

control_3_ie_solucion.R

Asus

2019-10-02

```
library(readxl)
library(ggplot2)
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
```

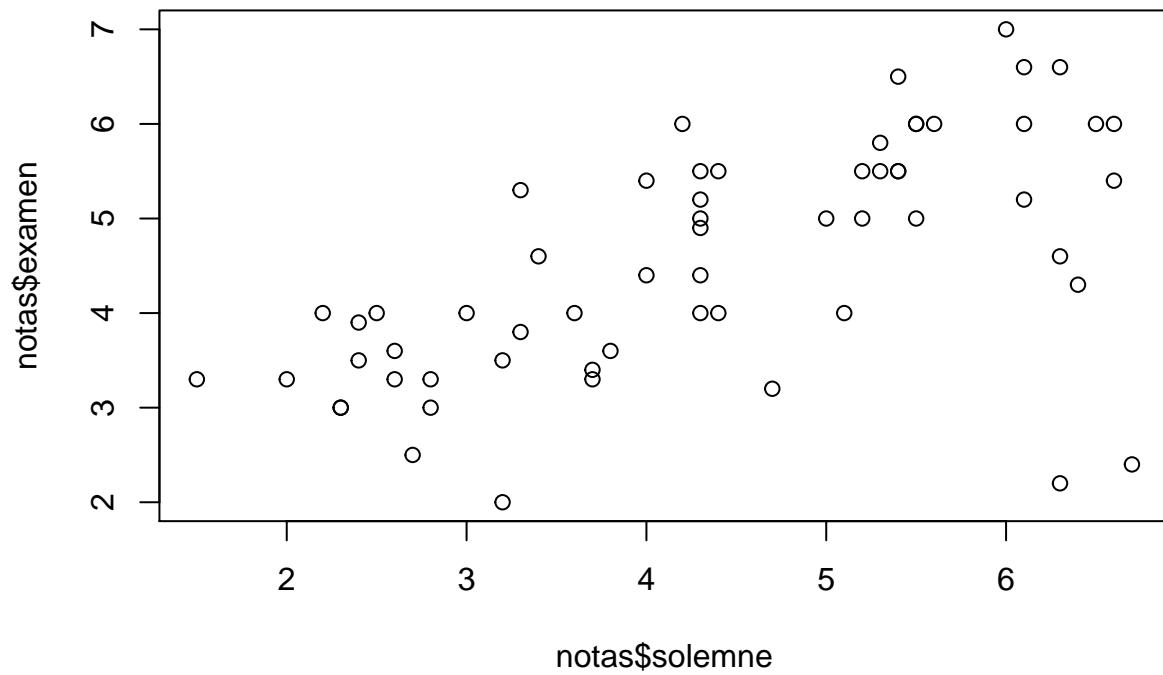
```
# 1) Use la funcion read_excel para importar la base de datos "nombres60.xlsx"
#   guarde el tibble resultante un objeto de nombre "notas"
```

```
# Sol: (no olvidar las comillas y la extensión .xlsx del archivo)
notas <- read_excel("nombres60.xlsx")
notas # no es necesario pero ayuda a entender la solución
```

