APA Formatted Manuscripts in R

Ryan Derickson[1,2] & Boris Yanovsky[2,3]

[1] National Center for Organization Development

[2] University of Cincinnati

[3] Facebook

Author Note

Complete departmental affiliations for each author (note the indentation, if you start a new paragraph). Enter author note here.

Correspondence concerning this article should be addressed to Ryan Derickson, 11500 Northlake Dr. Suite 230, Cincinnati OH 45249. E-mail: ryan.derickson@va.gov

Abstract

Enter abstract here (note the indentation, if you start a new paragraph).

Keywords: R, Markdown, Reproducible Research

APA Formatted Manuscripts in R

Background

Wouldn't it be nice if you could spend more time on substantive work and less time on tedious formatting? R isn't just for running analyses - write your whole manuscript in R from start to finish and let R handle the markup! This has many advantages:

1. **Simplify your workflow**. All your work is in one location - no more copy/paste or importing work from your analytic platform into your writing platform.

2. **Focus your effort**. R separates the content from the markup, so you can concentrate on the writing. You don't have to divide your attention between markup and content, because R will handle almost all the markup for you.

3. **Reduce errors**. Because your analyses are in the same workflow as your manuscript, you *know* that your manuscript has the correct and current output. If you modify your analyses, all your in-text results automatically update. No more forgetting which code file does what or forgetting to update your paper if you change your analyses.

4. **Collaborate better**. Contributors can update analyses *and* content. This makes writing very efficient and also creates total transparency in what was done, so others can catch coding errors and see results from analyses even if they are omitted from the final manuscript.

Math

If your writing includes equations, you probably already know about LaTeX. Including equations or symbols is incredibly easy and efficient, compared to WYSIWYG text editors like Word. And, the results look great. For example, the 3PL IRT model is defined as:

$$P(Y_{is} = 1|\theta_s) = c_i + (1 - c_i)\frac{exp(1.7a_i(\theta_s - b_i))}{1 + exp(1.7a_i(\theta_s - b_i))} \quad (1)$$

Use symbols in-line also. For example, θ is ability, a is slope, b is location, and c is the lower asymptote. Refer to equations later, without having to remember which number they are, by calling their labels. This was equation 1.

$$P(Y_{is} = 1|\theta_s) = \frac{exp(1.7a_i(\theta_s - b_i))}{1 + exp(1.7a_i(\theta_s - b_i))} \tag{2}$$

If c is 0, the 3PL reduces to the 2PL model, which is equation 2. R knows the 2PL model is equation 2 and you can rearrange the order without having to go back and update the numbers each time.

Analyses

Include R code chunks that contain your analyses, and they will evaluate when you compile the document. For example, reference tables created in R this way: Table 1, Table 2. Tables are pushed to the end of the manuscript by default.

Include in-line results also, like this: ...the coefficient for horsepower in Model 2 was -0.03. (If you're not familiar with modeling in R, you can extract parameters from the model object). You can perform R operations in-line like the previous example, but it's usually better to do them in the R chunk to make it clearer what you're pushing to the text.

papaja supports several basic types of figures (barplots, heatmaps, etc.) but you're not limited to those - anything can be included. Figure 1 shows a CFA, Figure 2 shows a conditional inference tree with survival analyses in the terminal nodes.

Methods

Download the APA6 template:

```
devtools::install_github("crsh/papaja")
```

For the analyses in this paper, you'll also need:

```
install.packages("lavaan", "rmarkdown", "devtools", "semPlot",
                 "knitr", "party", "TH.data", "afex")
```

To start a new manuscript, File / New / Markdown - from template - APA 6th Edition. References pull from a .BibTeX file. Most libraries allow you to save references in .BibTeX format when you're searching PhychInfo or other databases. .BibTeX entries look something like this:

```
@Article{R-party_a,
  title = {Unbiased Recursive Partitioning: A Conditional Inference Framework},
  author = {Torsten Hothorn and Kurt Hornik and Achim Zeileis},
  journal = {Journal of Computational and Graphical Statistics},
  year = {2006},
  volume = {15},
  number = {3},
  pages = {651--674},
}
```

The first tag ("R-party_a") is the identifier for that source. You can name it whatever you want. Cite the source in text, Hothorn, Hornik, and Zeileis (2006), or parenthetically (Hothorn et al., 2006). When you compile the manuscript, R takes the references you used and creates an APA-formatted reference page automatically. R also helpfully adds citations to your .BibTeX file based on sources cited by the authors of the packages you used in your analyses! They do not appear in your References page unless you cite them yourself, but the entries are there for you to use if you wish. For example, by loading the "party" package, five sources were automatically added to our .BibTeX file.

