

# Class 11: Population Scale Analysis

Eric Jordahl

2022-11-2

## Table of contents

Population Scale Analysis . . . . .	1
<b>Q13</b>	<b>1</b>
<b>Q14</b>	<b>2</b>

## Population Scale 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

First we need to read the data provided

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

```
table(data$geno)
```

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

## 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, fill=geno)) + geom_boxplot(notch=TRUE)
```

