Final Draft of Chapter for: Dorling, D. (2010) Using statistics to describe and explore data, 2nd Edition of Key Methods in Geography (Eds. N. Clifford, S., French and G. Valentine), London: Sage, pp. 374-385.

21

Using statistics to describe and explore data

Danny Dorling

Synopsis

In this chapter I argue that simple statistics are better than complex statistics in the study of geography, that statistics are of themselves dull and that the majority of university geography students and staff are ill-equipped to use them, let alone enliven them. Five rough and ready rules are suggested for when and where to use statistics in the study of geography. One origin of the word statistics is that it implies 'facts about the nation-state'. The development of both 'state-istics' and the academic subject of statistics is briefly summarized to provide a context for the argument presented here that geographers should concentrate on simple statistics. Finally, the chapter ends with three examples of how things may not be as they seem when viewed statistically.

The chapter is organized into the following sections:

INTRODUCTION

Statistics are duller than ditch water. In and of themselves they tend to be of interest only to people who are not very interesting. To me, it is only when statistics are set in a wider context that they begin to come to life. In geography this wider context obviously also varies widely. Suppose you are swamped by a series of floods and want to know how likely they are to happen again and how large they could be. Statistics give you techniques to make good guesses – guesses that don't necessarily rely on you knowing much about floods. Suppose you are interested in poverty and how poverty rates compare between countries. Poverty rates are statistics. You need to understand the statistics if you are interested in the issue of poverty. You could, of course, ignore the statistical study of floods or poverty while still being interested in either subject. Were you to do this, however, you would be missing out on a great deal that has been learnt about these things. Statistics on floods might miss out the nature, causes and meaning of floods. Similarly, statistics on poverty may dehumanize suffering. But if you are concerned about getting wet or about why higher levels of poverty persist in some places and not others (and hence one key aspect of what reduces poverty – variation in, say, policies that influence poverty), you are unlikely to get far without the numbers and methods which make up statistics.

Statistics are dull because they are so general – because so many different things can be turned into a percentage; because similar techniques can be used to study so many different processes. The generality of statistics is also their main weakness as well as being a strength. Often issues and problems in geography are shoe-horned into a question that can be addressed, perhaps answered, statistically. A researcher turns an interesting research question into a series of dull hypotheses, the answer to none of which quite gets to addressing his or her original question. It is, for instance, much easier with classical statistical tests to suggest that certain things are (probably) not true – rather than to assert what is true. Statistics can also have the disadvantage that many students have built up a resistance to them, an inherent dislike of their use and feel a chill wind travel down their spine when the word is mentioned. Ask yourself this (assuming you are a student studying for a geography degree): what proportion of your fellow students are likely to be reading this book and, of those, what proportion would skip this chapter? Are you in a minority of a minority to have got this far? Perhaps I am wrong, but hopefully if you read a little further I will try to show you that statistics are not quite as painful as they may be thought to be – particularly if you only use them when you need to.

There is at least one other probable reason why you may have got this far through this chapter. You bought or read this book a year or so ago when you were first being introduced to methods in geography. You read about the interesting ones, the novel and the new, but you skipped this chapter. Later, in a seminar or tutorial, someone asks 'what statistics are you using?' for your research project. That feeling of panic returns. Or you referred to some statistic in an essay and your marker irritatingly scribbled 'explain!' next to it. In this short chapter I am going to show you only one statistical method (calculating and interpreting a confidence limit). No statistical test and no actual statistics – of the '95% of statistics are wrong' variety. Instead I'll follow this introduction with five simple rules for using statistics in geography and then talk a little about the origins of statistics, give some common misinterpretations of statistics and end with possibilities for the future use of statistics in geography.

