Actividad 10

2024-08-30

La recta de mejor ajuste

```
M = read.csv("Estatura-peso_HyM.csv")
head(M)
```

```
##   Estatura  Peso Sexo
## 1     1.61 72.21    H
## 2     1.61 65.71    H
## 3     1.70 75.08    H
## 4     1.65 68.55    H
## 5     1.72 70.77    H
## 6     1.63 77.18    H
```

Matriz de correlacion

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

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
```

A partir de la matriz de correlación podemos notar que existe una alta correlación entre la estatura y peso de los hombres con 0.84, mientras que las muejres tienen una menor correlación con 0.52, sin embargo sigue siendo suficiente para indicar que existe alguna relación entre estas variables.

Obtener medidas

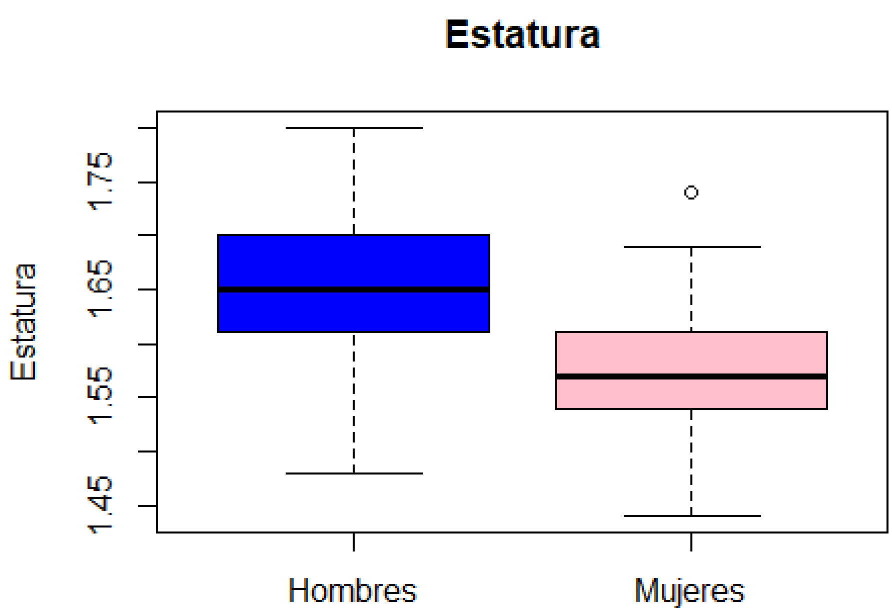
```
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)

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

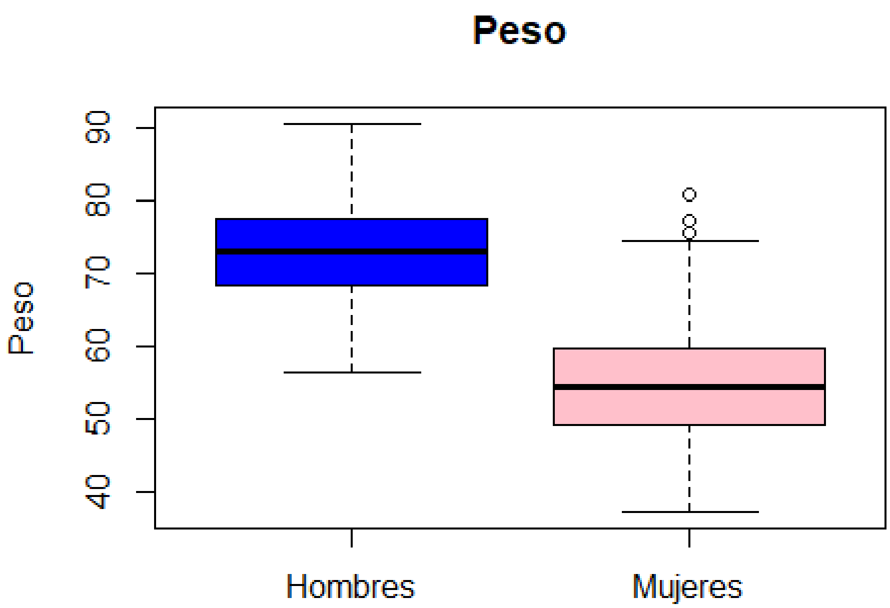
```
##               Minimo      Q1 Mediana      Media      Q3 Máximo   Desv Est
## H-Estatura    1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## H-Peso       56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408
## M-Estatura    1.44  1.5400   1.570  1.572955  1.6100   1.74 0.05036758
## 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("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")
```



