7. Regresión Lineal

Ozner Leyva

2024-08-30

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

```
M = read.csv("C:/Users/ozner/Desktop/Computer Science/R/Estatura-peso_HyM.csv")
data <- data.frame(Estatura = M$Estatura, Peso = M$Peso)
correlacion <- cor(data)
print(correlacion)
```

```
##           Estatura      Peso
## Estatura 1.0000000 0.8032449
## Peso     0.8032449 1.0000000
```

Podemos ver una correlacion alta por lo que si aumenta una una variable, la otra deberia aumentar en cierta medida.

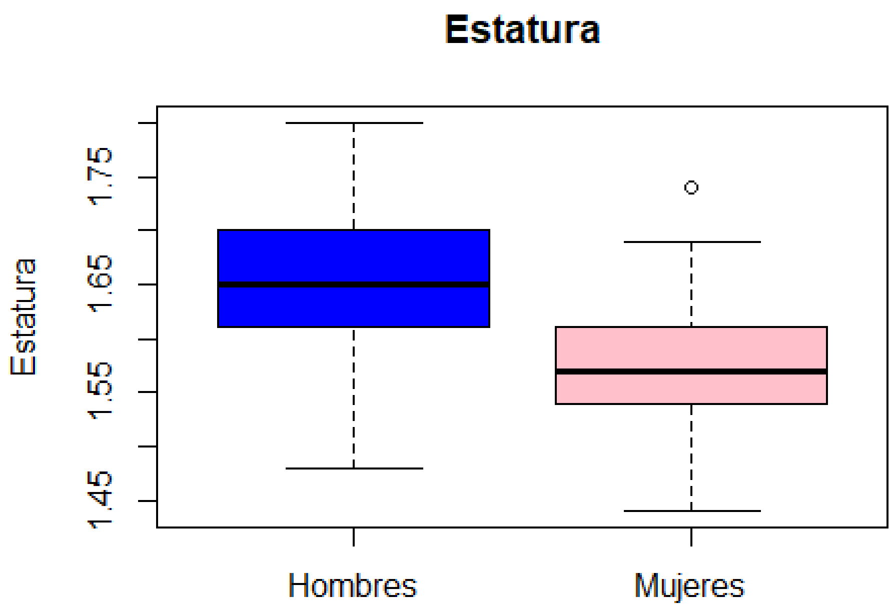
```
MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)

n=4
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")
```

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

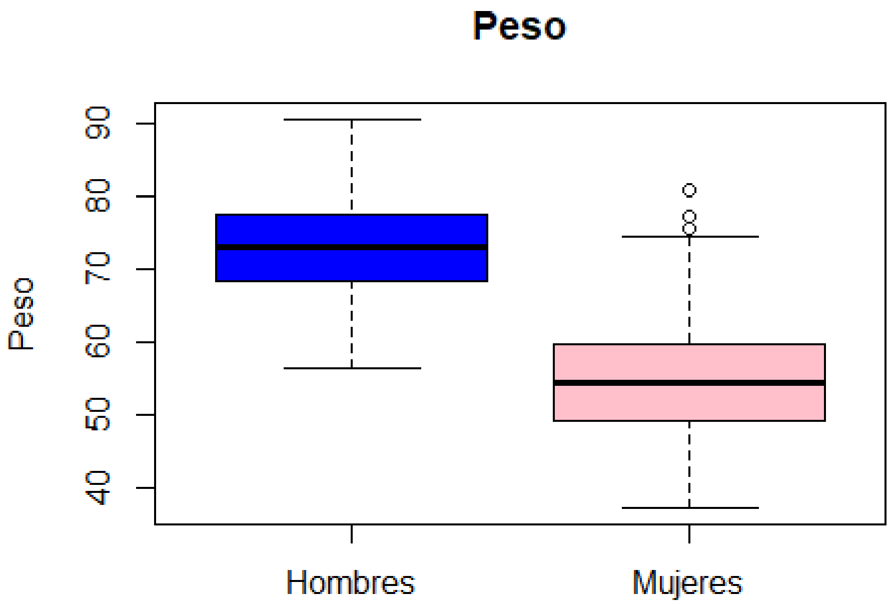
```
resumen <- summary(data)
print(resumen)
```

```
##     Estatura          Peso      
##  Min.   :1.440   Min.   :37.39  
##  1st Qu.:1.560   1st Qu.:54.49  
##  Median :1.610   Median :64.53  
##  Mean   :1.613   Mean   :63.97  
##  3rd Qu.:1.660   3rd Qu.:73.22  
##  Max.   :1.800   Max.   :90.49
```

```r
desviacion_estandar <- apply(data, 2, sd)
print(desviacion_estandar)

##    Estatura        Peso
##  0.06929171 11.54161456

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")
```



