

10Es_A01571214_Lautaro_Coteja

A01571214 - Lautaro Coteja

2024-09-06

R Markdown

Estadística - Actividad / Tarea 10: Regresión Lineal

Parte 1

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

```
M = read.csv("C:/Users/lauda/Downloads/Estatura-peso_HyM.csv")

# Separar Los datos por sexo
MM = subset(M, M$Sexo == "M")
MH = subset(M, M$Sexo == "H")

# Crear un dataframe con las variables por sexo
M1 = data.frame(MH$Estatura, MH$Peso, MM$Estatura, MM$Peso)

head(M1)

##      MH.Estatura MH.Peso MM.Estatura MM.Peso
## 1          1.61   72.21          1.53   50.07
## 2          1.61   65.71          1.60   59.78
## 3          1.70   75.08          1.54   50.66
## 4          1.65   68.55          1.58   56.96
## 5          1.72   70.77          1.61   51.03
## 6          1.63   77.18          1.57   64.27

n = 4
d = matrix(NA, ncol = 7, nrow = n)

# Calcular las medidas descriptivas para cada variable
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", "Maximo", "Desv Est")
```

```
# Mostrar el dataframe con las medidas descriptivas
```

```
m
```

```
##           Minimo      Q1 Mediana      Media      Q3 Maximo      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
```

```
# Calcular la matriz de correlacion para hombres, mujeres y combinado
```

```
correlacion_hombres = cor(MH[, c("Estatura", "Peso")])
```

```
correlacion_mujeres = cor(MM[, c("Estatura", "Peso")])
```

```
correlacion_combinado = cor(M[, c("Estatura", "Peso")])
```

```
correlacion_hombres
```

```
##           Estatura      Peso
## Estatura 1.0000000 0.8468348
## Peso     0.8468348 1.0000000
```

```
correlacion_mujeres
```

```
##           Estatura      Peso
## Estatura 1.0000000 0.5244962
## Peso     0.5244962 1.0000000
```

```
correlacion_combinado
```

```
##           Estatura      Peso
## Estatura 1.0000000 0.8032449
## Peso     0.8032449 1.0000000
```

```
# Regresion hombres
```

```
A_hombres = lm(MH$Peso ~ MH$Estatura)
```

```
# Regresion mujeres
```

```
A_mujeres = lm(MM$Peso ~ MM$Estatura)
```

```
# Regresion combinada
```

```
A_combinado = lm(Peso ~ Estatura, data = M)
```

```
# Validacion del modelo para hombres
```

```
summary(A_hombres)
```

```
##
```

```
## Call:
```

```
## lm(formula = MH$Peso ~ MH$Estatura)
```

```
##
```

```
## 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 ***
## MM$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

# Validacion del modelo para mujeres
summary(A_mujeres)

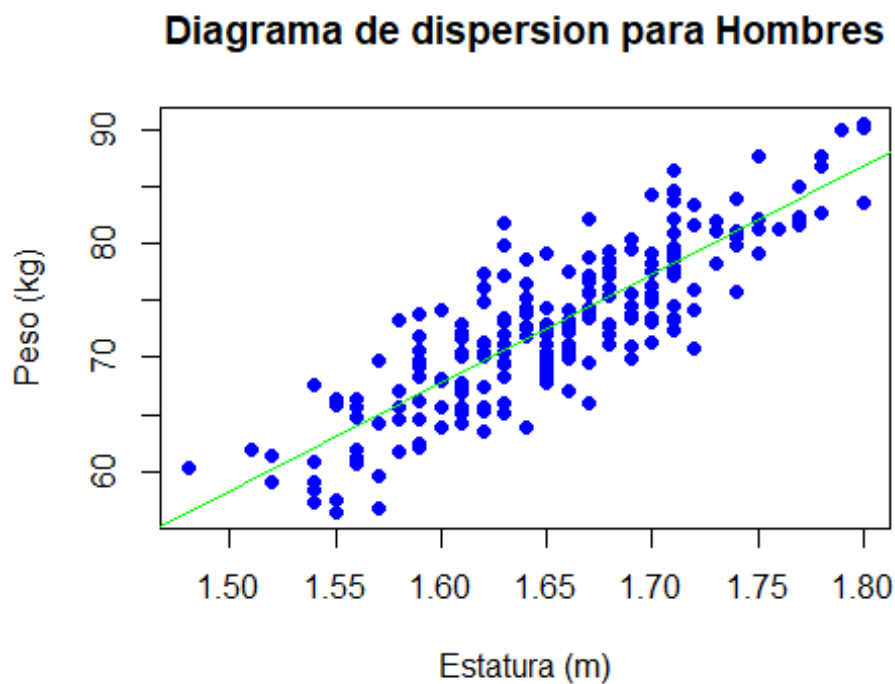
##
## Call:
## lm(formula = MM$Peso ~ MM$Estatura)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -4.1942   0.4004   4.2724  17.9114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -72.560      14.041  -5.168 5.34e-07 ***
## MM$Estatura   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

# Validacion del modelo combinado
summary(A_combinado)

##
## Call:
## lm(formula = Peso ~ Estatura, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.8653  -3.7654   0.6706   5.0142  15.6006
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -151.883      7.655  -19.84  <2e-16 ***
## Estatura    133.793      4.741   28.22  <2e-16 ***
## ---
```

