Author : Corey S. Sparks

Title of Work : Applied Demographic Data Analysis with R

Response to Reviewer comments

Below, I respond to the two reviews in line in **bold** text

Reviewer 2

Original comments

Based on the materials provided by the author(s), we would like to have your advice on the following points:

1. Overall, do you feel that the textbook as outlined in the proposal will be a useful book for its intended audience?

*Yes. I feel that this book will be very useful in the field. An applied demography textbook that introduces data analysis in R will fill a gap between the formal demography textbooks that seldom offer guidance in model fitting etc., let alone programming. It should benefit both social science and non social science students in general. I believe the textbook would be suitable for intermediate through advanced graduate student learners. I think the book would also be useful for practitioners, applied researchers, data scientists, etc. There is great coverage of modern tools (i.e., extending base-R to the tidyverse…and beyond) that should prove useful for those who use R, but are not up-to-speed on the most current libraries.*

**I thank the reviewer for their comments, and agree that the audience will likely include a diverse array of practitioners and students in social science disciplines.**

2. Does it cover the topics mentioned adequately?

*Yes. The coverage seems appropriate for a one semester course. I would like to see some coverage of modern statistical demography and formal demography linkages. However, that might not be the place in this book, which is covering much-needed preliminary material. A great strength in my view is the focus on the generalized linear model (the Swiss Army Knife of statistical modeling).*

**I agree, the book is devised to be a one semester course worth of material, with potential to be supplementary in other classes as well. In terms of the linkage with formal demography, I was hoping in the “macro” section of the book to link statistical models with traditional measures such as the life table and age specific fertility rates.**

3. What is the relative strength of this proposal compared to textbooks that are currently used in similar courses?

*First, the book has a strong programming component that is often neglected in statistics textbooks. Normally, programming examples are relegated to a website and are not well integrated. Second, it is strongly focused on applied analysis geared to getting the learner engaged with data, designing suitable methods of summarization/presentation of statistical results and relationships. Third, it is straightforward and engaging. The focus on the GLM and it’s variants is also a key strength as noted earlier.*

**I agree with the reviewer’s comments that integration of code and content is necessary for learning in this context.**

4. Which parts of the concept would the author need to address before it is suitable for the market segment stated in the proposal? How could the author best address those concerns?

*As a whole this project is solid. The book will be useful to wide audience.*

**I agree with the reviewer again that the book will hopefully find a wide audience.**

5. Which types of courses would benefit from having this textbook available? Are those self-study or classroom courses?

*This book would be ideal in a post-regression methods and statistics course. Many (if not most) graduate social science programs require one or two semesters of statistics (i.e.., basic concepts and linear models). This training, in addition to the natural self-selection that occurs in higher-level statistics courses, makes this book a good choice for applying more in-depth data-science (data handling, processing, and analyzing, presenting, and reasoning from results/output etc.). It is also suitable as a companion book or supplementary reading. It is also written in a way that a motivated autodidact can appreciate.*

**This is my intention that this book be an additional resource beyond traditional social science statistics courses that often do not approach complex survey data and I have done a short survey on twitter and the majority of respondents said they had no coverage of complex survey data in formal courses.**

[**https://twitter.com/CoreySparks1/status/1369342203965956098**](https://twitter.com/CoreySparks1/status/1369342203965956098)

6. If different from the readership stated in the proposal, who do you feel will be the primary and secondary audience for this textbook?

*I would say that the proposal describes the intended readership well.*

**I agree with the reviewer.**

7. Are the planned pedagogical features of this textbook suitable and sufficient, and are they presented in a way that is engaging and attractive to the intended readership?

*Yes. In every way. The introductory chapter covers the essentials of R programming in an informative and engaging manner.*

**I thank the reviewer and think the first chapter is thorough without being overwhelming**

8. Would this proposed textbook appeal to a global market? If there is a geographical bias that would make this more suitable for a limited audience, please indicate which audience this would be.

*I believe it is suitable for a global market. For example, there appear to be examples using DHS data. I believe it would greatly benefit demographers in the developing world as well as those in the developed world.*

**I agree with the reviewer. I am attempting to use data from a diverse array of sources so to find interested in a diverse audience.**

9. Is the proposal written in an engaging style and is the English usage in the sample materials in accordance with the intended level and readership?

