BUSINESS STATISTICS

Using EXCEL & SPSS

0

Nick Lee & Mike Peters

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Kirsty Smy
Editorial assistant: Molly Farrell
Production editor: Nicola Marshall
Copyeditor: Neville Hankins
Proofreader: Andy Baxter
Indexer: Gary Kirby
Marketing manager: Alison Borg
Cover design: Francis Kenney
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound in Great Britain by Ashford Colour Press Ltd

MIX
Paper from responsible sources
FSC® C011748
FSC www.fsc.org

First published 2016

Library of Congress Control Number: 2015939926

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-84860-219-9
ISBN 978-1-84860-220-5 (pbk)

1

DEMYSTIFYING QUANTITATIVE DATA ANALYSIS

CONTENTS

learning objectives

This is the first chapter, and our learning objectives are simple really; however, there are quite a few, and they do cover some pretty foundational issues, so please try to bear the following objectives in mind when studying this chapter:

- ☑ Understand why people might be scared or turned off by studying quantitative methods.
- ☑ Understand how quantitative analysis for business is primarily to help you make decisions in your future career, even in what you think might be the most exciting and creative professions.
- ☑ Understand exactly what data is, and what is an element (or data point), a variable and an observation, and how these together make a data set.
- ☑ Understand the difference between qualitative and quantitative variables, and discrete and continuous data.

(Continued)

(Continued)

- ☑ Begin to see how we often approximate the amount of qualities (like happiness) by numbers, but that these numbers are not the exact same thing.
- ☑ Learn the names and properties of the four different scales of measurement – nominal, ordinal, interval and ratio.
- ☑ Understand the difference between cross-sectional and longitudinal data.
- ☑ Understand what a sample is, what a population is, and the relation between the two.
- ☑ Learn why variation is important in quantitative analysis.
- ☑ Learn the concepts of BEDMAS, exponential notation, powers/exponents and logarithms.
- ☑ Understand what equations and functions are, and how they can often be expressed in sigma notation.

If you are reading this, then you are probably beginning a basic, introductory or otherwise foundational quantitative methods course. It will probably be concerned with business studies in some way, but may not be. Many of you will be quite apprehensive at this stage. Like so many people in this world, you may not feel confident with numbers, mathematics and statistics. If you are anything like me, you might have not done so well at these in the past, and may even have forgotten much of what you learnt previously. Or perhaps you might not really see the point of studying quantitative analysis anyway; after all, your course is probably in something else like human resources, work psychology, general management or (like my undergraduate degree) marketing. So the aim of this chapter is to get you off on the right foot in your quantitative analysis studies from now on – whether you are straight out of school, coming back after some work experience, or whatever. I will go through some really key concepts of number theory, mathematics and statistics, and hopefully give you the basic tools to approach quantitative analysis with some confidence. But at the same time, I will do my best to help you understand why quantitative analysis is important for business studies and many other areas of life.

But the first thing to remember is, if you are scared, you are not alone. At this stage, most students are apprehensive about beginning a quantitative course. In fact, this chapter begins with the stories of two students who are pretty typical in my experience – of course they are not real people, but more a combination of characters I have met (parts might even be based on me, but I am not telling which).

NUMEROPHOBIA[1]

Quantitative data analysis is frightening. Yes it is. Go ahead, admit it, you are *scared* of numbers. If you aren't, (a) you are lying to yourself, (b) you don't even know enough to be scared yet, or (c) you might be one of the lucky people who were always quite good at it. If you are among the latter, then you'd better get used to getting a lot of late-night visits or calls from your colleagues (actually, if you play your cards right, you can work that to your advantage). But even if you are pretty good at it, don't get lazy – because, unlike many subjects, quantitative analysis can get very tough very quickly, and if you don't lay the foundations effectively, you *will* come unstuck at some stage.

However, I'll let you in on a little secret. It's not just students who are scared of quantitative analysis – your lecturers may be too! I know it's hard to believe, but it's quite probably true. In fact, you could call the quantitative analysis course in many business school (or other university) departments 'the graveyard shift'. Certainly in my own experience it's where bright new lecturers start out teaching (unless they claim some kind of stress-related psychiatric condition or something like that). I started out my career teaching market research, and did it for nearly 10 years. Occasionally, at 9 a.m. on a Monday morning,

[1] Fear of numbers. Not to be confused with hexakosioihexekontahexaphobia, which is fear of the number 666.

I would walk into the lecture theatre and see tumbleweeds blowing across the floor, so desolate was the environment. Perhaps I am exaggerating, but when I wake up screaming '*Central Limit Theorem*' in the middle of the night, I'm not so sure.

In fact, the only thing more scary than quantitative analysis is actually writing a book about it. Box 1.1 shows some of the most common sources of fear that individuals have about quantitative analysis, and some alternative ways of thinking about them.

box 1.1

Fear of Numbers

'I'm scared of being wrong.' Well, this is totally acceptable. In fact, when we learn something new, we are all scared of being wrong. However, what is needed is for you to try to divorce your feelings of success from being 'right' immediately. Try to break a numerical task up into small steps, and go back to the last part which you did get right, then go forwards from there.

'How do I know if I have it wrong?' Try to look at a problem as a set of steps which need to be followed in order to get to the answer, or a decision which needs to be carefully worked through to get to the end. I sometimes like to think of a problem as a recipe, in the knowledge that if I follow the steps, then I can't fail. All that is needed is a calm head and a knowledge of what steps to follow. Of course, just like cooking, with practice it gets easier.

'I can only do simple sums.' Well, everything in maths is built upon these simple foundations. In fact, it is a lot harder to learn the simple things without any prior knowledge than it is to build in little steps on top of those foundations – so you have already done the hardest part and you don't even remember it. Try to think of each new step as a simple layer on top; don't panic as it gets more complex, just make sure you learn it before moving on.

'I can't remember my times tables.' Neither can I. The trick in maths is not to fixate on memorizing those kind of things, but to understand the rules of the game. Once you do that, you don't need to memorize hundreds of numbers. Of course, memorizing basic things can help, but you can't rely on that for everything, so make it your task to understand the rules. Focusing on memorization will actually usually lead to a point where you get stuck because you didn't bother learning the concepts.

'But I don't know the answer.' Try to think of maths and quantitative analysis as a set of rules of increasing complexity. Each rule depends on your knowledge of previous rules. So, if you take your learning slowly, and try to understand all the steps which lead up to the solution, you'll actually know the answer in the end.

Finally, always remember that maths is very simple when you break it down to its component parts – it is a set of consistent rules about what to do. Understanding maths and quantitative analysis is the process of taking your time, and being confident with each small new step before moving on to the next. Remember, there was a time when you didn't know how to add numbers, or even what a number was.

If you are anything like I was when I arrived at university to study my first[2] undergraduate mathematics course, you don't know a heck of a lot about quantitative analysis. You might have studied it at school – you might even have got pretty good grades as well – but after the summer break you probably forgot most of it, or never really *learnt* it properly in the first place. Now there's a whole bunch of stuff that you need to know before you can approach a data set with some confidence and analyse it in a meaningful way. The

[2] To be honest, despite my future career being heavily based around applied mathematics and statistics, it remains my *only* proper university mathematics course.

early chapters of the book lay down these foundational concepts in both a theoretical and a practical way. As a student of an applied discipline (like management, marketing and the like) you are also probably not very interested in quantitative analysis. However, I would not be exaggerating when I say that the difference between someone who is a success in business and someone who is not, is often the ability to take data and manipulate it to draw useful conclusions.

The simple reason for this is that the business world is awash with data and numbers, and I'm going to show you how different professions use quantitative analysis in the next section. If you're a person who can cut through all those intimidating numbers, you are automatically in demand. Trust me. That's of course not to say there won't be plenty of examples from here on in, but this book is about explaining key quantitative analysis concepts as clearly as possible, and teaching you to think correctly about them – it's not a book about 'your' specialist subject (e.g. market research, financial mathematics, etc.). So use this book as the foundation stone for your course, and supplement it with the unique subjects you'll need as you move through the course and specialize more and more tightly. Then, at any time you can come back to this book and refresh your memory. In fact, I still do that with my own first-year undergraduate statistics books.

HOW IS QUANTITATIVE ANALYSIS USED IN BUSINESS?

When I first went to university as a 19 year old, I wanted to be in advertising (obviously, that didn't work out too well). I figured it was all about drawing cool pictures and playing 'creative games' in the middle of the day, and I thought that going to university to study business and marketing would be a good start to this, for some reason. One thing I did *not* expect was how important quantitative analysis skills would be in my degree. In fact, I was *forced* to study two quantitative modules in my first year – *two*! Maths *and* Stats! You might be having the same issues. Then, as I specialized in marketing, I had to study yet more quantitative techniques. Not to put too fine a point on it, I struggled for quite a few years with numbers. I just never seemed to be able to get my head around all the concepts. While this was likely to be because I am just fundamentally lazy and had what I thought were better things to do, it may also have been partly because I didn't see quite how business was not the 'seat of the pants' career I expected, but one where the smart and successful decisions are most often made based on the analysis of data. After all, I was quite keen on studying strategic decisions, or designing creative marketing campaigns, but I didn't understand how *those* decisions depended on an earlier process of data analysis. In this section, I'm going to try to give a brief picture of how quantitative analysis is vital to almost all business careers, which may help you to understand the importance it has to your own success. In fact, now that I know multiple creative directors from advertising agencies, I realize that even what I thought was a completely non-numerical field is often heavily dependent on data analysis to make decisions.

