Assignment1

Kowshik Sarker

7/19/2020

## Github Repo Link - “<https://github.com/kowshiksarker/IBA-Repo>”

## 1.From the USJudgeRatings dataset find the mean and standard delivation of the oevrall ratings of the judges

overallratings<-USJudgeRatings  
overallratings<-transform(overallratings,Mean=apply(overallratings,1,mean),SD=apply(overallratings,1,sd))  
overallratings<-as.data.frame(overallratings[,c(13:14)])  
overallratings

## Mean SD  
## AARONSON,L.H. 7.291667 0.6459079  
## ALEXANDER,J.M. 8.150000 0.5807519  
## ARMENTANO,A.J. 7.616667 0.2657180  
## BERDON,R.I. 8.458333 0.5484828  
## BRACKEN,J.J. 5.733333 0.8172385  
## BURNS,E.B. 8.116667 0.6873312  
## CALLAHAN,R.J. 8.858333 0.5991787  
## COHEN,S.S. 5.458333 0.7786449  
## DALY,J.J. 8.516667 0.4281744  
## DANNEHY,J.F. 7.891667 0.4294994  
## DEAN,H.H. 7.458333 0.4144182  
## DEVITA,H.J. 7.125000 0.3744693  
## DRISCOLL,P.J. 7.366667 0.6443225  
## GRILLO,A.E. 6.683333 0.3613946  
## HADDEN,W.L.JR. 7.850000 0.4562695  
## HAMILL,E.C. 7.450000 0.3060006  
## HEALEY.A.H. 6.866667 0.4942089  
## HULL,T.C. 7.400000 0.3541956  
## LEVINE,I. 7.808333 0.2712206  
## LEVISTER,R.L. 6.608333 1.0663944  
## MARTIN,L.F. 7.091667 0.5017394  
## MCGRATH,J.F. 6.783333 0.4489044  
## MIGNONE,A.F. 5.841667 0.7140898  
## MISSAL,H.M. 7.458333 0.5282188  
## MULVEY,H.M. 8.450000 0.3205110  
## NARUK,H.J. 8.783333 0.3270622  
## O'BRIEN,F.J. 7.941667 0.3654594  
## O'SULLIVAN,T.J. 8.483333 0.3950451  
## PASKEY,L. 8.066667 0.2570226  
## RUBINOW,J.E. 8.791667 0.5822501  
## SADEN.G.A. 7.775000 0.5879471  
## SATANIELLO,A.G. 7.800000 0.3074824  
## SHEA,D.M. 8.191667 0.4718596  
## SHEA,J.F.JR. 8.500000 0.4199567  
## SIDOR,W.J. 5.808333 0.6921092  
## SPEZIALE,J.A. 8.183333 0.1749459  
## SPONZO,M.J. 7.841667 0.3553701  
## STAPLETON,J.F. 7.683333 0.4667749  
## TESTO,R.J. 7.108333 0.5264950  
## TIERNEY,W.L.JR. 7.983333 0.3186144  
## WALL,R.A. 7.016667 0.7505553  
## WRIGHT,D.B. 7.941667 0.3941812  
## ZARRILLI,K.J. 7.425000 0.4673426

## 2.Read the Aids2.csv file

Before reading the file we need to set our working directory by setwd() command and keep the file in this directory and then execute the below commands.

aids<- read.csv("Aids2.csv")  
head(aids)

## X state sex diag death status T.categ age  
## 1 1 NSW M 10905 11081 D hs 35  
## 2 2 NSW M 11029 11096 D hs 53  
## 3 3 NSW M 9551 9983 D hs 42  
## 4 4 NSW M 9577 9654 D haem 44  
## 5 5 NSW M 10015 10290 D hs 39  
## 6 6 NSW M 9971 10344 D hs 36

## 3.Create a subset of the data without the state “Other”

subset\_aids<-subset(aids,aids$state!="Other")  
head(subset\_aids)

## X state sex diag death status T.categ age  
## 1 1 NSW M 10905 11081 D hs 35  
## 2 2 NSW M 11029 11096 D hs 53  
## 3 3 NSW M 9551 9983 D hs 42  
## 4 4 NSW M 9577 9654 D haem 44  
## 5 5 NSW M 10015 10290 D hs 39  
## 6 6 NSW M 9971 10344 D hs 36

## 4.Add a new variable called ‘agebracket’

## if age is below 20, agebracket is “0-20”

## if age is between 20 to 40, agebracket is “20-40”

## if age is between 40 to 60, agebracket is “40-60”

## if age is above 60, agebracket is “>60”

subset\_aids$agebracket<-ifelse(subset\_aids$age<"20",c("0-20"),  
 ifelse(subset\_aids$age>="20" & subset\_aids$age<"40",c("20-40"),  
 ifelse(subset\_aids$age>="40" & subset\_aids$age<="60",c("40-60"),  
 ifelse(subset\_aids$age>"60",c(">60"),"NA"))))  
   
head(subset\_aids)

## X state sex diag death status T.categ age agebracket  
## 1 1 NSW M 10905 11081 D hs 35 20-40  
## 2 2 NSW M 11029 11096 D hs 53 40-60  
## 3 3 NSW M 9551 9983 D hs 42 40-60  
## 4 4 NSW M 9577 9654 D haem 44 40-60  
## 5 5 NSW M 10015 10290 D hs 39 20-40  
## 6 6 NSW M 9971 10344 D hs 36 20-40

## 5.Sort the data from high to low based on the variable “diag” and then low to high based on “death”

subset\_aids<-subset\_aids[order(-subset\_aids$diag,subset\_aids$death),]  
head(subset\_aids)

## X state sex diag death status T.categ age agebracket  
## 1654 1654 NSW M 11503 11504 A hs 56 40-60  
## 1755 1755 NSW M 11503 11504 A hs 32 20-40  
## 1650 1650 NSW M 11502 11504 A hs 39 20-40  
## 1680 1680 NSW M 11502 11504 A hs 26 20-40  
## 2011 2011 QLD M 11502 11504 A hsid 36 20-40  
## 2654 2654 VIC M 11502 11504 A hs 33 20-40

## 6.Calculate and add one more variable which is (diag^2/death) and name it as “dd”

subset\_aids$dd<-subset\_aids$diag^2/subset\_aids$death  
head(subset\_aids)

## X state sex diag death status T.categ age agebracket dd  
## 1654 1654 NSW M 11503 11504 A hs 56 40-60 11502  
## 1755 1755 NSW M 11503 11504 A hs 32 20-40 11502  
## 1650 1650 NSW M 11502 11504 A hs 39 20-40 11500  
## 1680 1680 NSW M 11502 11504 A hs 26 20-40 11500  
## 2011 2011 QLD M 11502 11504 A hsid 36 20-40 11500  
## 2654 2654 VIC M 11502 11504 A hs 33 20-40 11500

## End Of The Assignment