HW2_Vikas Sanil

Vikas Sanil

Due Date 2/9

```
#Using tigerstats library for data.
library(tigerstats)
```

1. Changing the author field and file name. (5 points)

(a) Change the `author:` field on the Rmd document from Your Name Here to your own name.

(b) Rename this file to "HW2_YourHameHere.Rmd", where YourNameHere is changed to your own name.

2. Measure of location and variability (20 points)

A certain polymer is used for evacuation systems for aircraft. It is important that the polymer be resistant to the aging process. Twenty specimens of the polymer were used in an experiment. Ten were assigned randomly to be exposed to an accelerated batch aging process that involved exposure to high temperatures for 10 days. Measurements of tensile strength of the specimens were made, and the following data were recorded on tensile strength in psi:

```
No aging: 227 222 218 216 218 217 225 229 228 221
Aging: 219 214 218 203 215 211 209 204 201 205
```

```
# You can use the following code to create a data frame
strength = c( 227 ,222, 218, 216, 218, 217, 225, 229, 228,221,219,214,218,203,215,211,209,204,201,205)
aging<-as.factor(c(rep(0,10),rep(1,10)))
polymerData<-data.frame(strength,aging)
noAgingPolymerData<-polymerData$strength[polymerData$aging==0]
agedPolymerData<-polymerData$strength[polymerData$aging==1]
```

```
meanOfNoAgingPolymer<-mean(noAgingPolymerData)
meanOfAgingPolymer<-mean(agedPolymerData)
```

```
cat('The sample mean tensile strength of no aging polymer is ', meanOfNoAgingPolymer , fill = TRUE)
```

(a) Calculate the sample mean tensile strength of the two samples. (5 points)

```
## The sample mean tensile strength of no aging polymer is  222.1
```

```r
cat('The sample mean tensile strength of aging polymer is ', meanOfAgingPolymer, fill = TRUE )
```

```
## The sample mean tensile strength of aging polymer is  209.9
```

```r
medianOfNoAgingPolymer<-median(noAgingPolymerData)
medianOfAgingPolymer<-median(agedPolymerData)
```

```r
cat('For No aging polymer  median tensile strength of ', medianOfNoAgingPolymer, ', is **lower** than tl
```

(b) Calculate the median for both. Discuss your observation with the mean and median of each group. (5 points)

For No aging polymer median tensile strength of 221.5 , is **lower** than the mean tensile strength 222.1 .

```r
cat('For Aging polymer  median tensile strength of ', medianOfAgingPolymer, ', is **greater** than the r
```

For Aging polymer median tensile strength of 210 , is **greater** than the mean tensile strength 209.9 .

```r
varOfNoAgingPolymer<-var(noAgingPolymerData)
varOfAgingPolymer<-var(agedPolymerData)
sdOfNoAgingPolymer<-sd(noAgingPolymerData)
sdOfAgingPolymer<-sd(agedPolymerData)
```

```r
cat('The sample variance tensile strength of no aging polymer is ', varOfNoAgingPolymer )
```

(c) Calculate the sample variance as well as standard deviation in tensile strength for both samples. (5 points)

```
## The sample variance tensile strength of no aging polymer is  23.65556
```

```r
cat('The sample deviation tensile strength of no aging polymer is ', sdOfNoAgingPolymer )
```

```
## The sample deviation tensile strength of no aging polymer is  4.863698
```

```r
cat('The sample variance tensile strength of aging polymer is ', varOfAgingPolymer )
```

```
## The sample variance tensile strength of aging polymer is  42.1
```

```r
cat('The sample deviation tensile strength of aging polymer is ', sdOfAgingPolymer )
```

```
## The sample deviation tensile strength of aging polymer is  6.488451
```

```r
cat('With Aging we can see the tensile strength varies largely for the Polymer.')
```

