Regresión Lineal

Erika Martínez Meneses

2024-08-30

Estatura y Peso

Regresión Lineal

Analiza la base de datos de estatura y peso de los hombres y mujeres en México y obten el mejor modelo de regresión para esos datos.

```
file.choose()

## [1] "C:\\Users\\erika\\Documents\\Agos-Dic2024\\Estadística\\Estatura-peso_HyM.csv"

library(readr)
M <- read_csv("C:\\Users\\erika\\Documents\\Agos-Dic2024\\Estadística\\Estatura-peso_HyM.csv")

## Rows: 440 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Sexo
## dbl (2): Estatura, Peso
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1 = data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)
```

La recta de mejor ajuste

Análisis descriptivo

1. Matriz de correlación.

```
cor(M1)
```

```
##               MH.Estatura     MH.Peso  MM.Estatura    MM.Peso
## MH.Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872
## MH.Peso     0.8468347920 1.000000000 0.0035132246 0.02154907
## MM.Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621
## MM.Peso     0.0472487231 0.021549075 0.5244962115 1.00000000
```

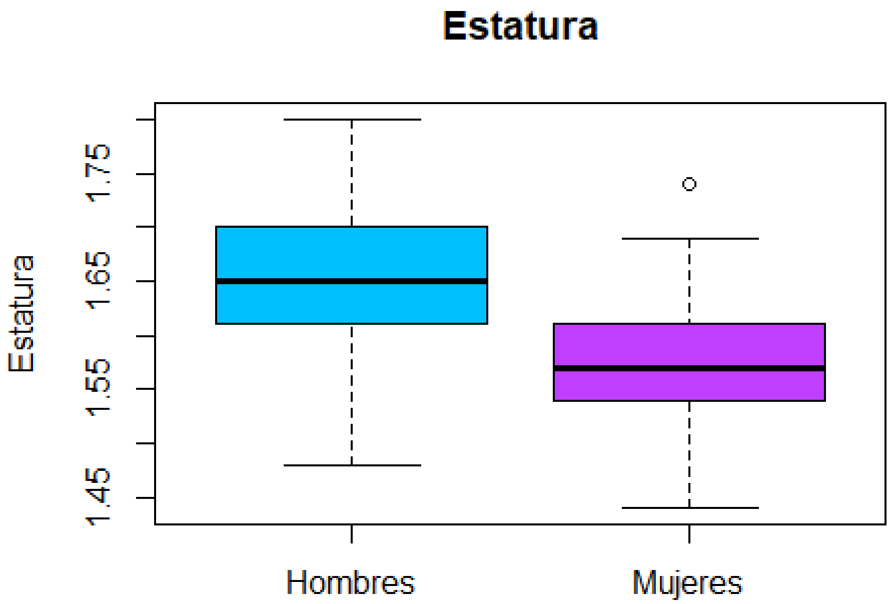
Se observa que existe correlación entre la estatura y peso cuando los datos son del mismo género (hombre o mujer) sin embargo la correlación entre la estatura y peso en hombres es más fuerte (0.8468) que la correlación existente entre la estatura y peso de mujeres (0.5244). Esto sugiere que la variabilidad en el peso de los hombres se explica mejor por la estatura en comparación con las mujeres. Las correlaciones cruzadas (estatura de hombres con peso de mujeres, y viceversa) son muy bajas, lo que indica que no hay relación entre las variables de distintos géneros.

2. Obtén medidas (media, desviación estándar, etc) que te ayuden a analizar los datos.

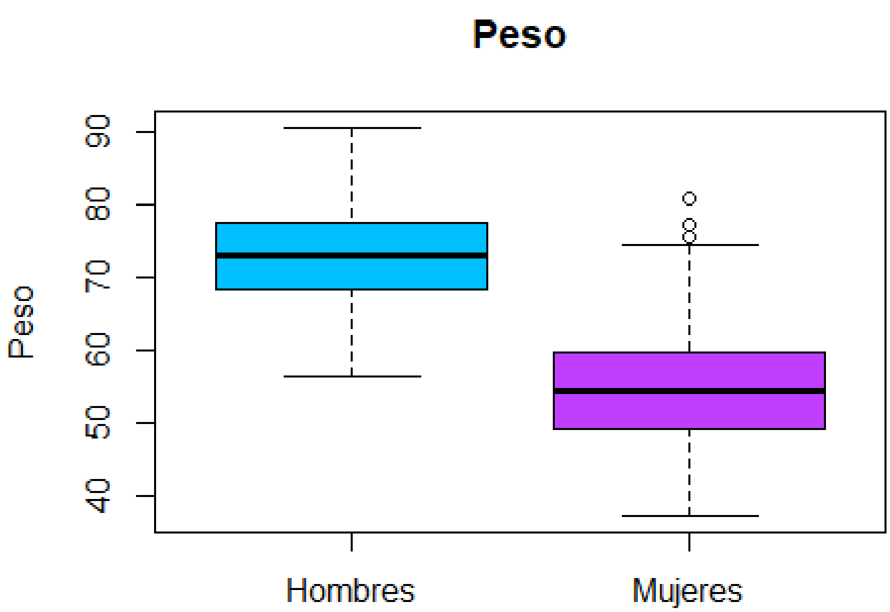
```r
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)

row.names(m)=c("H-Estatura","H-Peso","M-Estatura","M-Peso")
names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
m

##             Minimo      Q1 Mediana     Media      Q3 Máximo   Desv Est
## H-Estatura   1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## H-Peso      56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408
## M-Estatura   1.44  1.5400   1.570  1.572955  1.6100   1.74 0.05036758
## M-Peso      37.39 49.3550  54.485 55.083409 59.7950  80.87 7.79278074
```

Boxplot

```r
boxplot(M$Estatura~M$Sexo, ylab="Estatura", xlab="",
col=c("#00BFFF","#BF3EFF"), names=c("Hombres", "Mujeres"),
main="Estatura")
```




```r
boxplot(M$Peso~M$Sexo, ylab="Peso",xlab="", names=c("Hombres",
"Mujeres"), col=c("#00BFFF","#BF3EFF"), main="Peso")
```



La media y la desviación estándar de la estatura y el peso para ambos géneros indican que, en promedio, los hombres son más altos y pesados que las mujeres.

