Using R Markdown to Generate Report in ASA Format

Anonymous[1], Banana Slug[2]

Department of Statistical Science, UC Santa Cruz[1]

College 9/10 Dining Hall[2]

Abstract

This is a sample using R Markdown to generate reports in the form of the two-column ASA template. Latex is nice, but you still will have to keep track of all figures and tables yourself. With R Markdown, you will be able to compile your entire exam paper with one button (or two, ctrl+b, if you use Sublime Text). This saved me valuable time during numerous takehome exams during my first year. This template can be download from **my github**.

KEY WORDS: R Markdown, ASA template

1. Introduction

1.1 Use R Markdown with ASA template

The following steps is modified based on Chunyi's guide. Thanks Chunyi!

Step 1: acquire the template

Download the asaproc.cls file and prefix.tex file into your working directory. asaproc.cls file is for the latex ASA format, and prefix.tex file contains title and author informtaion. You need to modify it with your own title and name. Just open prefix.tex file in RStudio and replace the content in title and author.

Step 2: prepare the .rmd file

Download the asa_template.Rmd file. Keep everything above the *maketitle* and you can modify things below it to your own content. This template provide some examples about include graphs, tables, equations and reference. When you write your own content, just follow those examples.

Step 3: knit the .rmd file and you're done!

2. Some Easy Examples

2.1 Problem 1-2

The plot is shown in Figure 1.

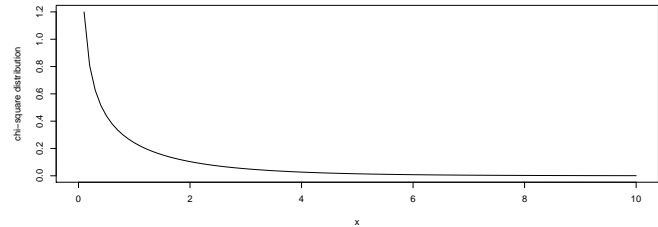The result shows the distribution of chi-square distribution.

Figure 1: Plot1

A funny story: Firstly I use this code to generate figure with caption and label but it just does not work. The solution turned out to be a little stupid but funny. The R Markdown compiler does not show captions for two consecutive figures or for a figure not separated from the text in a new paragraph. Therefore, just **use enough space** by pressing "Enter", and it will works! How amazing is coding!

2.2 Problem 1-3

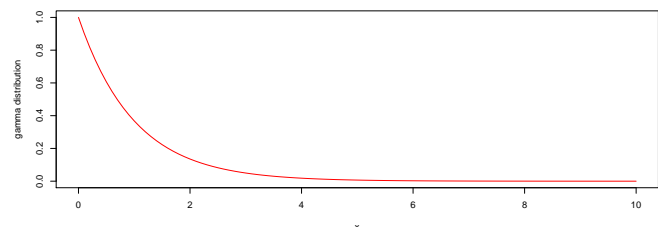Firstly, display curve of gamma with shape 1 and rate 1 in Figure 2.

Figure 2: Plot2

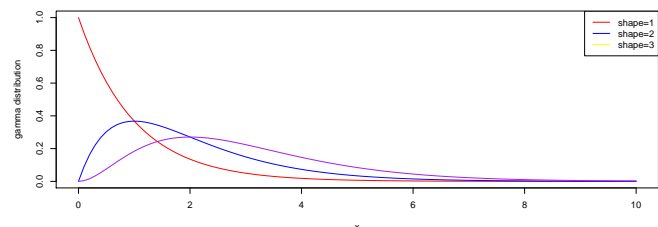Then, we add gamma distribution with shape parameters equal to 2 and 3, and shows in Figure 3.

Figure 3: Plot3

Table 1: President and opponent height

winner	opponent	time
185	175	2008
182	193	2004
182	185	2000

2.3 Problem 1-6

This example is used to generate table with caption and label, which is shown in Table 1.

3. Implementation Details

3.1 Chunk Options

This template frees you from switch between RStudio and your latex IDE by writing your report basically the same way as you do in latex in the main stream, and include your R code for computation, generating graphs and tables in chunks. The following R chunks are usually useful based on my experience. You can find more about chunck options from the **Definitive Guide**. (You can click the bold text and it should open your default browser and take you to the guide. This is also an example of including web links in your e-version report.)

- "cache": usually set to "TRUE". If caching is enabled, the same code chunk will not be evaluated the next time the document is compiled. It saves a lot of time.

- "eval": sometimes set to "FALSE" to code you have used but does not generate any tables or figures in your report.

