STA 602 Lab 1

Yicheng Shen

05 September, 2022

Here's a list of resources for learning rmarkdown:

1. R markdown introduction
2. R for Data Science
3. Introduction to R markdown
4. Another introduction to R markdown
5. R markdown cheatshett
6. Data visualization with ggplot2 cheat sheet
7. Learn LaTex in 30 minutes

Exercise 1

1. Create a code chunk and set the header parameter to TRUE and print out the top rows of the table with `head()` as above.

```
data <- read.table(file = url("http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/azdiabetes.dat"),
                   header = TRUE)
head(data)
```

```
##   npreg glu bp skin  bmi   ped age diabetes
## 1     5  86 68   28 30.2 0.364  24       No
## 2     7 195 70   33 25.1 0.163  55      Yes
## 3     5  77 82   41 35.8 0.156  35       No
## 4     0 165 76   43 47.9 0.259  26       No
## 5     0 107 60   25 26.4 0.133  23       No
## 6     5  97 76   27 35.6 0.378  52      Yes
```

Exercise 2

2. Generate a sequence of 100 equispaced real numbers from 0 to 1 and store it in a variable called `seq2`.

```
seq2 <- seq(from = 0, to = 1, by = 1 / 99)  # If use 1/100 would give 101 numbers, so 1/99
length(seq2)
```

```
## [1] 100
```

Exercise 3

3. Sort the entries in `seq3` from greatest to least.

```
seq3 <- seq(from = -3, to = 3, by = .5)
sort(seq3, decreasing = T)
```

```
##  [1]  3.0  2.5  2.0  1.5  1.0  0.5  0.0 -0.5 -1.0 -1.5 -2.0 -2.5 -3.0
```

Exercise 4

4. Find the variance of each row of `mat5`

```
mat5 <- matrix(seq(1, 100, 1), nrow = 4, ncol = 25, byrow = T)
apply(X = mat5, MARGIN = 1, FUN = var)
```

```
## [1] 54.16667 54.16667 54.16667 54.16667
```

Exercise 5

5. Generate 500 samples from a Beta distribution with shape parameter $[a, b] = [0.5, 0.5]$ and store the samples in a variable called `W`

```
set.seed(32507)
W <- rbeta(500, 0.5, 0.5)
head(W) # output is big hence not shown completely
```

```
## [1] 0.579297945 0.094230524 0.005047281 0.650463616 0.343816786 0.999999462
```

Exercise 6

6. Browse online resources (some below), or use code from above to make a few plots of your own.

```
set.seed(253)
norm_samples <- rnorm(10000)

a <- norm_samples %>%
  data.frame(x = .) %>%
  mutate(`x^2` = x^2, `x^3` = x^3, `abs(x)` = abs(x)) %>%
  ggplot2::ggplot() +
  geom_point(aes(x,`x^2`),color = "blue") +
  geom_point(aes(x,`abs(x)`),color = "red") +
  geom_point(aes(x,`x^3`),color = "orange") +
  labs(title = "Transformation of Normal Variable", y = "") + theme_bw()

b <- norm_samples %>%
  data.frame(x = .) %>%
  mutate(`Cumulative Sums` = cumsum(x), time = 1:10000) %>%
  ggplot2::ggplot() + geom_line(aes(time, `Cumulative Sums`)) +
```

```
  labs(title = "Random Time Series Plot") + theme_bw()

set.seed(466)
norm_samples_2 <- rnorm(10000)
poission_samples_2 <- rpois(10000, 10)

c <- data.frame(x = norm_samples, y = norm_samples_2, z = poission_samples_2) %>%
  ggplot2::ggplot() + geom_jitter(aes(x,y,color = z)) +
  labs(title = "Random Scatterplot") + theme_bw()

d <- bayesrules::plot_beta_binomial(alpha = 20, beta = 20, y = 40, n = 100) +
  labs(title = "Beta Binomial Model") + theme_bw()

grid.arrange(a, b, c, d, nrow = 2)
```