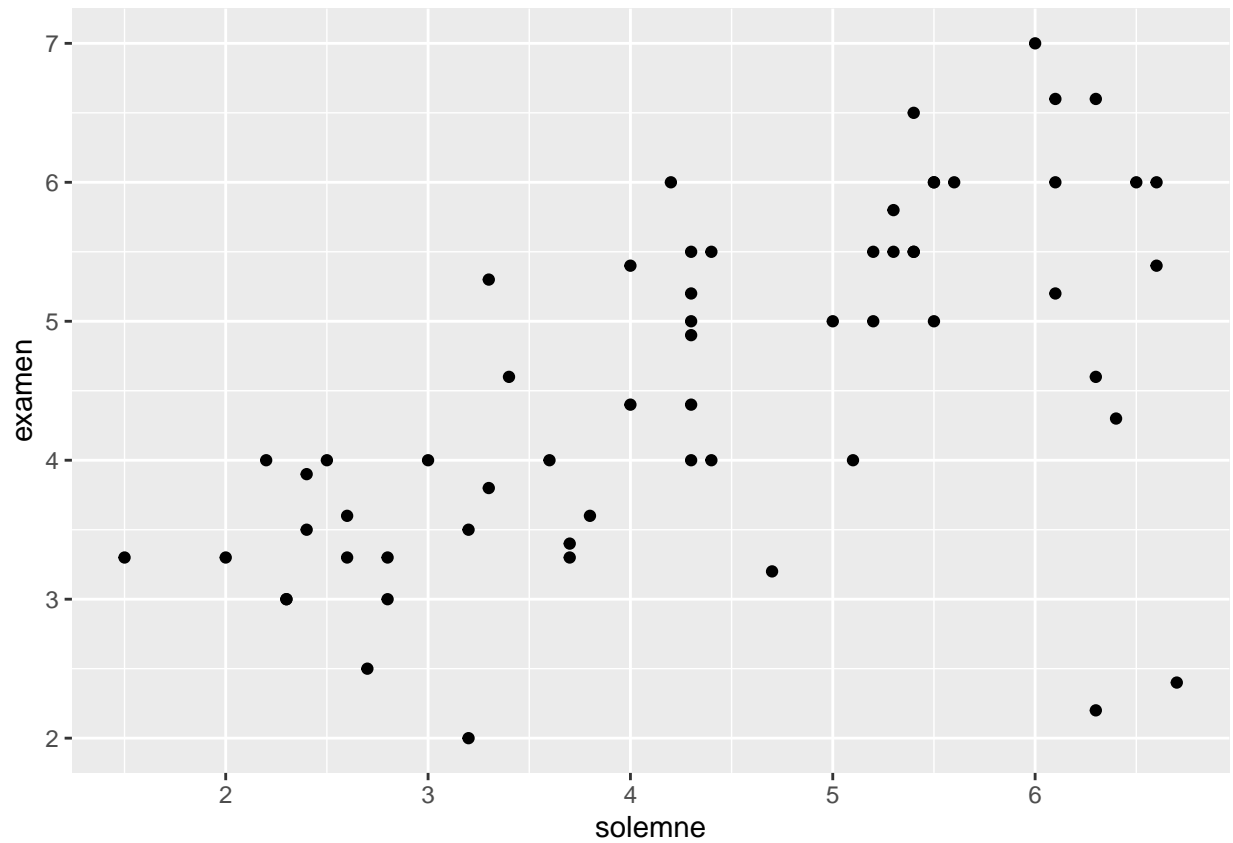
```
## # A tibble: 60 x 4
##   apellido sexo   soleme examen
##   <chr>    <chr>   <dbl>  <dbl>
## 1 Mazzanti mujer    5.1    4
## 2 Naccarato mujer    4.3    5.5
## 3 Morel    hombre    5.2    5
## 4 Vacca    hombre    4.2    6
## 5 Sartori  mujer    2.3    3
## 6 Schiavi  mujer    3.2    2
## 7 Beta     hombre    6.6    5.4
## 8 Zunino   mujer    6.4    4.3
## 9 Kokokokok mujer    4.7    3.2
## 10 Grosso  hombre    6.7    2.4
## # ... with 50 more rows
```

```
# 2) Cree un diagrama de dispersión, con las notas de la prueba
# soleme en el eje horizontal y las del examen en el eje vertical
# Puede usar los grafico de base o los de ggplot2
```

```
# Sol:
# Opción 1: base. Note que requiere vectores, como p.e. notas$examen
plot(notas$soleme, notas$examen)
```

