

Homework 3

Adeline Guthrie

September 18, 2019

Problem 3

Part a

```
#Store URL
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/Sensory.dat"

#Read in Data, skip first line with just "Operator", treat the second row as the header,
#and fill empty spaces with "NA"
sensory <- read.table(url, header = T, sep = " ", skip = 1, fill = T)

#Shift data values to the right and add the item number
c <- 0
for( i in 1:30){
  if(i%%3 != 1){
    sensory[i,] <- lag(sensory[i,])
    sensory[i,1] <- c
  }else{
    c <- c + 1
  }
}

#Give each observation its own row
#Rename columns and remove X's in front of operator numbers
sensory_tidy <- sensory %>%
  gather(X1, response, -Item) %>%
  transmute(Item = Item, Operator = parse_number(X1), Response = response) %>%
  group_by(Operator)
```

Warning: package 'bindrcpp' was built under R version 3.4.4

```
#Final Dataset:
kable(sensory_tidy)
```

Item	Operator	Response
1	1	4.3
1	1	4.3
1	1	4.1
2	1	6.0
2	1	4.9
2	1	6.0
3	1	2.4
3	1	3.9
3	1	1.9

Item	Operator	Response
4	1	7.4
4	1	7.1
4	1	6.4
5	1	5.7
5	1	5.8
5	1	5.8
6	1	2.2
6	1	3.0
6	1	2.1
7	1	1.2
7	1	1.3
7	1	0.9
8	1	4.2
8	1	3.0
8	1	4.8
9	1	8.0
9	1	9.0
9	1	8.9
10	1	5.0
10	1	5.4
10	1	2.8
1	2	4.9
1	2	4.5
1	2	5.3
2	2	5.3
2	2	6.3
2	2	5.9
3	2	2.5
3	2	3.0
3	2	3.9
4	2	8.2
4	2	7.9
4	2	7.1
5	2	6.3
5	2	5.7
5	2	6.0
6	2	2.4
6	2	1.8
6	2	3.3
7	2	1.5
7	2	2.4
7	2	3.1
8	2	4.8
8	2	4.5
8	2	4.8
9	2	8.6
9	2	7.7
9	2	9.2
10	2	4.8
10	2	5.0
10	2	5.2
1	3	3.3

Item	Operator	Response
1	3	4.0
1	3	3.4
2	3	4.5
2	3	4.2
2	3	4.7
3	3	2.3
3	3	2.8
3	3	2.6
4	3	6.4
4	3	5.9
4	3	6.9
5	3	5.4
5	3	5.4
5	3	6.1
6	3	1.7
6	3	2.1
6	3	1.1
7	3	1.2
7	3	0.8
7	3	1.1
8	3	4.5
8	3	4.7
8	3	4.7
9	3	9.0
9	3	6.7
9	3	8.1
10	3	3.9
10	3	3.4
10	3	4.1
1	4	5.3
1	4	5.5
1	4	5.7
2	4	5.9
2	4	5.5
2	4	6.3
3	4	3.1
3	4	2.7
3	4	4.6
4	4	6.8
4	4	7.3
4	4	7.0
5	4	6.1
5	4	6.2
5	4	7.0
6	4	3.4
6	4	4.0
6	4	3.3
7	4	0.9
7	4	1.2
7	4	1.9
8	4	4.6
8	4	4.9

Item	Operator	Response
8	4	4.8
9	4	9.4
9	4	9.0
9	4	9.1
10	4	5.5
10	4	4.9
10	4	3.9
1	5	4.4
1	5	3.3
1	5	4.7
2	5	4.7
2	5	4.9
2	5	4.6
3	5	2.4
3	5	1.3
3	5	2.2
4	5	6.0
4	5	6.1
4	5	6.7
5	5	5.9
5	5	6.5
5	5	4.9
6	5	1.7
6	5	1.7
6	5	2.1
7	5	0.7
7	5	1.3
7	5	1.6
8	5	3.2
8	5	4.6
8	5	4.3
9	5	8.8
9	5	7.9
9	5	7.6
10	5	3.8
10	5	4.6
10	5	5.5

```
#Summarize data
kable(summarise(sensory_tidy, Mean = mean(Response), Std_dev = sd(Response),
  Min = min(Response), Median = median(Response),
  Max = max(Response)))
```