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

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

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

Hipotesis:

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

```
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
```

```
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
```

```
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
```

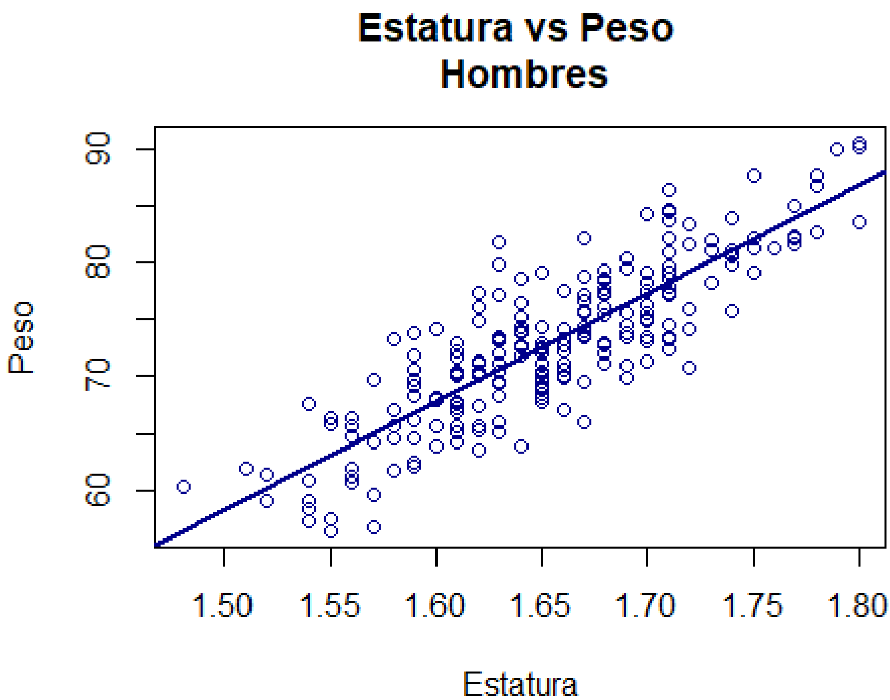
```r
summary(modelo2)

##
## 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
```

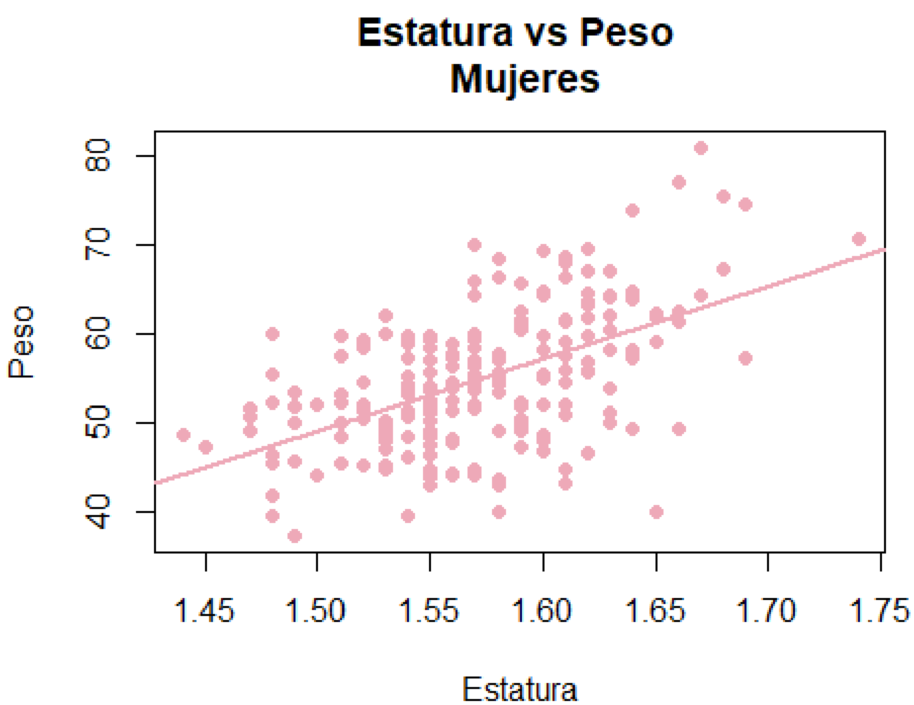
A 0.03 si es significativo.