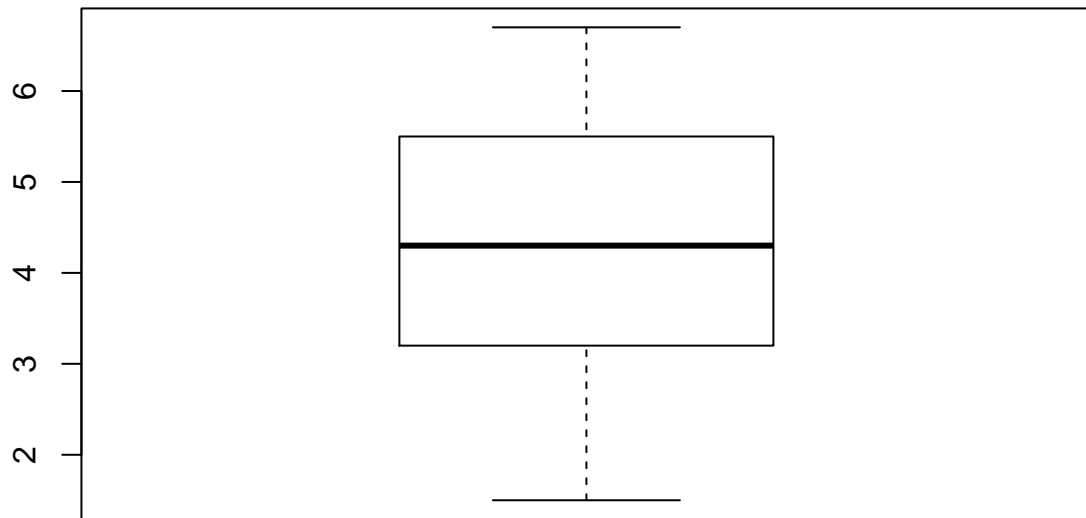


```
# Opción 2: ggplot2. Requiere el data frame y el nombre de las cols  
ggplot(notas) +  
  geom_point(aes(x = solemne, y = examen))
```

