

# control\_3\_ie\_Agustin\_Villalba.R

*Asus*

*2019-10-17*

```
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
```

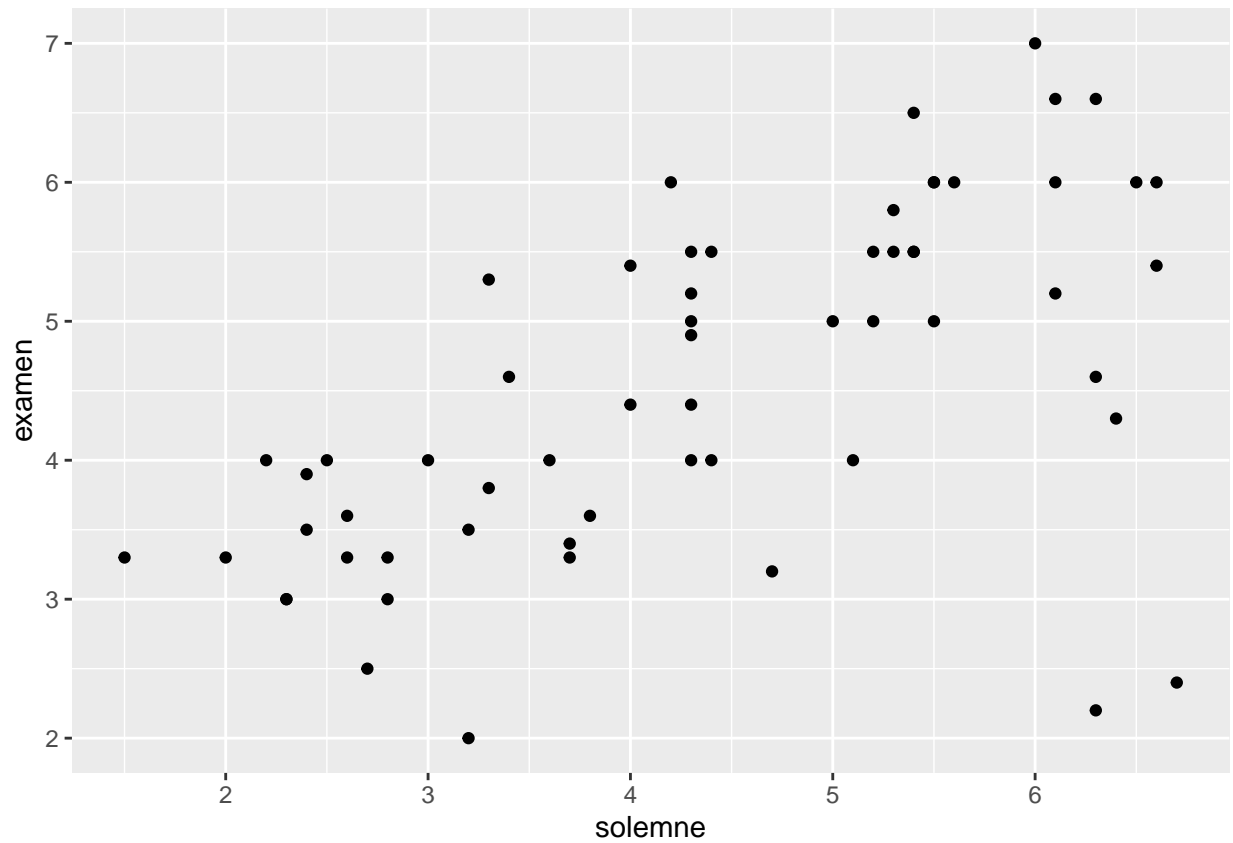
```
library(dplyr)
# View(a)
```

- 1) Use la funcion read\_excel para importar la base de datos “nombres60.xlsx” guarde el tibble resultante un objeto de nombre “notas”

```
notas <- read_excel("nombres60.xlsx")

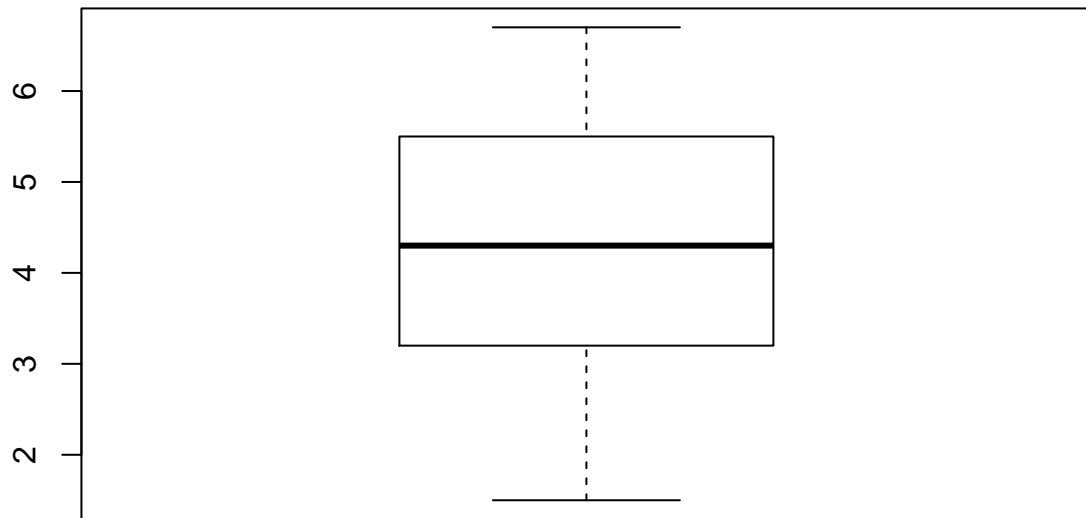
# 2) Cree un diagrama de dispersión, con las notas de la prueba
# solemne en el eje horizontal y las del examen en el eje vertical
# Puede usar los grafico de base o los de ggplot2

ggplot(data=notas)+geom_point(aes(x=solemne,y=examen))
```