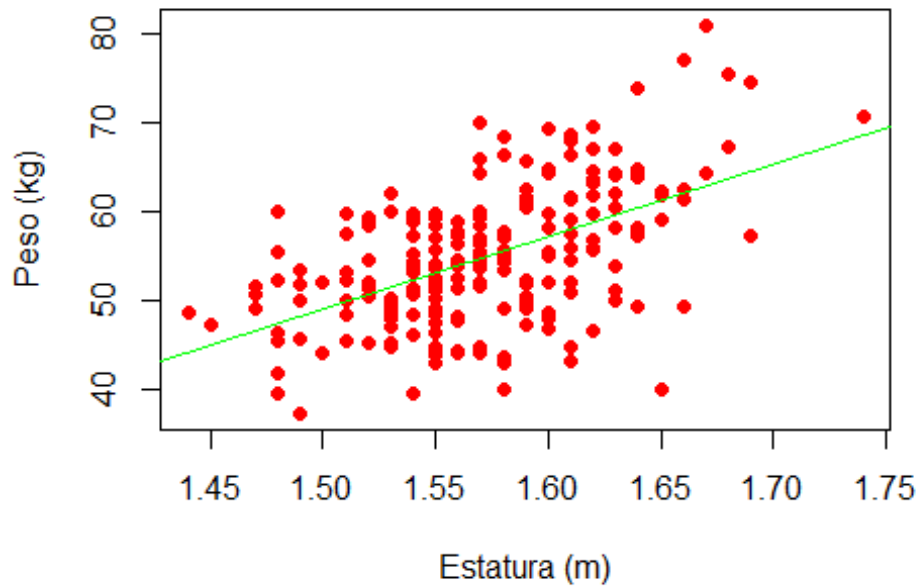
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.883 on 438 degrees of freedom
## Multiple R-squared:  0.6452, Adjusted R-squared:  0.6444
## F-statistic: 796.5 on 1 and 438 DF,  p-value: < 2.2e-16

# Diagrama de dispersion con recta de mejor ajuste para hombres
plot(MH$Estatura, MH$Peso, main="Diagrama de dispersion para Hombres",
     xlab="Estatura (m)", ylab="Peso (kg)", pch=19, col="blue")
abline(A_hombres, col="green")
```



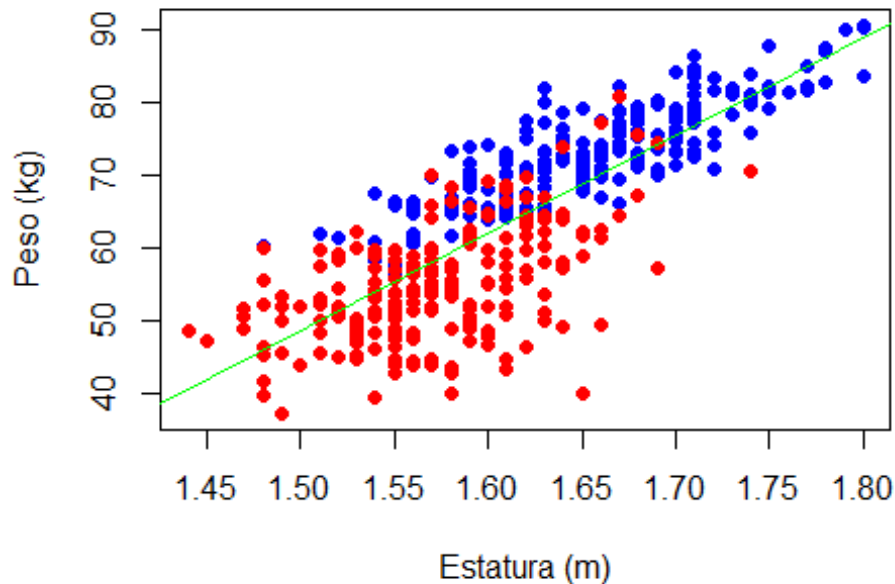
```
# Diagrama de dispersion con recta de mejor ajuste para mujeres
plot(MM$Estatura, MM$Peso, main="Diagrama de dispersion para Mujeres",
     xlab="Estatura (m)", ylab="Peso (kg)", pch=19, col="red")
abline(A_mujeres, col="green")
```

Diagrama de dispersión para Mujeres