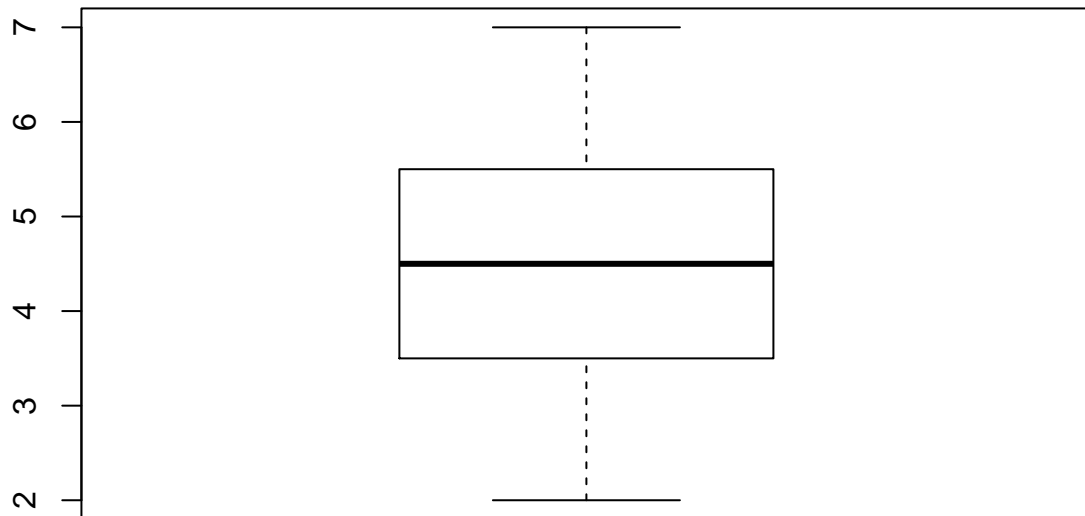


```
# 3) Cree dos diagramas boxplot: uno para las notas de la solemne
# y otro para las notas del examen
# Puede usar los grafico de base o los de ggplot2

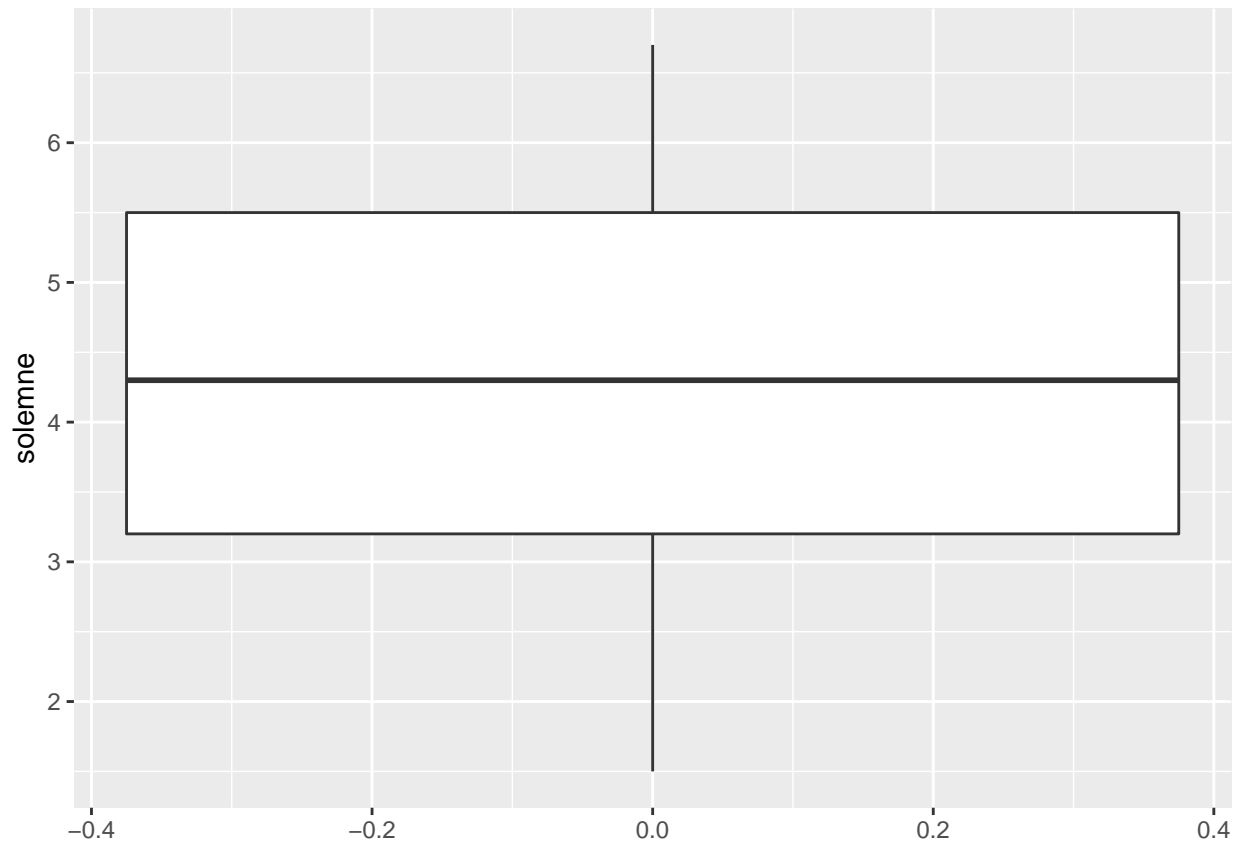
# Sol:
# Opción 1: base. Note que requiere vectores, como p.e. notas$examen
boxplot(notas$solemne)
```