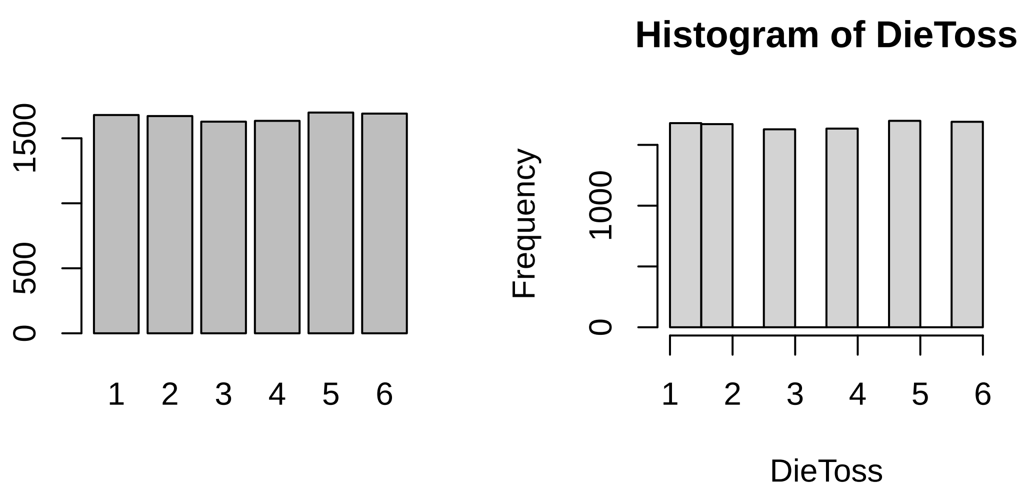
(d) Does there appear to be any evidence that aging affects the variability in tensile strength? (5 points)

```
## With Aging we can see the tensile strength varies largely for the Polymer.
```

3. Experiment of Tossing A Die (20 points)

Suppose we have a fair die. Let $n = 10000$ be the number of tossing.

```r
# To generate a fair die
n <- 10000
x <- c(1:6)
DieToss <- sample(x, size=n, replace=TRUE)
par(mfrow=c(1,2))
barplot(table(DieToss))
hist(DieToss)
```




```r
dieToss10<-sample(x, size=10, replace=TRUE)
dieToss1000<-sample(x, size=1000, replace=TRUE)
dieToss100000<-sample(x, size=100000, replace=TRUE)
```

```r
meanDieToss10<-mean(dieToss10)
meanDieToss1000<-mean(dieToss1000)
meanDieToss100000<-mean(dieToss100000)
overAllMean<- mean(c(dieToss100000, dieToss1000, dieToss10))

meanDieToss10
```

(a) Modify the number of repetition to 10 times, 1000 times, and 100000 times in the handout2-2. Calculate the mean of the results instead of sum. (10 points)

```
## [1] 3.7
```

```r
meanDieToss1000
```

```
## [1] 3.492
```

```r
meanDieToss100000
```

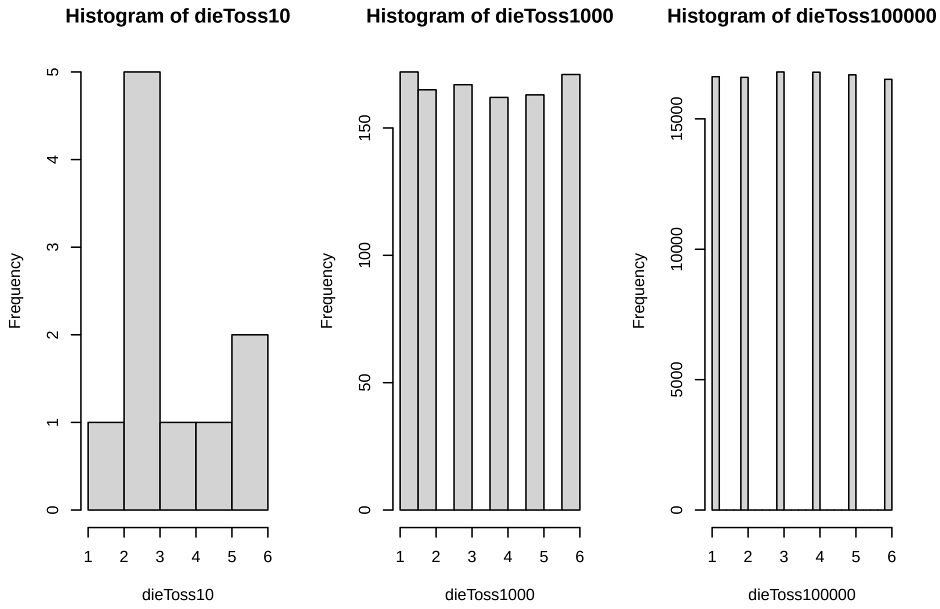
```
## [1] 3.49888
```

```r
overAllMean
```

