

Lesson 2 Data analysis: Climate data for 70 glaciers at their equilibrium line altitudes

Complete the following tasks using the accompanying Excel work book.

Copy the columns of data into an Excel spreadsheet of your own that you can manipulate (the version attached to the website is read only). You will shortly analyse these data in a scatter graph.

- A scatter graph is a way of presenting two sets of data to see if there is a relationship between those sets of data. In other words, are changes in one variable related to changes in the other variable? In this example, we are looking to see if at the ELA of glaciers, there is a relationship between the average summer temperature (variable 1) and the total precipitation (accumulation of snow) in a year (variable 2).
- A scatter graph also allows us to test a **hypothesis** about the relationship between the variables. A hypothesis is a statement about what we are expecting to find based on our previous knowledge. Suggest a hypothesis for the relationship between temperature and precipitation at the ELA of glaciers in the space below:

Our stated hypothesis can be referred to as **H1**. The opposite possibility (i.e. the relationship we are expecting does not exist) is referred to in statistics as the **null hypothesis**, or **H0**.

Now create a **scatter graph** inside your Excel spreadsheet comparing the temperature values with the precipitation (accumulation) values.

- Highlight the temperature column and the precipitation column together. By keeping the temperature column on the left side when you highlight, temperature will become the X axis (horizontal), and precipitation the Y axis (vertical).
- Go to the 'Insert' menu, and select 'scatter' from the chart menu (choose the scatter that does not show points connected by a line).
- After your chart appears you can right click on any of the axis values to format the axis (it helps visually to set the vertical axis crossing the horizontal axis at -4.0).
- You could also alter the title of the chart, add axes labels, and change the appearance of the grid lines.

Describe the nature of the correlation between the values that you have plotted. Is it 'positive' or 'negative'? Does it appear to be weak or strong?

Now add a trendline and R-squared value to your scatter graph.

- Right click on any of the data points, and you will see a box allowing you to add a trendline.
- Click on 'Add Trendline', and select linear.
- Also select 'Display R-squared value on chart'

The trendline is a 'line of best fit' that summarises the relationship between the two sets of data. If the points are close to the trendline, then we say that the correlation is a strong one. If there appears to be a trend, but many points are not very close to the line, any correlation that exists is said to be weak.

In statistics, 'r' stands for the Pearson Product-Moment Correlation Coefficient. It is calculated to measure the strength of the correlation, and more specifically to **test the significance** of the correlation. It is applied in a similar way to the Spearman's Rank Correlation Coefficient which is more often encountered in A-level studies. Like the Spearman's Rank coefficient, the fewer the data points there are, the higher the Pearson coefficient has to be for the correlation to be statistically significant.

By 'statistically significant' we mean whether we can be sure at a certain level of confidence (e.g. 95% confidence, or 99% confidence) that the correlation observed between the variables could not have occurred by chance. Put in another way, if the correlation is statistically significant, then we can say with a high degree of confidence that there is a real relationship between the variables that we're studying.

Click on any of the links below to learn more about how significance testing of correlations works.

<http://www.le.ac.uk/bl/gat/virtualfc/Stats/pear.html>

<http://geographyfieldwork.com/SpearmanRank.htm>

Without worrying about how the Pearson coefficient is calculated, when we square it, we get a value that tells us *the percentage* by which the variation in one variable can be explained by the other.

Write down the 'R-squared' value that Excel calculated for your scatter graph. _____

Multiply this decimal value by 100 and you have the percentage variation in total precipitation that is accounted for by the summer temperature at the equilibrium line altitude.

In the space below, describe whether your hypothesis is supported by the data or not. If your H1 is supported, how much confidence do you think you can have that this relationship could not have occurred by chance?

Print out or save your chart so that you can answer the **Lesson 2 Plenary questions**.