```
# Diagrama de dispersion con recta de mejor ajuste para ambos sexos
plot(M$Estatura, M$Peso, main="Diagrama de dispersión para Hombres y
Mujeres", xlab="Estatura (m)", ylab="Peso (kg)", pch=19, col=ifelse(M$Sexo ==
"H", "blue", "red"))
abline(A_combinado, col="green")
```

Diagrama de dispersión para Hombres y Mujeres



```
# Prueba de normalidad para Los residuos de hombres
```

```
shapiro.test(A_hombres$residuals)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  A_hombres$residuals
```

```
## W = 0.99356, p-value = 0.4597
```

```
# Prueba de normalidad para Los residuos de mujeres
```

```
shapiro.test(A_mujeres$residuals)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  A_mujeres$residuals
```

```
## W = 0.99659, p-value = 0.9144
```

```
# Prueba de normalidad para Los residuos del modelo combinado
```

```
shapiro.test(A_combinado$residuals)
```

```
##
```

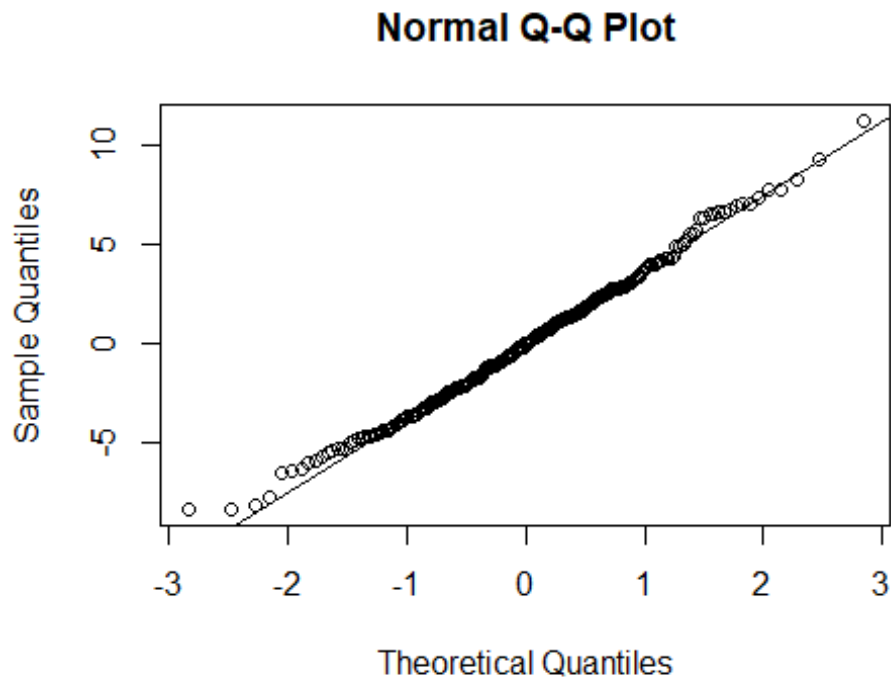
```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  A_combinado$residuals
```

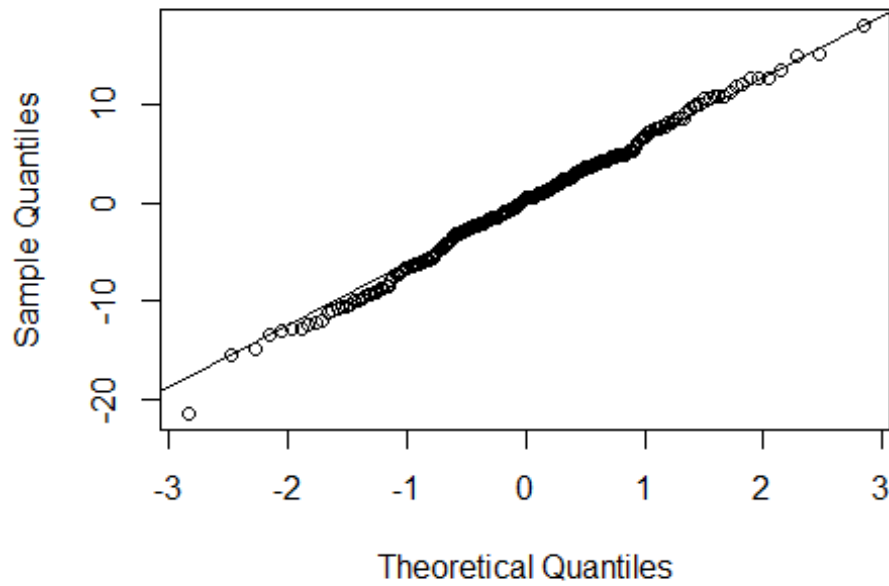
```
## W = 0.97683, p-value = 1.803e-06
```