FIVE RULES

Often there is little point in using statistics

All too often statistics are included in a study or student essay to try to show that the writer or student had been working hard. An essay on the geography of employment in Britain might begin: 'Thirty million people work in Britain today.' So what? Statistics in and of themselves are dull (unless dull things excite you!). 'There are thirty million people working in Britain and forty million jobs' would be a little more interesting – many people would have to have two jobs – but that is only interesting if it is new to you as a reader. 'There was a positive correlation between the number of people working in each county and the number of jobs' – hardly surprising, and what's your point? Perhaps what you were trying to say was:

> More people are working then ever before at more and more tasks in Britain and at even more in the USA. Work, rather than removing human drudgery, has added to it. As we work harder we have created more work. Huge numbers of people have to work at two or three jobs a week compared to our labours in the recent past. The proportion of both children and the 'retired' who work has risen. For the affluent, childcare and housework are increasingly subcontracted as work for others, as for many is cooking when we eat out. It is difficult to see a smooth end to the apparent unstoppable commodification of human time. Work turns time into money. We now have both more money and less time.

Only the very crudest of statistics were used in that argument. Things had simply gone up or down – were more or less. It would have helped the argument if I had listed some sources for this information. It is probably not all true and what is described is certainly not true for many people. However, the argument did not need numbers or tests to strengthen it. There is a time and place for statistics.

If you do use statistics make sure they can be understood

Most people, including a majority of human geography lecturers and a fair share of physical geography lecturers, are innumerate – as are most students of geography. By innumerate I do not mean innumerate in the sense of 'could not get an A grade at GCSE maths'. By innumerate I mean do not have an innate feel for numbers, mathematics and simple algebra in the same way that the minority of students who are dyslexic do not have an innate feel for words and their spelling. Because dyslexia is a minority condition it is given a name. Because innumeracy is the norm in America, Britain and most of the western world, we do not call it as such but save the word for people with almost no ability to handle numbers. If you are in the innumerate majority you may have thought you had some weakness and everyone else could understand maths and statistics and did not need to learn examples by rote as you did. You are wrong: you are normal. If you are in the numerate minority (that is, you could walk into an exam and get an A in maths without trying) you have a problem – many people (including often those reading or marking your work) might not understand you because they do not think like you. Statistics in geography are primarily a form of communication – a way of trying to convey information convincingly. You will only convey information convincingly if, to begin with, the person you are conveying that information to can understand the language you are using.

Do not overuse statistics in your work or methods

How many facts should there be to the page? Occasionally a great many simple statistics all pointing to the same things or suggesting a trends is going one way using many examples can be a powerful tool in making an argument. Generally, however, I'd try to stick to two or three facts on a page unless you are writing a reference book. You are trying to build up a picture in the mind of the reader/marker – explaining an argument through an essay. Similarly, how many numbers should tables of statistics contain? Part of the history of geography has been that multiple regression was a favoured statistical technique for several decades. Multiple regression allows dozens of variables to be entered into a model to predict an outcome. Each variable can have numerous parameters associated with it – for example, its 'effect', 'significance', 'error' and so on. When computers first began to be used in geography, a huge amount of work was required to enter the data and variables to undertake this technique. One result was that huge tables of 'results' were often printed in papers which in fact only referred to a few of the numbers in the tables of sometimes hundreds of figures. Have a look at a journal like *Environment and Planning A* in the 1970s and 1980s for these giant tables. The tables give the impression of thoroughness, but if you want people to *read* the numbers in your tables I'd aim to have only a dozen numbers there. Finally, concerning tests and techniques, there should be little need to use more than one statistical test or technique on a set of data. If you are using more than one, the chances are that only one is appropriate.

