Homework 5

Michael Ackerman

November 10, 2015

Question 1

```
#load data and packages
setwd("/Users/michaeltackerman/Documents/Bios6301/datasets")
h <- read.csv("haart.csv", stringsAsFactors = F)
library(lubridate)
library(knitr)

#convert dates
h$init.date <- mdy(h$init.date)
h$date.death <- mdy(h$date.death)
h$last.visit <- mdy(h$last.visit)

#Fix a date
i <- which(year(h$init.date)>2016)
h$init.date[i] <-mdy("04081998")

# 1) Table
table(format(h$init.date, "%Y"))
```

```
##
## 1998 2000 2001 2002 2003 2004 2005 2006 2007
##    1    5   17   60  270  292  207  104   44
```

```
# 2) Create indicator variable
h$dead1 <- ifelse(!is.na(h$date.death) & h$date.death-h$init.date <=365,1,0)
# How many people died within the first year?
sum(h[,"dead1"])
```

```
## [1] 92
```

```
h$followup <- with(h, ifelse(death==1,
                      ifelse( difftime(date.death, init.date,unit="days")<=365,
                              difftime(date.death, init.date,unit="days"), 365),
                      ifelse(difftime(last.visit, init.date,unit="days")<=365,
                             difftime(last.visit, init.date,unit="days"),365)))

#Quantiles
quantile(h$followup)
```

```
##    0%   25%   50%   75%  100%
##   0.0 329.5 365.0 365.0 365.0
```

```r
# 4) Loss of Follow up
h$lost <- with(h, ifelse(is.na(date.death) & is.na(last.visit),1,0))
# Amount lost
sum(h$lost)
```

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

```r
#5)
#Code from class
reg_list <- strsplit(as.character(h[,'init.reg']),',')
all_drugs <- unique(unlist(reg_list))
reg_drugs <- matrix(nrow=nrow(h),ncol=length(all_drugs))
for(i in seq_along(all_drugs)){
  reg_drugs[,i] <- +sapply(reg_list, function(x) all_drugs[i] %in% x)
# + makes T/F into 1/0
}
colnames(reg_drugs) <- all_drugs

# created a matrix of only unique pairings of drugs
uniques <- reg_drugs[!duplicated(reg_drugs), ]

results <- integer(47)

#this isnt the most effecient way to do this but it compares all combos
for(j in seq_along(uniques[,1])){
  for(i in seq_along(reg_drugs[,1])){
    if(identical(uniques[j,], reg_drugs[i,])){
      results[j] <-  results[j]+1 }
  }
}
# Print out the drug combos that are greater than 100
cbind(uniques[results>100,],results[results>100])
```

```
##      3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV
## [1,]   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## [2,]   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
##      FPV
## [1,]   0 421
## [2,]   0 284
```

```r
#6)
# If this question had been read before, I would have made my cleaning more generalizable; instead, it'
h <- read.csv("haart.csv",stringsAsFactor =F)
h2 <- read.csv("haart2.csv",stringsAsFactor =F)
h <- rbind(h,h2)

#convert dates
h$init.date <- mdy(h$init.date)
h$date.death <- mdy(h$date.death)
h$last.visit <- mdy(h$last.visit)

#Fix a date
```

```
i <- which(year(h$init.date)>2016)
h$init.date[i] <-mdy("04081998")

# Create indicator variable for death in first year
h$dead1 <- ifelse(!is.na(h$date.death) & h$date.death-h$init.date <=365,1,0)

# follow up
h$followup <- with(h, ifelse(death==1,
                      ifelse( difftime(date.death, init.date,unit="days")<=365,
                              difftime(date.death, init.date,unit="days"), 365),
                      ifelse(difftime(last.visit, init.date,unit="days")<=365,
                             difftime(last.visit, init.date,unit="days"),365)))

#Loss of Follow up
h$lost <- with(h, ifelse(is.na(date.death) & is.na(last.visit),1,0))

kable(h[-(6:999),-(3:7)])
```

	male	age	init.reg	init.date	last.visit	death	date.death	dead1	followup	lost
1	1	25.00000	3TC,AZT,EFV	2003-07-01	2007-02-26	0	NA	0	365	0
2	1	49.00000	3TC,AZT,EFV	2004-11-23	2008-02-22	0	NA	0	365	0
3	1	42.00000	3TC,AZT,EFV	2003-04-30	2005-11-21	1	2006-01-11	0	365	0
4	0	33.00000	3TC,AZT,NVP	2006-03-25	2006-05-05	1	2006-05-07	1	43	0
5	1	27.00000	3TC,D4T,EFV	2004-09-01	2007-11-13	0	NA	0	365	0
1000	0	40.00000	3TC,D4T,NVP	2003-07-03	2008-02-29	0	NA	0	365	0
1001	0	27.00000	3TC,AZT,NVP	2003-12-01	2004-01-05	0	NA	0	35	0
1002	1	38.72142	3TC,AZT,NVP	2002-09-26	2004-03-29	0	NA	0	365	0
1003	1	23.00000	3TC,DDI,EFV	2007-01-31	2007-04-16	0	NA	0	75	0
1004	0	31.00000	3TC,D4T,NVP	2003-12-03	2007-10-11	0	NA	0	365	0

