

Informe de la Práctica de Probabilidades y Estadística II

24 de Mayo - 2021

Grupo nº: 91

Miembros del grupo de clase: GII-4F2M

Alumnos:

Alejandro Ayuso Expósito	(190238) (alejandro.ayusoex@alumnos.upm.es)
Hernán Calvo Aguiar	(190090) (h.calvoa@alumnos.upm.es)
Rubén García Vilches	(190169) (ruben.garcia.vilches@alumnos.upm.es)
Vinh Le Quang	(190286) (vinh.lequang@alumnos.upm.es)

Índice:

Parte 1: Identificación de Modelo y Muestreo.	3
Ajuste del Modelo	3
Breve descripción de las variables sleeptime y steps:	3
Ajuste de todos los datos a:	4
Muestreo	6
Se toman muestras del conjunto de datos con tamaño 200 para la variable Age.	6
Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Medias muestrales y representar hist y boxplot. Ajustar a la distribución normal cada uno de los vectores de medias (variable aleatoria muestral)	6
Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Varianzas muestrales y representar hist y boxplot. Ajustar a la distribución normal cada uno de los vectores de varianzas (variable aleatoria muestral)	7
Parte 2: Estimación Clásica (puntual, intervalos).	10
Parte 3: Estimación Bayesiana	14
Parte 4: Contrastes	15
a) Tomar dos muestras de tamaño 200 de la variable IMC: Sample1 y Sample2.	15
b) Tomar una muestra de tamaño 200 de la variable IMC, Sample, con nivel de significación $\alpha = 0,05$.	16
Anexo:	18

DataSet:

g91

Inicializar el script

Establecemos el seed, el working directory, importamos datos del csv y cargamos los paquetes.

```
#script para inicializar los datos

#set seed e importar Data
set.seed(2021)
setwd("Path")
Data = read.csv( file=paste("PYE2DataSet","91",".csv",sep=""), header=TRUE)

#importación de Librerías
librerias <- c("tidyverse", "broom", "plyr",
              "ggplot2", "lattice", "Rmisc", "DescTools", "MASS", "car",
              "EstimationTools", "ISwR", "IMTest", "boot", "vcd", "rcompanion", "FSA",
              "psych", "e1071", "TeachingDemos", "lmtest");

lapply(librerias, require, character.only = TRUE)
```

1. Parte 1: Identificación de Modelo y Muestreo.

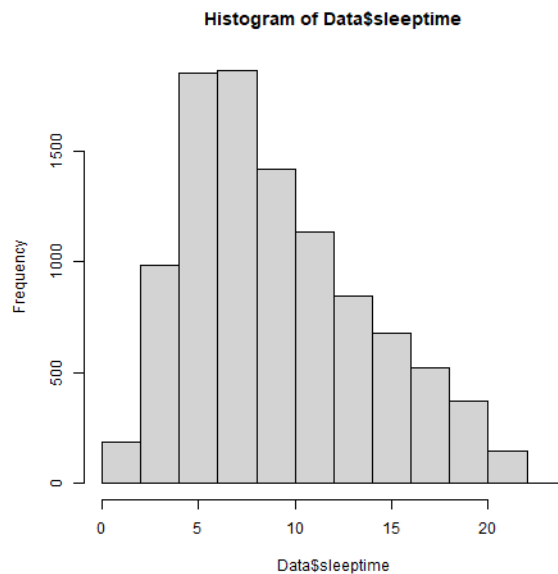
a. Ajuste del Modelo

i. Breve descripción de las variables sleeptime y steps:

Summary de sleeptime:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.8542	5.5239	8.1516	9.1012	12.1139	22.1035

Histograma de sleeptime:

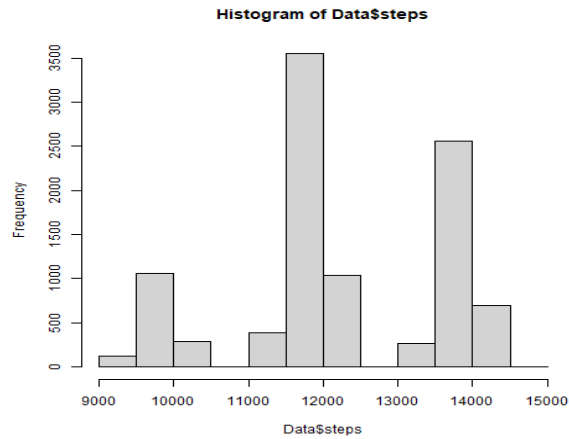


Skewed right.

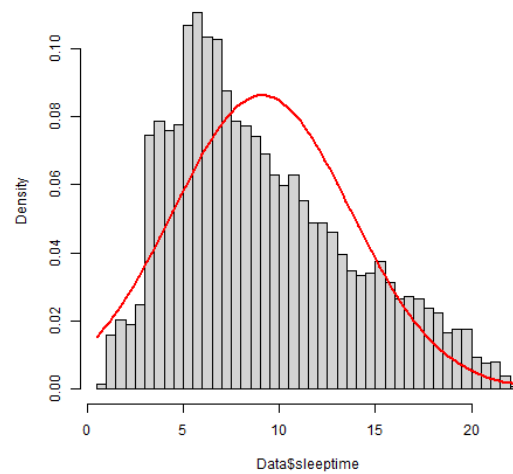
Summary de steps:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
9085	11628	11939	12223	13695	14579

Histograma de steps:



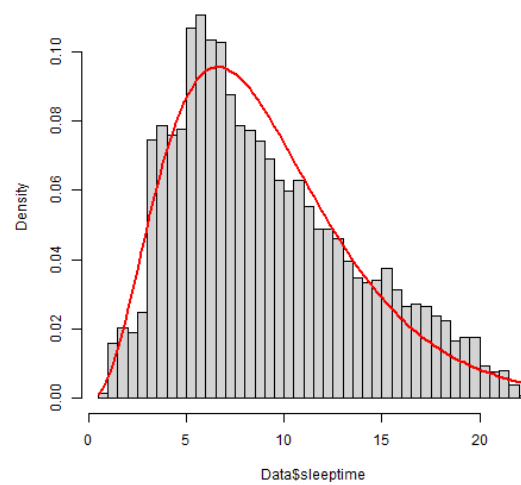
- ii. Ajuste de todos los datos a:
1. Distribucion Normal:



Test de Kolmogorov-Smirnov:
 $p\text{-value} < 2.2e-16$

Como podemos observar por el histograma, la curva normal no se ajusta nada. El test de Kolmogorov-Smirnov subraya lo poco bueno que es este ajuste.

