Regresión Lineal

Erika Martínez Meneses

2024-08-30

Estatura y Peso

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.

```
file.choose()

## [1] "C:\\Users\\erika\\Documents\\Agos-Dic2024\\Estadística\\Estatura-peso_HyM.csv"

library(readr)
M <- read_csv("C:\\Users\\erika\\Documents\\Agos-Dic2024\\Estadística\\Estatura-peso_HyM.csv")

## Rows: 440 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Sexo
## dbl (2): Estatura, Peso
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

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

La recta de mejor ajuste

Análisis descriptivo

1. Matriz de 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
```

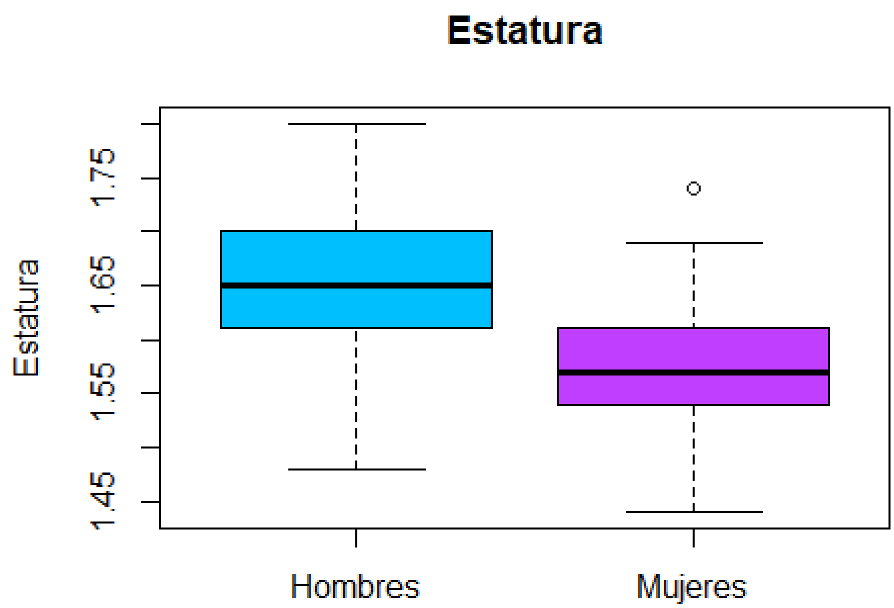
Se observa que existe correlación entre la estatura y peso cuando los datos son del mismo género (hombre o mujer) sin embargo la correlación entre la estatura y peso en hombres es más fuerte (0.8468) que la correlación existente entre la estatura y peso de mujeres (0.5244). Esto sugiere que la variabilidad en el peso de los hombres se explica mejor por la estatura en comparación con las mujeres. Las correlaciones cruzadas (estatura de hombres con peso de mujeres, y viceversa) son muy bajas, lo que indica que no hay relación entre las variables de distintos géneros.

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

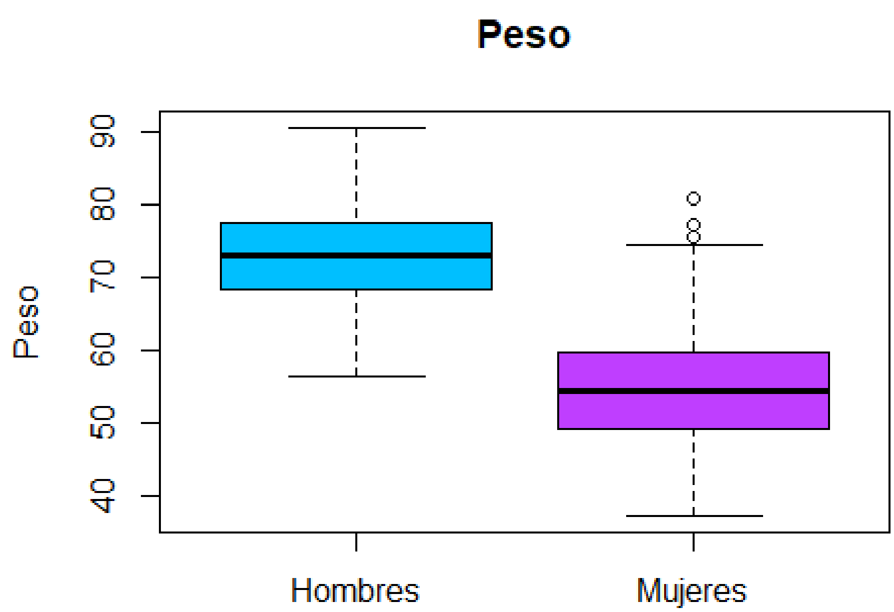
```r
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)

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("#00BFFF","#BF3EFF"), names=c("Hombres", "Mujeres"),
main="Estatura")
```




```
boxplot(M$Peso~M$Sexo, ylab="Peso",xlab="", names=c("Hombres",
"Mujeres"), col=c("#00BFFF","#BF3EFF"), main="Peso")
```



La media y la desviación estándar de la estatura y el peso para ambos géneros indican que, en promedio, los hombres son más altos y pesados que las mujeres.