- "results": sometimes set to "hide" to hide the output of the code. Useful when run MCMC. You may need to keep track of the loop, but you do not wnat to show it in your final report.

3.2 Reference

Some examples of reference is given here. The idea is to use a self-defined latex command \labelText (in the prefix.tex file if you are interested) and \nameref to refer, such as [1,2,4]. You can also refer to some text using this method, just replacing the number after \labelText with the text you want to refer.

3.3 Equations

I give some examples of writing equations here. Hopefully it covers the commonly used types of equations.

- Simplest form:

$$\mathcal{M}_1 : y_{ij} = \mu + \delta_i + \epsilon_{ij} \tag{1}$$

- Equation with matrix and bold greeks:

$$\boldsymbol{\omega}_t \sim N(\mathbf{0}, \begin{pmatrix} diag(\omega_{t1}, \omega_{t2}) & \mathbf{0}_{2 \times p} \\ \mathbf{0}_{p \times 2} & \mathbf{0}_{p \times p} \end{pmatrix})) \tag{2}$$

- Hierarchical model:

$$y_{kt} = \gamma_0 + \gamma_1(5 + t) + \epsilon_{kt}, k = 1, \cdots, 20, t = 1, \cdots, 5$$
$$\epsilon_{kt}|\sigma^2 \overset{i.i.d.}{\sim} N(0, \sigma^2)$$
$$\pi(\boldsymbol{\gamma}, \sigma^2) = p(\boldsymbol{\gamma}|\sigma^2)\pi(\sigma^2) \tag{3}$$

- Equations aligned:

$$\boldsymbol{\mu}_\gamma^* = (\mathbf{V}_\gamma^{-1} + X^T X)^{-1}(\mathbf{V}_\gamma^{-1}\boldsymbol{\mu}_\gamma + X^T \mathbf{y})$$
$$\mathbf{V}_\gamma^* = ((\mathbf{V}_\gamma)^{-1} + X^T X)^{-1}$$
$$a_{\sigma^2}^* = a_{\sigma^2} + \frac{n}{2}$$
$$b_{\sigma^2}^* = b_{\sigma^2} + \frac{1}{2}[\boldsymbol{\mu}_\gamma^T \mathbf{V}_\gamma^{-1}\boldsymbol{\mu}_\gamma + \mathbf{y}^T \mathbf{y} - (\boldsymbol{\mu}_\gamma^*)^T (\mathbf{V}_\gamma^*)^{-1}\boldsymbol{\mu}_\gamma^*] \tag{4}$$

- Equations aligned with a left bracket:

$$\begin{cases} \alpha_i = a_0 + a_i & i = 1, 2, 3 \\ \beta_i = b_0 + b_i & i = 1, 2, 3 \end{cases} \tag{5}$$

- Long equations:

$$p(\boldsymbol{\gamma}, \sigma^2|data) \propto p(\mathbf{y}|\boldsymbol{\gamma}, \sigma^2)p(\boldsymbol{\gamma}|\sigma^2)\pi(\sigma^2)$$
$$\propto (\frac{1}{\sigma^2})^{n/2}\exp\{-\frac{(\mathbf{y} - X\boldsymbol{\gamma})^T(\mathbf{y} - X\boldsymbol{\gamma})}{2\sigma^2}\}$$
$$\times (\frac{1}{\sigma^2})^{p/2}\exp\{-\frac{(\boldsymbol{\gamma} - \boldsymbol{\mu}_\gamma)^T\mathbf{V}_\gamma^{-1}(\boldsymbol{\gamma} - \boldsymbol{\mu}_\gamma)}{2\sigma^2}\}$$
$$\times (\frac{1}{\sigma^2})^{a_{\sigma^2}+1}\exp(-b_{\sigma^2}/\sigma^2) \tag{6}$$
$$\propto (\frac{1}{\sigma^2})^{a_{\sigma^2}^*+1}\exp(-\frac{b_{\sigma^2}^*}{\sigma^2})$$
$$\times (\frac{1}{\sigma^2})^{p/2}\exp\{-\frac{(\boldsymbol{\gamma} - \boldsymbol{\mu}_\gamma^*)^T(\mathbf{V}_\gamma^*)^{-1}(\boldsymbol{\gamma} - \boldsymbol{\mu}_\gamma^*)}{2\sigma^2}\}$$
$$\propto p(\sigma^2|data)p(\boldsymbol{\gamma}|\sigma^2, data)$$

To refer to a equation, you first need to label it with something like *eq:foo* (replace foo with your own name). Then the refering is down with latex command \ref, such as (1), (4).