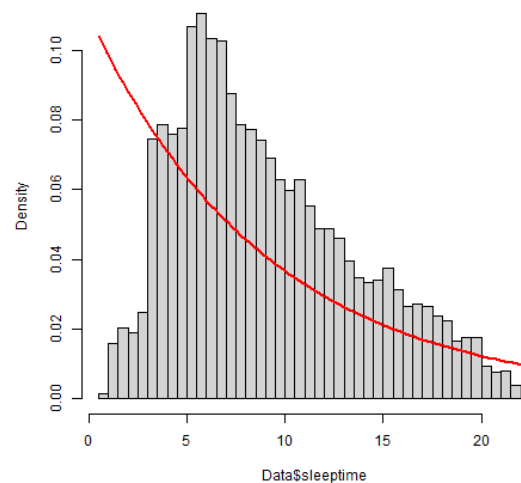
2. Distribución Gamma:



Test de Kolmogorov-Smirnov:
p-value = 4.154e-07

A ojo, podríamos pensar que este ajuste es más válido. Pero el test nos demuestra lo contrario.

3. Distribución Exponencial:



Test de Kolmogorov-Smirnov:
p-value < 2.2e-16

No hay mucho que comentar, tanto el test como la relación visual entre curva e histograma es válido.

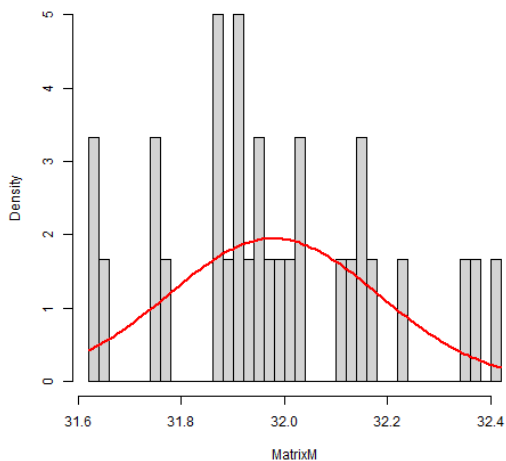
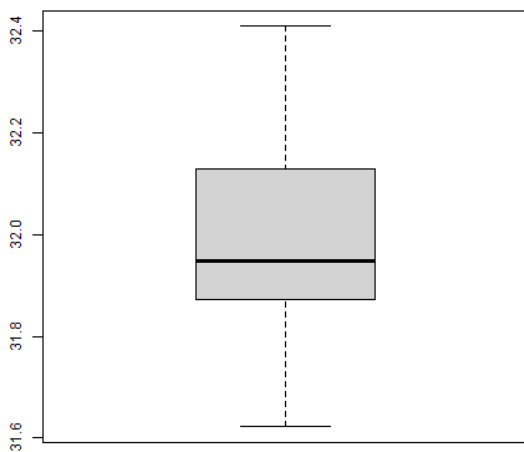
b. Muestreo

- i. Se toman muestras del conjunto de datos con tamaño 200 para la variable Age.

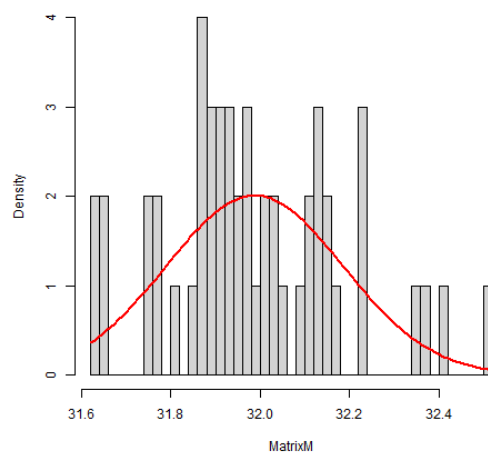
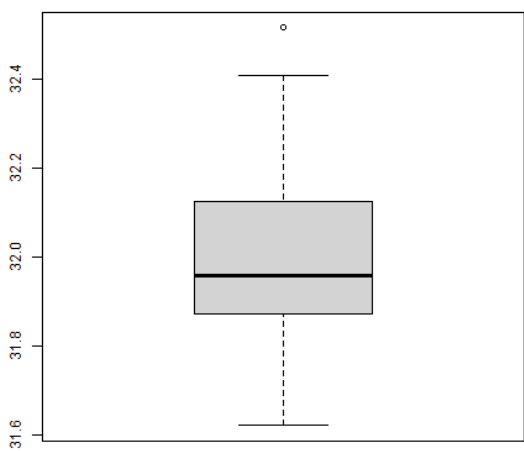
Tomaremos n muestras de tamaño 200 según soliciten los siguientes apartados.

- ii. Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Medias muestrales y representar hist y boxplot. Ajustar a la distribución normal cada uno de los vectores de medias (variable aleatoria muestral)

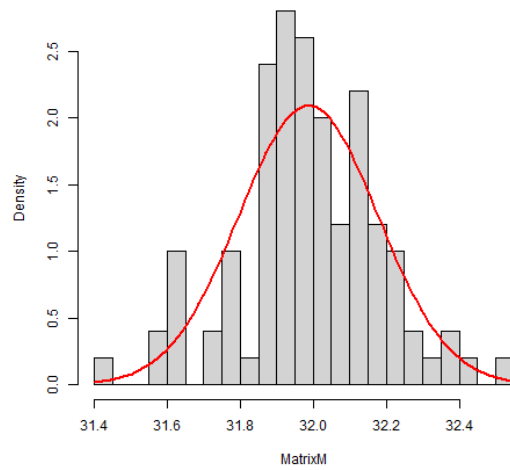
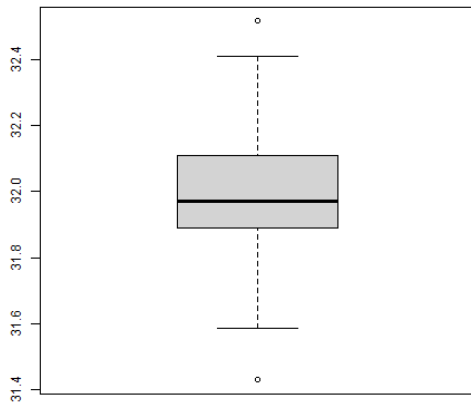
30 Muestras:



50 Muestras:



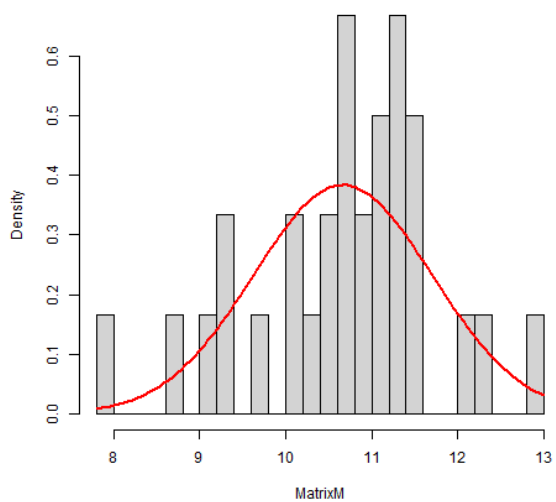
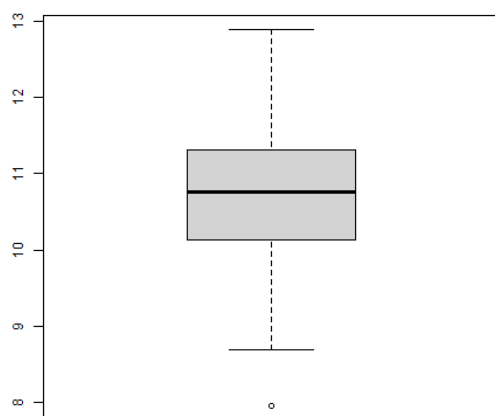
100 Muestras:



