

# Tarea 10 - Regresión Lineal

Héctor Hibrán Tapia Fernández - A01661114

2024-09-03

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

```
df = read.csv("./Estatura-peso_HyM.csv")
df_numerico <- df[sapply(df, is.numeric)]
matriz_correlacion <- cor(df_numerico, use = "complete.obs")
print(matriz_correlacion)
```

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

La matriz muestra la relación (lineal) que tienen dos variables, 1 es una relación (lineal) perfecta y 0 sería que no tienen relación (lineal) y valores negativos sería que serían que tienen una relación (lineal) negativa.

Ambas variables tienen una relación (lineal) positiva muy fuerte lo que dice que están relacionadas y dando sentido a los números es verdad, la estatura tiene que ver mucho con el peso y el peso con la estatura.

## 2. Obtén medidas (media, desviación estándar, etc)

# que te ayuden a analizar los datos.

```
df_hombres = subset(df, df$Sexo == "H")
df_mujeres = subset(df, df$Sexo == "M")

df_medidas = data.frame(
  "H-Estatura" = df_hombres$Estatura,
  "H-Peso" = df_hombres$Peso,
  "M-Estatura" = df_mujeres$Estatura,
  "M-Peso" = df_mujeres$Peso
)

n = 4

d = matrix(NA, ncol = 8, nrow = n)

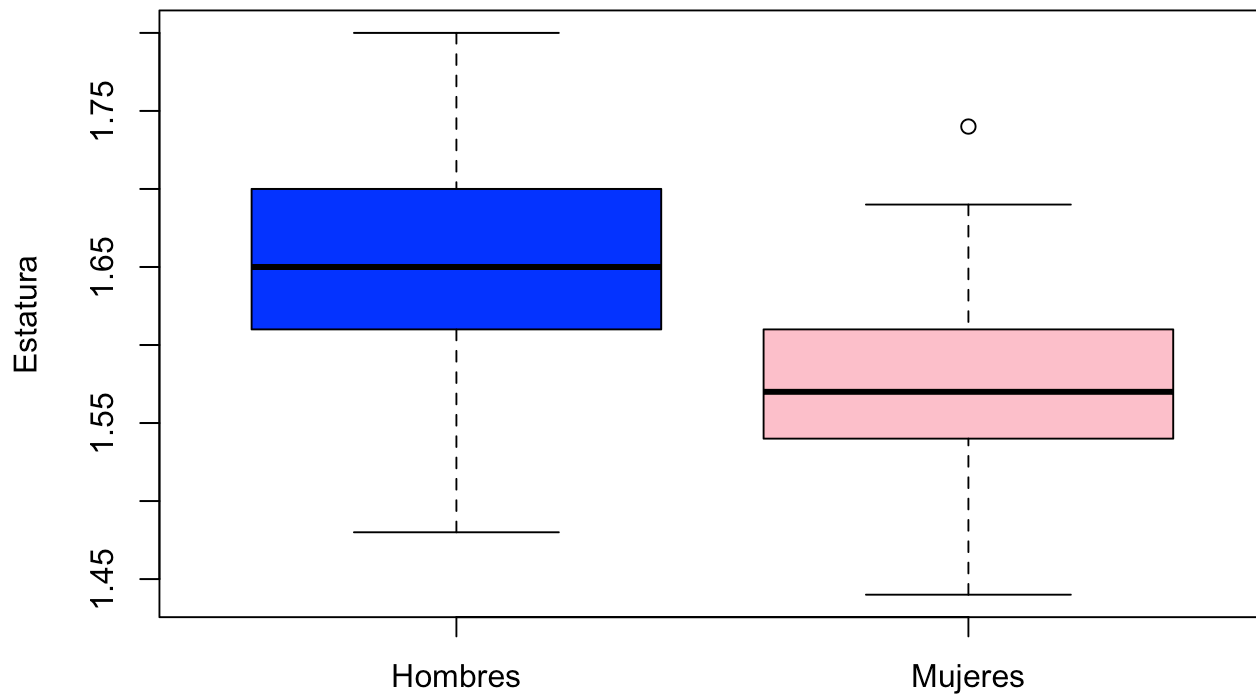
for (i in 1:n) {
  d[i, ] <- c(
    as.numeric(summary(df_medidas[, i])),
    sd(df_medidas[, i]),
    sd(df_medidas[, i]) / mean(df_medidas[, i])
  )
}

medidas = as.data.frame(d)
row.names(medidas) = c("H-Estatura", "H-Peso", "M-Estatura", "M-Peso")
names(medidas) = c("Minimo", "Q1", "Mediana", "Media", "Q3", "Máximo", "Desv Est", "CV")
print(medidas)
```

```
##           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
##              CV
## H-Estatura  0.03732833
## H-Peso      0.09471004
## M-Estatura  0.03202100
## M-Peso      0.14147238
```