Operator	Mean	Std_dev	Min	Median	Max
1	4.593333	2.239140	0.9	4.55	9.0
2	5.063333	2.045429	1.5	4.95	9.2
3	4.166667	2.098494	0.8	4.15	9.0
4	5.193333	2.132334	0.9	5.40	9.4
5	4.266667	2.143206	0.7	4.60	8.8

Part b

```
#Store URL
url2 <- "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat"

#Read in data, skip the first row since the labels are messy, fill empty spaces with "NA"
jumps <- read.table(url2, header = F, sep = " ", skip = 1, fill = T)

#Resize the original matrix to give each observation its own row
#Convert to tibble to use dplyr functions
#Remove rows of "NA"
#Name columns
#Sort by Year

jumps_tidy <- jumps %>%
  resize(nrow = 24, ncol = 2, byrow = T) %>%
  as_tibble(.name_repair = "unique") %>%
  na.omit() %>%
  rename(Year = "..1", Long_Jump = "..2") %>%
  arrange(Year)

## New names:
## * `` -> `..1`
## * `` -> `..2`

#Final Dataset:
kable(jumps_tidy)
```

Year	Long_Jump
-4	249.75
0	282.88
4	289.00
8	294.50
12	299.25
20	281.50
24	293.13
28	304.75
32	300.75
36	317.31
48	308.00
52	298.00
56	308.25
60	319.75
64	317.75
68	350.50
72	324.50
76	328.50
80	336.25
84	336.25
88	343.25
92	342.50

```
#Summarize data
kable(summarise(jumps_tidy, Mean = mean(Long_Jump), Std_dev = sd(Long_Jump),
               Min = min(Long_Jump), Median = median(Long_Jump),
               Max = max(Long_Jump)))
```

Mean	Std_dev	Min	Median	Max
310.2873	24.36121	249.75	308.125	350.5

Part c

```
#Store URL
url3 <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat"

#Read in data, skip the first row since the labels are messy, fill empty spaces with "NA"
weights <- read.table(url3, header = F, sep = " ", skip = 1, fill = T)

#Resize the original matrix to give each observation its own row
#Convert to tibble to use dplyr functions
#Name columns
weights_tidy <- weights %>%
  resize(nrow = 63, ncol = 2, byrow = T) %>%
  as_tibble(.name_repair = "unique") %>%
  rename(Body_Wt = "..1", Brain_Wt = "..2")

## New names:
## * `` -> `..1`
## * `` -> `..2`

#Final Dataset:
kable(weights_tidy)
```

Body_Wt	Brain_Wt
3.385	44.50
521.000	655.00
2.500	12.10
0.480	15.50
0.785	3.50
55.500	175.00
1.350	8.10
10.000	115.00
100.000	157.00
465.000	423.00
3.300	25.60
52.160	440.00
36.330	119.50
0.200	5.00
10.550	179.50
27.660	115.00

Body_Wt	Brain_Wt
1.410	17.50
0.550	2.40
14.830	98.20
529.000	680.00
60.000	81.00
1.040	5.50
207.000	406.00
3.600	21.00
4.190	58.00
85.000	325.00
4.288	39.20
0.425	6.40
0.750	12.30
0.280	1.90
0.101	4.00
62.000	1320.00
0.075	1.20
0.920	5.70
6654.000	5712.00
0.122	3.00
1.000	6.60
3.500	3.90
0.048	0.33
0.005	0.10
6.800	179.00
192.000	180.00
0.060	1.00
35.000	56.00
3.000	25.00
3.500	10.80
4.050	17.00
160.000	169.00
2.000	12.30
0.120	1.00
0.900	2.60
1.700	6.30
0.023	0.40
1.620	11.40
2547.000	4603.00
0.010	0.30
0.104	2.50
0.023	0.30
1.400	12.50
4.235	50.40
187.100	419.00
250.000	490.00
NA	NA

```
#Summarize data
kable(summary(weights_tidy))
```