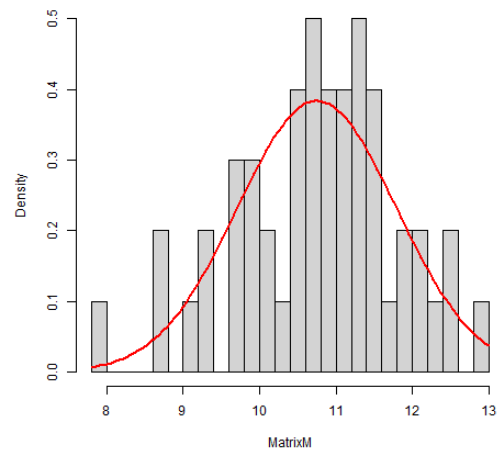
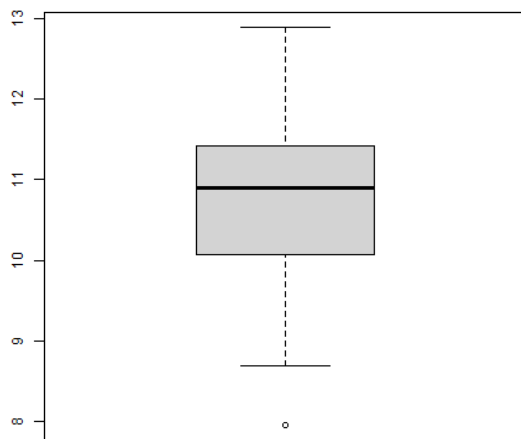
Este muestreo no reduce ni la presencia de datos atípicos ni hace que el ajuste se acerque más a la curva normal, salvo en las 100 muestras.

- iii. Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Varianzas muestrales y representar hist y boxplot. Ajustar a la distribución normal cada uno de los vectores de varianzas (variable aleatoria muestral)

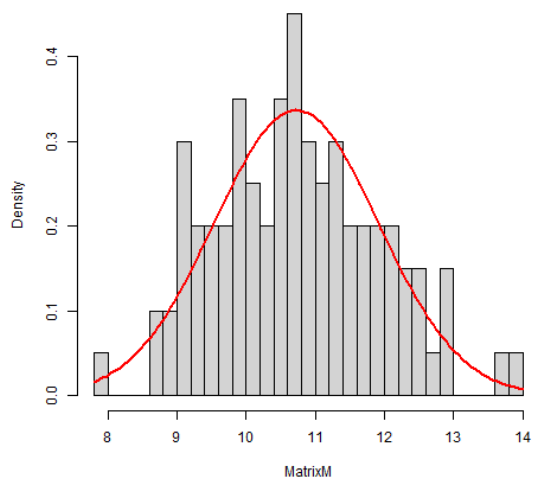
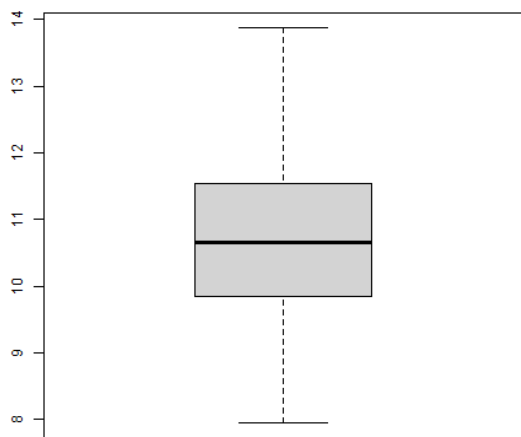
30 Muestras:



50 Muestras:



100 Muestras:

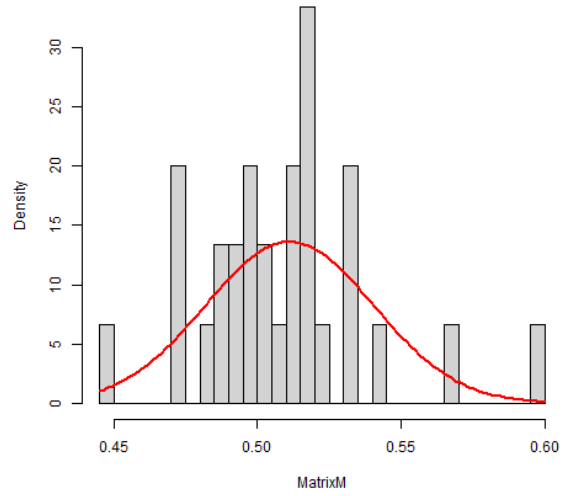
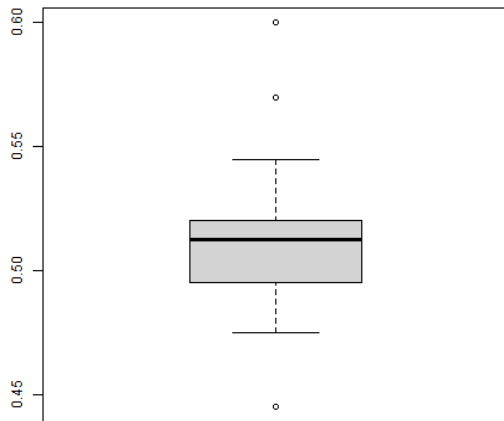


A mayor número de muestras podemos observar cómo el histograma se acerca cada vez más a la normal. Tenemos algún dato que se sale de la descripción de la curva normal, pero son principalmente dentro del intervalo \pm -sigma.

