Notes on Biostatistics

Phùng Khánh Lâm

2019-11-10

Contents

Preface

Bookdown

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter **??**. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure **??**. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table **??**.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (**?**) in this sample book, which was built on top of R Markdown and **knitr** (**?**).

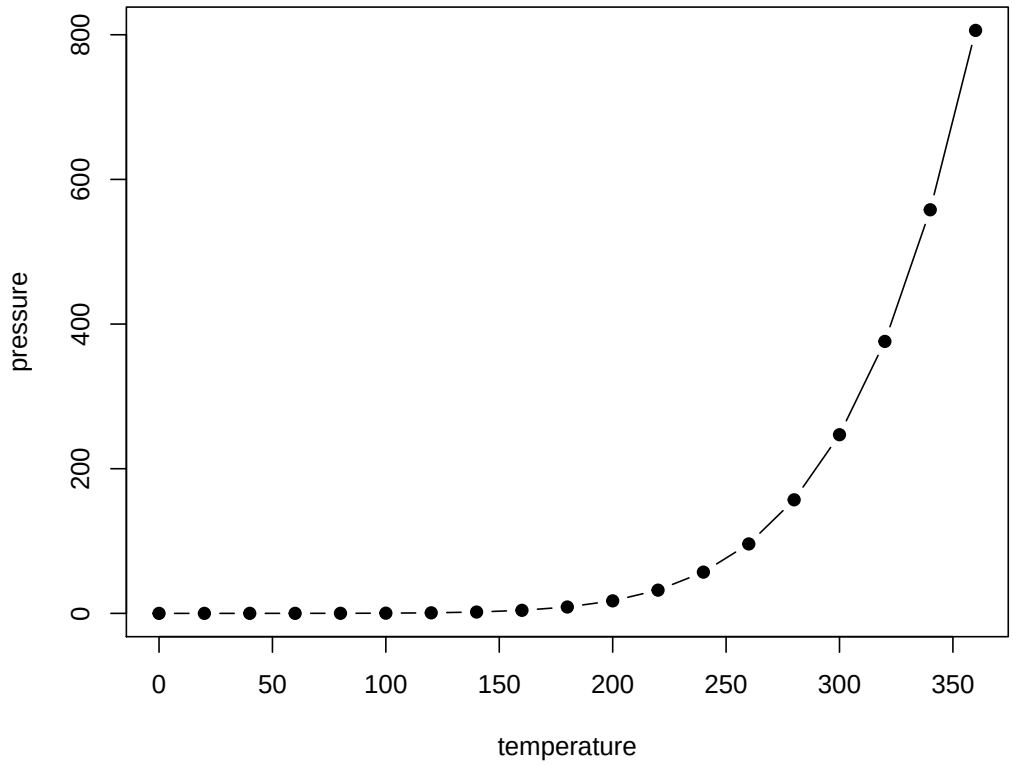

Figure 1: Here is a nice figure!

Table 1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 1

New

- Bianchi, M. T., Alexander, B. M., & Cash, S. S. (2009). Incorporating Uncertainty Into Medical Decision Making: An Approach to Unexpected Test Results. Medical Decision Making, 29(1), 116–124. https://doi.org/10.1177/0272989X08323620
- Nobles AL, Leas EC, Althouse BM, et al. Requests for Diagnoses of Sexually Transmitted Diseases on a Social Media Platform. JAMA. 2019;322(17):1712–1713. doi:https://doi.org/10.1001/jama.2019.14390
- Steve Ruberg (2019) No. 13: Unconsciously Biased and Consciously Unbiased. https://analytixthinking.blog/2019/11/07/no-13-unconsciously-biased-and-consciously-unbiased/
- renv: Project Environments for R. https://blog.rstudio.com/2019/11/06/renv-project-environments-for-r/
- The frontier of simulation-based inference. https://deepai.org/publication/the-frontier-of-simulation-based-inference
- Non-randomized studies using causal-modelling may give different answers than RCTs: a meta-epidemiological study. DOI: https://doi.org/10.1016/j.jclinepi.2019.10.012.
- Four mistakes to avoid when talking with patients about risk. https://www.aafp.org/journals/fpm/blogs/inpractice/entry/risk_communication_mistakes.html#.XDYKYKwH6CA.twitter
- 5 mistakes of communicating risk. https://www.youtube.com/watch?v=jAY3H92MXGU
- **?**'s keynote on Machine learning: calm down!. https://twitter.com/Georg___Heinze/status/1192796427619065857
- What logic gives us the authority to say something about an individual by observing other individuals, however similar? https://twitter.com/yudapearl/status/1192768005316341760. http://ftp.cs.ucla.edu/pub/stat_ser/r375-reprint.pdf
- Just released today, and it is shameful: https://ncd.gov/sites/default/files/NCD_Quality_Adjusted_Life_Report_508.pdf. Recommendation against QALYs and #healtheconomics to support resource allocation decisions. https://twitter.com/dollendorf/status/1192181668985102338
- stats might kill personalized medicine. https://twitter.com/KertViele/status/1192194206049226754
- RStudio Package Manager. https://blog.rstudio.com/2019/11/07/package-manager-v1-1-no-interruptions/