La recta del mejor ajuste

Encuentra la ecuación de regresión de mejor ajuste

Dos rectas

```
Modelo1H = lm(Peso~Estatura, data = MH)
Modelo1H

##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Coefficients:
## (Intercept)     Estatura
##      -83.68        94.66
```

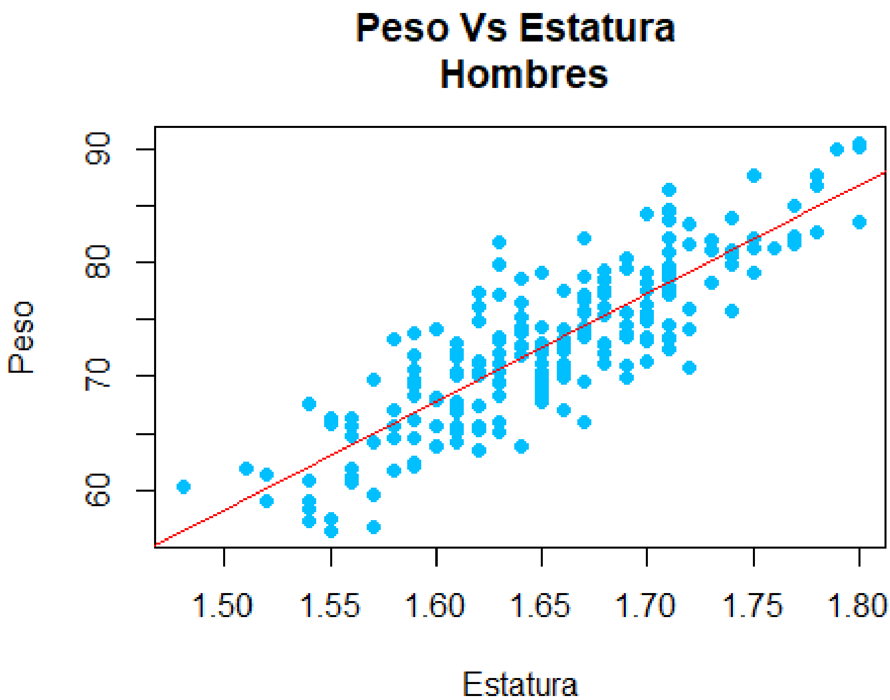
Modelo 1 Hombres * Estatura = -83.68 + 94.66

Hipótesis * $H_0: \beta_1 = 0$ * $H_1: \beta_1 \neq 0$

```
summary(Modelo1H)

##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3881 -2.6073 -0.0665  2.4421 11.1883
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -83.685      6.663  -12.56   <2e-16 ***
## Estatura      94.660      4.027   23.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.678 on 218 degrees of freedom
## Multiple R-squared:  0.7171, Adjusted R-squared:  0.7158
## F-statistic: 552.7 on 1 and 218 DF,  p-value: < 2.2e-16
```

Grafica

```
plot(MH$Estatura, MH$Peso, col = "#00BFFF", main = "Peso Vs Estatura \n
Hombres", ylab = "Peso", xlab = "Estatura", pch = 19)
abline(Modelo1H, col = "red", lwd = 1.6)
```



Rechazo B_0, sí es significativa

- El 71% de la variabilidad esta siendo explicada
- Nos queda una t grande lo que significa que en la distribución se encuentra 46.72 y 23.51 veces lejos de la β hipotética.
- Podemos observar en los cuartiles que hay simetría.

```
Modelo1M = lm(Peso~Estatura, data= MM)
Modelo1M

##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Coefficients:
## (Intercept)     Estatura
##      -72.56        81.15
```

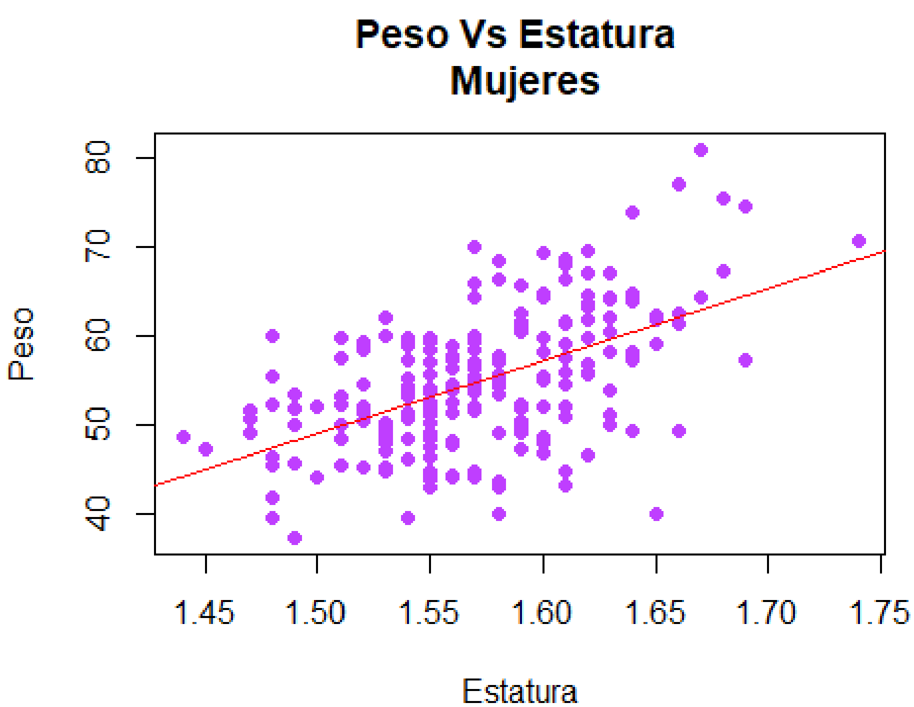
Modelo 1 Mujeres * Estatura = -72.56 + 81.15

```
summary(Modelo1M)

##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -21.3256  -4.1942   0.4004   4.2724  17.9114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -72.560     14.041  -5.168 5.34e-07 ***
## Estatura      81.149      8.922   9.096  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.65 on 218 degrees of freedom
## Multiple R-squared:  0.2751, Adjusted R-squared:  0.2718
## F-statistic: 82.73 on 1 and 218 DF,  p-value: < 2.2e-16
```

Grafica

```r
plot(MM$Estatura, MM$Peso, col = "#BF3EFF", main = "Peso Vs Estatura \n
Mujeres", ylab = "Peso", xlab = "Estatura", pch = 19)
abline(Modelo1M, col = "red", lwd = 1.6)
```



Rechazo B_0, sí es significativa

- El 27.51% de la variabilidad esta siendo explicada
- El R^2 nos sale bajo
- Nos queda una t grande lo que significa que en la distribución se encuentra 66.859 y 9.096 veces lejos de la β hipotética.

