

Guia 7

Agustin Muñoz Gonzalez

6/5/2020

Preparamos el entorno.

Limpiamos los registros.

```
rm(list=ls())
```

1 Calentando motores

1 Descargar de esta página un conjunto de $n = 500$ observaciones, con todas las variables y leer el archivo en R.

```
alturas=read.csv('alturas_n_500.csv',header = T)
attach(alturas)
```

2. Identificar el nombre de las columnas del data.frame.

```
names(alturas)
```

```
## [1] "altura"          "genero"          "contextura_madre" "altura_madre"
```

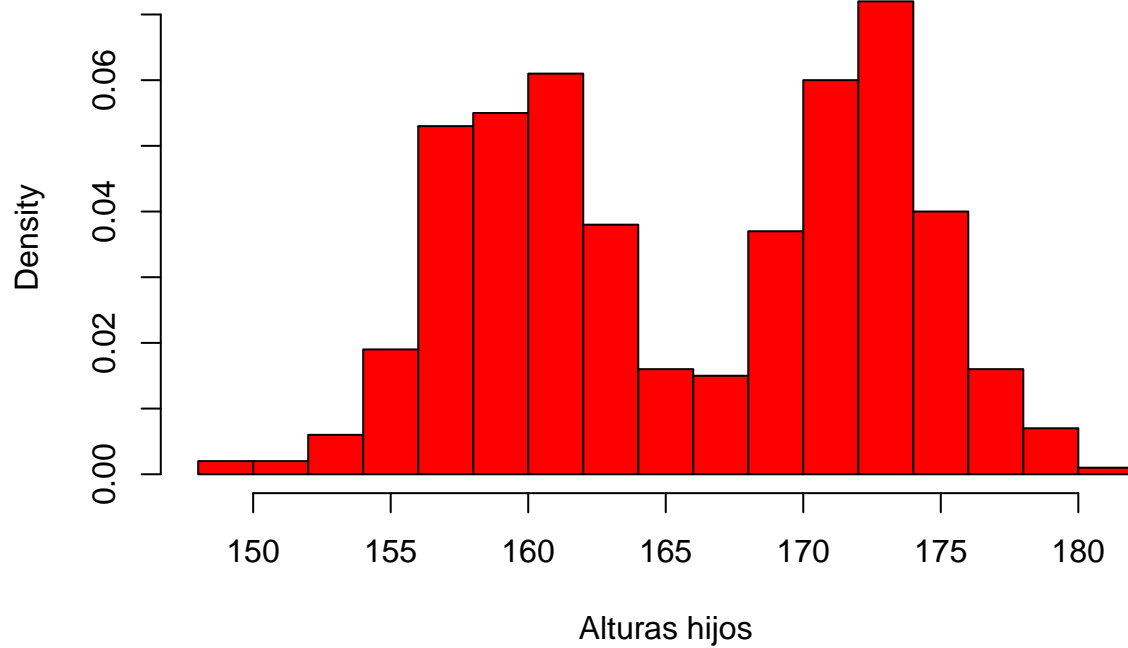
3. Calcular el promedio de las alturas de los hijos para predecir la altura de un nuevo individuo.

```
promedio_total=mean(altura)
```

4. Realizar un histograma de las alturas de los hijos. ¿Cuántas modas se observan? ¿A qué se puede atribuir?

```
hist(altura,freq=F,main="Histograma de densidad",
     nclass=15, col="red",xlab="Alturas hijos")
```

Histograma de densidad



Calculemos la moda.

```
aux=c()
for(i in unique(sort(altura))){
  aux=c(aux, sum(altura==i))
}
frecuencias=data.frame('alturas'=unique(sort(altura)), 'frecuencias'=aux)
moda=frecuencias$altura[frecuencias$frecuencias==max(frecuencias$frecuencias)]
moda
```

