NIGEL WALFORD

PRACTICAL STATISTICS FOR GEOGRAPHERS AND EARTH SCIENTISTS

SECOND EDITION

with website

WILEY

Practical Statistics for Geographers and Earth Scientists

Practical Statistics for Geographers and Earth Scientists

Second Edition

Nigel Walford

WILEY

This edition first published 2025

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, New Era House, 8 Oldlands Way, Bognor Regis, West Sussex, PO22 8NQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data Applied for:
Paperback: 9781119526971

Cover Design: Wiley
Cover Image: © vectorfusionart/Adobe Stock

Set in 9.5/12.5pt STIXTwoText by Straive, Pondicherry, India

To Ann

Contents

Preface to the Second Edition

Some 12 years on from publication of the first edition of *Practical Statistics for Geographers and Earth Scientists* provides the opportunity to reflect on what has changed and what has remained the same. The majority of undergraduate and postgraduate students on degree programmes in Geography, Earth and Environmental Sciences are still expected to carry out a piece of independent research in a similar way to architecture students design a new structure or engineering students fashion the prototype of a new product. The underlying aim of such projects persists, namely to enable students to produce new knowledge and understanding, but is now accompanied by an increased emphasis on how successfully completing a project develops practical skills and experience that are pertinent in the 21st century workplace.

More than 60 years ago the quantitative revolution in Geography and cognate disciplines established the basis of quantitative research that was subsequently re-enforced by the arrival of Geographical Information Systems and the 'spatialising' of statistical analysis, sometimes by researchers in other disciplines. Reaction against a perceived dominance by the quantitative paradigm contributed to the 'cultural turn' in human geography and the ascent of qualitative approaches. At the risk of inviting challenge and dispute, it could be argued that protagonists on both sides of the quantitative versus qualitative debate have now seen an opportunity for pragmatic compromise in the form of mixed methods research. Although this text is firmly grounded in statistical analysis of both spatial and non-spatial varieties, it also emphasises that a combination of approaches to addressing research questions will often produce a more holistic interpretation of the data and understanding of the topic.

Writing the second edition has provided the opportunity to update and revise the content of the first, although the fundamentals of statistical analysis have remained extant for decades, if not centuries in some cases. Three specific changes are worth highlighting. First, non-spatial and spatial statistical techniques are now separated into different sections of the book with their own chapters. This still allows readers to examine similarities and differences between the two groups of techniques, but also avoids confusion. Second, an entirely new concluding chapter focuses on four practical ways of helping students to carry out their independent research project and write their dissertation or report. The Chapter 12 sets out to:

- review the results obtained from previous chapters and identify unanswered questions arising from the analyses;
- examine how researchers have presented and discussed their findings in published journal articles;
- outline how to use information technology and software are used for quantitative analysis;
- describe three mini projects showing how to develop on previously published research.

The third change is that selected reading and references have now been added to the end of each chapter to make following up on points in other sources more direct and straightforward. The journal articles examined in Chapter 12 are all available on open access enabling readers to relate the discussion in the text to the articles themselves.

This book has been written in recognition that three aims or learning outcomes are common to many undergraduate and postgraduate degrees in Geography, Earth and Environmental Sciences:

1) development of a level of understanding and competence in statistical analysis.
2) completion of a carefully defined, independent research investigation involving qualitative and/or quantitative analysis.
3) interpretation of research findings and statistical analyses presented in journal articles and other published material.

Many features of the first edition have been retained to promote these aims including the use of self-assessment or reflection questions scattered through most of the chapters. Some of these could form the basis for discussion in class, others are intended for independent study. The *Statistical Analysis Planner and Checklist* between book sections II, *Exploring Geographical Data*, and III, *Testing Times*, has been kept to provide guidance on selecting techniques for different styles of project. The format of the chapters includes boxed sections where the spatial or non-spatial techniques are explained and applied to one of a series of datasets. These datasets are representative of teaching undergraduate and postgraduate students for over 35 years in practical computer laboratory workshop sessions or relate directly to group projects carried out during field work. The chapters in the book progress from consideration of issues related to formulating research questions, collecting data and summarising information, applying statistical techniques to test hypotheses, analysing associations and relationships between variables and investigating the inherently geospatial aspects of geographical data. Overall, the focus is on the practicalities of choosing statistical techniques for different styles of quantitative research, including those that could be incorporated into a mixed methods approach, but even the comprehensive introduction given here cannot hope to include everything.

