Actividad 11

Oskar Arturo Gamboa Reyes

2024-09-03

Problema 1

Leer datos

```
M=read.csv("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)

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
```

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 hay una alta correlacion entre el peso y la altura de los hombres con 0.84, aunque menor tambien se observa bastente correlacion entre el peso y la altura de las mujeres con 0.52.

Medidas para analizar

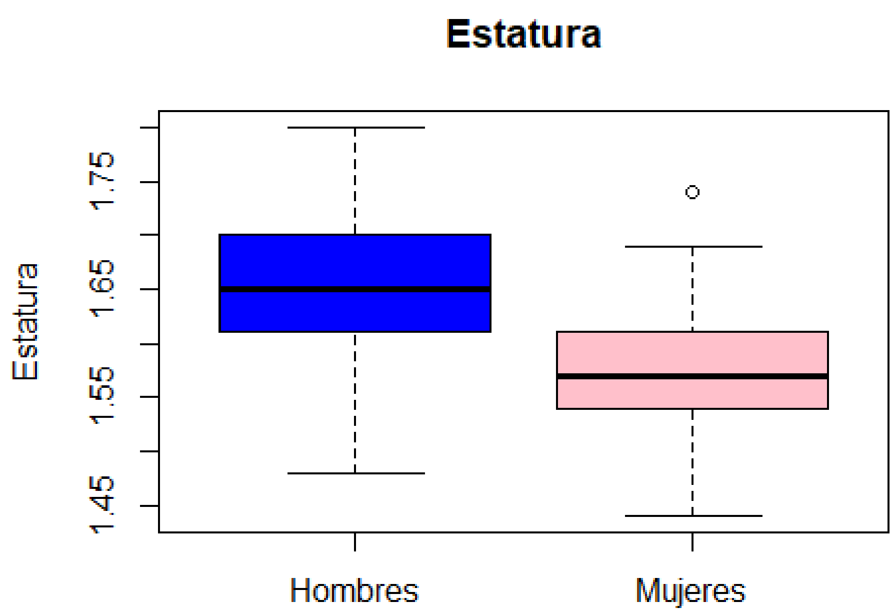
```
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)
```

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

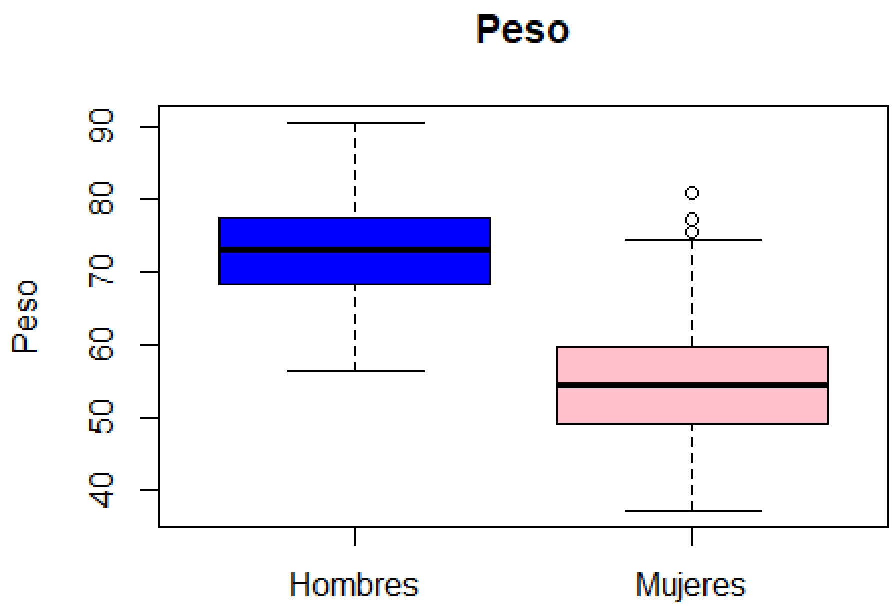
```
##             Minimo      Q1 Mediana     Media      Q3 Máximo   Desv Est
## H-Estatura    1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## H-Peso       56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408
## M-Estatura    1.44  1.5400   1.570  1.572955  1.6100   1.74 0.05036758
##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

Gráficas para datos

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




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



Rectas de mejor ajuste

Dos rectas

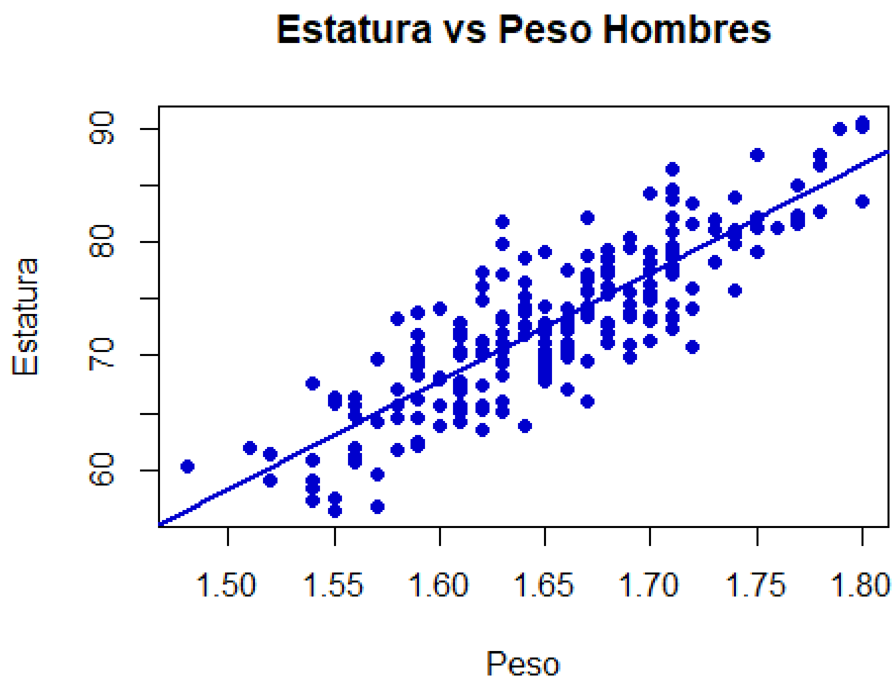
```
Modelo1H = lm(Peso~Estatura, MH)
Modelo1H