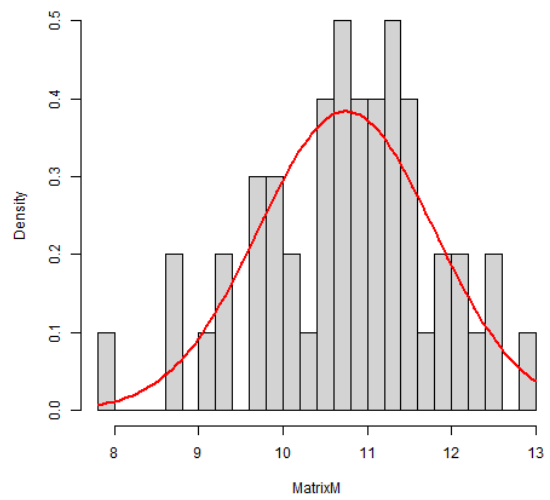
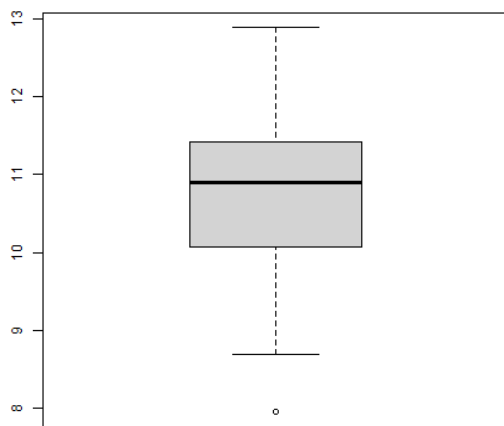
Los boxplots varían con mucha más fuerza, pero conforme aumentan las muestras, se reduce la presencia de datos atípicos.

- iv. Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 proporción muestral de Mujeres/Varones y representar hist y boxplot. Ajustar a la distribución normal cada uno de los vectores de proporciones (variable aleatoria muestral)

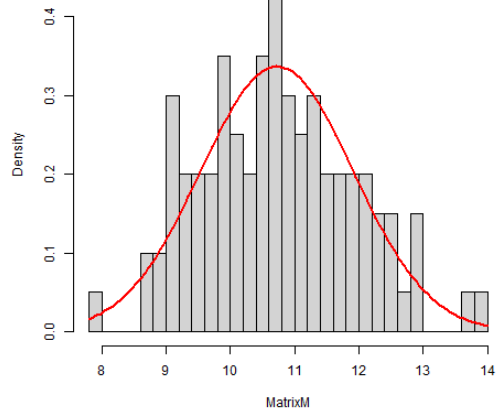
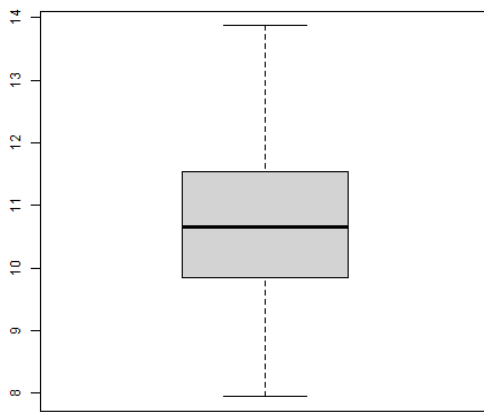
30 Muestras:



50 Muestras:



100 Muestras:



Para estas muestras los datos se acercan a la curva normal muy rápidamente. Podemos destacar los datos atípicos de las 30 muestras y la presencia de datos fuera de la curva en la parte izquierda de la curva de los 3 histogramas.

2. Parte 2: Estimación Clásica (puntual, intervalos).

a. Puntual:

- i. Estimar media y varianza de las variables sleeptime y steps. Primero con todos los datos y segundo con una muestra de tamaño 200:

```
> estimacion("Mujeres y Varones")
Estimacion media sleeptime Mujeres y Varones, con todos los datos: 9.101236
Estimacion varianza sleeptime Mujeres y Varones, con todos los datos: 21.41806
Estimacion media sleeptime Mujeres y Varones, con muestra de 200: 9.278468
Estimacion varianza sleeptime Mujeres y Varones, con muestra de 200: 22.45069
Estimacion media steps Mujeres y Varones, con todos los datos: 9.101236
Estimacion varianza steps Mujeres y Varones, con todos los datos: 21.41806
Estimacion media steps Mujeres y Varones, con muestra de 200: 12323.72
Estimacion varianza steps Mujeres y Varones, con muestra de 200: 1815074
```

- ii. Estimar media y varianza de las variables sleeptime y steps entre las Mujeres. Primero con todos los datos y segundo con una muestra de tamaño 200:

```
> estimacion("Mujeres")
Estimacion media sleeptime Mujeres, con todos los datos: 10.07005
Estimacion varianza sleeptime Mujeres, con todos los datos: 20.40148
Estimacion media sleeptime Mujeres, con muestra de 200: 10.18539
Estimacion varianza sleeptime Mujeres, con muestra de 200: 20.28736
Estimacion media steps Mujeres, con todos los datos: 10.07005
Estimacion varianza steps Mujeres, con todos los datos: 20.40148
Estimacion media steps Mujeres, con muestra de 200: 13269.52
Estimacion varianza steps Mujeres, con muestra de 200: 853936.2
```

- iii. Estimar la media y la varianza de las variables sleeptime y steps entre los Varones. Primero con todos los datos y segundo con una muestra de tamaño 200:

```
> estimacion("Varones")
Estimacion media sleeptime Varones, con todos los datos: 8.122294
Estimacion varianza sleeptime Varones, con todos los datos: 20.54246
Estimacion media sleeptime Varones, con muestra de 200: 7.525649
Estimacion varianza sleeptime Varones, con muestra de 200: 18.69168
Estimacion media steps Varones, con todos los datos: 8.122294
Estimacion varianza steps Varones, con todos los datos: 20.54246
Estimacion media steps Varones, con muestra de 200: 11269.83
Estimacion varianza steps Varones, con muestra de 200: 913949.6
```

b. Intervalo:

- i. Estimación del intervalo de confianza para la media, varianza, proporción, al nivel de confianza 90 %, 95 % y 99 %, para las variables sleeptime y step entre según niveles {"M", "V"} del factor Sex, con una muestra de tamaño 200. Primero suponer normalidad y segundo usar Bootstrap para el caso de poblaciones de distribución general o arbitraria. Para la media suponer primero varianza conocida y segundo desconocida.
- ii. Estimación del intervalo de confianza para la diferencia de medias, razón de varianzas, proporción, al nivel de confianza 90%, 95% y 99%, para las variables sleeptime y steps según niveles {"M", "V"} del factor Sex, con una muestra de tamaño 200. Primero suponer normalidad y segundo usar Bootstrap para el caso de poblaciones de distribución general o arbitraria. Para la diferencia de medias suponer primero varianzas conocidas y segundo desconocidas e iguales.

