

# Actividad 11

José Carlos Sánchez Gómez

2024-09-03

```
M =  
read.csv("C:\\Users\\jcsg6\\Documentos\\Uni\\SeptimoSemestre\\Estadistica  
\\Estatura-peso_HyM.csv")  
MM = subset(M,M$Sexo=="M")  
MH = subset(M,M$Sexo=="H")  
M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)
```

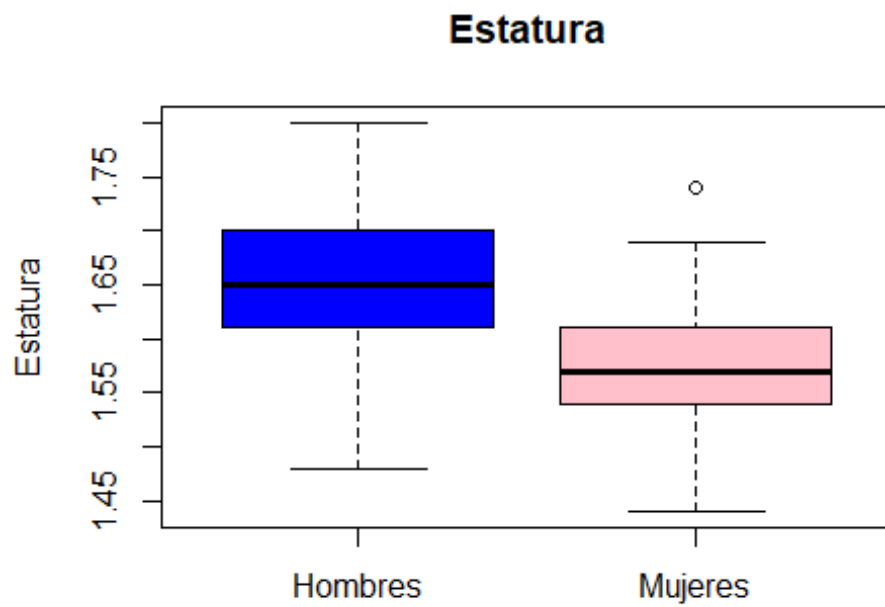
## correlacion

```
cor(M1)
```

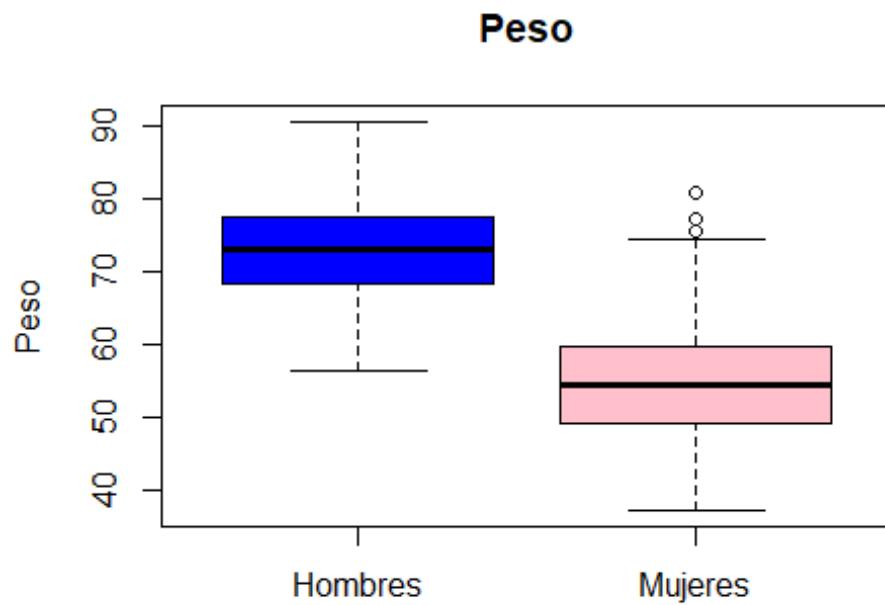
```
##           MH.Estatura    MH.Peso  MM.Estatura    MM.Peso  
## MH.Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872  
## MH.Peso     0.8468347920 1.0000000000 0.0035132246 0.02154907  
## MM.Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621  
## MM.Peso     0.0472487231 0.021549075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

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