La recta del mejor ajuste

Encuentra la ecuación de regresión de mejor ajuste

Dos rectas

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

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

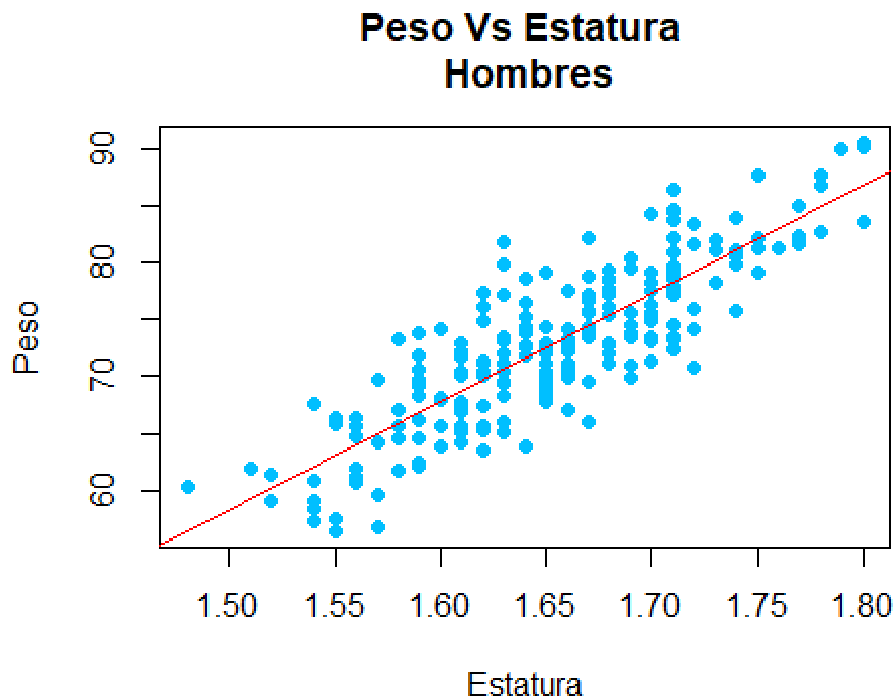
Modelo 1 Hombres * Estatura = -83.68 + 94.66

Hipótesis * $H_0: \beta_1 = 0$ * $H_1: \beta_1 \neq 0$

```
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
```

Grafica

```
plot(MH$Estatura, MH$Peso, col = "#00BFFF", main = "Peso Vs Estatura \n
Hombres", ylab = "Peso", xlab = "Estatura", pch = 19)
abline(Modelo1H, col = "red", lwd = 1.6)
```



