Lecture 13: Preparing journal artical with *papaja*

Hu Chuan-Peng[1] & All Students[1,2]

[1] Nanjing Normal Unviersity

[2] Collaborators' Affiliations

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Author Note: This is for demonstration only.

The authors made the following contributions. Hu Chuan-Peng: Conceptualization, Writing - Original Draft Preparation, Supervision; All Students: Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Hu Chuan-Peng, #122 Ninghai Rd, Gulou District, Nanjing. E-mail: hcp4715@hotmail.com

Abstract

Psychological science has encountered a serious replication crisis.

To improve the credibility of the field, researchers actively reform the current practices and increase the opennes, transparency, and reproducibility of published. Using R language for data analyses is highly recommended. To increase computational reproduciblity of results, Aust & Barth developed *papaja*, which combines data analyses and manuscript preparation into a single RMD file.

This chapter aims to demonstrate how to use *papaja*.

We will introduce the package and key elements of the it.

After the lecture, we expected students able to create an example APA manuscript using open data or examplary data we had provided at the beginning of the class.

This demo and practice will further enhance the student's experience in computational reproducibility.

By spreading the ideas of reproducbility and teaching papaja, we aimed to improve the credibility of Psychological Science as a whole.

Keywords: Reproducibility, R, Teaching, Demonstration

Word count: X

Lecture 13: Preparing journal artical with *papaja*

R language has been widely used in psychological science and social science in general. However, for most students with a background of psychology, learning R is not easy.

In the past decade, an increasingly number of psychological researchers had switched from SPSS to R. The most common suggestions from those who successfully adopted R is: learning by doing.

To fully integrate R in the life-cycle of psychological research, using R for preparing journal article is necessary.

Fortunately, we now have *papaja* (Aust & Barth, 2024), which enable us to prepare journal article that formatted as required by APA.

Below we will demonstrate how to use *papaja* to generate an APA formatted PDF. Note that we will use the data from perceptual matching task, which was from (Hu, Lan, Macrae, & Sui, 2020). The data analysis for this dataset has been showed in previous lectures.

The prerequisite knowledge of using *papaja* include: (1) R language; (2) R Markdown, which share grammar with Latex for formulas, see R Markdonw Cookbook; (3) APA style.

Here are some examples of common R Markdown syntax:

- `‘#‘: header`
- `‘##‘: sub-header`
- `‘###‘: sub-sub-header`
- `‘*text*‘: italic text`
- `‘**text**‘: bold text`
- `‘~~text~~‘: strikethrough text`

- `[link](https://example.com)`: hyperlink
- `![caption](image.png)`: image with caption
- `$math$`: inline math (e.g., $y_i = \alpha + \beta X_i$)
- `$$math$$`: block math
- font color: Roses are \textcolor{red}{red}, violets are \textcolor{blue}{blue}.

One key advantage of using *papaja* is that you can use variable names instead of a specific values in the text. Which means you do not need to update the text when your input (and thus the value) updated.

Now, let do this using a simple example.

We sampled 150 number from a normal distribution, and we found that mean of the these data are: M = -0.06, SD = 2.01.

Here we will try to reproduce what has been reported in Hu et al. (2020). Thus, we adopted the code from: https://github.com/hcp4715/moralSelf_ddm.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

The sample size of the study was determined in a dynamic way (Schoenbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). Specifically, we kept collecting data and analyzing the strength of evidence for the critical hypothesis, including the interaction between Self-Relevance × Valence on RT data and two Bayes factor paired t-tests (good-self vs. bad-self, good-self vs. good-other). We stopped recruiting new participants

when both paired t-tests reached $BF_{10} \leq 0.1$ or $BF_{10} \geq 10$. Participants who were already recruited at that moment continued to complete the experiment. See https://osf.io/w6hrj/ for the change of Bayes factor during the data collection. In total, 44 college students (25 females, age: 20.91 ± 2.58) were recruited. Two participants were excluded from data analysis because of procedural failures, leaving 42 participants (24 females, age: 20.71 ± 2.38).

Material and Procedure

The data was collected using the same settings as described in the pilot study, with several differences:

- In the shape-categorization task, the shapes were presented for 100 ms, instead of 200 ms in the pilot experiment, and feedback was Chinese character 'Correct' or 'Incorrect', instead of happy or sad symbolic faces.
- There were only two different types of blocks in the categorization task in the confirmatory study because the importance judgments resulted in unbalanced trials between participants.
- There were more trials per condition: 72 experimental trials for the matching task.
- The questionnaires were different from the pilot study.

Data analysis

We used R (Version 4.4.3; R Core Team, 2024) and the R-packages *afex* (Version 1.4.1; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *emmeans* (Version 1.11.1; Lenth, 2024), *ggplot2* (Version 3.5.2; Wickham, 2016), *patchwork* (Version 1.3.0; Pedersen, 2024) and *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024) for all our analyses. The results from Frequentist hypothesis testing (i.e., ANOVA and t-tests) will be reported below.

