Bodo Winter (2019). *Statistics for Linguists: An Introduction using R*. New York and London: Routledge Taylor & Francis Group. Pp. xi+310.

Reviewed for *Phonology* by Shigeto Kawahara (Keio University)

1 General introduction

Statistical skills did not used to be a prerequisite for doing phonological analyses in the generative tradition.[1] In fact, the field was dominated by a strong belief that grammatical knowledge (=competence) is independent of statistical patterns observed in the lexicon (e.g. Chomsky 1957; Halle 1978). This situation is rapidly changing in recent years, due to a growing body of studies demonstrating that phonological knowledge is—or at least can be—stochastic, partly reflecting statistical trends in the lexicon (e.g. Ernestus & Baayen 2003; Hayes & Londe 2006; Pierrehumbert 2020). Another reason for the increased use of statistical methods is the rise of interest in deploying experimental techniques as well as corpus data for phonological argumentation, a general approach that is now known as "laboratory phonology" (Beckman & Kingston 1990; Pierrehumbert et al. 2000). It is probably safe to say that familiarity with statistical techniques has now become a desirable skill for theoretical phonologists to have. One can of course take the position that phonological knowledge is strictly dichotomous, that (lexical) statistics has nothing to do with phonological competence, or that evidence from experimental and corpus studies is irrelevant for phonological theorization. Nevertheless, in order to evaluate how quantitative data/analyses bear upon questions about phonological competence, I would say that it is crucial that we have some basic understanding of statistics. To take an example, MaxEnt Harmonic Grammar has recently been used to model various aspects of linguistic knowledge (e.g. Breiss & Hayes 2020; Goldwater & Johnson 2003; Hayes & Wilson 2008; Shih 2017), and it is mathematically equivalent to (multinomial) logistic regression (Jurafsky & Martin 2019). Whether we endorse it or reject it as a theory of grammar, we need to understand it, and to do so, it is necessary to know how regression works in general.

[1]Bodo Winter and Gillian Gallagher provided very detailed comments on previous versions of this review, for which I am very thankful. Donna Erickson and Seunghun Lee also helped me with the preparation of this final version. All errors that may have remained in this document are my responsibility.

There are already several statistics textbooks written for linguists, including Baayen (2008), Gries (2013), Johnson (2008), and Levshina (2015). Nevertheless, Winter's book is a very welcome addition to the field. I thoroughly enjoyed reading the book myself and learned a lot from it.

2 The structure of the book

The structure of the book is as follows. Chapter 1 is a general introduction to R, followed by an introduction to `tidyverse` in Chapter 2. Even if the reader is already familiar with `tidyverse`, section 2.10 is a must-read. It explains what reproducible research is, why it is important, and why we should not be scared to make our code and data available. Chapter 3 explains various basic statistical measures like means, medians and distributions. Rather than talking about "more traditional" statistical testing like ANOVA, Winter moves onto linear regression model in Chapter 4. Chapter 5 explains various linear and non-linear transformations, including centering, standardization, log-transformation and Pearson correlation analyses. One topic that is of particular importance is centering, which is essential when we interpret multiple regression models with significant interaction terms, which is covered in Chapter 6. Winter spends considerable time explaining how to interpret the coefficients in regression modeling in Chapters 7 and 8. In this first half of the book, it is made very clear that what is important is *to understand the data* using statistical modeling rather than asking whether an effect of interest is statistically significant or not. Modeling comes first; significance testing comes next.

Once this message is registered, Chapter 9 very cautiously introduces statistical inferences, taking t-tests as an example. Chapter 10 explains common misunderstandings about statistical testing, such as how p-values have been misused. Chapter 11 explains significance testing in the context of regression models. Chapter 12 is devoted to logistic regression, a method that is useful to model binary categorical responses. Poisson regression, suited to analyze count data, is discussed in Chapter 13. Any readers who use ANOVA for categorical data (by averaging over participants or items) should carefully consult these chapters. The next two chapters are devoted to explaining linear mixed effect modeling, which has become the *de facto* standard analytical strategy in linguistics and other neighboring disciplines. Chapter 16 offers future perspectives—what I found most useful is the discussion in section 16.7, which argues in detail why it is important to construct statistical models based on our theoretical considerations. In Appendix A, interested readers can learn how more traditional statistical tests such as ANOVA correspond to linear modeling. Appendix B lists recommendations for future reading. The book ends with an index for keywords and one for R functions, both of which seem comprehensive and useable. The e-book version is searchable so the reader may opt for that option if s/he is looking for particular keywords or functions.

3 Distinguishing features of the book

There are many features that make this book special, which are listed in ((1)).

