Class 10: Population Analysis

Divya Shetty (A15390408)

2/17/2022

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.

```
#read text file
exp <- read.table("rs8067378_ENSG00000172057.6.txt")
head(exp)
```

```
##    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
```

What are the sample sizes for the different genotypes?

```
table(exp$geno)
```

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

There are 108 samples for A/A, 233 samples for A/G, and 121 samples for G/G.

What are the median expression levels for the different genotypes?

```
#genotype: A/A
summary(exp[exp$geno == "A/A", "exp"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.40   27.02   31.25   31.82   35.92   51.52
```

```
#genotype: A/G
summary(exp[exp$geno == "A/G", "exp"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.075  20.626  25.065  25.397  30.552  48.034
```

```
#genotype: G/G
summary(exp[exp$geno == "G/G", "exp"])
```

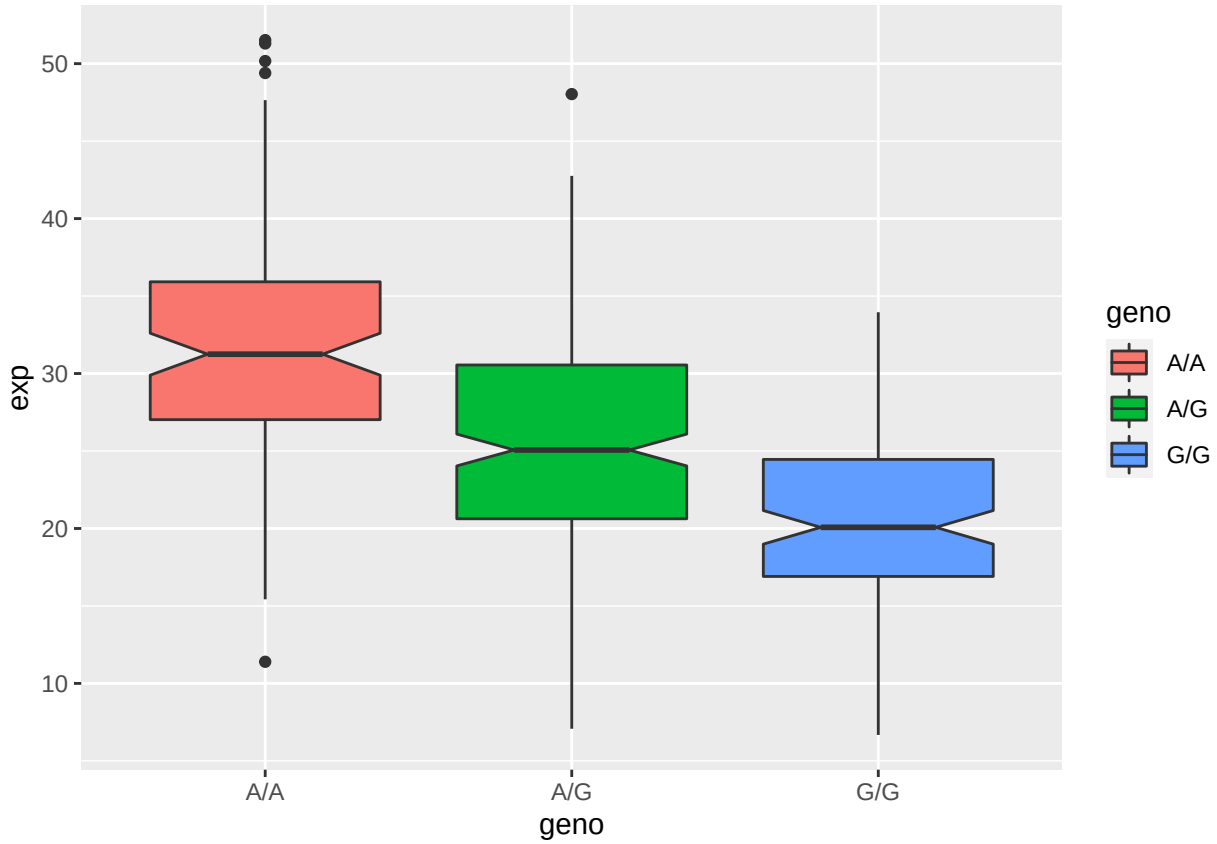
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.675  16.903  20.074  20.594  24.457  33.956
```

The median expression level for A/A is 31.25 and the mean is 31.82. The median expression level for A/G is 25.065 and the mean is 25.397. The median expression level for G/G is 20.074 and the mean is 20.0594.

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)

#make a boxplot
ggplot(exp, aes(x = geno, y = exp, fill = geno)) +
  geom_boxplot(notch = TRUE)
```



There is a noticeable difference in the median level of expression of ORMDL3 between A/A and G/G, where the median expression level for A/A is higher than that of G/G. This would indicate that the SNP is associated with the expression of ORMDL3.