April 2024

Nigel Walford
Wivenhoe

Acknowledgements

My initial point of contact at Wiley, Andrew Harrison, moved on to pastures new during the course of writing this book. I would like to thank him and his successors at Wiley for their patience and gentle reminders over the last few years and to others in the production and editorial offices for steering the manuscript through to publication.

Several of the original figures and diagrams, which have been included in this second edition, were produced by Claire Ivison formerly the cartographer in the Department of Geography, Geology and the Environment at Kingston University. I offer my sincere thanks to her.

I also acknowledge the support and assistance from colleagues at Kingston University and to the students whose questions prompted me to attempt this book as they attempted to understand the complexities of statistical analysis.

Finally, I extend a very sincere and heartfelt thank you to Ann Hockey, who has survived me writing a textbook four times over more than 30 years.

About the Companion Website

This book is accompanied by a companion website.

www.wiley.com/go/PracticalStatistics2e

This website contains:

- A Word file giving details of the variables in the dataset.
- An Excel file containing the columns of data.

Section I

First Principles

1

What's in a Number?

Chapter 1 provides a brief review of the development of quantitative analysis in Geography, Earth and Environmental Science and related disciplines. It also discusses the relative merits of using numerical data and how numbers can be used to represent qualitative characteristics. A brief introduction to mathematical notation and calculation is provided to a level that will help readers to understand subsequent chapters. Overall, this introductory chapter is intended to define terms and to provide a structure for the remainder of the book.

Learning Outcomes

This chapter will enable readers to:

- Outline the difference between quantitative and qualitative approaches and their relationship to statistical techniques;
- Describe the characteristics of numerical data and scales of measurement;
- Recognise forms of mathematical notation and calculation that underlie analytical procedures covered in subsequent chapters;
- Plan their reading of this text in relation to undertaking an independent research investigation in Geography and related disciplines.

1.1 Introduction to Quantitative Analysis

Quantitative analysis is one of the two main approaches to researching and understanding the world around us. In simple terms, it is the processing and interpretation of data about things, sometimes called phenomena, which are held in a numerical form. In other words, from the perspective of Geography and other Earth Sciences, it is about investigating the differences and similarities between people and places that can be expressed as numerical quantities rather than words. In contrast, qualitative analysis recognises the uniqueness of all phenomena and the important contribution towards understanding that is provided by unusual, idiosyncratic cases as much as by those conforming to some numerical pattern. Using the two approaches together has become popular in recent years to develop a more robust understanding of how processes work that lead to variations in the distribution of phenomena over the Earth's surface than by employing either methodology on its own.

Practical Statistics for Geographers and Earth Scientists, Second Edition. Nigel Walford.

Companion website: www.wiley.com/go/PracticalStatistics2e

If you are reading this book as a student on a university or college course, there will be differences and similarities between you and the other students taking the same course in terms of such things as your age, height, school level qualifications, home town, genetic make-up, parental annual income and so on. You will also be different because each human being, and for that matter each place on the Earth, is unique. There is no one else exactly like you, even if you have an identical twin, nor is there any place precisely the same as where you are reading this book. You are different from other people because your upbringing, cultural background and physical characteristics have moulded your own attitudes, values and feelings and how you respond to new situations and events. In some ways, it is the old argument of nature versus nurture, but we are unique combinations of both sets of factors. You may be reading this book on a laptop in your room in a university hall of residence, and there are many such places in the various countries of the world and those in the same institution often seem identical, but the one where you are now is unique. Just as the uniqueness of individuals does not prevent analysis of people as members of various groups, so the individuality of places does not inhibit investigation of their distinctive and shared characteristics.