3. Encuentra la ecuación de regresión de mejor ajuste:

Hombres

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

##
## Call:
## lm(formula = Estatura ~ Peso, data = MH)
##
## Coefficients:
## (Intercept)         Peso
##    1.101770     0.007576
```

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

```
summary(Modelo1H)

##
## Call:
## lm(formula = Estatura ~ Peso, data = MH)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.091473 -0.020942  0.001445  0.024020  0.082089
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1017704  0.0235832   46.72   <2e-16 ***
## Peso        0.0075758  0.0003223   23.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03291 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
```

Mujeres

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

```
##
## Call:
## lm(formula = Estatura ~ Peso, data = MM)
##
## Coefficients:
## (Intercept)         Peso
##     1.38622      0.00339
```

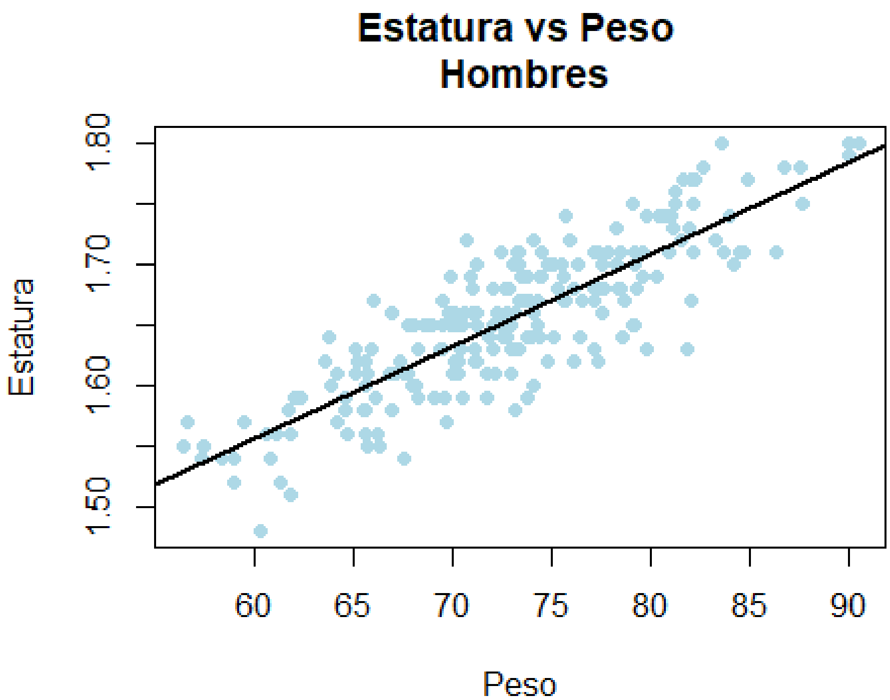
```
summary(Modelo1M)
```

```
##
## Call:
## lm(formula = Estatura ~ Peso, data = MM)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11162 -0.02611 -0.00174  0.02806  0.12814
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3862212  0.0207336  66.859   <2e-16 ***
## Peso        0.0033900  0.0003727   9.096   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04298 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