```
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(M$Estatura~M$Sexo, ylab="Estatura", xlab="",  
col=c("blue", "pink"), names=c("Hombres", "Mujeres"), main="Estatura")
```



```
boxplot(M$Peso~M$Sexo, ylab="Peso",xlab="", names=c("Hombres",  
"Mujeres"), col=c("blue","pink"), main="Peso")
```



## la recta de mejor ajuste

Encuentra la ecuación de regresión de mejor ajuste:

```
modelo1H = lm(Peso ~ Estatura, data = MH)
modelo1H

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

cat("Variacion explicada (r^2): ", summary(modelo1H)$adj.r.squared)

## Variacion explicada (r^2):  0.7158316

modelo1M = lm(Peso ~ Estatura, data = MM)
modelo1M

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

cat("Variacion explicada (r^2): ", summary(modelo1M)$adj.r.squared)

## Variacion explicada (r^2):  0.271771
```

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

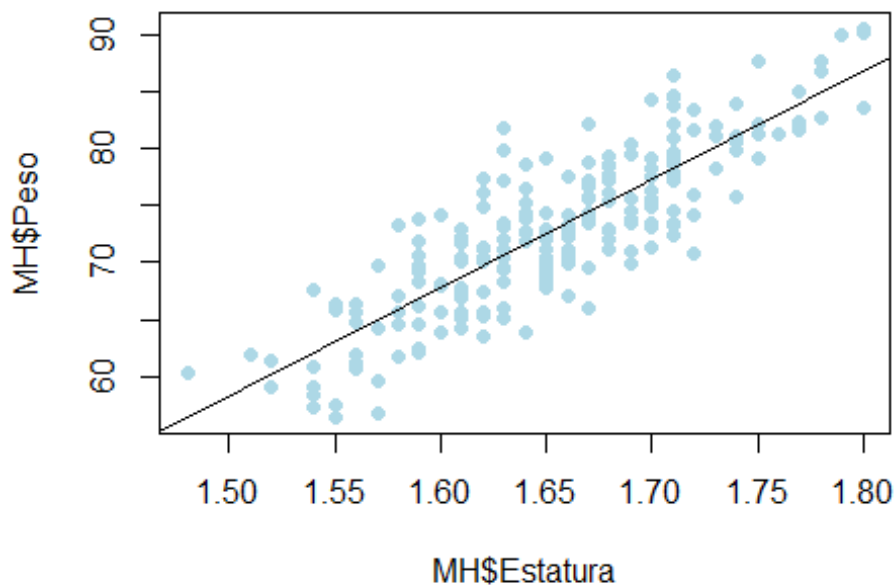
```
summary(modelo1M)

##
## 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
```

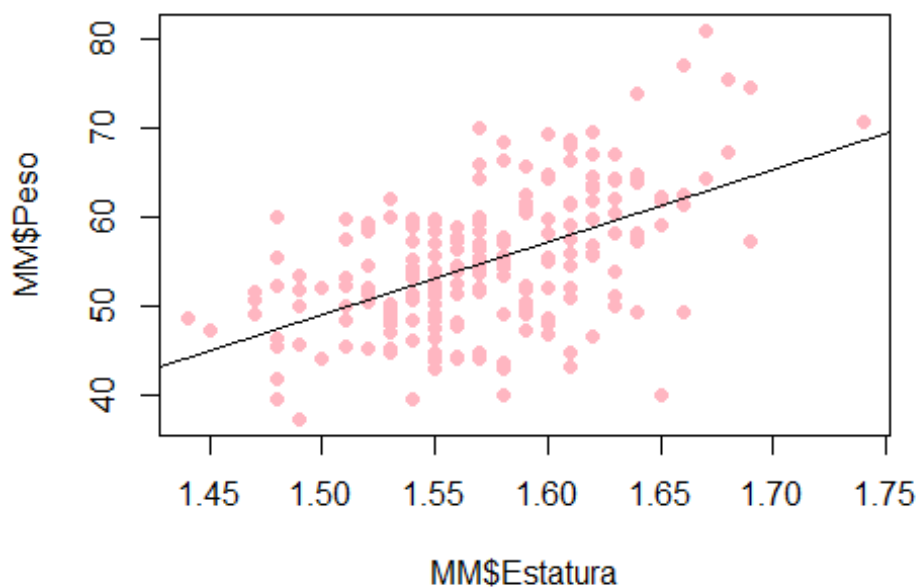
```
#Grafico de Hombres
```