As I've already alluded to, the key way in which quantitative analysis is used in business is to help *make decisions*. In almost all cases, if you want to maximize your chances of making an effective decision, you need to base that decision on *information*. Different professions or situations need different types of information, but in almost all cases a large part of that information is numerical, either in the form of data which is specifically collected for the purpose (e.g. consumer satisfaction scores), or as part of the inherent nature of the profession (e.g. stock prices and trends as data for investment decisions). Some examples of how quantitative analysis is used in business are given below, along with some references to the appropriate part of this book where they are introduced. While almost all of these specialized fields will have dedicated courses and books about them, they all rely on the base material covered in this book – without having that under control, it is very hard to move on to more specialized methods, as I know from hard experience. In fact, numeric skills are useful in making decisions in all the different aspects of your life, now and in the future. Box 1.2 gives some interesting examples.

box 1.2

Decisions, Decisions...

Life is about decisions, and many of those decisions are about numbers. I've split things into three categories (your university life, your future career and your actual life in general), and below are some very simple examples of how good quantitative skills can help you out.

University: One of the more useful skills in university life is time allocation. In order to allocate time, at minimum you must know basic mathematics. You will also find that your numeric skills will come in handy in allocating your finances, or dividing up who will pay what at dinner (the cleverest mathematician can sometimes make a profit here).

Career: Whatever career you have in mind, numeric skills will serve you well. Apart from the specific demands of the job, if you have to decide between jobs the decision often comes down to finance, and relying on headline salary figures is often misleading. Sometimes you will need to weigh up more complex packages against one another. People without numeric skills often 'can't be bothered' weighing up things like distance, travel costs, benefits and the like, along with salary.

Life: Life is full of decisions which require numbers, so much so that most of the time you do not even realize. Every time you measure something, or convert one unit (e.g. kilograms) to another (e.g. pounds), you actually perform quite a complex procedure. Of course, quantitative skills can also lead you to maximize your financial potential as well. In fact, this book may end up making you a lot of money!

think it over 1.1

Can everything that varies (i.e. what are called variables) be reduced to numbers?

Think about feelings. When you say you love something or at the opposite end hate something, how can you measure it accurately?

What about when you get exam results? For example, you score 98%, top of the class, and you are asked to complete a questionnaire about the quality of teaching in your class – what will your response be? What if you got 35%, a fail – would this make you feel different about the quality of teaching?

think it over 1.2

Some people just 'know' something is right. When you ask them how they 'know' it is right, they reply with statements such as 'it feels right', 'my gut instinct tells me it's right' or 'just intuition'. How, if possible, can you account for this?

Accounting

It is no surprise that accounting is highly quantitative. Balance sheets, cash flows, incomings and outgoings are inherently numerical. There is a vast difference between a profit and loss account that shows a profit of £24.6 million and one that uses the words 'quite a lot'. Many of the basic accounting tools and techniques are reliant on quite simple mathematics (the skill is in their correct application), but there are also lots of statistical methods employed in many large firms. For example, an auditing firm needs to validate whether or not a company's figures are accurate. However, for a large company it is impossible to examine every single transaction to see whether it tallies with the 'accounts receivable' amount shown on the balance sheet. Thus, auditors will extract a sample of the total accounts to analyse for accuracy. From this sample, the auditors can make a judgement with a certain level of confidence as to whether or not the full figure is accurate. As well as this, accountants often need to work out things such as depreciation and amortization of assets or finances. You'll be pretty confident with the basic mathematics by the end of this book, and sampling is covered in Chapters 6 and 7.

Economics

If you watch or listen to the news, you will likely hear the results of the quantitative analysis done by economists every day. They are often concerned with forecasting future values of numbers such as the gross domestic product (GDP) of a country, or the inflation of prices on important goods such as fuel or food. Despite what many may think sometimes, economists do not pull these forecasts out of, uh, 'thin air', but instead rely on complex mathematical models. These models are generally based around a regression framework, which is introduced in Chapters 10 and 11, and expanded upon in Chapter 15.

Finance and Banking

Finance is generally considered to be a highly mathematical field. Much of financial mathematics concerns techniques and models for more accurately forecasting future profits (or some other monetary value such as share price). The more accurately one is able to do that, the better the decisions about where to invest can be made. In fact, with the amounts of money being invested by the large investment banks today, even a tiny increase in the performance of these models can lead to a huge increase in profits. The foundational mathematics of this is similar to that of economics, but finance is also the most obvious business application of all that calculus you learnt at school and didn't understand why. Finance also uses a lot of statistical information – particularly in comparing the key figures of one potential investment target against market averages over a longer term. This helps analysts decide how much emphasis to place on things like short-term fluctuations of stock price. Critical forecasting concepts are introduced in Chapter 15, but those of you studying finance subjects will be using a lot of specialized texts beyond this one.

think it over 1.3

In the banking crisis of 2008–9, do you think investment decisions were made purely on the available numeric data? If they were, do you think the crisis would have happened? If they were not, what factors do you think caused the problems?

Human Resource Management

Human resource (HR) professionals are also reliant on numerical data. In particular, in order to design and conduct successfully any large-scale HR initiative, data must be collected about the organization first. In fact, in my own organization I seem to be responding to a questionnaire about staff satisfaction or opinions almost every week. Members of the HR department collect and analyse this information in order to determine in which direction to take the overall HR strategy – or at least I assume they do. In doing so, they must understand sampling, and especially sampling and non-sampling error (see Chapters 6 and 7), to understand how accurate their results are. They will probably also need to understand the correlation and regression methods in Chapters 10 and 11 to discover if some factors are linked to others, and also they will probably need to compare different groups of employees (e.g. males and females, or different departments) which may necessitate *t*-test or ANOVA analysis methods, which are covered in Chapters 12–14.

Production

There are many areas of production which require quantitative skills. Perhaps most interesting is the need for increasing standards of quality. As competition has intensified, quality has assumed an ever more prominent place in the mind of the consumer. Modern statistical methods have allowed the organization to monitor accurately the quality of a production process. For example, control charts, which are covered in Chapter 16, help managers to make decisions about whether or not a process is performing within an acceptable range of quality. Using control charts and other methods of quality control involves concepts from throughout this book, including sampling methods, and many of the basic statistical foundations covered in the early chapters, such as confidence intervals (Chapter 7) and statistical hypothesis testing (Chapter 8).

Marketing

I was shocked to discover quite how quantitative that marketing, my own field of choice as an undergraduate, was. Effective marketing decisions should almost always involve research which looks at customers – often done through administering questionnaires to a sample of consumers. You have probably participated in this yourself on your own main street. But even more sophisticated methods are now becoming the norm. For example, major market research agencies collect data from supermarket scanners and sell their analysis of this to other organizations. Such analysis is often based on the regression and forecasting methods introduced in Chapters 10, 11 and 15. In this way, consumer goods firms can get reasonably accurate information on key indicators such as market share, and track how effective their promotions are. This in turn helps in developing future strategies. Many firms now also have gigantic consumer databases (Google for instance has an almost unimaginable store of consumer information) and sophisticated data mining methods allow firms to use this data effectively.

Retailing

Of course, retailers themselves can also use the scanner data I referred to above, to help them to understand how consumers may behave in their stores. For example, does positioning a product in a different area of the store result in changed sales results? Or do various methods of promotional discounts result in increased profit? Such analysis relies on regression (Chapters 10 and 11) and also sometimes on the ANOVA methods in Chapter 14. Retailers may also be interested in understanding traffic flows over time and the profit potential of various locations. In order to understand these issues, knowledge of forecasting (Chapter 15) is very helpful.

Strategy

Tying it all together is strategy. In order to make strategic decisions with a high chance of success, managers need information. This information is most likely to be in numerical form, perhaps drawn from one of the areas already mentioned. For example, in order to make investment decisions, strategic managers should have some knowledge of financial analysis. In order to make decisions about what business the firm should be in, forecasting and economic analysis will be useful. Naturally, many other situations will require excellent quantitative skills. While of course the strategic manager will use many more specialized managers for advice, in order to make a decision the manager must *understand* the advice given to him or her by the – say – financial expert. Thus, even at the very highest levels of the organization, the foundational skills and tools presented in this book are vital to success.

THIS SECTION IS REALLY IMPORTANT!

Sorry, I didn't mean to shock you. I just wanted to make sure I caught your attention, because this section actually *is* really important. The common thread running through all the above examples is the idea of *data*. Quantitative analysis manipulates numbers, but those numbers are not just any numbers without meaning, they are data. Data is information in its raw form. It is the information that you have collected or obtained in some way to answer the question you have. In the context of this book, the data we will be working with is numerical, but it does not have to be, as we will see soon. The different subsections to follow all describe key ways in which you can describe the characteristics of data, and they are all important, so do make sure to read through them as they will come in handy later on.