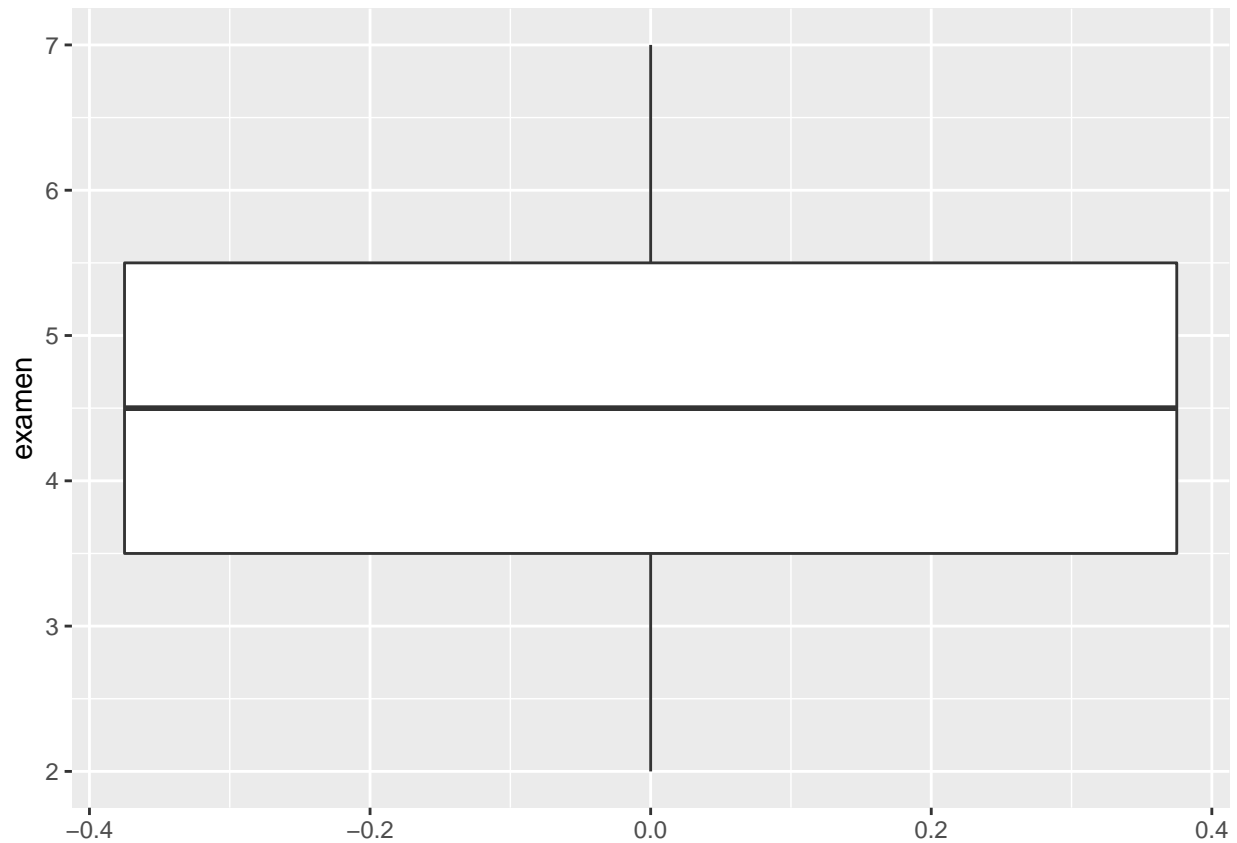
```
boxplot(notas$examen)
```



```
# Opción 2: ggplots. Requiere el data frame y el nombre de las cols  
ggplot(notas) +  
  geom_boxplot(aes(y = solemne))
```



```
ggplot(notas) +  
  geom_boxplot(aes(y = examen))
```



```
# 4) Compute la media de las notas de la prueba solemne y
# la desviacion estándar de las notas de examen
```

```
# Sol:
```

```
# Opción 1, usando vectores
```

```
mean(notas$solemne)
```

```
## [1] 4.37
```

```
sd(notas$examen)
```

```
## [1] 1.234544
```

```
# Opción 2, usando una data frame y summary (base)
```

```
notas %>%
```

```
  summarise(mean_sol = mean(solemne),
```

```
            sd_examen = sd(examen))
```

```
## # A tibble: 1 x 2
```

```
##   mean_sol sd_examen
```

```
##   <dbl>    <dbl>
```

```
## 1     4.37     1.23
```

```
# 5) Compute los quantiles 0.25 y 0.75 de las notas del examen
```

```
# Sol:
```

```
# Opción 1. Usando vectores y quantile
```

```
quantile(notas$examen, 0.25)
```

```
## 25%
```

```
## 3.5
```

```
quantile(notas$examen, 0.75)
```

```
## 75%
```

```
## 5.5
```

```
# Opción 1b. Usando vectores y summary
```

```
summary(notas$examen)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.000   3.500   4.500   4.538   5.500   7.000
```

```
# Opción 2. Usando todo el data frame y summarise
```

```
notas %>%
```

```
  summarise(examen_25 = quantile(examen, 0.25),
            examen_75 = quantile(examen, 0.75))
```

```
## # A tibble: 1 x 2
```

```
##   examen_25 examen_75
```

