ANALYSING QUANTITATIVE SURVEY DATA

for BUSINESS and MANAGEMENT STUDENTS

JEREMY DAWSON

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

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: Lyndsay Aitken
Production editor: Sarah Cooke
Marketing manager: Catherine Slinn
Cover design: Francis Kenney
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound by CPI Group (UK) Ltd,
Croydon, CR0 4YY

First published 2017

Library of Congress Control Number: 2016935698

British Library Cataloguing in Publication data

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

ISBN 978-1-47390-750-8
ISBN 978-1-47390-751-5 (pbk)

1

AN INTRODUCTION TO CLASSICAL TEST THEORY AND QUANTITATIVE SURVEY DATA

1.1 ABOUT THIS BOOK

This book covers types of analysis that apply predominantly to data gathered via quantitative surveys. It is intended for non-experts who may be conducting survey research for the first time – for example for a student dissertation. As such, it has relatively few details of the mathematical underpinning of these methods (although some are essential to aid understanding), and concentrates more on the key principles of when and how the analysis should be done, and how it can be interpreted.

Although this text covers all of the key issues specific to analysis of quantitative survey data using classical test theory, many aspects of a broader study may be covered elsewhere. In particular, most types of analysis that would be used to describe data, or to test hypotheses, are covered in another book in this series (Scherbaum and Shockley, 2015). However, it will cover aspects of analysing survey data that other basic guides to data analysis might miss – in particular, data cleaning, reliability analysis, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Although it does not assume you have read a previous book in this series on questionnaires (Ekinci, 2015), it would be advisable to do this if you are starting by constructing and administering your own questionnaire as this book is concerned with data analysis rather than collection. There will be some points in this book where I refer back to that text rather than explaining something fully again.

1.2 SURVEY DATA AND QUESTIONNAIRES

Surveys have long been used as an important method of data gathering in social science and other fields. As such, a wide range of methods have been developed over recent decades to enable the appropriate analysis of such data; in particular, in disciplines such as psychology this has become an important area of research in its own right (the field of *psychometrics* is devoted to studying the theory and technique of psychological measurement, which is predominantly survey-based).

One of the reasons surveys are so popular is their flexibility. Then can be used to collect both quantitative and qualitative data (this volume is concerned only with the quantitative part); they are sometimes the only valid way of collecting some quantitative data (e.g. opinions, attitudes, perceptions – things that cannot be measured directly); and they allow a large amount of data to be collected using consistent and relatively inexpensive methods.

Quantitative data within questionnaires can take on different forms, however – as it can with any source of data. Before going into detail on the types of methods that can be used, it is essential to understand these different data types.

The first distinction it is important to understand is that between *categorical* and *numerical* data. Categorical data refers to variables that can take on different categories – for example, sex, nationality and occupation. Each value represents a different thing, or category, but this is not a numerical quantity (although we may choose to use numbers as labels for these categories). Numerical data refers to variables that have a meaningful numerical value, whether on a naturally occurring or constructed scale – for example age, income, number of children, well-being measured on a scale from 1 to 10.

However, within these two major groups, there are sub-groups that it is important to understand: these are described and exemplified within Table 1.1.

Table 1.1 Six different types of questionnaire data

Major type	*Sub-type*	*Description*	*Examples*
Categorical	Nominal	Separate categories that have no natural ordering: just separate groups	Nationality (e.g. British, German, Chinese) Occupational group (e.g. teacher, doctor, lawyer)
	Binary	A special case of nominal data in which there are only two categories	Sex* Questions requiring a 'Yes/No' response
	Ordinal	Separate categories that have a natural and consistent ordering – so it is always possible to say whether one category is higher than another	Job grade (e.g. grades 1, 2, 3, 4, 5) Class of degree awarded (e.g. First, Upper second, Lower second, third)
Numerical	Continuous: interval	A number representing a quantity, but can be measured on any scale, not necessarily a naturally occurring one (and therefore the value 0 usually has no natural interpretation as it would for a ratio variable)	Job satisfaction (measured on a scale from 1–7) Temperature (if measured in °Celsius or °Fahrenheit – if measured in Kelvin this would be a ratio variable)