M y V:

```

> estimacionConf("Mujeres y Varones")
Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 90% : [8.618143,9.694688]
Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 90% : [8.636011,9.67682]
Estimacion intervalo sleeptime Mujeres y Varones de la varianza al nvl de confianza 90% : [16.94597,19.65143]
Estimacion intervalo sleeptime Mujeres y Varones de la proporcion, al nvl de confianza 90% : [0.479395,0.5994914]

Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 95% : [8.515024,9.797807]
Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 95% : [8.535425,9.777406]
Estimacion intervalo sleeptime Mujeres y Varones de la varianza al nvl de confianza 95% : [16.44825,19.77563]
Estimacion intervalo sleeptime Mujeres y Varones de la proporcion, al nvl de confianza 95% : [0.4683438,0.6100945]

Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 99% : [8.313484,9.999347]
Estimacion intervalo sleeptime Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 99% : [8.337408,9.975423]
Estimacion intervalo sleeptime Mujeres y Varones de la varianza al nvl de confianza 99% : [15.53078,19.87536]
Estimacion intervalo sleeptime Mujeres y Varones de la proporcion, al nvl de confianza 99% : [0.4469448,0.6304168]

Estimacion intervalo steps Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 90% : [12182.8,12502.88]
Estimacion intervalo steps Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 90% : [12193.43,12492.25]
Estimacion intervalo steps Mujeres y Varones de la varianza al nvl de confianza 90% : [1396841,1619849]
Estimacion intervalo steps Mujeres y Varones de la proporcion, al nvl de confianza 90% : [0.479395,0.5994914]

Estimacion intervalo steps Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 95% : [12152.15,12533.54]
Estimacion intervalo steps Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 95% : [12164.55,12521.13]
Estimacion intervalo steps Mujeres y Varones de la varianza al nvl de confianza 95% : [1355814,1630087]
Estimacion intervalo steps Mujeres y Varones de la proporcion, al nvl de confianza 95% : [0.4683438,0.6100945]

Estimacion intervalo steps Mujeres y Varones de la media, con varianza conocida, al nvl de confianza 99% : [12092.22,12593.46]
Estimacion intervalo steps Mujeres y Varones de la media, con varianza desconocida, al nvl de confianza 99% : [12107.7,12577.98]
Estimacion intervalo steps Mujeres y Varones de la varianza al nvl de confianza 99% : [1280188,1638308]
Estimacion intervalo steps Mujeres y Varones de la proporcion, al nvl de confianza 99% : [0.4469448,0.6304168]

```

M:

```

> estimacionConf("Mujeres")
Estimacion intervalo sleeptime Mujeres de la media, con varianza conocida, al nvl de confianza 90% : [9.597135,10.67368]
Estimacion intervalo sleeptime Mujeres de la media, con varianza desconocida, al nvl de confianza 90% : [9.576511,10.6943]
Estimacion intervalo sleeptime Mujeres de la varianza al nvl de confianza 90% : [19.54555,22.66603]

Estimacion intervalo sleeptime Mujeres de la media, con varianza conocida, al nvl de confianza 95% : [9.494016,10.7768]
Estimacion intervalo sleeptime Mujeres de la media, con varianza desconocida, al nvl de confianza 95% : [9.468485,10.80233]
Estimacion intervalo sleeptime Mujeres de la varianza al nvl de confianza 95% : [18.97147,22.80928]

Estimacion intervalo sleeptime Mujeres de la media, con varianza conocida, al nvl de confianza 99% : [9.292477,10.97834]
Estimacion intervalo sleeptime Mujeres de la media, con varianza desconocida, al nvl de confianza 99% : [9.255822,11.01499]
Estimacion intervalo sleeptime Mujeres de la varianza al nvl de confianza 99% : [17.91326,22.92432]

Estimacion intervalo steps Mujeres de la media, con varianza conocida, al nvl de confianza 90% : [13055.67,13375.74]
Estimacion intervalo steps Mujeres de la media, con varianza desconocida, al nvl de confianza 90% : [13106.26,13325.15]
Estimacion intervalo steps Mujeres de la varianza al nvl de confianza 90% : [749475.1,869130.4]

Estimacion intervalo steps Mujeres de la media, con varianza conocida, al nvl de confianza 95% : [13025.01,13406.4]
Estimacion intervalo steps Mujeres de la media, con varianza desconocida, al nvl de confianza 95% : [13085.11,13346.3]
Estimacion intervalo steps Mujeres de la varianza al nvl de confianza 95% : [727462.3,874623.4]

Estimacion intervalo steps Mujeres de la media, con varianza conocida, al nvl de confianza 99% : [12965.09,13466.32]
Estimacion intervalo steps Mujeres de la media, con varianza desconocida, al nvl de confianza 99% : [13043.46,13387.94]
Estimacion intervalo steps Mujeres de la varianza al nvl de confianza 99% : [686884.9,879034.4]

```

V:

```
> estimacionConf("Varones")
Estimacion intervalo sleeptime Varones de la media, con varianza conocida, al nvl de confianza 90% : [7.563061,8.639607]
Estimacion intervalo sleeptime Varones de la media, con varianza desconocida, al nvl de confianza 90% : [7.535282,8.667386]
Estimacion intervalo sleeptime Varones de la varianza al nvl de confianza 90% : [20.04918,23.25008]

Estimacion intervalo sleeptime Varones de la media, con varianza conocida, al nvl de confianza 95% : [7.459942,8.742725]
Estimacion intervalo sleeptime Varones de la media, con varianza desconocida, al nvl de confianza 95% : [7.425873,8.776794]
Estimacion intervalo sleeptime Varones de la varianza al nvl de confianza 95% : [19.46032,23.39702]

Estimacion intervalo sleeptime Varones de la media, con varianza conocida, al nvl de confianza 99% : [7.258403,8.944265]
Estimacion intervalo sleeptime Varones de la media, con varianza desconocida, al nvl de confianza 99% : [7.210487,8.99218]
Estimacion intervalo sleeptime Varones de la varianza al nvl de confianza 99% : [18.37484,23.51502]

Estimacion intervalo steps Varones de la media, con varianza conocida, al nvl de confianza 90% : [11099.98,11420.06]
Estimacion intervalo steps Varones de la media, con varianza desconocida, al nvl de confianza 90% : [11152.06,11367.98]
Estimacion intervalo steps Varones de la varianza al nvl de confianza 90% : [729298.1,845732]

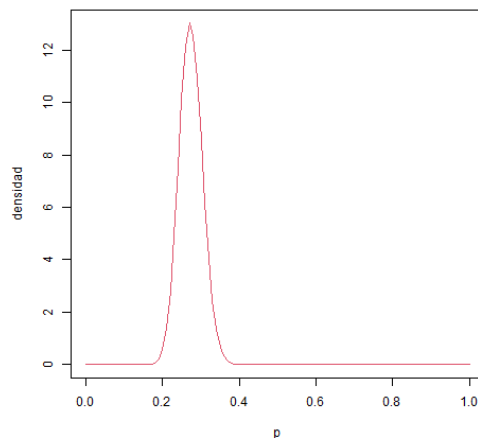
Estimacion intervalo steps Varones de la media, con varianza conocida, al nvl de confianza 95% : [11069.33,11450.72]
Estimacion intervalo steps Varones de la media, con varianza desconocida, al nvl de confianza 95% : [11131.2,11388.85]
Estimacion intervalo steps Varones de la varianza al nvl de confianza 95% : [707877.9,851077.1]

Estimacion intervalo steps Varones de la media, con varianza conocida, al nvl de confianza 99% : [11009.41,11510.64]
Estimacion intervalo steps Varones de la media, con varianza desconocida, al nvl de confianza 99% : [11090.12,11429.93]
Estimacion intervalo steps Varones de la varianza al nvl de confianza 99% : [668392.9,855369.4]
```

3. Parte 3: Estimación Bayesiana

- a. La proporción p de individuos de nacionalidad española en la población está entre 25 % y el 35 %. En una muestra de 200 personas hay nE españoles. Suponer que $p \sim \beta(\alpha = 5, \beta = 10)$, con función de densidad $f(x; \alpha; \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}$ y moda $= \frac{\alpha-1}{\alpha+\beta-1}$.
- i. La información disponible a priori es equivalente a la observación de r_0 elementos con el atributo estudiado en una muestra de n_0 elementos.