```
##      <dbl>      <dbl>
```

```
## 1         3.5         5.5
```

```
# Opción 2b. Usando todo el data frame summary (base)
```

```
summary(notas)
```

```
##      apellido      sexo      solemne      examen
## Length:60      Length:60      Min.   :1.50      Min.   :2.000
## Class :character Class :character 1st Qu.:3.20      1st Qu.:3.500
## Mode  :character Mode  :character Median :4.30      Median :4.500
##                                     Mean  :4.37      Mean   :4.538
##                                     3rd Qu.:5.50      3rd Qu.:5.500
##                                     Max.   :6.70      Max.   :7.000
```

```
# 6) Cree una nueva tibble que contenga unicamente las observaciones
```

```
# de las personas que obtuvieron un 4 o más en el examen
```

```
# azul en el examen opcion 1
```

```
# Sol:
```

```
# Opcion 1: Indexado logico (base)
```

```
notas_exazul <- notas[notas$examen >= 4, ]
```

```
# Opcion 2: usando filter (filter)
```

```
notas_exazul <- notas %>% filter(examen >= 4)
```

```
notas_exazul
```



```
## # A tibble: 40 x 4
##   apellido sexo   solemne examen
##   <chr>    <chr>   <dbl> <dbl>
## 1 Mazzanti mujer    5.1    4
## 2 Naccarato mujer    4.3    5.5
## 3 Morel    hombre    5.2    5
## 4 Vacca    hombre    4.2    6
## 5 Beta     hombre    6.6    5.4
## 6 Zunino   mujer    6.4    4.3
## 7 Erti     mujer    3.3    5.3
## 8 Bernardini hombre    5.4    6.5
## 9 Quirino  mujer    6.1    5.2
## 10 Dea     hombre    4      5.4
## # ... with 30 more rows
```

