November 10, 2023 Class 12 Population Scale Analysis

Savannah Bogus A69027475

Population Scale Analysis

```r
data<-read.table("rs8067378_ENSG00000172057.6.txt")
```

```r
summary(data)
```

```
    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(data$geno)
```

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

There are 108 of A/A genotype, 233 of A/G genotype, and 121 of G/G genotype.

```r
library(dplyr)
```

```
Attaching package: 'dplyr'
```
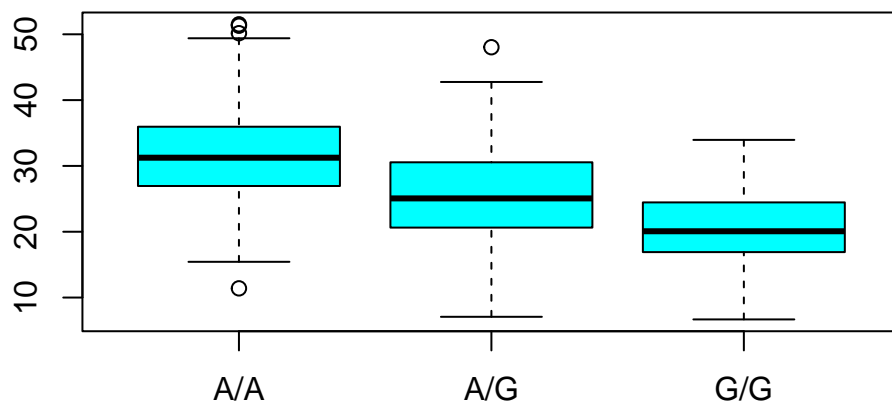
```
The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
AA<-data%>%
  filter(geno=="A/A")

AG<-data%>%
  filter(geno=="A/G")
GG<-data%>%
  filter(geno=="G/G")

boxplot(AA$exp,AG$exp,GG$exp,names=c("A/A","A/G","G/G"),col="cyan")
```



```r
median(AA$exp)
```

```
[1] 31.24847
```

```
median(AG$exp)
```

[1] 25.06486

```
median(GG$exp)
```

[1] 20.07363

The median expression levels for the A/A genotype is 31.24847, 25.06486 for A/G, and 20.07363 for the G/G genotype. Based on my boxplot, that looks correct.

Q14

This SNP does affect the expression of ORMDL3, with A/A having a higher expression level, G/G having a lower expression level, and A/G being somewhere in between.