

# Regresión Lineal

Catherine Rojas

2024-08-30

## La recta de mejor ajuste (Primera entrega)

```
library(readr)

## Warning: package 'readr' was built under R version 4.3.3

data <- read_csv("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.
```

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

```
dataM = subset(data, data$Sexo=="M")
dataH = subset(data, data$Sexo == "H")
data1 = data.frame(dataH$Estatura, dataH$Peso, dataM$Estatura,
dataM$Peso)

# Matriz de correlación
correlation_matrix <- cor(data1)

# Mostrar resultados
print("Matriz de correlación:")

## [1] "Matriz de correlación:"

correlation_matrix

##           dataH.Estatura dataH.Peso dataM.Estatura dataM.Peso
## dataH.Estatura  1.0000000000 0.846834792  0.0005540612 0.04724872
## dataH.Peso      0.8468347920 1.000000000  0.0035132246 0.02154907
```

```
## dataM.Estatura  0.0005540612 0.003513225  1.0000000000 0.52449621
## dataM.Peso      0.0472487231 0.021549075   0.5244962115 1.00000000
```

## Interpretación

- Hombres: Existe una fuerte relación entre estatura y peso, lo que sugiere que a mayor estatura, mayor peso.
- Mujeres: Existe una correlación positiva moderada entre estatura y peso.
- Relaciones cruzadas: Las estaturas y pesos entre hombres y mujeres no muestran correlaciones significativas entre sí, indicando que estos grupos se comportan de manera independiente en términos de las variables consideradas.

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

```
# Función para calcular todas las medidas estadísticas
calcular_medidas <- function(df, variable) {
  medidas <- data.frame(
    Medida = c("Media", "Mediana", "Mínimo", "Máximo", "Q1", "Q3",
"Desviación Estándar"),
    Valor = c(mean(df[[variable]], na.rm = TRUE),
median(df[[variable]], na.rm = TRUE),
min(df[[variable]], na.rm = TRUE),
max(df[[variable]], na.rm = TRUE),
quantile(df[[variable]], 0.25, na.rm = TRUE),
quantile(df[[variable]], 0.75, na.rm = TRUE),
sd(df[[variable]], na.rm = TRUE))
  )
  return(medidas)
}

# Calcular medidas para estatura y peso de hombres
medidas_estatura_hombres <- calcular_medidas(dataH, "Estatura")
medidas_peso_hombres <- calcular_medidas(dataH, "Peso")

# Calcular medidas para estatura y peso de mujeres
medidas_estatura_mujeres <- calcular_medidas(dataM, "Estatura")
medidas_peso_mujeres <- calcular_medidas(dataM, "Peso")

# Crear un solo data.frame consolidado con todas las medidas
resultados <- data.frame(
  Medida = medidas_estatura_hombres$Medida,
  Hombres_Estatura = medidas_estatura_hombres$Valor,
  Mujeres_Estatura = medidas_estatura_mujeres$Valor,
  Hombres_Peso = medidas_peso_hombres$Valor,
  Mujeres_Peso = medidas_peso_mujeres$Valor
)
```

)

resultados

```
##           Medida Hombres_Estatura Mujeres_Estatura Hombres_Peso
## 1           Media      1.65372727      1.57295455      72.857682
## 2          Mediana      1.65000000      1.57000000      72.975000
## 3           Mínimo      1.48000000      1.44000000      56.430000
## 4           Máximo      1.80000000      1.74000000      90.490000
## 5             Q1      1.61000000      1.54000000      68.257500
## 6             Q3      1.70000000      1.61000000      77.522500
## 7 Desviación Estándar      0.06173088      0.05036758      6.900354
## Mujeres_Peso
## 1      55.083409
## 2      54.485000
## 3      37.390000
## 4      80.870000
## 5      49.355000
## 6      59.795000
## 7       7.792781
```

**Observaciones** \* Los hombres son en promedio, más altos y pesan más que las mujeres. \* Los datos de peso presentan mayor variabilidad en las mujeres, lo que sugiere que el peso de las mujeres varía más ampliamente que el de los hombres.