The collection of data to be used for a given task or study is known as the data set, and Table 1.1 shows a typical data set. In this case it is a set of figures describing the characteristics of various companies. A data set like this is rarely of use by itself; it needs to be *analysed* or somehow presented. You can see this in Table 1.1. Even with only 20 companies it is hard to get a feel for the data, but imagine that there were 100, 1000 or even 100000 firms! That raw data set is essentially meaningless by itself, without analysis.

think it over 1.4

Is there a difference between data and information? If you just read the number 10 would it have any meaning? But if you read £10 would that be different?

There are three components to every data set. The elements of a data set are the entities (i.e. 'things') about which the data is collected. In Table 1.1, the elements are companies, but they may be individuals, countries or anything that you have collected data about. Sometimes in more academic circles you can hear elements referred to as data points, but this is not common in business circles. The data set shown in Table 1.1 contains 20 elements, or data points. Possibly the most important term to understand right now though is *variable*. A variable is any specific characteristic of the elements you have collected data on. In the case of Table 1.1, the variables are *firm* (the name of the company), *turnover* (the value of the annual sales of the company) and *employees* (number of employees of the company). Finally, the set of measurements of each variable for one element is called an observation, or sometimes a case. In other words, looking at Table 1.1, observations run *along the rows*, whereas variables run *down the columns*. Or at least that is how a good data set should usually look.

Now we have the basics out of the way, let's start to think a little more deeply about data. Perhaps the most important thing to keep in mind when you are dealing with a numerical data set is that *the numbers in the data set are not always numbers in the real world*. What is that supposed to mean? Well think about it carefully. In Table 1.1 it is pretty self-explanatory: the number of employees is a number, and the turnover is also a number, so no worries there. But what about if we consider the data set in Table 1.2? Here I have added some average customer satisfaction scores. Imagine that each company asked randomly selected customers to say how satisfied they were on a scale of 1 to 5, 1 being 'not at all satisfied' and 5 being 'very satisfied', like that in Figure 1.1. Then we took that score and added it to the data set from Table 1.1 to get Table 1.2. We could use that data set to examine lots of interesting questions, as you will see later in the book.

Table 1.1 An example data set

Firm	Turnover	Employees
Jones Bros	14000000.00	74.00
Dataforce	14000000.00	1000.00
Smiths	13000000.00	130.00
AMBD Ltd	12000000.00	69.00
X-FS	12000000.00	40.00
Signum	12000000.00	30.00
Korg	10000000.00	48.00
Imatix	10000000.00	48.00
147.com	10000000.00	48.00
Stevensons	10000000.00	48.00
Analysis Ltd	10000000.00	48.00
Marketize	10000000.00	48.00
Jones Partners	10000000.00	48.00
Endoff	10000000.00	300.00
BDM Co	9000000.00	40.00
Kozelek Inc.	7000000.00	700.00
Chambers	7000000.00	3000.00
Controls	5000000.00	32.00
Star Products	4000000.00	20.00
Nick's Emporium	3303000.00	30.00

think it over 1.5

In statistics we talk about data all the time. Quite often it is given in a numeric form. Would you say these numbers are purely data (no meaning) or do they need to be understood in the context of the situation which is being analysed?

Strongly Disagree		Neither Agree nor Disagree		Strongly Agree
1	2	3	4	5

Figure 1.1 Example Scale

However, you should not confuse the *number* in the 'satisfaction' column with the *real customer satisfaction* of a consumer. The number is simply an indicator of the satisfaction. The most basic illustration of this is that the 1–5 scale only contains whole numbers. What if my real satisfaction is not 1 or 2, but somewhere between? The scale cannot capture this. So the number is a crude approximation of what is probably quite a complex thing.

Table 1.2 Example data set 2

Firm	Turnover	Employees	Satisfaction
Jones Bros	14000000.00	74.00	1.00
Dataforce	14000000.00	1000.00	4.00
Smiths	13000000.00	130.00	5.00
AMBD Ltd	12000000.00	69.00	2.00
X-FS	12000000.00	40.00	3.00
Signum	12000000.00	30.00	4.00
Korg	10000000.00	48.00	5.00
Imatix	10000000.00	48.00	3.00
147.com	10000000.00	48.00	1.00
Stevensons	10000000.00	48.00	2.00
Analysis Ltd	10000000.00	48.00	2.00
Marketize	10000000.00	48.00	1.00
Jones Partners	10000000.00	48.00	1.00
Endoff	10000000.00	300.00	4.00
BDM Co	9000000.00	40.00	5.00
Kozelek Inc.	7000000.00	700.00	3.00
Chambers	7000000.00	3000.00	3.00
Controls	5000000.00	32.00	2.00
Star Products	4000000.00	20.00	5.00
Nick's Emporium	3303000.00	30.00	4.00

This is the essence of what we are doing when we use numbers to represent many concepts that we are interested in. We are no longer playing with 'real things', but instead with numerical *representations* of things. Using numbers allows us to use the amazingly useful tools of mathematics and statistics to answer our questions, but we should never forget that we need to link these tools and the answers they provide back eventually to the 'real world'. The numbers by themselves are not useful. Figure 1.2 illustrates this very simply.

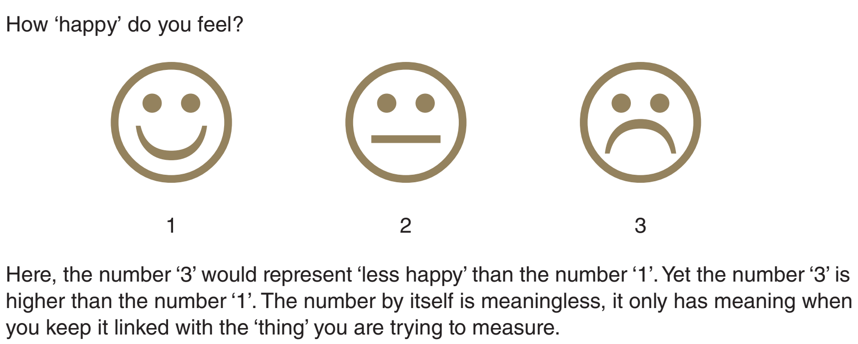

Figure 1.2 Linking Numbers to the Real World

Qualitative and Quantitative Variables

The most obvious way to understand the difference between numerical data and the real-world qualities it often represents is to consider the difference between *qualitative* and *quantitative* variables. This is a book about quantitative analysis, but often the things we want to analyse do not come naturally in numerical form like, say, profit (a quantity of money) or age (a quantity of time) does. Qualitative variables are those like the 'firm' variable in Table 1.1. These types of variables are often labels or names of things. There is not much we can do with such variables in terms of mathematics or statistics, other than count them, work out their proportions or use them as labels. Nevertheless, they can be important in your analysis, for instance to summarize the data set – which will be discussed later when we come to Chapter 4 on descriptive statistics.

On the other hand, quantitative variables contain numerical data which indicates *how much*, or *how many*, of a variable there is. These variables are like the 'employees' (how many) or 'turnover' (how much) variables in Table 1.1. With such variables we can perform all kinds of useful mathematical and statistical analyses, which will make up the bulk of the tools explained in this book.

But what if I were to use numbers instead of words to represent qualitative variables in our data set? Take a qualitative variable like biological gender – male or female. Instead of writing 'Male' or 'Female' in the spreadsheet, I could make '1' represent 'Male' and '2' represent 'Female'. Then I have numbers in the spreadsheet, not words, so surely I can use my fancy mathematical tools on them? Unfortunately not. You see, the numerical value in that case is still just a label; it does not represent anything quantitative about the variable. There is not 'more' of anything to do with females than males, *even though the number for 'Female' is bigger than for 'Male'*. This is the crux of the distinction between quantitative and qualitative.

You might at this point be thinking back to that customer satisfaction example earlier, or the happiness example in Figure 1.2. Surely, these are qualitative concepts as well? Well, yes and no. Satisfaction is not *naturally* numerical, like profit, or share price, might be. But it is quantitative in that it is meaningful to consider 'how much' of it there is. In such a case we can *measure* it with a number, not just assign a label to it.

This concept is illustrated well by an explanation of the different scales of measurement.

think it over 1.6

In the satisfaction survey, if a person rated their satisfaction as 4, does that mean they are twice as satisfied as someone who responded with 2?

Scales of Measurement

So, we now know that we can represent 'quantities' of things in the real world with numbers in our data set. But we cannot take off our thinking caps yet. You already know many basic rules for working with numbers: you know how to count, add, subtract, multiply, divide (at a minimum, you can do these on a calculator) and the like. Now, these operations and rules all apply to numbers – give us a number and we can operate on it. But the same rules do *not* always apply to the real-world variables we are working with. For example, remember above that I labelled 'Males' with '1' and 'Females' with '2'? Well, I can certainly do all kinds of things with those *numbers*, but many of them will just not be logical if I was to do them with the real-world concept of biological gender. For example, if I had 3 males and 3 females, I could take the average of the gender variable and come up with 1.5 as my answer. But a value of 1.5 does not make sense in the context of my real-word concept, which can only take the value of male (1) or female (2). What is 1.5 meant to be? So, we need *rules* that can tell us what we can logically do with the different qualitative and quantitative variables we have in our data set, even though they may all be using numbers. Most of the time, we tend to reject rules, but in the case of quantitative analysis, rules are good because they make life easier. Try to think of them as instructions for a recipe. If you do not know the rules, the recipe will not work!