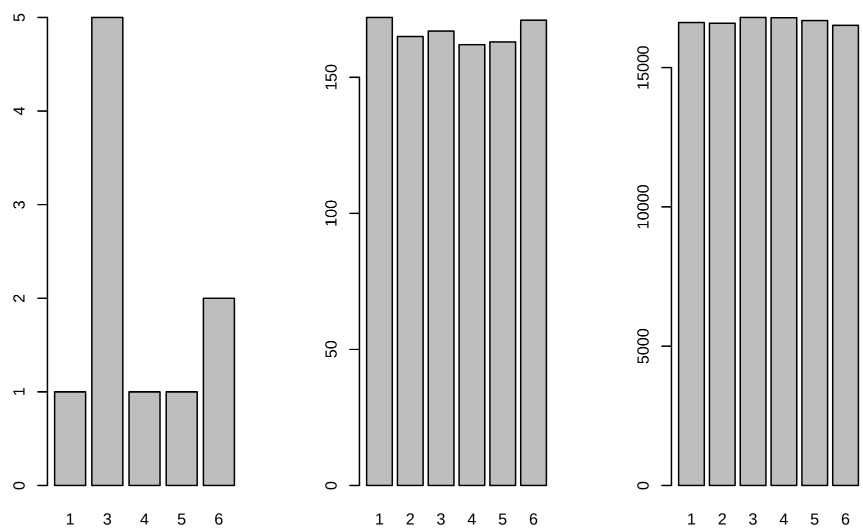
```
## [1] 3.498832
```

```r
par(mfrow=c(1,3))
hist(dieToss10)
hist(dieToss1000)
hist(dieToss100000)
```

(b) Save the results of each repetition and use par to display the histogram plot of each repetition. (10 points)




```
par(mfrow=c(1,3))
barplot(table(dieToss10))
barplot(table(dieToss1000))
barplot(table(dieToss100000))
```



4.Normal Distribution with pnorm and qnorm (20 points)

```
par(mfrow=c(1,3))
pnormGC((-0.3), region="below", graph=TRUE)
```

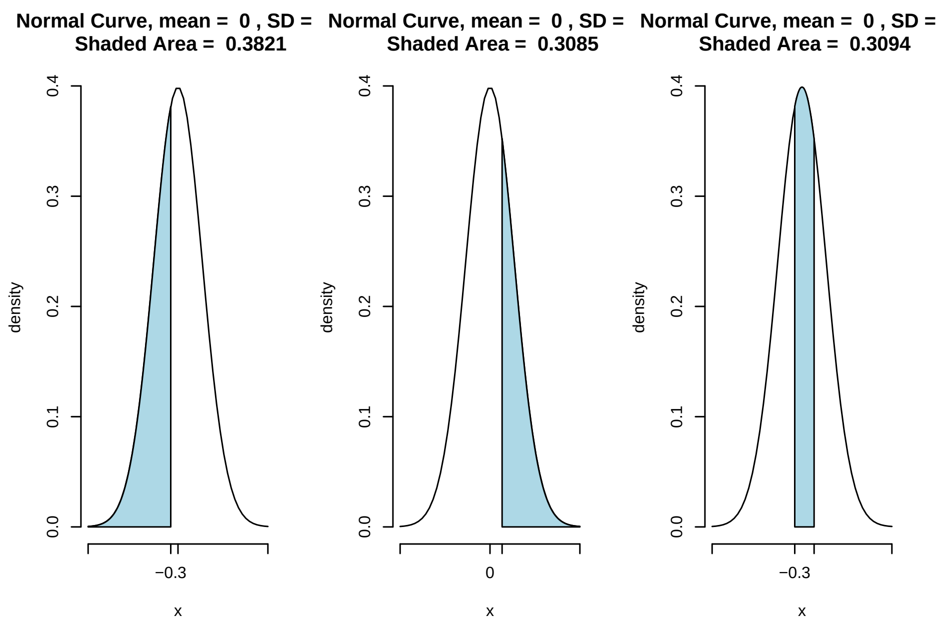
(a) Suppose $Z \sim \mathcal{N}(0,1)$, please use `pnormGC` to find $P(X \leq -0.3)$, $P(X \geq 0.5)$, $P(-0.3 \leq X \leq 0.5)$ (10 points)

```
## [1] 0.3820886
```

```
pnormGC((0.5), region="above", graph=TRUE)
```

```
## [1] 0.3085375
```

```
pnormGC(c(-0.3,0.5), region="between", graph=TRUE)
```




```
## [1] 0.3093739
```

```
z0<-qnorm((1-0.025), mean = 0, sd=1)
z0
```