### 3. Encuentra la ecuación de regresión de mejor ajuste:

```
# Ajustar el modelo de regresión para hombres
modelo_hombres <- lm(Peso ~ Estatura, data = dataH)

# Ajustar el modelo de regresión para mujeres
modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)

# Ajustar el modelo de regresión general
modelo <- lm(Peso ~ Estatura + Sexo, data)

# Mostrar los resultados del modelo
summary(modelo_hombres)

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

```
summary(modelo_mujeres)
```

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

```
summary(modelo)
```

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

# Obtener la ecuación de regresión para hombres
cat("Ecuación de regresión para hombres modelo: Peso =",
coef(modelo_hombres)[1], "+", coef(modelo_hombres)[2], "* Estatura\n")

## Ecuación de regresión para hombres modelo: Peso = -83.68454 + 94.66024
* Estatura

# Obtener la ecuación de regresión para mujeres
cat("Ecuación de regresión para mujeres modelo : Peso =",
coef(modelo_mujeres)[1], "+", coef(modelo_mujeres)[2], "* Estatura\n")

## Ecuación de regresión para mujeres modelo : Peso = -72.56045 +
81.14911 * Estatura

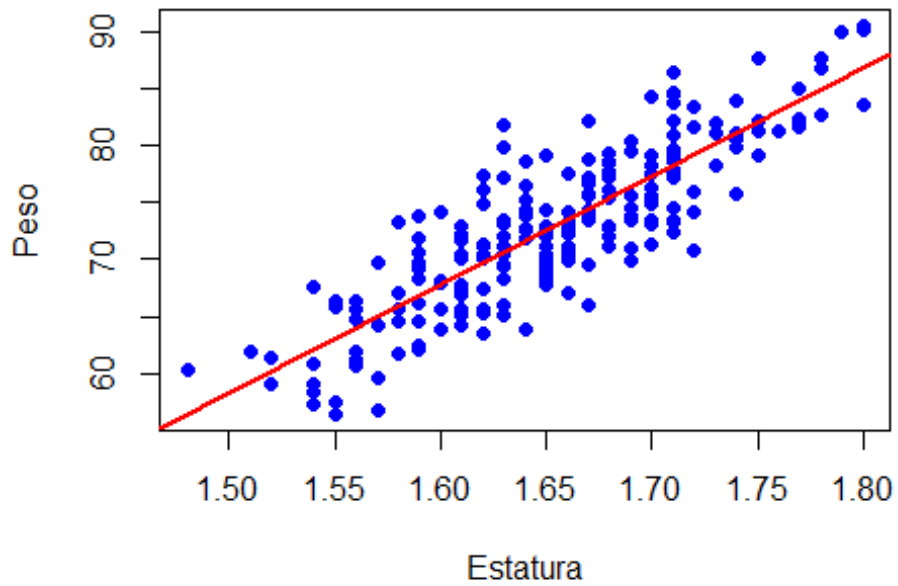
# Obtener la ecuación de regresión para el modelo general
cat("Ecuación de regresión para el modelo general: Peso =",
coef(modelo)[1], "+", coef(modelo)[2], "* Estatura\n")

## Ecuación de regresión para el modelo general: Peso = -74.7546 +
89.26035 * Estatura
```

#### 4. Realiza la regresión entre las variables involucradas

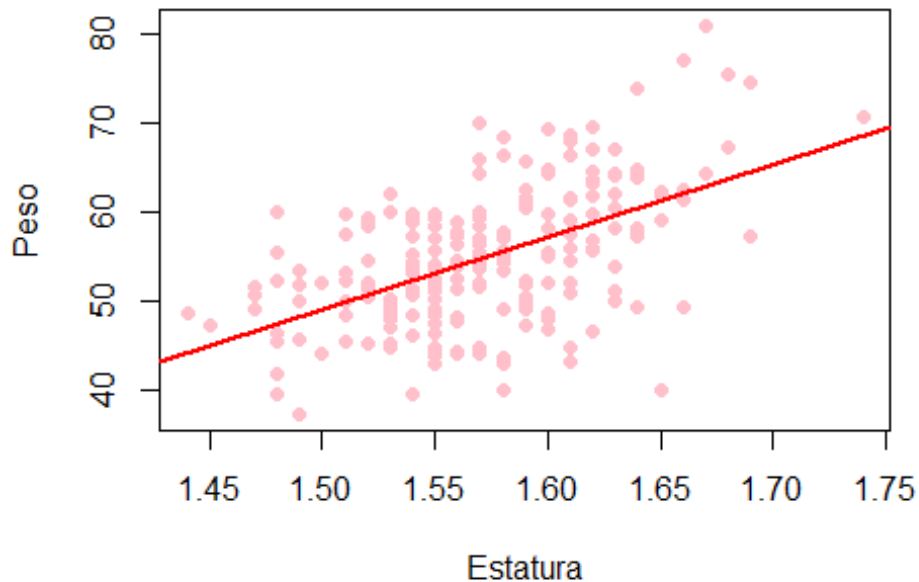
```
# Gráfico para hombres
plot(data1$dataH.Estatura, data1$dataH.Peso, main = "Hombres: Estatura vs
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "blue")
abline(modelo_hombres, col = "red", lwd = 2)
```