Quantitative analysis concentrates on those influences that are important in producing differences and similarities between individual phenomena and to disregard those producing unusual outcomes. Confusion between quantitative and qualitative analysis may arise because sometimes numbers are used to represent the qualitative characteristics of people and places. For example, areas in a city may be assigned to a series of qualitative categories, such as downtown, suburbs, shopping mall, commercial centre and housing estate, and these may be given numerical codes, or a person may agree or disagree with the UK's exit from the European Union with these opinions assigned numerical values such as 1 or 2. Just as qualitative characteristics can be turned into numerical codes, numerical measurements can be converted into descriptive labels. This could be simple, such as grouping household income values into low, medium and high ranges. In contrast, a selection of different socio-economic and demographic numerical counts for various locations may have combined them in an 'analytical melting pot' to produce geodemographic descriptions or labels of what neighbourhoods are like, such as *Old people, detached houses; Larger families, prosperous suburbs; Older rented terraces;* and *Council flats, single elderly*.

The major focus of this book is on quantitative analysis in Geography and other Earth Sciences, although this emphasis does not imply that it is in some definitive sense 'better', or even more scientific, than qualitative analysis. Nor are they mutually exclusive since researchers from many disciplines now appreciate the advantages of combining both forms of analysis in a mixed methods approach. This book concentrates on quantitative analysis because many students, and perhaps researchers, find dealing with numbers and statistics difficult and are a barrier to understanding the 'real' Geography or Earth Science topics that interest them. Why should we bother with numbers and statistics, when what really interests us is why migrants are risking their lives crossing the English Channel to seek asylum in the UK, why we are experiencing a period of global temperature increase, or why people live in areas vulnerable to natural hazards?

We can justify working with numbers and statistics to answer such research questions in a few diverse ways. As with most research, when we try to explain things such as migration, global warming and exposure to natural hazards, the answers often seem like common sense and even rather obvious. Maybe this is a sign of 'good' research because it suggests the process of explaining such things is about building on what has become common knowledge and understanding. If this is true, then ongoing academic study both relies upon and questions the work of previous generations of researchers and puts across complex issues in ways that can be understood. Answers to research questions are commonly stated with a high degree of certainty, whereas they are often underlain by an analysis of numerical information that is anything but certain. The results of the research are likely to be correct, but they may be false. Using statistical techniques gives us a way of expressing this uncertainty and of hedging our bets against the possibility that our set of results has only arisen by chance. At some time in the future, another researcher might come along and contradict our findings.

But what do we really mean by the phrase 'the results of the research'. For the 'consumers' of research, whether the public at large or professional groups, the results or outcomes of research are often some pieces of crucial information. The role of such information is to either confirm facts that are already known or believed or to fulfil the unquenchable need for new information. For the academic, these detailed factual results may be of less direct interest than the implications of the research findings, about some overarching theory. The student undertaking a research investigation as part of their course sits somewhere, perhaps uncomfortably, between these two positions. Realistically many students recognise that their research endeavours are unlikely to contribute significantly to theoretical advance, policy decisions or commercial product development, although obviously there are exceptions. Yet they also recognise that their tutors are unlikely to be impressed simply by the presentation of new information. Such research investigations (often called independent projects) are typically included in undergraduate degree programmes to provide students with training that prepares them for a career where such skills as collecting and assimilating information will prove useful, whether this be in academia or more typically in other professional fields. Students face a dilemma to which there is no simple answer. They need to demonstrate that they have carried out their research in a rigorous scientific manner using appropriate quantitative and qualitative techniques, but they do not want to overburden the assessors with an excess of detail that obscures the implications of their results.