```
plot(MH$Estatura,MH$Peso, col = "lightblue", pch = 19)  
abline(modelo1H)
```



```
# Grafico de Mujeres
```

```
plot(MM$Estatura, MM$Peso,col = "lightpink", pch = 19)  
abline(modelo1M)
```



```
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

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

cat("Variacion explicada (r^2): ", summary(Modelo2)$adj.r.squared)

## Variacion explicada (r^2):  0.7826698

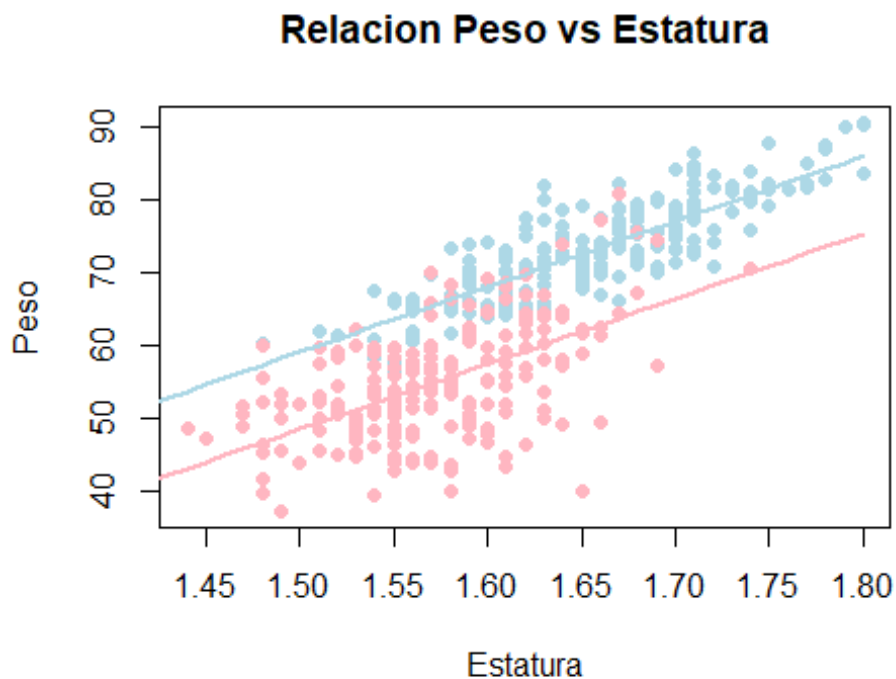
b0 = Modelo2$coefficients[1]
b1 = Modelo2$coefficients[2]
b2 = Modelo2$coefficients[3]

YM = function(x){b0 + b2 + b1 * x}
YH = function(x){b0 + b1 * x}

colores = c("lightblue", "lightpink")
plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =
"Peso", xlab = "Estatura", main = "Relacion Peso vs Estatura")

x = seq(1.40, 1.80, 0.01)
predH = YH(x)
predM = YM(x)

lines(x, predM, col = "lightpink", lwd = 2)
lines(x, predH, col = "lightblue", lwd = 2)
```



La informacion que proporciona  $B_0$  es que los hombres pesan más que las mujeres en cualquier peso, ya que inicia desde un punto más arriba del eje de las y, por aproximadamente de 10 kilos. Mientras que el  $B_1$  representa que los pesos aumentan conforme la altura aumenta. Por esto mismo la  $B_1$  es positiva.

### Continuacion de la tarea

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

anova(Modelo3)

## Analysis of Variance Table
##
## Response: Peso
##              Df Sum Sq Mean Sq    F value    Pr(>F)
## Estatura      1  37731   37731 1306.5938 <2e-16 ***
## Sexo          1   8097    8097  280.3892 <2e-16 ***
## Estatura:Sexo  1     61     61    2.1085  0.1472
## Residuals    436 12590     29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("Variacion explicada (r^2): ", summary(Modelo3)$adj.r.squared)

## Variacion explicada (r^2):  0.7832197

Modelo3

##
## Call:
```

