

Class 12 Homework Q13 and 14

Sabrina Koldinger A16368238

Homework

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

```
data= read.table("Homeworkdata.txt")
head(data)
```

	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

Find the number of samples for each genotype.

```
table(data$geno)
```

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

Find Median: create df of only genotype and expression level

```
genoexp=data[,c("geno","exp")]
head(genoexp)
```

```
   geno      exp
1  A/G 28.96038
2  A/G 20.24449
3  A/A 31.32628
4  A/A 34.11169
5  G/G 18.25141
6  A/A 32.89721
```

create a subset data frame for the genotypes

```
AA= subset.data.frame(genoexp, subset = geno <="A/A")
median(AA$exp)
```

```
[1] 31.24847
```

make it a function

```
medians= function(x){y=subset.data.frame(genoexp, subset = geno <=x)
median(y$exp)}
medians("G/G")
```

```
[1] 25.11561
```

```
medians("A/G")
```

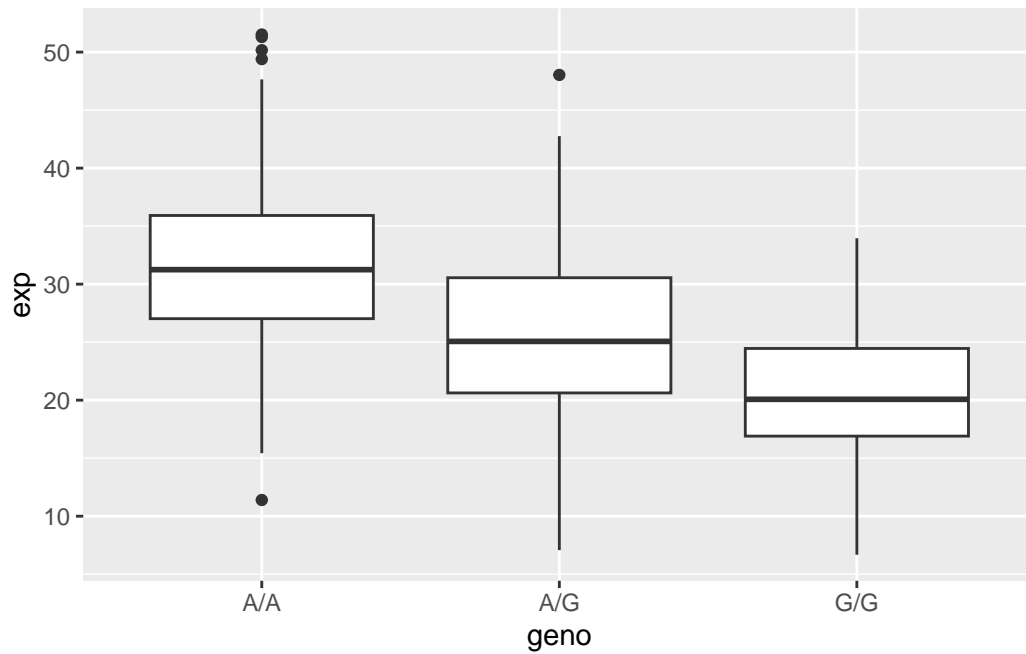
```
[1] 27.43637
```

The sample sizes for each genotype are 108 for AA, 233 for AG, and 121 for GG. The median expression level for each genotype is 31.25 for AA, 27.44 for AG, and 25.12 for GG.

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(data)+ aes(x=geno, y=exp)+ geom_boxplot()
```



It can be assumed that ORMDL3 gene will be expressed more with an A for the SNP in comparison to having a G at the same position. AA is expressed more than GG, so yes the SNP has an effect on expression. If you have one allele that has the G, it will decrease the expression level of the ORMDL3 gene.