

Lab 01 – MATH 240 – Computational Statistics

Brendan Mariano
Affiliation
Mathematics
bmariano@colgate.edu

Abstract

This document provides a basic template for the 2-page labs we will complete each week. Here, you should provide a succinct summary about what you did and why it might be helpful.”

Keywords: What topics does the lab cover with respect to class?

1 Instructions

For this lab, you will

1. Install **R** and **RStudio**
2. Install tinytext (if necessary):
`install.packages("tinytext")`
3. Create a GitHub account [here](#), and email me your username.
4. Install **GitHub desktop**
5. Accept the LAB 1 assignment [here](#)
6. Recreate this document (except put your name/info at the top) to get used to writing in L^AT_EX and to see the types of things we can do when creating a document to convey statistical information. Make sure to commit and push your work using GitHub desktop as you finish each section.

Remark: You will find the class Sweave cheatsheet to be *incredibly* (`\emph{incredibly}`) helpful.

2 Word Processing Tasks

2.1 Centering Text

We can center text in Sweave.

2.2 Bold, Italics and Underlining

We can **bold**, *italicize*, underline, and *emphasize* text in Sweave.

Note, I did a column break here so that the list wasn't broken across columns

2.3 Lists, and Numbered Lists

We can write an unordered list in Sweave.

- first item
- second item
- third item

We can write a numbered list in Sweave.

1. first item
2. second item
3. third item

We can write a lettered list in Sweave.

- a. first item
- b. second item
- c. third item

2.4 Submissions

This part of the midterm is due Sunday November 14 by 5p. I will not accept late submissions. Note that you may use this template to help build your introduction and methods sections, and you can use the work you did as a group during the datathon. Still, I expect this submission to be your own summary and extension of that work without collaboration.

2.5 Typing Mathematical Equations

We can write a one line equation that is centered like this

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \epsilon_i.$$

This can be written in the text, as $\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \epsilon_i$.

When we need to show multiple steps, we can create a multi-line equation that is centered like this:

$$\begin{aligned} 8(x-5) + x &= 9(x-5) + 5 \\ 8x - 40 + x &= 9x - 45 + 5 && \text{(Distributing)} \\ 9x - 40 &= 9x - 40 && \text{(Combining like terms)} \\ 9x &= 9x && \text{(Adding 40 to both sides)} \\ x &= x && \text{(Dividing both sides by 9)} \end{aligned}$$

The equality holds for any x.

Note, I did a page break here so that the next section started on a clean page.

2.6 Running R Code

Code chunks can be entered into Sweave; e.g., here are some comments.

```
# R code goes here
# Output is automatically printed in the pdf
```

Below, you can see that we can do algebra with R.

```
8*(9-5) + 9 # 8(x-5) + x for x = 9
## [1] 41
```

Below, you can see that we can do algebra with R.

```
8*(9-5) + 9 # 8(x-5) + x for x = 9
```

Alternatively, we can produced the output without the code

```
## [1] 41
```

We can also call object values from R directly.

```
result <- 8*(9-5) + 9 # 8*(9-5) + 9 # 8(x-5) + x for x = 9
result.with.error <- result + rnorm(1, mean = 0, sd = 0.1)
result.with.error
## [1] 41.26796
```

The result is 40.920042. Note that I did not type the result, but I used the `\Sexpr{}` command.

2.7 Plotting

We can also plot with R.

```
#Plot a histogram of random exponential data
hist(rexp(100))
```

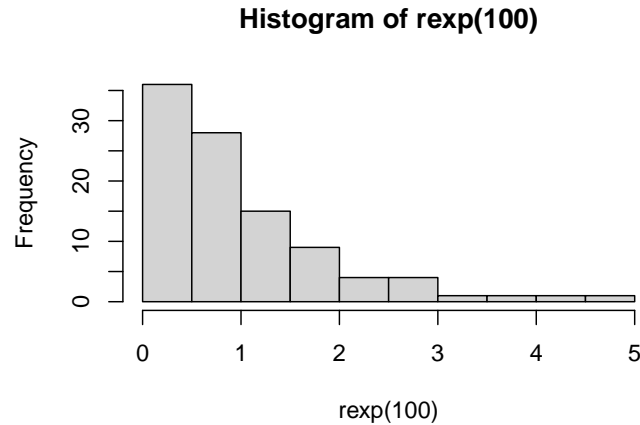


Figure 1: A histogram of random normally distributed data, $n = 100$.