Realiza la regresión entre las variables involucradas

Un modelo

```
Modelo2 = lm(Peso~Estatura+Sexo, M)
Modelo2

##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = M)
##
## Coefficients:
## (Intercept)     Estatura        SexoM
##      -74.75        89.26       -10.56
```

Peso = -74.75 + 89.26*Estatura -10.56SexoM

Mujeres: Y = -73.75 - 10.56 + 89.26*Estatura

Hombres: Y = -73.75 + 89.26*Estatura

Va a ser 1 para mujer y 0 para hombre

```
summary(Modelo2)

##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.9505  -3.2491   0.0489   3.2880  17.1243
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -74.7546     7.5555  -9.894   <2e-16 ***
## Estatura     89.2604     4.5635  19.560   <2e-16 ***
## SexoM       -10.5645     0.6317 -16.724   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.381 on 437 degrees of freedom
## Multiple R-squared:  0.7837, Adjusted R-squared:  0.7827
## F-statistic: 791.5 on 2 and 437 DF,  p-value: < 2.2e-16
```

a) Significancia del modelo con alfa de 0.03

```
if (summary(Modelo2)$fstatistic[1] > qf(1-0.03, df1 =
summary(Modelo2)$fstatistic[2], df2 = summary(Modelo2)$fstatistic[3])) {
  print("El modelo es significativo con alfa de 0.03.")
} else {
  print("El modelo no es significativo con alfa de 0.03.")
}

## [1] "El modelo es significativo con alfa de 0.03."
```

b) Significancia de $\widehat{B}_\iota$ con un alfa de 0.03.

```
coef(summary(Modelo2))[, 4] < 0.03 # Devuelve TRUE si los coeficientes son significativos

## (Intercept)    Estatura       SexoM
##        TRUE        TRUE        TRUE
```

Los parámetros del modelo (Estatura, Sexo) son significativos tanto para $\alpha = 0.05$ como para $\alpha = 0.03$ ya que sus valor p (Pr(>|t|)) son pequeños. Asimismo podemos observar que el p value del modelo es bajo lo que significa que el modelo es significativo.

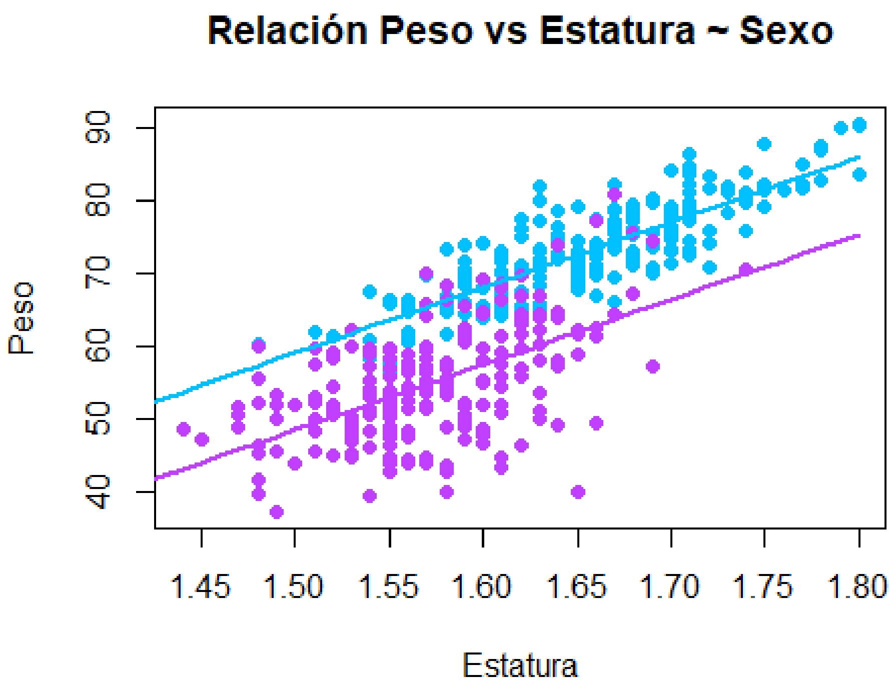
y los modelos quedarían:

Mujeres: * Estatura = -74.75 + 89.26 P + -10.56 SexoM = -85.31 + 0.0052296 P

Hombres: * Estatura = -74.75 + 89.26 P

c) Verificación del porcentaje de variación explicada por el modelo (R^2)

```
paste("El modelo explica el", round(summary(Modelo2)$r.squared * 100, 2), "% de la variabilidad del peso.")

## [1] "El modelo explica el 78.37 % de la variabilidad del peso."

b0 = Modelo2$coefficients[1]
b1 = Modelo2$coefficients[2]
b2 = Modelo2$coefficients[3]
Ym = function(x){b0+b2+b1*x}
Yh = function(x){b0+b1*x}

colores = c( "#00BFFF","#BF3EFF")

plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab = "Peso", xlab = "Estatura", main = "Relación Peso vs Estatura ~ Sexo")
x = seq(1.40, 1.80, 0.01)
lines(x, Ym(x), col = "#BF3EFF", lwd = 2)
lines(x, Yh(x), col = "#00BFFF", lwd = 2)
```



Aparentemente centimetro más en mujeres debería tener el mismo efecto que un centímetro más en hombres.

Hombres: El modelo de regresión lineal indica que por cada centímetro adicional en la estatura, el peso aumenta en promedio 94.66 kg.

Mujeres: Similarmente, el peso aumenta en 81.15 kg por cada centímetro adicional de estatura, pero el R^2 es mucho más bajo, indicando una relación menos fuerte.

$\widehat{B_0}$ representa el peso esperado cuando la estatura es cero, se usa como punto de referencia en la recta.

- Para Hombres: En el modelo de regresión lineal para hombres, $\widehat{B_0}$ es -83.68. Esto significa que, en teoría, si un hombre tuviera una estatura de 0 cm, el peso predicho sería -83.68 kg. Aunque esta situación es imposible y no tiene un sentido práctico (nadie puede tener 0 cm de estatura), el intercepto aún es importante para definir la ecuación de la recta de regresión. En la práctica, el valor del intercepto no suele ser interpretado directamente cuando no tiene sentido contextual, sino que se usa para calcular predicciones de peso a partir de estaturas razonables.
- Para Mujeres: Similarmente, para mujeres, $\widehat{B_0}$ es -72.56. Esto indica que, para una estatura de 0 cm, el peso predicho sería -72.56 kg, lo cual también carece de significado práctico. Sin embargo, este intercepto es necesario para construir la ecuación de la recta de regresión.

$\widehat{B_1}$ representa el cambio promedio en el peso por cada incremento de una unidad en la estatura (cm).

- Para Hombres (Modelo1H): La pendiente $\widehat{B_1}$ es 94.66. Esto significa que, para hombres, por cada incremento de 1 cm en la estatura, se espera que el peso promedio aumente en 94.66 gramos (0.9466 kg). Este coeficiente es estadísticamente significativo, lo que indica que existe una relación fuerte y positiva entre la estatura y el peso para los hombres en los datos analizados.
- Para Mujeres (Modelo1M): La pendiente $\widehat{B_1}$ es 81.15. Para mujeres, esto significa que, por cada incremento de 1 cm en la estatura, se espera que el peso promedio aumente en 81.15 gramos (0.8115 kg). Aunque también es una relación positiva entre estatura y peso, el coeficiente de determinación (R^2) es menor que en el modelo de los hombres, lo que sugiere que la estatura no explica tanto la variación en el peso de las mujeres como lo hace en los hombres.