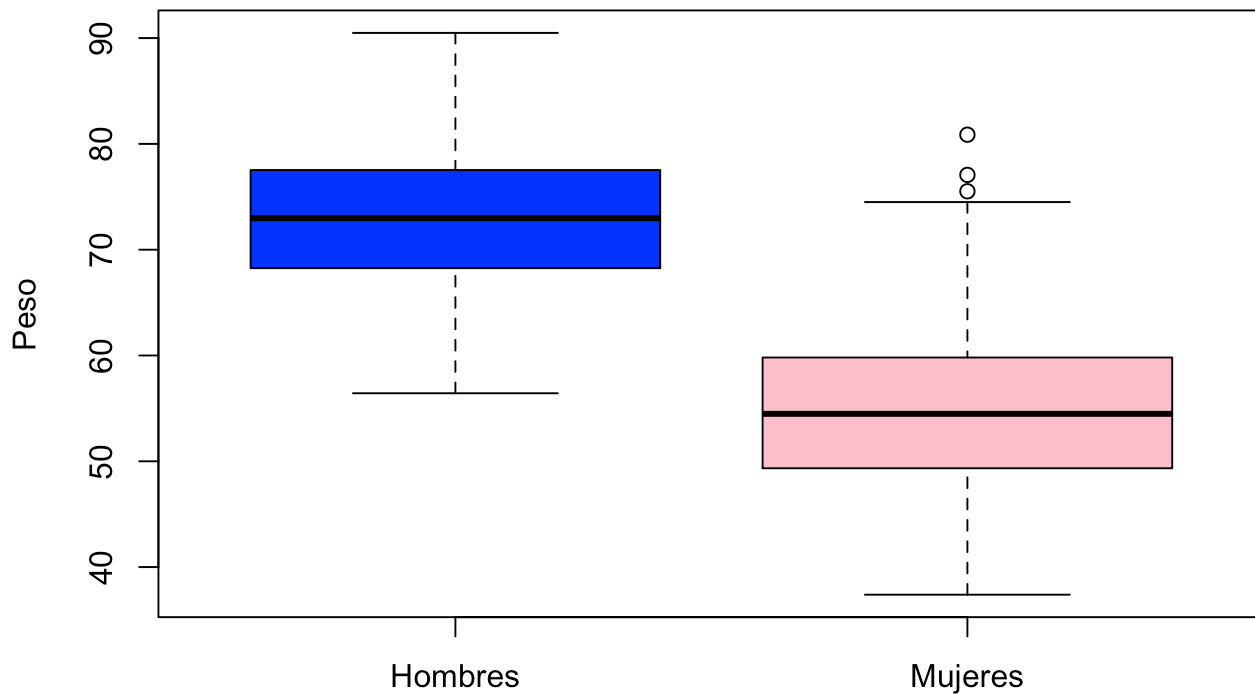
```
boxplot(df$Estatura~df$Sexo, ylab="Estatura", xlab="", col=c("blue","pink"), names=c("Hombres", "Mujeres"), main="Estatura")
```

## Estatura



```
boxplot(df$Peso ~ df$Sexo, ylab = "Peso", xlab = "", names = c("Hombres", "Mujeres"), col = c("blue", "pink"), main = "Peso")
```

## Peso



**3. Encuentra la ecuación de regresión de mejor ajuste haciendo el ajuste por separado para Hombres y Mujeres por separado e incluyendo hombres y mujeres como variable Dummy. Para los 3 modelos:**

**3.1 Realiza la regresión entre las variables involucradas.**

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

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

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

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

Hipótesis:

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

```
sH = summary(Modelo1H)
print(sH)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = df_hombres)
##
## 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
```

```
sM = summary(Modelo1M)
print(sM)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = df_mujeres)
##
## 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
```