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

```r
plot(MH$Estatura, MH$Peso, col="blue4", main = "Estatura vs Peso \n
Hombres", ylab="Peso", xlab="Estatura")
abline(modelo1H, col="blue4", lwd=2)
```




```r
plot(MM$Estatura, MM$Peso, col="pink2", pch=19,main = "Estatura vs Peso
\n Mujeres", ylab="Peso", xlab="Estatura")
abline(modelo1M, col="pink2", lwd=2)
```

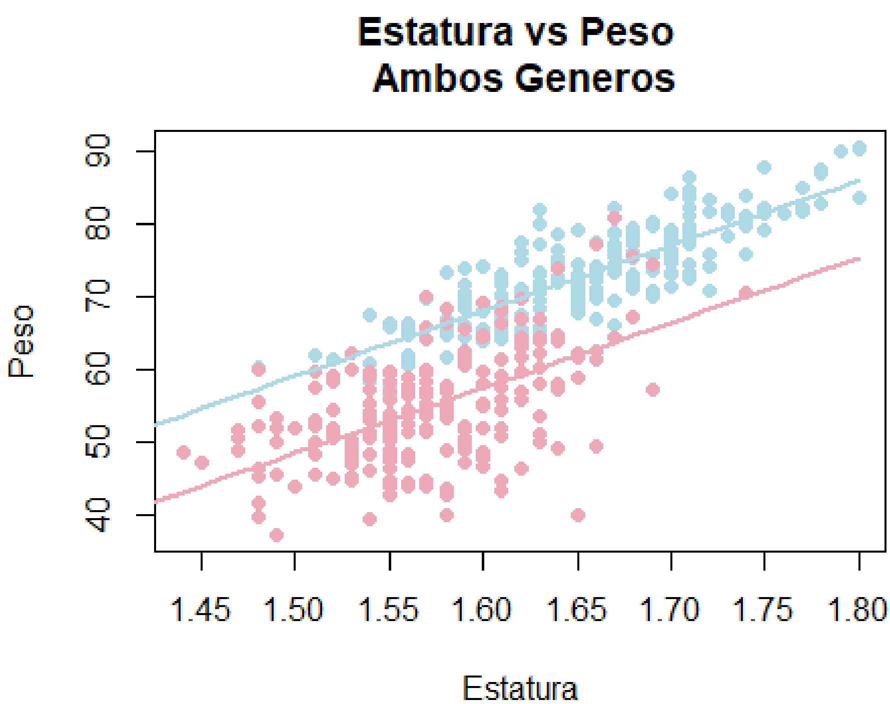


```r
b0 = modelo2$coefficients[1]
b1 = modelo2$coefficients[2]
b2 = modelo2$coefficients[3]

ym = function(x){b0+b2+b1*x}
yh = function(x){b0+b1*x}

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

plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19, main =
"Estatura vs Peso \n Ambos Generos", ylab="Peso", xlab="Estatura")
x = seq(1.40, 1.80, 0.01)
lines(x, ym(x), col="pink2", lwd=2)
lines(x, yh(x), col="lightblue", lwd=2)
```



Interpretación de resultados

Podemos ver que los dos modelos nos dan una significancia adecuada para este problema, sin embargo, en mi opinión es mejor usar un modelo separado para hombres y para mujeres. Se observa que existe una mayor variación en los datos de las mujeres, que al usar un modelo compartido no logra adaptarse completamente a las medidas. Esto ocurre ya que la relación entre el peso y la estatura se mantiene igual (la pendiente de la recta) y lo unico que puede cambiar es la intersección en "y" que determinan la magnitud de estas medidas y no la manera en que se relacionan.

Explicación de beta 0 Beta 0 nos explica la relación entre la estatura y el peso, basicamente es el número que multiplica a la estatura para obtener el peso esperado.

Explicación de beta 1 Beta 1 explica la magnitud de las medidas o el peso inicial, por ejemplo, como las mujeres suelen pesar y medir menos, el modelo que explica el comportamiento de las mujeres va a tener una beta 1 menor.