If you find a complex statistic useful then explain it clearly

Sometimes, for some problems, only a complex statistical method or complex statistics get to the heart of the issue you are interested in – and it is the issue rather than the method that we are interested in. A typical case in geography might be that you have access to a whole series of river-water samples and know the level of a particular pollutant in each sample. You also know whether the sample was taken from water near the surface of the river or near its bed, whether from a pool in the river, a riffle, or a bend and what both the slope and velocity of the river were at each sample point. You want to know whether some rivers are more polluted than others or whether they just appear to be so because of the nature of the way the samples of water were collected from each river. Your data are arranged in at least two geographical levels: that of the river and that of geographical location within each river. Each individual sample point also has a number of characteristics you want to keep taking into account. There are a number of techniques you could use to study these data. They range from various types of ANOVA (analysis of variance) tests to MLM (multi-level models). The one you are likely to choose will depend most on what you have been taught and remembered, or perhaps on what you have read. My recommendation is that, when using a complex technique in a complex situation like this, explain every step simply – but above all understand why you are using these techniques. The above techniques are interesting for their abilities to identify 'interactions'. Within a particular set of rivers you might expect more pollution in riffles than in pools and vice versa in another set of rivers. If you are not interested in interactions, why are you using these complicated techniques? Furthermore, check first that there is not a very simple pattern in your data that simple statistics could undercover more persuasively. For example: 'Pollution levels are on average ten times higher in the pools of rivers than when measured at any other point.'

Recognize and harness the power of statistics in geography

Statistics have a political power across much of geography. In physical geography their use implies you are (or are becoming) a competent 'scientist'. Many scientific (and, in particular, medical) journals use statisticians to referee papers before publication as well as referees who understand the substantive subject of the paper. Statistics has become a language of scientific credibility – rather like Greek, and then Latin, was the language of religious credibility in the Christian church for most of its history. Languages are used both for their ability to communicate and to exclude the uninitiated. The power of statistics (in particular, statistical methods) in human geography has fallen in recent years. This fall was partly the result of researchers seeing through the way statistics have been used to exclude and also partly because fewer numerate human geographers choose to teach (or research) geography than once did. However, within human geography, statistics placed carefully can still add authority to an argument that, without them, appears little more than a considered rant. This is particularly true when you are arguing against the generally accepted case. As I argue (or rant!) below, statistics are increasingly used in general debates outside of geography. It is possible both to appreciate the weaknesses of statistics and to use them with effect. But it helps first to know where they have come from.

Before considering the origins of statistics, in the pursuit of honesty and to sum up, it is probably worth explaining the situation in which I find myself most commonly introducing geography students to statistics (I don't teach statistics in lectures as I can't find a way of doing so that would keep 150 undergraduates awake or at least attentive for longer than a few minutes).

Here's the scenario: 'So, you've thought of your research questions, done your survey/questionnaire/measuring, got your data, and are ready for some "statistics" . . .' It would have helped if the students had thought about what they were going to do with their data before they collected them but, for many, even if they had done that they would not have known the possibilities: what are these statistics things and why do you need them? For most undergraduate students in geography, and much of the rest of the social or earth sciences, statistics are the things that change a 2.2 dissertation into a 2.1 dissertation – or at least the things they think might do that. They are the things 'taught' by the lecturer who didn't get to teach what he or she wanted to teach – or, even worse – by the lecturer who actually chose to teach statistics! They were the bane of your life while revising for A level. They are the things you dread in exams. You had to learn a few, learn to use a few techniques and learn some more general statistical methods. A dislike of statistics was the reason you studied geography rather than psychology, economics or biology. Students who can cope with statistics don't tend to opt for a geography degree. However, they find them difficult rather than repulsive. If you could not stand statistics you might have picked English, history, art or drama instead. This chapter is for these kinds of student. If you have read this far then it might well be for you too. Above I tried to explain why I think it is worth using some simple statistics. Below I define statistics more fully, give some reasons for not using complex statistics too often and then I end this short chapter by giving a few simple statistical problems.

WHAT ARE STATISTICS?