```
# QQ plot para Los residuos de hombres  
qqnorm(A_hombres$residuals)  
qqline(A_hombres$residuals)
```



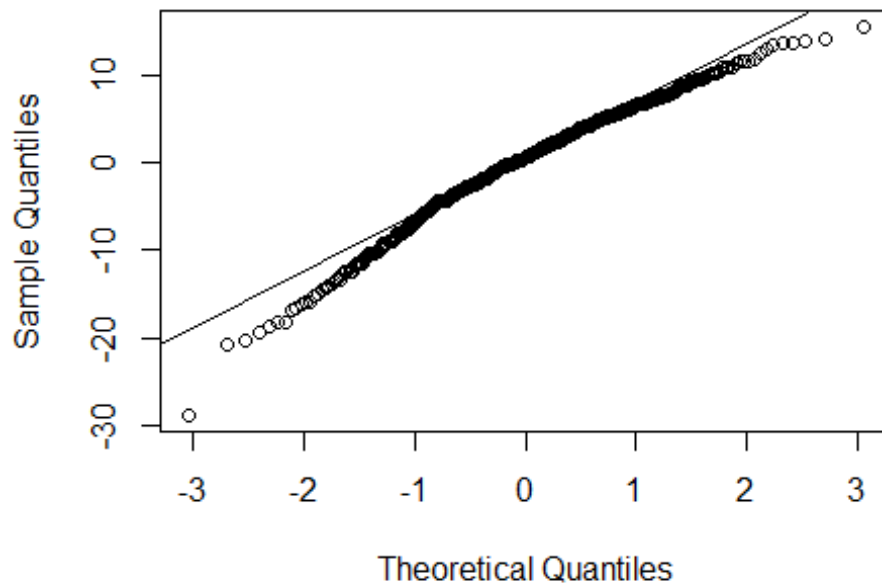
```
# QQ plot para Los residuos de mujeres  
qqnorm(A_mujeres$residuals)  
qqline(A_mujeres$residuals)
```

Normal Q-Q Plot

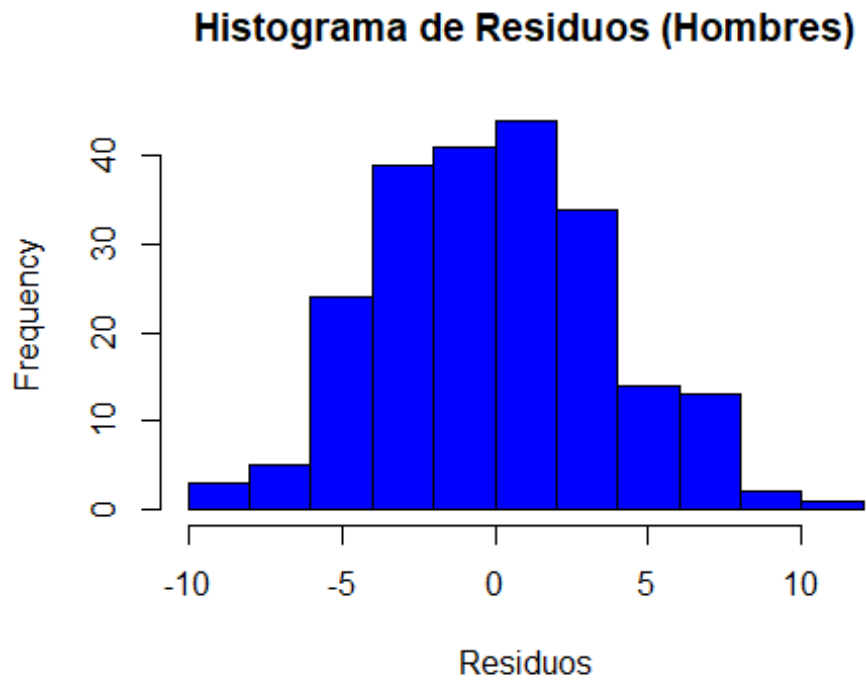