## 3.2 Significancia del modelo:

- Valida la significancia del modelo con un alfa de 0.03 (incluye las hipótesis que pruebas)

Para Hombres:

```
p_valor <- sH$coefficients[2, 4]

if (p_valor < 0.03) {cat("Se rechaza la hipótesis nula (H0). Hay evidencia suficiente pa
ra afirmar que existe una relación lineal significativa entre Estatura y Peso.\n")} else
{cat("No se rechaza la hipótesis nula (H0). No hay evidencia suficiente para afirmar que
existe una relación lineal significativa entre Estatura y Peso.\n")}
}
```

```
## Se rechaza la hipótesis nula (H0). Hay evidencia suficiente para afirmar que existe u
na relación lineal significativa entre Estatura y Peso.
```

Para Mujeres:

```
p_valor = sM$coefficients[2, 4]

if (p_valor < 0.03) {cat("Se rechaza la hipótesis nula (H0). Hay evidencia suficiente pa
ra afirmar que existe una relación lineal significativa entre Estatura y Peso.\n")} else
{cat("No se rechaza la hipótesis nula (H0). No hay evidencia suficiente para afirmar que
existe una relación lineal significativa entre Estatura y Peso.\n")}
}
```

```
## Se rechaza la hipótesis nula (H0). Hay evidencia suficiente para afirmar que existe u
na relación lineal significativa entre Estatura y Peso.
```

- Validar la significancia de  $\beta_i$  con un alfa de 0.03 (incluye las hipótesis que pruebas)

Para Hombres:

```
coeficientes <- sH$coefficients

for (i in 1:nrow(coeficientes)) {
  coeficiente <- coeficientes[i, 1]
  error_estandar <- coeficientes[i, 2]
  valor_t <- coeficientes[i, 3]
  valor_p <- coeficientes[i, 4]

  cat(paste("Coeficiente:", rownames(coeficientes)[i], "\n"))
  cat(paste("  Valor del coeficiente:", round(coeficiente, 4), "\n"))
  cat(paste("  Error estándar:", round(error_estandar, 4), "\n"))
  cat(paste("  Valor t:", round(valor_t, 4), "\n"))
  cat(paste("  Valor p:", round(valor_p, 4), "\n"))

  if (valor_p < 0.03) {
    cat("  Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.\n\n")
  } else {
    cat("  Resultado: No se rechaza la hipótesis nula (H0). El coeficiente no es significativo.\n\n")
  }
}
```

```
## Coeficiente: (Intercept)
##  Valor del coeficiente: -83.6845
##  Error estándar: 6.6635
##  Valor t: -12.5587
##  Valor p: 0
##  Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.
##
## Coeficiente: Estatura
##  Valor del coeficiente: 94.6602
##  Error estándar: 4.0266
##  Valor t: 23.5089
##  Valor p: 0
##  Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.
```

Para Mujeres:

```

coeficientes <- sM$coefficients

for (i in 1:nrow(coeficientes)) {
  coeficiente <- coeficientes[i, 1]
  error_estandar <- coeficientes[i, 2]
  valor_t <- coeficientes[i, 3]
  valor_p <- coeficientes[i, 4]

  cat(paste("Coeficiente:", rownames(coeficientes)[i], "\n"))
  cat(paste(" Valor del coeficiente:", round(coeficiente, 4), "\n"))
  cat(paste(" Error estándar:", round(error_estandar, 4), "\n"))
  cat(paste(" Valor t:", round(valor_t, 4), "\n"))
  cat(paste(" Valor p:", round(valor_p, 4), "\n"))

  if (valor_p < 0.03) {
    cat(" Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.\n\n")
  } else {
    cat(" Resultado: No se rechaza la hipótesis nula (H0). El coeficiente no es significativo.\n\n")
  }
}

```

```

## Coeficiente: (Intercept)
## Valor del coeficiente: -72.5604
## Error estándar: 14.0408
## Valor t: -5.1678
## Valor p: 0
## Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.
##
## Coeficiente: Estatura
## Valor del coeficiente: 81.1491
## Error estándar: 8.9218
## Valor t: 9.0956
## Valor p: 0
## Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.