(b) Please use `qnorm` to find the critical values $z_{0.025}$ and $z_{0.005}$. Which value is larger? Why? (10 points) [1] 1.959964

```
z1<-qnorm((1-0.005), mean = 0, sd=1)
z1
```

[1] 2.575829

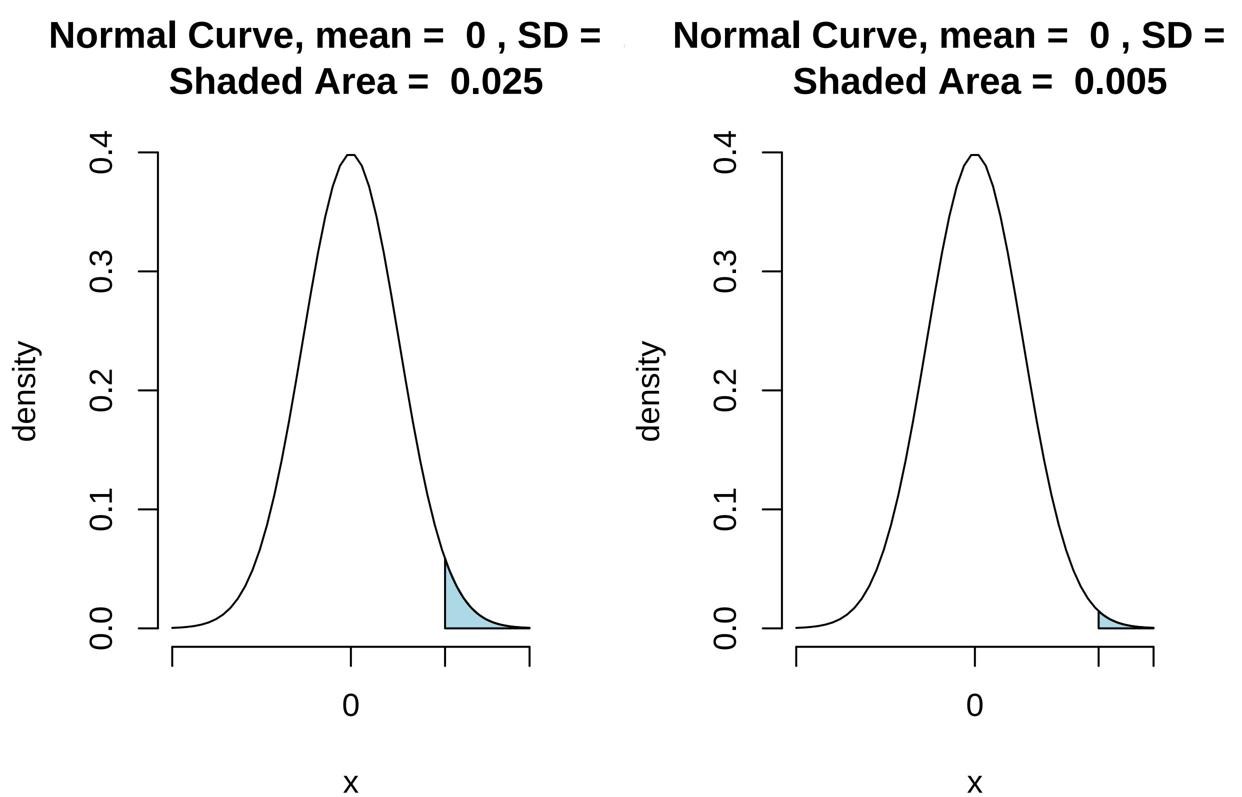
```r
cat('The critical value of $z_{0.005}$ is greater than $z_{0.025}$ as it is away from the mean value co
```