En el modelo combinado que incluye la variable Sexo, la pendiente para Estatura (89.26) es el efecto promedio de la estatura sobre el peso cuando se considera ambos géneros. Además, la variable SexoM (un indicador binario que toma el valor 1 para mujeres y 0 para hombres) tiene un coeficiente de -10.56. Esto implica que, en promedio, el peso de las mujeres es aproximadamente 10.56 kg menos que el de los hombres, manteniendo constante la estatura. Este modelo muestra que el efecto de la estatura sobre el peso es similar en hombres y mujeres, pero hay una diferencia base entre ambos géneros.

Regresión Lineal - Con interacción

Propón un nuevo modelo. Esta vez toma en cuenta la interacción de la Estatura con el Sexo

```
Modelo3 = lm(Peso~Estatura*Sexo, M)
Modelo3
```

```
##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Coefficients:
##    (Intercept)        Estatura            SexoM  Estatura:SexoM
##         -83.68           94.66            11.12          -13.51
```

Peso: -83.68 + 94.66*Estatura + 11.12SexoM -13.51EstaturaSexoM

- Mujeres: Y = -83.68 + 11.12 + 94.66*Estatura -13.51Estatura
- Hombres: Y = -83.68 + 94.66*Estatura

Obtén el modelo e interpreta las variables Dummy

```r
summary(Modelo3)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -83.685      9.735  -8.597   <2e-16 ***
## Estatura          94.660      5.882  16.092   <2e-16 ***
## SexoM             11.124     14.950   0.744    0.457
## Estatura:SexoM   -13.511      9.305  -1.452    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16
```

Significancia del modelo

Valida la significancia del modelo con un alfa de 0.03 (incluye las hipótesis que pruebas)

- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$

```r
if (summary(Modelo3)$fstatistic[1] > qf(1-0.03, df1 =
summary(Modelo3)$fstatistic[2], df2 = summary(Modelo3)$fstatistic[3])) {
  print("El modelo es significativo con alfa de 0.03.")
} else {
  print("El modelo no es significativo con alfa de 0.03.")
}
```

```
## [1] "El modelo es significativo con alfa de 0.03."
```

Validala significancia de $\widehat{B_i}$ con un alfa de 0.03 (incluye las hipótesis que pruebas)

Hipótesis * $H_0: \beta_i = 0$ * $H_1: \exists \beta_i \neq 0$

```r
coef(summary(Modelo3))[, 4] < 0.03 # Devuelve TRUE si los coeficientes
son significativos
```

```
##    (Intercept)       Estatura          SexoM Estatura:SexoM
##           TRUE           TRUE          FALSE          FALSE
```

Se rechaza $H_0: \beta_i = 0$

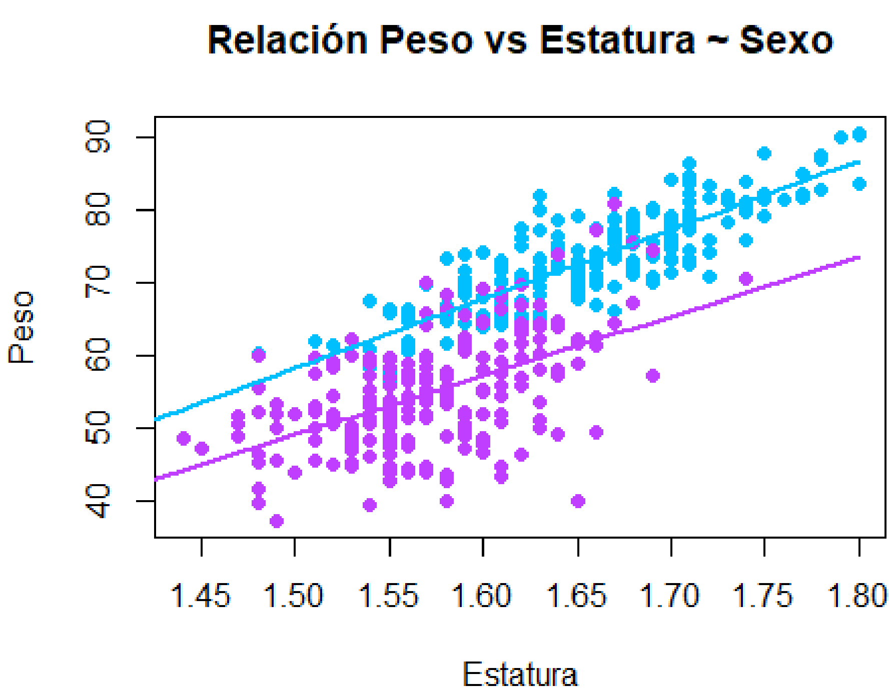
En los parámetros del modelo el intercepto y la estatura son significativos para $\alpha = 0.03$ ya que sus valor p (Pr(>|t|)) son pequeños, sin embargo SexoM y Estatura:SexoM tienen valores p mayores a 0.03 por lo cual no son significativos para el modelo.

Indica cuál es el porcentaje de variación explicada por el modelo.

```
paste("El modelo explica el", round(summary(Modelo3)$r.squared * 100, 2),
"% de la variabilidad del peso.")

## [1] "El modelo explica el 78.47 % de la variabilidad del peso."
```

Dibuja el diagrama de dispersión de los datos y la recta de mejor ajuste.

```
b0 = Modelo3$coefficients[1]
b1 = Modelo3$coefficients[2]
b2 = Modelo3$coefficients[3]
b3 = Modelo3$coefficients[4]
Ym = function(x){b0+b2+(b1+b3)*x}
Yh = function(x){b0+b1*x}

colores = c( "#00BFFF","#BF3EFF")

plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =
"Peso", xlab = "Estatura", main = "Relación Peso vs Estatura ~ Sexo")
x = seq(1.40, 1.80, 0.01)
lines(x, Ym(x), col = "#BF3EFF", lwd = 2)
lines(x, Yh(x), col = "#00BFFF", lwd = 2)
```



$\widehat{B_0}$ representa el peso esperado cuando la estatura es cero, se usa como punto de referencia en la recta.

$\widehat{B_1}$ representa el cambio promedio en el peso por cada incremento de una unidad en la estatura (cm).

$\widehat{B_2}$ representa el cambio en el peso en base al género.

$\widehat{B_3}$ representa el cambio promedio en el peso por cada incremento de una unidad en la estatura (cm) de la interacción.