In the 1950s and 1960s, several academic disciplines 'discovered' quantitative analysis and few geography students of the last 50 years will not have heard of the 'quantitative revolution'. There was, and still is, a belief that the principles of rigour, replication and respectability enshrined in scientific endeavour sets it apart from, and superior to, other forms of more discursive academic enquiry. The adoption of the quantitative approach was seen implicitly, and in some cases explicitly, as providing the passport to recognition as a scientific discipline. Geography and other Earth Sciences were not alone, and perhaps were more sluggish than some disciplines, in seeking to establish their scientific credentials. The classical approach to geographical enquiry followed from the colonial and exploratory legacies of the 18th and 19th centuries that permeated regional geography into the early 20th century and concentrated on the gathering of information about places. Using this information to classify and categorise places seemed to correspond with the inductive scientific method that recognised pattern and regularity in the occurrence of phenomena with any preconception or theory about what would be found. However, the difficulty of establishing laws about intrinsically unique places and regions led other geographers to search for ways of applying the deductive scientific method. The deductive method involves devising hypotheses with reference to existing conditions and testing them using empirical evidence obtained through the measurement and observation of phenomena.

Geography and to a lesser extent the other Earth Sciences have emerged from a period of self-reflection on the scientific nature of their endeavour with an acceptance that various philosophies can coexist and further their collective enterprise. Thus, many university departments include physical geographers and Earth scientists, adhering to positivist scientific principles, working alongside human geographers following a range of traditions including humanism, Marxism and structuralism as well as more positivist social science. A further important development over the past 40 years has been the growing influence of information and communications technologies on the conduct of academic and other types of research. The most obvious illustration of this being the use of computers for collecting, storing, processing and visualising data. The development of Geographical Information Systems (GIS) and remote sensing has especially prompted the emergence of new directions in geographical enquiry and fostered an expanded range of quantitative techniques for the analysis of spatial data. Students in academic departments need to be equipped with the skills not only to undertake quantitative and qualitative research investigations but also to handle geographical and spatial data in a digital environment.

Johnston (1979) commented that statistical techniques provide a way of testing hypotheses and the validity of empirical measurements and observations. However, the term statistics has several meanings. In general usage, statistics typically refers to information obtained from censuses, surveys and administrative processes, such as recording waiting times in hospital accident and emergency departments, that are published in books, newspapers, on the Internet or other media. The associated term 'official statistics', usually reserved for information collected, analysed and published by national, regional or local government, are commonly regarded as having more authority and being more reliable than those statistics disseminated by commercial or other types of organisation that wish to put across a certain message. This belief may be founded upon a presumption of impartiality and rigour, although recent events may have increased people's scepticism about the neutrality of method or intent in all 'official statistics'. Statistics also refers to a branch of mathematics that may be used in quantitative investigations to substantiate or refute the results of scientific research. Used in this way the term statistics has a double meaning: first as a series of **techniques** ranging from simple summarising measures to complex models involving many variables and second as the numerical **quantities** produced by these techniques. All these meanings of the term statistics are relevant to this text, since published statistics may well contribute to research investigations, and statistical techniques and the measures associated with them are an essential part of quantitative analysis. Such techniques are applied to numerical data and serve two general purposes:

- to confirm or otherwise the significance of research results towards the accumulation of knowledge with respect to a particular area of study;
- to establish whether empirical connections between distinctive characteristics for a given set of phenomena are likely to be genuine or spurious.

Different areas of scientific study and research have over the years carved out their own particular niches. For example, in simplistic terms the Life Sciences are concerned with living organisms, the Chemical Sciences with organic and inorganic matter, Political Science with national and international government, Sociology with social groups and Psychology with individuals' mental condition. Subdivision has often occurred when these broad categories became too general, which led to the emergence of fields in the Life Sciences such as cell biology, biochemistry, and physiology. Geography and the other Earth Sciences do not seem to fit easily into this straightforward partitioning of scientific subject matter, since their broad perspective leads to an interest in all the things covered by other academic disciplines, even to the extent of Earth Scientists transferring their allegiance to examine **terrestrial** processes on other planetary and celestial bodies. No doubt if, or when, ambient intelligent life is found on other planets, geographers will be there investigating the interactions and associations between members of these non-human societies and their environment. It is commonly argued that the unifying focus of Geography is its concern for the spatial and earthly context in which those phenomena of interest to other disciplines make their home. Thus, for example human geographers are concerned with the same social groups as the sociologist but emphasise their spatial juxtaposition and interaction rather than the social ties that bind them. Similarly, geochemists focus on the chemical properties of minerals not only for their individual characteristics but also for how assemblages combine to form different rocks in distinctive locations. Other disciplines may wonder what Geography and Earth Science add to their own academic endeavour, but geographers and Earth scientists are equally certain that if their area of study did not exist, it would soon need to be invented.