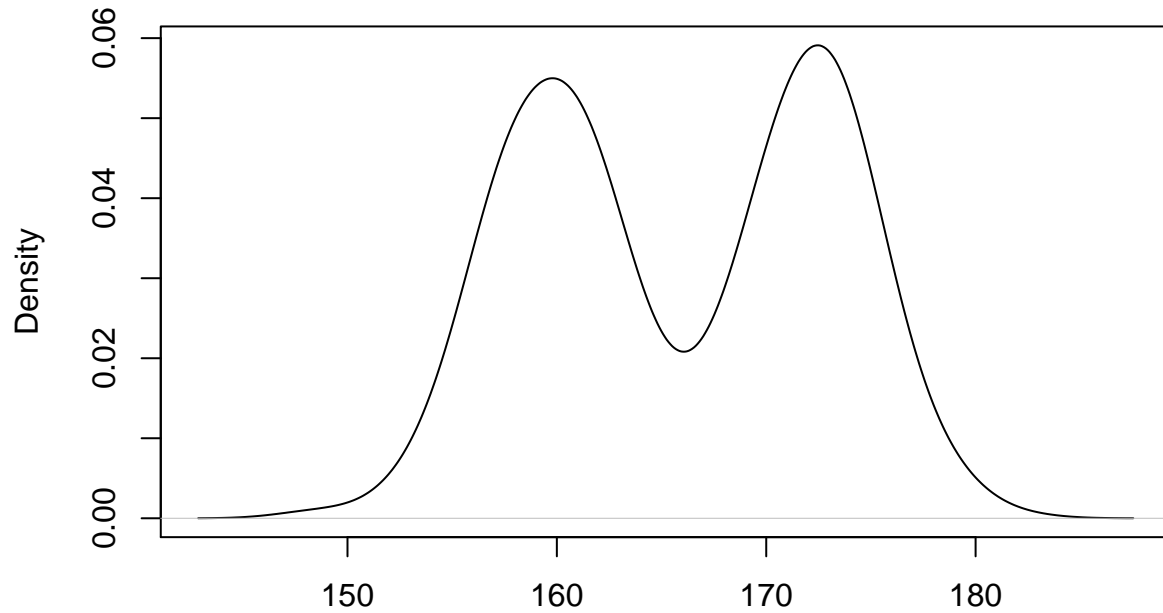
```
## [1] 161.4 173.7
```

Tenemos entonces 2 modas. Se le puede atribuir al hecho de que en general las mujeres miden menos que los hombres y por tanto es de esperar (si hay una cantidad similar de hombres y de mujeres) que haya una concentración de valores para la altura de las mujeres y una concentración de valores para la altura de los hombres.

5. Explor el comando `plot(density(variable))`, utilizando como variable la columna de alturas con las que se realizó el histograma. ¿Qué está pasando? ¿Cuántas modas observa? ¿A qué se puede atribuir?

```
plot(density(altura))
```

density.default(x = altura)

