 77.53

= − 72.1 + 77.53 = 5.43 5 oktas

1. Comment on the accuracy of the prediction in question 13.

Not very accurate as there is a lot of scatter at a daily mean pressure of 1030 hPa.

**Report**

The scatter diagram and correlation coefficient suggest there is a slight negative correlation between daily mean total cloud cover and daily mean pressure for Leuchars in 2015 between May and October. i.e. as daily mean total cloud cover decreases daily mean pressure increases in Leuchars May to October in 2015.

Further, the regression model suggests for every 10 hPa increase in daily mean pressure the daily mean total cloud cover decreases by approximately one eighth.

However, there is a lot of scatter about the line of regression for high values of daily mean pressure.

**Investigation 4 (b)**

Plot a scatter diagram to investigate the relationship between daily mean total cloud cover against daily mean pressure for Leuchars 2015, split by season.

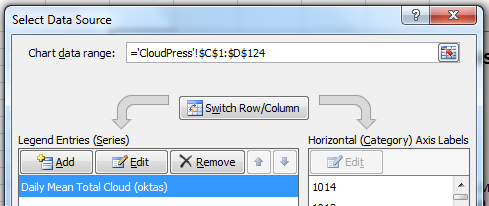
*To plot the scatter diagram daily mean total cloud cover against daily mean pressure for Leuchars 2015, split by season*

*In the worksheet* ***CloudPress*** *select* ***E1*** *type* ***Spring/Summer Enter***

*Select* ***F1*** *type* **Autumn Enter** (these will be used as series names)

*Plot a scatter diagram for Spring/Summer only, cells* ***C1:D124****.*

*Move the chart to the top of the worksheet. Either drag or* ***Cut*** *and* ***Paste.***

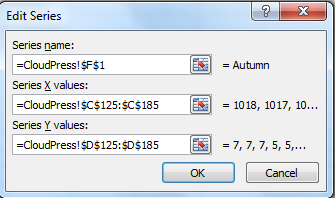
**

*Add a label for Spring/Summer*

*Right click on the* ***chart*** *select* ***Select Data***

*Select* ***Edit***

***Series name*** *select* ***E1 OK OK***

*Add Autumn data and label*

*Right click on the* ***chart*** *select* ***Select Data***

*Select* ***Add***

***Series name*** *select* ***F1***

***Series X values*** *select* ***C125:C185***

***Series Y values*** *select* ***D125:D185****(delete ={y})*

***OK OK***

*Format the chart by adding a title and labels.*



**Report**

1. Comment of the split of the data between Spring/Summer and Autumn.

The data points for spring/summer and autumn see fairly well mixed. They are not distinct.

**Investigation 4 (c)**

Investigate if there is a correlation between daily mean visibility and daily mean temperature. Use the same random sample i.e. Leuchars 2015.

**The data**

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

Select the **Information** worksheet.

1. Read the information in cell **A12.**

Explain how daily mean temperature is measured.

Air temperatures are recorded by thermometers in a louvered screen 1.25 metres above short grass, except at some Weather Centre’s and Climate Data Logger stations, where observations are made from a non-standard roof top exposure.

Values are noted in degrees and tenths (Degrees Celsius) and values below 0 Deg C are preceded by a minus sign. A reading which is not available is listed as ‘n/a’.

The daily mean air temperature (0900-0900 GMT) is the average of the hourly temperature readings during this period.

1. Read the information in cell **A18.**

Define daily mean visibility.

Visibility is defined as the greatest distance at which an object can be seen and recognized in daylight, or at night could be seen and recognized if the general illumination were raised to daylight level. Values are noted in decametres (Dm)

1. What type of variable is daily mean temperature?

It is a continuous variable.

**Process**

Plot a scatter diagram of daily mean visibility against daily mean temperature.

1. Which variable should be plotted on the vertical axis?

Daily mean visibility.



1. Comment on the correlation between the two variables.

The scatter diagram suggests there is no correlation between daily mean visibility and daily mean temperature.

1. Explain why no line of regression should be calculated or fitted to the data.

There is no correlation between the two variables. This suggests an increase or decrease in daily mean temperature will cause no or very little change to daily mean visibility.

**Investigation 4 (d)**

Investigate if there is a correlation between daily mean air temperature and daily mean pressure in Beijing May to October 2015.

**The data**

The data are provided in the Excel workbook **Edexceldataset.xlsx.**

Copy the data for Beijing 2015 into a new workbook

Rename the workbook **Beijing2015**

Delete rows 1 – 5

Save workbook as **Beijing2015**

**Process**

Plot a scatter diagram to investigate if there is a correlation between daily mean air temperature and daily mean pressure in Beijing, May to October 2015. Daily mean pressure is the explanatory variable.

1. Which variable should be plotted on the vertical axis?

Daily mean air temperature.



1. Comment on the correlation between the two variables.

The scatter diagram suggests there is a strong negative correlation between daily mean air temperature and daily mean pressure in Beijing May to October 2015.

1. Give an interpretation of the correlation between the two variables.

As daily mean pressure increases (rises) daily mean air temperature decreases (falls).

Or

As daily mean air temperature decreases (falls) daily mean pressure increases (rises).

1. Find the product moment correlation coefficient in Excel.

Correlation coefficient = − 0.733

1. Add a line of regression to the plot.



1. Write the line of regression using the names of the variables.

Daily mean air temperature = − 0.5 × Daily mean pressure + 556.8

1. Interpret the gradient of the line of regression for daily mean air temperature against daily mean pressure.

The regression model suggests for every hPa increase in daily mean pressure the daily mean air temperature decreases by 0.5 degree C.

Or

The regression model suggests for every 10 hPa increase in daily mean pressure the daily mean air temperature decreases by 5 degrees C.

1. Interpret the intercept of the line of regression for daily mean air temperature against daily mean pressure.

The intercept represents the daily mean temperature when the daily mean pressure is zero.

One reason this value is not useful in this example is that the intercept is not in the range of the data.

1. Does this regression model seem to fit the data? Give a reason for your answer.

Yes quite well as the data seem to follow a linear pattern and there is not too much scatter about the line.

1. Use the regression model (line of regression) to predict the daily mean air temperature for a daily mean pressure of 1005 hPa.

Daily mean air temperature = − 0.5 × 1005 + 556.8 = 54.3 degrees C

1. Comment on the accuracy of the predicted daily mean air temperature in question 10.

A reasonable prediction as there is a little scatter about a daily mean pressure of 1005 hPa.

1. Plot scatter diagram to show the relationship between daily mean air temperature against daily mean pressure for Beijing, May to October 2015, split by season.



**Report**

The scatter diagram and correlation coefficient suggest there is a strong negative correlation between daily mean air temperature and daily mean pressure in Beijing May to October 2015. i.e. as daily mean air temperature decreases (falls) daily mean pressure increases (rises).

The distributions of the data for spring/summer and autumn seem to be two distinct populations. Spring/Summer days have higher daily mean air temperatures with lower daily mean pressure whereas autumn days have lower daily mean air temperatures with higher daily mean pressure.

The correlation between the two variables seems stronger in autumn than in spring/summer.