Chapter 2

Initial data analysis

Chapter 3

Data cleaning

Chapter 4

Descriptive analysis

Chapter 5

Graphics

Chapter 6

Causal Inference

6.1 Reading list

- Daniel Westreich (2019) Epidemiology by Design: A Causal Approach to the Health Sciences. Oxford University Press. https://global.oup.com/academic/product/epidemiology-by-design-9780190665760?cc=vn&lang=en&

Chapter 7

Study design

Chapter 8

Hypothesis Testing

Chapter 9

P value

Chapter 10

Bayesian data analysis

Chapter 11

Prediction models

Chapter 12

Medical decision making

12.1 Reading list

12.1.1 Risk communication

- Sisk BA, Baker JN (2018) Microethics of Communication—Hidden Roles of Bias and Heuristics in the Words We Choose. JAMA Pediatr;172(12):1115–1116. doi:https://doi.org/10.1001/jamapediatrics.2018.3111.
- Steinhardt Joe (2019) The role of numeric and statistical content on risk perception in infographics about road safety, Journal of Risk Research, DOI:10.1080/13669877.2019.1596147
- Kunneman M, Stiggelbout AM, Pieterse AH (2019) Do clinicians convey what they intend? Lay interpretation of verbal risk labels used in decision encounters. Patient Educ Couns. 2019 Aug 26. pii: S0738-3991(19)30377-5. doi: 10.1016/j.pec.2019.08.035.
- NICE guidance on risk communication: https://twitter.com/BerksMaternity/status/1171542152842817536/photo/1. Figure
- Okan Y, Smith SG, Bruine de Bruin W (2019) How is cervical cancer screening information communicated in UK websites? Cross-sectional analysis of content and quantitative presentation formatsBMJ Open 2019;9:e029551. doi: 10.1136/bmjopen-2019-029551.
- Yen, Renata W, Barr, Paul J, Cochran, Nan, Aarts, Johanna W, Légaré, France, Reed, Malcolm, O'Mallley, A James, Scalia, Peter, Guérard, Genevieve Painchaud, Backer, Grant, Reilly, Clifford, Elwyn, Glyn and Durand, Marie-Anne (2019) Medical students' knowledge and attitudes towards shared decision-making: results from a multinational cross-sectional survey. Medical Decision Making. http://sro.sussex.ac.uk/id/eprint/85728/
-

Chapter 13

Intensive Longitudinal data analysis

13.1 Reading list

13.1.1 Papers:

- T. Asparouhov, E. L. Hamaker, and B. Muthén. "Dynamic Structural Equation Models". In: *Structural Equation Modeling: A Multidisciplinary Journal* 25.3 (Dec. 2017), pp. 359-388. DOI: 10.1080/10705511.2017.1406803. <URL: https://doi.org/10.1080/10705511.2017.1406803>.
- T. Asparouhov and B. Muthén. "Comparison of Models for the Analysis of Intensive Longitudinal Data". In: *Structural Equation Modeling: A Multidisciplinary Journal* (Jul. 2019), pp. 1-23. DOI: 10.1080/10705511.2019.1626733. <URL: https://doi.org/10.1080/10705511.2019.1626733>.
- E. M. Chi and G. C. Reinsel. "Models for Longitudinal Data with Random Effects and AR(1) Errors". In: *Journal of the American Statistical Association* 84.406 (Jun. 1989), pp. 452-459. DOI: 10.1080/01621459.1989.10478790. <URL: https://doi.org/10.1080/01621459.1989.10478790>.
- L. M. Collins. "Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model". In: *Annual Review of Psychology* 57.1 (Jan. 2006), pp. 505-528. DOI: 10.1146/annurev.psych.57.102904.190146. <URL: https://doi.org/10.1146/annurev.psych.57.102904.190146>.
- S. C. Duncan, T. E. Duncan, and H. Hops. "Analysis of longitudinal data within accelerated longitudinal designs." In: *Psychological Methods* 1.3 (1996), pp. 236-248. DOI: 10.1037/1082-989x.1.3.236. <URL: https://doi.org/10.1037/1082-989x.1.3.236>.
- T. E. Duncan and S. C. Duncan. "An introduction to latent growth curve modeling". In: *Behavior Therapy* 35.2 (2004), pp. 333-363. DOI: 10.1016/s0005-7894(04)80042-x. <URL: https://doi.org/10.1016/s0005-7894(04)80042-x>.
- E. L. Hamaker and M. Wichers. "No Time Like the Present". In: *Current Directions in Psychological Science* 26.1 (Feb. 2017), pp. 10-15. DOI: 10.1177/0963721416666518. <URL: https://doi.org/10.1177/0963721416666518>.
- E. L. Hamaker, T. Asparouhov, A. Brose, et al. "At the Frontiers of Modeling Intensive Longitudinal Data: Dynamic Structural Equation Models for the Affective Measurements from the COGITO Study". In: *Multivariate Behavioral Research* 53.6 (Apr. 2018), pp. 820-841. DOI: 10.1080/00273171.2018.1446819. <URL: https://doi.org/10.1080/00273171.2018.1446819>.
- N. C. Jacobson, S. Chow, and M. G. Newman. "The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data". In: *Behavior Research Methods* 51.1 (Aug. 2018), pp. 295-315. DOI: 10.3758/s13428-018-1101-0. <URL: https://doi.org/10.3758/s13428-018-1101-0>.