(1) Distinguishing features of the book

a. Emphasis on statistical modeling over significance testing
b. Accessibility/friendliness
c. Emphasis on theory-driven construction of statistical modeling
d. Emphasis on reproducibility of research
e. Use of R, especially `tidyverse`
f. Critical assessment of the common practice in the field

Let me expand on each of these points. First, statistical modeling has an advantage over more traditional significance testing (e.g. t-test or ANOVA) in forcing us to better understand the data, rather than relying too much on p-values, which can be misunderstood and misused in several important ways. Reliance on significance tests can cause a harmful illusion that there is a dichotomous distinction between a particular effect being either significant or not, which has a number of undesirable consequences (see p. 165 of the book and references cited there). As Winter emphasizes throughout the book, statistical modeling instead encourages us to think more deeply about what the data actually look like, what mathematical functions would be useful to fit the data, and what meanings we can extract from the data. Moreover, statistical modeling allows us to make predictions about the data that we have not seen before—something that significance testing alone is unable to do. Although Winter does not make this point very clearly in the book (perhaps intentionally), this is an important advantage of statistical modeling, because making predictions about the future is arguably one important task for science in general (cf. Breiman 2001).

The tone of the book is informal and friendly for an introductory statistics book like this, and my assessment is that this strategy is working very well. Winter uses "you" to refer to the readers, and going through the book may make the readers feel like they are working together with Winter in an actual statistical course.[2] Winter consistently reminds the readers that it is ok, for example, to forget specific R commands, to make mistakes, or to get overwhelmed. This "hand-holding" approach makes the long journey—from simple addition to linear mixed effects modeling—easier to follow. The book can be used as a textbook for both undergraduate and graduate statistical classes, as well as for self-teaching. Statistics can be off-putting, especially for those who think that

[2]The reason may be that the book builds upon many statistics courses that Winter has taught over the last several years. This experience allowed Winter to spot and address some misunderstandings about statistics that are common among the linguistics community, which is quite useful (see §4 below for some actual examples). The book also builds on a short, previously circulated tutorial on linear mixed effects modeling (Winter 2013). This tutorial too is written in a friendly manner, and many students of mine found it helpful. Some readers may be familiar with this tutorial, and those who enjoyed reading it would most likely enjoy reading the book as well.

they are not good at numbers. Winter has done a wonderful job of lowering, or even eliminating, the psychological boundary of such readers. The mathematics underlying the statistics introduced in the book is deliberately kept minimal. However, Winter has many pointers to other books for those who would like to know more about such mathematical foundations (Appendix B).

We tend to believe that there is an objective way to build "the best" statistical model; Winter has done a good job of debunking this "myth." For example, in my own statistics training during my graduate study, I was told that we should include as many factors as possible in ANOVA to soak up variance, but Winter has informative discussion on why this is a bad practice, and argues that model construction should be a subjective, theory-driven process (see especially pp. 277-279). Winter discusses an interesting history of how linguists have adapted linear mixed effects model (p. 242). After Baayen (2008), it was common to include random intercepts only. There was a "swing-back" to include random slopes to maximize the random structure, due to the influential recommendation by Barr et al. (2013). Winter in many portions of the book stresses how important it is to let our theory guide which free varying parameters to include in our model, rather than blindly following what the field sets as its standard. This advice is not just about random slopes and intercepts, but also about which fixed factors to include as well as how interaction terms should be coded.

Winter repeatedly argues that it is important that the data and the code are published together with the paper, in order to promote "open and reproducible research practice" in linguistics. He claims that reproducible research, in which any third-party reader can reproduce the reported results, is essential in order for science to progress in a cumulative fashion. In addition, making the data publicly available may help other researchers conduct additional analyses on the published results, which may reveal new aspects of the data. However, some of us may feel shy about, and/or scared of, making the code and data publicly available—Winter discusses several reasons why we may feel that way, and explains very gently why we should not (pp. 47-50). To promote more reproducible research practice in linguistics, he emphasizes that using R and its markdown function is essential and provides some convincing reasons for why. Winter even goes so far as to say that "teaching SPSS to students is unethical" (xiv), because SPSS is not well suited for open and reproducible research. Making the research results open and reproducible is likely to become a prerequisite for journal publication in the near future; it in fact already has for some journals, to which many readers of *Phonology* may consider submitting their papers (e.g. *Language*, *Laboratory Phonology* and *Linguistics*).