3.4 Tables

3.4.1 Using Latex Code

You can use latex code to generate tables, like the following Table 2.

Table 2: Summary of inference result for model $\mathcal{M}_2$ and $\mathcal{M}_3$.

	Dependent variable:	
	Height	
	$\mathcal{M}_2$	$\mathcal{M}_3$
a_2	1.674	4.051***
	(−6.670, 10.018)	(2.586, 5.517)
a_3	1.823	9.651***
	(−6.521, 10.167)	(8.186, 11.117)
b_0	5.270***	5.716***
	(4.516, 6.024)	(5.300, 6.133)
b_2	0.297	
	(−0.730, 1.324)	
b_3	0.979*	
	(−0.049, 2.006)	
a_0	81.300***	77.728***
	(75.177, 87.423)	(74.226, 81.230)
R^2	0.907	0.903
Adjusted R^2	0.902	0.900
$\hat{\sigma}$	2.979 (df = 94)	3.006 (df = 96)
Note:	*p<0.1; **p<0.05; ***p<0.01	

The strength is you can use greek letters and subscript relatively easy in your table, but it will be really tedious if you have a large table. If you have a large table, I suggest use the following two ways. If you need to refer to the table, use \ref with something like *tab:foo*

3.4.2 Use Functions from Package knitr for General Tables

The R package knitr have a function *kable* that can generate tables directly. Table 3 is an example. Notice here to refer to the table, you need to use something like *tab:tabfoo* in \ref.

The default setting will have a empty line after every other five lines in the table. If you do not like it, see the code that generate Table 4.

Table 3: President and opponent height

winner	opponent	time
185	175	2008
182	193	2004
182	185	2000
188	187	1996
188	188	1992
188	173	1988

Table 4: President and opponent height

winner	opponent	time
185	175	2008
182	193	2004
182	185	2000
188	187	1996
188	188	1992
188	173	1988

The strength of this method is you do not need to input numbers by hand. Imagine how much time it can save for you if you have a very large table. However, if your table have complicated column names, like subscript or greek letters, this method may not be appealing. You can open the correspond .tex file and modified there, but do remember do this in the vary last step because if you knit the .Rmd file again, it will generate a new .tex file that replace the one you just revised!

3.4.3 Use Package stargazer for Linear Regression Summary Table

Table 5: Summary Statistics of Model 2 and Model 3

	Dependent variable:	
	y	
	(1)	(2)
X1	1.43	1.73
	(−1.65, 4.51)	(−1.45, 4.91)
X2		1.39
		(−1.79, 4.57)
Constant	3.57***	3.27***
	(2.28, 4.86)	(1.80, 4.74)
R^2	0.05	0.10
Note:	*p<0.1; **p<0.05; ***p<0.01	

The R package stargazer provide very fancy summary table for linear regression models, just like Table 2. Check the code that provide Table 5 as an example.

Another example is given in Table 6. You can check out the code that provide this table in the .Rmd file.

Table 6: Summary Statistics of Model 1

|	*Dependent variable:*
	y
mother2	-0.3333333
	$(-4.7353810, 4.0687150)$
mother3	-1.0000000
	$(-5.4020480, 3.4020480)$
mother4	-2.0000000
	$(-6.4020480, 2.4020480)$
mother5	-1.5000000
	$(-6.4216390, 3.4216390)$
mother6	-3.5000000
	$(-8.4216390, 1.4216390)$
mother7	0.0000000
	$(-6.2254360, 6.2254360)$
Constant	5.0000000^{**}
	$(1.8872820, 8.1127180)$
Observations	17
R^2	0.2156504
Adjusted R^2	-0.2549593
Residual Std. Error	2.7507580 (df = 10)
F Statistic	0.4582362 (df = 6; 10)
Note:	*p$<$0.1; **p$<$0.05; ***p$<$0.01

