# Project 6 Solutions

## Kevin Choe

## Collaborators: N/A

## TA help: Hilda Ibriga and Katie Brinkers guided with problems 1-5, explained new concepts

## Online resources used: Stat 190 Example Book (All problems)

### Question 1

```
dat <- read.csv("/class/datamine/data/fars/7581.csv")

#Calculates mean number of motorists
tapply(dat$PERSON, dat$DRUNK_DR, mean)
```

```
       0        1        2        3        4        6 
2.615540 2.474079 3.660711 5.197917 5.250000 6.000000 
```

```
# Read in data that maps state codes to state names
state_names <- read.csv("/class/datamine/data/fars/states.csv")

# Create a vector of state names called v
v <- state_names$state

# Set the names of the new vector to the codes
names(v) <- state_names$code

# Create a new column in the dat dataframe with the actual names of the states
dat$mystates <- v[as.character(dat$STATE)]
```

I prefer to use the tapply function because it is easier and simpler to write.

### Question 2

```
#Sorts states by highest average number of drunk drivers per accident

tapply(dat$DRUNK_DR, dat$mystates, mean)
```

```
             Alabama               Alaska              Arizona 
           0.2136050            0.5223022            0.4126347 
            Arkansas           California             Colorado 
           0.2650494            0.4863834            0.5326633 
         Connecticut             Delaware District of Columbia 
           0.4621138            0.5642023            0.3153409 
             Florida              Georgia               Hawaii 
           0.2898366            0.3309584            0.4952652 
               Idaho             Illinois              Indiana 
           0.4049811            0.3366005            0.2717200 
```

```
                Iowa               Kansas             Kentucky 
           0.3609572            0.3133971            0.3637387 
           Louisiana                Maine             Maryland 
           0.3241348            0.4916084            0.3422666 
       Massachusetts             Michigan            Minnesota 
           0.3308242            0.4713560            0.4492386 
         Mississippi             Missouri              Montana 
           0.1688661            0.2078921            0.5269231 
            Nebraska               Nevada        New Hampshire 
           0.4146229            0.5127907            0.6094050 
          New Jersey           New Mexico             New York 
           0.4286125            0.3184573            0.1983089 
      North Carolina         North Dakota                 Ohio 
           0.2678010            0.2887538            0.3161686 
            Oklahoma               Oregon         Pennsylvania 
           0.3484964            0.4692250            0.3793978 
        Rhode Island       South Carolina         South Dakota 
           0.4188830            0.3052830            0.5132450 
           Tennessee                Texas                 Utah 
           0.4159967            0.1852601            0.3385707 
             Vermont             Virginia           Washington 
           0.5126263            0.3426975            0.5498288 
       West Virginia            Wisconsin              Wyoming 
           0.1672332            0.5350330            0.4110644 
```

```
sort(tapply(dat$DRUNK_DR, dat$mystates, mean), decreasing = TRUE)
```

```
       New Hampshire             Delaware           Washington 
           0.6094050            0.5642023            0.5498288 
           Wisconsin             Colorado              Montana 
           0.5350330            0.5326633            0.5269231 
              Alaska         South Dakota               Nevada 
           0.5223022            0.5132450            0.5127907 
             Vermont               Hawaii                Maine 
           0.5126263            0.4952652            0.4916084 
          California             Michigan               Oregon 
           0.4863834            0.4713560            0.4692250 
         Connecticut            Minnesota           New Jersey 
           0.4621138            0.4492386            0.4286125 
        Rhode Island            Tennessee             Nebraska 
           0.4188830            0.4159967            0.4146229 
             Arizona              Wyoming                Idaho 
           0.4126347            0.4110644            0.4049811 
        Pennsylvania             Kentucky                 Iowa 
           0.3793978            0.3637387            0.3609572 
            Oklahoma             Virginia             Maryland 
           0.3484964            0.3426975            0.3422666 
                Utah             Illinois              Georgia 
           0.3385707            0.3366005            0.3309584 
       Massachusetts            Louisiana           New Mexico 
           0.3308242            0.3241348            0.3184573 
                Ohio District of Columbia               Kansas 
           0.3161686            0.3153409            0.3133971 
      South Carolina              Florida         North Dakota 
           0.3052830            0.2898366            0.2887538 
```

```
             Indiana       North Carolina             Arkansas
           0.2717200            0.2678010            0.2650494
             Alabama             Missouri             New York
           0.2136050            0.2078921            0.1983089
               Texas          Mississippi        West Virginia
           0.1852601            0.1688661            0.1672332
```

New Hampshire has the highest average number of drunk drivers per accident.

### Question 3

```r
#Sorts total number of fatalities for each day of the week.

sort(tapply(dat$FATALS, dat$DAY_WEEK, sum), decreasing = TRUE)
```

```
    7     1     6     5     4     2     3     9
72253 56985 56406 41802 38737 37115 36441     3
```

```r
#Sorts proportion of fatalities over the total number of people in the accidents

sort(tapply(dat$FATALS, dat$DAY_WEEK, sum), decreasing = TRUE)/sort(tapply(dat$PERSONS, dat$DAY_WEEK, su
```

```
        7         1         6         5         4         2         3
0.4289692 0.4219423 0.4319915 0.4512842 0.4509598 0.4440018 0.4486371
        9
1.0000000
```

The numbers are suprising to me because Sundays, Saturdays, and Fridays have the highest number of fatalities compared to other days.

I was expecting a smaller proportion for the days with higher number of fatalities. I was expecting a high proportion on day 5.

### Question 4

```r
#Sorts average number of crashes involving drunk drivers that occur on straight, curvy, and unknown roa

sort(tapply(dat$DRUNK_DR, dat$ALIGNMNT, mean), decreasing = TRUE)
```

```
        2         1         9
0.4729582 0.3143146 0.2764798
```

The average for straight roads is 0.31 and the average for curvy roads is 0.47

### Question 5

```r
#Finds the total number of fatalities in time interval

tapply ( dat$FATALS, cut(dat$HOUR, breaks=c(0,6,12,18,24,99), include.lowest=T), sum )
```

```
  [0,6]  (6,12] (12,18] (18,24] (24,99]
  93151   49764   96375   98715    1737
```

```r
#Finds the average number of fatalities in time interval

tapply ( dat$FATALS, cut(dat$HOUR, breaks=c(0,6,12,18,24,99), include.lowest=T), mean)
```

```
   [0,6]   (6,12]  (12,18]  (18,24]  (24,99]
1.133293 1.123037 1.128671 1.140331 1.087664
```

## Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.
> Accountable together - We are Purdue.