RegresionLineal

Cleber Perez

2024-08-30

Leer archivo

```
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
```

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

Larecta de mejor ajuste

Analisis descriptivo

1. Obtén la matriz de correlación de los datos que se te proporcionan. Interpreta.

```
cor(M1)
```

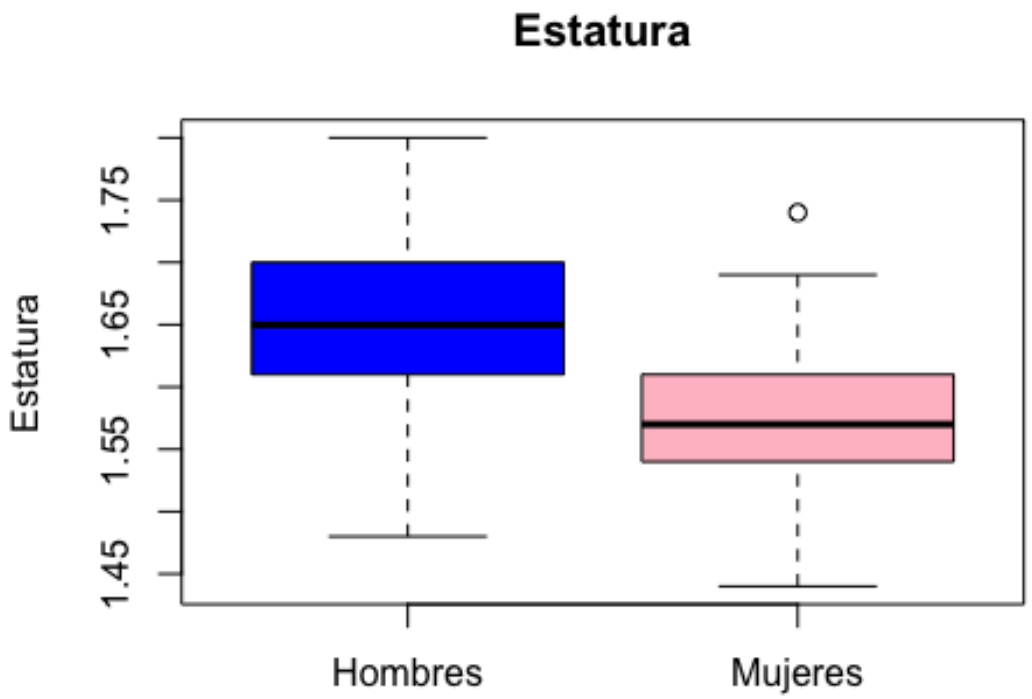
```
##              MH.Estatura     MH.Peso  MM.Estatura    MM.Peso
## MH.Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872
## MH.Peso     0.8468347920 1.000000000 0.0035132246 0.02154907
## MM.Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621
## MM.Peso     0.0472487231 0.021549075 0.5244962115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

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

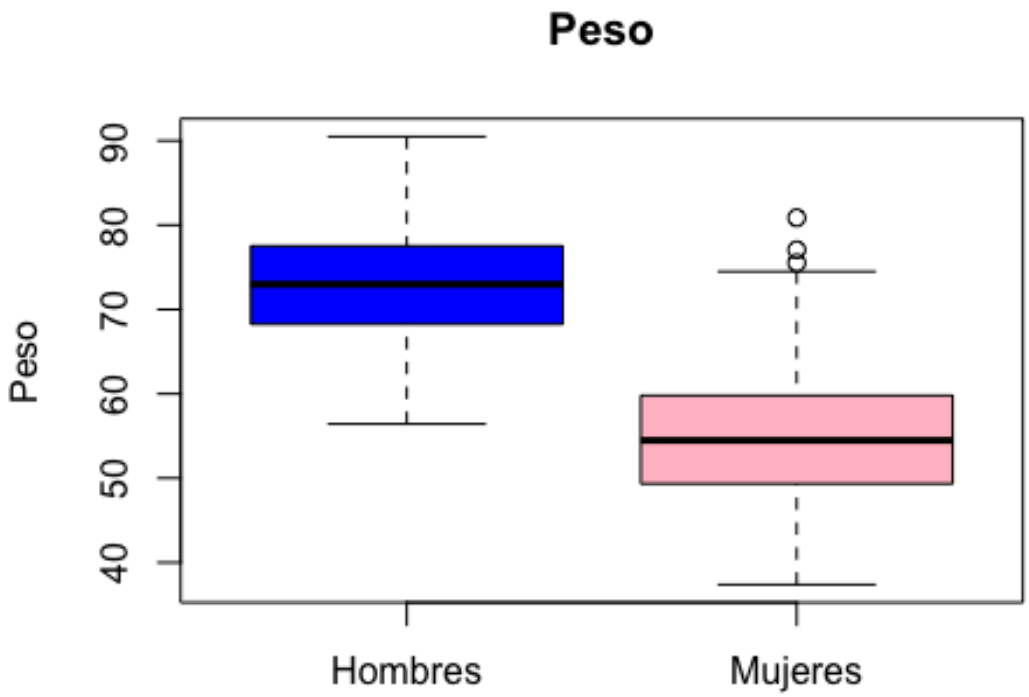
```
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)