Body_Wt	Brain_Wt
Min. : 0.005	Min. : 0.10
1st Qu.: 0.600	1st Qu.: 4.25
Median : 3.342	Median : 17.25
Mean : 198.790	Mean : 283.13
3rd Qu.: 48.203	3rd Qu.: 166.00
Max. :6654.000	Max. :5712.00
NA's :1	NA's :1

Part d

```
#Store URL
url4 <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat"

#Read in data, skip first row and use the second row as header, remove the
#comment character = "#"
tomatoes <- read.table(url4, header = T, sep = "", skip = 1,
                      comment.char = "")

#Split the data in each of the six cells by ","
I_X10000 <- do.call("rbind", strsplit(toString(tomatoes[1,1]), ","))
I_X20000 <- do.call("rbind", strsplit(toString(tomatoes[1,2]), ","))
I_X30000 <- do.call("rbind", strsplit(toString(tomatoes[1,3]), ","))
P_X10000 <- do.call("rbind", strsplit(toString(tomatoes[2,1]), ","))
P_X20000 <- do.call("rbind", strsplit(toString(tomatoes[2,2]), ","))
P_X30000 <- do.call("rbind", strsplit(toString(tomatoes[2,3]), ","))

#Combine the vectors of split data
I <- cbind(I_X10000, I_X20000, I_X30000)
P <- cbind(P_X10000, P_X20000, P_X30000)

tomatoes <- rbind(I,P)

#Convert to tibble to use dplyr and tidyr functions
#Create column with variety names (I = Ife\#1, P = PusaEarlyDwarf)
#Name columns
#Give each observation its own row
#Fix density labels and remove unnecessary column
#Sort by variety and then density
#Convert Response type to Numeric

tomatoes_tidy <- tomatoes %>%
  as_tibble(.name_repair = "unique") %>%
  mutate("Variety" = c("I", "P")) %>%
  rename("10000_1" = "..1", "10000_2" = "..2", "10000_3" = "..3",
        "20000_1" = "..4", "20000_2" = "..5", "20000_3" = "..6",
        "30000_1" = "..7", "30000_2" = "..8", "30000_3" = "..9") %>%
  gather(X1, Response, -Variety) %>%
  separate(X1, c("Density", "temp")) %>%
  select(Variety, Density, Response) %>%
```



```

arrange(Variety, Density) %>%
transform(Response = as.numeric(Response))

```

```

## New names:
## * `` -> `..1`
## * `` -> `..2`
## * `` -> `..3`
## * `` -> `..4`
## * `` -> `..5`
## * ... and 4 more

```

```

#Final Dataset:
kable(tomatoes_tidy)

```

Variety	Density	Response
I	10000	16.1
I	10000	15.3
I	10000	17.5
I	20000	16.6
I	20000	19.2
I	20000	18.5
I	30000	20.8
I	30000	18.0
I	30000	21.0
P	10000	8.1
P	10000	8.6
P	10000	10.1
P	20000	12.7
P	20000	13.7
P	20000	11.5
P	30000	14.4
P	30000	15.4
P	30000	13.7

```

#Summarize data by variety
kable(tomatoes_tidy %>%
  group_by(Variety) %>%
  summarise(Mean = mean(Response), Std_dev = sd(Response), Min = min(Response),
    Median = median(Response), Max = max(Response)))

```

Variety	Mean	Std_dev	Min	Median	Max
I	18.11111	1.985223	15.3	18.0	21.0
P	12.02222	2.603257	8.1	12.7	15.4

```

#Summarize data by density
kable(tomatoes_tidy %>%
  group_by(Density) %>%
  summarise(Mean = mean(Response), Std_dev = sd(Response), Min = min(Response),

```

```
Median = median(Response), Max = max(Response))
```

Density	Mean	Std_dev	Min	Median	Max
10000	12.61667	4.148453	8.1	12.70	17.5
20000	15.36667	3.189775	11.5	15.15	19.2
30000	17.21667	3.205256	13.7	16.70	21.0