```
# QQ plot para los residuos del modelo combinado  
qqnorm(A_combinado$residuals)  
qqline(A_combinado$residuals)
```

Normal Q-Q Plot

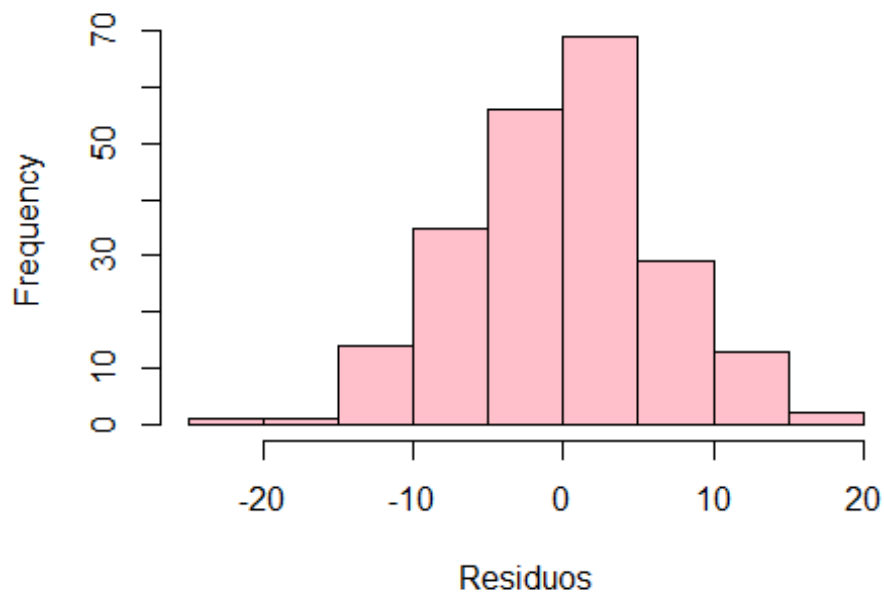



```
# Histograma para los residuos de hombres  
hist(A_hombres$residuals, main = "Histograma de Residuos (Hombres)", xlab =  
"Residuos", col = "blue")
```



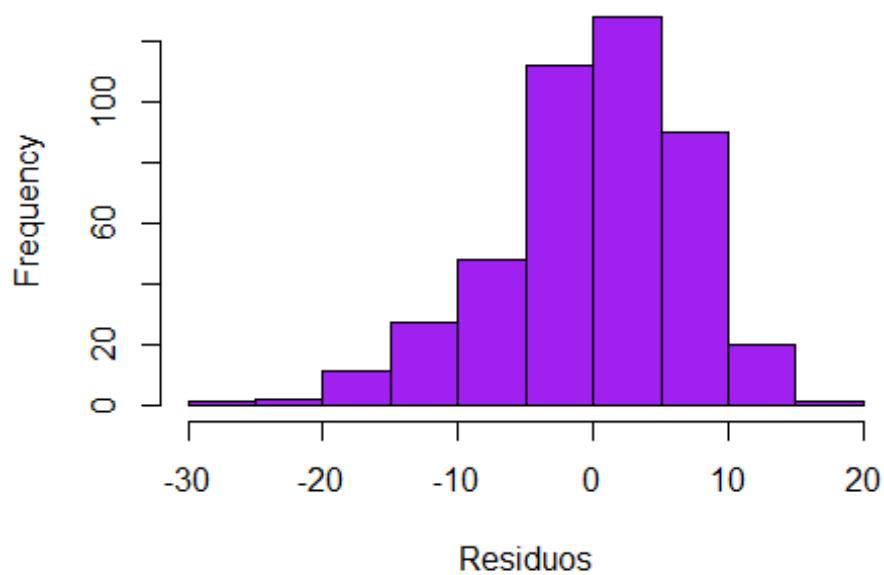
```
# Histograma para los residuos de mujeres  
hist(A_mujeres$residuals, main = "Histograma de Residuos (Mujeres)", xlab =  
"Residuos", col = "pink")
```

Histograma de Residuos (Mujeres)