Major type	*Sub-type*	*Description*	*Examples*
	Continuous: ratio	A number, which theoretically can take on any possible value within a given range, and measured on a consistent scale such that the value 0 represents an absence of the quality being observed. The accuracy of the number is constrained only by the measurement available	Age Height
	Discrete	Often referred to as a 'count' variable, this is usually a count score that can take on only whole numbers. It is most commonly used when the count is generally low: if the count is very high (e.g. population of countries) then it would usually be treated as a continuous (ratio) variable instead	Number of previous jobs Number of children

*Of course this is not necessarily a simple binary variable as it fails to take into account transgender people – however, it is often measured in a binary way.

The distinction between these different types and sub-types is crucial for making decisions about analysis. Some types of analysis (e.g. correlations) only make sense with continuous data; some (e.g. chi-squared tests) only make sense with categorical data; some (e.g. one-way ANOVA) require a mixture of the two. However, these types of analysis are covered in Scherbaum and Shockley (2015), and will only be mentioned briefly in this book.

For the majority of this book we will concern ourselves with one particular type of data that is very common in questionnaires, but doesn't fit neatly into the categorization in Table 1.1: Likert scales.

1.3 LIKERT SCALES

Likert scales, named after the twentieth-century American psychologist Rensis Likert, are a method of measuring a variable (construct) that cannot be directly measured, by asking respondents to what extent they agree with a series of statements. Each statement is known as an 'item'; technically, a Likert scale is the summation (or average) of the different items, although it is sometimes used to refer to an individual item as well.

For example, a set of three items relating to extraversion (Lang et al., 2011) is:

- I see myself as someone who is talkative.
- I see myself as someone who is outgoing, sociable.
- I see myself as someone who is reserved.

For each item, the respondent would typically choose from one of the following options:

- Strongly disagree.
- Disagree.
- Neither agree nor disagree.
- Agree.
- Strongly agree.

Sometimes a different set of response categories might be used: for example, options ranging from 'Very dissatisfied' to 'Very satisfied', or from 'Not at all' to 'All the time'. Technically these should be referred to as Likert-type scales, as the original definition of Likert scales refers to items asking about the level of agreement only; however, it is common for all such scales to be referred to as Likert scales, and so this book will use the term 'Likert scales' to refer to all sets of items with such rating scales.

The number of response options may vary: in the above example there were five, but this may be four, six, seven or indeed any higher number. Note that if there are an odd number of items, then a symmetrical scale will yield a neutral or average item (e.g. 'Neither agree nor disagree') whereas an even number of items (e.g. six) will not.

In any case, each item should be considered as an *ordinal* variable. That is, the respondent chooses from one of a number of ordered categories. However, when taken together this changes. The purpose of asking three separate questions about extraversion here is not that the individual items are themselves of particular interest, but that between them they should give a better overall indication of the level of extraversion. Therefore a single score for the construct (extraversion) needs to be created.

For this purpose, a number is assigned to each of the responses – typically these would be 1, 2, 3, 4, 5 for the above example, although for the third question (which measures the extent to which someone is reserved – the opposite of what we would

1. To what extent do you agree with the following?

	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
a. I see myself as someone who is talkative	☐	☐	☐	☐	☐
b. I see myself as someone who is outgoing, sociable	☐	☐	☐	☐	☐
c. I see myself as someone who is reserved	☐	☐	☐	☐	☐

Figure 1.1 Lang et al.'s (2011) extraversion scale in questionnaire form

expect for an extravert) a response of 'strongly disagree' means higher extraversion, and therefore we would code these responses as 5, 4, 3, 2, 1 respectively. This is often referred to as a 'negatively worded question'.[1]

The overall score for extraversion, then, would be calculated as the average – the arithmetic mean – or, alternatively, the sum of these three item scores. So, for example, if someone responded 'Strongly agree' to 'I see myself as someone who is talkative' this would be scored 5; if they responded 'Neither agree nor disagree' to 'I see myself as someone who is outgoing, sociable' this would be scored 3; and if they responded 'Disagree' to 'I see myself as someone who is reserved', this would be scored 4 (because it is a negatively worded question). The overall extraversion score for this individual, then, would be $(5 + 3 + 4) \div 3 = 4.0$.