First of all, there is the *nominal* scale. This scale refers to the situation where we are using numbers to label a qualitative element. The simplest example is biological gender as above. Biological gender is a characteristic of an element in a data set – you can be either male or female. Similarly, the 'colour of your underpants' and many other things. In such cases we *could* use words as labels. But we often like to assign a numeric code to the different possible values of the variable (e.g. 1 = 'red', 2 = 'orange', 3 = 'blue', etc.); this can make life easier for us in the long term. However, in such cases the numbers are only *labels*, there is no logic behind their order or value. The only mathematical thing you can do with this sort of variable is *count* it. You can thus examine the comparative amounts or proportions of each different value. This is the simplest rule.

If there is some meaning to the *order* of numbers we use to represent the variable, we can term the scale *ordinal*. In this case we are now working with quantitative concepts. For example, we could ask for ratings of consumer satisfaction on a scale of 'excellent', 'so-so' or 'rubbish'. In this case, we know that 'excellent' is better than 'so-so' which is better than 'rubbish'. So there is a meaning to the order. But we do not know whether the difference in satisfaction between 'excellent' and 'so-so' is the same 'amount' (also called magnitude) as the difference between 'so-so' and 'rubbish'. We can label these as '3', '2' and '1' from excellent to rubbish respectively if we want. However, even though the differences *between the numbers* are the same (i.e. 3 is one bigger than 2, which is one bigger than 1), this still does not mean the differences *between the values* of excellent, so-so and rubbish are the same.

If we can see that our variable has some kind of fixed unit of measurement (i.e. the differences between values are consistent) we can use an *interval* scale. Consider the typical temperature scale in degrees Celsius. This is a great example of an interval scale. The difference between 20 and 30 degrees Celsius is 10 degrees. This is exactly the same amount of heat that there is between 60 and 70 degrees Celsius. So just like an ordinal scale, we can rank the scores from low to high (or vice versa), but the *differences between the scores* are also meaningful. This scale enables us to use more mathematical techniques again, and is yet more powerful. But it still lacks the full range of properties of the standard number sequence. For example, consider the statement 'it was twice as cold as it was yesterday, and yesterday it was 0 degrees Celsius'. So how cold does that make it? Another interesting example is the Western date in years. We know the difference between AD 1950 and 2000 is the same as that between AD 150 and 200, that is 50 years. But think about what the date 'AD 0' represents.

What I am getting at is that an interval scale lacks a meaningful zero point. Even though it may have 'zero' in the range, this zero point does not mean there is 'nothing'. Zero degrees Celsius simply means that water freezes; AD 0 does not represent the beginning of time, just the beginning of a certain number sequence representing time. To be *meaningful* the zero point must indicate no value of the variable at all. In such cases, we have a *ratio* scale. The term ratio refers to the fact that once we have a meaningful zero, ratios of the variable have meaning. For example, the ratio temperature scale is the Kelvin scale. Zero Kelvin refers to a total absence of temperature (which is in fact around –273 degrees Celsius). If we think about 80 degrees Celsius, this cannot mean there is 'twice as much temperature' as 40 degrees Celsius, the ratio is just nonsense without the meaningful zero. But, 200 degrees Kelvin *is* twice as much temperature as 100 degrees Kelvin. In a business context, *profit* is an excellent example of a ratio scale. Although things get a bit more complicated if you start thinking about negative profit (i.e. loss). Even so, there is such a thing as 'no' or 'zero' profit.

One key thing to realize here is that the most powerful scale we can use is defined by the real-world variable in question. For example, for a variable like gender, we can only ever use a nominal scale. However, for heat, we can use ratio if we wish, but also we could use interval (Celsius), or even ordinal (say 'hot', 'cold', 'brrrrrrr!'). This means you should always think carefully about what you are trying to measure and why. Furthermore, the scale you use has impacts on what you can do with the data mathematically. For example, even though ordinal scales have some meaning to their magnitude, gaps between the values are *not* meaningful, so you cannot take the mean of the ordinal scale. The important point to keep in your mind is – again – that the numbers you see in front of you, that you will be working with throughout your courses, are only *representations* of real-world concepts, properties and objects.

The Temporal Dimension

Another key distinction between different types of data concerns time. In case you did not know, 'temporal' is just a fancy way of saying 'time' and something which has always impressed me – so I figured it might impress you too. Anyway, imagine these two different types of research project: (a) we collect data about the balance sheets of 100 companies at the end of the 2008–9 financial year; and (b) we collect data on the profit of one company over the course of a decade, collecting data each month.

The first situation results in what is called a cross-sectional data set. This data set is collected at a single point in time – or at least, an approximately single point in time. Perhaps the best way to think about it is that the *intention* is for the data to represent a 'snapshot' of the situation at one point in time. Of course, it is usually impossible to collect data at exactly the same point in time across many cases.

On the other hand longitudinal data is collected over a long time period, which is why this type of data set is sometimes called time series data. The intention here is for the data to represent how the data changes over time. For example, Figure 1.3 shows a time series for the amount of sales calls made by a salesperson in one month.

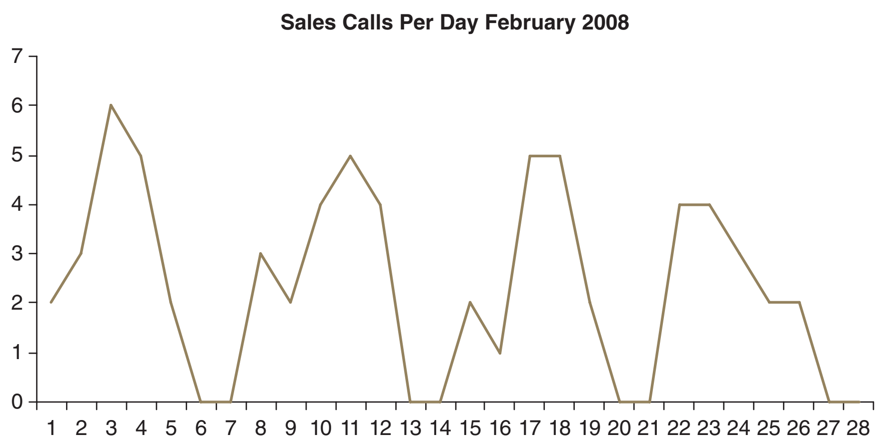

Figure 1.3 Example Time Series

While this is not a concrete rule, you often find that cross-sectional data is collected about many variables at one time, whereas time series data can be collected on a single variable over many time periods. Most of the analysis tools in this book will look at cross-sectional data, but time series data is specifically covered in the later chapters, for example on forecasting. Whether you need cross-sectional or time series data depends on your research goals, but that is outside the scope of this book, which is mainly aimed at telling you what to *do* with the data when you have it.

think it over 1.7

Which is better, cross-sectional data or longitudinal data? Why do we have these two techniques?

Samples and Populations

The concept of sampling, and the linked concept of the population, will come up again and again in this book. However, it will be handy for you to have come across it before getting to later chapters. I already snuck in a reference to sampling earlier on, when we talked about what auditors sometimes do, so it should not be totally new to you (you probably learned about it before as well).

In essence, it is usually impossible for us to collect data on every element we are interested in. For example, perhaps we want to know the turnover of every company with less than 100 employees in the UK. Theoretically, this may be possible to collect if we have the resources (maybe we are the UK government), but most likely it is impossible. Even if it is possible, what about if we wanted to know about every company in the *world* with less than 100 employees? Or what if we wanted to collect satisfaction scores from every purchaser of Heinz Baked Beans? The population is a term which refers to the set of all of the elements we are interested in. In one of the earlier situations, the population was 'every company in the UK with less than 100 employees', and in another 'every company in the world with less than 100 employees'. It is important that analysts and researchers understand clearly exactly what population they are interested in, even when they cannot collect data on it in its entirety.

In fact, most of the time it is just not feasible to collect data on every element. If we can collect data on every element, we are doing what is called a census. Such exercises are done by many national governments, and their relative infrequency should give you an idea of how resource intensive they are. More commonly, we have to take what is called a sample. A sample is any subset of the entire population. There are many different ways of collecting samples, and I will go through sampling theory in Chapter 7. In fact, most of the specific research courses you take later on in your career (e.g. market research) will spend a long time talking about sampling. This is basically because we are more interested in applying the results of the analysis of our sample back to the population than we are in the sample results. As a consequence, there will be a lot of information in this book which is related to the issue.