The problem that all this raises for geographers and Earth scientists is how to define and specify the units of observation, the things that are of interest to them, and them alone. One solution that emerged during the era of quantification was that geography is pre-eminently the science of spatial analysis and its focus is to discover the laws that govern the processes giving rise to spatial patterns of diverse types of human and physical phenomena (Hartshorne, 1958). A classic example of this approach in human geography was the search for regions or countries where the collections of settlements conformed to the spatial hierarchy theorised by central place theory (e.g. Getis and Getis, 1966; Smith, 1986). Physical geographers and Earth scientists also became interested in spatial patterns. They had more success in associating the spatial occurrence of environmental phenomena with underlying explanatory processes, such as the lines of earthquakes and volcanoes associated with the movement of tectonic plates around the Earth. According to this spatial analytic approach, once the geographer and Earth scientist have identified some spatially distributed phenomena, such as settlements, hospitals, earthquakes or volcanoes, then their investigation can proceed by quantifying variables such as adjacency, connectivity and distance and from these describe the spatial pattern.

This foray into spatial theorising in the 1950s and 1960s became progressively undermined, when it became apparent that exception rather than conformity to proposed spatial laws was the norm. Other approaches, or possibly paradigms, emerged, particularly in human geography, that sought to escape from the straight jacket of positivist science. Advocates of Marxist, behavioural, politico-economic, structural and cultural geography have held sway at various times over the intervening 50 years. However, it is probably fair to say that none of these have entirely rejected using numerical quantities as a way of expressing geographical difference. Certainly, in physical geography and Earth Science where many would argue that positivism inevitably still forms the underlying methodology, quantitative analysis has never receded into the background. Despite the vagaries of all these different approaches, most geographers and Earth scientists still hold on to the notion that what interests them and what they feel other people should be reminded of is that the Earth, its physical phenomena, its environment and its inhabitants display differences and similarities and are unevenly distributed over space.

From the practical perspective of this text, we need to agree what constitutes legitimate data for the study of Geography and Earth Science. Let us simply state that a geographical or Earth sciences dataset is a collection of data items, facts if you prefer, that relate to a group of spatially distributed phenomena (things). Such a dataset should relate to at least one discrete, defined and delimited portion of the Earth, its surface or its atmosphere. This definition deliberately provides a wide scope for using many diverse types and sources of data to investigate and hopefully answer geographical and Earth science research questions.

The spatial focus of Geography and the Earth Sciences has two significant implications. First, the location of the phenomena or observation units, in other words where they are on or near the Earth's surface is regarded as important. For instance, investigations into landforms composed of calcareous rocks need to recognise whether these are in temperate or tropical environments. The nature of such landforms including their features and structures is in part dependent upon prevailing climatic conditions either now or in the past. Second, the spatial arrangement of phenomena may be important, which implies that a means of numerically quantifying the position of different occurrences of the same category of observation may be required. In this case, spatial variables such as area, proximity, slope angle, aspect and volume may form part of the data items captured for the phenomena under investigation.

