

DaigleRExam.R

2011home

Thu Mar 1 12:36:29 2018

```
## Chris Daigle #####
```

```
# Exam 1, 1 Mar ###
```

```
# Question 1 ####
```

```
countyData <- read.csv("/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Sp  
summary(countyData)
```

```
##      State      County      TotalPop  
## Length:3220      Length:3220      Min.      :      85  
## Class :character      Class :character      1st Qu.:    11218  
## Mode  :character      Mode  :character      Median :    26035  
##                                         Mean   :    99409  
##                                         3rd Qu.:    66430  
##                                         Max.    :10038388  
##      Men      Women      Hispanic      White  
## Min.      :    42      Min.      :    43      Min.      : 0.000      Min.      : 0.00  
## 1st Qu.:    5637      1st Qu.:    5572      1st Qu.: 1.900      1st Qu.:64.10  
## Median :   12932      Median :   13057      Median : 3.900      Median :84.10  
## Mean      :   48897      Mean      :   50512      Mean      :11.012      Mean      :75.43  
## 3rd Qu.:   32993      3rd Qu.:   33488      3rd Qu.: 9.825      3rd Qu.:93.20  
## Max.      :4945351      Max.      :5093037      Max.      :99.900      Max.      :99.80  
##      Black      Native      Asian      Pacific  
## Min.      : 0.000      Min.      : 0.000      Min.      : 0.000      Min.      : 0.00000  
## 1st Qu.: 0.500      1st Qu.: 0.100      1st Qu.: 0.200      1st Qu.: 0.00000  
## Median : 1.900      Median : 0.300      Median : 0.500      Median : 0.00000  
## Mean      : 8.665      Mean      : 1.724      Mean      : 1.229      Mean      : 0.08273  
## 3rd Qu.: 9.600      3rd Qu.: 0.600      3rd Qu.: 1.200      3rd Qu.: 0.00000  
## Max.      :85.900      Max.      :92.100      Max.      :41.600      Max.      :35.30000  
##      IncomePerCap      Unemployment  
## Min.      : 5878      Min.      : 0.000  
## 1st Qu.:20238      1st Qu.: 5.500  
## Median :23460      Median : 7.600  
## Mean      :23982      Mean      : 8.094  
## 3rd Qu.:27053      3rd Qu.: 9.900  
## Max.      :65600      Max.      :36.500
```

```
# Question 1, Part 1 ####
```

```
# Find the county that has the largest population in each state
```

```
maxStateCounty <- aggregate(countyData[, c("County", "TotalPop")], list(countyData$State), max, na.rm=
```

```
maxCounty <- maxStateCounty[, c(2,3)]
```

```
maxCounty
```

```
##      County      TotalPop  
## 1      Winston      659026  
## 2 Yukon-Koyukuk Census Area      299107  
## 3      Yuma      4018143  
## 4      Yell      390463  
## 5      Yuba      10038388
```

## 6	Yuma	655024
## 7	Windham	939983
## 8	Sussex	549643
## 9	District of Columbia	647484
## 10	Washington	2639042
## 11	Worth	983903
## 12	Maui	984178
## 13	Washington	417501
## 14	Woodford	5236393
## 15	Whitley	926335
## 16	Wright	452369
## 17	Wyandotte	566814
## 18	Woodford	755809
## 19	Winn	444690
## 20	York	286119
## 21	Worcester	1017859
## 22	Worcester	1556116
## 23	Wexford	1778969
## 24	Yellow Medicine	1197776
## 25	Yazoo	245874
## 26	Wright	1001327
## 27	Yellowstone	153692
## 28	York	537655
## 29	White Pine	2035572
## 30	Sullivan	403972
## 31	Warren	926330
## 32	Valencia	673943
## 33	Yates	2595259
## 34	Yancey	990288
## 35	Williams	162500
## 36	Wyandot	1263189
## 37	Woodward	754480
## 38	Yamhill	768418
## 39	York	1555072
## 40	Yauco	371400
## 41	Washington	630459
## 42	York	474903
## 43	Ziebach	178942
## 44	Wilson	937750
## 45	Zavala	4356362
## 46	Weber	1078958
## 47	Windsor	159711
## 48	York	1128722
## 49	Yakima	2045756
## 50	Wyoming	190781
## 51	Wood	955939
## 52	Weston	95431

Question 1, Part 2

Select a subset of counties whose population belong to the top 10% and compute the mean and median of

```
sub <- countyData[,c(1,2,3)]
```

```
quant <- quantile(sub$TotalPop, probs = c(0.25, 0.5, 0.75, 0.9), na.rm = TRUE)
```

```
interestVector <- rep(NA, length(sub$TotalPop))
```

```

interestVector[sub$TotalPop >= quant[4]] <- TRUE
topTen <- sub$County[which(sub$TotalPop >= quant[4])]
avgTopTenCounty <- mean(sub$TotalPop[which(sub$County == topTen)])
medTopTenCounty <- median(sub$TotalPop[which(sub$County == topTen)])
avgTopTenCounty

```

```
## [1] 85308
```

```
medTopTenCounty
```

```
## [1] 85838
```

```
# Question 2 ####
```

```

oneHundred <- rnorm(100, 0, 1)
tenByTen <- matrix(sample(oneHundred), nrow=10, ncol = 10)

```

```

row.names(tenByTen) <- c(1,2,3,4,5,6,7,8,9,10)
rowMins <- rep(NA, 10)
for (i in 1:10){
  rowMins[i] <- min(tenByTen[i,])
}

```

```
# Question 3 ####
```

```

regression <- function(X,Y) {
  Z = X ^ 2
  bHatOne <- (sum(Z ^ 2) * sum(X*Y) - sum(X*Z)*sum(Z*Y)) / (sum(X^2)*sum(Z^2) - sum(X*Z)^2 )
  bHatTwo <- (sum(X ^ 2) * sum(Z*Y) - sum(X*Z)*sum(X*Y)) / (sum(X^2)*sum(Z^2) - sum(X*Z)^2 )
  bHatZero <- mean(Y) - bHatOne*mean(X) - bHatTwo*mean(Z)
  cat("intercept:", bHatZero)
  print("")
  cat("betaHatOne:", bHatOne)
  print("")
  cat("betaHatTwo:", bHatTwo)
}

```

```

X <- (countyData$Men / countyData$Women)
Z <- X ^ 2
Y <- countyData$Unemployment
regTest <- lm(countyData$Unemployment ~ X + Z)

```

```
regression(X, Y)
```

```

## intercept: 0.02699652[1] ""
## betaHatOne: 13.74382[1] ""
## betaHatTwo: -5.613248

```

```
coefficients(regTest[1])
```

```

## (Intercept)          X          Z
## 22.106582 -21.149591  7.081153

```

```
# The function takes in the values as requested, but testing against the builtin "lm", my function does
```