

# STATISTICS WITHOUT MATHEMATICS





Los Angeles | London | New Delhi  
Singapore | Washington DC

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

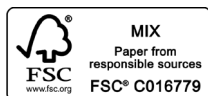
SAGE Publications Inc.  
2455 Teller Road  
Thousand Oaks, California 91320

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: Jai Seaman  
Assistant editor: James Piper  
Production editor: Victoria Nicholas  
Copyeditor: Richard Leigh  
Proofreader: Andy Baxter  
Indexer: David Rudeforth  
Marketing manager: Sally Ransom  
Cover design: Shaun Mercier  
Typeset by: C&M Digitals (P) Ltd, Chennai, India  
Printed in India at Replika Press Pvt Ltd



© David J Bartholomew 2016

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

**Library of Congress Control Number: 2015931451**

**British Library Cataloguing in Publication data**

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

ISBN 978-1-4739-0244-2

ISBN 978-1-4739-0245-9 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the Egmont grading system. We undertake an annual audit to monitor our sustainability.

# O



## BEFORE WE BEGIN

### Summary

This book aims to present the basic ideas of Statistics without using the language of mathematics, which is the normal medium for that purpose. The chapter therefore begins by explaining the relationship between Statistics and mathematics and goes on to introduce the key notions of number, variation and measurement. It concludes with a brief outline of the origins of the subject.

This chapter is not an essential part of the main book but is intended to prepare the reader for what follows.

### Role of Mathematics in Statistics

There is a widespread belief that Statistics is necessarily mathematical and, therefore, difficult. This book aims to show that the key ideas of Statistics can be expressed without any use of mathematical symbolism at all. This is not the same as saying that one can be a statistician without a good command of mathematics; merely that one can understand what the subject is all about without a knowledge of mathematics. It must be realised that mathematical language, in which introductions are usually expressed, is both precise and concise – indeed, these are among its principal advantages. When both are dispensed with there is, necessarily, the risk of imprecision and a certain vagueness. This is something we shall have to learn to live with, but its ill effects are minimised if we concentrate on ideas – as we shall do.

The book is aimed particularly at students of social science who are required to take a Statistics course as part of their degree course and who

find the mathematical aspect completely off-putting. Even if such courses are disguised under beguiling titles such as ‘quantitative methods’ or ‘research methods’, closer examination will quickly reveal that there is some statistical content which will not lose its capacity to instil fear in spite of all assurances to the contrary.

This book is certainly not intended as a substitute for those standard texts which the student will meet – quite the opposite. Once the basic ideas have been grasped, the student will be better prepared to move on to the more serious treatments. I recall, many years ago, carrying out an experiment when I first encountered mathematical resistance among students. I sat down with one student and went very slowly through the ideas until she had grasped what it was all about. This took a great deal of time and it would not have been practicable to repeat the experiment with others, or with this student a second time. I still remember her remark at the end of it all: ‘Is that all there is to it?’ It was indeed all there was to it, and the same could be said of most other parts of the course. The basic reason for her difficulty was that she did not appreciate that most of the ideas were already very familiar from everyday experience and could be classed as what we often call common sense.

Something which is common to all expositions of Statistics, whether mathematical or not, is the sequential character of the subject. In a subject like geography, where one might be reading a book on industrial towns, the order of chapters might not matter too much. One could dip in here and there according to one’s interest and still obtain some benefit. But with introductory Statistics it is different. What comes later often depends in a crucial fashion on what has gone before. Missing a chapter or lecture effectively derails the whole enterprise. The particular advantage of a book over a course of lectures is that one can retrace one’s steps and make sure that each step has been properly digested before moving on.

Arithmetic occupies, perhaps, a half-way position between the mathematical and the non-mathematical as regards its capacity to deter students. Arithmetic deals with numbers and the particular rather than the general. As far as possible, arithmetic will also be avoided, but there are occasions where the benefits seem to outweigh the disadvantages. This will occur particularly when we come to think about summary measures in Chapter 4.

One frequently meets the term *mathematical Statistics*. Strictly speaking, this is a branch of mathematics centred on the proof of theorems starting from well-specified axioms. It bears much the same relationship

to practical Statistics as does geometry to surveying and architecture. Geometry belongs to an idealised world, but one which is sufficiently close to the real world for many of its results to be transferable. The fit is perhaps less close in the case of Statistics, and this means that the results of mathematical Statistics need to be applied with particular care. Applied statisticians should not be intimidated when it is pointed out that their methods are ‘inadmissible’ or ‘biased’ because these terms relate to an idealised world which may not exactly match the real world in which they are operating. One cannot, of course, suppose that the extraction of the core ideas of Statistics in a book such as this can be done without making a subtle change in the subject itself, but this is a topic for the philosophically minded which lies outside our remit.

