Lab11HW

AUTHOR

Joe Kesler

> 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.

First, I will bring in the table

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

```
    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
```

Now I will determine the sample size.

```
summary(rs)
```

```
    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
```
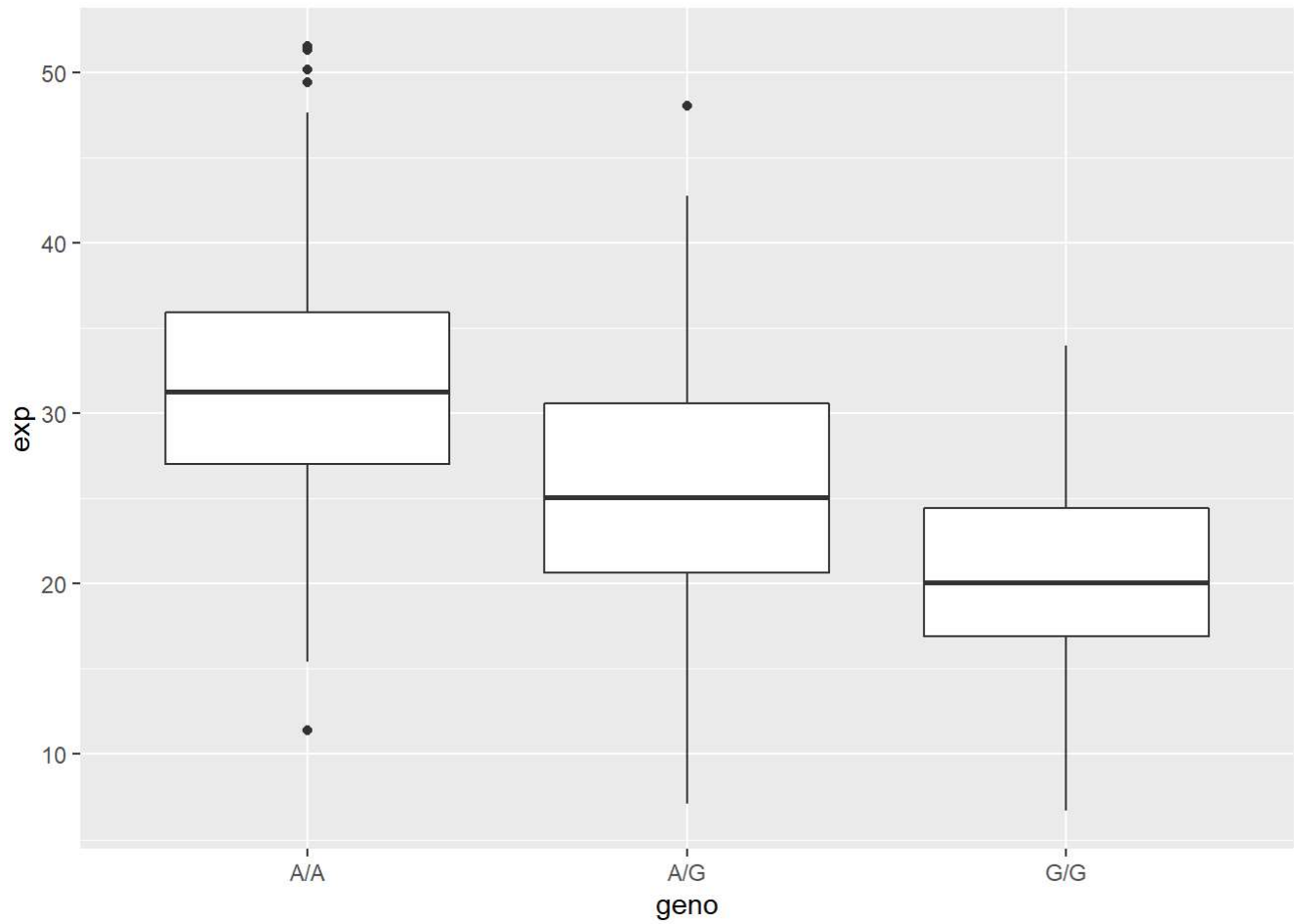
The samples size is for each genotype is 462 and the median is 25.116

> 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)
```

```
Warning: package 'ggplot2' was built under R version 4.1.3
```

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

Those that are A/A clearly express more than G/G judging by the boxplot, so I would conclude that ORMDL3 does affect expression