The critical value of $z_{0.005}$ is greater than $z_{0.025}$ as it is away from the mean value compared to $z_{0.025}$. Below graphs demonstrates the same.

```r
par(mfrow=c(1,2))
pnormGC(z0, region="above", graph=TRUE)
```

[1] 0.025

```r
pnormGC(z1, region="above", graph=TRUE)
```



[1] 0.005

Type your answer here

5. Working With Data (35 points)

- Obtaining the wine quality dataset

(a) Import the `winequality-red` data set and name it `winequalRed`. (5 points)

```
# here is a hint for the r version
# -- change these commands as needed and delete these comments before submitting your work --
# if you downloaded the data set as a .csv file then you can read it in as follows:
# winequalRed <- read.csv("~/Documents/datasets/winequality-red.csv", sep=";")
# To view the data set
#   View(winequalRed)
```

```
winequalRed <- read.csv("~/datasets/winequality-red.csv", sep=";")
```

(b) Create a table of the `quality` and `alcohol` attributes from the `winequalRed` data set. (5 points) Do not save the output from the code.

```
# hint: if you have two data columns named X and Y in your data frame, you can use code like this to cre
table(my.data.set$X, my.data.set$Y)
```

```
table_QA<-table(winequalRed$alcohol, winequalRed$quality)
names(dimnames(table_QA))<-c("Alcohol","Quality")
```

```
wineequalRed20RowDF<-winequalRed[1:20,]
wineequalRed20RowDF
```

(c) Save the first twenty records of the `winequalRed` data set as their own data frame. (5 points)

```
##    fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1            7.4            0.700        0.00            1.9     0.076
## 2            7.8            0.880        0.00            2.6     0.098
## 3            7.8            0.760        0.04            2.3     0.092
## 4           11.2            0.280        0.56            1.9     0.075
## 5            7.4            0.700        0.00            1.9     0.076
## 6            7.4            0.660        0.00            1.8     0.075
## 7            7.9            0.600        0.06            1.6     0.069
## 8            7.3            0.650        0.00            1.2     0.065
## 9            7.8            0.580        0.02            2.0     0.073
## 10           7.5            0.500        0.36            6.1     0.071
## 11           6.7            0.580        0.08            1.8     0.097
## 12           7.5            0.500        0.36            6.1     0.071
## 13           5.6            0.615        0.00            1.6     0.089
## 14           7.8            0.610        0.29            1.6     0.114
## 15           8.9            0.620        0.18            3.8     0.176
## 16           8.9            0.620        0.19            3.9     0.170
## 17           8.5            0.280        0.56            1.8     0.092
## 18           8.1            0.560        0.28            1.7     0.368
## 19           7.4            0.590        0.08            4.4     0.086
## 20           7.9            0.320        0.51            1.8     0.341
##    free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                   11                   34  0.9978 3.51      0.56     9.4
## 2                   25                   67  0.9968 3.20      0.68     9.8
## 3                   15                   54  0.9970 3.26      0.65     9.8
## 4                   17                   60  0.9980 3.16      0.58     9.8
```

```
## 5                   11                    34  0.9978 3.51      0.56     9.4
## 6                   13                    40  0.9978 3.51      0.56     9.4
## 7                   15                    59  0.9964 3.30      0.46     9.4
## 8                   15                    21  0.9946 3.39      0.47    10.0
## 9                    9                    18  0.9968 3.36      0.57     9.5
## 10                  17                   102  0.9978 3.35      0.80    10.5
## 11                  15                    65  0.9959 3.28      0.54     9.2
## 12                  17                   102  0.9978 3.35      0.80    10.5
## 13                  16                    59  0.9943 3.58      0.52     9.9
## 14                   9                    29  0.9974 3.26      1.56     9.1
## 15                  52                   145  0.9986 3.16      0.88     9.2
## 16                  51                   148  0.9986 3.17      0.93     9.2
## 17                  35                   103  0.9969 3.30      0.75    10.5
## 18                  16                    56  0.9968 3.11      1.28     9.3
## 19                   6                    29  0.9974 3.38      0.50     9.0
## 20                  17                    56  0.9969 3.04      1.08     9.2
##    quality
## 1        5
## 2        5
## 3        5
## 4        6
## 5        5
## 6        5
## 7        5
## 8        7
## 9        7
## 10       5
## 11       5
## 12       5
## 13       5
## 14       5
## 15       5
## 16       5
## 17       7
## 18       5
## 19       4
## 20       6
```

```r
dFDensityPH<-data.frame(winequalRed$density, winequalRed$pH)
```