Discrete and Continuous Data

The final distinction I will cover in this section is between discrete and continuous data. This is covered in some depth later on in this book (e.g. when I introduce ideas of sampling, probability, distributions and the like), but it will help you a lot if you get the idea bubbling away in your head right now. That said, it is quite a difficult thing to get your head around at first, so do not worry too much if you do not 'click' with it immediately, I will be going through this again at various points of the book. It is important that you meet it right now, though, because having this idea somewhere in your mind throughout the rest of your reading will help you ultimately to understand it when you *really* need it.

Consider this: I collect data on the age of 1000 people, asking them to give me their age to the closest full year. I can then count the frequency of occurrence of each year and I could plot it on a graph or something like that. So, basically, I would have something where maybe there were 45 people who were aged 30, 15 were aged 19, 20 were aged 50, and so forth. This kind of data is discrete. In other words, there is separation between the possible values which an element can take – you can be either 20 or 21, but not both, and not 21.5. Get it?

Now, work with me here and imagine the following situation. Try really hard because it is quite challenging to get your head around. Imagine we made the way we measured age finer and finer – first to months, then weeks, then days, minutes, seconds, and keep going. You can imagine that as the way to measure age got more fine-grained, the number of people of each age would get smaller and smaller. OK? So, at the same time imagine that we kept increasing the number of people we were measuring as we made the measure finer. Imagine if you will that in fact we kept *infinitely* making the measure finer and finer, while increasing the amount of people we were measuring at the same time. Eventually (remember, *infinity*) there would be no distinction between one age and the next.[3] This kind of data is continuous. In other words, there is no separation between the possible values of a variable each element can take. There are many real-life situations where data is essentially continuous if we do not artificially divide it into units (e.g. time), but the real benefit of this concept is evident when it comes to the statistical theories which we will come to eventually in Chapter 5.

think it over 1.8

Does continuous data really exist or is it just a convenient way of thinking about data? How many numbers are there between 0 and 1? What about 0 and 0.00000001? If I wanted to measure the temperature of a room accurately, what are the decisions I would have to make?

A RECAP OF KEY MATHEMATICAL AND STATISTICAL CONCEPTS

One of the things which kind of scared (or should that be 'scares'?) me about quantitative analysis is that I always felt a little bit lost in the terminology of basic foundational concepts. In other words, I never felt that I had learned the 'language' of mathematics and statistics. So, just seeing an equation or function on a sheet of paper was enough to scare me away. As a consequence, I think it is really important for beginning students to have a grasp of these things before moving on to anything else. I will begin by discussing the concept of statistics itself. After this, there will be a set of critical mathematical tools, concepts and operations which you should understand intimately. Finally, I will conclude with a section about the basics of mathematical functions and symbols, which is something I always struggle with myself.

What Are Statistics?

Since I mentioned statistical theory above, now would seem to be a good time to introduce the idea of statistics. First let's think about why we might need statistics. If you go back to Table 1.1, you will see that I made the point that it was hard really to draw any conclusions about raw data without analysing it. A huge set of raw numbers in rows and columns is pretty meaningless to us. What we really need is some way of summarizing the characteristics of the set, or extracting a smaller set of information from it. For example, we might want to calculate the average turnover of the companies in Table 1.1. Actually, it is very imprecise to use the word average, since there are many different ways of calculating average – but I will deal with this in Chapter 4. Anyway, moving swiftly onwards, if we were to take the average of every company in the population, this would be called a parameter. However, when we are dealing with the analysis of a sample, it is called a statistic.

Most of the time, we are actually not very interested in the statistic by itself. We are in fact more interested in the population parameter. Recall the accounting example above. The auditors were not interested in the sample statistic itself, they were interested in whether it was *the same as the value the company they were auditing claimed for the parameter.* This task is what we call statistical inference. It is the task of understanding the *likelihood of our sample statistics being the same as the population parameters*. Or at least that is it in a nutshell. Most of Chapters 5–8 are concerned with this, and the rest of the book after Chapter 8 generally depends on that information as well. So it is kind of important.

[3] I know infinity is a tough concept, but just run with it for now.

Variation and Statistics

Building on the previous brief discussion of what statistics are, I want now to talk about the concept of variation in data. You can see from lots of examples already in this chapter that the variables take on different values for different elements. In fact, the clue is in the name – we call them variables, remember. The idea of variation is an absolutely vital one in statistical inference and other quantitative analysis fields. Again, variation will be covered in much more depth later on, but basically try to imagine the situation where you wanted to see whether one factor (maybe something simple like gender) was associated with a change in another (something more controversial like, say, intelligence test performance). Continue to imagine the situation where we have 50 males and 50 females, but that each of those elements scored exactly the same on our intelligence test. So how could we tell whether differences in test scores were associated with differences in gender? Simple answer: you cannot! So it is easy to see how the concept of variation is key in quantitative analysis methods. In fact, we can measure it in many ways, which will be introduced in Chapter 4.

In theory, variation can be partitioned into at least two types. The first is systematic variation, which is basically variation which we think is due to something important – usually what we are testing for. In the case above, systematic variation is that which is due to the gender effect (i.e. if females are smarter than males). The second is unsystematic variation, which is the variation due to everything else – sometimes called error or random variation.

When we are 'doing statistics' – or in other words when we are performing statistical inference – we are basically trying to work out the chances that the numbers we observed in our sample are the same in the population. We do this by first calculating what are called test statistics. There are many (hundreds at least) different test statistics, each used for different purposes. I will be covering many of them in the later sections of the book, but the basic concept is simple. A test statistic is simply a number with specific characteristics. Most importantly, we know how often the different values of it occur in any given situation.

As I said, there are many of these, but it is easy to understand when you think about something a bit more simple, like a characteristic of people. Take something like reaction time, for example. Studies have shown the average reaction time for all kinds of tasks; let's take 'pushing a button in response to a sound' as an example. We know that the average reaction time for a human to push the button after hearing the sound is some amount of time, which we can represent with the symbol x. We also know the distribution of reaction time, in that we know the probability of the various different reaction times. This makes basic logical sense if you think about it. If the *average* time is x, then, if we take any given person, it is *most likely* that they will have a reaction time of x or close to it. Therefore it is *less likely* that they will have a reaction time either a lot slower or quicker than x. Flip that logic around and, once you know x, then if you were to measure your own reaction time, you could work out the probability of observing that time, given that reaction time has a known distribution. In other words, are you around the average x, or are you much lower or higher? This will be covered in much more depth in Chapters 6–8.

Moving back to variation, these test statistics are basically calculated by working out how much of the variance in your data is systematic, versus that which is not. Of course, it is more complex than that, and each statistic works it out differently. Because we know the distribution of these test statistics, we can calculate how likely it was that we observed that particular value for the test statistic. This is the whole point of statistical inference, and I will begin to discuss it in depth in Chapters 7 and 8.

think it over 1.9

If I used a reliable statistical technique to measure the reaction times of 1000 people, could I use the answer to predict the reaction time of one person?

Basic Mathematical Concepts and Tools

Now, let's get some critical mathematical revision out of the way. Of course, it is impossible to cover absolutely every basic mathematical and numerical concept here; perhaps one day I will write a book about that too (good lord, I am only at the very beginning of this one so far!). I do have to assume some basic knowledge. Nevertheless, there are some pretty handy things which might have slipped your mind, or been taught while you were, ahem, 'sick' at school, which will be quite useful here. These are basic mathematical tools which are very simple but the cause of great trouble if one does not know them.

First, when I was younger at school in New Zealand, I was taught the acronym BEDMAS, which stands for Brackets, Exponents, Division, Multiplication, Addition, Subtraction. It refers to the order in which you should calculate in an equation. In the UK system, I am told it is in fact referred to as BODMAS,[4] which is the same thing, but Exponent is termed 'Order'. I think that is confusing, because 'order' refers to a specific mathematical thing, but this is within an acronym about order in a different sense. It seems a bit silly to me, especially when there is a ready-made alternative. So let's work with BEDMAS. Basically, BEDMAS tells you how to calculate in any simple equation, and is quite important. First, however, let me define exponents (aka orders, aka indices). This is simple: for example, 5 squared is written mathematically as 5^2, where 5 is the coefficient and 2 is the exponent (also termed 5 to the order of 2, or 5 to the power of 2). So let's see an example, in Figure 1.4.

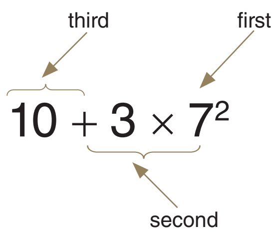

Figure 1.4 BEDMAS

think it over 1.10

Why are exponents useful? What about algebra in general, using letters to represent quantities?

Of course, the answer here is 157 if you follow the order shown in Figure 1.4. If you place brackets in that equation, you can actually change the answer. If you bracket it to make $(10 + 3) \times 7^2$ you will get 637, whereas if you bracket it to make $10 + (3 \times 7^2)$ you will get 157 again. The latter is because, in this case, the brackets do not change anything (you were already doing the exponent before the multiplication before the addition anyway).

think it over 1.11

Why are there rules like BEDMAS? What happens if we violate these rules?