En este nuevo modelo en donde consideramos la interacción entre Estatura y Sexo, la pendiente cambia para la mujer y para el hombre, para el hombre la pendiente para estatura es 94.66, este es el efecto promedio de estatura sobre el peso en hombres, mientras que para mujeres a parte del 94.66 consideramos -13.51 de la interacción obteniendo que el efecto promedio de la estutura sobre el peso para las mujeres (la pendiente para estatura) es 81.15. Además, la variable SexoM tiene un coeficiente de 11.12. Este modelo muestra que el efecto de la estatura sobre el peso es diferente en hombres y mujeres, sin embargo no hay una diferencia extravagante, al contrario es pequeña y demuestra la flexibilidad existente en este tercer modelo, asimismo hay una diferencia base entre ambos géneros.

Considero que este último modelo es más apropiado por su flexibilidad y el modelo explica el 78.47 % de la variabilidad del peso, que es más que los modelos anteriores.

Regresión Lineal - Análisis de los resultados

Analiza si el (los) modelo(s) obtenidos anteriormente son apropiados para el conjunto de datos. Realiza el análisis de los residuos:

Normalidad de los residuos

Prueba de Hipótesis

- H_0 = La muestra proviene de una distribución normal
- H_1 = La muestra no proviene de una distribución normal

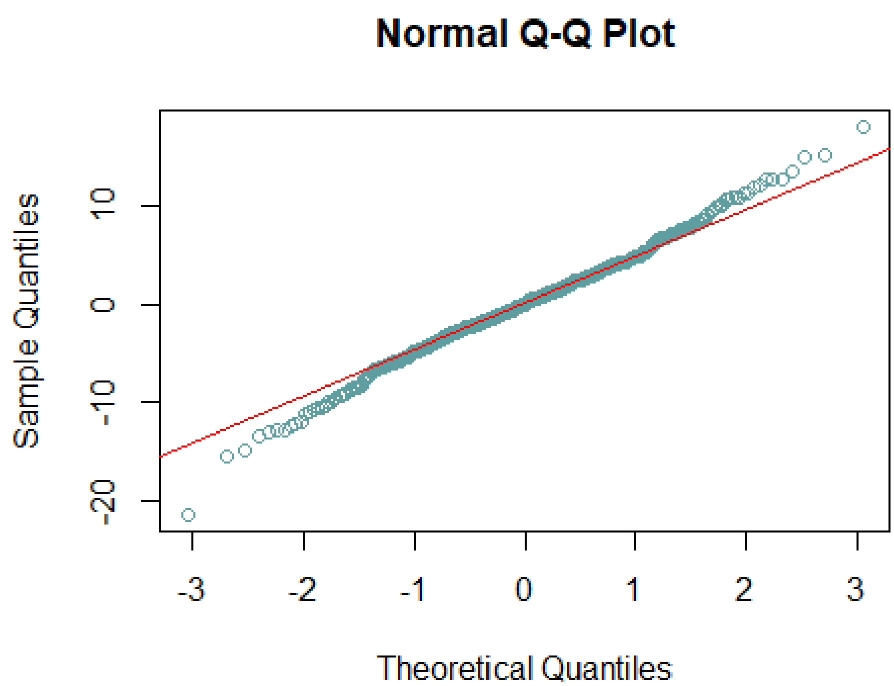
Regla de decisión: Se rechaza H_0 si valor p < α

```r
library(nortest)
ad.test(Modelo3$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  Modelo3$residuals
## A = 0.8138, p-value = 0.03516
```

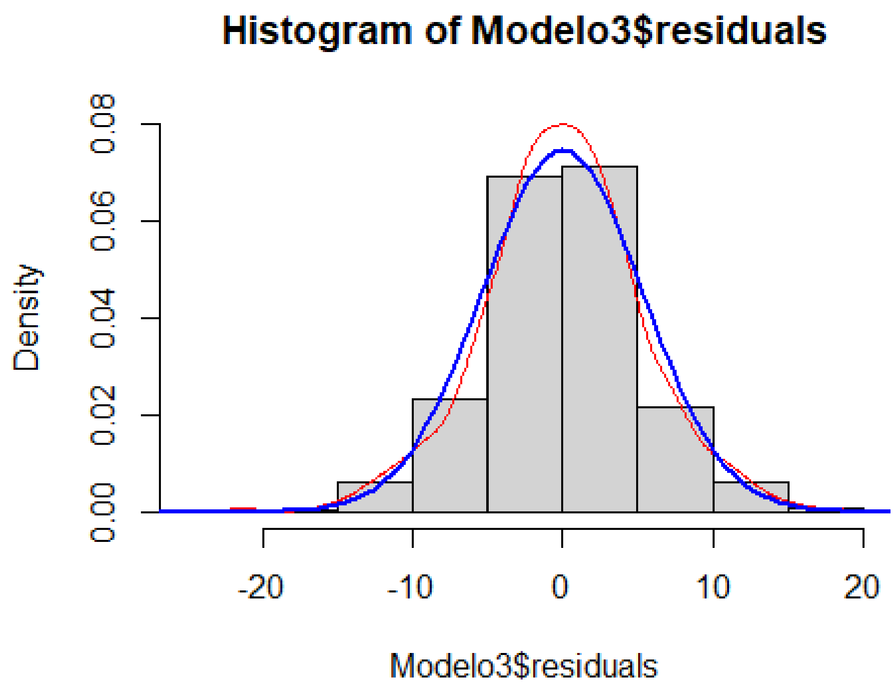
Aceptamos H_0 ya que el valor p = 0.03516 > $\alpha = 0.03$ por lo que podemos decir que nuestros datos provienen de una distribución normal

```
qqnorm(Modelo3$residuals, col = "#5F9EA0")
qqline(Modelo3$residuals, col = "red")
```



Esto lo podemos reforzar mediante el qqplot ya que podemos observar en nuestra gráfica que la mayoría de los datos se encuentran en la línea, únicamente en las colas se separan un poco de la línea sin embargo como la mayoría de los datos se encuentran en la línea y no hay una separación muy significativa podemos argumentar que se observa normalidad.

```
hist(Modelo3$residuals,freq=FALSE, ylim = c(0,.08))
lines(density(Modelo3$residual),col="red")
curve(dnorm(x,mean=mean(Modelo3$residuals),sd=sd(Modelo3$residuals)),
from=-30, to=30, add=TRUE, col="blue",lwd=2)
```



Nuevamente observamos a los datos comportarse con normalidad, los datos se distribuyen con forma de campana, concentrandose la mayoría en el centro. Respecto al sesgo podemos observar que hay simetría y la curtosis nos indica que existe una distribución normal (mesocúrtica).

