

Class11

Max Strul

Table of contents

#Data

```
url <- "https://bioboot.github.io/bgggn213_W19/class-material/rs8067378_ENSG000000172057.6.t
data <- read.table(url)
```

#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.

```
#factor(data[,2])
ag <- sum(data[,2]=="A/G")
gg <- sum(data[,2]=="G/G")
aa <- sum(data[,2]=="A/A")
ag
```

```
[1] 233
```

```
gg
```

```
[1] 121
```

```
aa
```

```
[1] 108
```

```

ag_exp <- data$exp[data[,2]=="A/G"]
gg_exp <- data$exp[data[,2]=="G/G"]
aa_exp <- data$exp[data[,2]=="A/A"]
median(ag_exp)

```

```
[1] 25.06486
```

```
median(gg_exp)
```

```
[1] 20.07363
```

```
median(aa_exp)
```

```
[1] 31.24847
```

```
summary(data)
```

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

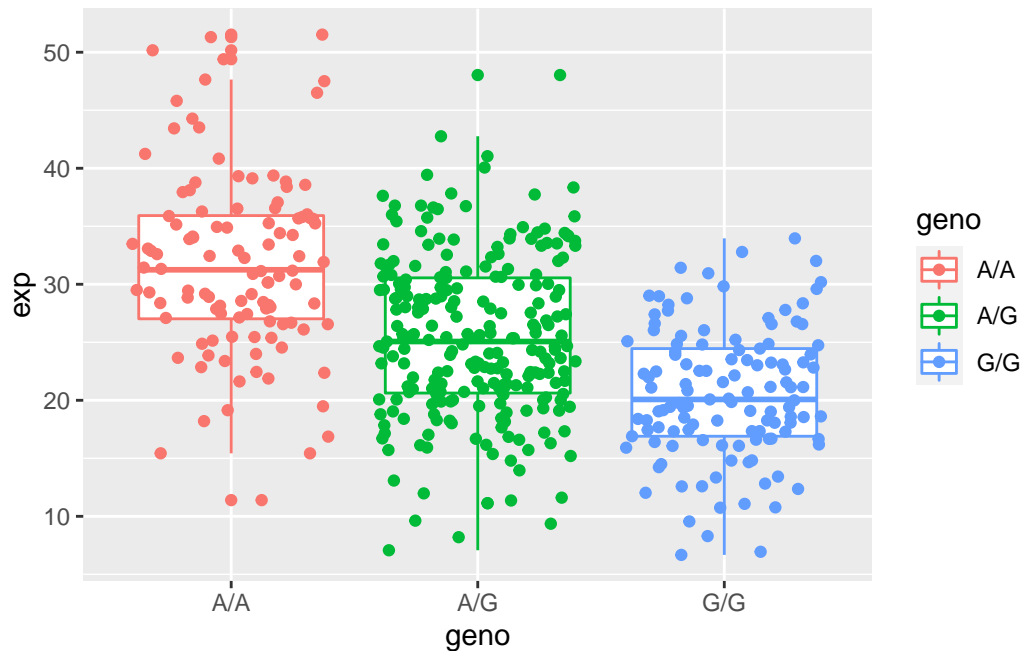
#Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

```

library("ggplot2")

ggplot(data)+
  aes(geno,exp,col=geno)+
  geom_boxplot()+
  geom_point(position="jitter")

```



```
t.test(data$exp[data$geno=="A/G"],data$exp[data$geno=="A/A"])
```

Welch Two Sample t-test

```
data: data$exp[data$geno == "A/G"] and data$exp[data$geno == "A/A"]
t = -7.2293, df = 187.27, p-value = 1.198e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.174207 -4.669471
sample estimates:
mean of x mean of y
 25.39680  31.81864
```

```
t.test(data$exp[data$geno=="A/G"],data$exp[data$geno=="G/G"])
```

Welch Two Sample t-test

```
data: data$exp[data$geno == "A/G"] and data$exp[data$geno == "G/G"]
```

```
t = 6.9667, df = 290.79, p-value = 2.172e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.446174 6.160009
sample estimates:
mean of x mean of y
25.39680 20.59371
```

```
t.test(data$exp[data$geno=="G/G"],data$exp[data$geno=="A/A"])
```

Welch Two Sample t-test

```
data: data$exp[data$geno == "G/G"] and data$exp[data$geno == "A/A"]
t = -12.214, df = 191.65, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-13.037619 -9.412243
sample estimates:
mean of x mean of y
20.59371 31.81864
```

All groups are statistically significantly different from one another and it does appear that an SNP can alter expression levels of (ORMDL3)