```
# Histograma para los residuos del modelo combinado  
hist(A_combinado$residuals, main = "Histograma de Residuos (Combinado)", xlab  
= "Residuos", col = "purple")
```

Histograma de Residuos (Combinado)



```

# Prueba t para verificar si La media de Los residuos es cero (hombres)
t.test(A_hombres$residuals)

##
## One Sample t-test
##
## data: A_hombres$residuals
## t = 4.5495e-16, df = 219, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.4876507 0.4876507
## sample estimates:
## mean of x
## 1.125698e-16

# Prueba t para verificar si La media de Los residuos es cero (mujeres)
t.test(A_mujeres$residuals)

##
## One Sample t-test
##
## data: A_mujeres$residuals
## t = -3.9979e-16, df = 219, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.881609 0.881609
## sample estimates:
## mean of x
## -1.788342e-16

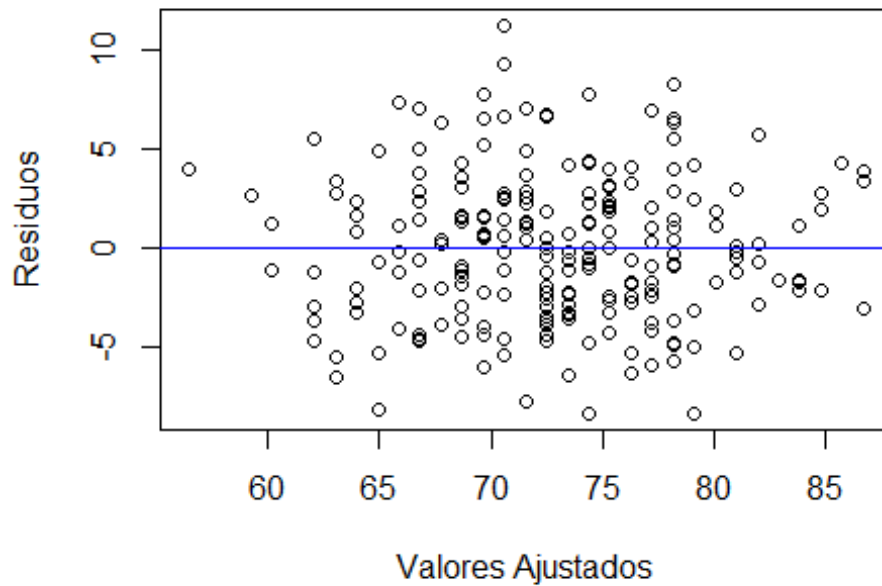
# Prueba t para verificar si La media de Los residuos es cero (combinado)
t.test(A_combinado$residuals)

##
## One Sample t-test
##
## data: A_combinado$residuals
## t = 2.7844e-15, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.6441362 0.6441362
## sample estimates:
## mean of x
## 9.12569e-16

# Grafico (hombres)
plot(A_hombres$fitted.values, A_hombres$residuals, main = "Residuos vs
Valores Ajustados (Hombres)", xlab = "Valores Ajustados", ylab = "Residuos")
abline(h = 0, col = "blue")