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

Modelo1M = lm(Peso~Estatura, MM)
Modelo1M

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

Hipótesis:

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

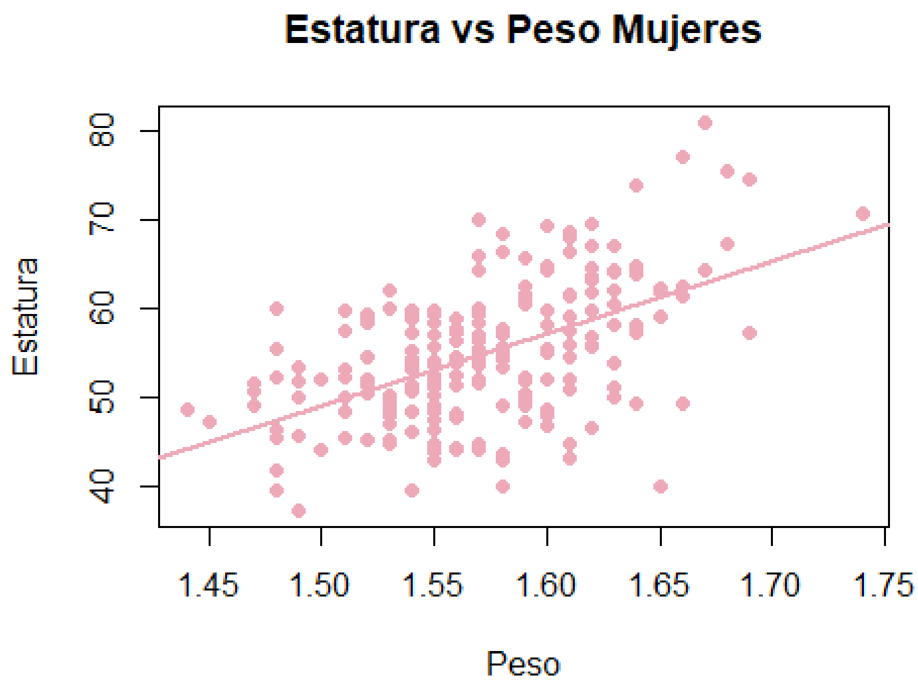
Hombres

```
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

plot(MH$Estatura,MH$Peso, col="blue3", main="Estatura vs Peso Hombres",
ylab="Estatura", xlab = "Peso",pch=19)
abline(lm(Peso~Estatura, MH), col="blue3",lwd=2)
```



Mujeres

```
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

plot(MM$Estatura,MM$Peso, col="pink2", main="Estatura vs Peso Mujeres",
ylab="Estatura", xlab = "Peso", pch=19)
abline(lm(Peso~Estatura, MM), col="pink2", lwd=2)
```



Un modelo

```r
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
```

A 0.05 si es significativo y los modelos quedarían:

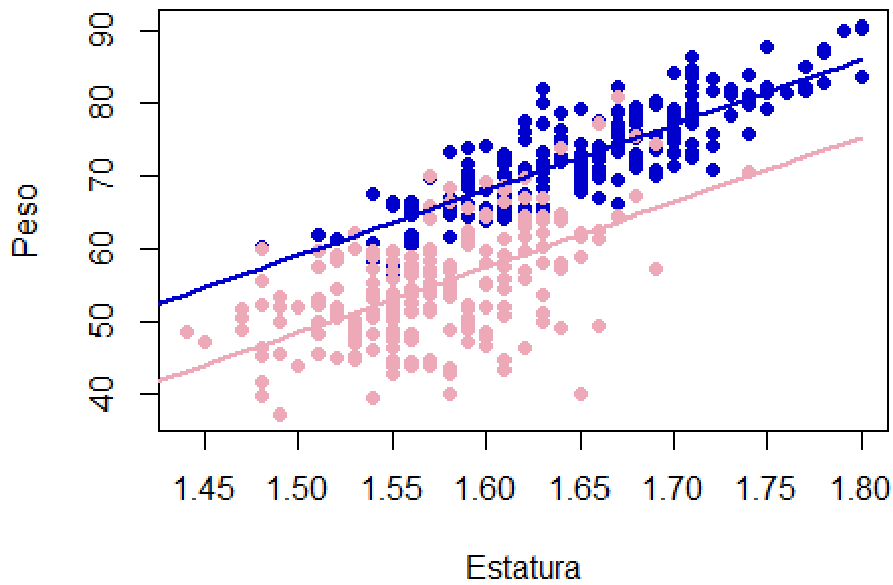
Hombre:

Peso = -74.7546 + 89.2604E

Mujeres:

Peso = -85.3191 + 89.2604E

```
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("blue3", "pink2")

plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch=19, xlab =
"Estatura", ylab = "Peso")

x= seq(1.40,1.80,0.01)
lines(x, Ym(x), col="pink2",lwd=2)
lines(x, Yh(x), col="blue3",lwd=2)
```



Modelo con interacción

Hipótesis:

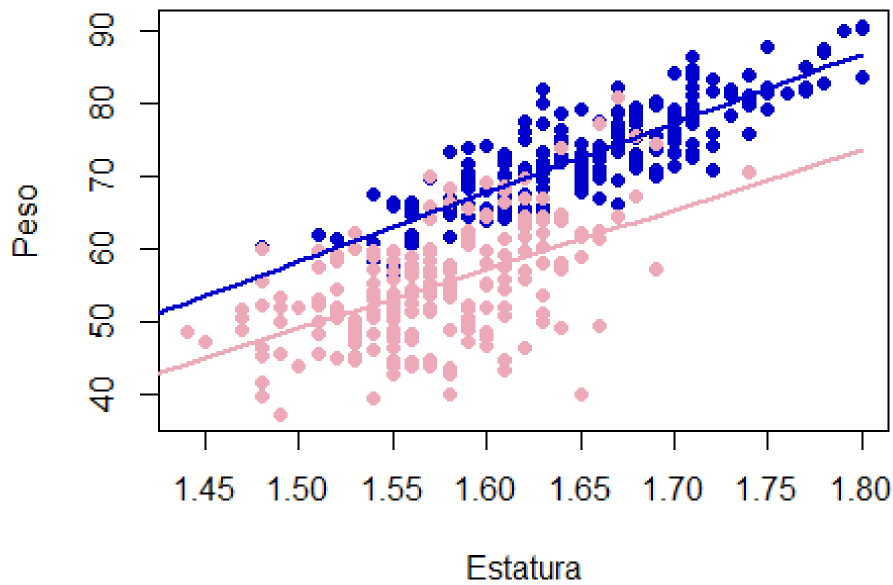
- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$

```
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 = −83.6845 + 94.6602Estatura + 11.1241SexoM − 13.5111(Estatura * SexoM)

```
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

b0 = Modelo3$coefficients[1]
b1 = Modelo3$coefficients[2]
b2 = Modelo3$coefficients[3]
b3 = Modelo3$coefficients[4]

Yh = function(x){b0+b1*x}
Ym = function(x){b0+b1*x+b2+b3*x}

colores = c("blue3", "pink2")

plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch=19, xlab = 
"Estatura", ylab = "Peso")

x= seq(1.40,1.80,0.01)
lines(x, Ym(x), col="pink2",lwd=2)
lines(x, Yh(x), col="blue3",lwd=2)
```



Conclusión

¿Qué información proporciona β̂0 sobre la relación entre la estatura y el peso de hombres y mujeres? Interpreta y compara entre este modelo con los 3 modelos anteriores.

Esta beta nos describe la magnitud incial del modelo, en el contexto del problema no tiene mucho sentido ya que una altura de 0 es imposible, sin embargo si describe un valor inicial del peso que afecta a toda la población. Todos los modelos tienen una beta 0 similar, ya que lo que afectan estos modelos es principalmente la pendiente.

¿Cómo interpretas β̂i en la relación entre la estatura y el peso de hombres y mujeres? Interpreta y compara entre este modelo con los 3 modelos anteriores.

Estos describen la pendiente de nuestra función, podemos ver que hay una relación diferente entre el peso y estatura de los hombres y de las mujeres, en el modelo anteriore esto no se tomaba en cuenta, lo que hacía a nuestro modelo menos significante. Los primeros dos modelos se ajustan de la misma manera que el último, ya que fueron hechos independientemente.

Indica cuál(es) de los modelos probados para la relación entre peso y estatura entre hombres y mujeres consideras que es más apropiado y explica por qué.

Podemos notar que el modelo que toma en cuenta la interacción es un mejor modelo que describe mejor el comportamiento de los datos, ya que ahora si existe una variación entre las pendientes de las gráficas que se ajustan dependiendo a sus datos dependiendo de la variable sexo. Los dos primeros modelos explican bien el comportamiento pero estan independientes por loq ue no nos dejan apreciar las diferencias entre los seoxs. El último modelo es mejor.