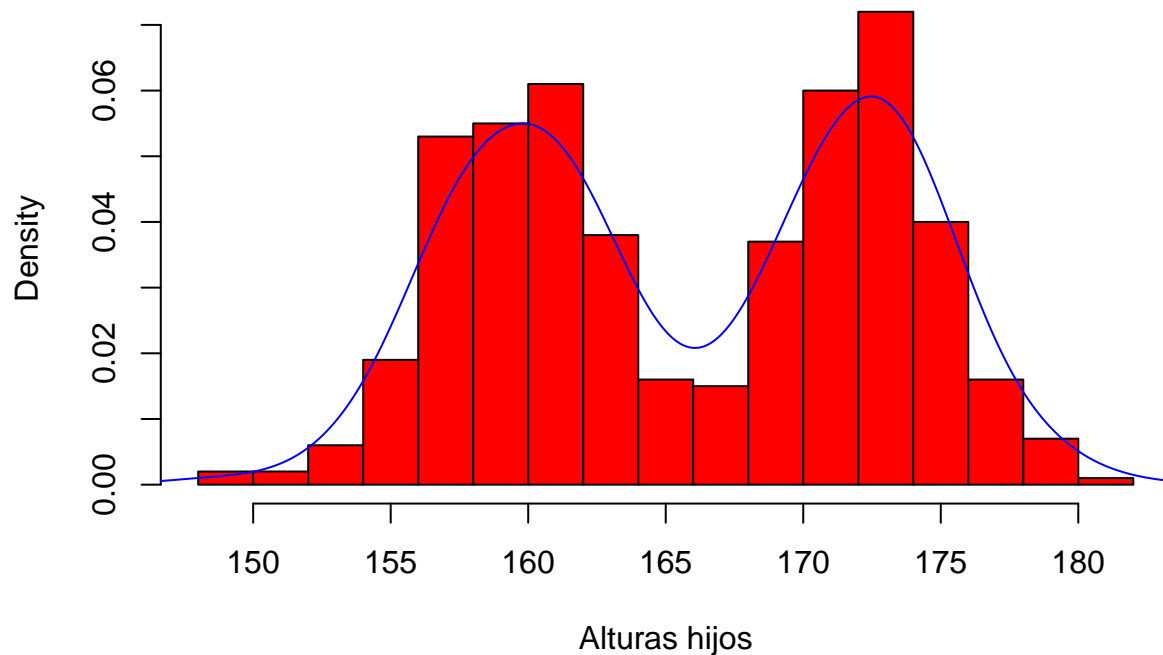


N = 500 Bandwidth = 1.844

Grafiquemoslo superpuesto con el histograma para entender un poco mejor qué está pasando.

```
hist(altura,freq=F,main="Histograma de densidad",  
     nclass=15, col="red",xlab="Alturas hijos")  
points(density(altura),type='l',col='blue')
```

