

class11

Steven Gan

2022-11-02

Table of contents

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

Section 1. Proportion of G/G in a population

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv",  
               row.names = 1)  
head(mxl)
```

	Genotype..forward.strand.	Population.s.	Father	Mother
NA19648 (F)	A A	ALL, AMR, MXL	-	-
NA19649 (M)	G G	ALL, AMR, MXL	-	-
NA19651 (F)	A A	ALL, AMR, MXL	-	-
NA19652 (M)	G G	ALL, AMR, MXL	-	-
NA19654 (F)	G G	ALL, AMR, MXL	-	-
NA19655 (M)	A G	ALL, AMR, MXL	-	-

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

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

Section 4. Population Scale Analysis

```
expr <- read.table(file = "https://bioboot.github.io/bgg213_W19/class-material/rs8067378_
row.names(expr) = expr$sample
expr <- expr[-1]
```

Q13:

```
table(expr$geno)
```

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

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
```

```
expr %>%
  group_by(geno) %>%
  summarise(Mean = mean(exp), Median = median(exp), Std = sd(exp))
```

```
# A tibble: 3 x 4
  geno   Mean Median   Std
<chr> <dbl>  <dbl> <dbl>
1 A/A    31.8    31.2  7.91
2 A/G    25.4    25.1  6.99
3 G/G    20.6    20.1  5.67
```

Q14:

```
library(ggplot2)
```

```
p <- ggplot(expr) + aes(geno, exp, fill = geno) +  
  geom_boxplot(notch = T) +  
  geom_jitter(alpha = 0.2, width = 0.2) +  
  labs(x = "Genotype", y = "Expression")
```

```
# install.packages("ggpubr")  
library(ggpubr)
```

```
compare_means(exp ~ geno, data = expr)
```

```
# A tibble: 3 x 8
```

	.y.	group1	group2	p	p.adj	p.format	p.signif	method
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>
1	exp	A/G	A/A	1.09e-11	2.2e-11	1.1e-11	****	Wilcoxon
2	exp	A/G	G/G	6.68e-10	6.7e-10	6.7e-10	****	Wilcoxon
3	exp	A/A	G/G	1.99e-23	6e-23	< 2e-16	****	Wilcoxon

```
comp <- list(c("A/A", "A/G"), c("A/G", "G/G"), c("A/A", "G/G"))  
p + stat_compare_means(comparisons = comp)
```