Here are more sources to make an example reference page: Zeileis, Hothorn, and Hornik (2008), Epskamp and Simon Stuber (2017), Dahl (2016), Genz and Bretz (2009).

Applications

1. Collaborating with colleagues
2. Teaching - no more horribly formatted homework, or having to guess what analyses students performed
3. Own work - keep your work tidy; your future self will thank you.
4. Publishing - many journals require APA formatting, and those that require variations can usually be easily accommodated. Other journals have their own LaTeX templates which can be used. All the benefits of reproducibility transfer to these situations.

Example

We'll test models predicting who survived the Titanic sinking, to illustrate some of the ideas and advantages of reproducible research in R. Change the models, and see how accurately you can predict who survived the Titanic. Also note how any part of the analyses you do here could be pulled into the manuscript, if this was the research topic.

Notes

- Can be compiled to .PDF or .DOC, but a bit of manual formatting might be necessary with .DOC.
- Most markup can be either Markdown syntax (e.g., "**" = bold), or LaTeX(e.g., "\textbf{}")
- Complete reference guide is available at: https://github.com/crsh/papaja/blob/master/vignettes/papaja_vignette.Rmd

References

Dahl, D. B. (2016). *Xtable: Export tables to latex or html*. Retrieved from https://CRAN.R-project.org/package=xtable

Epskamp, S., & Simon Stuber. (2017). *SemPlot: Path diagrams and visual analysis of various sem packages' output*. Retrieved from https://CRAN.R-project.org/package=semPlot

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Heidelberg: Springer-Verlag.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514.

Table 1

Model with displacement, horsepower, and weight

Predictor	b	95% CI	$t(28)$	p
Intercept	37.11	[32.78, 41.43]	17.58	$< .001$
Disp	0.00	[−0.02, 0.02]	-0.09	.929
Hp	-0.03	[−0.05, −0.01]	-2.72	.011
Wt	-3.80	[−5.98, −1.62]	-3.56	.001

Table 2

Model excluding displacement

Predictor	*b*	95% CI	*t*(29)	*p*
Intercept	37.23	[33.96, 40.50]	23.28	$< .001$
Hp	-0.03	[−0.05, −0.01]	-3.52	.001
Wt	-3.88	[−5.17, −2.58]	-6.13	$< .001$