## Hombres: Estatura vs Peso



```
# Gráfico para mujeres  
plot(data1$dataM.Estatura, data1$dataM.Peso, main = "Mujeres: Estatura vs  
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "pink")  
abline(modelo_mujeres, col = "red", lwd = 2)
```

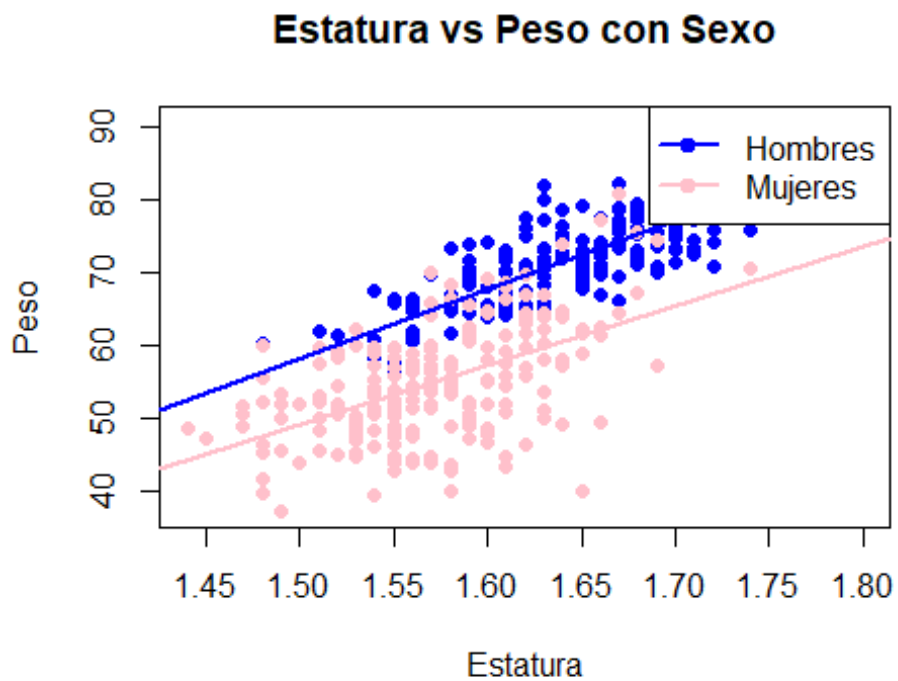
## Mujeres: Estatura vs Peso



```
# Gráfico de estatura vs peso con distinción por sexo
plot(data$Estatura, data$Peso, col = ifelse(data$Sexo == "H", "blue",
"pink"), pch = 19,
      xlab = "Estatura", ylab = "Peso", main = "Estatura vs Peso con
Sexo",
      xlim = c(min(data$Estatura), max(data$Estatura)),
      ylim = c(min(data$Peso), max(data$Peso)))

# Añadir líneas de regresión por sexo
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "H")), col =
"blue", lwd = 2)
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "M")), col =
"pink", lwd = 2)

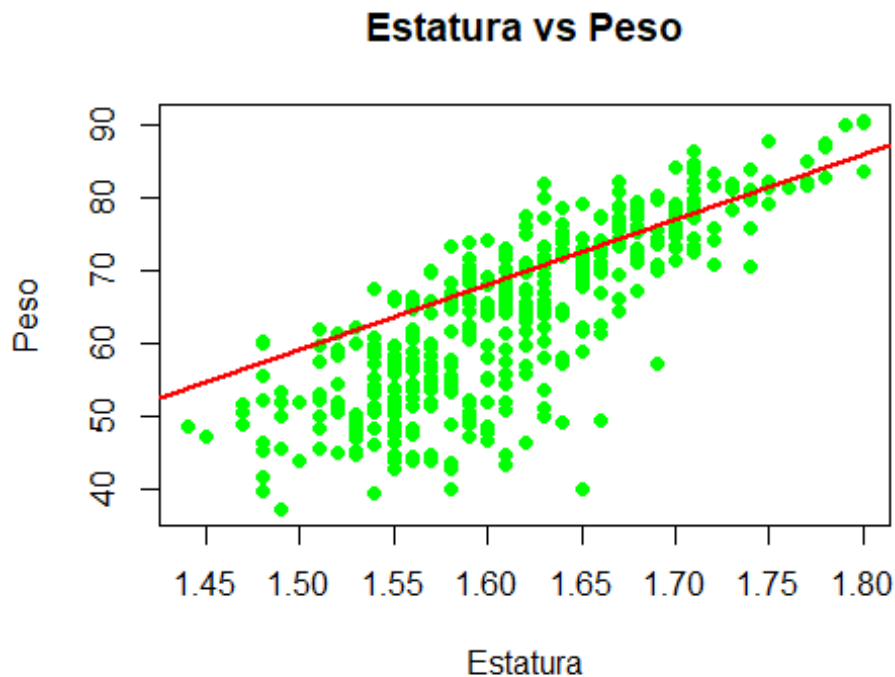
# Añadir Leyenda
legend("topright", legend = c("Hombres", "Mujeres"), col = c("blue",
"pink"), lwd=2,pch=19)
```



```
# Gráfico general
plot(data$Estatura, data$Peso, main = "Estatura vs Peso", xlab =
"Estatura", ylab = "Peso", pch = 19, col = "green")
abline(modelo, col = "red", lwd = 2)

## Warning in abline(modelo, col = "red", lwd = 2): only using the first
two of 3
## regression coefficients
```