## Numbers

Statistics is about extracting meaning from numbers – especially from collections of numbers. It is therefore crucial to know something about the numbers from which everything starts and on which it depends. This aspect may be lost sight of in the body of the book, so it is worth making the point clear before we begin. The numbers with which we shall be mainly concerned here are measures of some quantity such as length, money or time. That is, they are quantities – of something. But numbers can also mean different things according to their context. The number 3, for example, could mean a variety of things according to the way it is used. It could be the rank order of a pupil in a class examination. It could be the weight in kilograms of a bag of tomatoes, and in this case it might be the actual weight or simply the weight to the nearest kilogram. It might be the time, as for example in ‘3 o’clock’, which might be when a meeting starts. Even if it refers to a weight in kilograms, the same amount could be expressed, equivalently, in pounds or ounces. Usually all of these matters have been sorted out before statistical analysis begins, and so they are easily taken for granted. We assume that this is true for this book also. But it must never be forgotten that what we are about to do only makes sense if the collections of numbers dealt with are of the same kind. We must not mix different units of length or, more generally, mix numbers referring to different kinds of things. An important thing is to distinguish numbers which are measurements of some kind from numbers which are simply being used as codes. A good example of the latter is provided by the numbers which footballers have on the back of their shirts. These numbers may well convey a meaning

which is perfectly understood by *aficionados* of the game, but they do not measure any quantitative characteristic of the players. Similarly the scoring system in tennis may seem idiosyncratic (love, 15, 30, 40, game), until it is remembered that the numbers and names are merely labels and not quantities on which we may do arithmetic.

## Variation

Since we are claiming that the key idea of Statistics is *variation*, we need to note what it means to say that different kinds of number are varying. Virtually everyone has a nationality and people differ according to their nationality. We may therefore legitimately say that nationality is something which varies from one individual to another. People also differ in age and this enables us to say that age varies from one individual to another. But the variation in age differs from the variation in nationality. In the case of age, differences can be specified in quantitative terms – they measure something. In the case of nationality, the differences cannot be so specified numerically. Hence, although it makes sense to add up ages and calculate an average age, there are no numbers we can sensibly add up to form an ‘average’ nationality. It is true that by showing an excessive zeal for quantitative analysis some have assigned numbers to nationality in some way, but the average nationality calculated from such numbers would have no real meaning. Rather than detain the reader with formal definitions of levels of measurement, we shall proceed with this example to serve as a warning that it is important to be clear about what the numbers mean before we start on any analysis. Broadly speaking, we started by thinking of numbers as ‘amounts of something’, which makes it meaningful to speak of the variation between individuals in quantitative terms. Notice that ordinal numbers, which are equivalent to ranks, do not fall into this category because knowing that two individuals are ranked 7th and 13th, say, tells us nothing about how far apart they are. Is the difference between them, for example, greater or less than that between the individuals ranked 8th and 14th?

## Measurement

Before going any farther we shall pause, to briefly notice some of the different sorts of number that arise in statistical work. The simplest kind of measurement is a *count*. Many measurements that we encounter are counts. The number of people in a household, the population of a

town and the number of eggs in a bird's nest are all counts and they have a direct and obvious meaning intelligible to everyone. Counts must be whole numbers and they cannot be negative. Because such numbers measure a quantity and vary among themselves, we are justified in including them among what we collectively call *variables*.

Next we come to the sort of measurements we make with instruments, such as a ruler or a pressure gauge. These operations also yield numbers but, unlike counts, they need not be whole numbers and their accuracy is limited only by the precision of the measuring instrument or our ability to use it. Essentially we are comparing the thing to be measured with some standard represented by the length of intervals on a ruler, for example. We normally record what we have measured to the accuracy determined by the quality of our instrument or our eyesight. More sophisticated measuring systems do not rely directly on human perception, but the principle is the same.

A fundamental feature of all such measures is that the scale of measurement is often *arbitrary*. The choice between yards, feet and inches and metres and centimetres as measures of length is a very important practical matter but it is usually of no theoretical significance. Length, time, weight and pressure are all physical quantities, but they have no *natural* scale of measurement. To them we may add money. Monetary values may be measured in dollars, rupees or pounds and although the pound, the rupee and the dollar may have great national significance, that position does not arise from anything inherent in that choice of unit. The arbitrariness of such scales is very important because the degree of variability exhibited by any quantity depends on the scale of measurement. The variation in the time taken to run a kilometre by a group of athletes yields a much larger number when expressed in milliseconds rather than in hours. If we do not want our conclusions to be equally arbitrary, the arbitrariness in the scale of such measurements must be made irrelevant for the purpose in hand. Our conclusion may be formally stated as follows. *We should not do anything with the numbers or draw any conclusions from them which depend on an irrelevant unit of measurement.* Thus the end-product of any calculations we make with money should not depend on whether we work in pounds, dollars or guilders or anything else.

The arbitrariness of the unit of measurement can be very limiting, but worse is to come! We may measure air temperature in degrees Fahrenheit or Celsius, but when we do this it is not only the scale of the unit of measurement which differs. The two scales also have different *origins*. The origin is the temperature to which we assign the value zero.

