

Class12: Population Scale Analysis

Richard Gao - PID:A16490010

2024-02-18

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("ExpressionResults.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
```

```
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
```

```
# Determines sample size for each genotype
table(data$geno)
```

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

```
# Corresponding median expression levels for each genotype
median(data$exp[data$geno == "A/A"])
```

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

```
median(data$exp[data$geno == "A/G"])
```

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

```
median(data$exp[data$geno == "G/G"])
```

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

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)
```

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

