Preamble

General information

This learning module on data analysis will enable you to plan and analyse data collected as part of a scientific investigation and arrive at the correct conclusions. Some of the notes have are based on the book by Triola (2009); however some people may prefer the text book by Kirkup (2002). Triola's book however is more relevant to this particular course. However, through experience I have found that Excel can be quite limited and sometimes requires a lot of effort (and patience), therefore I have tried to include information in how to use MATLAB to do the analysis in place of Excel; an excellent reference for those getting started in MATLAB is Hanselman and Littlefield (2001). By all means feel free to look at these books; however, don't expect to find the solutions to the Excel projects: they do not contain them. In order to solve these problems you will have to work at them yourself and may find it necessary to ask myself or the demonstrators for pointers. Also, don't be afraid to speak to each other if you are having problems understanding certain aspects of what we are doing.

A module on data analysis and statistics is seen as a necessity for environmental and earth scientists these days. If it is your desire to published peer reviewed articles on your work then they are subject to criticism if the methods used to summarise your data or arrive at you conclusions are not statistically robust.

I'll mention the text book in class, but my personal opinion is that the notes you receive in lectures are adequate to pass the course; however, if you want to learn more about statistical methods, with lots and lots of examples that are relevant to the environmental sciences, including ANalysis Of VAriance (ANOVA) or statistical process control (which will allow you to say whether global warming is out of statistical control), then by all means get the book.

I will not follow the book exactly, there is far too much information to cover in a lecture, since statistics is quite a large subject. Consequently it is expected that you will read them in your own time. Whether you want to read before the lecture or after is up to you entirely: I wont lecture you on the best way for you to learn! You are all different and all learn in different ways; however, I think you'll agree that it is impossible for me to pour the knowledge into your mind, and so *you* still have to do the work.

In this module there will be no end of semester exam, but you will be tested as you go through the material (see the schedule on the handouts to be sure you don't miss any assignments). The aim of the practicals are for you consolidate your knowledge by 'learning by doing'. If there is one thing I hope you take from this course, it is not to memorise a bank of statistical procedures and formulae. Rather, it is that I hope you'll remember when you come to plan an investigation or analyse data, that there are certain rules that should be adhered to and that you'll be able to look them up and use them throughout your careers. Hence, while I won't lecture

you on note keeping (you're all adults!). I will say that it may be a good idea to save the spreadsheets that you work on in the practicals so that you can come back to them should you need to.

Where will I use statistical methods?

Here are some examples where you may use statistical methods:

- Throughout your degree
- In your final year dissertation: projects that use robust statistical methods are generally the ones that attain the highest marks.
- If you go on to get a job in: hazards, research into environmental health, toxicology, forecasting, consultancy, etc.
- Some BIOL modules that environmental science student take have used ANOVA techniques; while I won't be covering it here, the methods developed in this course will make using ANOVA that much easier. I can offer notes and guidance to those interested.

Why use statistical methods?

Basically, because if you can not show that your results are significant or meaningful then why should anyone take any notice or what you are saying?

- You will get better project marks if you are competent at data analysis.
- If you want to publish any results you will have a hard time unless you show the results are significant.

The example I like to use is say you have a bucket full of different coins: 1, 2, 5, 10, 50 pence pieces, say. You are asked to take out 5 coins, one by one at random, and by chance you manage to take out the coins in ascending order (e.g. you take out a 1p, a 2p,...,50p). You could wrongly conclude that the value of the coin you take out depends on the order you take them out. Or course if you were to repeat this experiment again you would have little chance of taking the coins out in order of monetary value.

MATLAB vs Excel

Its no secret that I am a MATLAB advocate, and now python. Python is excellent, but MATLAB is probably the easier language to pick up if this is the first time you've done any programming. I would rarely use Excel for any type of analysis. Excel is very good for keeping track of spreadsheets and accounts—which is what it was designed for. Throughout the course I will give examples in both MATLAB and Excel in the hope that you'll see how powerful MATLAB is. However, it is *mostly* up to you which on you choose to make use of in the practicals. I am pretty sure I wont convert the most stalwart Excel warriors to MATLAB.