Esto es el intervalo de confianza 95%: 0.2135745 0.3334059.



- La estatura (variable height) del grupo de españoles, franceses e italianos sigue una $N(170; 7)$. Estimar con la muestra anterior, la estatura media con varianza conocida.

171.174

4. Parte 4: Contrastes

a) Tomar dos muestras de tamaño 200 de la variable IMC: Sample1 y Sample2.

Cuantiles:

25%	50%	75%
24.14895	24.70300	25.16196

- Contrastar si la media μ_1 de Sample1 es $Q_1 \leq \mu_1$, con varianza desconocida (Q_1 : cuartil 1 de la muestra)

data: sample4_1

t = 10.35, df = 199, p-value < 2.2e-16

alternative hypothesis: true mean is greater than 24.14895

95 percent confidence interval:

24.5902 Inf

sample estimates:

mean of x

24.67404

Rechazamos la hipotesis nula ya que pvalor es menos de 0.01. Concluimos que la media de la muestra es mayor que la media de Q_1

- Contrastar si la media μ_1 de Sample1 es $\mu_1 \leq Q_3$, con varianza desconocida

data: sample4_1

t = -9.6175, df = 199, p-value < 2.2e-16

alternative hypothesis: true mean is less than 25.16196

95 percent confidence interval:

-Inf 24.75788

sample estimates:

mean of x

24.67404

Rechazamos la hipotesis nula ya que pvalor es menos de 0.01. Concluimos que la media de la muestra es menor que la media de Q_3

- Contrastar si la varianza σ^2 de Sample1 es mayor que 1.0, con media desconocida.

data: sample4_1

X-squared = 102.44, df = 199, p-value = 1

alternative hypothesis: true variance is greater than 1

95 percent confidence interval:

0.4398153 Inf

sample estimates:

var of sample4_1

0.5147648

No rechazamos la hipotesis nula ya que pvalor es mayor que 0.2.

- Contrastar si $\mu_1 - \mu_2 = 0$, con Sample1 y Sample2 respectivamente, con varianzas desconocidas.

data: sample4_1 and sample4_2

t = -0.85658, df = 199, p-value = 0.3927

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.21031409 0.08293304

sample estimates:

mean of the differences

-0.06369052

No rechazamos que las medias son iguales (hipotesis nula) con una confianza del 95% ya que el pvalor es mayor que 0.2. Se estima que la diferencia de las medias es 0.06369052

- Contrastar si $\sigma_1^2/\sigma_2^2 = 1$, con Sample1 y Sample2 respectivamente.

data: sample4_1 and sample4_2

F = 0.94691, num df = 199, denom df = 199, p-value = 0.7008

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.7166066 1.2512207

sample estimates:

ratio of variances

0.9469071

No rechazamos que las varianzas son iguales (hipotesis nula) con una confianza del 95% ya que el pvalor es mayor que 0.2. Se estima que la ratio de las varianzas es 0.973572 (casi 1).

b) Tomar una muestra de tamaño 200 de la variable IMC, Sample, con nivel de significación $\alpha = 0,05$.

- Contrastar la normalidad de Sample, mediante el test de Pearson y el test de Kolmogorov-Smirnov

Pearson chi-square normality test

data: sample4_1

P = 9.27, p-value = 0.8134

One-sample Kolmogorov-Smirnov test

data: sample4_1

D = 0.035058, p-value = 0.9666

alternative hypothesis: two-sided

No rechamos la hipotesis nula porque los pvalores son mayores que 0.2, es decir, no rechazamos que sea normal.

- Contrastar la independencia de Sample, mediante el test de Durbin-Watson. Se trata de ver si hay dependencia de IMC respecto a algunas variables de conjunto de datos. Sugerencia: paquete lmtest y función dwtest(), es decir, se toma una muestra de tamaño 200 del conjunto de datos y se proponen algunas variables independientes de las que pueda depender IMC, y tras hacer el test se saca una conclusión.

data: sample_dw

DW = 2.0893, p-value = 0.7379

alternative hypothesis: true autocorrelation is greater than 0

No rechazamos la hipótesis nula, es decir, existe una dependencia entre el IMC con sleeptime+steps+weight

- Contrastar la homogeneidad de Sample, mediante el test de Wilcoxon. Se trata de ver si varias muestras provienen de la misma población, es decir, tomamos dos muestras de tamaño 200 de la variable IMC (de la misma población) y tras hacer el test se saca una conclusión. Sugerencia: paquete stats y función wilcox.test()

data: sample4_1 and sample4_2

W = 19088, p-value = 0.4305

alternative hypothesis: true location shift is not equal to 0

No rechazamos la hipótesis nula, es decir, las dos muestras son homogéneas

Anexo:

```
#source("inicializar.R",local=TRUE);

#PARTE 1
###Ajuste de modelo

#Sleeptime y steps
sink("Resultados 1A_1.txt");
summary(Data$sleeptime)
png(filename = paste(paste("Histograma", "sleeptime", sep="_"), ".png",
sep=""));
hist(Data$sleeptime);
dev.off();

summary(Data$steps)
png(filename = paste(paste("Histograma", "steps", sep="_"), ".png", sep=""));
hist(Data$steps);
dev.off();
sink()

sink("Resultados 1A_2.txt")
slptm <- Data$sleeptime
###D.Normal

normal <- fitdistr(slptm, "normal")
  png(filename = paste(paste("1B", "Histograma", "D.Normal", "sleeptime",
sep="_"), ".png", sep=""))
  #histograma
  hist(Data$sleeptime, pch=20, breaks=35, prob=TRUE, main="")
  curve(dnorm(x, normal$estimate[1], normal$estimate[2]), col="red", lwd=2,
add=T)
  dev.off()
  #Test de K-S
  normtest<-ks.test(slptm,"pnorm")
  normtest

###D.Gamma
gamma <- fitdistr(slptm, "gamma",lower = c(0, 0))
```

```

    png(filename = paste(paste("1B", "Histograma", "D.Gamma", "sleeptime",
sep="_"), ".png", sep=""))
    #histograma
    hist(x=Data$sleeptime, breaks=35, prob=TRUE, main="")
    curve(dgamma(x, gamma$estimate[1], gamma$estimate[2]), col="red", lwd=2,
add=T)
    dev.off()
    #Test de K-S

gammatest<-ks.test(slptm,"pgamma",shape=gamma$estimate[1],rate=gamma$estimate[
2])

    gammatest

##D.Exponencial
expo <- fitdistr(slptm, "exponential")
    png(filename = paste(paste("1B", "Histograma", "D.Exponencial",
"sleeptime", sep="_"), ".png", sep=""))
    #histograma
    hist(x=Data$sleeptime, breaks=35, prob=TRUE, main="")
    curve(dexp(x, expo$estimate[1]), col="red", lwd=2, add=T)
    dev.off()
    #Test de K-S
    expotest<-ks.test(slptm,"pexp",rate=expo$estimate[1])
    expotest

sink()

###Muestreo
muestreoTipo <- function(n,tipo,method = NA){
  MatrixM = matrix(NA,nrow = n,ncol = 1)
  #Tomamos n muestras
  for(i in 1:n){
    if(tipo == "meanAge" | tipo == "varAge"){
      set.seed(2021)
      SampleM <- replicate(n, Data$Age[sample(1:dim(Data)[1],200)],)
      MatrixM[i,] <- method(SampleM[,i])
    } else if(tipo == "propSex"){
      set.seed(2021)
      SampleM <- replicate(n, Data$Sex[sample(1:dim(Data)[1],200)],)
      MatrixM[i,] <- table(SampleM[,i])[1]/200
    }
  }
}