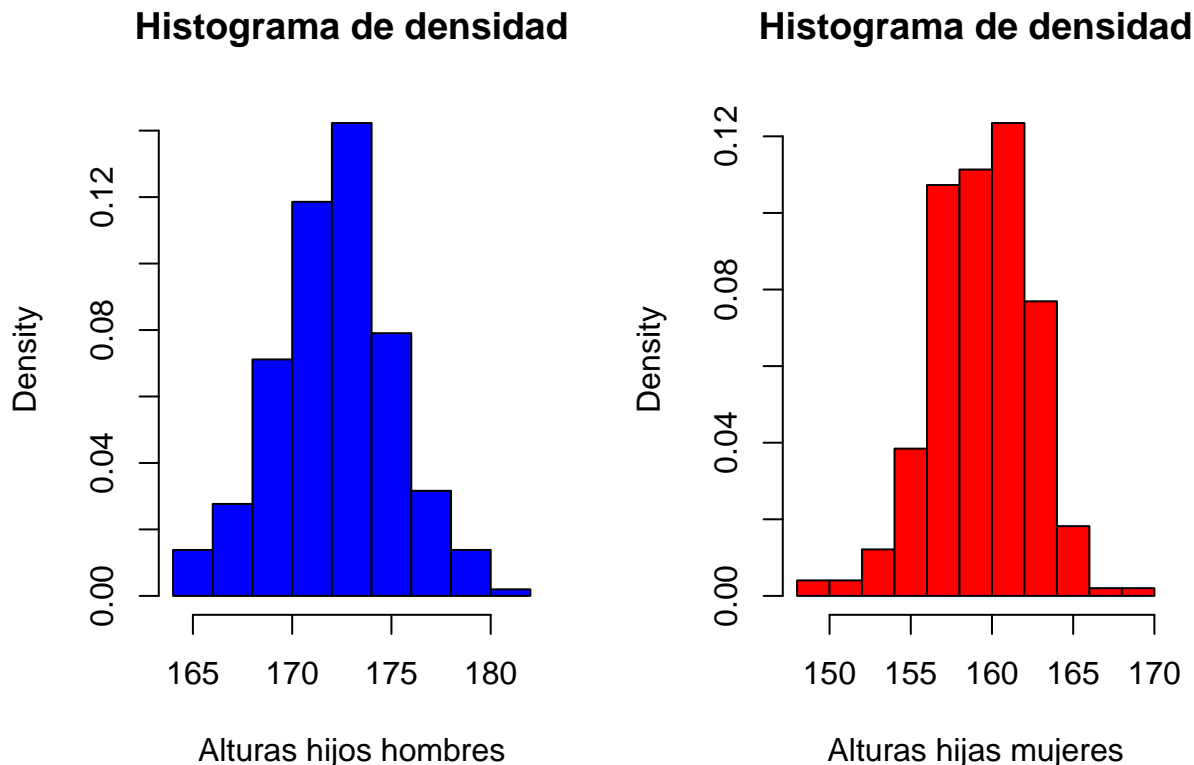
Histograma de densidad



Se puede observar el mismo fenómeno que en el Ejercicio anterior, se observan dos modas cercanas a los valores que mencionamos antes. Dado que la moda son los valores mas recurrentes, las densidades son mas altas en esos valores, con lo cual este fenómeno se lo atribuimos al mismo hecho que en el Ejercicio anterior.

6. Realizar ahora un histograma de alturas por cada género. Es decir, un histograma para las alturas correspondientes al género Masculino y otro para las alturas correspondientes al género Femenino.

```
par(mfrow=c(1,2))
hist(altura[genero=='M'],freq=F,main="Histograma de densidad",
     nclass=10, col="blue",xlab="Alturas hijos hombres")
hist(altura[genero=='F'],freq=F,main="Histograma de densidad",
     nclass=10, col="red",xlab="Alturas hijas mujeres")
```



7. Superponer en cada histograma del ítem anterior una función de densidad normal con los parámetros que consideres pertinentes.

