

# Tarea3

*Silvia Rodríguez*

*14 de noviembre de 2018*

## Ejercicio 1

### Parte a)

```
library(rmarkdown)
library(tidyverse)
library(dplyr)
library(tidyr)
library(tibble)

genes<- read.table("//164.73.246.107/grupos/SERIES/SILVIA/Maestria_Y0/Reproducibilidad/Tarea 3/datos_rn

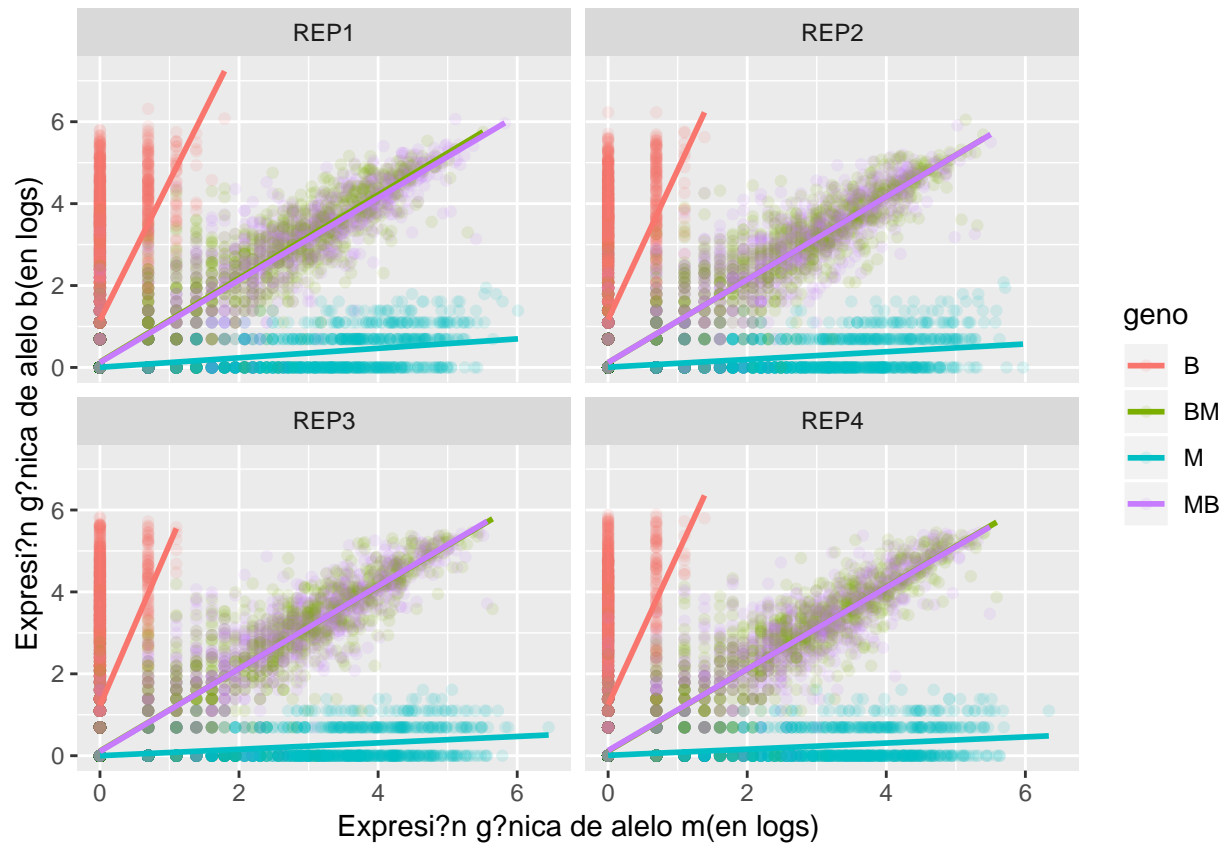
genest<-as_tibble(genes) %>%
  gather(key=nacho, value=silvia, -GeneID) %>%
  separate(col=nacho , into= c("Rep","geno","bmttotal"),remove=TRUE) %>%
  spread(key=bmttotal,value=silvia)
```

### Parte b

No puedo hacer que s?lo grafique la recta de regresi?n que pase por la mayor nube de puntos correspondiente a geno=BM.

```
library(ggplot2)

genest %>%
  mutate(b1=b+1,m1=m+1) %>%
  mutate (logb=log(b1)) %>%
  mutate (logm=log(m1)) %>%
  ggplot(aes(logm,logb, color=geno))+
  geom_point(alpha=1/8)+
  geom_smooth(method =lm, se=FALSE)+
  facet_wrap(~ Rep)+
  labs(x="Expresi?n g?nica de alelo m(en logs)", y = "Expresi?n g?nica de alelo b(en logs)" )+
  theme(legend.position="right")
```