Verificación de media cero

Prueba de Hipótesis

- $H_0: \mu_e = 0$
- $H_1: \mu_e \neq 0$

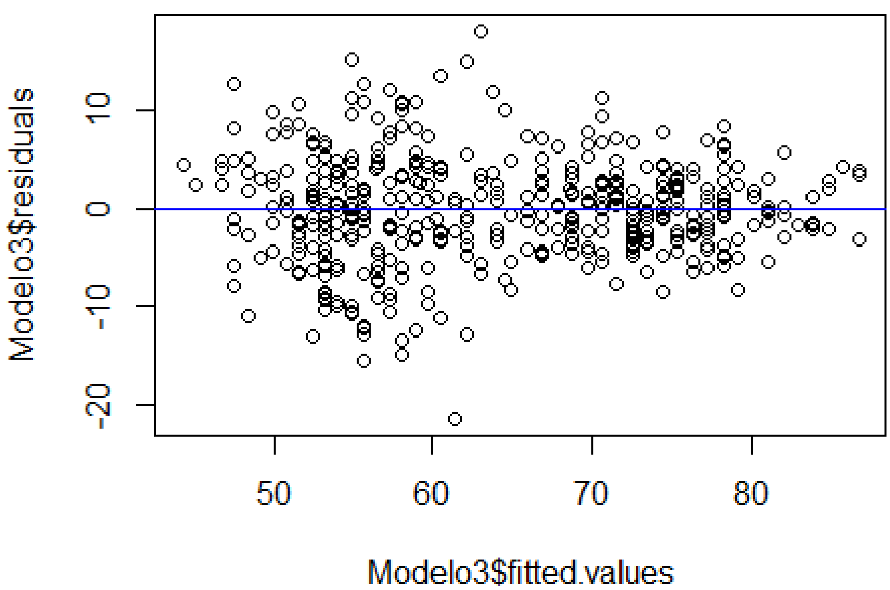
Regla de decisión * Se rechaza si valor p < α

```
t.test(Modelo3$residuals)

##
##  One Sample t-test
##
## data:  Modelo3$residuals
## t = -8.5817e-16, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.5017741  0.5017741
## sample estimates:
##     mean of x
## -2.190956e-16
```

Acepttamos H_0 ya que nuestro valor p = 1 > $\alpha = 0.03$ entonces podemos concluir que $\mu_e = 0$. Los residuos tienen media cero: el modelo es bueno.

Homocedasticidad e independencia

```
plot(Modelo3$fitted.values,Modelo3$residuals)
abline(h=0, col="blue")
```



Los residuos no muestran ninguna estructura evidente lo que indica que el modelo de regresión es adecuado y existe linealidad y homocedasticidad.

Pruebas de hipótesis para independencia

Test de Durbin-Watson y Prueba Breusch-Godfrey

- H_0: Los errores no están autocorrelacionados.
- H_1: Los errores están autocorrelacionados.

Regla de decisión: Se rechaza H_0 si valor p < α

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
dwtest(Modelo3)
```

```
##
##  Durbin-Watson test
##
## data:  Modelo3
## DW = 1.8646, p-value = 0.07113
## alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(Modelo3)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  Modelo3
## LM test = 1.3453, df = 1, p-value = 0.2461
```

No rechazo H_0 ya que valor p = 0.07 y 0.2461 > α = 0.03 lo que significa que los errores no están autocorrelacionados.

Pruebas de hipótesis para homocedasticidad

Prueba de Breusch-Pagan y White

- H_0: La varianza de los errores es constante (homocedasticidad)
- H_1: La varianza de los errores no es constante (heterocedasticidad)

Regla de decisión: Se rechaza H_0 si valor p < α

```r
library(lmtest)
bptest(Modelo3)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Modelo3
## BP = 59.211, df = 3, p-value = 8.667e-13
```