*Yes. Highly readable and engaging, but also moderately rigorous.*

**I agree, I hope that the book will be a good combination of rigor and application**

10. Do you have other suggestions for improvement to the author(s)?

*No.*

Reviewer 3

Based on the materials provided by the author(s), we would like to have your advice on the following points:

1. Overall, do you feel that the textbook as outlined in the proposal will be a useful book for its intended audience?   
*Yes.*

**Agreed**

2. Does it cover the topics mentioned adequately?  
*Yes.*

**Agreed**

3. What is the relative strength of this proposal compared to textbooks that are currently used in similar courses?

*Bringing the macro and micro perspective of demography into the same framework with R being the main demonstration tool.*

**I agree, this is rarely done in a demographic perspective, and many studies are working in this “meso” space with a blend of individual and contextual information.**

4. Which parts of the concept would the author need to address before it is suitable for the market segment stated in the proposal? How could the author best address those concerns?

*The market segment is very broad. It presents the book as data analysis for social scientists in general. However, the topics in the proposal cover only a segment of the area of statistical demography. The book should be more specific about what it is, either in the form of changing the title, subtitle or the description.*

**I think that by approaching this project as fitting well within the social science statistics cannon is appropriate. Many areas of social science use demographic data, but are often not exposed to appropriate use and analysis of the data. In this regard this book will fill a gap in instruction. I also agree that I could adjust the description to better summarize what this project is hoping to do.**

5. Which types of courses would benefit from having this textbook available? Are those self-study or classroom courses?

*Courses that cover survey analysis and statistical analysis of demographic data. It could be suited for both, self-study as well as classroom courses.*

**Agreed, I hope this book will find use in this area.**

6. If different from the readership stated in the proposal, who do you feel will be the primary and secondary audience for this textbook?

*Data scientists/analysts (non-demographers) who work with demographic data and want to learn more about statistical models to analyze it.*

*I’m not sure that I agree that “professionals/instructors who wish to learn how to use R” would be the primary users, as there is a huge amount of other R resources that cover learning R in much more depth than this book will.*

**I agree and perhaps my mention that people looking to learn R will use this book may be a miss-statement. This is very much aimed at students learning to analyze demographic data and applied researchers working in the demographic area.**

7. Are the planned pedagogical features of this textbook suitable and sufficient, and are they presented in a way that is engaging and attractive to the intended readership?

*I think so.*

8. Would this proposed textbook appeal to a global market? If there is a geographical bias that would make this more suitable for a limited audience, please indicate which audience this would be.

*The proposal reads like the book will not be geographically biased.*

**Agreed, I am trying to use data from a variety of sources in the hopes to avoid focusing only on one geographic area.**

9. Is the proposal written in an engaging style and is the English usage in the sample materials in accordance with the intended level and readership?

*Yes.*

10. Do you have other suggestions for improvement to the author(s)?

*I have a few general comments regarding the first Chapter:*

* *Code style:*
  + *The code would be more readable if spaces are used around all operators as suggested in many style guides, see e.g.* [*https://jef.works/R-style-guide*](https://jef.works/R-style-guide/)
  + *For a long time, it has been recommended to use TRUE and FALSE instead of T and F, see e.g.* [*https://www.r-bloggers.com/2012/12/r-tip-avoid-using-t-and-f-as-synonyms-for-true-and-false/*](https://www.r-bloggers.com/2012/12/r-tip-avoid-using-t-and-f-as-synonyms-for-true-and-false/)

**Both of these comments are taken to heart and I have already been revising the first chapter to include these recommended formatting changes.**

* *I would recommend to carefully review the sequence of difficulty of the code. For example, details about functions could go after other R basics, or the use of pipes already on page 21 could be overwhelming.*

**Agreed, I can do some reorganization of the presentation to accommodate a more appropriate sequence.**

* *Demographic data often come in the “wide” format. Showing a conversion from “wide” to “long” in order to use ggplot would be useful.*

**This is going to be included.**

* *The book should be useful also for users who already know R and therefore can skip Chapter 1. As there will be users who might have different preferences regarding processing datasets (e.g. data.table users), my suggestion would be to make the code less dependent on dplyr. In fact, demographic data (especially micro data) often contain millions of rows. For such datasets, processing via data.table can increase the memory/time efficiency more than tenfold, in addition to shorter syntax. It would be a pity to make the book less valuable for users who do not want to pass on those benefits.*

**I agree with some of these assessments, but disagree with others. The dplyr and tidyverse packages are essential to modern data processing, and I agree that they can be nominally slower that data.table, based on a short poll on twitter,(** [**https://twitter.com/CoreySparks1/status/1356723970855297034**](https://twitter.com/CoreySparks1/status/1356723970855297034)**) the majority of 123 respondents voted for dplyr or a mixture. It is my goal to us a mixture of these in the book, but dplyr is often seen as a more readable and intelligible form of programming, writ large.**

* *The section on Variable Types is missing the (important) type “logical”.*

**This is noted and I agree, as I use this widely, and I will add this to the chapter.**