[4] Actually, when I looked at my friends' daughter Lucy's maths homework last week, they used BIDMAS, where the I stood for 'indices' – which is again another term for exponents/order. BIDMAS works OK for me too (it's definitely an improvement on BODMAS), so you might have heard of that term.

Scientists and other quantitative gurus often represent numbers in exponential notation. This is because such people often work with huge or very small numbers which are rather unwieldy to write out. Think about the mass of the Earth, which is around 5973600000000000000000000 kilograms, or the mass of an electron, which is something around 0.00000000000000000000000000000091093822 kilograms. Imagine having to write out those kinds of numbers repeatedly, or even once. Mistakes will happen, I promise you. Instead, if you use exponential notation you can express the number as *something multiplied by 10 to the power of something else*. Or more scientifically, as a coefficient multiplied by $10^{\text{exponential}}$, where the coefficient is a number between 1 and 10, with the relevant exponent. So in this notation, the Earth's mass is 5.9736×10^{24} kilograms, and an electron's mass is $9.1093822 \times 10^{-31}$ kilograms.

What confuses some people (myself included) is that calculators often express this a bit differently. Usually, they use an 'e' or 'E' to represent the 'times 10 to the power of' bit. So the Earth's mass would be 5.9736e24. I always used to think the 'e' represented 'error' and that I had done something wrong. This caused much heartache at school and university (as if I didn't have enough already).

Exponents or powers (sometimes called 'orders' as well, as we found out above) are basically representative of growth, with x, x^2, x^3, being named the first, second, third (and so on in sequence) powers of x. Of course, you will probably remember that these lower powers (second and third) are called *squared* and *cubed*. To introduce for the first time some algebraic notation, using n to stand for 'any number', you could say that x^n is the nth power of x. Incidentally, if you read Above and Beyond Box 1.1 you might wonder about the power of 0. Well, any non-zero number raised to the power of 0 equals one, which can be proved mathematically if you want.[5] Do not even ask about 0^0, when things get even more complicated!

You can think of logarithms as the inverse of exponentials. Logarithms (called 'logs' by those who want to look cool) are important because they help to describe a lot of interesting properties in the real world, as you will see later on. A logarithm is the power (explained above) to which you must raise a given number in order to attain a specific other number. A logarithmic expression contains two main elements, the *base* and the *power* (or exponent). For example, the logarithm of 10000 to the base 10 is 4, because 10^4 equals 10000. This would be expressed as $\log_{10}10000 = 4$. The logarithm to the base 2 of 128 is 7, because 2^7 equals 128 ($\log_2 128 = 7$). The main bonus from the development of logarithms is the fact that they can reduce multiplication to addition, which makes complex calculations much easier if one has to do them by hand. They do this by virtue of the formula $\log(xy) = \log x + \log y$. Handy. Base 10 and 2 are the most common simple logarithmic bases, but you will also very commonly come across what is called the natural logarithm, which is a logarithm to the base e, where e is an irrational constant (see Above and Beyond Box 1.1) which is *approximately* equal to 2.718281828. This logarithm is super-useful in mathematics, statistics, economics and various other fields. It is sometimes noted as ln, Ln, $\log_e$ or even just log. However, log can also mean base 10 – which is confusing. It is called 'natural' because it appears so often in mathematics.

Finally, Above and Beyond Box 1.1 goes into some very interesting different types of numbers, which might come in handy somewhere along the line…

above and beyond

Box 1.1 Strange Numbers

In this box, and others throughout the book, there are terms and concepts which are not strictly necessary to understand the material here, but which might help you build your confidence in mathematics and statistics by illustrating interesting terms and concepts you might have heard about but not understood. They also may put more familiar terms into a new and interesting context.

[5] Don't email asking me to do it, find it for yourself!

Consider the number 0. What actually *is* it? It is generally accepted to have been invented at the same time by the Chinese and Hindi cultures, around the sixth century AD. Being more concerned with practical uses of maths than rules, the Chinese came up with it simply as a 'placeholder' to represent the missing place in numbers such as 'one hundred and one' (i.e. 101). By contrast, the Hindi civilization considered it in a more philosophical manner, concerning their idea of 'The Void'. This latter idea is vital to the meaning of zero, because it is very weird (and is a good example of how seemingly simple mathematical concepts become very odd if we think them through to their extreme). On the one hand, zero is just like any number, in that it can be added and subtracted without any weird things happening. But what about multiplying? Multiply anything by zero and you get zero, and divide anything by zero and you get infinity! And also, you do not *count* using zero, you start at one. That for example is why I used to get so confused about centuries – I was born in the twentieth century, because we started counting in the first, not the 'zeroth'. The more you investigate zero, the weirder it gets. But there are plenty of other odd numbers which you could find out about.

Prime numbers are those which cannot be divided by any other number apart from themselves and one. What this means is that if you divide a prime number by any other number apart from itself or one, you get a fraction. You may think this has no real point, apart from being interesting, but there are applications for this, for example in creating encryption algorithms – such as those useful for Internet applications. There are an infinite number of primes, and there is an ongoing competition for finding the largest known prime.

Rational numbers are another name for fractions. They are numbers which can be represented by ratios of two integers (whole numbers), like 4/5. They are used for calculation, but not counting. They have their own arithmetic, which can be quite hard to follow. In this book, fractions will appear very infrequently, as we will work with decimals mainly.

Irrational numbers are those which *cannot* be represented by the ratio of two integers. This sounds weird to most people (hence the name I guess). But you almost certainly know one – pi or π, which is the ratio of the circumference of a circle to its diameter. As you also probably know, if you were to calculate it as a decimal, it would never end. The decimal expansion of an irrational number neither repeats nor terminates. Another example is the square root of two.

Complex numbers are produced when a real number is added to what is termed an imaginary unit, which is basically the square root of -1. Like most of the above, they do not really concern us here other than for interest's sake, but they are almost essential to most higher-level mathematics applications. For example, in solving some equations, even though the answer may be a real number, complex numbers are necessary or useful during the solving.

think it over 1.12

Are complex numbers called complex numbers because they are complicated? Why are they called complex numbers? Can a rational number also be a complex number?

Equations

Equations cause much pain and suffering among the less quantitatively oriented students. Which is somewhat unfortunate since they are the fundamental core of mathematics (and thus business statistics and the like). You will see equations throughout this book on almost every page. Some basic concepts will help you overcome any fear of equations and have you playing with them like a pro. Well, almost.

Equations contain variables, which in this case mean something slightly different from the data set example given earlier. Variables in equations are 'quantities which are unknown' and usually represented by *letters*. Representing unknown numbers with letters is sometimes the point where your mind warps to the point of no return and mathematics becomes 'stupid'. But just think of the letters as numbers which we do not know (yet). If the purpose of the equation is to determine what number that letter represents, we call it an unknown, otherwise it is just a variable. A constant in an equation is simply a number we do know, represented by – you guessed it – an actual number. Sometimes constants are called parameters. Equations are basically essential for anything more than extremely basic maths, and are actually very useful in everyday life as well, which makes it a shame that so many students switch off as soon as a letter appears in an equation. They are essential, at least in part because they allow us to solve mathematical problems far easier than otherwise, and given that a lot of our lives is concerned with solving mathematical problems, understanding them is useful. This is of course quite apart from how integral they are to this book's material. Back to Basics Box 1.1 is a simple explanation of one of the most basic equations possible, in a way which might clarify things for some.

think it over 1.13

The term modelling is used quite a bit in statistics and mathematics. What do we mean when we say we have created a statistical model?

back to basics

Box 1.1 Equations

In contrast with the Above and Beyond boxes scattered throughout, Back to Basics boxes aim to present absolutely key foundational concepts in an alternative manner which might be easier for some to understand. Given that the 'omigod, letters are actually *numbers*!' stuff has been the downfall of many a maths student, why not start here?[6] This first one will concern pints of beer, which I am quite partial to.

Consider that you have £37 in your student account left after the first term. You want to go out on the town for the last Friday night of term, and during this night you want to pluck up the courage to confess your love to the girl/boy/whatever you have been infatuated with all term. You know from previous experiences that it will take the consumption of a minimum of 5 pints of beer in order for you to get up the courage to do this. Yet you also know that on Saturday morning you have to buy a train ticket home for the holidays, which will cost £9. There are a variety of bars around town where you could purchase your pints, and where the beer is of varying prices. *You need to know how expensive the pints can be* to enable you to purchase enough to give you the courage for your confession, as well as leave £9 in your pocket for the trip home to Mum.

This is an extremely simple equation, which you can set up as follows:

$$5x + 9 = 33$$

where x is the maximum average price of a pint which would enable you to buy five. Our objective is therefore to discover what x should be, so in the end we would like an equation to look something like:

$$x = ??$$

where ?? is the answer. Right?

There are various ways of solving this. Because it is so simple, we could do it by trial and error, but that is not the point here (and they are not always so simple!).

[6] That said, if you think expressing numbers with words is nonsensical, you are in good company. One of the greatest philosophers and mathematicians ever, René Descartes, originally thought the same. And he went on to make huge contributions to mathematics in the end.