```r
gqtest(Modelo3)
```

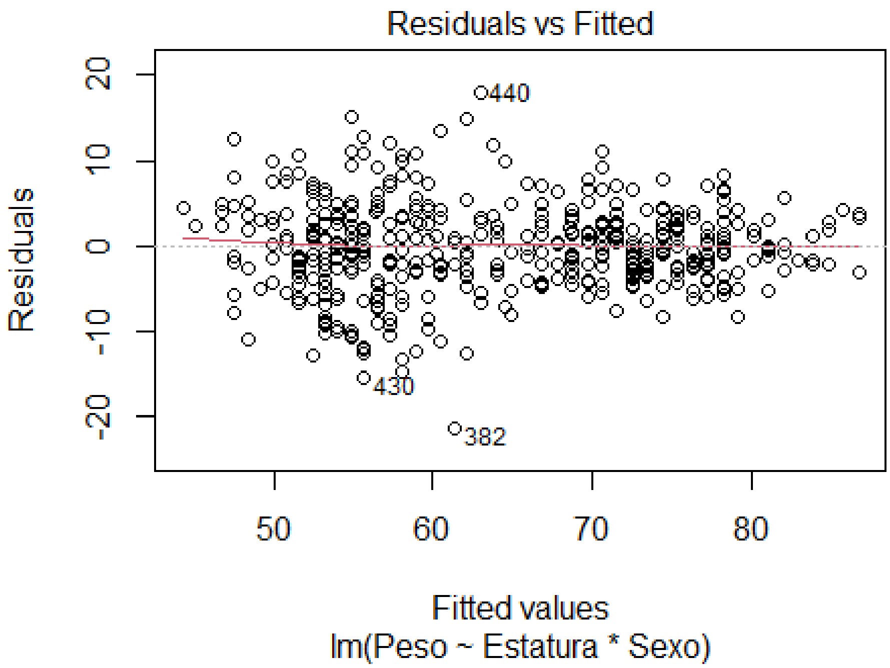
```
##
##  Goldfeld-Quandt test
##
## data:  Modelo3
## GQ = 3.2684, df1 = 216, df2 = 216, p-value < 2.2e-16
## alternative hypothesis: variance increases from segment 1 to 2
```

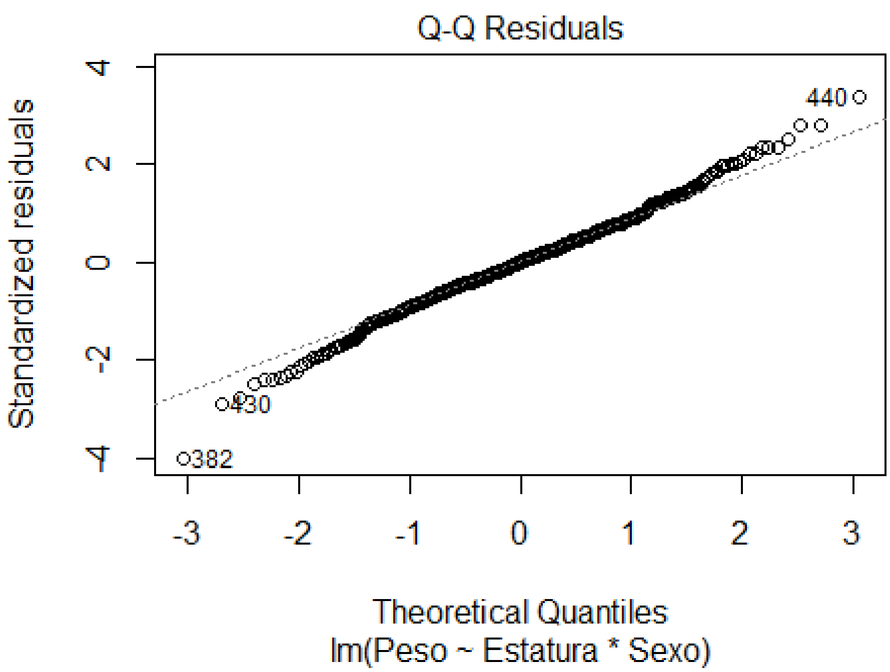
Apesar de que en la gráfica nos aparenta homocedasticidad al hacer las pruebas de hipótesis rechazamos H_0 ya que mis valores p para ambas pruebas son muy pequeñas

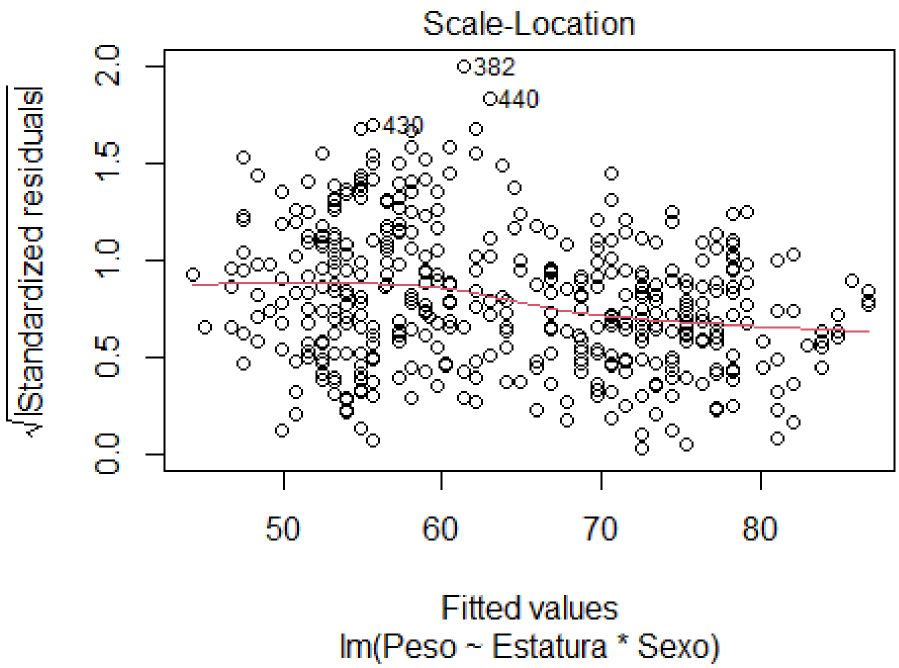
< α = 0.03 lo que significa que la varianza de los errores no es constante y existe heterocedasticidad.

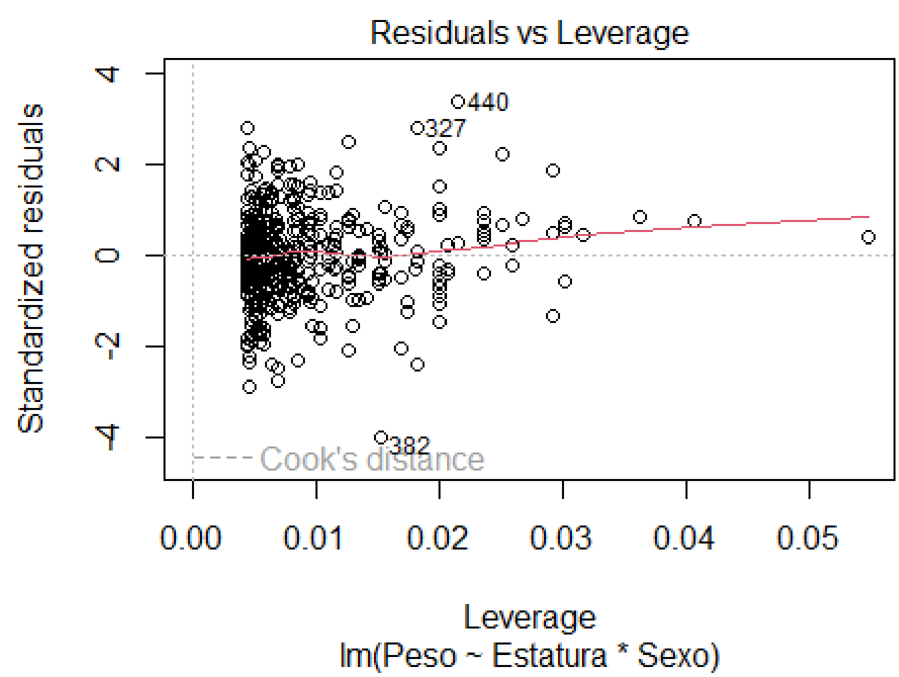
Utiliza el comando: plot(modelo). Observa las gráficas obtenidas y contesta:

```r
plot(Modelo3)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Peso ~ Estatura * Sexo)
440
430
382


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Peso ~ Estatura * Sexo)
440
430
382





Las gráficas obtenidas mmediante plot(Modelo3) son similares a las gráficas que ya teníamos, en lo que difieren es que se agregan las gráficas escaladas y podemos ver como fluctuan los datos, sin embargo a grandes rasgos podemos observar que sí se cumple que la media se encuentra en el cero como lo habíamos observado anteriormente por lo que las conclusiones únicamente considerando las gráficas realizadas sería la misma que con las gráficas anteriores. Sin embargo este análisis nos deja como aprendizaje el no confiarnos únicamente de las gráficas ya que si hubieramos observado solo estas o las gráficas anteriores confiariamos en que nuestro modelo cumple hocedasticidad, lo cual es falso ya que al realizar la prueba de hipótesis observamos que esto se rechaza y que en realidad nuestro modelo tiene heterocedasticidad. Comparando resultados encontramos que los mejores modelos son los que únicamente considera a los hombres o únicamente a las mujeres.