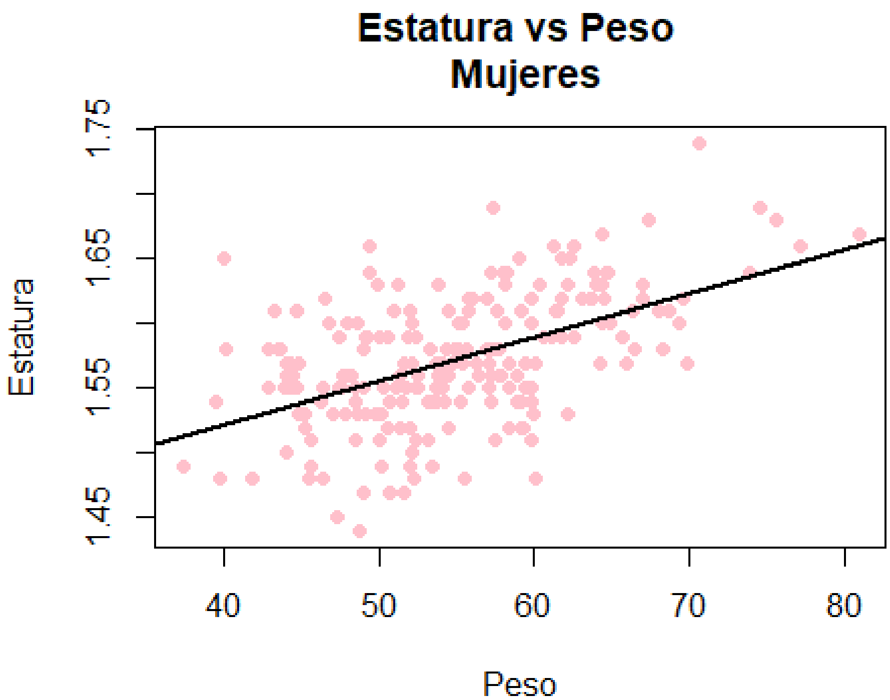
Hombres

```
plot(MH$Peso, MH$Estatura, col="lightblue", main="Estatura vs Peso \n
Hombres", ylab="Estatura", xlab="Peso", pch=19)
abline(Modelo1H, col="black", lwd=2)
```



Mujeres

```
plot(MM$Peso, MM$Estatura, col="pink", main="Estatura vs Peso \n
Mujeres", ylab="Estatura", xlab="Peso", pch=19)
abline(Modelo1M, col="black", lwd=2)
```



Modelo 2

```
Modelo2 = lm(Estatura~Peso, M)
Modelo2

##
## Call:
## lm(formula = Estatura ~ Peso, data = M)
##
## Coefficients:
## (Intercept)         Peso
##    1.304850     0.004822
```

Verifica el modelo:

Verifica la significancia del modelo con un alfa de 0.03.

Verifica la significancia de $\hat{\beta}i$ con un alfa de 0.03.