```

```

    }
    normal <- fitdistr(MatrixM, "normal")
    png(filename = paste(paste("1B", tipo, "ajuste", n, sep="_"), ".png",
sep=""))
    hist(x=MatrixM, breaks=35, prob=TRUE, main="")
    curve(dnorm(x, normal$estimate[1], normal$estimate[2]), col="red", lwd=2,
add=T)
    dev.off()
    png(filename = paste(paste("1B", tipo, "boxplot", n, sep="_"), ".png",
sep=""))
    boxplot(MatrixM)
    dev.off()
}

```

```

prevwd <- getwd()
setwd(paste(getwd(), "/", "output", sep=""))

```

```
#1B
```

```

muestreoTipo(30,"meanAge",mean)
muestreoTipo(50,"meanAge",mean)
muestreoTipo(100,"meanAge",mean)
muestreoTipo(30,"varAge",var)
muestreoTipo(50,"varAge",var)
muestreoTipo(100,"varAge",var)
muestreoTipo(30,"propSex")
muestreoTipo(50,"propSex")
muestreoTipo(100,"propSex")

```

```
setwd(prevwd)
```

```

#####
#####
#####
#####

```

```
##PARTE 2. Estimacion Clasica
```

```
#a
```

```

estimacion <- function(){
  for( genero in c("Mujeres", "Varones", "Mujeres y Varones")){
    if(genero == "Mujeres"){
      DataX <- Data[Data[, "Sex"] == "M",]; #Dataframe con solo mujeres de
Data
    } else if(genero == "Varones"){

```

```

    DataX <- Data[Data[, "Sex"] == "V",]; #Dataframe con solo mujeres de
Data
  } else{
    DataX <- Data;
  }
  for(column in c("sleeptime", "steps")){
    cat("Estimacion media ", column, " ", genero, ", con todos los datos: ",
mean(DataX$sleeptime), "\n", sep="")
    cat("Estimacion varianza ", column, " ", genero, ", con todos los datos:
", var(DataX$sleeptime), "\n", sep="")
    set.seed(2021)
    SampleM <- DataX[,column][sample(1:dim(DataX)[1],200)];
    cat("Estimacion media ", column, " ", genero, ", con muestra de 200: ",
mean(SampleM), "\n", sep="")
    cat("Estimacion varianza ", column, " ", genero, ", con muestra de 200:
", var(SampleM), "\n", sep="")
  }
}

sink("Resultados 2A.txt");
estimacion()
#estimacion("Mujeres y Varones")
#estimacion("Mujeres")
#estimacion("Varones")
sink();

#b
#Estimacion de intervalo de confianza para media, varianza, proporcion
# al nvl de confianza 90, 95, 99
# para las variables sleeptime y steps
# segun sex Mujeres y Varones

# Primero suponer normalidad y segundo usar Bootstrap para
# el caso de poblaciones de distribucion general o arbitraria.
# Para la media suponer primero varianza conocida y segundo desconocida.

estimacionConf <- function(genero){

  if(genero == "Mujeres"){

```

```

    DataX <- Data[Data[, "Sex"] == "M",];
  } else if(genero == "Varones"){
    DataX <- Data[Data[, "Sex"] == "V",];
  } else{
    DataX <- Data;
  }
#checkseed();
set.seed(2021)
SampleM <- DataX[sample(1:dim(DataX)[1],200),];
bootR <- 1000
for(column in c("sleeptime", "steps")){
  for(conflvl in c(0.90,0.95,0.99)){

    #Media
    # Varianza conocida
    intervals <- z.test(SampleM[,column], sd = sd(Data[,column]), conf.level
= conflvl)$conf.int
    cat("Estimacion intervalo ", column, " ", genero, " de la media, con
varianza conocida, al nvl de confianza ", conflvl*100,
    "% : [", intervals[1], ",", intervals[2], "]", "\n", sep="")

    #bootstrap
    Mboot = boot(SampleM[,column],function(x,i) mean(x[i]),R=bootR)
    #print(mean(Mboot$t[,1]))
    bci200 <- boot.ci(Mboot,conf = conflvl,type=c("norm","basic"))$normal
    cat("Estimacion intervalo ", column, " ", genero, " de la media,
mediante Bootstrap, al nvl de confianza ", conflvl*100,
    "% : [", bci200[2], ",", bci200[3], "]", "\n", sep="")

    # Varianza desconocida
    intervals <- t.test(x=SampleM[,column], conf.level=conflvl)$conf.int
    cat("Estimacion intervalo ", column, " ", genero, " de la media, con
varianza desconocida, al nvl de confianza ", conflvl*100,
    "% : [", intervals[1], ",", intervals[2], "]", "\n", sep="")

    #bootstrap

    #Varianza
    df <- length(SampleM[,column]) - 1
    varSample <- var(SampleM[,column])
    lower = varSample * df / qchisq((1-conflvl)/2, df, lower.tail = FALSE)
    upper = varSample * df / qchisq(conflvl/2, df, lower.tail = FALSE)
  }
}