Question 2

```
toRegress = h[,c("death", "weight", "hemoglobin", "cd4baseline")]
toRegress = toRegress[complete.cases(toRegress),]

x <- toRegress[2:4]
y <- toRegress[1]



estimate_logistic <- function(x, y, MAX_ITER=10) {

  logistic <- function(x) 1 / (1 + exp(-x))

    n <- dim(x)[1]
    k <- dim(x)[2]

    x <- as.matrix(cbind(rep(1, n), x))
    y <- as.matrix(y)

    # Initialize fitting parameters
```

```
    theta <- rep(0, k+1)

    J <- rep(0, MAX_ITER)

    for (i in 1:MAX_ITER) {

        # Calculate linear predictor
        z <- x %*% theta
        # Apply logit function
        h <- logistic(z)

        # Calculate gradient
        grad <- t((1/n)*x) %*% as.matrix(h - y)
        # Calculate Hessian
        H <- t((1/n)*x) %*% diag(array(h)) %*% diag(array(1-h)) %*% x

        # Calculate log likelihood
        J[i] <- (1/n) %*% sum(-y * log(h) - (1-y) * log(1-h))

        # Newton's method
        theta <- theta - solve(H) %*% grad
    }

    return(theta)
}
estimate_logistic(x, y)
```

```
##                      [,1]
## rep(1, n)    3.585111170
## weight      -0.046283908
## hemoglobin  -0.351255050
## cd4baseline  0.002094131
```

Question 3

```
setwd("/Users/michaeltackerman/Documents/Bios6301/datasets")
addr <- readLines("addr.txt")

this <- lapply(addr, function(s) {unlist(strsplit(s, split= "[ ]{2,}"))})
streetno <- unlist(lapply(this, function(s) {strsplit(s[3],split="[ ]{1}")[[1]][1]} ))
streetname <- unlist(lapply(this, function(s) {gsub("[0-9 ]{1,} ","",s[3])} ))
info <- data.frame()

for(i in seq_along(this)){
  for(j in c(1,2)){
    info[i,j] <- this[[i]][j]
  }
    info[i,3] <- streetno[i]
    info[i,4] <- streetname[i]

  for(j in 5:7){
    info[i,j] <- this[[i]][j-1]
  }
```

```
}

names(info) <- c("lastname", "firstname", "streetno", "streetname", "city", "state", "zip")

kable(info)
```

lastname	firstname	streetno	streetname	city	state	zip
Bania	Thomas M.	725	Commonwealth Ave.	Boston	MA	O2215
Barnaby	David	373	W. Geneva St.	Wms. Bay	WI	53191
Bausch	Judy	373	W. Geneva St.	Wms. Bay	WI	53191
Bolatto	Alberto	725	Commonwealth Ave.	Boston	MA	O2215
Carlstrom	John	933	E. 56th St.	Chicago	IL	60637
Chamberlin	Richard A.	111	Nowelo St.	Hilo	HI	96720
Chuss	Dave	2145	Sheridan Rd	Evanston	IL	60208-3112
Davis	E. J.	933	E. 56th St.	Chicago	IL	60637
Depoy	Darren	174	W. 18th Ave.	Columbus	OH	43210
Griffin	Greg	5000	Forbes Ave.	Pittsburgh	PA	15213
Halvorsen	Nils	933	E. 56th St.	Chicago	IL	60637
Harper	Al	373	W. Geneva St.	Wms. Bay	WI	53191
Huang	Maohai	725	W. Commonwealth Ave.	Boston	MA	O2215
Ingalls	James G.	725	W. Commonwealth Ave.	Boston	MA	O2215
Jackson	James M.	725	W. Commonwealth Ave.	Boston	MA	O2215
Knudsen	Scott	373	W. Geneva St.	Wms. Bay	WI	53191
Kovac	John	5640	S. Ellis Ave.	Chicago	IL	60637
Landsberg	Randy	5640	S. Ellis Ave.	Chicago	IL	60637
Lo	Kwok-Yung	1002	W. Green St.	Urbana	IL	61801
Loewenstein	Robert F.	373	W. Geneva St.	Wms. Bay	WI	53191
Lynch	John	4201	Wilson Blvd	Arlington	VA	22230
Martini	Paul	174	W. 18th Ave.	Columbus	OH	43210
Meyer	Stephan	933	E. 56th St.	Chicago	IL	60637
Mrozek	Fred	373	W. Geneva St.	Wms. Bay	WI	53191
Newcomb	Matt	5000	Forbes Ave.	Pittsburgh	PA	15213
Novak	Giles	2145	Sheridan Rd	Evanston	IL	60208-3112
Odalen	Nancy	373	W. Geneva St.	Wms. Bay	WI	53191
Pernic	Dave	373	W. Geneva St.	Wms. Bay	WI	53191
Pernic	Bob	373	W. Geneva St.	Wms. Bay	WI	53191
Peterson	Jeffrey	5000	Forbes Ave.	Pittsburgh	PA	15213
Pryke	Clem	933	E. 56th St.	Chicago	IL	60637
Rebull	Luisa	5640	S. Ellis Ave.	Chicago	IL	60637
Renbarger	Thomas	2145	Sheridan Rd	Evanston	IL	60208-3112
Rottman	Joe	8730	W. Mountain View Ln	Littleton	CO	80125
Schartman	Ethan	933	E. 56th St.	Chicago	IL	60637
Spotz	Bob	373	W. Geneva St.	Wms. Bay	WI	53191
Thoma	Mark	373	W. Geneva St.	Wms. Bay	WI	53191
Walker	Chris	933	N. Cherry St.	Tucson	AZ	85721
Wehrer	Cheryl	5000	Forbes Ave.	Pittsburgh	PA	15213
Wirth	Jesse	373	W. Geneva St.	Wms. Bay	WI	53191
Wright	Greg	791	Holmdel-Keyport Rd.	Holmdel	NY	O7733-1988
Zingale	Michael	5640	S. Ellis Ave.	Chicago	IL	60637

** Question 4 **