Rechazo B_0, sí es significativa

- El 71% de la variabilidad esta siendo explicada
- Nos queda una t grande lo que significa que en la distribución se encuentra 46.72 y 23.51 veces lejos de la β hipotética.
- Podemos observar en los cuartiles que hay simetría.

```
Modelo1M = lm(Peso~Estatura, data= MM)
Modelo1M

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

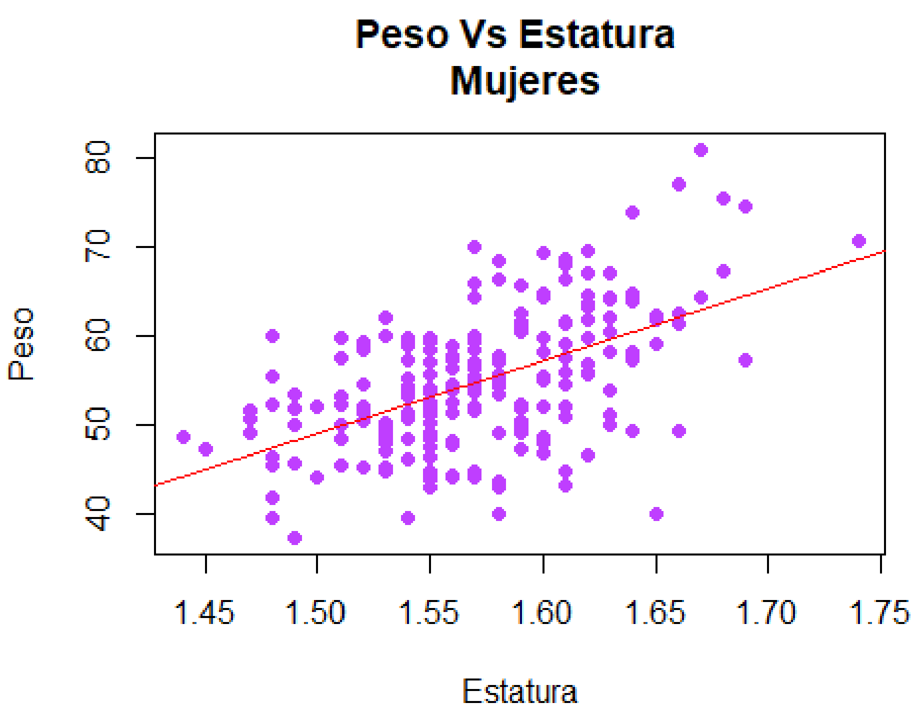
Modelo 1 Mujeres * Estatura = -72.56 + 81.15

```
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
```

Grafica

```
plot(MM$Estatura, MM$Peso, col = "#BF3EFF", main = "Peso Vs Estatura \n
Mujeres", ylab = "Peso", xlab = "Estatura", pch = 19)
abline(Modelo1M, col = "red", lwd = 1.6)
```



Rechazo B_0, sí es significativa

- El 27.51% de la variabilidad esta siendo explicada
- El R^2 nos sale bajo
- Nos queda una t grande lo que significa que en la distribución se encuentra 66.859 y 9.096 veces lejos de la β hipotética.

Realiza la regresión entre las variables involucradas

Un modelo

```
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
```

Va a ser 1 para mujer y 0 para hombre

```
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) Significancia del modelo con alfa de 0.03

```
if (summary(Modelo2)$fstatistic[1] > qf(1-0.03, df1 =
summary(Modelo2)$fstatistic[2], df2 = summary(Modelo2)$fstatistic[3])) {
  print("El modelo es significativo con alfa de 0.03.")
} else {
  print("El modelo no es significativo con alfa de 0.03.")
}
```

```
## [1] "El modelo es significativo con alfa de 0.03."
```

b) Significancia de $\widehat{B_i}$ con un alfa de 0.03.

```
coef(summary(Modelo2))[, 4] < 0.03 # Devuelve TRUE si los coeficientes
son significativos
```

```
## (Intercept)    Estatura       SexoM
##        TRUE         TRUE        TRUE
```

Los parámetros del modelo (Estatura, Sexo) son significativos tanto para $\alpha = 0.05$ como para $\alpha = 0.03$ ya que sus valor p (Pr(>|t|)) son pequeños. Asimismo podemos observar que el p value del modelo es bajo lo que significa que el modelo es significativo.

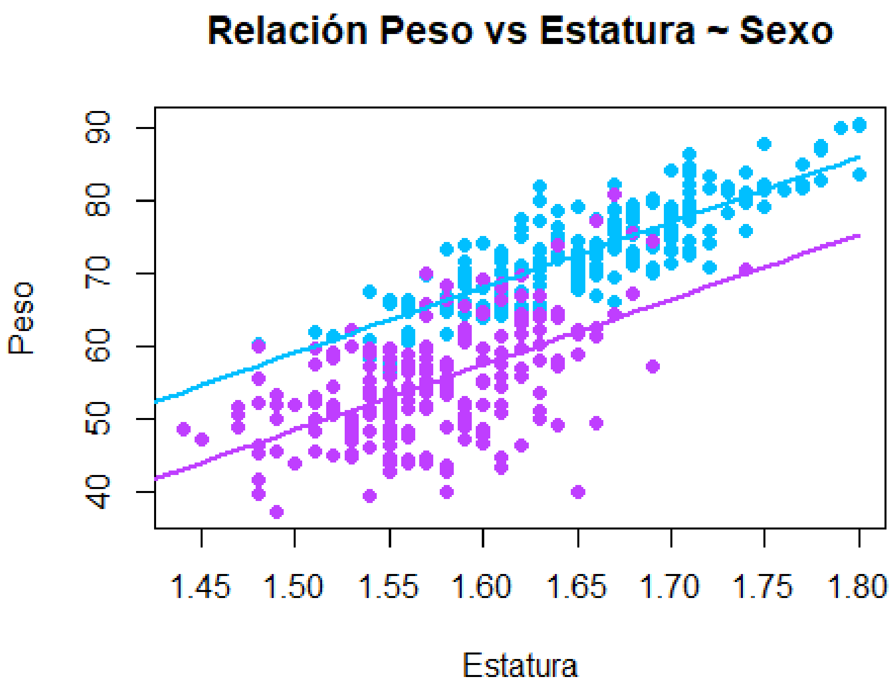
y los modelos quedarían:

