

AMS597 Quiz 2 Suggested Solution

Make sure your current working directory in your R session is the same directory that contains the downloaded data file.

```
# Read in the data file
mydat <- read.delim('Quiz2Data.txt', header=T)
```

```
## Part (a)
```

```
# Create a table of the Race variable data
table(mydat$Race) ## AA, 209
```

```
##
##    AA Asian White
##  209   148   143
```

Thus, we observe that the racial group with highest frequency is **AA**, with **count = 209**.

```
## Part (b)
```

```
# Create vector containing the different Race variable values in the dataset
urace <- unique(mydat$Race)
# Create vector to store the new variable to be added to the dataset
MarkerStd <- rep(NA,dim(mydat)[1])
# For each Race value, compute the sample sd of Marker2 and
# divide the corresponding Marker1 values by the Marker2 sd for that Race value
for(i in 1:length(urace)){
  cur.sd <- sd(na.omit(mydat$Marker2[mydat$Race==urace[i]]))
  MarkerStd[which(mydat$Race==urace[i])] <- mydat$Marker1[which(mydat$Race==urace[i])]/cur.sd
}
# Add the MarkerStd column to the dataset
mydat2 <- data.frame(mydat,MarkerStd=MarkerStd)

head(mydat2)
```

```
##    Race Marker1 Marker2 MarkerStd
## 1    AA   106.0    54.7  1.8577222
## 2 White   110.0   119.0  1.8441159
## 3 Asian    86.1    32.0  1.6190188
## 4 White    24.0   171.0  0.4023526
## 5 White   110.0   170.0  1.8441159
## 6    AA    53.4    95.6  0.9358714
```