Finally, another aspect which makes this book special as a textbook is that Winter challenges some common practices in the field. For example, he argues against the convention in which we are asked to conduct statistical tests on all possible comparisons, instead of limiting to those comparisons that are theoretically motivated (p. 187). He argues against the use of stepwise regression

(pp. 276-277), which is the standard strategy in some branches of linguistics, primarily because in this approach, we are relegating the critical portion of the thinking process to a machine. He tells the readers a story when he was asked to include interaction terms by a reviewer, and refused to follow that advice because including the interaction terms was not theoretically motivated (p. 278). This general strategy to include some critical remarks about common practices in the field, I believe, nurtures critical thinking and conveys the message to the readers that we need to argue for a particular statistical analysis, just like we argue for a particular linguistic analysis.

4 General recommendation

Taken together, these features make the book informative and useful for a wide range of readers, including beginning students up to professional linguists, i.e. anybody who is eager to learn statistical techniques in modern linguistics research. Especially if one of the following applies, I would say that it is worth reading the book.

(2) Potential target audience of the book (non-exhaustive)

a. I want an accessible introduction to linear mixed effects modeling, the *de facto* standard approach for statistical analyses in modern linguistics and neighboring disciplines.
b. I want an intuitive understanding of linear mixed effects modeling rather than simply "running the appropriate R code" that I obtained from somewhere else.
c. I am not sure why linear mixed effects models are called "mixed" or what the difference is between random slopes and random intercepts.
d. I have not carefully examined the correlation between random slopes and random intercepts in linear mixed effects modeling.
e. I do not know what an offset term, often used in Poisson regression, means.
f. My default statistical strategy is ANOVA.
g. I am not sure how ANOVA differs from linear regression, logistic regression or Poisson regression.
h. I do not know why it is important to center/standardize continuous explanatory variables.
i. I feel more comfortable using spreadsheet software (e.g. Excel) than command-line-based software (e.g. R).
j. I want to learn (more about) R/R markdown/`tidyverse`/`ggplot2`.
k. I am confused with different styles of writing R code.
l. I feel scared of making my data and code publicly available.

m. I am not sure if the difference between exploratory investigation and confirmatory investigation has any practical consequences.

Also, Winter identifies, and corrects, a number of misunderstandings about statistical analyses. For example, it seems common to mistakenly think that we can interpret regression coefficients just as ANOVA main effects, and relatedly, that if we are dealing with a categorical variable, choosing between sum-coding and treatment-coding does not affect the interpretation of the results (for which, see pp. 134-146). It is a frequently-observed thinking trap to believe that the lower the p-values, or the higher the R^2 values, the more correct the proposed hypothesis is (p. 171). I used to think that intercept terms are not very interesting to interpret, but Winter's book, in a couple of places, told me that intercepts convey interesting information, sometimes as much as slope coefficients (e.g. p. 205). Relatedly, I used to think that it is not interesting or important to examine the estimates for random effects, because they just represent random variability coming from inter-speaker/inter-item differences; however, Winter reminds us that we can learn much from random effects, sometimes more than we can learn from fixed effects (e.g. p. 272). One may think that we can safely interpret model results as long as we do not get a warning message or a convergence issue, but the story is not that simple (pp. 257-260). This list can go on. Readers who are interested in finding out more about these issues are encouraged to actually read the book.

In short, the book is extremely informative on various ends. On the other hand, the book is not for readers who want to learn more traditional significance tests like t-tests or ANOVA. Neither is the book appropriate for readers who simply want to know which test to apply given the nature of their data (what Winter refers to as "a cookbook approach": pp. 275-276), or those who want to know the nuts and bolts of the mathematical foundation of statistical modeling. There are a plenty of other books out there for these purposes (e.g. Levshina 2015, see also those listed in Appendix B).

To recap, the most important message of this book is *to engage more with the data*. It is as dangerous to apply a linear mixed effects model as to apply ANOVA, if we are not carefully examining the nature of the data, not considering what mathematical function should be used to fit the data or which factors should be included in the model, not examining the resulting coefficients, or not thinking about what the model predicts about new data. After reading the book, readers will be convinced that they should at least be able to defend why they chose the model that they did for their analyses, going beyond merely saying that "I am following the standard practice in our field."

Although the book is written in a very friendly manner, Winter does not allow us to be "lazy"; he consistently encourages the reader to go beyond mindlessly applying particular statistical tests that the field sets as "the standard." Relatedly, another general strategy that Winter deploys in the book is that he encourages the readers to generate a toy example and reanalyze the constructed example with regression modeling. This strategy I think is very helpful—especially, this exercise

in the context of linear mixed effects model (pp. 245-253) helps us understand the general structure of linear mixed effects modeling.