Statistics are many things and mean different things to different people. My favourite definition is that they are 'facts about the state' – meaning the nation-state people inhabit and usually involving aspects of their lives. This is a very human geography definition but it fits the origins of statistics where the word began to be used in earnest two hundred years ago to describe the numerical portrayal of countries which was beginning in earnest then. (For more on these, consult William Playfair's writing and a little of what has been written on their origins – see the further reading section at the end of the chapter.) The first modern census – collecting useful statistics about the entire population of a country – was undertaken in the USA in 1790. It is perhaps telling that the counting of people in this way as a new technique began in the new world. Statistics as a word has come, in recent times, to mean more often than not numerical facts about the state. Most statistics that people will encounter in a day in the country I am writing in (Britain) are about that country. The same is even more the case of the USA. Two centuries ago, as today, economists in various guises were and are responsible for producing the majority of these kinds of statistics. There use in political arguments, however, has involved a far wider set of disciplines than economics (Dorling and Simpson, 1999, provide examples of these).

Of course, statistics has a second meaning far removed from simply being about facts. Just over one hundred years ago a small group of people who were interested in chance (probabilities) created the academic discipline of statistics. They were also, incidentally, quite interested in society, and many had views about, say, 'criminals' and 'races' which would be rightly seen by most today as abhorrent. This meaning of the word is a world away from the first definition. Most large universities have either a statistics department in them or statistics is part of their school of mathematics. Statistics in this sense has been dehumanized partly due to a wish to forget part of the history of statistics. You may have taken part of an A level in these kind of statistics. Half my first university degree was taken in a statistics department. There the working definition of statistics was very different from that used across the campus in geography. One of my first textbooks was entitled *Statistics: A Guide to the Unknown* (Tanur, 1989), the implication being that for those things we know about there is little need for statistics, but for anything we don't it has a purpose. The introductory examples I was taught as an undergraduate ranged from assessing the likelihood that Shakespeare was the author of each play attributed to him, to calculating how many whales were likely to be living in the seas around Japan in the future, given current trends (and not knowing how many whales were actually in the sea). The one area that examples were not often drawn from was the study of human societies, and so I often wondered whether this was because we know too much about people and how they organize their lives to make them a suitable subject for statistical study. Many human geographers would argue that people's lives are too intricate to summarize meaningfully statistically. But again, perhaps the lack of statisticians now studying societies is partly due to embarrassing links between the origins of the study of statistics and movements in the first few decades of the last century which sought to claim that some people were superior to others? These are elitist movements which remain strong but which operate far more covertly today. You are almost certainly reading this because you are a product of the continued success of such elitist movements. You were selected to be 'reading' for the degree that leads you to read this book because you passed a series of tests and went to a series of schools for

those deemed to be ‘more able’ (including comprehensive schools in affluent areas). Most sensible social researchers know that you did not pass those tests because of some inherent position you hold on a supposed ‘normal curve’ of ability, but because of where and to whom you were born, the wealth and caring of your family. But you are also where you are because statistics were misinterpreted in the past to create an education system in Britain that remains highly segregated and shockingly elitist.

WHY USE SIMPLE STATISTICS

I have not found that the use of complex statistics has greatly enhanced my understanding of people and how their lives are organized over space. Nor have I generally found that reading other people’s accounts of such models has helped me much either. Perhaps its just me, but I think it’s worth saying.

From looking at complex models of who chooses how to vote for political parties, to why some people in some places are more likely to be ill than others – to me, the results of complex statistical tests and models depend in most cases more on how the tests and models are put together than on the data. This is a very personal and generalized view (which is why I have written it and why most of this chapter is in the first person). But it is half the explanation I have as to why I think simple statistics are usually better. When you have less control over the numbers and how you manipulate them it is harder to condition the answers you get to what you want to find.