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

```
##
```

```
## Coefficients:
```

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

```
##      -83.68         94.66         11.12         -13.51
```

Formula mujeres:  $-83.68 + 11.12409 + (-13.51 * x) + (94.66 * x)$  Formula hombres:  $-83.68 + (11.12 * x)$

```
b0 = Modelo3$coefficients[1]
```

```
b1 = Modelo3$coefficients[2]
```

```
b2 = Modelo3$coefficients[3]
```

```
b3 = Modelo3$coefficients[4]
```

```
YM_inter = function(x){b0 + b2 + b3 * x + b1 * x}
```

```
YH_inter = function(x){b0 + b1 * x}
```

```
plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =  
"Peso", xlab = "Estatura", main = "Relacion Peso vs Estatura con  
Interacción")
```

```
predH_inter = YH_inter(x)
```

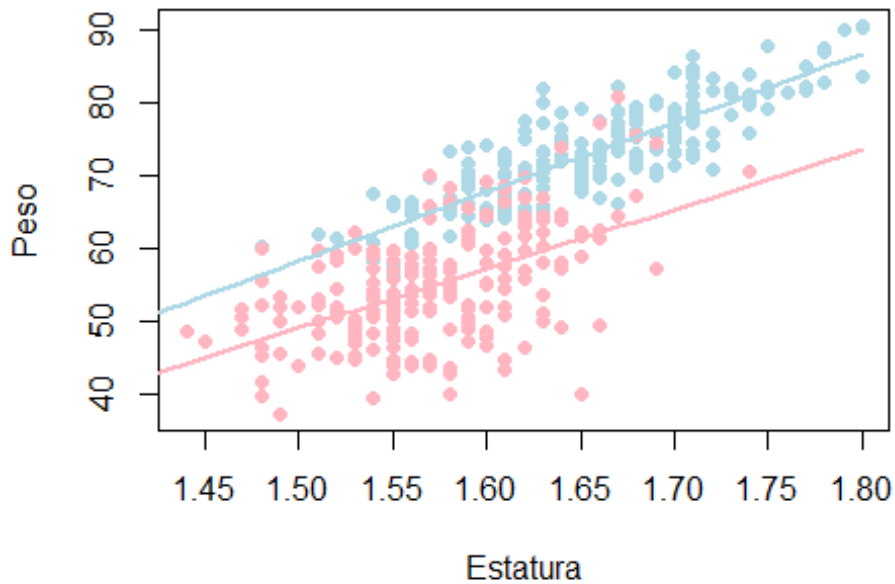
```
predM_inter = YM_inter(x)
```

```
lines(x, predM_inter, col = "lightpink", lwd = 2)
```

```
lines(x, predH_inter, col = "lightblue", lwd = 2)
```



## Relacion Peso vs Estatura con Interacción



Lo que representa  $B_0$  dentro del problema es el valor promedio de una persona cuando mide 0 m. Lo que representa  $B_1$  es el incremento en peso por unidad adicional de estatura en las personas. Lo que representa  $B_2$  es la diferencia entre mujeres y hombres cuando la altura de estos es 0. Lo que representa  $B_3$  es la diferencia en la relación entre estatura y peso entre hombres y mujeres.

Uno podría pensar que este modelo es mejor que el anterior ya que cuenta con una varianza  $r^2$  mayor, sin embargo, viendo el  $F$  de la intersección entre sexo y estatura es muy bajo, por lo que podemos entender que no hay una fuerte relación entre estas dos variables, por lo que no es necesario hacer esta operación, y nos podríamos quedar con el modelo anterior puesto que la varianza del modelo anterior es menor por una décima de diferencia.

Por lo anterior es que considero que el Modelo 2 es el mejor, pues ofrece resultados prácticamente iguales, mientras que las operaciones para obtenerlo son menos complejas.