Class 11: Population Analysis HW

Kaitlyn Madriaga, A17217752

Population Scale Analysis

```r
file <- read.table(url("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENS

head(file)
```

```
  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
```

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

```r
summary(file)
```

```
   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
```
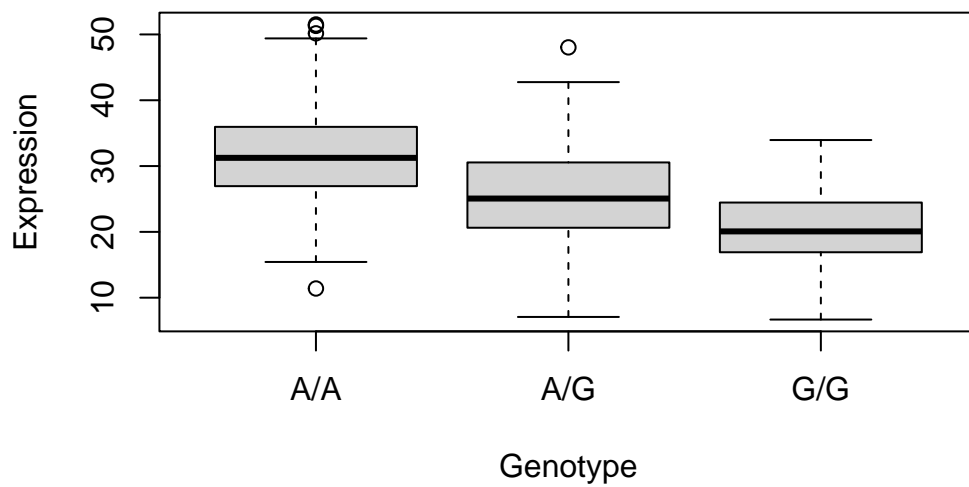
```r
table(file$geno)
```

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

There are 462 samples– 108 A/A, 233 A/G, 121 G/G.

```r
box <- boxplot(file$exp ~ file$geno, xlab = "Genotype", ylab = "Expression")
```

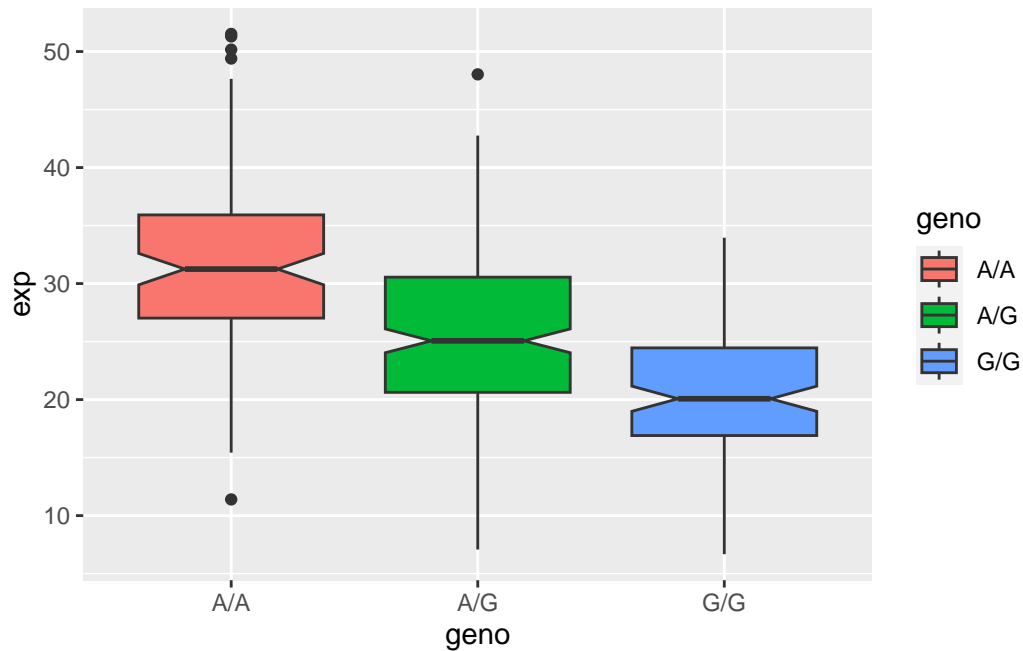

```r
box$stats
```

```
          [,1]     [,2]      [,3]
[1,] 15.42908  7.07505   6.67482
[2,] 26.95022 20.62572  16.90256
[3,] 31.24847 25.06486  20.07363
[4,] 35.95503 30.55183  24.45672
[5,] 49.39612 42.75662  33.95602
```

The median for A/A is 31.2487, the median for A/G is 25.06486, and the median for G/G is 20.07363.

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?

```
library(ggplot2)
```

```
ggplot(file) + aes(geno, exp, fill = geno) +
  geom_boxplot(notch = TRUE)
```


The expression values of A|A and G|G are very different, as their medians/notches do not overlap each other. We can infer that with G|G genotype have a lower expression of the ORMDL3.