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

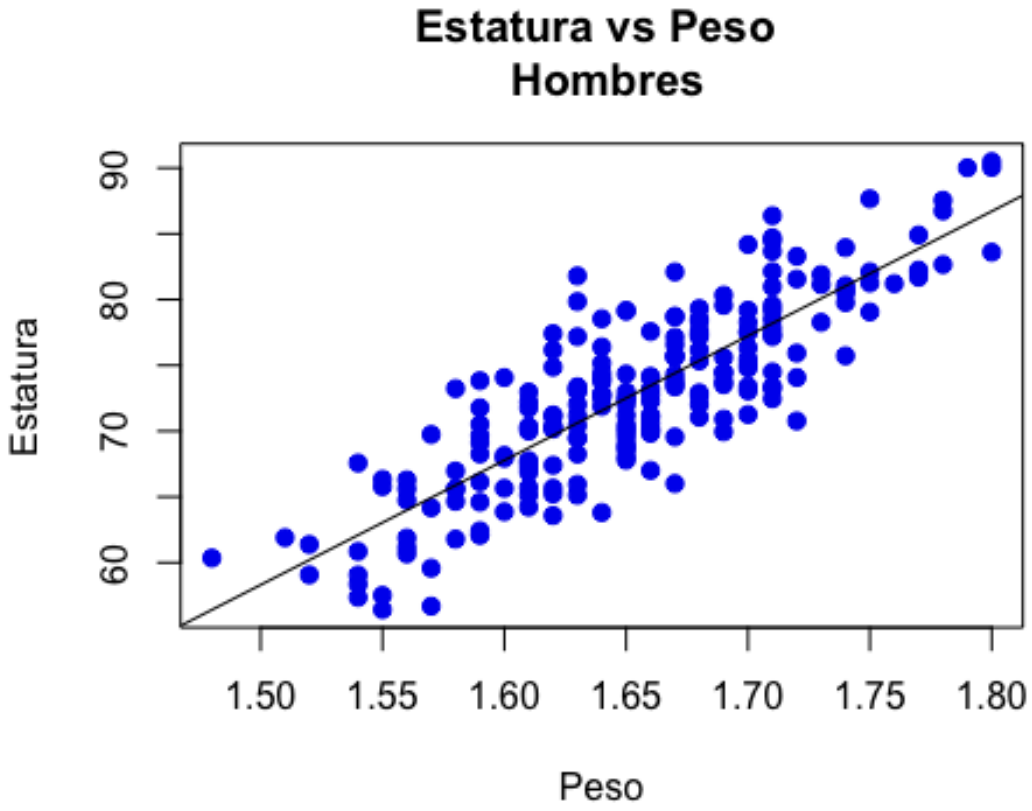
Encuentra la ecuación de regresión de mejor ajuste:

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

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

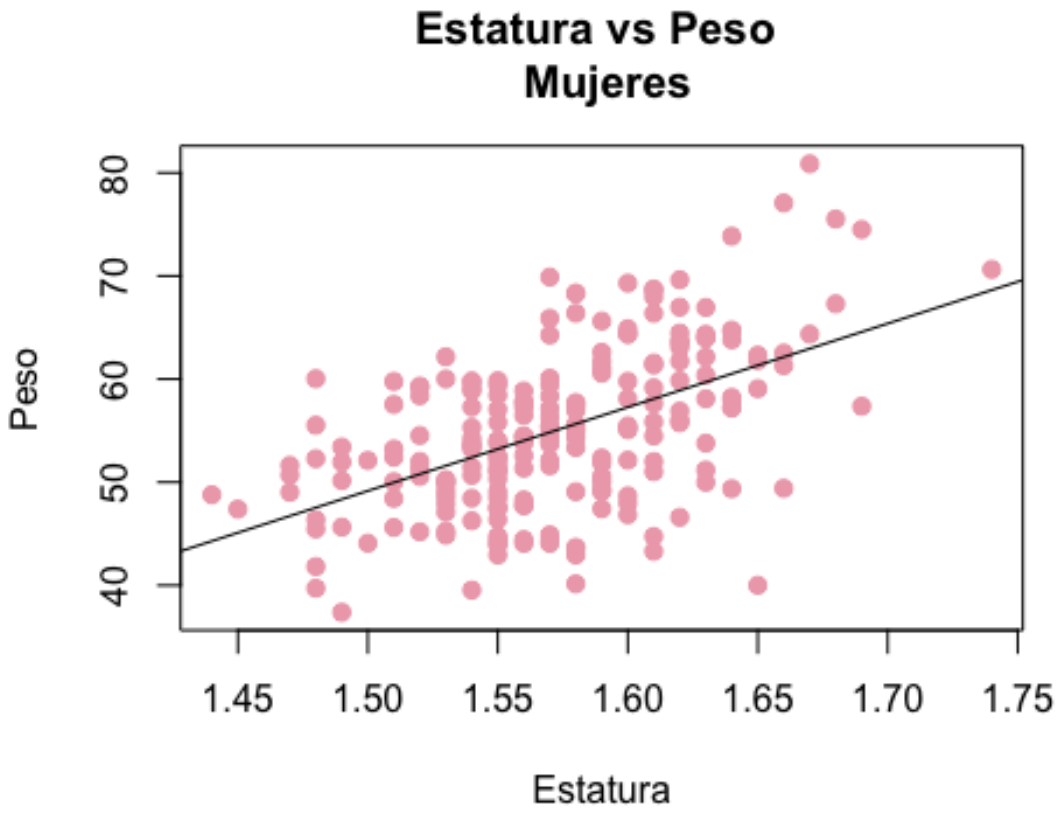
```
summary(Modelo1H)

##
## 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

plot(MH$Estatura, MH$Peso, col="blue2", main="Estatura vs Peso \n Hombres",
ylab="Estatura", xlab="Peso", pch=19)
abline(Modelo1H)
```




```
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
```

```r
plot(MM$Estatura, MM$Peso, col="pink2", main="Estatura vs Peso \n Mujeres",
     ylab="Peso", xlab="Estatura", pch=19)
abline(Modelo1M)
```



El 71% de la variabilidad esta explicada por el modelo y lo demas esta en los errores, el peso es significativo

Realiza la regresión entre las variables involucradas

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
```

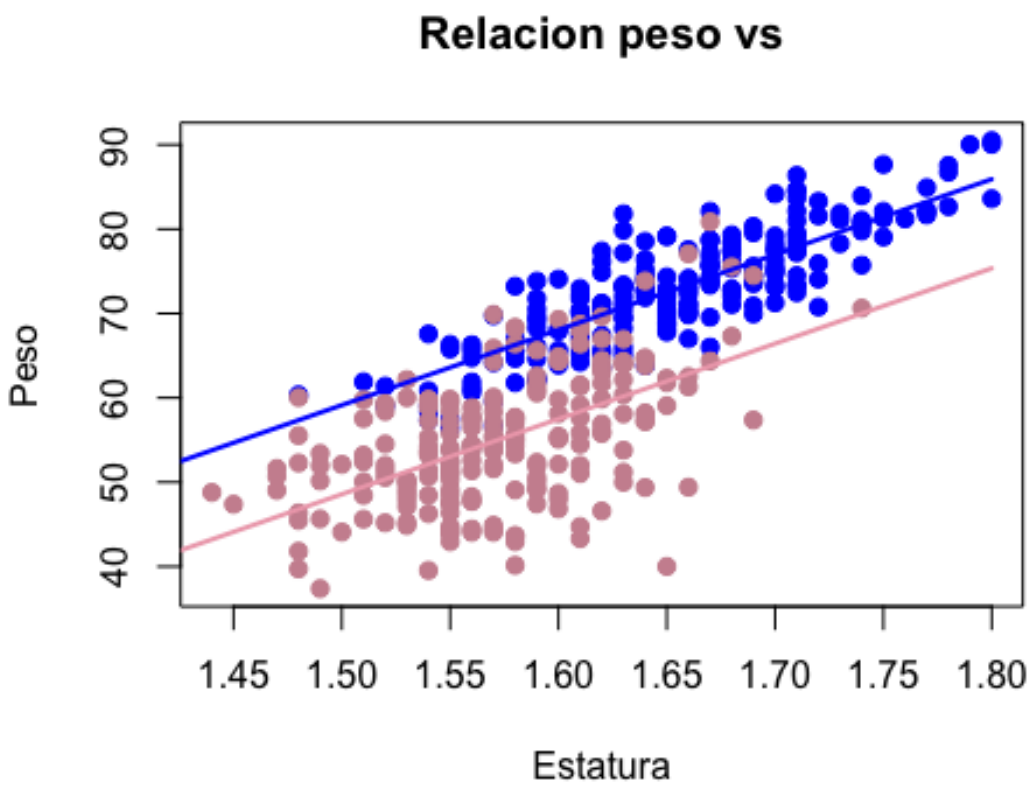
```r
summary(Modelo2)
```