```

- Indica cuál es el porcentaje de variación explicada por el modelo.

```

R2 = sH$r.squared
porcentaje_variacion_explicada <- R2 * 100
cat("Porcentaje de variación explicada por el modelo es:", round(porcentaje_variacion_explicada, 2), "%\n")

```

```
## Porcentaje de variación explicada por el modelo es: 71.71 %
```

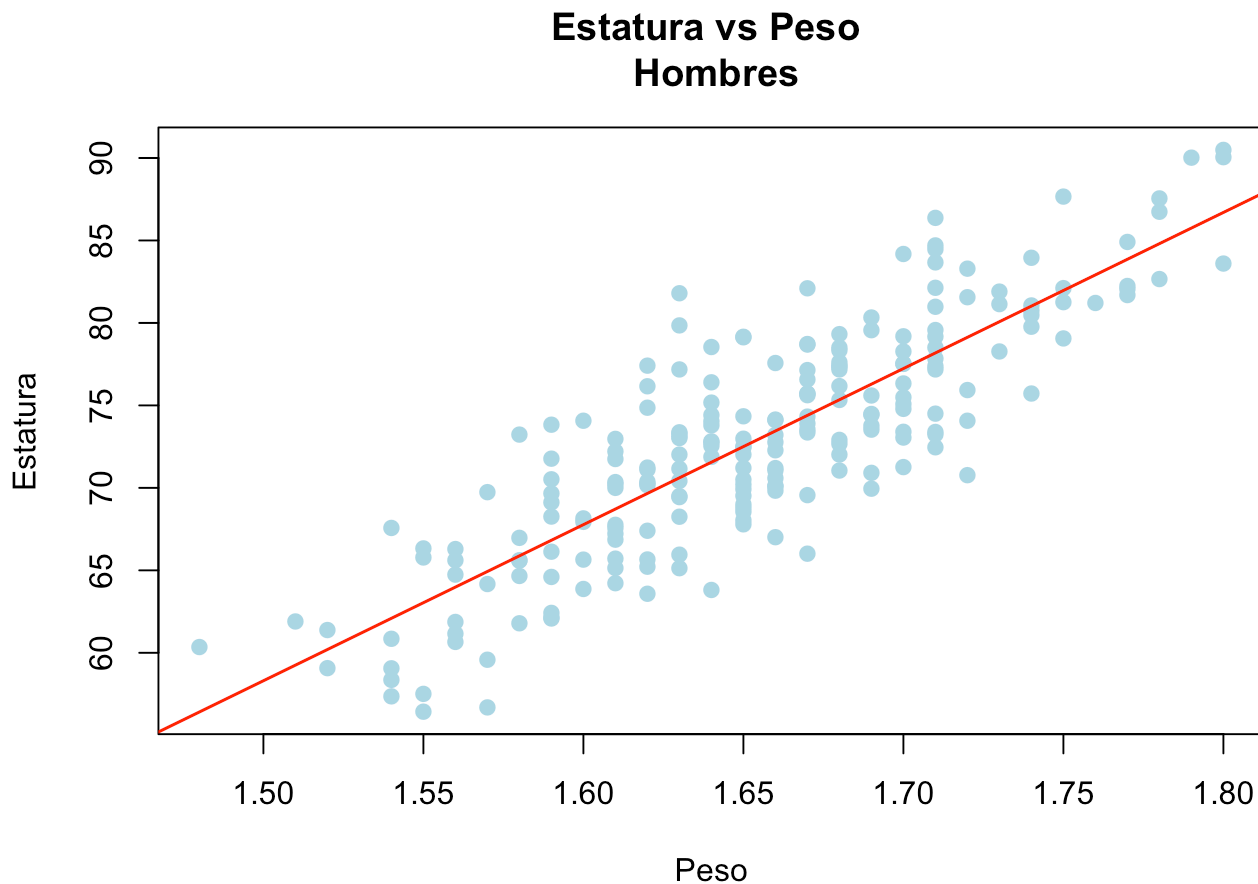


```
R2 = sM$r.squared
porcentaje_variacion_explicada <- R2 * 100
cat("Porcentaje de variación explicada por el modelo es:", round(porcentaje_variacion_explicada, 2), "%\n")
```

```
## Porcentaje de variación explicada por el modelo es: 27.51 %
```