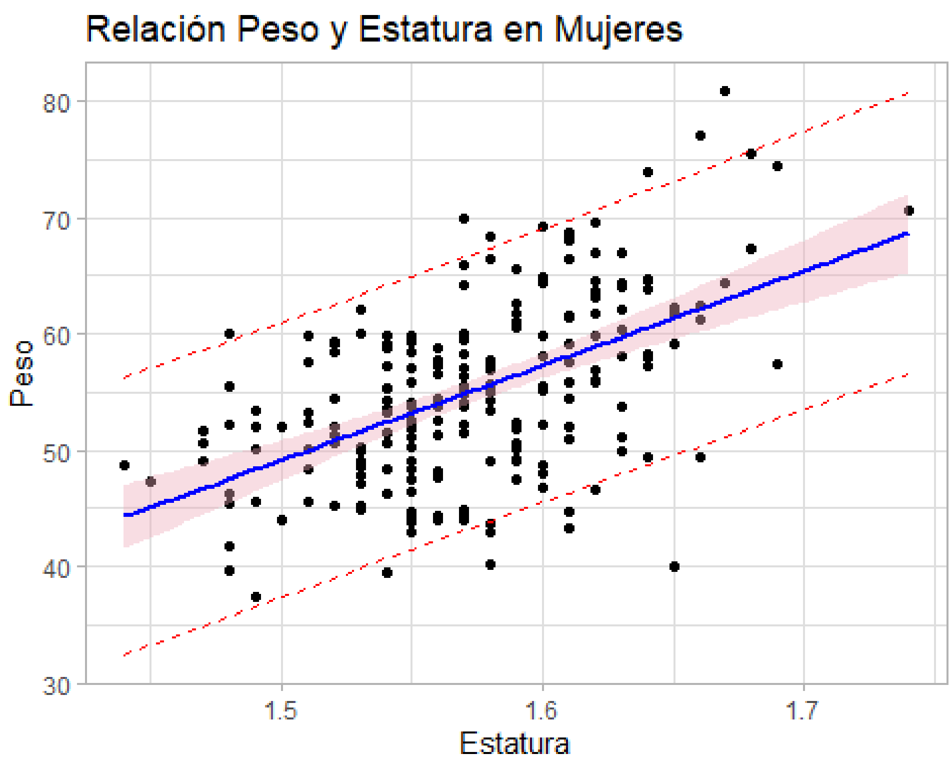
Intervalos de confianza

Con los datos de las estaturas y pesos de los hombres y las mujeres construye la gráfica de los intervalos de confianza y predicción para la estimación y predicción de Y para el mejor modelo seleccionado.

```
Ip = predict(object=Modelo3,interval="prediction",level=0.97)
```

```
## Warning in predict.lm(object = Modelo3, interval = "prediction", level
= 0.97): predictions on current data refer to _future_ responses
```

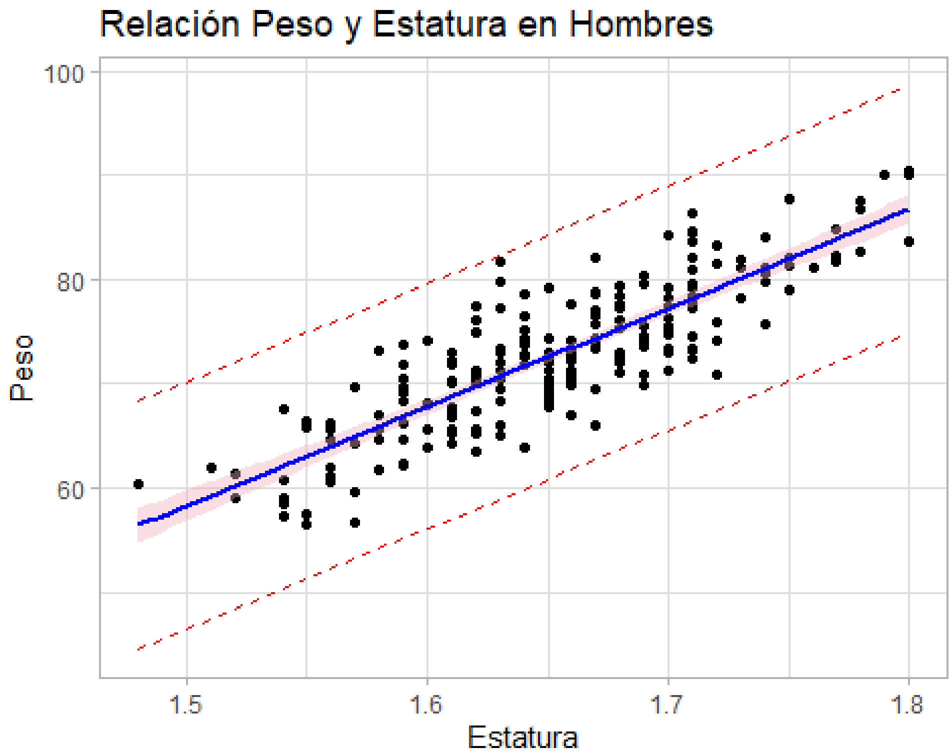
```
M2 = cbind(M,Ip)
M2m = subset(M2, Sexo == "M")
M2h = subset(M2, Sexo == "H")
library(ggplot2)
ggplot(M2m,aes(x=Estatura,y=Peso))+
  ggtitle("Relación Peso y Estatura en Mujeres")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_line(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  theme_light()
```



La gráfica muestra la relación entre el peso y la estatura de mujeres. Cada punto negro representa un individuo. La línea azul representa la línea de regresión lineal que mejor se ajusta a los datos, indicando una relación positiva entre la estatura y el peso. La banda sombreada en color rosa alrededor de la línea azul muestra el intervalo de confianza del 97% para la estimación de la línea de regresión. Esto implica que hay un 97% de probabilidad de que la verdadera línea de regresión esté dentro de esa banda. Las líneas rojas discontinuas muestran los límites de predicción para el 97% de los datos. Esto significa que, para una estatura dada, el 97% de los pesos esperados deberían caer dentro de estas líneas.

Existe una correlación positiva entre la estatura y el peso en mujeres. A medida que la estatura aumenta, el peso tiende a incrementarse. La dispersión de los puntos indica que, aunque hay una tendencia positiva, hay variabilidad considerable en el peso para una misma estatura. Los intervalos de confianza y de predicción son más amplios en los extremos, lo que sugiere mayor incertidumbre en las estimaciones para estaturas más bajas y más altas.

```
ggplot(M2h,aes(x=Estatura,y=Peso))+
  ggtitle("Relación Peso y Estatura en Hombres")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_line(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  theme_light()
```



Similar a la gráfica de las mujeres, esta gráfica muestra la relación entre el peso y la estatura, pero para hombres. Se observa una relación positiva más clara entre la estatura y el peso en hombres, con menor dispersión de puntos en comparación con la gráfica de mujeres, lo que sugiere una relación más fuerte. Los intervalos de confianza y los límites de predicción son más estrechos en esta gráfica, lo que indica mayor certeza en la predicción de peso a partir de la estatura para hombres.

Comparación entre Hombres y Mujeres: Aunque ambos géneros presentan una correlación positiva entre peso y estatura, la relación es más fuerte y predecible en los hombres que en las mujeres. La dispersión en la gráfica de las mujeres es mayor, lo que sugiere que el peso de las mujeres puede estar influenciado por más factores aparte de la estatura en comparación con los hombres.