This is no problem with length or weight, for example, because we all know what is meant by a zero length or zero weight. Temperature is not quite like that. There is in fact a natural zero point which occurs when all molecules are at rest, but this is so far removed from the scale we need in meteorology, for example, that it is almost irrelevant. Absolute zero occurs when all the molecules in a piece of matter are at rest because, to a physicist, temperature is a measure of the activity of the molecules and this has a natural origin. On the Celsius scale it is at  $-273$  degrees. Although some parts of the universe are almost as cold as that, it would be very inconvenient to measure the air temperature on our planet using such an origin because all the resulting numbers would be large and close together. Instead we choose an arbitrary origin to suit our own convenience. If we extend the principle enunciated above to cope with the new situation it becomes: *We should check whether anything we do with temperatures depends on either the unit of measurement or on the origin from which the measurements are made.*

This brings us back to ordinal measurements. As we have seen, these are rather different in that they do not attempt to place individuals on a scale in the same sense as the quantities we have just discussed. The number assigned to an individual merely tells us where that individual comes in the rank order of the set of individuals with which it is to be compared. This precludes doing any arithmetical operation on them which goes beyond their ordinal character. The natural way to assign numbers is to use the numbers 1, 2, 3, ... so that the 7th largest member has the number 7 and so on. But any other set satisfying the basic order restriction would serve as well. We could take only the odd numbers or, to take a rather bizarre example, we could use only the prime numbers, provided only that the order condition was met. Our general principle when applied to ordinal numbers, therefore, says that *we should do nothing with the ordinal numbers which makes the result depend on what particular set of labels we happen to have chosen to use.*

It should be clear, by now, that the numbers with which we have to deal are not all the same. As we have already noted, books on Statistics for social scientists sometimes speak about ‘levels of measurement’, and this involves a more systematic treatment of the same idea. The guiding principle underlying this discussion is that our analysis should not require more of the numbers than their level of measurement justifies. In practice it says that all our data should consist of measures of size of the same kind expressed in the same units.

There is much to be said for the idea that the early stages in education should concentrate on instilling information – even rote learning – on the grounds that illumination will follow, and cannot be had until a sufficient



foundation has been absorbed. But it does not always seem to work like that with Statistics where, among the readership we have in mind, the incentive to learn the language first is entirely lacking. It therefore seems worth beginning at the other end by pointing out that much of the learning has already been done and all that one has to do additionally is to uncover it and expose its relevance.

Those already well versed in the subject may be having doubts about our assertion that some of the ideas of Statistics are already familiar, and they will claim that others are unfamiliar and counterintuitive. This may sometimes be true, especially if one tries to swallow too much at one bite; we also have to recognise that common sense is not so very common. These are problems we shall have to meet and try to solve as we go. As we have already noted, there are no exercises in this book of the usual kind, but rather we offer an invitation to look around the world and see examples of Statistics everywhere. For this reason our examples are drawn from a very wide field of common experience and not restricted to the social sciences alone.

## Origins of Statistics

Although the name ‘Statistics’ has a long history, the name is, in effect, an Anglicisation of the German word for ‘state science’. Modern Statistics can be more accurately dated from the work of Karl Pearson at University College London around 1900. Variation was in fact the key element in early Statistics, especially in the tradition stemming from Karl Pearson’s predecessors, Charles Darwin and Francis Galton. Stigler, for example, makes the following remark about Darwin’s *Origin of Species*:<sup>1</sup>

Accordingly, chapters 1, 2 and 5 were exclusively concerned with variation, starting with variation in domestic plants and animals. Darwin developed a wealth of information on dogs, pigeons, fruit and flowers. By starting with domestic populations he could exploit his reader’s knowledge of widespread experience in selective breeding and agriculture to improve the breed and the crop. The substantial variation was convincingly argued and, what’s more, the variations that he presented were demonstrably heritable.

---

<sup>1</sup>Stigler, S.M. (2010) Darwin, Galton and the Statistical Enlightenment. *Journal of the Royal Statistical Society, Series A*, 173, 469–82.

The early volumes of the journal *Biometrika*, founded and edited by Pearson, are full of frequency distributions of biological and anthropological variables. Since then there has often been a tendency to overlook or even ignore distributional matters by focusing on particular aspects of distributions such as the average. As the subject has grown, its spread has increased enormously and now hardly any branch of science – or social science – is beyond its reach.

It may seem slightly odd that in a book directed primarily to social science students, we should trace the origins of the subject to biology. Nevertheless, that is where the origins actually lie. The founding father of social Statistics in the UK in a social science context was Sir Arthur Bowley, the first professor of Statistics at the London School of Economics and Political Science, and possibly the first professor of Statistics in the world. In 1910 he published his *Elements of Statistics* which ran through many editions. Although it is devoid of any discussion of frequency distributions until Part II, and even then only briefly, Bowley makes profuse acknowledgement in the Preface of his debt to his academic neighbour Karl Pearson and his associates at University College London.

## In Conclusion

It was Pierre Simon de Laplace, speaking of probability theory, who said that at bottom it was common sense reduced to calculation. We are claiming that something similar may be said of Statistics, though it might be more accurate, as far as this book is concerned, to reverse the statement and say that it is calculation reduced to common sense!