3.5 Figures

- **Simple Figures**
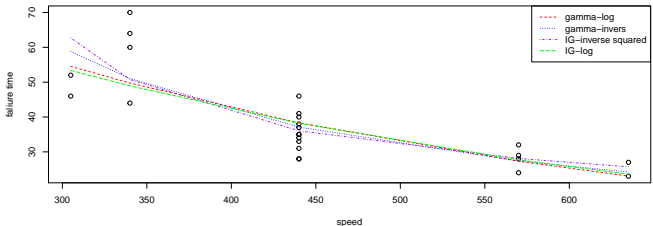
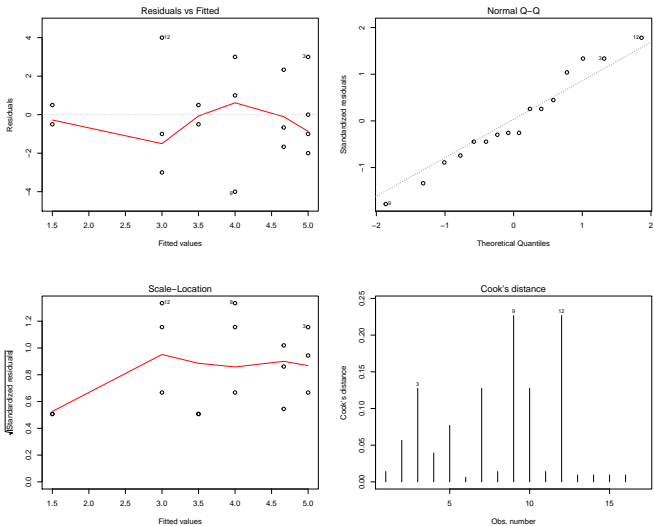Figure 4: GLM EDA For Choose Link Function.

- **Arrange Figures**

Figure 5: Residual plot of the linear regression model.

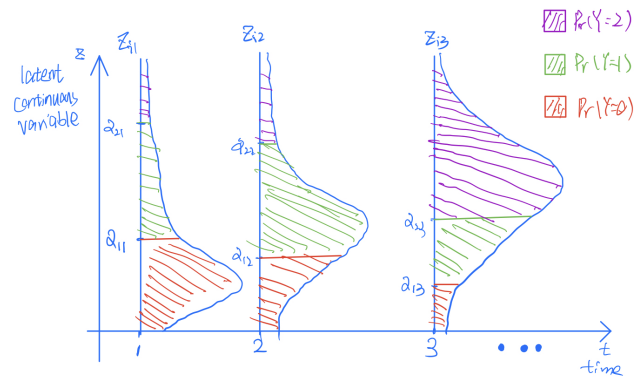- **Include Figures that are not generated by R**

Figure 6: Illustrate modeling ideas for latent continuous variable Z.

To refer to a figure, use $\ref\{fig:foo\}$ such as Figure 4, Figure 6.

REFERENCES

1. Gelman, Andrew, Carlin, John B., Stern, Hal S. and Rubin, Donald B. (2014), "Bayesian Data Analysis (3rd Editon)", *Chapman and Hall/CRC*, ISBN 978-1-4398-9820-8.
2. Sansó, Bruno (2020), "Lecture Notes on Intermediate Bayesian Statistical Modeling", available at UCSC course website.
3. Goldstein, H. (1979), "The Design and Analysis of Longitudinal Studies", London: Academic Press.
4. Verbeke, G., and Molenberghs, G. (2000), "Linear Mixed Models for Longitudinal Data", New York: Springer-Verlag.
5. Albert, J. (2009), "Bayesian Computation with R (2nd Edition)", *Springer*, ISBN 978-0-387-92297-3.
6. Faraway, J. (2014). "Linear Models with R (2nd Edition)", *Chapman and Hall/CRC*, ISBN 978-1-4398-8734-9.

7. Hoff, P. (2009), "A First Course in Bayesian Statistical Methods", *Springer Texts in Statistics*, ISBN 978-0-387-92299-7.

8. Robert, Christian P. (2007), "The Bayesian Choice (2nd Edition)", *Springer Texts in Statistics*, ISBN 978-0-387-71598-8.

9. Gelfand, Alan and Ghosh, Sujit (1998), "Model choice: A minimum posterior predictive loss approach", *Biometrika, Volume 85, Issue 1*, Pages 1–11.

10. Lund, Robert (2020), "Lecture Notes on Linear Statistical Models", available at UCSC Canvas.

A. Include R code in Appendix

Just set the chunk option *echo* to **TRUE** and *eval* to **FALSE**. See below.

```r
## fit linear models
lm1=lm(height~group,data=dat)
L1=lm(height ~ group*age,data=dat)
lm2=lm(height~group+age,data=dat)

## plot residual
par(mfrow=c(1,2))
plot(lm2,which=1:2)

## test residual
n=nrow(dat)
sample.res.corr=function(h){
  numerator=sum(lm2$residuals[1:(n-h)]
                *lm2$residuals[(h+1):n])
  denominator=sum(lm2$residuals^2)
  return(numerator/denominator)
}

Q=sapply(seq(1,5,by=1), sample.res.corr)
Q=n*sum(Q^2)

res.vs.norm=qqnorm(lm2$residuals,
                   plot.it = FALSE)
(cor(res.vs.norm$x,res.vs.norm$y))^2
```