So, let's solve it by using basic algebraic manipulation. Think of the equation like a scale with the equals sign as the pivot. The rule is: you *must keep it balanced*. In other words, everything you do to one side of the 'pivot' (i.e. the equals sign) must be done to the other. So let's work this out…

First, subtract 9 from both sides: $5x = 24$

Then, divide by 5 on both sides: $x = 4.8$

So, the maximum average pint price is £4.80 to still leave funds for the train trip. At the time of writing this book, that would certainly be fine, and in fact would probably leave you with funds left over to buy the object of your desire a drink. Bonus!

So, never let it be said that mathematics is not romantic.

An equation such as that in Back to Basics Box 1.1 is termed a linear equation. It contains only a variable to the power of 1, and as you will see later in the book, it can be plotted as a straight line on a graph. A quadratic equation contains a variable to the power of 2 (i.e. x^2). Because of the properties of multiplication, they always have two possible answers, or 'roots', although they can be the same. For example, $x^2 = 25$ could have the answer 5 or –5. A cubic equation unsurprisingly has a single variable to the power 3. Cubic equations always have three roots, and two or all of them may be equal. A quartic equation is one with a variable to the power of 4, and you can imagine by now no doubt that there are four roots here. You can also term these equations first-, second-, third- and fourth-degree equations respectively, and there is basically no limit to the degrees as we go on.

You will no doubt have noticed that so far we have been dealing with equations with one variable. However, equations can have more variables than this, and in fact there is no limit to how many variables there can be in an equation. Nevertheless, in most cases, if there is more than one variable in a single equation, this equation by itself is unsolvable, or *insoluble*. The only way to solve such equations is if we have additional equations which contain the same variables. If we have the same number of equations as we do variables, it is usually possible to solve the equations and discover the values of the variables. In such cases, we say we are dealing with a system of simultaneous equations, like the following:

$$5x + xy + 7 = 0$$

$$x + 2xy = 0$$

The first thing to remember is that these systems are internally consistent. In other words, x and y are the same in both equations. So, we can solve this one quite easily, even though it might look quite complicated to the uninitiated at first. But remember the rules from Back to Basics Box 1.1: everything you do on one side of a single equation has to be done to the other. However, you do not have to do the same thing to both equations at the same time.

To begin then, multiply the first equation by 2 to get:

$$10x + 2xy + 14 = 0$$

Do you see what has happened here? If we subtract the second equation from the new first equation, we get:

$$9x + 14 = 0$$

This leaves us with a very simple piece of manipulation to do, like we did in the box above:

$$9x = -14$$

$$x = -14/9 = -1.56$$

The next thing to do is *substitute* that value for x into the first equation, replacing x with -1.56, to give $y = -0.51$. Note, though, that these values have been rounded to two decimal places, so they are not absolutely exact.

That said, most simultaneous equations cannot be solved as easily as this, and require complex computer models to get approximate solutions (usually by trying over and over again with different values, to try to get as close as possible, termed an iterative method).

think it over 1.14

If you plotted a graph of the two equations given above, could you predict where they would cross? What is the significance of this value?

Sigma Notation

One thing that always confused me (for what you would think was a surprisingly long time) was what I later found out could be termed sigma notation, or summation notation. This appears all over statistics and involves a rather gratuitous use of the Greek letter $\sum$ (sigma). $\sum$ is used in many statistical contexts to indicate 'the sum of'. So $\sum x$ means the *sum of all x values*, and $\sum x^2$ means the *sum of all the squared x values*. In the latter case, you would square each x first, and only then sum the squared values (remember, BEDMAS). You are normally in the situation where you have a set of x values (maybe some data), and the sigma notation shows you that you have to add them together in some specific way. Here is a simple example.

If x takes the values of 2, 5, 10, 256, then:

$$\sum x = 2 + 5 + 10 + 256 = 273$$

$$\sum x^2 = 2^2 + 5^2 + 10^2 + 256^2 = 65665$$

This is simple enough in this instance, but it is in fact a rather informal version of sigma notation. I always got more confused when things got more general. Consider the more formal notation in Figure 1.5.

Figure 1.5 Formal Sigma Notation

$$\sum_{i=m}^{n} x_i$$

The index variable appears at the bottom of the $\sum$ (termed the *subscript*). The index variable i is called the index of summation, and m is the lower bound of summation. In this case $i = m$ means that the index i begins equal to m. The n represents the upper bound of summation. In this case, we find successive values of i by adding one to the previous value of i, starting out equal to m and finishing up when $i = n$. If it is an infinite series, the infinity symbol ∞ is used as n. The x_i part of the formula is changeable, and it depends on what you are summing. As you know, the i part of the formula is the index of summation and varies between m and n, but the x part can be replaced by pretty much any relevant expression.

Functions

Functions are kind of similar to equations in concept, and you should also be aware that the word 'function' means different things to different people. So do not get it confused with a subroutine in computer science, or, more importantly, a party. Turning up to mathematics classes dressed in your party clothes might be seen as inappropriate. Anyway, a function is a way of expressing some dependence between two variables or quantities. The concept of functions is vital to most of the more advanced mathematical and statistical operations in this book, so let's get it straight now.

As I just said, a function expresses some dependence between two variables. One of these is called the independent variable, sometimes called the argument or the input. You might also hear it called the predictor, or exogenous variable in some contexts. The independent variable is usually given. The other variable is called the dependent variable, or sometimes the value or output. It may also be called the criterion or endogenous variable in some situations. A function basically tells you what output you get for a specific input.

According to convention, we represent variables by letters at the end of the alphabet, such as x, y and z. Functions can also contain constants, which are of a constant value (hence the name), and where necessary these are represented by letters at the start of the alphabet, such as a, b and c.

As a very simple example of a function consider the following:

$$f(x) = x^2$$

This is the standard notation to describe a function; it can be read as 'the function of x is x squared'. So, in other words, for every input we put into the function, we get an output of the input number squared. Thus the input of 2 results in the output of 4, which can be written as $f(2) = 4$.

A different example you might recognize occurs in cooking. I like cooking – it is very relaxing after a long day writing about statistics. In New Zealand, I learnt to cook using grams and kilograms as units of weight, but in the UK (when the European Union is not looking) we often use ounces and pounds. So, of course, I need to convert these units. If I have x grams, I can use a formula to express easily the conversion from grams to ounces, for example as:

$$f(x) = 0.053x$$

This case shows that 1 gram equals 0.053 ounces. I think you can pretty easily work out other values, and this is of course an extremely simple formula.

We use the f notation above when we need to give the function a name for some purpose. This is quite common in mathematics. For example, we might need to use a long function repeatedly, and it would be handy just to use $f(x)$ for simplicity. So I could equally define $f(x) = x^3 + 5x + 3$. Then $f(2) = 21$. However, sometimes we do not need to name a function (which is much more common in more applied situations like those in this book). In such cases we sometimes drop the f business and use the form:

$$y = x^2$$

This defines the dependent variable as y and the independent variable as x. It might help you to work out what is going on here. The confusion may arise when you wonder whether you have an equation or a function. If you have an equation like $y = x^2$ to deal with, then you already know that this is insoluble without any other information. However, if it is a function it is not meant to be solvable by analysis, instead it simply *describes what happens when you input a value of x*. Or in other words, the value of y which results from a given value of x.

As you might have already guessed, functions can be shown in many different ways. So far I have given you functions in the form of formulae, but sometimes it is more useful to show them as a plot or graph. Figure 1.6 shows $f(x) = x^2$ as a plot, and you will probably recognize it. We plot the inputs on the x axis and the corresponding outputs on the y axis. Plots and formulae are the most useful representations for our purposes, but functions can instead have their properties described in words (as was done before), or be represented by algorithms, or even by their relationships to other functions (e.g. the inverse of some other function).

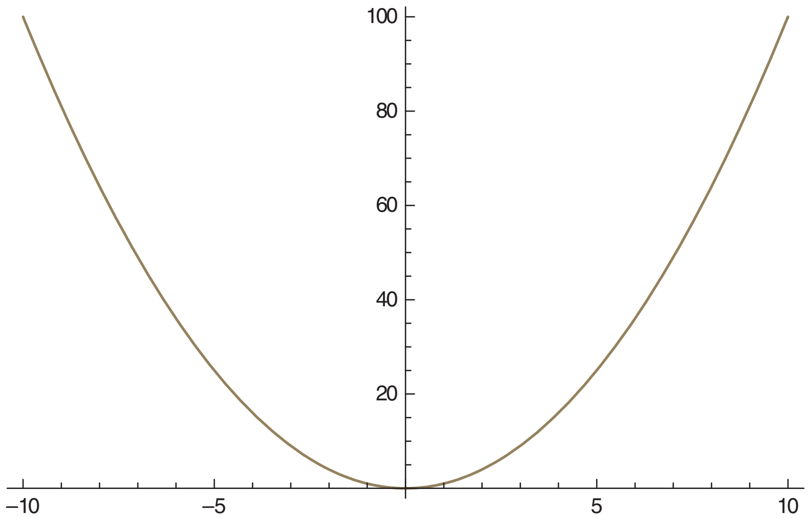

Figure 1.6
Plot of $f(x) = x^2$

think it over 1.15

Look at Figure 1.6. If you drew a vertical line up from –5 and another one up from +5 until they hit the curve, and then joined the two end points to form a rectangle, where would the horizontal line cross the vertical axis? What is significant, mathematically speaking, about this value?