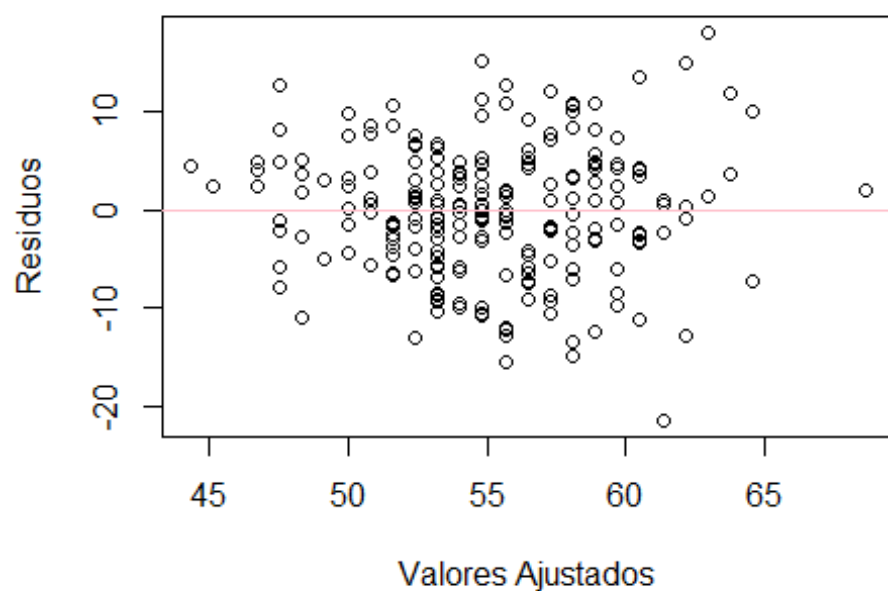
```

Residuos vs Valores Ajustados (Hombres)



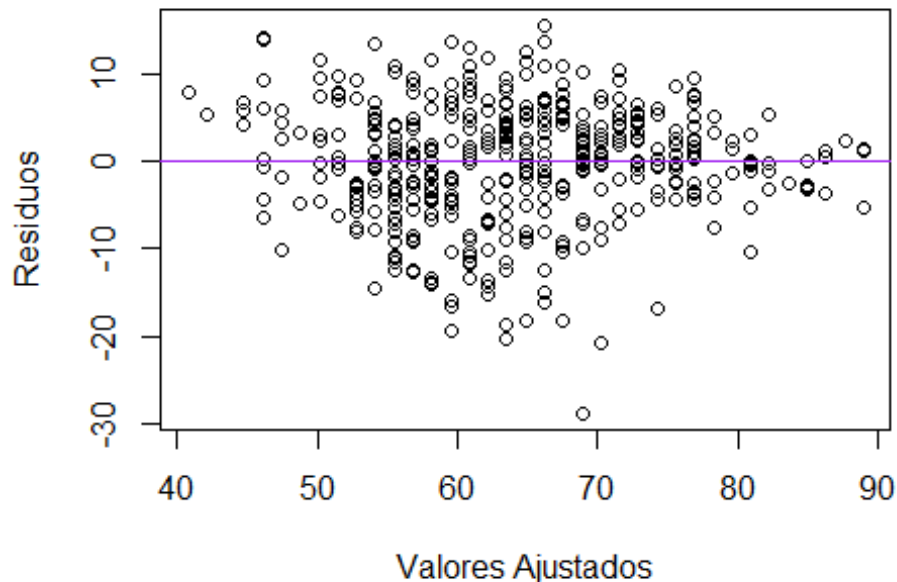
```
# Grafico (mujeres)
plot(A_mujeres$fitted.values, A_mujeres$residuals, main = "Residuos vs
Valores Ajustados (Mujeres)", xlab = "Valores Ajustados", ylab = "Residuos")
abline(h = 0, col = "pink")
```

Residuos vs Valores Ajustados (Mujeres)



```
# Grafico (combinado)
plot(A_combinado$fitted.values, A_combinado$residuals, main = "Residuos vs
Valores Ajustados (Combinado)", xlab = "Valores Ajustados", ylab =
"Residuos")
abline(h = 0, col = "purple")
```

Residuos vs Valores Ajustados (Combinado)



Conclusion En

base a el analisis hecho hasta ahora, se concluye que la estatura es un predictor importante del peso tanto en hombres como en mujeres. Por los modelos, se nota que un modelo especifico para cada sexo es mejor que un modelo combinado.

Parte 2 CON INTERACCION

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

Hipotesis (Modelo)

H0: El modelo no es significativo, todos los coeficientes son igual a 0.

H1: Al menos uno de los coeficientes no es igual a 0.

```
summary(modelo_interaccion)
```

```
##
```

```
## 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

# Interpretacion variables dummy
# El modelo incluira variables dummy para el sexo, estas son codificadas de
manera que uno de los niveles actua como referencia y el otro como variable
dummy.

