Reviewer Comments

Reviewer #1

> "There are many published books that cover this material (theory) in great detail. I would recommend to the author of just briefly introducing these topics (referencing other books to learn about the theory) and focus the majority of the book on providing multiple case studies from these topics (e.g. in each chapter, provide a complete data analysis starting from the preprocessing the data, exploratory data analysis, applying models / inference, summarizing the results). The author can highlight various functions from R packages throughout the data analysis in a supporting way. This would be the most ideal way for an instructor to directly use the examples in this book in the classroom."

Yes, that is a good point. I had planned on focusing primarily on running the reader through the analyses and keep mathematical concepts to a minimum (as I am also not a mathematician and wouldn't feel confident explaining them properly). I would, however, like to avoid jumping too much between packages or using too many different packages, as I think that could confuse the reader.

> "Currently, the provided chapter in the proposed book feels like the author is just demonstrating the functions of various R packages in a set of disconnected datasets. The vignette of an R package can be used demonstrate the functions, without providing greater context of a data analysis. Instead, I think the book would be greatly enhanced and be much more widely used in the classroom if the author would frame each chapter as a complete data analysis of one or two datasets, demonstrating how the R functions can be used in various stages of the data analysis."

I understand where the reviewer is coming from. My reasoning behind using the example data given in the various packages was that it would be easier for the reader to go back to the package documentation and explore additional options that I have not covered. I could, however, frame these package examples more like case studies and explain their background and the data that was used. That would have the advantage that the examples with these packages don't overly complicate things as they are usually not "messy".

In cases where I will use several packages for one continous analysis, I would then showcase maybe one or two of these examples for the whole pipeline.

> "I am concerned about the target audience. The author states she will "primarily address users without mathematical background but with a strong foundation in biology. A basic knowledge of R is assumed". In my experience, the students who have a strong foundation in biology without a mathematical background do not have a basic knowledge in R. If this book is to be used in the classroom, I would encourage the author to introduce the basics of R or at a minimum provide a set of resources that can be used to learning R prior to starting this book."

Yes, I could write a short chapter at the beginning where I would mainly point to resources for learning the basics of R.

> "I believe the author needs to incorporate the use of messy, highthroughput data examples. It is common for books such the one proposed here to use "small" or contrived data examples instead what is commonly found when analyzing real data. Including the length of time to run various functions on different sizes of data would also be valuable."

I am not sure if using big messy datasets is a good way to teach beginners the concepts of statistical genetics. I would suggest that I keep the simple example case studies for the chapters and include exercises for using more complicated datasets. This could tie in nicely with reviewer #2's suggestion to have exercises with downloadable data.

> "I would recommend that the author create a corresponding GitHub repository to provide online exercises. A good example of this from a recently published Chapman & Hall book is here:

> http://mdsrbook.github.io. This provides a venue for the exercises to be updated if mistakes are found and provides tangible examples that can be reproduced in code provided online. Without this, instructors who will use this book will have to read the code in the book, and translate it by hand based to R."

Good idea! I'll do that!

> "I would recommend the above examples be written in a reproducible document such as R Markdown or R Notebooks."

Yes, I'm going to use bookdown (has also been used for the example chapter).

> "I would recommend the author include a section on simulating genetic data. Methods developers and instructors who will use this in the classroom may want students to be able to generate their own genetic data to test out proposed methods or apply different genetic algorithms in R packages."

I'm not really sure about this one as I'm not convinced it will help students understand the statistical genetics concepts beyond using example data and case studies. But it could of course be included, if you think that would be helpful.

Reviewer #2

> "hands on exercises with solutions, many many examples done, and if possible data downloadable to also have R exercises."

See above.

> "more comprehensive."

Yes, 300 pages should work!

> "I would love to have comprehensive book covering concepts and modern statistical tools for estimation and analysis, with tons of examples, and homework."

I will try to find a good balance between explaining the basics and core functions that have been around for a while and modern tools/ new techniques.

> "The second type of book that I ambition a lot to find a helpful and comprehensive one is for a initial level statistical genetics class."

That's what I would aim for.

> "I think that the organization of the chapters could be better, I find it a little mixed up, I would suggest to go to other tables of contents on other books for a better organization. I would also suggest not to focus/organize the book on analysis."

I will think about the narrative of my book some more. But maybe I could change the order of chapters to:

1. Intro
2. Evol. Genetics
3. Pop. Genetics
4. Quant. Genetics etc.

This would allow me to introduce core Genetics concepts as they pertain to evolution, then focus on population structure, etc. and finally go to analysis of traits within populations.

> "Read many other books, I don't agree with the document shared with me that the problem with the other books are always that the other books are old (7 yr or more). The fundamental

> principles of genetics do not change. The type of data available change. Thus, a great book would be one that keep a strong conceptual section in each chapter, and has a second section of 'how to do analysis and inference of... e.g. LD' "

I'm not sure it is feasible to include a whole lot of details on the concepts AND case studies/ exercises in 300 pages. Since he also states that the drawback of many books is that they are too expensive. I think I would try to keep the book as concise as possible on the background but provide sufficient resources for digging deeper. These resources will mainly be papers, because as opposed to books, students usually have access to them via their institution (at least, that's the case in Germany) and wouldn't need to spend additional money. For core concepts, that are central to understanding and evaluating the analysis, I will of course explain them in more detail.

> "However, many people do not realize that their problem is not that they don't know how to do certain analysis with R, but they first need to know genetics/statistics, second learn how to do it and how to interpret the results and evaluate their quality."

That is a good point. I will incorporate the genetics/statistics concepts that I think are relevant for each chapter.

John

> "One good reason to provide solutions or hints if there is no unique answers is that working the problems is the best way to make sure you know the difficulty level. The most common mistake with exercises is not including enough easy ones; it is discouraging to get stuck on the first one."

I will keep that in mind!

> "'Statistical Genetics Analyses" does not sound quite right because of the two nouns. Foulkes used "Applied Statistical Genetics..." Perhaps "Statistical Genetics: Analyses with R"?"

"Statistical Genetics: Analyses with R" sounds good. Since I am not a native speaker, I fully trust you on what sounds good!