

Class12

Aaron (PID A17544470)

Table of contents

Section 1. Proportion of G/G in population	1
Section 4. Population Scale Analysis	2

Section 1. Proportion of G/G in population

Download a CSV file from Ensemble.

Here we read this CSV file

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(mxl)
```

	Sample..	Male.	Female.	Unknown.	Genotype..	forward.	strand.	Population.s.	Father
1					NA19648	(F)		A A ALL, AMR, MXL	-
2					NA19649	(M)		G G ALL, AMR, MXL	-
3					NA19651	(F)		A A ALL, AMR, MXL	-
4					NA19652	(M)		G G ALL, AMR, MXL	-
5					NA19654	(F)		G G ALL, AMR, MXL	-
6					NA19655	(M)		A G ALL, AMR, MXL	-
	Mother								
1		-							
2		-							
3		-							
4		-							
5		-							
6		-							

```
table(mx1$Genotype..forward.strand.) / nrow(mx1) * 100
```

```

      A|A      A|G      G|A      G|G
34.3750 32.8125 18.7500 14.0625

```

Section 4. 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. Hint: The `read.table()`, `summary()` and `boxplot()` functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the `boxplot()` function to an R object and examining this object. There is also the `medium()` and `summary()` function that you can use to check your understanding.

There is 108 for A|A, 233 for A|G, and 121 for G|G

```
url <- read.table("rs8067378_ENSG00000172057.6.txt")
```

```
head(url)
```

```

      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(url)
```

```

      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

```

```
nrow(url)
```

```
[1] 462
```

```
table(url$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(url) + aes(geno, exp, fill = geno) +  
  geom_boxplot(noth = T)
```

Warning in `geom_boxplot(noth = T)`: Ignoring unknown parameters: `noth`

