

## 10. Regresión Lineal

Eliezer Cavazos

2024-08-30

Analiza la base de datos de estatura y peso de los hombres y mujeres en México y obten el mejor modelo de regresión para esos datos.

```
oPesoEstatura =  
read.csv("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\Estatura-peso_HyM.csv"  
) #Leer la base de datos  
  
oMujeres = subset(oPesoEstatura,oPesoEstatura$Sexo=="M")  
oHombres = subset(oPesoEstatura,oPesoEstatura$Sexo=="H")  
  
M1=data.frame(oHombres$Estatura,oHombres$Peso,oMujeres$Estatura,oMujeres$Peso  
)
```

### La recta de mejor ajuste (Primera entrega)

#### Analisis Descriptivo

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

```
cor(M1)  
  
##                oHombres.Estatura oHombres.Peso oMujeres.Estatura  
## oHombres.Estatura      1.0000000000      0.846834792      0.0005540612  
## oHombres.Peso          0.8468347920      1.000000000      0.0035132246  
## oMujeres.Estatura      0.0005540612      0.003513225      1.0000000000  
## oMujeres.Peso          0.0472487231      0.021549075      0.5244962115  
##                oMujeres.Peso  
## oHombres.Estatura      0.04724872  
## oHombres.Peso          0.02154907  
## oMujeres.Estatura      0.52449621  
## oMujeres.Peso          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

## La recta de mejor ajuste

Encuentra la ecuación de regresión de mejor ajuste:

Hombres:

```
ModeloHombres = lm(oHombres$Peso ~ oHombres$Estatura, oHombres)
```

ModeloHombres

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

```
summary(ModeloHombres)
```

```
##
## Call:
## lm(formula = oHombres$Peso ~ oHombres$Estatura, data = oHombres)
##
## 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 ***
## oHombres$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
```

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

Mujeres:

```
ModeloMujeres = lm(oMujeres$Peso ~ oMujeres$Estatura, oMujeres)
ModeloMujeres

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

summary(ModeloMujeres)

##
## Call:
## lm(formula = oMujeres$Peso ~ oMujeres$Estatura, data = oMujeres)
##
## 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 ***
## oMujeres$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
```

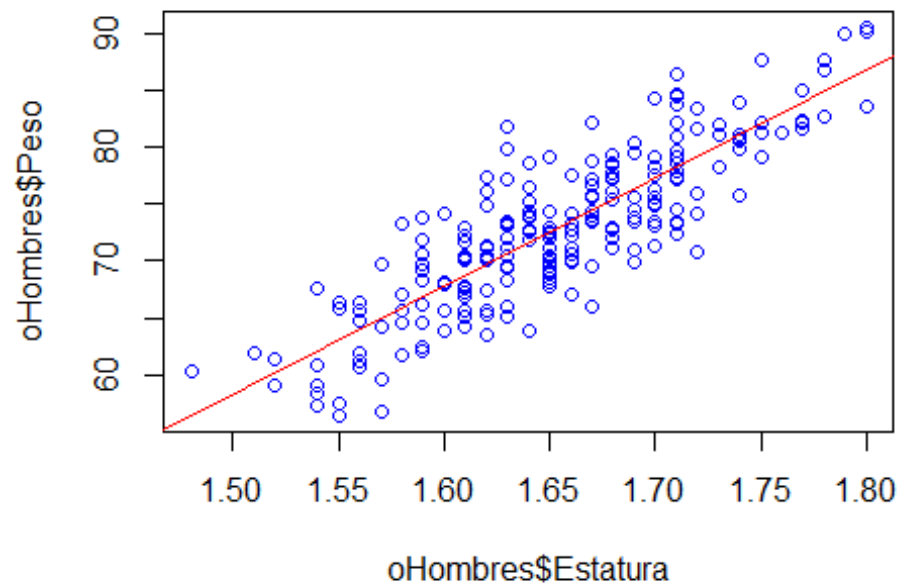
### Realiza la regresión entre las variables involucradas

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

Hombres:

```
plot(oHombres$Estatura, oHombres$Peso, col="blue", main="Estatus vs Peso
Hombres")
abline(ModeloHombres, col="red")
```

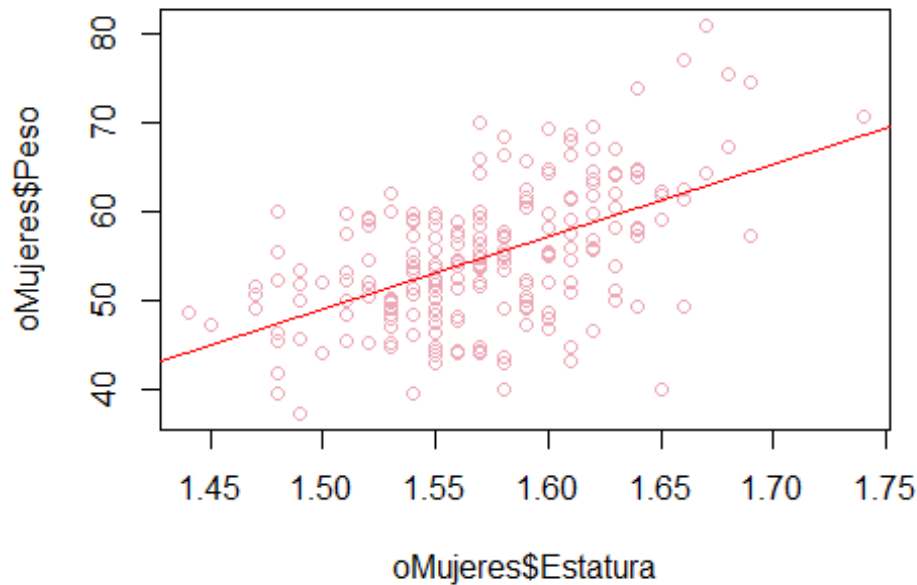
## Estatus vs Peso Hombres



Mujeres:

```
plot(oMujeres$Estatura, oMujeres$Peso, col="pink2",main="Estatus vs Peso  
Mujeres")  
abline(ModeloMujeres, col="red")
```

## Estatus vs Peso Mujeres



### Verificacion del modelo:

Verifica la significancia del modelo con un alfa de 0.03.

Verifica la significancia de  $\beta_i$  con un alfa de 0.03.

Verifica el porcentaje de variación explicada por el modelo

Hombres:

```
Modelo2 = lm(Peso~Estatura+Sexo, oPesoEstatura)
```

```
summary(Modelo2)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = oPesoEstatura)
##
## 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 quedarian:

Mujeres:  $\text{Estatura} = -72.56 + 81.15E$   $\text{Estatura} = -74.7546 + 89.2604E + -10.5645\text{SexoM}$   
 $\text{Estatura} = -85.3191 + 0.0052296E$

Hombres:  $\text{Estatura} = -72.56 + 81.15E$   $\text{Estatura} = -74.7546 + 89.2604E + -10.5645\text{SexoM}$   
 $\text{Estatura} = -74.7546 + 89.2604E$

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

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

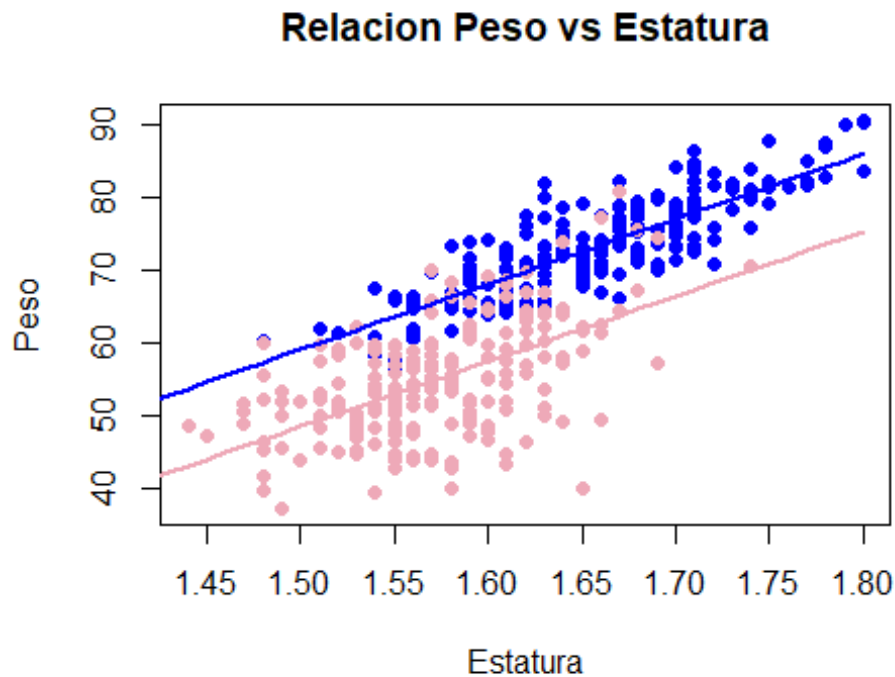
Ym = function(x){b0+b2+b1*x} #1.2848 + 0.0052296P
Yh = function(x){b0+b1*x} # 1.2727097 + 0.0052296P
colores = c("blue", "pink2")
plot(oPesoEstatura$Estatura,oPesoEstatura$Peso, data=oPesoEstatura,
col=colores[factor(oPesoEstatura$Sexo)],pch=19, ylab="Peso", xlab="Estatura",
main="Relacion Peso vs Estatura")

## Warning in plot.window(...): "data" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
a
## graphical parameter
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
a
## graphical parameter

## Warning in box(...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

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



Interpreta en el contexto del problema cada uno de los análisis que hiciste.

En los análisis se identificó la relación del Peso por Estatura tanto en Hombres como Mujeres para poder generar un modelo de regresión lineal y así poder realizar predicciones para identificar Peso por Estatura. ### Interpreta en el contexto del problema: ### ¿Qué información proporciona  $\beta_0$  sobre la relación entre la estatura y el peso de hombres y mujeres?  $B_0$  nos da el valor de Peso cuando la Estatura es 0 y el Sexo es una mujer, aunque esto no tiene mucho sentido cuando la estatura es 0 ya que el modelo está entrenado sobre estatura mayores a 1.45, con esto dicho no significa que esté mal, solo que al momento de hacer una predicción dará un mejor resultado mientras la estatura esté dentro de un valor “real”/aproximado. ### ¿Cómo interpretas  $\beta_1$  en la relación entre la estatura y el peso de hombres y mujeres?  $B_1$  va a ser el valor que se va a estar sumando (restando en el caso de la fórmula que tenemos) al valor de  $B_0$  para dar el peso indicado.  $B_1$  es también multiplicado por la Estatura, ya que siguiendo los datos mientras más Estatura en promedio el peso también debe aumentar.