Vamos a superponer en cada histograma una función de densidad normal con la media y el desvío de los datos de la altura de los hombres y las mujeres respectivamente.

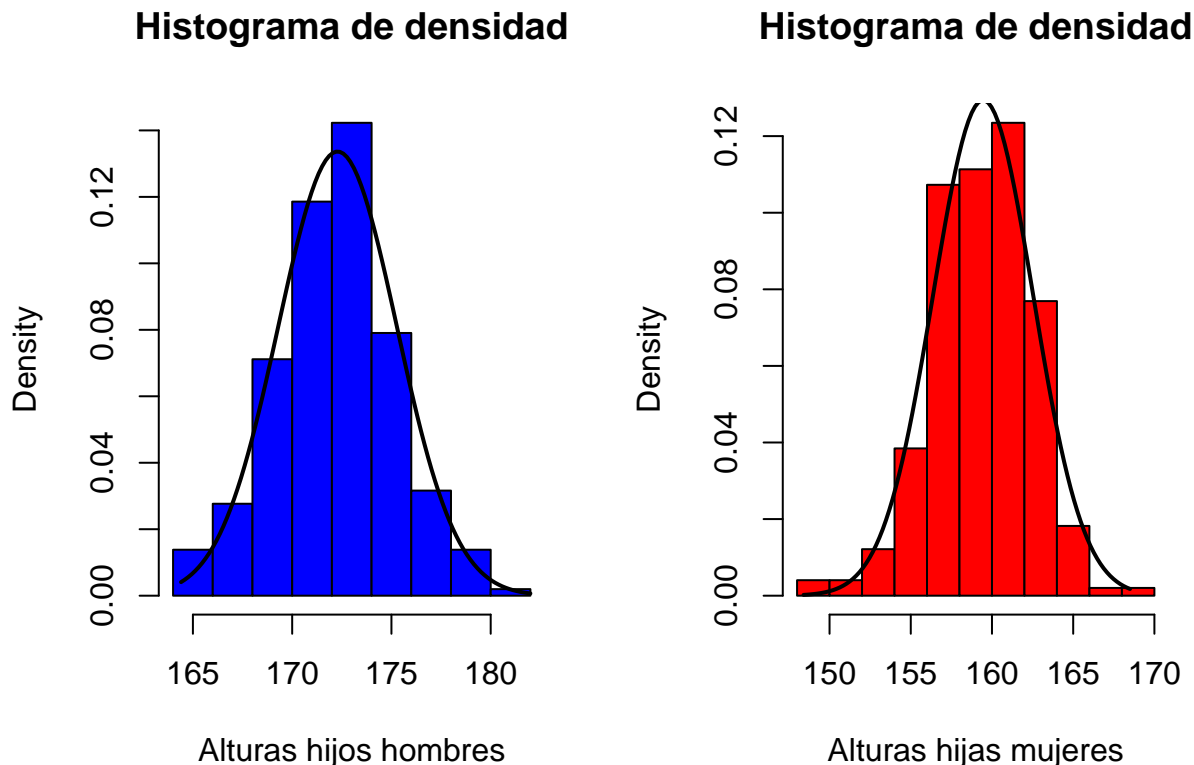
```
media_M=mean(altura[genero=='M'])
desvio_M=sd(altura[genero=='M'])
grilla_M=seq(range(altura[genero=='M'])[1],
             range(altura[genero=='M'])[2],length=100)
funn_M=dnorm(grilla_M,media_M,desvio_M)

media_F=mean(altura[genero=='F'])
desvio_F=sd(altura[genero=='F'])
grilla_F=seq(range(altura[genero=='F'])[1],
             range(altura[genero=='F'])[2],length=100)
```

```
funn_F=dnorm(grilla_F,media_F,desvio_F)

par(mfrow=c(1,2))
hist(altura[genero=='M'],freq=F,main="Histograma de densidad",
     nclass=10, col="blue",xlab="Alturas hijos hombres")
lines(grilla_M,funn_M,lwd=2)

hist(altura[genero=='F'],freq=F,main="Histograma de densidad",
     nclass=10, col="red",xlab="Alturas hijas mujeres")
lines(grilla_F,funn_F,lwd=2)
```