### Interpretaciones de los modelos

- Comparación entre Hombres y Mujeres:

La estatura es un predictor más fuerte del peso para los hombres que para las mujeres, como lo indica el  $R^2$  más alto en el modelo para hombres. Por otro lado, el modelo para mujeres tiene un  $R^2$  significativamente más bajo, lo que sugiere que la estatura no es tan buen predictor del peso en mujeres como lo es en hombres.

- Modelo General: Cuando se incluyen tanto la estatura como el sexo en el modelo, se explica una mayor proporción de la variabilidad en el peso (78.27%), lo que indica que ambos factores son importantes para predecir el peso.

## 5. Verifica el modelo:

### a) Verifica la significancia del modelo con un alfa de 0.03.

```
# Resumen del modelo para obtener valores p
summary_hombres <- summary(modelo_hombres)
summary_mujeres <- summary(modelo_mujeres)
summary_general <- summary(modelo)

# Extraer Los valores p del modelo para hombres y mujeres
p_value_hombres <- summary_hombres$coefficients[2, 4]
p_value_mujeres <- summary_mujeres$coefficients[2, 4]
p_value_general <- summary_general$coefficients[2, 4]
```

```

# Verificar la significancia con un alfa de 0.03
alpha <- 0.03

cat("Significancia del modelo para Hombres:\n")

## Significancia del modelo para Hombres:

if (p_value_hombres < alpha) {
  cat("El modelo es significativo (p-value =", p_value_hombres, "<",
alpha, ").\n")
} else {
  cat("El modelo no es significativo (p-value =", p_value_hombres, ">=",
alpha, ").\n")
}

## El modelo es significativo (p-value = 1.063532e-61 < 0.03 ).

cat("\nSignificancia del modelo para Mujeres:\n")

##
## Significancia del modelo para Mujeres:

if (p_value_mujeres < alpha) {
  cat("El modelo es significativo (p-value =", p_value_mujeres, "<",
alpha, ").\n")
} else {
  cat("El modelo no es significativo (p-value =", p_value_mujeres, ">=",
alpha, ").\n")
}

## El modelo es significativo (p-value = 5.997517e-17 < 0.03 ).

cat("\nSignificancia del modelo general:\n")

##
## Significancia del modelo general:

if (p_value_general < alpha) {
  cat("El modelo es significativo (p-value =", p_value_general, "<",
alpha, ").\n")
} else {
  cat("El modelo no es significativo (p-value =", p_value_general, ">=",
alpha, ").\n")
}

## El modelo es significativo (p-value = 1.179664e-61 < 0.03 ).