*# No era necesario asignar el resultado a un objeto, pero hace
más corto el próximo código*

7) Compute la proporción de gente que obtuvo un azul en el examen
Sol:
Opción 1, usando la nueva tibble
`nrow(notas_exazul)/nrow(notas)`

```
## [1] 0.6666667
```

Opción 2.a usando el vector de notas examen y la suma de TRUE y FALSE
`sum(notas$examen >= 4)/length(notas$examen)`

```
## [1] 0.6666667
```

Opción 2.b indexando el vector de notas
`length(notas$examen[notas$examen >= 4])/length(notas$examen)`

```
## [1] 0.6666667
```

Opción 3 usando la tibble original
`notas %>%
 summarise(
 n = n(),
 n_azul = sum(examen >= 4),
 proporción = n_azul/n
)`

```
## # A tibble: 1 x 3
##       n n_azul proporción
##   <int> <int>   <dbl>
## 1    60    40    0.667
```

*# 8) Suponga que tiene una población de 10000 observaciones provenientes
de una distribución exponencial con valor esperado igual a 1. Suponga que
de esta población extraemos muestras de tamaño igual a 25 observaciones.*

```
# Cree un histograma de las medias muestrales y calcule la desviación  
# estándar de las medias muestrales. Utilice 500 muestras para el histograma y  
# el cálculo de la desviación estándar.
```

```
# Sol:
```

```
# Adaptando el código de la Guía 2, para el caso de  
# 500 muestras, de tamaños muestrales 25 y datos de una exponencial  
# con valor esperado 1
```

```
muestras <- 500
```

```
n <- 25
```

```
set.seed(12345)
```

```
poblacion <- rexp(10000, rate = 1) # rate es 1/media
```

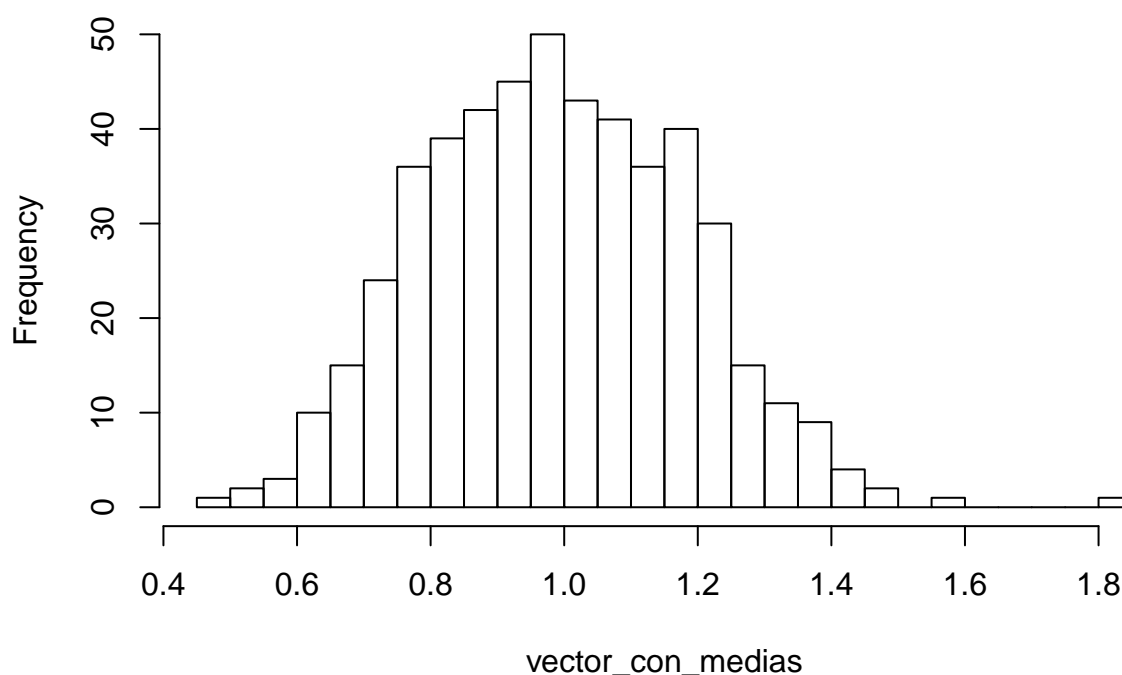
```
media_poblacional <- 1/1 # es 1
```

```
vector_con_medias <- vector(mode = "numeric", length = muestras)
```

```
for (i in 1:muestras) {  
  esta_muestra <- sample(poblacion, size = n)  
  media_muestral <- mean(esta_muestra)  
  # cuando i=1 cambiará el primer componente del vector  
  # cuando i = 2 cambiará el segundo componente del vector  
  vector_con_medias[i] <- media_muestral  
}
```

```
hist(vector_con_medias, breaks = 21)
```

Histogram of vector_con_medias



```
sd(vector_con_medias)
```

```
## [1] 0.1968365
```