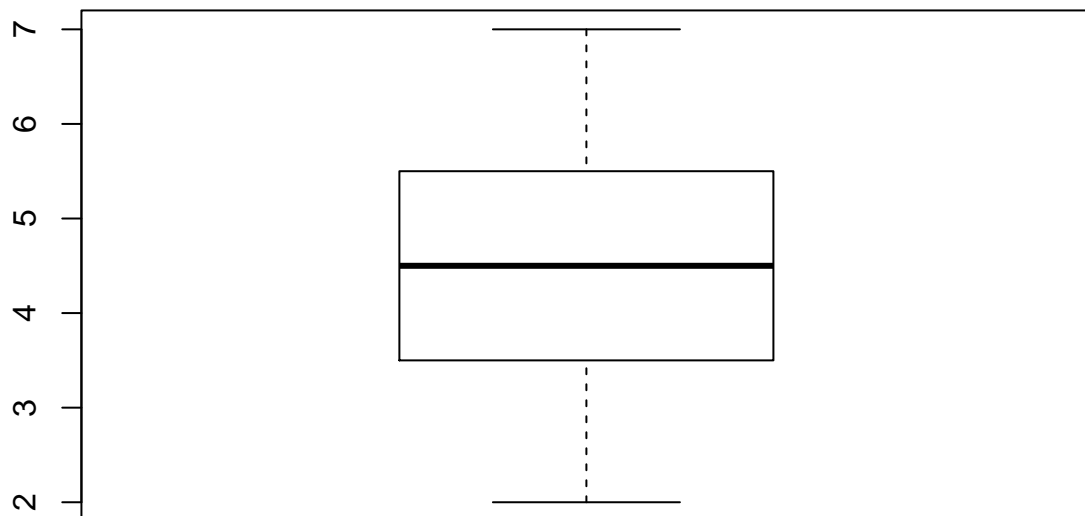


```
# 3) Cree dos diagramas boxplot: uno para las notas de la solemne
# y otro para las notas del examen
# Puede usar los grafico de base o los de ggplot2
```

```
boxplot(notas$solemne)
```



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



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

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

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

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

```
# 5) Compute los cuantiles 0.25 y 0.75 de las notas del examen  
quantile(notas$examen,.25)
```

```
## 25%
```

```
## 3.5
```

```
quantile(notas$examen, .75)
```

```
## 75%
```

```
## 5.5
```

```

# 6) Cree una nueva tibble que contenga unicamente las observaciones
# de las personas que obtuvieron un 4 o más en el examen
a <- filter(notas, examen>=4)

# 7) Compute la proporcion de gente que obtuvo un azul en el examen

#gente que saco azul = 40
#total gente =60
## proporcion=40/60, osea 2/3 de los estudiantes saco una azul en el examen.

# 8) Suponga que tiene una población de 10.000 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 de la media muestral.

n <- 25
muestra <- 10000
set.seed(12345)
poblacion <- rexp(10000, rate = 1)

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

for (i in 1:muestra) {
  esta_muestra <- sample(poblacion, size = n)
  media_muestral <- mean(esta_muestra)

  vector_con_medias[i] <- media_muestral
}

sd(vector_con_medias)

```

```
## [1] 0.2005679
```

```

# 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?

n <- 250
muestra <- 10000
set.seed(12345)
poblacion <- rexp(10000, rate = 1)

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

for (i in 1:muestra) {
  esta_muestra <- sample(poblacion, size = n)
  media_muestral <- mean(esta_muestra)

  vector_con_medias[i] <- media_muestral
}

```

```
sd(vector_con_medias)
```

```
## [1] 0.06265895
```

```
#asi la desviaciun estandar es mayor cuando las muestra es de 25, sd(25)= 0.2005679,  
#en cambien la sd(250)=0.06265895.  
#lo cual tiene logica ya que una mayor muestra generara valores mas precisos es decir  
#con menor error
```