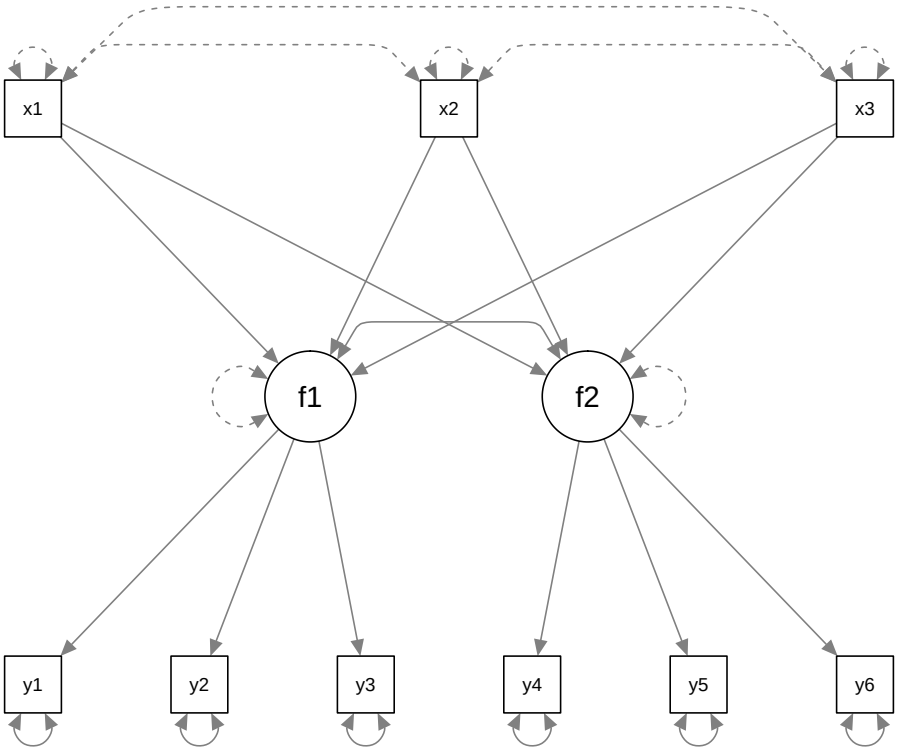

Figure 1

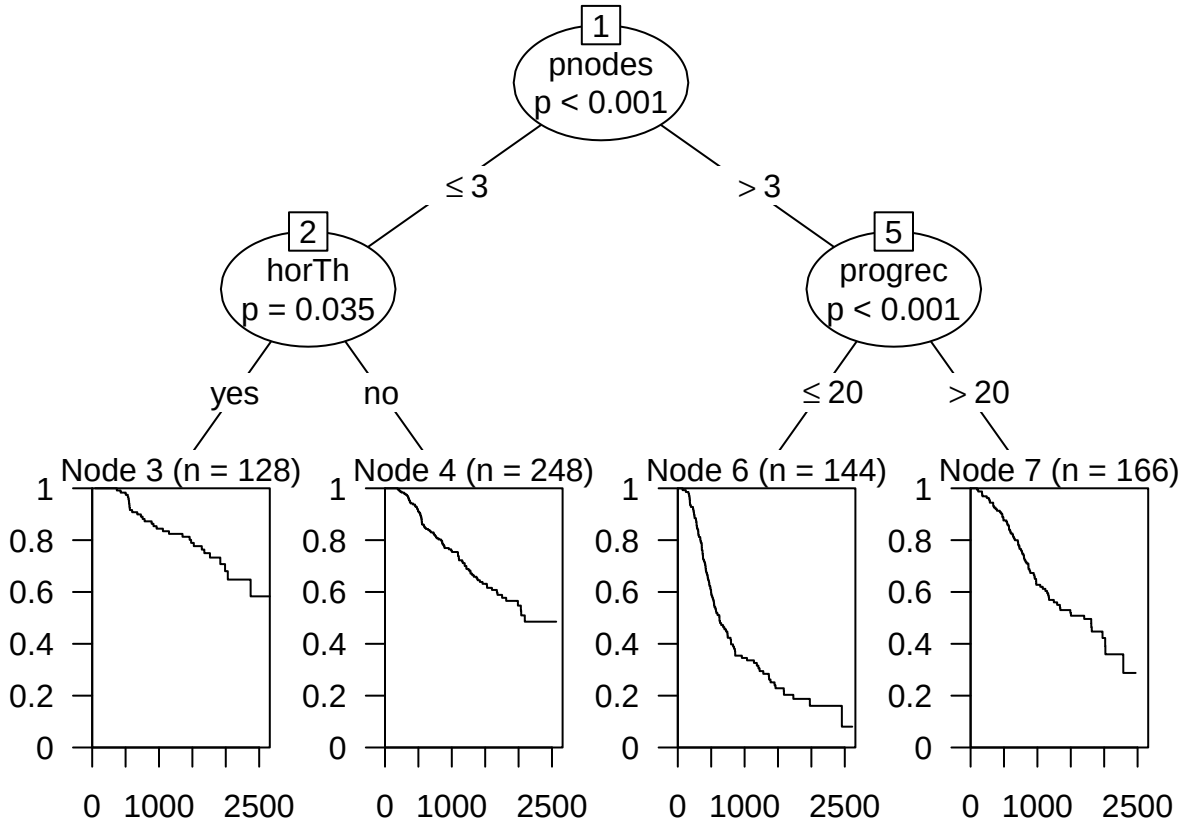
1
pnodes
p < 0.001
≤ 3
> 3
2
horTh
p = 0.035
5
progrec
p < 0.001
yes
no
≤ 20
> 20
Node 3 (n = 128)
Node 4 (n = 248)
Node 6 (n = 144)
Node 7 (n = 166)
1
0.8
0.6
0.4
0.2
0
0 1000 2500

Figure 2