A second reason for tending to use simple statistics is that you are more likely to have something to compare with. Take, for instance, researchers who study pollen in geography departments. These are people who are trying to construct past climate records from the historical record of pollen in a soil core. They tend to call themselves things like Quaternary scientists interested in palaeoclimatic reconstructions rather than people looking at specks of pollen in old mud. Note that how we choose to define ourselves is often as much to confuse as enlighten! What little I know of them and their use of some statistics suggests to me that they mostly tend to use a similar method of cluster analysis to study some aspects of their data. They do that because it worked pretty well for the first person who used it and who taught the second and, although it may not be perfect, it makes more sense if almost everyone uses a method he or she can understand rather than each invent his or her own. Cross-comparison and communication between studies are then possible. Statistics is a language that has different dialects in different places – even within different parts of one small academic discipline. A ‘total fertility rate’ and ‘bank-full discharge’ are two common simple geographical statistics but hardly anyone will know the meaning of both. If you do want to know the meaning of these terms, key them into a search engine on the Internet – but be careful to confirm the definition you are given on more than one website. Anyone can write almost anything on the web! Wikipedia, much maligned by lecturers and the publishers of textbooks, is a far better source than most others that can be accessed for free. If you want to know whether the quality control on it is better than in text-books, then: if this sentence is still here when you are reading it – that will tell you that no one chose to delete it after the point it was typed, tell your lecturer that Wikipedia is a great source – you read it here! It is particularly good on statistics by the way (less so on opinion, in my opinion).

Even if you use a simple statistic in one part of geography, it is often peculiar to that part of geography. You are choosing to speak to a very small audience. There are, however, even simpler statistics. On the physical side of the subject these are your basic rates, weights and measures. On the human side simple statistics are the stuff of social summaries. They appear in everyday settings most commonly delivered through newspapers, radio, television or the web. These are the statistics most worth using and which put a geographer into a very good position to question the facts. Many, if not most people, accept facts at face value, which is part of the reason why they are such powerful tools to use in argument (see Dorling and Simpson, 1999, if you are interested in how such things are questioned). In short, simple statistics are more easily understood and more convincing. Within a subject such as geography in the first decade of the twenty-first century – where so few geographers feel (or are) numerate – the use of complex statistical techniques has to be questioned: whom are they being presented to? So, if you as a student find them hard to understand I would say to you, don't worry, and ask yourself why you were being asked to understand them in the first place! If you find a good reason for that, *then* try to understand them.

THREE STATISTICAL EXAMPLES

It is hoped this chapter so far has given you some food for thought about the use of statistics in geography today. Rather than list a series of techniques and refer you on to books about their use, I've tried to give you a more personal view about one of the supposedly most impersonal of subjects taught within geography. My view may be a little odd: I use numbers but believe they are deadly dull; I like both to undermine them and use them in the arguments I make. I think our legacy of complicated statistical techniques in geography is more of a historical accident of what was found to work at certain times than a particularly useful set of tools – at least where the study of the geography of society is concerned. I'd like to end this chapter, however, with three examples of how statisticians think differently – which, hopefully, show some things to be learnt from statistics.

Example 1: a simple prisoner's dilemma

You are playing a quiz game on TV. In front of you are three doors. Behind one door is the prize and behind the other two doors nothing. You pick a door. The quiz host then picks another door. If you picked the right door the host picks a wrong door; if you picked a wrong door the host picks the right door. You then have a chance to change your mind over which door you enter. Which do you go through?

The answer is relatively simple, or it is if you find this kind of thing easy: you pick the door the host picked. Your chance of winning the prize is 2 out of 3. Why? Because 2 out of 3 times you will have initially picked the wrong door. Why does this matter to a geographer? Well, go back to where we started this chapter. There's been a large flood and you are interested in the probability of another large flood happening. The fact that you are suddenly interested in the probability of the flood is not independent of the flood occurring. In fact, the chance of the flood

having occurred after the event was 100%. You cannot say ‘that was the 1 in 100 year flood’, just that ‘this is what it would be like’ (or perhaps, better, might be like!). Probabilities are usually conditional. In geography we often make simple statistical mistakes as geographers learnt most of their statistics during the discipline’s classical phase (when statisticians didn’t worry about conditional probabilities so much).

Example 2: another conditional probability

You are worried you have a disease. One per cent of students have the disease. You take a test for the disease. For people who have the disease the test is accurate 95% of the time. The test is positive for you. What is the chance you have the disease?