# R²
summary(modelo_interaccion)$r.squared

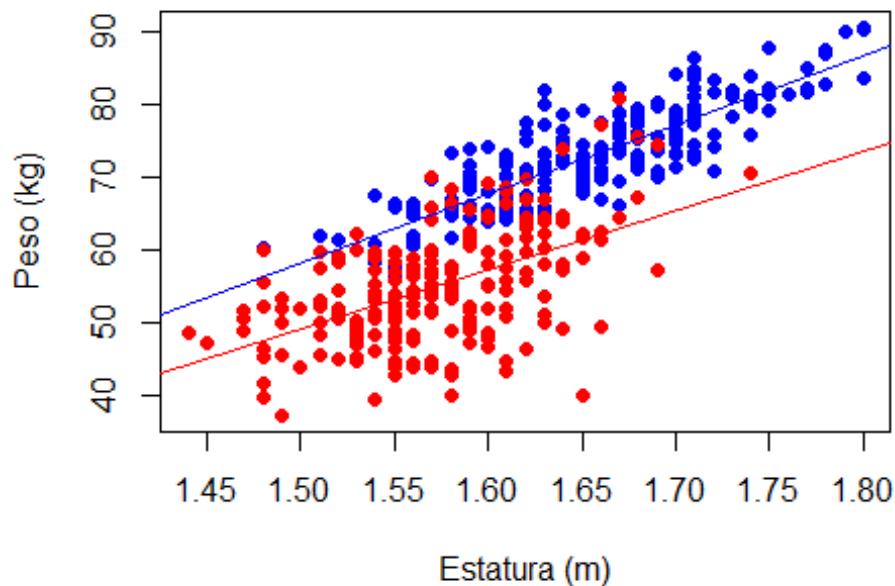
## [1] 0.7847011

# Explicacion
# Nos indica el porcentaje de variacion en el peso que es explicado por la
estatura y la interaccion con el sexo. En base a los resultados, nos dice que
la variabilidad en el peso se explica por la estatura, el sexo y la
interaccion entre estatura y sexo. Tiene mayor R² que los otros modelos, lo
que dice que este tiene el mejor ajuste general, aunque la mejora no vale
tanto debido a que el termino de interaccion no es significativo.

# Hipotesis (Significancia)
# H0:  $\beta_i = 0$  (el coeficiente no es significativo)
# H1:  $\beta_i \neq 0$  (el coeficiente es significativo)

# Diagrama de dispersion con rectas de mejor ajuste diferenciadas por sexo
plot(M$Estatura, M$Peso, col=ifelse(M$Sexo == "H", "blue", "red"), pch=19,
xlab="Estatura (m)", ylab="Peso (kg)", main="Peso vs Estatura con Interacción
por Sexo")
abline(lm(Peso ~ Estatura, data=M[M$Sexo == "H", ]), col="blue")
abline(lm(Peso ~ Estatura, data=M[M$Sexo == "M", ]), col="red")
```

Peso vs Estatura con Interacción por Sexo



Intercepto

En el modelo para hombres da un valor negativo lo cual no es una situación realista, En el para mujeres es lo mismo que el de los hombres, y en el combinado nos da un valor que no es tan preciso y es mas generalizado, igual que en el modelo con interaccion, pero dice que la relacion entre peso y estatura es diferente en base al sexo.

Coefficientes

En el modelo para hombres dice que por cada metro adicional de estatura, el peso aumenta para el hombre, en el caso del modelo para mujeres la relacion es al reves, y en el combinado indica que en promedio, por cada metro adicional de estatura, el peso aumenta tambien, y en el modelo con interaccion dice que el efecto de la estatura en el peso es diferente para las mujeres.

Conclusion

En base a los analisis realizados, dado que el termino de interaccion no es significativo, el modelo combinado sin interaccion puede ser suficiente para describir la relacion entre estatura y peso. Como extra es un modelo mas sencillo y con buena R^2 . Pero si se desea un analisis mas detallado para cada sexo, los modelos separados son los mas convenientes, especialmente porque el R^2 es mas alta en los hombres. Asi que se termina diciendo que

la estatura es un predictor significativo del peso en ambos sexos, pero la fuerza de esta relacion varia entre hombres y mujeres.