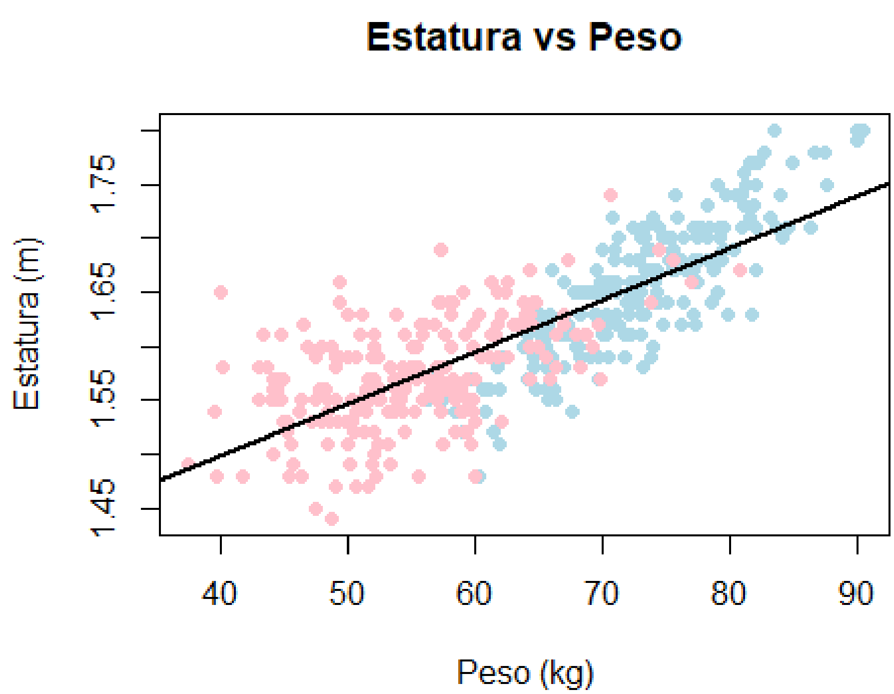
Verifica el porcentaje de variación explicada por el modelo.

```
summary(Modelo2)

##
## Call:
## lm(formula = Estatura ~ Peso, data = M)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
```

```
## -0.115881 -0.026479 -0.001523  0.027500  0.152206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3048497  0.0111068  117.48   <2e-16 ***
## Peso        0.0048224  0.0001709   28.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04132 on 438 degrees of freedom
## Multiple R-squared:  0.6452, Adjusted R-squared:  0.6444
## F-statistic: 796.5 on 1 and 438 DF,  p-value: < 2.2e-16
```

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

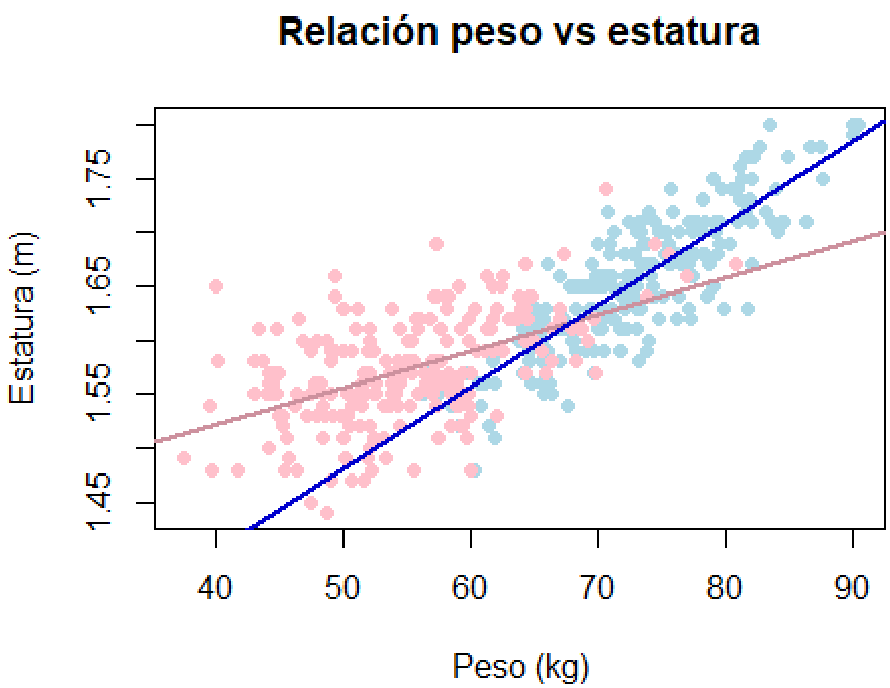
```
colors <- ifelse(M$Sexo == "H", "lightblue", "pink")
plot(M$Peso, M$Estatura, main="Estatura vs Peso",
    xlab="Peso (kg)", ylab="Estatura (m)", col=colors, pch=19)
abline(Modelo2, col="black", lwd=2)
```



Con 0.05 si es significativo.

