



UNIVERSITY OF LEEDS

This is a repository copy of *An introduction to secondary data analysis with IBM SPSS statistics (1st ed.)*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/123294/>

Version: Accepted Version

Article:

Homer, MS orcid.org/0000-0002-1161-5938 (2018) An introduction to secondary data analysis with IBM SPSS statistics (1st ed.). *Educational Review*, 70 (2). pp. 251-252. ISSN 0013-1911

<https://doi.org/10.1080/00131911.2017.1330503>

© 2017 Matt Homer. This is an Accepted Manuscript of an article published by Taylor & Francis in *Educational Review* on 30 May 2017, available online:
<http://www.tandfonline.com/10.1080/00131911.2017.1330503>. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

An Introduction to Secondary Data Analysis with IBM SPSS Statistics

first edition.

JOHN MACINNES, 2016.

SAGE Publications Ltd, London

£28.99

ISBN: 9781446285770

Review

The use and re-use of secondary data has become a priority with many funders, governments and other interested parties both in the UK and more widely. In the UK, for example, the ESRC has had a number of secondary data initiatives to encourage researchers to make the most of the growing availability of high quality, often longitudinal, datasets. Further, the UK Data Service provides a gateway to numerous sets of secondary data for both researchers and educators to use. In part, students and experienced researchers are being asked whether their research questions can to an extent be answered with data already available. In addition, there has been a shift of emphasis with a move away from the usual process of research questions leading to data generation, and towards the opposite approach, where researchers are also being asked to be creative in considering what questions the existing data can answer.

Secondary data can, of course, be both qualitative and quantitative in nature, but John MacInness's timely book is focussed entirely on the latter. The recognition that quantitative skills in the graduate/researcher workforce are weak, particularly in the UK social sciences, has led to the Q-step initiative which aims to provide greater focus in the social sciences on the importance of developing such skills. The book is apposite in this context, and aims to be a 'comprehensive introduction to

secondary data and big data sources', with a focus on employing SPSS as the main statistical software tool.

The book consists of 13 chapters and uses a range of publically available survey and administrative datasets as data sources. It begins with a nice overview of what secondary data analysis is, and, in the second chapter, discusses what statistical analysis can and cannot do for researchers – key issues such as measurement error, reliability/validity and experimental design are touched on here.

Chapter 3 discusses some online tools (e.g. Gapminder and Nesstar) that readers can use to carry out secondary data analysis from scratch in '5 minutes'. This and subsequent chapters end with a set of exercises in data analysis using some of the tools and datasets discussed in the chapter. The basics of using SPSS software is introduced in Chapter 4 and then in the next chapter, the key issue of data documentation is considered. This is a challenging area for new researchers as the documentation for secondary datasets itself can be quite complex and dense, and yet it is obviously crucial for the researcher to understand exactly what the sample is, what the variables are and so on. Chapter 6 focuses on replicating existing published analysis, an effective pedagogic approach which, when carried out successfully, will bolster readers' confidence. The next couple of chapters, 7 and 8, develop the analytic approach well with material on preparing your data (cataloguing, documenting, removing variable/cases, using syntax for repetitive tasks, cleaning) and then on managing and manipulating your data (summarizing, restructuring, aggregating, merging). Chapters 9 to 12 develop the researcher's analytic toolkit, focussing on ordinary least squares regression for a continuous outcome, and binary logistic regression for a dichotomous response. There is some useful material in these chapters, for example, in encouraging the use of spreadsheets to produce additional quality graphs, and introductions to causal diagrams and residual diagnostics. The book finishes in Chapter 13 by reviewing key issues and highlighting easily committed 'statistical sins'.

As an academic researcher, I found the relative lack of references a bit frustrating but this is perhaps a minor quibble. That said, many interesting and important data analysis issues are discussed, such as descriptive analysis as the necessary start of the process, effect sizes over p-values, and the importance of theory, and of theory-building in quantitative research. There are so many things that could be covered in such a book, that it is unsurprising that some things are left untouched – for example, it would have been interesting to see a consideration of the sample vs. population arguments that some argue render statistical hypothesis testing irrelevant when using administrative datasets where the whole population has been ‘sampled’. Another issue that could have been considered would be the potential contribution of mixed methods to a quantitative secondary data analysis, and the issues that might then arise.

This would certainly be a useful book to support an introductory undergraduate quantitative analysis course for any social science discipline, including in education, even though the examples are not generally focussed on education-specific datasets. The book dips into a number of important issues that could be used to open up meaningful discussion of key issues in quantitative data analysis and to help teach the strengths and pitfalls of using secondary data sets in research. Its unique approach, focussing on secondary data and related issues, certainly complements other well-known, perhaps more methods-focussed, SPSS-based books such as those of Andy Field (2013) and Julie Pallant (2010).

References

Field AP. 2013. *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock “n” roll*. London, UK: Sage Publications.

Pallant J. 2010. *Spss Survival Manual: A step by step guide to data analysis using SPSS*. 4 edition. Maidenhead: Open University Press.

Word Count

806 (not including references)

Reviewer

Matt Homer, School of Education, University of Leeds, m.s.homer@leeds.ac.uk