

## class 12

Gaeun Jun (A16814573)

2024-11-11

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.

```
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
head(expr)
```

```
##      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
```

```
nrow(expr)
```

```
## [1] 462
```

There are 462 individuals.

```
table(expr$geno)
```

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

The sample size for the A/A genotype is 108, for the A/G genotype is 233, and for the G/G genotype is 121.

```
data_AG <- expr[expr$geno == "A/G", ]
```

```
median_AG <- median(data_AG$exp, na.rm = TRUE)
```

```
print(median_AG)
```

```
## [1] 25.06486
```

The median expression levels of A/G is 25.06.

```
data_GG <- expr[expr$geno == "G/G", ]
```

```
median_GG <- median(data_GG$exp, na.rm = TRUE)
```

```
print(median_GG)
```

```
## [1] 20.07363
```

The median expression levels of G/G is 20.07.

```
data_AA <- expr[expr$geno == "A/A", ]
median_AA <- median(data_AA$exp, na.rm = TRUE)
print(median_AA)
```

```
## [1] 31.24847
```

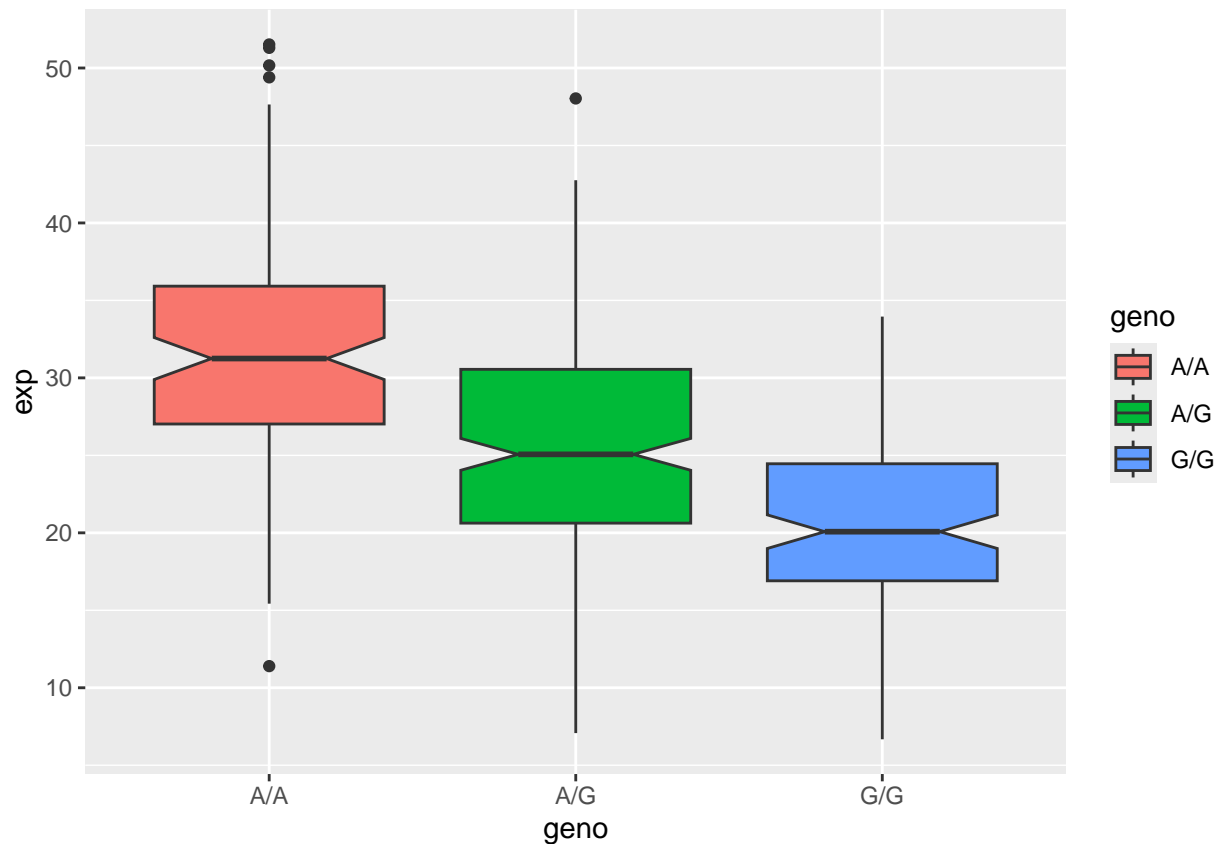
The median expression levels of A/A is 31.25.

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 ORM DL3?

```
library(ggplot2)
```

Let's make a boxplot.

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



The expression values of A/A are greater overall than the expression values of G/G. Since the median expression of A/A is higher than the median expression of G/G, we can infer that A/A genotype is associated with a higher expression of ORM DL3. Because there is a difference in the distribution, we can also infer that SNP affects the expression of ORM DL3.