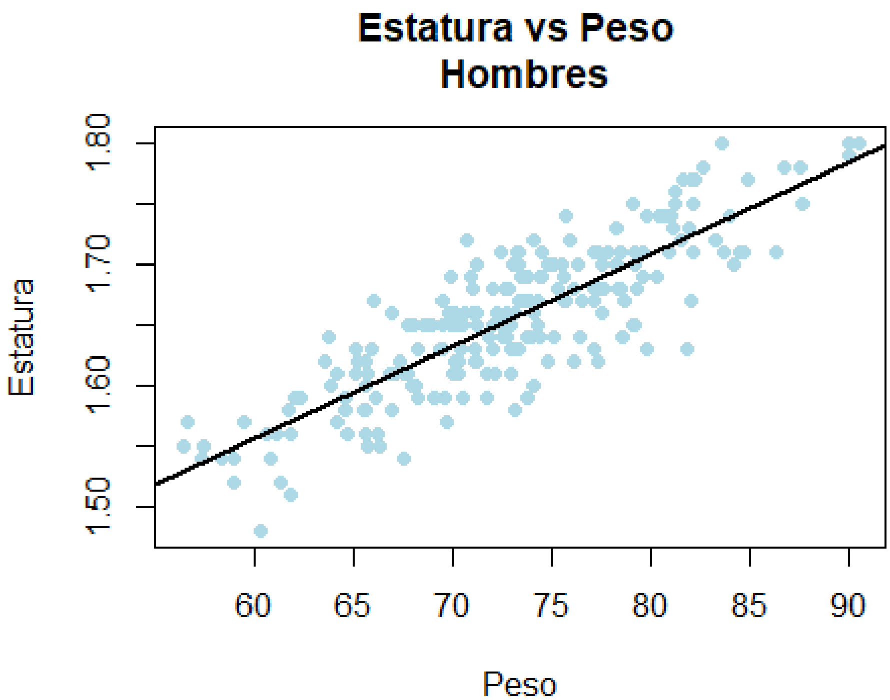
```
b0 <- Modelo2$coefficients[1]
b1 <- Modelo2$coefficients[2]
b2 <- Modelo2$coefficients[3]

Ym <- function(x){b0 + b2 + b1*x}
Yh <- function(x){b0 + b1*x}
```

```
colors <- ifelse(M$Sexo == "H", "lightblue", "pink")

plot(M$Peso, M$Estatura, col=colors, pch=19,
     xlab="Peso (kg)", ylab="Estatura (m)", main="Relación peso vs 
estatura")

x <- seq(1.40, 1.80, 0.01)

lines(x, Ym(x), col="pink3", lwd=2)
lines(x, Yh(x), col="blue3", lwd=2)

abline(Modelo1M, col="pink3", lwd=2)
abline(Modelo1H, col="blue3", lwd=2)
```



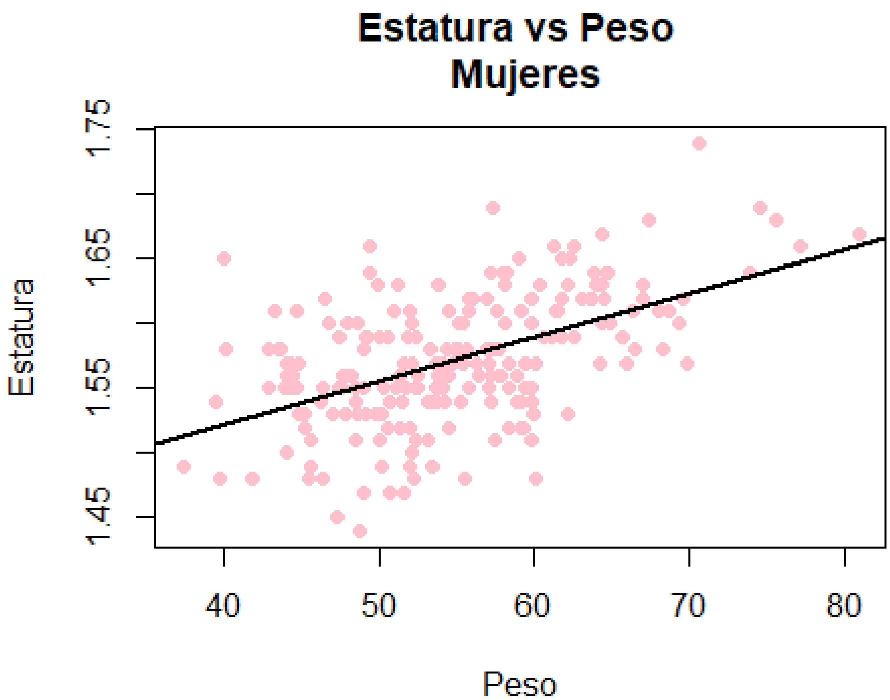
Hombres

```
plot(MH$Peso, MH$Estatura, col="lightblue", main="Estatura vs Peso \n 
Hombres", ylab="Estatura", xlab="Peso", pch=19)
abline(Modelo1H, col="black", lwd=2)
```