Throughout the book Winter uses data from his own research for illustration. Winter is a cognitive linguist by profession, but because his research interests include a wide range of topics, some of the data are of direct interest for many readers of *Phonology*, such as response time data, speech error counts, and vowel durations. Generally, the topics chosen in the book are concepts that are rather easy to understand, so any student of linguistics would be able to understand what is being analyzed. All the data and the R code used in the book are made available on its open science foundation (osf) website (`https://osf.io/34mq9`). The book is replete with practical advice regarding how to use R for quantitative analyses as well as what values are to be reported in a published paper, and how. Each chapter comes with a number of exercises, and the solutions are made available on its osf repository, another feature which makes the book useful as a textbook for actual classes and self-teaching.

5 Final remarks

There are a few typos but not to the extent that it hinders the readability. The book discusses the use of some colors but they do not appear (at least not clearly) in the printed book, but this is a very minor problem (although it could be confusing to some beginning readers).[3]

If there is a second edition of the book, which one hopes that there will be, given Winter's ability to explain statistical concepts so clearly, there are a couple of topics that I wish to be covered in further detail, which includes measures like log-likelihood, (null and residual) deviance and AIC/BIC as well as other statistical methods including ordinal logistic regression, Generalized Additive Models (GAMs), and hierarchical Bayesian regression models. Since log-likelihood and the other related measures are key notions to interpret the model's fit to the data and its ability to make predictions, in a book of this sort which focuses on statistical modeling, they could have been discussed in further depth. A brief discussion on ordinal logistic regression, an extension of logistic regression to deal with ordinal data, can be added at the end of Chapter 12, since it is very common for linguists to use an ordinal scale for their experiments (e.g. "1" is completely unacceptable and "5" is completely acceptable).[4] GAMs and Bayesian regression modeling are becoming increasingly common in linguistic analyses, and I would love for Winter to provide an intuitive understanding of these analytical strategies. For the time being, a reader who is interested in the intuitive illustration of Bayesian modeling, a tutorial offered by Franke & Roettger (2019),

[3] Winter informed me that it was the publisher's decision not to use colors. Interested readers can run the relevant R code, available on the osf repository, on their own and check out the colors themselves.

[4] The `clmm` function in the `ordinal` package (Christensen 2019) can be used to implement mixed effects ordinal logistic regression analyses.

which itself is based on Winter's (2013) work, is very readable, and a more advanced reader can consult Vasishth et al. (2018). Those who are interested in GAMs can consult the references that Winter cites on p. 155, especially Winter & Wieling (2016) and Wieling (2018).

In conclusion, I wish there was a book like this when I started learning statistics myself. If I was to teach a statistical course myself, this book would undoubtedly be the first candidate to use as a textbook. If a student is looking for a statistics textbook, then this book is my first recommendation.

References

Baayen, Harald R. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.

Beckman, Mary & John Kingston. 1990. An introduction. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 1–16. Cambridge: Cambrdige University Press.

Breiman, Leo. 2001. Statistical modeling: The two cultures. *Statistical Science* 16(3). 199–231.

Breiss, Canaan & Bruce Hayes. 2020. Phonological markedness effects in sentential formation. *Language* 96. 338–370.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Christensen, Rune Haubo Bojesen. 2019. `ordinal`. R package.

Ernestus, Mirjam & Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79(1). 5–38.

Franke, Michael & Timo B. Roettger. 2019. Bayesian regression modeling (for factorial designs): A tutorial. Ms. `https://doi.org/10.31234/osf.io/cdxv3`.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory* 111–120.

Gries, Stefan Th. 2013. *Statistics for linguistics with R*. Berlin: Mouton.

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 294–303. Cambridge: MIT Press.

Hayes, Bruce & Zsuzsa Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Johnson, Keith. 2008. *Quantitative methods in linguistics*. Oxford: Wiley-Blackwell.

Jurafsky, Daniel & James H. Martin. 2019. *Speech and language processing (3rd edition, draft)*. `https://web.stanford.edu/~jurafsky/slp3/`.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2020. 70+ years of probabilistic phonology. In B. Elan Dresher & Harry van der Hulst (eds.), *Oxford handbook on the history of phonology*, Oxford University Press.

Pierrehumbert, Janet, Mary Beckman & Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Phillip Carr & Gerard Docherty (eds.), *Phonological knowledge: Conceptual and empirical issues*, 273–303. Oxford: Oxford University Press.

Shih, Stephanie S. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 34(2). 243–268.

Vasishth, Shravan, Bruno Nicenboim, Mary Beckman, Fangfang Li & Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71. 147–161.

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70. 86–116.

Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications. Ms. arXiv:1308.5499, `https://arxiv.org/pdf/1308.5499.pdf`.

Winter, Bodo & Martijn Wieling. 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1. 7–18.