2.8 Tables

Below, we load and take a peek at some data about the death rates per 1000 in Virginia in 1940 (Molyneaux et al., 1947).

```
data(VADeaths)
head(VADeaths) # Take a peek of the data
```

```
##      Rural Male Rural Female Urban Male Urban Female
## 50-54      11.7       8.7      15.4       8.4
## 55-59      18.1      11.7      24.3      13.6
## 60-64      26.9      20.3      37.0      19.3
## 65-69      41.0      30.9      54.6      35.1
## 70-74      66.0      54.3      71.1      50.0
```

If we want to print this nicely, we can do so using the `xtable` package (Dahl et al., 2019), which we can reference using the label (Table 1).

```
library(xtable)
sleep.table<-xtable(VADeaths,
                    label = "VADeaths.tab", caption = "Death Rates per 1000 in Vi
```

| Rural Male | Rural Female | Urban Male | Urban Female |
|------------|--------------|------------|--------------|
| 11.70 | 8.70 | 15.40 | 8.40 |
| 18.10 | 11.70 | 24.30 | 13.60 |
| 26.90 | 20.30 | 37.00 | 19.30 |
| 41.00 | 30.90 | 54.60 | 35.10 |
| 66.00 | 54.30 | 71.10 | 50.00 |

Table 1: Death Rates per 1000 in Virginia (1940).

References

- Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.
- Horst, A. M., Hill, A. P., and Gorman, K. B. (2020). *palmer-penguins: Palmer Archipelago (Antarctica) penguin data*. R package version 0.1.0.
- Molyneaux, L., Gilliam, S. K., and Florant, L. C. (1947). Differences in virginia death rates by color, sex, age and rural or urban residence. *American Sociological Review*, 12(5):525–535.

3 Appendix

The palmerpenguins package for R (Horst et al., 2020) provides data on adult foraging penguins near Palmer Station, Antarctica. Figure 2 is too big to fit nicely in our column-based-template above, so I've placed it here in the abstract by saving it and presenting it scaled to 0.75.

```
library(palmerpenguins)
pdf("figure/penguins.pdf", width = 8, height = 5)
plot(penguins)
dev.off()
```

```
## pdf
## 2
```

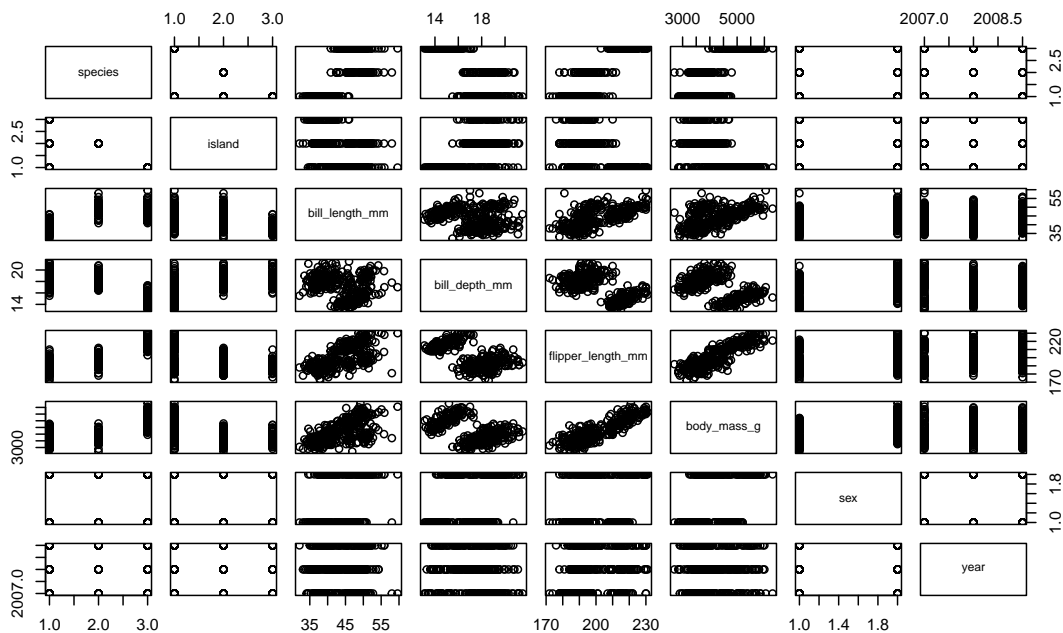


Figure 2: Figure 2: Data on adult foraging penguins near Palmer Station, Antarctica.