Mujeres: * Estatura = -74.75 + 89.26 P + -10.56 SexoM = -85.31 + 0.0052296 P

Hombres: * Estatura = -74.75 + 89.26 P

c) Verificación del porcentaje de variación explicada por el modelo (R^2)

```r
paste("El modelo explica el", round(summary(Modelo2)$r.squared * 100, 2),
"% de la variabilidad del peso.")

## [1] "El modelo explica el 78.37 % de la variabilidad del peso."

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( "#00BFFF","#BF3EFF")

plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =
"Peso", xlab = "Estatura", main = "Relación Peso vs Estatura ~ Sexo")
x = seq(1.40, 1.80, 0.01)
lines(x, Ym(x), col = "#BF3EFF", lwd = 2)
lines(x, Yh(x), col = "#00BFFF", lwd = 2)
```



Aparentemente centimetro más en mujeres debería tener el mismo efecto que un centímetro más en hombres.

Hombres: El modelo de regresión lineal indica que por cada centímetro adicional en la estatura, el peso aumenta en promedio 94.66 kg.

Mujeres: Similarmente, el peso aumenta en 81.15 kg por cada centímetro adicional de estatura, pero el R^2 es mucho más bajo, indicando una relación menos fuerte.

$\widehat{B_0}$ representa el peso esperado cuando la estatura es cero, se usa como punto de referencia en la recta.

- Para Hombres: En el modelo de regresión lineal para hombres, $\widehat{B_0}$ es -83.68. Esto significa que, en teoría, si un hombre tuviera una estatura de 0 cm, el peso predicho sería -83.68 kg. Aunque esta situación es imposible y no tiene un sentido práctico (nadie puede tener 0 cm de estatura), el intercepto aún es importante para definir la ecuación de la recta de regresión. En la práctica, el valor del intercepto no suele ser interpretado directamente cuando no tiene sentido contextual, sino que se usa para calcular predicciones de peso a partir de estaturas razonables.
- Para Mujeres: Similarmente, para mujeres, $\widehat{B_0}$ es -72.56. Esto indica que, para una estatura de 0 cm, el peso predicho sería -72.56 kg, lo cual también carece de significado práctico. Sin embargo, este intercepto es necesario para construir la ecuación de la recta de regresión.

$\widehat{B_1}$ representa el cambio promedio en el peso por cada incremento de una unidad en la estatura (cm).

- Para Hombres (Modelo1H): La pendiente $\widehat{B_1}$ es 94.66. Esto significa que, para hombres, por cada incremento de 1 cm en la estatura, se espera que el peso promedio aumente en 94.66 gramos (0.9466 kg). Este coeficiente es estadísticamente significativo, lo que indica que existe una relación fuerte y positiva entre la estatura y el peso para los hombres en los datos analizados.
- Para Mujeres (Modelo1M): La pendiente $\widehat{B_1}$ es 81.15. Para mujeres, esto significa que, por cada incremento de 1 cm en la estatura, se espera que el peso promedio aumente en 81.15 gramos (0.8115 kg). Aunque también es una relación positiva entre estatura y peso, el coeficiente de determinación (R^2) es menor que en el modelo de los hombres, lo que sugiere que la estatura no explica tanto la variación en el peso de las mujeres como lo hace en los hombres.

En el modelo combinado que incluye la variable Sexo, la pendiente para Estatura (89.26) es el efecto promedio de la estatura sobre el peso cuando se considera ambos géneros. Además, la variable SexoM (un indicador binario que toma el valor 1 para mujeres y 0 para hombres) tiene un coeficiente de -10.56. Esto implica que, en promedio, el peso de las mujeres es aproximadamente 10.56 kg menos que el de los hombres, manteniendo constante la estatura. Este modelo muestra que el efecto de la estatura sobre el peso es similar en hombres y mujeres, pero hay una diferencia base entre ambos géneros.