

Homework 2

Hanao Li

February 26, 2019

Question 1

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.4
```

```
setwd("C:/Users/36576/Desktop/gr5243")
data <- fread('Homework 2 Data.csv')
dim(data)
```

```
## [1] 2426922    13
```

```
length(unique(data$id))
```

```
## [1] 10000
```

```
colnames(data)
```

```
## [1] "id"          "begin"       "end"         "age"
## [5] "diabetes"    "hypertension" "kidney_disease" "ace"
## [9] "beta.blocker" "statin"      "hospital"    "heart.attack"
## [13] "death"
```

There are 2426922 rows, 13 columns and 10000 unique patients. All of the column names match up with the description of the data set.

Question 2

```
#Correct Death Time
data[death == 1, begin:= as.integer(end - 1)]

###Check Begin
class(data$begin)
```

```
## [1] "integer"
```

```
#Begin < 0
data[begin < 0, .N]
```

```
## [1] 0
```

```
#Begin > End
data[begin >= end, .N]
```

```
## [1] 0
```

```
###Check End
class(data$end)
```

```
## [1] "integer"
```

```
#End < 0
data[end < 0, .N]
```

```
## [1] 0
```

```
###Check Unique Baseline
data[begin ==0, .(length(unique(id)))]
```

```
##      V1
## 1: 10000
```

Begin and End looks okay

```
###Check Age  
class(data$age)
```

```
## [1] "integer"
```

```
summary(data$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
## 29.00   59.00   66.00   65.51   72.00   100.00
```

Age looks okay

```
###Check diabetes  
class(data$diabetes)
```

```
## [1] "integer"
```

```
summary(data$diabetes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
## 0.0000   0.0000   0.0000   0.1943   0.0000   1.0000
```

Diabetes looks okay

```
###Check hypertension  
class(data$hypertension)
```

```
## [1] "integer"
```

```
summary(data$hypertension)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
## 0.0000   0.0000   1.0000   0.5983   1.0000   1.0000
```

Hypertension looks okay

```
###Check kidney_disease  
class(data$kidney_disease)
```

```
## [1] "integer"
```

```
summary(data$kidney_disease)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
## 0.0000   0.0000   0.0000   0.1012   0.0000   1.0000
```

Kidney disease looks okay

```
###check Ace  
class(data$ace)
```

```
## [1] "integer"
```

```
summary(data$ace)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
## 0.0000   0.0000   1.0000   0.7284   1.0000   1.0000
```

```
###Check BB  
class(data$beta.blocker)
```

```
## [1] "character"
```

```
table(data$beta.blocker)
```

```
##
##           0           1 didn't take      False      Filled      N
##      937036      1417185      4103      4116      6353      3934
##           no           No           nope Not filled Picked up      True
##      4067      4139      4035      4104      6386      6234
##      Utilized      yeah      yes      Yes
##      6330      6308      6270      6322
```

Beta blocker looks not okay. We will change "Filled, Picked up, True, Utilized, yeah, yes, Yes" to 1, "didn't take, False, N, no, No, nope, Not filled" to 0

```
data[data$beta.blocker == "didn't take", ]$beta.blocker <- "0"
data[data$beta.blocker == "False", ]$beta.blocker <- "0"
data[data$beta.blocker == "N", ]$beta.blocker <- "0"
data[data$beta.blocker == "no", ]$beta.blocker <- "0"
data[data$beta.blocker == "No", ]$beta.blocker <- "0"
data[data$beta.blocker == "nope", ]$beta.blocker <- "0"
data[data$beta.blocker == "Not filled", ]$beta.blocker <- "0"
data$beta.blocker <- as.integer(data$beta.blocker)
```

```
## Warning: NAs introduced by coercion
```

```
data[is.na(data$beta.blocker), ]$beta.blocker <- 1
summary(data$beta.blocker)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000  1.0000  0.6022  1.0000  1.0000
```

Beta blocker looks okay now

