# Class 12 Homework

## Snehita Vallumchetla(A16853399)

Q13:Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

The sample size for the A/A genotype is 108 individuals with median expression levels of 31.2. The sample size for the A/G genotype is 233 individuals with median expression levels of 25.1. The sample size for the G/G genotype is 121 individuals with median expression levels of 20.1.

(my work is shown below)

```
df <- read.table("expression.txt", header = TRUE, sep = "", stringsAsFactors = FALSE)
head(df)
```

```
   sample geno      exp
1 HG00367  A/G 28.96038
2 NA20768  A/G 20.24449
3 HG00361  A/A 31.32628
4 HG00135  A/A 34.11169
5 NA18870  G/G 18.25141
6 NA11993  A/A 32.89721
```
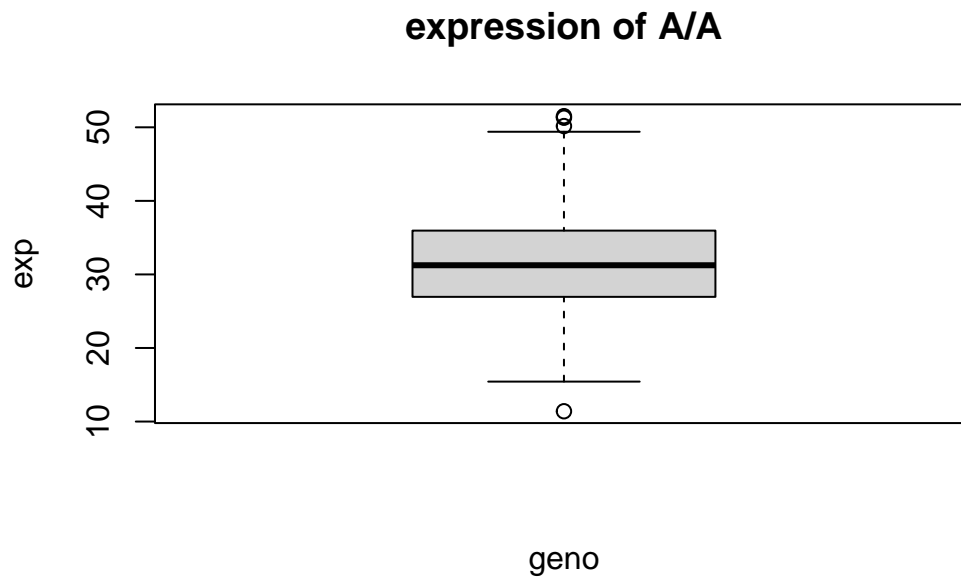
```
summary(df)
```

```
   sample              geno                exp
 Length:462         Length:462         Min.   : 6.675
 Class :character   Class :character   1st Qu.:20.004
 Mode  :character   Mode  :character   Median :25.116
                                       Mean   :25.640
                                       3rd Qu.:30.779
                                       Max.   :51.518
```
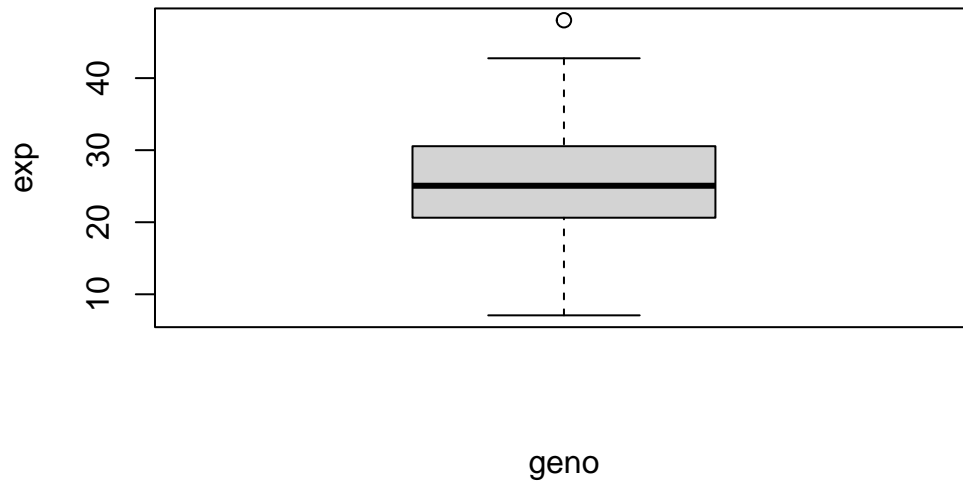
```r
table(df$geno)
```

```
A/A A/G G/G
108 233 121
```

```r
aa <- boxplot(exp ~ geno, data = df[df$geno == "A/A", ], main = 'expression of A/A')
```
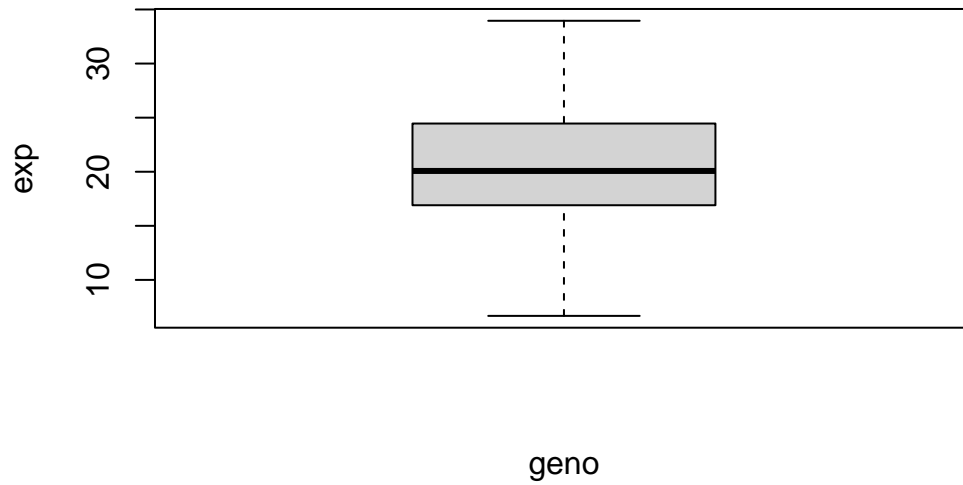
**expression of A/A**



```r
ag <- boxplot(exp ~ geno, data = df[df$geno == "A/G", ], main = 'expression of A/G')
```

## expression of A/G



```r
gg <- boxplot(exp ~ geno, data = df[df$geno == "G/G", ], main = 'expression of G/G')
```

## expression of G/G

```
# The 3rd row of the stats column provides us with the median expression level of each of the
aa$stats
```

```
          [,1]
[1,]  15.42908
[2,]  26.95022
[3,]  31.24847
[4,]  35.95503
[5,]  49.39612
```

```
ag$stats
```

```
          [,1]
[1,]   7.07505
[2,]  20.62572
[3,]  25.06486
[4,]  30.55183
[5,]  42.75662
```

```
gg$stats
```

```
          [,1]
[1,]   6.67482
[2,]  16.90256
[3,]  20.07363
[4,]  24.45672
[5,]  33.95602
```

> Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

The median expression levels of the A/A genotype is greater than the median expression levels of the G/G phenotype. Since there is a difference in median expression levels between the A/A and G/G genotypes, this is indicates that the SNP does effect the expression of ORMDL3, likely lowering its expression.

```
boxplot(exp ~ geno, data = df,
        main = "Expression by Genotype",
        xlab = "Genotype",
        ylab = "Expression",
        col = c("lightpink", "lightgreen", "lightblue"))
```

**Expression by Genotype**