Class 11 Homework

Izabelle Querubin

2023-05-28

[Extra Credit] Homework Class #11 (Population Analysis)

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.

```
# Read the file into a data frame
ormdl3 <- read.table("ORMDL3.txt", header = TRUE)

# Calculate sample size for each genotype
sample_sizes <- table(ormdl3$geno)

# Calculate median expression levels for each genotype
medians <- tapply(ormdl3$exp, ormdl3$geno, median)

# Display the sample sizes and median expression levels
cat("Sample sizes for each genotype:\n")
```

```
## Sample sizes for each genotype:
```

```
print(sample_sizes)
```

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

```
cat("\nMedian expression levels for each genotype:\n")
```

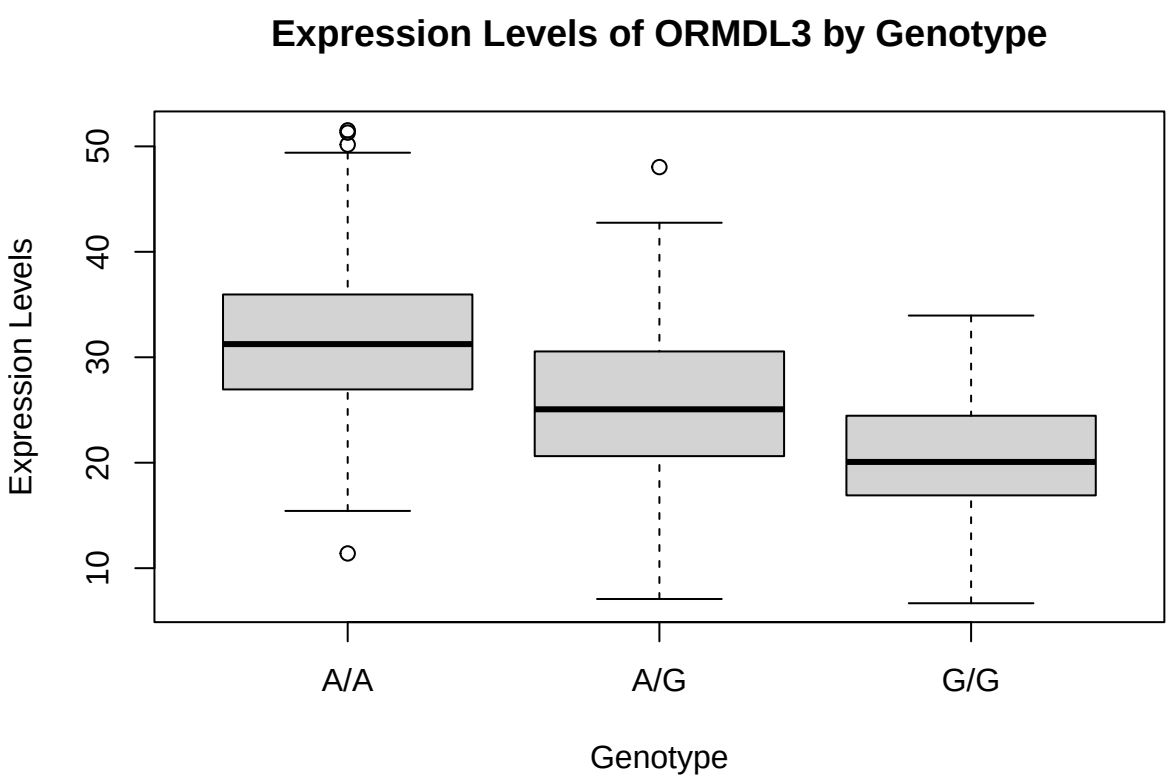
```
##
## Median expression levels for each genotype:
```

```
print(medians)
```

```
##      A/A      A/G      G/G
## 31.24847 25.06486 20.07363
```

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?

```
# Create boxplots for each genotype
boxplot(formula = exp ~ geno, data = ormdl3,
        xlab = "Genotype", ylab = "Expression Levels",
        main = "Expression Levels of ORMDL3 by Genotype")
```



Based on the boxplot, A/A (31.2) has a relatively higher median expression value than G/G (21.1). The difference in expression levels can suggest that the SNP (single nucleotide polymorphism) has an effect on the expression of ORMDL3 but more statistical analysis is required to make a verdict, like hypothesis testing.