8. Indicar con qué valor se puede predecir la altura de un hijo Masculino, es decir, la altura de un hijo. Indicar con qué valor se puede predecir la altura de una hija Femenino, es decir, la altura de una hija.

Con los promedios de las alturas de los hijos masculinos y femeninos respectivamente.

```
promedio_M=mean(altura[genero=='M'])
promedio_F=mean(altura[genero=='F'])
```

9. Indicar ahora con qué valor se puede predecir la altura de un hijo (Masculino), sabiendo además que su mamá es de contextura pequeña. Comparar el valor de la predicción con la predicción obtenida en el ítem anterior. ¿Qué se puede observar?

Con el promedio de las alturas de los datos para genero=='M' y contextura_madre=='bajita'.

```
promedio_M_bajita=mean(altura[genero=='M' & contextura_madre=='bajita'])
promedio_M
```

```
## [1] 172.2862
```

```
promedio_M_bajita
```

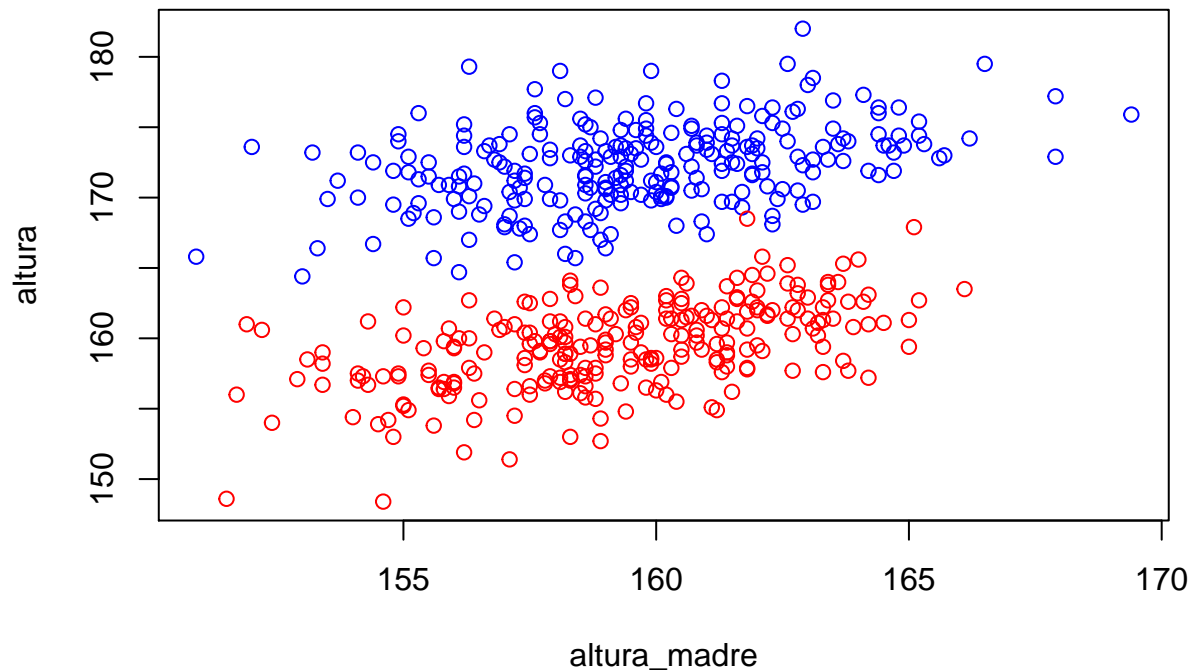
```
## [1] 170.5679
```

Observamos que cuando agregamos el filtro `contextura__madre=='bajita'` obtuvimos un promedio de altura mas bajo, lo cual tiene sentido.

2. Vamos a considerar ahora la altura de la mamá

10. Graficar altura de mamá (en el eje x) vs. altura del hijo (eje y), utilizando un color por cada género. ¿Qué se puede observar?

```
plot(altura_madre, altura, type='n')
points(altura_madre[genero=='M'], altura[genero=='M'], col='blue')
points(altura_madre[genero=='F'], altura[genero=='F'], col='red')
```



Observamos que las mujeres miden menos.

11. Indicar si hay alguna madre de hijo varón cuya altura sea 156 cm.

Una forma es sumar todos los valores de verdad de `altura_madre[genero=='M']==156` y si eso da más que 0, entonces sí existe tal madre.

```
sum(altura_madre[genero=='M']==156)
```

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

12. Vamos ahora a predecir la altura de un hijo correspondiente a una mamá que mide $x = 156$ cm haciendo promedio móvil centrado en 156 con ventana de tamaño $h = 0.5$ (cm).

a) Indicar cuántos casos hay donde la madre registra una altura entre 155.5 y 156.5 cm., inclusive.

```
sum(155.5 <= altura_madre[genero=='M'] & altura_madre[genero=='M'] <= 156.5)
```

```
## [1] 20
```

b) Calcular el promedio de la altura de los hijos cuyas madres registran una altura entre 155.5 y 156.5 cm.

```
mean(altura[155.5 <= altura_madre & altura_madre<= 156.5])
```

```
## [1] 163.6581
```

13. Promedio móvil centrado en 156 con ventana de tamaño $h = 1$ (cm).

a) Indicar cuántos casos hay donde la madre registra una altura entre 155 y 157 cm., inclusive.

```
sum(155 <= altura_madre[genero=='M'] & altura_madre[genero=='M'] <= 157)
```

```
## [1] 36
```

b) Calcular el promedio de la altura de los hijos cuyas madres registran una altura entre 155 y 157 cm.

```
mean(altura[155 <= altura_madre & altura_madre<= 157])
```

```
## [1] 164.7652
```

14. Repetir los items anteriores pero considerando ahora que la altura de la mamá es 159 cm. Es decir, nos movemos a calcular promedios en otro lado. Por eso hablamos de promedios móviles.