As you might have guessed, the answer is not 95%. There are two possibilities: either you have the disease and tested positive or you didn’t have the disease but the test gave you a false positive. Take 10 000 students. Of the 100 who had the disease, 95 would test positive. Of the 9900 who did not have the disease, 495 (5%) would test false positive. Thus the chance of you having the disease, having tested positive for it, is 95/(95 + 495) or roughly 1 in 6. You are unlikely to have the disease. Note that almost 6% of students would think they had the disease had they all taken the test. Of course, if you think there was a particular reason for you to be worried and could quantify that, more conditional probabilities would be introduced!

Now, substitute for disease and students, heavy metal pollution and soil samples. You are working in the labs and have a test for heavy metal pollution in a sample of soil that is 95% accurate. About 6 of your 100 soil samples appear to be polluted after you have tested them. Three of these are located in the same village – have you found a cluster? (Answer: no, probably not.)

Example 3: classical confidence limits

There is only one even mildly complicated statistic I now work out on a regular basis and even that has a more complicated interpretation than is usually taught to geography students. I often work out rates in areas – comparing how many things have happened in an area compared to how many you might have expected to have happened in that place given certain assumptions. Most commonly, this is how many people died in an area compared to how many you would have expected to die there given the ages and sexes of the people living there and national mortality rates. Given these things I might work out, for instance, that 120 people died in an area where you would expect 100 people to have died (this is called the indirect method of standardizing rates). Thus it looks as if 20% more people have died than could have been expected. However, it might just have been a bad year in that place. So with what confidence can I say that 20% more people died there? The equations to approximate the confidence limits for this statistic are as follows:

Given a standardized rate of $100 * O/E$, where O = the number observed and E = the number expected:

Lower confidence limit (95%) = $100*O*(1 - 1/(9*O) - 1.96/(3*\sqrt{(O)}))^3/E$ Upper confidence limit (95%) = $100*(O + 1)*(1 - 1/(9*(O + 1)) + 1.96/(3*\sqrt{(O + 1)}))^3/E$

Thus our standardized rate is $100*120/100 = 120$ with a lower confidence limit of approximately:

$$100*120*(1 - 1/(9*120) - 1.96/(3*\sqrt{(120)}))^3/100 = 120*(1 - 1/1080 - 1.96/32.86)^3 = 99.49$$

And an upper confidence limit of approximately:

$$100*(121*(1 - 1/(9*121)) + 1.96/(3*\sqrt{(121)}))^3/100 = 121*(1 - 1/1089 + 1.96/33)^3 = 143.49$$

So what does this say? That we can be 95% sure that the real death rate in that area lies between 99 and 143? That we can't be 95% sure that the death rate in that area is not average (100)? What it actually says it that given, say, lots of years of data, 95% of the time the true death rate of that area will lie within these limits – so that if we were to work out confidence limits for the Standardized Mortality Ratio (SMR) of an area each year for 20 years, on average the true rate would lie within those limits for 19 of those years. This assumes there is an unchanging 'true' rate – that something about the area raises or lowers the mortality rates of the population living there.

Given this, these classical confidence rates do not tell us what we thought they told us. It is not the case that the true mortality rate of this area lies between 99 and 143 with 95% certainty. This has not, however, stopped researchers, including myself, from labelling areas on maps as 'significant' if their lower confidence limits exclude 100, but strictly speaking we have little idea what level of significance applies! (For more on this, see Congdon, 2001.)