## Ejercicio 3

Preliminares:

```
compara <- function(x, y) {
  j <- length(x)
  n <- length(y)
  # calculo el estadistico de la prueba
  sp <- sqrt(((j-1)*sd(x)^2 + (n-1)*sd(y)^2) / (j+n-2))
  tstat <- (mean(x) - mean(y)) / (sp*sqrt(1/j + 1/n))
  # calculo el p-valor
  2*(1 - pt( abs(tstat), df = n+j-2) )
}
```

```
genest %>%
  filter(geno == 'BM') %>%
  summarize(compara(b,m))
```

```
## # A tibble: 1 x 1
##   `compara(b, m)`
##   <dbl>
## 1         0
```

```
genest %>%
  filter(geno == 'BM') %>%
```

```

    group_by(GeneID) %>%
    summarize( pv=compara(b,m))%>%
    ungroup()%>%
    summarize(SS=sum(is.nan(pv)))

```

```

## # A tibble: 1 x 1
##       SS
##   <int>
## 1  1137

```

## Parte b

Se espera concluir err?neamente, que la diferencia es significativa con  $\alpha=0.05$  en  $0.05N=0.051363=68.15$

## Parte c

Creo la variable indicatriz “veo” que me indica cuando el pvalor es menor a 0.05 para poder responder. Pero no puedo sumar esos unos.

```

genest %>%
  filter(geno == 'BM') %>%
  group_by(GeneID) %>%
  summarize( pv=compara(b,m))%>%
  arrange(desc(pv))

```

```

## # A tibble: 2,500 x 2
##   GeneID      pv
##   <fct>    <dbl>
## 1 AC198403.3_FG001 1.00
## 2 AC199384.3_FG002 1.00
## 3 AC202435.3_FG001 1.00
## 4 AC204868.3_FG004 1.00
## 5 AC234521.1_FG004 1.00
## 6 GRMZM2G000586    1.00
## 7 GRMZM2G010505    1.00
## 8 GRMZM2G015925    1.00
## 9 GRMZM2G016922    1.00
## 10 GRMZM2G020574    1.00
## # ... with 2,490 more rows

```

```

genest %>%
  filter(geno == 'BM') %>%
  group_by(GeneID) %>%
  summarize( pv=compara(b,m))%>%
  arrange(pv) %>%
  mutate(veo=if_else(pv<0.05,1,0, missing=NULL))

```

```

## # A tibble: 2,500 x 3
##   GeneID      pv   veo
##   <fct>    <dbl> <dbl>
## 1 GRMZM2G104419 0.00000708 1.00
## 2 GRMZM2G128922 0.00000837 1.00
## 3 GRMZM2G094666 0.0000128 1.00

```

```
## 4 GRMZM2G129031 0.0000128 1.00
## 5 GRMZM5G885061 0.0000144 1.00
## 6 GRMZM2G425377 0.0000165 1.00
## 7 GRMZM2G058655 0.0000167 1.00
## 8 GRMZM5G842695 0.0000200 1.00
## 9 GRMZM2G398055 0.0000258 1.00
## 10 GRMZM2G143791 0.0000280 1.00
## # ... with 2,490 more rows
```

```
#summarise(n=n()) %>%
#mutate(prop=n/sum(veo))
```

## Parte d

Creo la variable indicatriz “veo\_nuevop” que me indica cuando el pvalor ajustado es menor a 0.05 para poder responder. Pero no puedo sumar esos unos.

```
genest %>%
  filter(geno == 'BM') %>%
  group_by(GeneID) %>%
  summarize(pv=compara(b,m)) %>%
  arrange(pv) %>%
  mutate(nuevo_p = p.adjust(pv, method="bonferroni")) %>%
  mutate(veo_nuevop=if_else(nuevo_p<0.05,1,0, missing=NULL))
```

```
## # A tibble: 2,500 x 4
##   GeneID          pv nuevo_p veo_nuevop
##   <fct>          <dbl> <dbl>     <dbl>
## 1 GRMZM2G104419 0.00000708 0.00964     1.00
## 2 GRMZM2G128922 0.00000837 0.0114     1.00
## 3 GRMZM2G094666 0.0000128 0.0174     1.00
## 4 GRMZM2G129031 0.0000128 0.0174     1.00
## 5 GRMZM5G885061 0.0000144 0.0197     1.00
## 6 GRMZM2G425377 0.0000165 0.0225     1.00
## 7 GRMZM2G058655 0.0000167 0.0227     1.00
## 8 GRMZM5G842695 0.0000200 0.0272     1.00
## 9 GRMZM2G398055 0.0000258 0.0352     1.00
## 10 GRMZM2G143791 0.0000280 0.0382     1.00
## # ... with 2,490 more rows
```

```
#summarise(n2=n()) %>%
# mutate(prop2=n2/sum(veo))
```