```
###Check Statin
class(data$statin)
```

```
## [1] "character"
```

```
table(data$statin)
```

```
##
##           0           1      appt Ask doctor BP Normal
##           60      535311      1891335      43      37      41
## diabetes high risk
##           48           47
```

Statin looks not okay. Since we don't know whether the patients take the medicine, we will change those non 0, 1s to NA

```
data$statin <- as.integer(data$statin)
```

```
## Warning: NAs introduced by coercion
```

```
summary(data$statin)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  1.0000  1.0000  0.7794  1.0000  1.0000    276
```

Statin looks okay now

```
###Check Hospital
class(data$hospital)
```

```
## [1] "integer"
```

```
summary(data$hospital)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.02045 0.00000 1.00000
```

Hospital looks okay

```
##Check Heart Attack
class(data$heart.attack)
```

```
## [1] "integer"
```

```
summary(data$heart.attack)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000000 0.000000 0.000000 0.001344 0.000000 1.000000
```

Heart Attack looks okay

```
##Check Death
class(data$death)
```

```
## [1] "integer"
```

```
summary(data$death)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000000 0.000000 0.000000 0.002508 0.000000 1.000000
```

Death looks okay

Question 3

```
library(DT)
```

```
## Warning: package 'DT' was built under R version 3.4.4
```

```
Summaries <- c("Age", "Diabetes", "Hypertension", "Kidney_Disease")
Average <- round(unlist(c(data[begin == 0, .(mean(age))], data[begin == 0, .(mean(diabetes))]*100, data[begin == 0, .(mean(hypertension))]*100, data[begin == 0, .(mean(kidney_disease))]*100]), 1)
Average <- c(Average[1], paste(Average[2:4], c("%")))
Std <- round(unlist(c(data[begin == 0, .(sd(age))], data[begin == 0, .(sd(diabetes))]*100, data[begin == 0, .(sd(hypertension))]*100, data[begin == 0, .(sd(kidney_disease))]*100]), 1)
Std <- c(Std[1], paste(Std[2:4], c("%")))
patientsummary <- data.table(Summaries, Average, Std)
datatable(patientsummary)
```

Show 10 entries

Search:

	Summaries	Average	Std
1	Age	65.5	10
2	Diabetes	19.4 %	39.6 %
3	Hypertension	60 %	49 %
4	Kidney_Disease	10.2 %	30.3 %

Showing 1 to 4 of 4 entries

Previous 1 Next

Question 4

```
#Heart Attack
data[heart.attack == 1, .N]
```

```
## [1] 3262
```

```
#Death
data[death == 1, .N]
```

```
## [1] 6086
```

```
#Hospital
data[, .(sum(diff(hospital) > 0))]
```

```
##      V1
## 1: 49636
```

```
#Number of Rows with Hospital
data[hospital == 1, .N]
```

```
## [1] 49642
```

As we can see here the difference is 6 between the number of rows with hospital and hospital records.

Question 5

```
#Death
Death <- unlist(c(data[ace == 1, .(sum(death))], data[ace == 0, .(sum(death))], data[beta.blocker == 1, .(sum(death))], data
[beta.blocker == 0, .(sum(death))], data[statin == 1, .(sum(death))], data[statin == 0, .(sum(death))]))

#Heart Attack
Heart_Attack <- unlist(c(data[ace == 1, .(sum(heart.attack))], data[ace == 0, .(sum(heart.attack))], data[beta.blocker == 1,
.(sum(heart.attack))], data[beta.blocker == 0, .(sum(heart.attack))], data[statin == 1, .(sum(heart.attack))], data[statin =
= 0, .(sum(heart.attack))]))

##Hospitalization
Hospitalization <- unlist(c(data[ace == 1, .(sum(hospital))], data[ace == 0, .(sum(hospital))], data[beta.blocker == 1, .(su
m(hospital))], data[beta.blocker == 0, .(sum(hospital))], data[statin == 1, .(sum(hospital))], data[statin == 0, .(sum(hospi
tal))]))