Con ventana $h=0.5$

```
sum(158.5 <= altura_madre[genero=='M'] & altura_madre[genero=='M'] <= 159.5)
```

```
## [1] 46
```

```
mean(altura[158.5 <= altura_madre & altura_madre<= 159.5])
```

```
## [1] 166.4831
```

Con ventana $h=1$

```
sum(158 <= altura_madre[genero=='M'] & altura_madre[genero=='M'] <= 160)
```

```
## [1] 70
```

```
mean(altura[158 <= altura_madre & altura_madre<= 160])
```

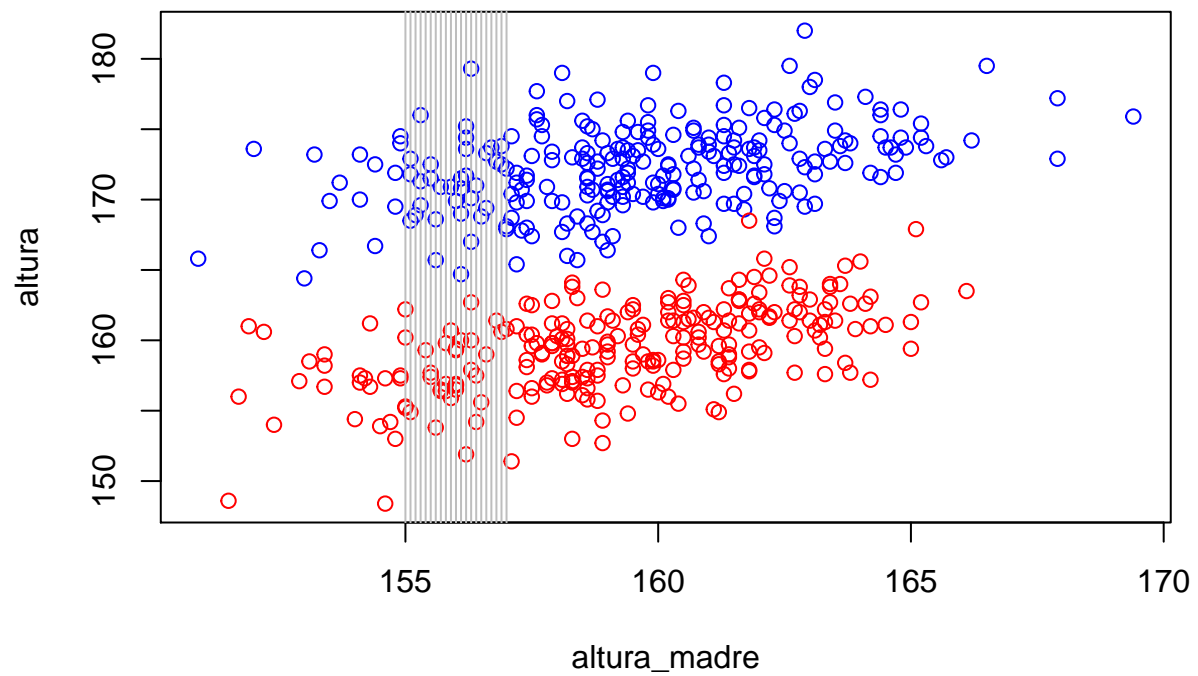
```
## [1] 165.7451
```

```
filtro=1
```

```
mean(altura[156-filtro <= altura_madre & altura_madre<= 156+filtro])
```

```
## [1] 164.7652
```

```
plot(altura_madre,altura,type='n')
points(altura_madre[genero=='M'],altura[genero=='M'],col='blue')
points(altura_madre[genero=='F'],altura[genero=='F'],col='red')
abline(v=seq(156-filtro,156+filtro,0.1),col='grey')
```

```
promedio_madre_ventana=function(altura_mama,filtro){  
  mean(altura[altura_mama-filtro <= altura_madre & altura_madre <= altura_mama+filtro])  
}
```