```

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

```

# Extraer Los valores p para el intercepto y La pendiente (Estatura)
p_value_intercepto_hombres <- summary_hombres$coefficients[1, 4]
p_value_intercepto_mujeres <- summary_mujeres$coefficients[1, 4]

```

```

p_value_intercepto_general <- summary_general$coefficients[1, 4]

# Verificar la significancia con un alfa de 0.03
alpha <- 0.03

cat("Significancia de los coeficientes para Hombres:\n")

## Significancia de los coeficientes para Hombres:

cat("Intercepto ( $\beta_0$ ):\n")

## Intercepto ( $\beta_0$ ):

if (p_value_intercepto_hombres < alpha) {
  cat("Significativo (p-value =", p_value_intercepto_hombres, "<", alpha,
  ").\n")
} else {
  cat("No significativo (p-value =", p_value_intercepto_hombres, ">=",
  alpha, ").\n")
}

## Significativo (p-value = 1.410453e-27 < 0.03 ).

cat("Pendiente ( $\beta_1$  - Estatura):\n")

## Pendiente ( $\beta_1$  - Estatura):

if (p_value_hombres < alpha) {
  cat("Significativo (p-value =", p_value_hombres, "<", alpha, ").\n")
} else {
  cat("No significativo (p-value =", p_value_hombres, ">=", alpha,
  ").\n")
}

## Significativo (p-value = 1.063532e-61 < 0.03 ).

cat("\nSignificancia de los coeficientes para Mujeres:\n")

##
## Significancia de los coeficientes para Mujeres:

cat("Intercepto ( $\beta_0$ ):\n")

## Intercepto ( $\beta_0$ ):

if (p_value_intercepto_mujeres < alpha) {
  cat("Significativo (p-value =", p_value_intercepto_mujeres, "<", alpha,
  ").\n")
} else {
  cat("No significativo (p-value =", p_value_intercepto_mujeres, ">=",
  alpha, ").\n")
}

```

```
## Significativo (p-value = 5.342762e-07 < 0.03 ).
cat("Pendiente ( $\beta_1$  - Estatura):\n")
## Pendiente ( $\beta_1$  - Estatura):
if (p_value_mujeres < alpha) {
  cat("Significativo (p-value =", p_value_mujeres, "<", alpha, ").\n")
} else {
  cat("No significativo (p-value =", p_value_mujeres, ">=", alpha,
  ").\n")
}

## Significativo (p-value = 5.997517e-17 < 0.03 ).
cat("\nSignificancia de los coeficientes General:\n")
##
## Significancia de los coeficientes General:
cat("Intercepto ( $\beta_0$ ):\n")
## Intercepto ( $\beta_0$ ):
if (p_value_intercepto_general < alpha) {
  cat("Significativo (p-value =", p_value_intercepto_general, "<", alpha,
  ").\n")
} else {
  cat("No significativo (p-value =", p_value_intercepto_general, ">=",
  alpha, ").\n")
}

## Significativo (p-value = 5.821123e-21 < 0.03 ).
cat("Pendiente ( $\beta_1$  - Estatura):\n")
## Pendiente ( $\beta_1$  - Estatura):
if (p_value_general < alpha) {
  cat("Significativo (p-value =", p_value_general, "<", alpha, ").\n")
} else {
  cat("No significativo (p-value =", p_value_general, ">=", alpha,
  ").\n")
}

## Significativo (p-value = 1.179664e-61 < 0.03 ).
```

### c) Verifica el porcentaje de variación explicada por el modelo

```
# Extraer el valor de R-cuadrado para cada modelo
r_squared_hombres <- summary_hombres$r.squared
r_squared_mujeres <- summary_mujeres$r.squared
r_squared_general <- summary_general$r.squared
```

```
# Mostrar el porcentaje de variación explicada por el modelo
cat("Porcentaje de variación explicada por el modelo para Hombres:",
r_squared_hombres * 100, "%\n")

## Porcentaje de variación explicada por el modelo para Hombres: 71.71292
%

cat("Porcentaje de variación explicada por el modelo para Mujeres:",
r_squared_mujeres * 100, "%\n")

## Porcentaje de variación explicada por el modelo para Mujeres: 27.50963
%

cat("Porcentaje de variación explicada por el modelo general:",
r_squared_general * 100, "%\n")

## Porcentaje de variación explicada por el modelo general: 78.36599 %
```

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

Los diagramas se realizaron previamente.