Variables <- c("With Ace", "Without Ace", "With BB", "Without BB", "With Statin", "Without Statin")
medicationssummary <- data.table(Variables, Death, Heart_Attack, Hospitalization)
datatable(medicationssummary)
```

Show **10** entries

Search:

	Variables	Death	Heart_Attack	Hospitalization
1	With Ace	2639	1475	22400
2	Without Ace	3447	1787	27242
3	With BB	2820	1333	20170
4	Without BB	3266	1929	29472
5	With Statin	2605	1536	23413
6	Without Statin	3479	1726	26229

Showing 1 to 6 of 6 entries

Previous

1

Next

Question 6

```
#Total person-years
data[, .(round((sum(end - begin))/365.25, 1))]
```

```
##      V1
## 1: 37061.1
```

```
#Average Per Person
data[, .(round((sum(end - begin))/365.25/10000, 1))]
```

```
##      V1
## 1: 3.7
```

Question 7

```
#Ace
acetime <- unlist(data[ace == 1, .(round((sum(end - begin))/365.25, 1))])
offace <- unlist(data[ace == 0, .(round((sum(end - begin))/365.25, 1))])
#BB
bbtime <- unlist(data[beta.blocker == 1, .(round((sum(end - begin))/365.25, 1))])
offbb <- unlist(data[beta.blocker == 0, .(round((sum(end - begin))/365.25, 1))])
#Statin
statintime <- unlist(data[statin == 1, .(round((sum(end - begin))/365.25, 1))])
offstatin <- unlist(data[statin == 0, .(round((sum(end - begin))/365.25, 1))])

Time <- c(acetime, bbtime, statintime)
OffTime <- c(offace, offbb, offstatin)
MedicationTime <- c("Ace", "Beta Blocker", "Statin")
medicationtotal <- data.table(MedicationTime, Time, OffTime)
datatable(medicationtotal)
```

Show 10 entries

Search:

	MedicationTime	Time	OffTime
1	Ace	26965.8	10095.4
2	Beta Blocker	22329.4	14731.8
3	Statin	28867.4	8180.6

Showing 1 to 3 of 3 entries

Previous

1

Next

```
#Missing Time
sum(data[, .(max(end)), id][[2]]) - data[, .(sum(end - begin))]
```

```
##      V1
## 1:    0
```

No time was missing from observation

Question 8

```
#Overall follow up time
Total_Time <- unlist(data[, .(round((sum(end - begin))/365.25, 1))])
#Each outcome
Event_Outcome <- unlist(c(data[heart.attack == 1, .N], data[death == 1, .N], data[, .(sum(diff(hospital) > 0))]))
#Ratio
Crude_Rate <- round(Event_Outcome / Total_Time * 100, 1)

vname <- c("Heart Attack", "Death", "Hospitalization")

Crude_Event_Rate <- data.table(vname, Total_Time, Event_Outcome, Crude_Rate)
datatable(Crude_Event_Rate)
```

Show 10 entries

Search:

	vname	Total_Time	Event_Outcome	Crude_Rate
1	Heart Attack	37061.1	3262	8.8
2	Death	37061.1	6086	16.4
3	Hospitalization	37061.1	49636	133.9

Showing 1 to 3 of 3 entries

Previous

1

Next

Question 9