Extremely common or useful functions are often given their own permanent name. You have almost certainly come across these in your prior mathematical career. For example, the function $f(x) = x^2$ is generally written as Square(x). Although, to be honest, this is not the greatest example since the permanent name is actually longer! Nevertheless, you will probably have come across the trigonometric functions sine, cosine and tangent, which are represented as sin(x), cos(x) and tan(x). In other words, you input x into the relevant function and get an output y. These are good examples of how function notation can be very useful to us.

think it over 1.16

Why do we call one variable independent and the other dependent? Is there a mathematical relationship between them?

SUMMARY

At this point I suspect you either have all kinds of confusing things going around in your head, or otherwise think I just totally wasted an hour of your life because you knew it all already. Whichever of those you find yourself closer to – or if you are instead somewhere in the middle – you need to keep in mind the following things which will really help you progress through the sometimes treacherous forests of quantitative analysis:

- Remember, a surprisingly large number of other people are scared of numbers, so if you are too, you are not alone. If you are not, then perhaps you have a fruitful avenue to get to know others?
- Quantitative analysis is vital to almost all business careers, almost certainly including the one you hope to pursue, or the one you will eventually find yourself in.
- The key to beginning to understand the relevance of quantitative analysis to your career is to remember that you need *data* to make effective decisions. Data is information, and you use it to answer important questions.
- The numbers that make up quantitative data are representations of things that are in the 'real world'. Sometimes numbers can behave differently from these things, and we need to understand the relationship between 'raw numbers' and the things they are supposed to represent, if we are going to do good analysis.

Finally, it is absolutely vital that you take seriously my words that learning how to do quantitative analysis is fundamentally a *step-by-step* process. If you do not master the early steps, the later steps will be very difficult, and in fact will end up being impossible and frustrating for you. If you did not do well at school, please take seriously the idea of doing some recapping in more depth of the basic concepts which are covered here – it will really help you out in the future. However, on the other hand this also means that maths and stats are not an impossible mystery that only a few 'clever' people can do. They are things which anyone who is prepared to sit down and work at can end up being pretty decent at – and most definitely decent enough to be very successful in business. So good luck. I am looking forward to accompanying you on this journey.

final checklist

You should have covered and understood the following basic content in this chapter. If not, go back and retry the relevant sections:

- ☑ Reasons why people might be scared or turned off by studying quantitative methods.
- ☑ How quantitative analysis for business is primarily to help you *make decisions* in your future career, even in what you think might be the most exciting and creative professions.
- ☑ What is data, and what is an element (or data point), a variable and an observation, and how these together make a data set.
- ☑ The difference between qualitative and quantitative variables, and discrete and continuous data.
- ☑ How we often approximate the amount of qualities (like happiness) by numbers, but that these numbers are not the exact same thing.
- ☑ The names and properties of the four different scales of measurement – nominal, ordinal, interval and ratio.
- ☑ The difference between cross-sectional and longitudinal data.
- ☑ What a sample is, what a population is, and the relation between the two.
- ☑ Why variation is important in quantitative analysis.
- ☑ What BEDMAS, exponential notation, powers/exponents and logarithms are.
- ☑ What equations and functions are, and how they can often be expressed in sigma notation.

EXERCISES

This section will contain useful exercises for you to do, to help you understand the content. However, they probably will not be like the ones you have seen in other similar books. Here, I am not trying to test your knowledge, but to help you further understand the content of the chapters. Your lecturers and teachers can test your knowledge, I am here to help you develop it.

With that in mind, I offer the following observation regarding the exercises in this chapter: from my own experience, it was not until I had to teach a subject that I started to understand it. In fact, I can remember fervently praying to the lecturing gods that students would not ask me about certain things (probability theory mainly!). Thus, when I want to learn something, I often put myself in the situation where I will have to teach it to someone else, and it helps a lot. The following questions are based on this theory. You can make them even *more* helpful by finding a friend who struggles with maths, and trying it for real.

Imagine the following scenario (this may be hard for some of you).

You have successfully obtained your qualification – top of the class. Some friends of yours have asked if you would help their school-age daughter with some maths. In particular, she is struggling with some of the notation.

1. First, your task is to explain to her what algebra is all about. She just does not get it – maths is about numbers, not letters. She is really stuck on the idea of variables. She wants to know if there is a difference between the variables she uses in maths, like x, y, z, and the ones she has seen in statistics, like firm, gender, etc. Can you explain to her the difference (if there is one)?
2. Next, she tells you she is confused about data – qualitative data, quantitative data. Does it mean some data is 'quality' and other data – well, that there is just lots of it? Give three examples of each type of data and explain why they are different.
3. Importantly, she tells you that she is doing a project based on the different attitudes of students and lecturers towards exams. What sort of scale do you think she should use to classify the variable, 'student or lecturer'?
4. For her project, to measure attitude she used a scale where 0 represented extreme hatred and 5 represented exquisite love of exams. In her sample five students circled 0 and three lecturers circled 5. Does this mean that hatred for exams scored 0 ($0 \times 5 = 0$) and exquisite love scored 15? You will need to explain your answer in simple terms here.
5. She has also been shown the symbol $\sum$, but it is all Greek to her. Can you give her an easy way to remember what the symbol means?
6. Her teacher has been banging on about functions (not bodily ones) as well. This is causing lots of problems, and she asks, with desperation in her voice, 'What does $f(x)$ tell me and what does it mean?'
7. As you walk home, you think back to your own performance in your quantitative courses. You remember that one time during your degree, you did a class test for quantitative analysis. You scored 80 in the test and your friend scored 40. Does that mean you are twice as clever as your friend? How would you classify this scale?

To access additional online resources please visit: **https://study.sagepub.com/leepeters**

The Ballad of Eddie the Easily Distracted: Part 2

Eddie felt a little better about life as he wandered up to the back of the lecture theatre to take his seat in QUAN101 for the second lecture. He actually felt that he had made a little progress at this numbers thing over the week, and some of the bits that he'd struggled with made some more sense – like formulas! Or formulae as his smug roommate had told him it should be said, last night at the Union. That had been a bit embarrassing, especially in front of those girls.

Despite his good mood, Eddie was quickly getting a little bored. Dr Jones seemed to be going on about *collecting* data, rather than analysing it. Eddie felt that all his hard-earned knowledge was ebbing away from him a little. Worse, this collection business seemed to require quite a lot of attention to detail and planning. Not his particular strength he thought glumly. Eddie looked at his lecture notes. Dr Jones was trying to explain the principle of 'garbage in, garbage out', which kind of made sense to him. After all, if you didn't know what it was you were analysing, how were you supposed to trust the results?

Pleased with himself for that insight, Eddie leaned back in his seat a little with a little smile. At this point, he knocked over his bottle of water, which fell to the floor with a clatter as Dr Jones finished her sentence. The entire class turned around and looked at Eddie, who was scrabbling around trying to stop the bottle from emptying all over the floor. As he felt his cheeks warm, Eddie heard them start to laugh. He managed to pick the bottle up – at least there was some water left – and sit down. As the class all turned back to look at Dr Jones (who Eddie felt was enjoying this a little too much) he noticed what's-her-name – Esha? – speaking to her friends, who all laughed again.

'That went well,' thought Eddie. 'Pride comes before a fall.'

Esha's Story: Part 2

Esha had had a pretty boring first week of classes. All the basic stuff she already knew. She'd moaned to her Mum about the whole thing, and Mum had just told her to have some sympathy – not everyone was as clever as her. While Esha had secretly enjoyed that a little, it didn't make her classes any more interesting. And when that clown at the back of QUAN101 knocked over his water bottle, she'd enjoyed it immensely, telling her friends that he'd probably done it as he fell asleep.

Frankly, that had been the highlight of the lecture itself. She just didn't see why it was necessary to think about collecting data before analysing it. After all, she was hardly going to be working on the high street getting people to fill in questionnaires for a living, was she now?

Still, she did find some appeal in the details of the task. All that careful planning and preparation for the task was something she did enjoy, and she could certainly see herself perhaps managing the process in the future. Maybe there was something to this after all? After she'd gone home, still giggling about the water bottle incident, she sat in her room with her friends having a coffee and talking about it. They'd set up a little study group to go over the week's material, and Esha enjoyed helping some of her friends understand things better.

'I guess,' Esha said, 'it kind of helps if you know where the data comes from – even if you didn't collect it yourself. I mean, it's what Dr Jones said: garbage in, garbage out.' Esha's friends nodded sagely. 'But also, I suppose it's quite likely that we might have to be in charge of getting data to answer questions in our jobs, even if we don't collect it, and also I think that understanding the spreadsheets is going to help us a lot.' Esha had a final insight: 'Hey, I bet we can use all of that project management stuff at the end to help with our group assignments.' She was rather pleased with herself at that, and enjoyed her coffee all the more.