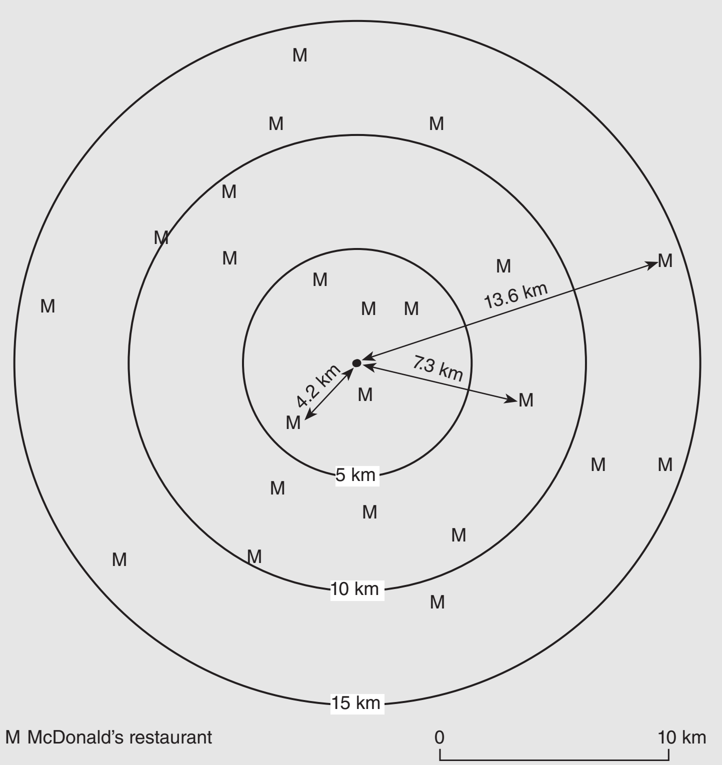
1.2 Nature of Numerical Data

We have already seen that quantitative analysis uses numbers in two ways, as shorthand labels for qualitative characteristics to save dealing with long textual descriptions or as actual measurements denoting differences in magnitude. Underlying this distinction is a division of data items into

attributes and **variables**. Williams (1984: 4) defines an attribute as 'a quality … whereby items, individuals, objects, locations, events, etc. differ from one another'. He contrasted these with variables that 'are *measured* … assigned numerical values relative to some standard – the *unit of measurement*' (Williams, 1984: 5). Examples of attributes and variables from Geography and other Earth Sciences are seemingly unbounded in number and diversity, since these fields of investigation cover such a wide range of subject areas. Relevant attributes include such things as rock hardness, soil type, land use, ethnic origin and housing tenure, whereas stream discharge, air temperature, population size, journey time and number of employees are examples of variables. The terms attribute and variable are sometimes confused and applied interchangeably, although we will endeavour to keep to the correct terminology here.

Numerical data are commonly categorised according to their **scale of measurement**. There are four scales usually known as **nominal**, **ordinal**, **interval** and **ratio** and can be thought of as a sequence implying a greater degree of detail as you progress from the nominal to the ratio. However, strictly speaking, the nominal is not a scale of measurement since it only applies to qualitative attributes and therefore is an assessment of qualitative difference rather than magnitude between observations. The other three scales provide a way of measuring the difference between observations to determine whether one is smaller, larger or the same as any other in the set. Box 1.1a,b summarise the features of each of these measurement scales and show how moving from the interval/ratio to the nominal scale results in information being thrown away. Such collapsing or recoding of the values for data items may be carried out for various reasons but may be necessary when combining a qualitative data item, such as housing tenure, with a quantitative one, such as household income.

Box 1.1a Scales of measurement: location of a selection of McDonalds restaurants in Pittsburgh, USA.



Box 1.1b Definitions and characteristics.

- Nominal scale attributes record a qualitative difference between observations.
- Ordinal scale variables rank one observation against others in the set with respect to the ascending or descending data values of a particular characteristic. These are particularly useful when analysing information obtained from secondary sources where the raw data are unavailable.
- Interval scale variables record differences of magnitude between observations in a set where the measurement units have an arbitrary zero.
- Ratio scale variables record differences of magnitude between observations in a set where the ratio between the values of any two observations remains constant irrespective of measurement units, which have an absolute zero.