You may notice that, by doing this, we are treating the ordinal measurement of the individual items in a more numerical way, and the eventual score for extraversion is no longer categorical, but actually resembles an interval variable. In fact, it is then usually treated in analysis as if it is a continuous, numerical variable.

Is this justified? Research suggests that it can be, but only if the Likert scale (or just 'scale') has good reliability and validity – concepts that are hugely important in survey research, and will be the focus for large sections of this book. These will be introduced formally in Chapter 2.

1.4 CLASSICAL TEST THEORY

Classical test theory (CTT) is the measurement theory that underlies the techniques being described in this book. It was developed by Mel Novick, who first published the codification in 1966 (Novick, 1966). It is based around the idea that any measured score consists of two parts – a *true* score and an *error*. The 'error' may represent measurement error, and other types of random or systematic error too – examples of these will be shown in the next chapter. As we will also see in Chapter 2, this can be expressed in a formal mathematical way that allows us to express concepts like reliability and validity in a more formal way.

A key concept in CTT is that, even though our measurement may be constrained by the tools we used, the underlying (true) score is on a continuum. Thus, when someone answers a question such as 'I see myself as someone who is talkative', their true perception of how talkative they are could fall anywhere within a given range – between the point where they would only answer 'strongly disagree' and the point where they would answer 'strongly agree'. If they consider themselves to be quite talkative, for instance, but perhaps not as much as some other people they know, then their true perception may fall somewhere between the points represented by 'agree' and

---

[1] In practice, we would normally start by coding everything as 1, 2, 3, 4, 5, and then recoding any negatively scored items later using computer software – see Chapter 3.

'strongly agree' – it would be up to the respondent to choose which the more appropriate answer is. Thus the item, which appears as an ordinal measurement, actually represents a blunt measurement of a continuous underlying measure.

1.5 TYPES OF ANALYSIS USING SURVEY DATA

When it comes to data analysis, there will normally be several possible stages. Which stages you use, and which types of analysis you apply at each stage, depend on the research questions and objectives, and the types of data available to you. As each type of analysis has its own type of output and possible conclusions, it is highly important that the correct methods of analysis are selected. This is why it is important to consider the type of analysis to be done, and therefore the types of data required, before data are even collected. However, sometimes this is not possible – for example, if using secondary data – so the procedures described in this book allow for the possibility of non-ideal situations also.

1.5.1 Stage 1 – data tidying

This stage does not (usually) lead to conclusions relating to the research objectives themselves, but is an essential first step before undertaking the other stages. After data entry, there will often be a number of tasks needing to be performed on a data set before it is ready to be analysed fully. Some of these are routine with any data set – for example, checking for ineligible values, noting missing data – but others are particularly useful with scale data from surveys.

Where data from multi-item scales have been collected, one aim of this stage is to get an overall score for each scale (e.g. a score for 'extraversion' in the example in section 1.3 above). However, before this can be calculated the following steps should be performed:

- Any negatively worded items should be *recoded*, so that all items have the same 'valence' (that is, a high score either consistently represents a high amount of the quality being measured, or consistently represents a low amount).
- The scale *reliability* should be checked.
- Sometimes (although not always) the scale *validity* should be checked by using factor analysis (either exploratory or confirmatory).
- If the reliability and validity are sufficient, then an overall scale score can be calculated as the average score across the items.

1.5.2 Stage 2 – descriptive analysis

Descriptive analysis is about saying what is in your data set – sometimes this is a sample from a larger population; other times, it might represent the complete set of data available for a given topic. Either way, there are different sets of procedures available for different purposes.

For any numerical data, there are a range of different statistics available that can describe each variable – the mean and median are often used to describe the average, or central, value of a variable; the standard deviation and interquartile range are among several statistics used to describe the spread of a variable; and more advanced statistics such as skewness and kurtosis describe the shape of a variable's distribution.

