

class11

Kalle Liimatta A59002114

Table of contents

Section 4 : Population Scale Analysis HOMEWORK	1
Q1 3 : Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.	1
The sample size for genotype A/A, A/G, and G/G are 108, 233, and 121, respectively.	2
The median expression levels for A/A, A/G, and G/G are 31.2, 25.1, and 20.1, respectively.	4
Q1 4 : 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?	4
Boxplot:	4
A/A is associated with significantly higher expression of ORMDL3. The SNP does affect the expression of ORMDL3.	4

Section 4 : Population Scale Analysis HOMEWORK

Q1 3 : Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

How many samples do we have?

```
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
```

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

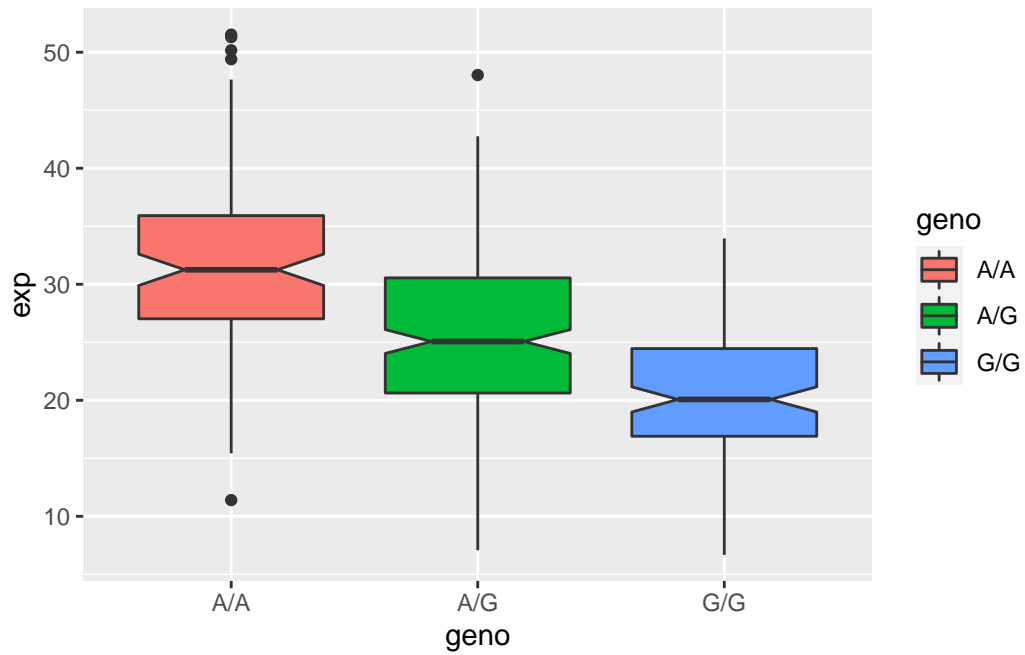
```
A/A A/G G/G
108 233 121
```

The sample size for genotype A/A, A/G, and G/G are 108, 233, and 121, respectively.

```
library(ggplot2)
```

let's make a boxplot

```
box <- ggplot(expr) + aes(x=geno, y=exp, fill=geno) +
  geom_boxplot(notch=TRUE)
box
```



```
AA <- subset(expr, geno=="A/A")  
median(AA$exp)
```

```
[1] 31.24847
```

```
AG <- subset(expr, geno=="A/G")  
median(AG$exp)
```

```
[1] 25.06486
```

```
GG <- subset(expr, geno=="G/G")  
median(GG$exp)
```

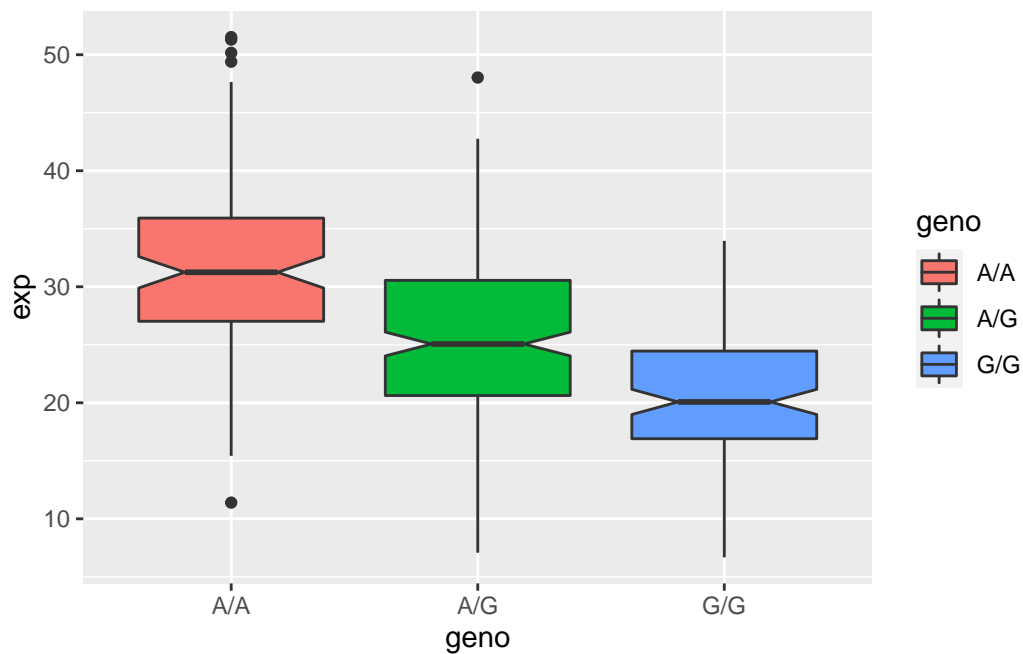
```
[1] 20.07363
```

The median expression levels for A/A, A/G, and G/G are 31.2, 25.1, and 20.1, respectively.

Q1 4 : 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?

Boxplot:

box



A/A is associated with significantly higher expression of ORMDL3. The SNP does affect the expression of ORMDL3.