*# 9) Repita el ejercicio anterior, pero esta vez tome muestras de tamaño igual 250.
Es ahora la desviación estándar de la media muestral mayor o menor
a la del ejercicio anterior?*

Sol:

*# El código es idéntico al de la pregunta anterior, excepto que
debemos cambiar el tamaño muestral (n) a 250*

```
muestras <- 500
```

```
n <- 250
```

```
set.seed(12345)
```

```
poblacion <- rexp(10000, rate = 1) # rate es 1/media
```

```
media_poblacional <- 1/1 # es 1
```

```
vector_con_medias <- vector(mode = "numeric", length = muestras)
```

```
for (i in 1:muestras) {  
  esta_muestra <- sample(poblacion, size = n)  
  media_muestral <- mean(esta_muestra)  
  # cuando i=1 cambiará el primer componente del vector
```

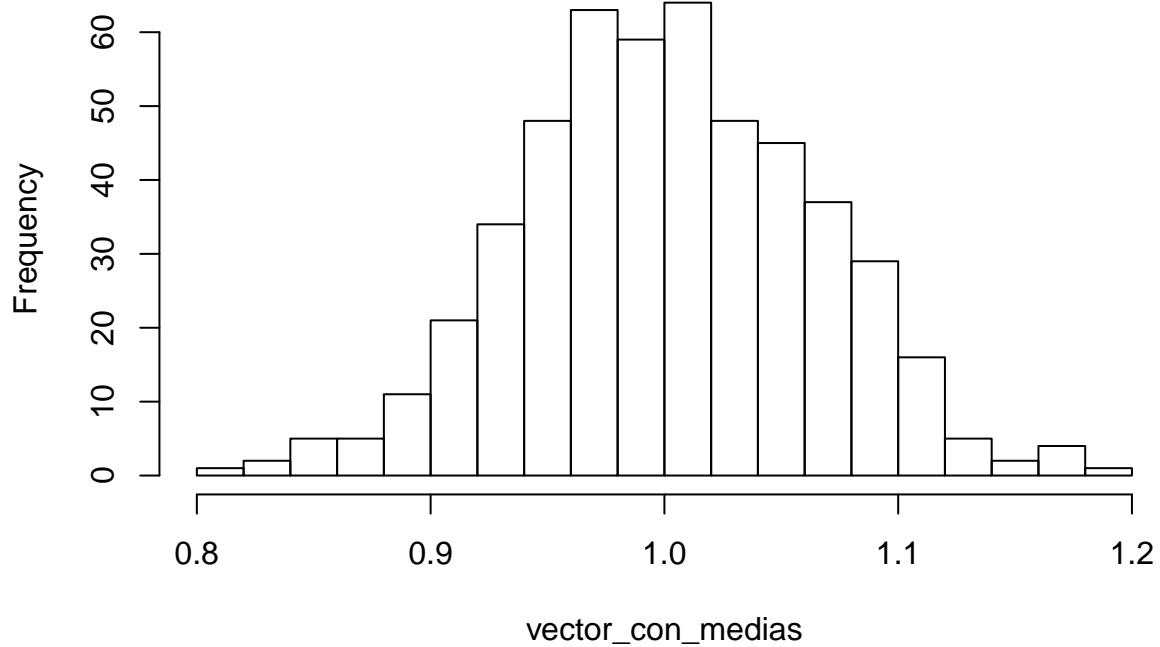
```

# cuando i = 2 cambiará el segundo componente del vector
vector_con_medias[i] <- media_muestral
}

hist(vector_con_medias, breaks = 21)

```

Histogram of vector_con_medias



```
sd(vector_con_medias)
```

```
## [1] 0.06320894
```

```

# Con tamaño muestrade 25, la desviación estándar de las medias
# muestrales era igual a 0.197, mientras que con tamaño muestral de 250,
# la misma desviación estándar es de 0.063, claramente menor que la primera.

```