Mujeres

```
plot(MM$Peso, MM$Estatura, col="pink", main="Estatura vs Peso \n
Mujeres", ylab="Estatura", xlab="Peso", pch=19)
abline(Modelo1M, col="black", lwd=2)
```



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

- Correlación Estatura - Peso: Existe una fuerte correlación positiva (r = 0.803) entre estatura y peso. Esto implica que individuos más altos tienden a pesar más, siendo el peso un buen predictor de la estatura en este conjunto de datos.

- Estadísticas Descriptivas: Los nos revelan diferencias notables entre los géneros:

Hombres: Estatura media de 1.613 m, peso medio de 63.97 kg Mujeres: Generalmente más bajas y ligeras Las variaciones en valores mínimos, máximos y cuartiles nos indican la diversidad dentro de cada grupo.

- Diagramas Caja: Los gráficos muestran distribuciones distintas para hombres y mujeres:

Hombres: Valores más altos en estatura y peso Ambos géneros: Presencia de valores atípicos, indicando casos que se desvían significativamente de la mediana

- Análisis de Regresión: La regresión lineal muestra una relación significativa entre estatura y peso para ambos géneros:

Hombres: R^2 ajustado ≈ 0.717 Mujeres: R^2 ajustado ≈ 0.271 Esto nos sugiere que el peso explica mejor la variabilidad de la estatura en hombres que en mujeres.

- Gráficos de Dispersión: Los gráficos confirman la relación positiva entre estatura y peso. Las líneas de regresión difieren entre géneros, que nos indican que el impacto del peso en la estatura varía según el sexo.

6. Interpreta en el contexto del problema:

¿Qué información proporciona $\hat{\beta}0$ sobre la relación entre la estatura y el peso de hombres y mujeres?

Modelo de hombres: El intercepto en el modelo masculino (aproximadamente 1.10 metros) representa la estatura teórica si el peso fuera cero. Aunque este valor no tiene sentido biológico (ya que nadie puede tener un peso de cero kilos) es útil matemáticamente para establecer el punto de partida de la línea de regresión.

Modelo de mujeres: De manera similar, el intercepto femenino (cerca de 1.38 metros) tampoco refleja una realidad física. Sin embargo, la comparación entre los interceptos entre hombres y mujeres ofrece información valiosa: sugiere que, en promedio y a igualdad de peso, las mujeres tienden a ser ligeramente más bajas que los hombres en este conjunto de datos.

¿Cómo interpretas $\hat{\beta}1$ en la relación entre la estatura y el peso de hombres y mujeres?

Para hombres: Por cada kilogramo adicional de peso, la estatura aumenta en promedio 0.0076 metros (0.76 cm). Esta relación relativamente fuerte se refleja en el alto valor de R^2 ajustado (0.717) mencionado anteriormente.

Para mujeres: Por cada kilogramo adicional de peso, la estatura aumenta en promedio 0.00339 metros (0.339 cm). Este incremento es menos de la mitad del observado en hombres, lo que nos indica una relación más débil entre peso y estatura en mujeres. La relación más débil se confirma con el menor valor de R^2 ajustado (0.271) en el modelo femenino.