For categorical data, frequencies and percentages in each category usually suffice, although if there are a large number of categories these might sometimes be grouped into broader categories first.

In either case, graphs can be used for helping to display this information: histograms for numerical variables (although bar charts may be preferred for discrete variables with relatively few values), and either bar charts or pie charts for categorical variables.

The descriptive analysis will often also include comparison of different variables, particularly if this is one of the research objectives. For example, if two categorical variables are to be compared, this is often best achieved by a cross-tabulation; if there are relatively few categories then a clustered bar chart may also be helpful.

If a categorical variable and a numerical variable are to be compared, then descriptive statistics of the numerical variable for each category of the categorical value are a useful way to display this. Various types of graph (e.g. bar charts, boxplots) can also help display this graphically.

To compare two numerical variables, a correlation is often used, accompanied by a scatter plot. More complex techniques may aim to summarize multiple variables at once, for example cluster analysis or principal components analysis (PCA).

This list of methods is not exhaustive, but gives some examples of the most typical procedures used.

1.5.3 Stage 3 – inferential analysis

Very often the data available represent a sample from a wider population (whether that is a population of organizations, or employees, or specific types of customer, or all people, or whatever); the research questions are not really so much about the sample as about the population, and so inferential analysis is used to draw inferences from the sample about the wider population.

In this type of analysis, the sample is usually chosen to be broadly representative of the whole population – for example, if wanting to learn about customers of a business, it might involve a completely random selection of people from a customer database (a 'simple random sample'), a random selection from within each of several types of customer (e.g. one-off customers, occasional customers, regular customers – this would be a 'stratified random sample'), or various other methods. The size of the sample can be quite small compared with the whole population, however, and still yield fairly accurate estimates for the whole population: for example, a sample of 100 will often yield quite good results, and a sample of 1,000 can give very accurate results indeed – even if the population size is in the millions.

This does depend on appropriate methods of analysis being used, however. A whole range of different techniques might be used here, depending on what the aim of the analysis is, and how many variables are involved:

- If a single variable is involved (e.g. estimating what proportion of a population is likely to vote for a particular candidate in a forthcoming election), then descriptive methods combined with confidence intervals are most likely to be used.
- If comparing two variables, then relatively simple techniques such as t-tests, one-way ANOVA, chi-squared tests and correlations can be employed.
- If more than two variables are involved, then more complex methods such as multiple regression, multi-way ANOVA, log-linear modelling and generalized linear models may be used.

For more on the types of analysis used in stages 2 and 3, see the volume in this series (Scherbaum and Shockley, 2015), which covers these in detail.

1.6 THE REMAINDER OF THIS BOOK

The majority of this book is concerned with the survey-specific parts of stage 1 analysis, as covered in section 1.5.1.

Chapter 2 will discuss the foundations of classical test theory, and the epistemological and methodological assumptions that underpin it. Chapter 3 will summarize the procedures used, describing the process of research from the end of data collection to the end of the analysis covered in this volume. The details of how to conduct these methods of analysis will be shown in Chapter 4, with some examples from the management literature shown in Chapter 5. Finally, Chapter 6 will examine some of the strengths and weaknesses of the methods, as well as discussing some of the more contentious decision-making criteria.

1.7 CHAPTER SUMMARY

In this chapter we have introduced quantitative survey data, classical test theory and an outline of the process for conducting analysis with this data. This chapter was not intended to give sufficient detail about these matters, but to serve as an introduction to the topic. These will now be expanded upon in subsequent chapters.

CHAPTER REFERENCES

Ekinci, Y. (2015) *Designing Research Questionnaires for Business and Management Students*. London: Sage.

Lang, F.R., John, D., Lüdtke, O., Schupp, J. and Wagner, G.G. (2011) Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavioral Research Methods*, 43(2), 548–567.

Novick, M.R. (1966) The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1–18.

Scherbaum, C. and Shockley, K. (2015) *Analysing Quantitative Data for Business and Management Students*. London: Sage.