```

#Number of Hospitalization Events
Hospitalization_Outcome <- unlist(c(data[ace == 1, .(sum(diff(hospital) > 0))], data[ace == 0, .(sum(diff(hospital) > 0))],
  data[beta.blocker == 1, .(sum(diff(hospital) > 0))], data[beta.blocker == 0, .(sum(diff(hospital) > 0))], data[statin == 1,
  .(sum(diff(hospital) > 0))], data[statin == 0, .(sum(diff(hospital) > 0))]))
#Hospitalization Rate
htime<- unlist(c(data[ace == 1, .((sum(end - begin))/365.25)], data[ace == 0, .((sum(end - begin))/365.25)], data[beta.block
er == 1, .((sum(end - begin))/365.25)], data[beta.blocker == 0, .((sum(end - begin))/365.25)], data[statin == 1, .((sum(end
- begin))/365.25)], data[statin == 0, .((sum(end - begin))/365.25))])
Hospitalization_Rate <- round(100 * Hospitalization_Outcome / htime, 1)

#Number of Heart Attack Events
Heart_Attack_Outcome <- unlist(c(data[ace == 1, .(sum(heart.attack))], data[ace == 0, .(sum(heart.attack))], data[beta.block
er == 1, .(sum(heart.attack))], data[beta.blocker == 0, .(sum(heart.attack))], data[statin == 1, .(sum(heart.attack))], data
[statin == 0, .(sum(heart.attack))]))
#Heart Attack Rate
Heart_Attack_Rate <- round(100 * Heart_Attack_Outcome / htime, 1)

#Number of Death
Death_Outcome <- unlist(c(data[ace == 1, .(sum(death))], data[ace == 0, .(sum(death))], data[beta.blocker == 1, .(sum(deat
h))], data[beta.blocker == 0, .(sum(death))], data[statin == 1, .(sum(death))], data[statin == 0, .(sum(death))]))
#Death Rate
Death_Rate <- round(100 * Death_Outcome / htime, 1)

Variables <- c("With Ace", "Without Ace", "With BB", "Without BB", "With Statin", "Without Statin")
Medication_Usage <- data.table(Variables, Hospitalization_Outcome, Hospitalization_Rate, Heart_Attack_Outcome, Heart_Attack_
Rate, Death_Outcome, Death_Rate)
datatable(Medication_Usage)

```

Show 10 ▾ entries

Search:

	Variables	Hospitalization_Outcome	Hospitalization_Rate	Heart_Attack_Outcome	Heart_Attack_Rate	Death_Outcome
1	With Ace	22358	82.9	1475	5.5	2639
2	Without Ace	26300	260.5	1787	17.7	3447
3	With BB	20067	89.9	1333	6	2820
4	Without BB	28946	196.5	1929	13.1	3266
5	With Statin	23366	80.9	1536	5.3	2605
6	Without Statin	25099	306.8	1726	21.1	3479

Showing 1 to 6 of 6 entries

Previous

1

Next

Question 10

```

#Hospitalization OR
Hospitalization_OR <- round(c(Hospitalization_Rate[1] / Hospitalization_Rate[2], Hospitalization_Rate[3] / Hospitalization_R
ate[4], Hospitalization_Rate[5] / Hospitalization_Rate[6]), 2)

#Heart Attack OR
Heart_Attack_OR <- round(c(Heart_Attack_Rate[1] / Heart_Attack_Rate[2], Heart_Attack_Rate[3] / Heart_Attack_Rate[4], Heart_A
ttack_Rate[5] / Heart_Attack_Rate[6]), 2)

#Death OR
Death_OR <- round(c(Death_Rate[1] / Death_Rate[2], Death_Rate[3] / Death_Rate[4], Death_Rate[5] / Death_Rate[6]), 2)

Variable_Name <- c("With Ace", "With BB", "With Statin")
OddsRatio <- data.table(Variable_Name, Hospitalization_OR, Heart_Attack_OR, Death_OR)
datatable(OddsRatio)

```

Show 10 ▾ entries

Search:

	Variable_Name	Hospitalization_OR	Heart_Attack_OR	Death_OR
1	With Ace	0.32	0.31	0.29
2	With BB	0.46	0.46	0.57
3	With Statin	0.26	0.25	0.21

Showing 1 to 3 of 3 entries

Previous

1

Next