CONCLUSION – POSSIBLE FUTURES

When this chapter was first being written it was government policy to attempt to kick start a renaissance in quantitative methods in the social sciences. Evidence-based 'this and that' were all the rage, and research training money was supposed to be being redirected slightly towards the modellers and away from more qualitative studies. The intended effect is that more of your lecturers in the future should have more of a grounding in statistics. What effect is this likely to have? Very little, I suspect. Only a minority of us are confident with numbers by the time we

(the third of children who go) turn up at university. The contortions of logic required to understand that significance tests and confidence limits are not quite what they seem to be – after all, you have only just managed to learn them – are a further great disincentive to use complex statistics. This should not, however, put you off using simple facts and quoting basic trends in sustaining the arguments you make in studying geography. Just because the study of statistics is becoming ever more complex does not mean the practice of using statistics within geography should cease altogether – just try to recognize the breadth of meaning to the word 'statistics'. Writing a short chapter on statistics within geography is rather like trying to define 'geography'. It means different things to different people. What you have just read has been my take on the subject. If you find contortions of logic of interest, please use the further reading section below. If you have been concerned about your insecurities with statistics, don't be – you are normal – just try to use a few more simple facts to strengthen your arguments and try to feel less intimated about the complex methods.

Summary

- Often there is little point in using statistics.
- If you do use statistics make sure they can be understood.
- Do not overuse statistics in your work or methods.
- If you find a complex statistic useful, explain it clearly.
- Recognize and harness the power of statistics in geography.
-

Further reading

For a chapter such as this the most useful set of further readings is probably a collection of the books and papers which inspired the point of view I've put forward here. Authors of chapters in this book were asked to supply five pieces of recommended further reading, so here they are. The more recent they are the easier they are to obtain.

- First, on the origins of statistics, why not go straight to the source? The links between geography and statistics which were there from the beginning are hard to miss, given Playfair's (1800) extremely descriptive title: A Geographical, Historical and Political Description of the Empire of Germany, Holland, the Netherlands, Switzerland, Prussia, Italy, Sicily, Corsica and Sardinia: With a Gazetteer of Reference to the Principal Places in those Countries, Compiled and Translated from the German: To which are Added, Statistical Tables of all the States of Europe. For students in the USA, a search for material on the 1790 census can begin with http://fisher.lib.virginia.edu/census/. If you are interested in worldwide statistics then take a look at www.worldmapper.org and the 400 or so tables of data in excel files available there.
- Two hundred years on and a hugely less ambitious work is the collection of papers that encouraged me to think that statistics are still useful in the study of society: Dorling and Simpson (1999). The book contains examples

of the use of statistics in studies across many aspects of life, mostly in Britain and mostly where political debate is currently raging.

- For statisticians, most have their basic reference book. Mine (before I stopped using complicated statistics!) was Breslow and Day (1980). Books such as Breslow and Day are very detailed and assume a great deal of knowledge, but its worth knowing where people using statistics refer to for their basic formulae.
- A much simpler, if now quite old introduction to the study of statistics is still available in many university libraries and, if you are interested in learning more about statistics in general, I'd start with Tanur (1989).
- Finally, for what may be both the most up-to-date and the most complex book on statistics written by a research professor in geography, see Congdon (2001). The first few pages help explain why 95% confidence limits are not what you thought they were!

Note: Full details of the above can be found in the references list below.

References

Breslow, N. and Day, N.E. (1980) The Analysis of Case-control Studies. Lyon: International Agency for Research on Cancer.

Congdon, P . (2001) Bayesian Statistical Modelling. Chichester: Wiley .

Dorling, D. and Simpson, S. (eds) (1999) Statistics in Society: The Arithmetic of Politics. London: Arnold.

Playfair , W . (1800) A Geographical, Historical and Political Description of the Empire of Germany, Holland, the Netherlands, Switzerland, Prussia, Italy, Sicily, Corsica and Sardinia: With a Gazetteer of Reference to the Principal Places in those Countries, Compiled and Translated from the German: To which are Added, Statistical Tables of all the States of Europe. London: J. Stockdale.

Tanur , J. (1989) Statistics: A Guide to the Unknown(3rd edn). Pacific Grove, CA: W adsworth.

Acknowledgements

I am grateful to Chris Keylock for suggesting some of the examples used here.