# Section 4 –
# Data Analysis

Royal Geographical Society with IBG
Advancing geography and geographical learning

The **data analysis** is the stage of the investigation where the researcher manipulates the data to make it more meaningful, easier to use and in a format whereby the research questions can be answered. It is commonly written up at the same time, and alongside, the data presentation section. It is highly likely that it will make sense to analyse some data before it is presented graphically and vice versa.

For numerical data, analysis usually involves the use of statistics, in which the data is put through tried and tested methods to show whether there are any relationships between variables within it and how significant these relationships actually are. Some statistical methods are relatively straightforward, such as calculating a percentage, or the mean (average) from a set of data. Others require a more in-depth mathematical understanding though, in most cases, it may be enough to simply use the formulae given to you and understand what the figures mean at the end of the calculation. Remember to use an appropriate form of statistical analysis. Properly explaining your results can positively contribute to gaining a good mark in your investigation. It is not uncommon for geographers to feel concerned about statistical analysis. You should not feel daunted by the prospect of using statistical methods as there is plenty of guidance available, along with step-by-step instructions, to help you.

As with every part of your Independent Investigation, it is important that the statistical methods chosen are **appropriate** both for the data in hand and the research questions or hypotheses the study aims to answer. A statistical method should certainly not be chosen simply because it is easy or the only one you know: it is worth spending some time exploring the different methods available to see which best suits the hypotheses. Each statistical method is designed with different aims. Therefore, you should not enter into the data analysis under the belief that with any one set of data, you can pick and choose from a menu of statistical techniques. Instead, you should be aware that each statistical technique has one purpose. If that purpose fits the aims of the study that statistical technique has effectively 'chosen you'. Some statistical techniques work best alongside certain data presentation techniques too.

A summary of some statistical techniques and their uses is highlighted below:

| Name of technique | Use | Use with data presentation techniques |
|---|---|---|
| Measures of Central Tendency | To show the most common observed data | Bar chart<br>Radial graph |
| Measures of Proportion | To find out what proportion of a whole the data is representing | Pie chart<br>Triangular graph<br>Composite bar chart<br>Proportional shape map |
| Dispersion | To show the frequency of data around a central idea | Box and Whisker graph<br>Line graph |

Nuffield Foundation
*This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.*

| | | |
|---|---|---|
| Spearman's Rank Correlation | To show the strength and nature of a correlation between two sets of data | Scatter graph |
| Chi-Squared Test | To show how closely observed data matches data that is expected | Bar chart<br>Histogram |
| Simpson's Diversity Index | To show the frequency of different categories as well as the abundance of those categories | Kite diagram<br>Choropleth map<br>Isoline map |
| Pearson's Product Moment | To show the strength of correlation between two variables that show a linear relationship | Scatter graph |
| Nearest Neighbour Analysis | To show the degree to which something is clustered or uniformly spaced | Choropleth map<br>Isopleth map |
| Mann Whitney U Test | To show whether two data samples are significantly different from one another | Choropleth map<br>Line graph<br>Scatter graph |

It is also good practice to **justify** to the reader the reasons for choosing one statistical method over another.

Non-numerical (**qualitative**) data, as well as **secondary data** can also be analysed. In the former, this may involve the conversion of 'wording' to numerical data (e.g. by counting the number of times certain negative or positive opinions are made during an interview) and the coding and indexing of transcripts can be used to make the wealth of 'word data' more manageable. Equally, secondary data can be subjected to the same statistical analyses as primary data, allowing the researcher to create a larger data set for analysis or more easily compare data from different time frames.

After reading your data analysis, the reader and marker of your Independent Investigation should be left with no surprises as to the type of **conclusions** you will be making. Therefore, the data analysis is the last stage at which the researcher can reveal any new or different connections in the data that have not already been explained; the data analysis section leads the reader naturally to the conclusion. However, conclusions can only be made if the data analysis shows that particular idea to be true: it is not good research practice to allow a hunch or personal instinct about a place and a situation to cloud one's analysis of the data. If the data does not show something to be true, no amount of data handling will make it so. Under no circumstances should the researcher be tempted to change or make up the raw data itself (such as removing outlier or anomaly data) in order to suit the conclusions that they wish to reach.

**Common Pitfalls:**

- **Carrying out more than one statistical test on the same set of data.** You will not gain extra marks for this – it is much better to choose the most appropriate statistical analysis method and justify your choice afterwards.
- **Only analysing some of the data.** If you intend to draw a conclusion from some data, or answer a research question from it, then it must be included in the data analysis.
- **Choosing a statistical analysis method on the ease of the mathematics required alone.** The best reason for choosing one statistical method over another is because it is most appropriate for the data you have and the nature of the conclusions you wish to draw.
- **Including pages of calculations.** It is best to put any workings in an appendix in case the reader and marker of your study wishes to check how a statistical method has been used.
- **Changing the data itself to suit a preconceived conclusion.** If the data analysis hasn't shown the conclusion you were expecting, there is likely to be a geographical reason for this, or there may be a limitation in the way you conducted your study. Both circumstances give you more interesting things to write about, compared to the situation of everything working out as you expected.
- **Making grand conclusions from under analysed data.** Unless you can prove on paper something you are saying, don't make a simplified statement.
- **Providing raw data as well as analysed data.** There is no reason to provide pages of raw data if you are going to ultimately show it in its manipulated form.