La recta de mejor ajuste se refiere al modelo de regresión lineal que mejor predice la variable dependiente (Peso) en función de la variable independiente (Estatura) y, en el caso del modelo general, también del Sexo. La mejor recta de ajuste es aquella que maximiza el valor de  $R^2$ , que indica el porcentaje de la variación en la variable dependiente explicada por el modelo.

La mejor recta de ajuste es la del modelo general con la siguiente ecuación:

$$\text{Peso} = -74.7546 + 89.26035 * \text{Estatura}$$

Es la mejor recta de ajuste ya que tiene un  $R^2$  del 78.27%, este modelo explica más de la variabilidad en el peso que los modelos separados para hombres y mujeres. Este modelo no solo considera la estatura, sino que también incluye el sexo, lo que mejora significativamente la capacidad del modelo para explicar el peso.

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

**Matriz de correlación:** La estatura es un buen predictor del peso, especialmente en hombres. Sin embargo, en mujeres, la estatura por sí sola no explica el peso tan bien, lo que sugiere la necesidad de considerar otros factores.

**Medidas estadísticas:** Estas medidas ayudan a entender la distribución y variabilidad de las estaturas y pesos en la población estudiada, lo que es importante para la construcción de modelos predictivos.

**Ecuaciones de regresión:** El modelo general es el más eficaz porque tiene en cuenta tanto la estatura como el sexo, lo que lo convierte en una herramienta más precisa para predecir el peso en la población. Este modelo sugiere que, para una misma estatura, las mujeres tienden a pesar menos que los hombres, lo cual es coherente con las observaciones anteriores.

**Verificación de la Significancia de los Modelos y Coeficientes:** La significancia estadística asegura que los resultados obtenidos no son debidos al azar y que las relaciones entre estatura, sexo y peso son consistentes y fiables para la población estudiada.

**$R^2$**  El alto  $R^2$  del modelo general sugiere que es el modelo más efectivo para predecir el peso en comparación con el resto de los modelos. En contextos de salud o nutrición, es importante entender cómo diferentes factores contribuyen al peso corporal.

A lo largo de los análisis, se observa que la estatura es un predictor fuerte del peso, especialmente en hombres, mientras que en mujeres otros factores podrían ser más influyentes. La inclusión del sexo en el modelo general mejora significativamente la capacidad predictiva del modelo, haciendo de este el más robusto y útil para predecir el peso en la población estudiada. Esto es relevante en contextos donde se busca entender mejor cómo las características físicas influyen en el peso corporal y, potencialmente, en la salud.

## 8. Interpreta en el contexto del problema:

### a) ¿Qué información proporciona $\beta_0$ sobre la relación entre la estatura y el peso de hombres y mujeres?

$\beta_0$  es el intercepto de la línea de regresión y tiene un valor teórico para el peso cuando la estatura es cero.

- Hombres: Un valor más negativo del intercepto (-83.68454) en comparación con el de las mujeres (-72.56045) sugiere que, para una misma estatura, los hombres tienden a tener un peso mayor en comparación con las mujeres cuando se analiza únicamente por la estatura.
- Mujeres: El intercepto menos negativo (-72.56045) sugiere que, al mantener la estatura constante, las mujeres suelen pesar menos que los hombres en este modelo.

### b) ¿Cómo interpretas $\beta_1$ en la relación entre la estatura y el peso de hombres y mujeres?

El coeficiente  $\beta_1$  en una regresión lineal representa la pendiente de la línea de regresión. Este coeficiente muestra cómo cambia la variable dependiente (en este caso, el peso) en respuesta a un cambio en la variable independiente (estatura).

- Para Hombres  $\beta_1 = 94.66$ : La estatura tiene un impacto muy fuerte en el peso, con un aumento de 94.66 kg por cada metro adicional de estatura. Este es el valor más alto entre los modelos, lo que indica una relación muy pronunciada entre estatura y peso en hombres.
- Para Mujeres  $\beta_1 = 81.15$ : La estatura también tiene un impacto positivo en el peso, pero con un aumento más moderado de 81.15 kg por cada metro adicional. Esto sugiere que, aunque la estatura es un buen predictor del peso, su efecto es menor en comparación con los hombres.
- Modelo General  $\beta_1 = 89.26$ : Este valor intermedio refleja el impacto promedio de la estatura sobre el peso cuando se consideran ambos sexos. Este modelo proporciona una estimación más general que se ajusta a ambos géneros.