Results

Sensitivity (*d'*)

A two-way repeated-measure ANOVA revealed a significant interaction between Identity and Valence. $F(1, 41) = 12.06$, $p = 0.0012$. We conducted follow-up simple main effect analyses. When the shapes were self-referential, *d'* was significantly larger for moral condition (2.5 ± 0.11) than for immoral condition (1.99 ± 0.1), $t(41) = 4.29$, $p < 0.001$. However, this was no clear evidence that moral condition (2.11 ± 0.116) and immoral condition (2.27 ± 0.107) were not significantly different when the shapes are other-referential (see figure 1).

```
## Warning: More than one observation per design cell, aggregating data using ‘fun_aggre
## To turn off this warning, pass ‘fun_aggregate = mean‘ explicitly.

## Anova Table (Type 3 tests)
##
## Response: RT_mean
##              Effect    df     MSE         F  ges p.value
## 1          Identity 1, 41 1181.37      1.06 .002    .309
## 2           Valence 1, 41  903.07 39.43 *** .051   <.001
## 3 Identity:Valence 1, 41 1156.92 13.97 *** .024   <.001
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’+’ 0.1 ’ ’ 1

## Identity = Self:
##  Valence emmean    SE df lower.CL upper.CL
##  Good       685  8.92 41      667      703
##  Bad        733 10.20 41      713      754
```

```
##
## Identity = Other:
##  Valence emmean    SE df lower.CL upper.CL
##  Good       710 10.70 41      688      731
##  Bad        719  9.53 41      700      738
##
## Confidence level used: 0.95

## Identity = Self:
##  contrast   estimate   SE df t.ratio p.value
##  Good - Bad    -48.7 6.92 41  -7.037  <.0001
##
## Identity = Other:
##  contrast   estimate   SE df t.ratio p.value
##  Good - Bad     -9.5 7.08 41  -1.342  0.1869
```

Reaction times

A two-way repeated-measure ANOVA revealed a significant interaction between Identity and Valence. $F(1, 41) = 13.97$, $p < 0.001$.We conducted follow-up simple main effect analyses. When the shapes were self-referential, *d'* was significantly larger for moral condition (685 $\pm$ 8.92) than for immoral condition (733 $\pm$ 10.20), $t(41) = -7.04$, $p < 0.001$. However, this was no clear evidence that moral condition (710 $\pm$ 10.70) and immoral condition (719 $\pm$ 9.53) were not significantly different when the shapes are other-referential.

Discussion

In this demonstration, we examplified how to use *papaja* to prepare an APA style manuscript. We highlighted the following details:

- Installation and start a manuscript from the template.
- YAML front
- Mathematical notations
- Citation
- Insert results from code block
- Plot

References

Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://doi.org/10.32614/CRAN.package.papaja

Hu, C.-P., Lan, Y., Macrae, C. N., & Sui, J. (2020). Good Me Bad Me: Does Valence Influence Self-Prioritization During Perceptual Decision-Making? *Collabra-Psychology*, *6*(1), 20. https://doi.org/10.1525/collabra.301

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from https://CRAN.R-project.org/package=patchwork

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schoenbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2024). *Afex: Analysis of factorial experiments*. Retrieved from https://CRAN.R-project.org/package=afex

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from https://CRAN.R-project.org/package=tidyr

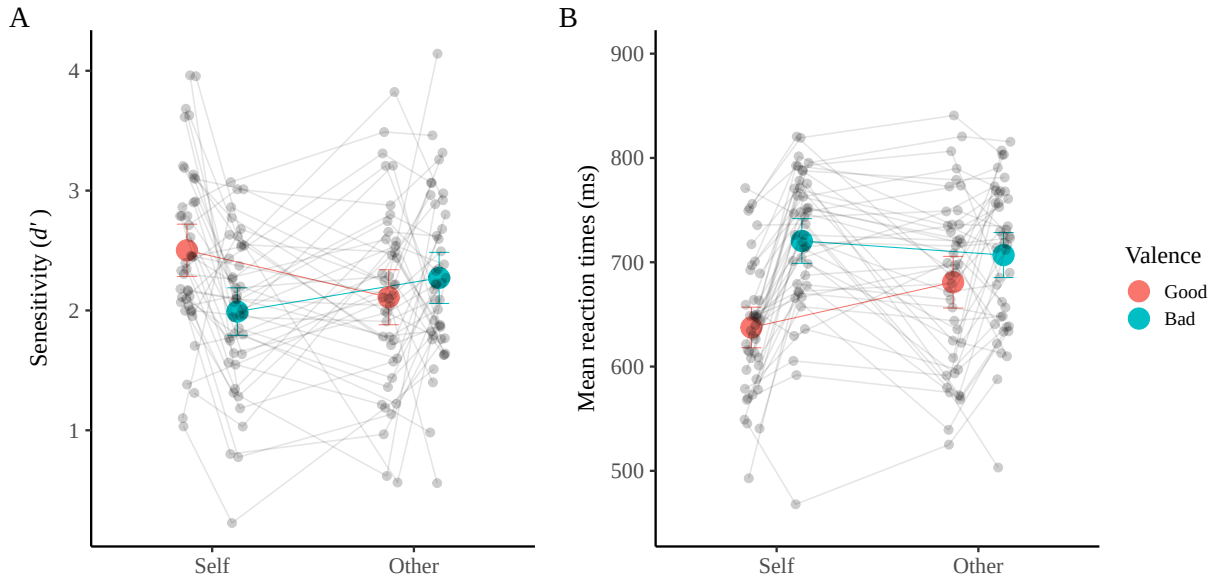

Figure 1. Interaction between identity and valence for (A) sensitivity and (B) reaction times.