(d) Save the `density` and `pH` records of the `winequalRed` data set as their own data frame. (5 points)

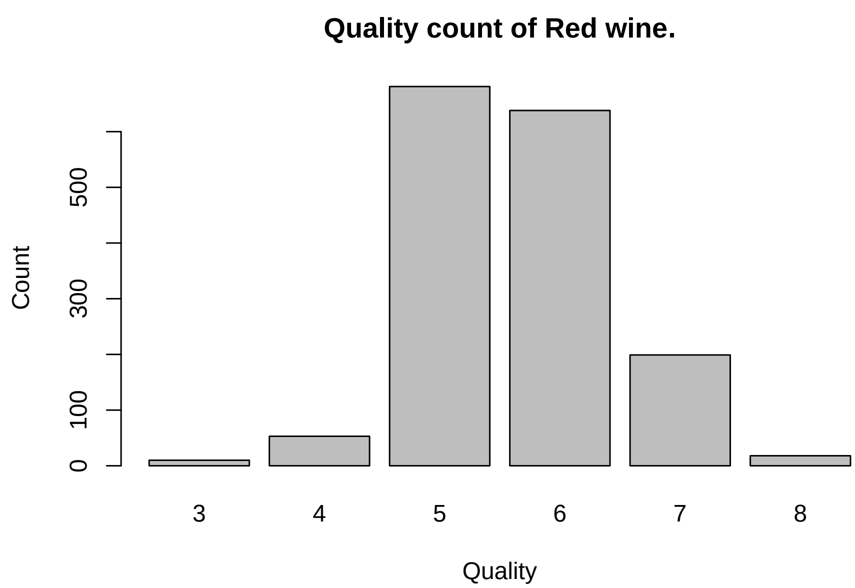
```r
lowQualWine<- subset(winequalRed,winequalRed$quality<6)
highQualWine<- subset(winequalRed,winequalRed$quality>5)
meanAlcoholLowQualWine<-mean(lowQualWine$alcohol)
meanAlcoholHighQualWine<-mean(highQualWine$alcohol)
cat('The mean alcohol level of high quality wine(',meanAlcoholHighQualWine,') is **greater** than low qu
```

(e) Separate the wine data into a low quality class (quality ≤ 5) and a high quality class (quality > 5), find the mean and standard deviation for the attributes `alcohol` for the two classes. Based on the statistical information, describe if there exists difference for `alcohol` between the low quality and high quality red wines. (10 points)

The mean alcohol level of high quality wine(10.85503) is **greater** than low quality wine(9.926478).

```
tableWineeQualRed<-table(winequalRed$quality)
barplot(tableWineeQualRed,
        main="Quality count of Red wine.",
        xlab="Quality",
        ylab="Count")
```

(f) Use `barplot` to investigate `quality` attribute. Show your result. (5 points)




```
        summary(winequalRed$quality)
```

Min. 1st Qu. Median Mean 3rd Qu. Max. 3.000 5.000 6.000 5.636 6.000 8.000

```
cat('Winequality-red data set has most of the data with the Quality number 5 and 6. Thus the mean quali
```

Winequality-red data set has most of the data with the Quality number 5 and 6. Thus the mean quality of data set is 5.636 and median of the data set is 6.