```
##
## 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
```

El sexo no es significativo A 0.05 si es significativo y los modelos quedarian:

Hombres:

```
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("blue", "pink3")
plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19,
xlab="Estatura", ylab="Peso", main="Relacion peso vs")
x = seq(1.40, 1.80, 0.01)
lines(x, Ym(x), col="pink2", lwd=2)
lines(x, Yh(x), col="blue", lwd=2)
```

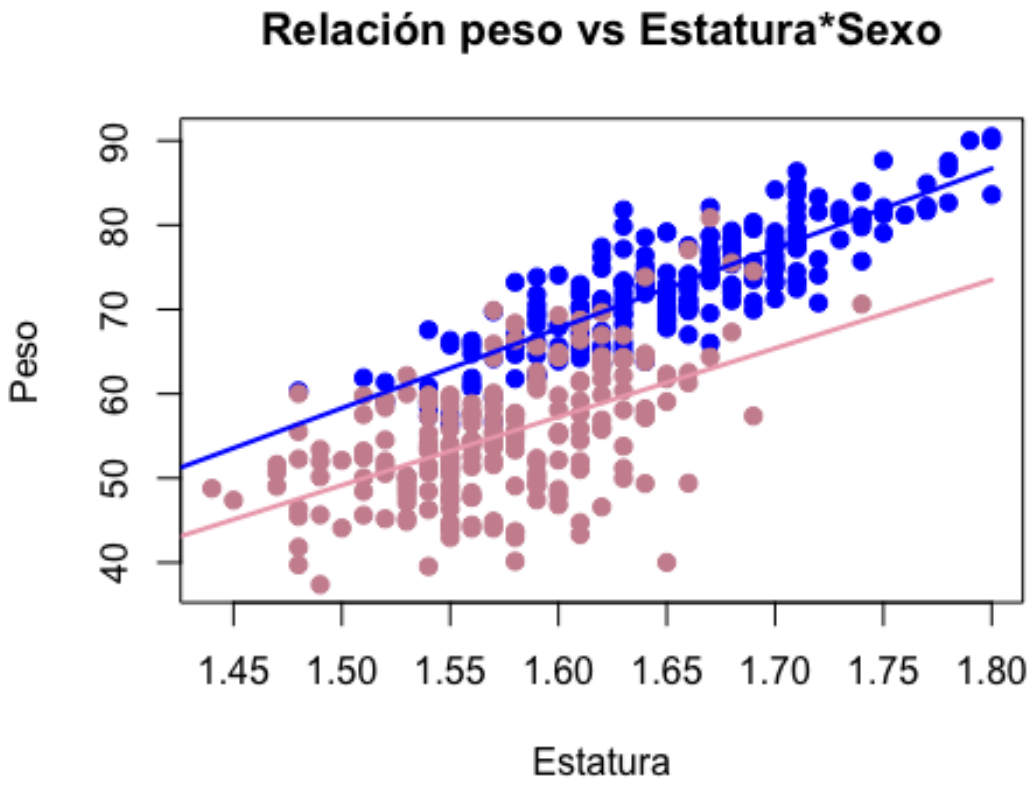


Nuevo Modelo Estatura-Sexo

```
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

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
```

```r
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("blue", "pink3")

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

x = seq(1.40, 1.80, 0.01)

lines(x, Ym(x), col="pink2", lwd=2)
lines(x, Yh(x), col="blue", lwd=2)
```


Relación peso vs Estatura*Sexo
Peso
40
50
60
70
80
90
1.45
1.50
1.55
1.60
1.65
1.70
1.75
1.80
Estatura