# Class 11 Extra Credit

## Max Gruber

## 5/23/23

## Section 4: Population Scale Analysis

**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**

```
genodata <- read.table("datafile.txt")

summary(genodata)
```

```
    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
```

```
table(genodata$geno)
```

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

```
AA <- genodata$exp[genodata$geno == "A/A"]
AG <- genodata$exp[genodata$geno == "A/G"]
GG <- genodata$exp[genodata$geno == "G/G"]
```

```
median(AA)
```
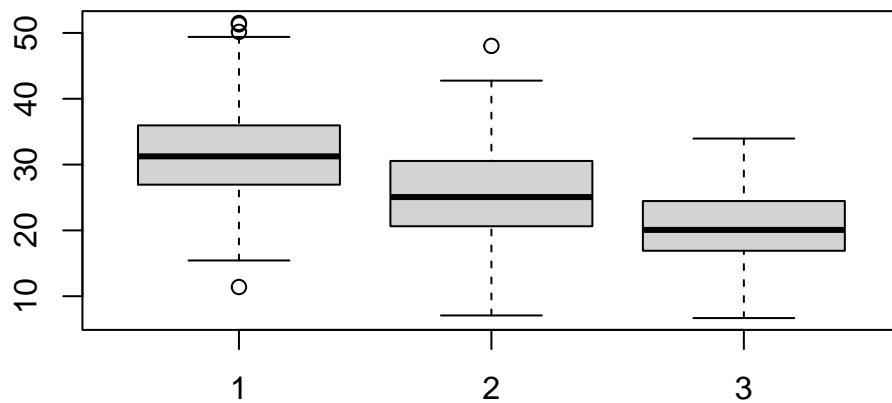
[1] 31.24847

```
median(AG)
```

[1] 25.06486

```
median(GG)
```

[1] 20.07363

The median values for A/A, A/G, and G/G genotypes are 31.248, 25.065, and 20.073, respectively.

**Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression of ORMDL3?**

```
boxplot(AA, AG, GG)
```

Yes, the SNP does affect the expression of ORMDL3, with A/A having the highest expression value followed by A/G and then G/G, respectively.