```

```

        cat("Estimacion intervalo ", column, " ", genero, " de la varianza al
nvl de confianza ", conflvl*100,
        "% : [", lower, ",", upper, "]", "\n", sep="")
        #bootstrap
        Mboot = boot(SampleM[,column],function(x,i) var(x[i]),R=bootR)
        #print(mean(Mboot$t[,1]))
        bci200 <- boot.ci(Mboot,conf = conflvl,type=c("norm","basic"))$normal
        cat("Estimacion intervalo ", column, " ", genero, " de la varianza,
mediante Bootstrap, al nvl de confianza ", conflvl*100,
        "% : [", bci200[2], ",", bci200[3], "]", "\n", sep="")

        if(genero == "Mujeres y Varones"){
            #Proporcion
            numMujeres <- table(SampleM$Sex)[1]
            intervals <- prop.test(numMujeres,200, conf.level = conflvl)$conf.int
            cat("Estimacion intervalo ", column, " ", genero, " de la proporcion,
al nvl de confianza ", conflvl*100,
            "% : [", intervals[1], ",", intervals[2], "]", "\n", sep="")
        }
        cat("\n",sep="")
    }
}

sink("Resultados 2B.txt");
estimacionConf("Mujeres y Varones")
estimacionConf("Mujeres")
estimacionConf("Varones")
sink();

#####
#####
#####
#####

#3
set.seed(2021)
SampleM = sample(Data$Nation,200)

r0 <- 5
n0 <- 10
r <- table(SampleM)[["SP"]] #52
n <- 200

```

```

aux1 <- r+r0
aux2 <- (n+n0)-(r+r0)

#Aproximaciones
#A posteriori antes de ajustar
post1 <- function(theta) {
  theta^(aux1)*(1-theta)^aux2 #57, 153
}
post2 <- function(theta) {
  theta^(aux1)*(1-theta)^aux2/k
}

k <- integrate(post1,Lower=0,upper=1)$value
round(integrate(post2,Lower=0,upper=1)$value,2)

#IC
qbeta(c(0.025,0.975),shape1=r+r0,shape2=(n+n0)-(r+r0))
png(filename = paste(paste("3", "densidad", sep="_"), ".png", sep=""))
curve(post2, col = 2,xlab = "p",ylab = "densidad")
dev.off()

SampleNations <- subset(Data, Nation == SampleM, select =
c("Nation","height"))
SampleSPFRIT <- subset(SampleNations,Nation == "SP" | Nation == "FR" | Nation
== "IT",select = c("Nation","height"))

SampleMean = mean(SampleSPFRIT$height)
n0 = (sd(SampleSPFRIT$height)^2)/(7^2)
meanPost = (200*SampleMean + n0*170)/(200 + n0)
cat("Estatura media:", meanPost)

```

```

#####
#####
#####
#####

```

```
#PARTE 4A
```



```

set.seed(2021)
sample4_1 <- sample(Data$IMC, 200)
sample4_2 <- sample(Data$IMC, 200)

quantile(sample4_1, prob=c(.25,.5,.75))
#Cuartil 1 media = 24.14895
t.test(sample4_1, alternative = "greater",mu = 23.61980)
# One Sample t-test
#
# data:  sample4_1
# t = 19.14, df = 199, p-value < 2.2e-16
# alternative hypothesis: true mean is greater than 23.6198
# 95 percent confidence interval:
#  24.56217      Inf
# sample estimates:
#  mean of x
# 24.65122

#Rechazamos la hipotesis nula ya que pvalor es menos de 0.01. Concluimos que
la media de la muestra es mayor que la media de Q1

t.test(sample4_1, alternative = "less",mu = 25.17701)

# One Sample t-test
#
# data:  sample4_1
# t = -9.757, df = 199, p-value < 2.2e-16
# alternative hypothesis: true mean is less than 25.17701
# 95 percent confidence interval:
#  -Inf 24.74028
# sample estimates:
#  mean of x
# 24.65122

#Rechazamos la hipotesis nula ya que pvalor es menos de 0.01. Concluimos que
la media de la muestra es menor que la media de Q3

library(TeachingDemos)
sigma.test(sample4_1, 1, alternative = "greater")

# data:  sample4_1

```

```
# X-squared = 115.58, df = 199, p-value = 1
# alternative hypothesis: true variance is greater than 1
# 95 percent confidence interval:
#  0.4962185      Inf
# sample estimates:
#   var of sample4_1
# 0.5807797
```

#No rechazamos la hipotesis nula ya que pvalor es mayor que 0.2

```
t.test(sample4_1, sample4_2, paired = TRUE, var.equal = TRUE)
```

```
# Paired t-test
#
# data:  sample4_1 and sample4_2
# t = 1.0592, df = 199, p-value = 0.2908
# alternative hypothesis: true difference in means is not equal to 0
# 95 percent confidence interval:
#  -0.06861783  0.22787214
# sample estimates:
#   mean of the differences
# 0.07962715
```

#No rechazamos que las medias son iguales (hipotesis nula) con una confianza del 95% ya que el pvalor es mayor que 0.2.

#Se estima que la diferencia de las medias es 0.07962715

```
var.test(sample4_1, sample4_2)
```

```
# F test to compare two variances
#
# data:  sample4_1 and sample4_2
# F = 0.97357, num df = 199, denom df = 199, p-value = 0.8503
# alternative hypothesis: true ratio of variances is not equal to 1
# 95 percent confidence interval:
#  0.7367862 1.2864551
# sample estimates:
#   ratio of variances
# 0.973572
```

```

#No rechazamos que las varianzas son iguales (hipotesis nula) con una
confianza del 95% ya que el pvalor es mayor que 0.2.
#Se estima que la ratio de las varianzas es 0.973572 (casi 1)

#PARTE 4B
library(nortest)
pearson.test(sample4_1)

# Pearson chi-square normality test
#
# data:  sample4_1
# P = 9.27, p-value = 0.8134

#No rechamos la hipotesis nula, es decir, no rechazamos que sea normal

library(stats)
ks.test(sample4_1, "pnorm", mean = mean(sample4_1), sd = sd(sample4_1))

# One-sample Kolmogorov-Smirnov test
#
# data:  sample4_1
# D = 0.035058, p-value = 0.9666
# alternative hypothesis: two-sided

#No rechamos la hipotesis nula, es decir, no rechazamos que sea normal

library(lmtest)

sample4_IMC <- sample(Data$IMC, 200)
sample4_sleeptime <- sample(Data$sleeptime, 200)
sample4_steps <- sample(Data$steps, 200)

sample4_1_frame <- data.frame (sample4_IMC,sample4_sleeptime,sample4_steps)
sample_dw <- lm(formula = sample4_IMC~sample4_sleeptime+sample4_steps, data =
sample4_1_frame)
dwtest(sample_dw)

# Durbin-Watson test
#
# data:  sample_dw
# DW = 1.9871, p-value = 0.4642

```

```
# alternative hypothesis: true autocorrelation is greater than 0

#No rechazamos la hipotesis nula, es decir, existe una dependencia entre el
IMC con sleeptime+steps+weight

wilcox.test(sample4_1,sample4_2)

# Wilcoxon rank sum test with continuity correction
#
# data:  sample4_1 and sample4_2
# W = 19199, p-value = 0.4887
# alternative hypothesis: true location shift is not equal to 0

#No rechazamos la hipotesis nula, es decir, las dos muestras son homogeneas
```