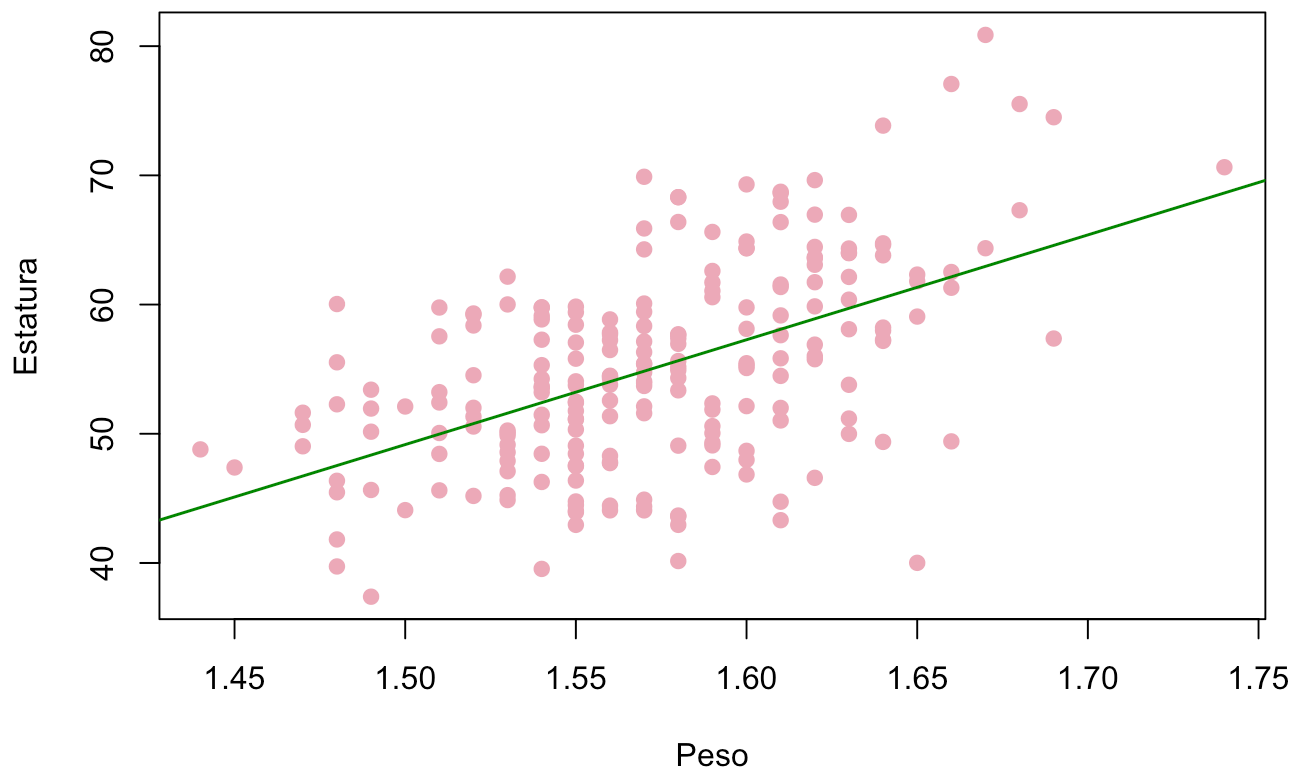
### 3.3 Dibuja el diagrama de dispersión de los datos y la recta de mejor ajuste.

```
plot(df_hombres$Estatura, df_hombres$Peso, col = 'lightblue', main = "Estatura vs Peso \n Hombres", ylab = "Estatura", xlab = "Peso", pch = 19)
abline(Modelo1H, col = 'red', lwd = 1.5)
```



```
plot(df_mujeres$Estatura, df_mujeres$Peso, col = 'pink2', main = "Estatura vs Peso \n Mujeres", ylab = "Estatura", xlab = "Peso", pch = 19)
abline(Modelo1M, col = 'green4', lwd = 1.5)
```

## Estatura vs Peso Mujeres



```
plot(df_hombres$Estatura, df_hombres$Peso, col = 'blue', main = "Estatura vs Peso \n Hom
bres y Mujeres", ylab = "Peso", xlab = "Estatura", pch = 19, xlim = c(1.45, 1.80), ylim
= c(40, 90))
```

```
points(df_mujeres$Estatura, df_mujeres$Peso, col = 'pink2', pch = 19)
```

```
abline(Modelo1H, col = 'blue', lwd = 2.5)
abline(Modelo1M, col = 'pink', lwd = 2.5)
```

## Estatura vs Peso Hombres y Mujeres