For most practical purposes, the interval and ratio scales are treated as equivalent with respect to various analytic procedures. The four scales of measurement are listed in order of increasing precision and it is possible to collapse the detail for a characteristic measured on the interval/ratio scales to the ordinal and to the nominal. This process results in a loss of detail as information is discarded at each stage, but this simplification may be necessary when analysing data from across different measurement scales. Box 1.1c illustrates that progressive loss of detail occurs with respect to the distance measurements between the centre of Pittsburgh and a selection of the MacDonald's restaurants in the city. Distance is measured on the ratio scale, with an absolute zero in this case indicating that a restaurant is located exactly at the centre. Sorting these raw data measurements in ascending order and assigning rank scores converts the information into an ordinal scale variable. There are two restaurants closest to the centre at 2.2 km from this point and the furthest is 18.4 km. Further loss of detail results from allocating a qualitative code to denote whether the restaurant is located in the inner, middle or outer concentric zone around the city centre. Although the detailed raw data of ratio/interval variables can be collapsed in this way, it is not possible to go in the other direction from qualitative attributes, through ordinal to interval/ratio scale variables.

Box 1.1c Collapsing between scales of measurement: distance of a selection of MacDonald's restaurants from centre of Pittsburgh, USA.

	Distance to centre		Rank score	Concentric zone
Restaurant No.	Ratio variable	Sorted distance	Ordinal variable	Nominal attribute
1	17.5	2.2	1	1 (inner)
2	15.1	2.2	1	1 (inner)
3	14.2	2.8	3	1 (inner)
4	10.3	3.6	4	1 (inner)
5	6.0	4.2	5	1 (inner)
6	8.2	6.0	6	2 (middle)
7	9.2	6.4	7	2 (middle)
8	10.8	7.0	8	2 (middle)
9	4.2	7.2	9	2 (middle)

(Continued)

Box 1.1c (Continued)

	Distance to centre		Rank score	Concentric zone
Restaurant No.	Ratio variable	Sorted distance	Ordinal variable	Nominal attribute
10	3.6	7.3	10	2 (middle)
11	2.2	8.1	11	2 (middle)
12	7.0	8.2	12	2 (middle)
13	2.2	9.2	13	2 (middle)
14	2.8	10.3	14	3 (outer)
15	11.2	10.8	15	3 (outer)
16	8.1	11.2	16	3 (outer)
17	6.4	11.2	16	3 (outer)
18	7.2	13.6	18	3 (outer)
19	7.3	13.9	19	3 (outer)
20	11.2	14.2	20	3 (outer)
21	13.9	15.1	21	3 (outer)
22	13.6	17.5	22	3 (outer)
23	18.4	18.4	23	3 (outer)

Most investigations will analyse a number of attributes and variables, and these are likely to be measured using one, two, three or all four of these scales of measurement. It is important to recognise the different scales, because some types of statistical quantity and technique are only used correctly with one measurement scale and not the others. In other words, statistical analysis is not just a question of 'pick and mix', more 'horses for courses' or fitting a 'round peg in a round hole'. An important distinction is between discrete and continuous measurements, which distinguishes nominal and ordinal data from interval and ratio scale variables. Discrete data values are when observations can only take certain values, usually integers (whole numbers), whereas theoretically any value, including negative ones for interval scale variables, is allowed with continuous measurements. Information technology has helped investigators to carry out research by using software to apply statistical analysis to numerical data, but we still usually need to 'tell' the software whether a given set of integers are nominal category labels for qualitative attributes, an ordinal sequence of integers or real data values on the interval or ratio scales. Investigators need to intervene to select appropriate types of analysis and not just tell the software to produce the same set of analyses for all their attributes and variables. Undoubtedly comments such as these may not be applicable in the future, but investigators will always be wise to confirm that software has made sensible decisions.

Investigations using attributes and variables involve holding the data in a numerical format. In the case of nominal attributes, this usually involves simple integer numbers such as 1, 2, 3 up to J, where J equals the number of categories, and there is a one-to-one correspondence between the number of categories and integers. However, in some cases, negative integers may be used. In surveys, people are often asked their opinion about a series of statements in terms of strongly disagree, disagree, neutral, agree or strongly agree. These qualitative labels may be linked to the numbers −2, −1, 0, 1 and 2, rather than 1, 2, 3, 4 and 5. The ordinal scale is based on ranking observations in order from smallest to largest, or *vice versa*. These are often converted into a rank score (first, second, third, etc.)