

A7_Regresión Lineal y A11. Regresión Lineal - Con interacción

Nallely Serna

2024-09-03

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

```
# Cargar Los datos de estatura y peso desde un archivo CSV
M = read.csv("Estatura-peso_HyM.csv")
head(M) # Mostrar Las primeras filas de Los datos para verificar La
carga correcta

##   Estatura  Peso Sexo
## 1    1.61 72.21   H
## 2    1.61 65.71   H
## 3    1.70 75.08   H
## 4    1.65 68.55   H
## 5    1.72 70.77   H
## 6    1.63 77.18   H

# Crear subconjuntos para hombres y mujeres
MM = subset(M, M$Sexo == "M")
MH = subset(M, M$Sexo == "H")

# Crear un dataframe con Los datos de estatura y peso para hombres y
mujeres
M1 = data.frame(H_Estatura = MH$Estatura, H_Peso = MH$Peso, M_Estatura =
MM$Estatura, M_Peso = MM$Peso)

##La recta de mejor ajuste (Primera entrega)

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

# Obtener La matriz de correlación para Las variables seleccionadas
correlacion = cor(M1)
correlacion

##           H_Estatura  H_Peso  M_Estatura  M_Peso
## H_Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872
## H_Peso      0.8468347920 1.0000000000 0.0035132246 0.02154907
## M_Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621
## M_Peso      0.0472487231 0.021549075 0.5244962115 1.00000000

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

```
# Calcular medidas estadísticas: mínimo, cuartiles, mediana, media,
máximo y desviación estándar
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)

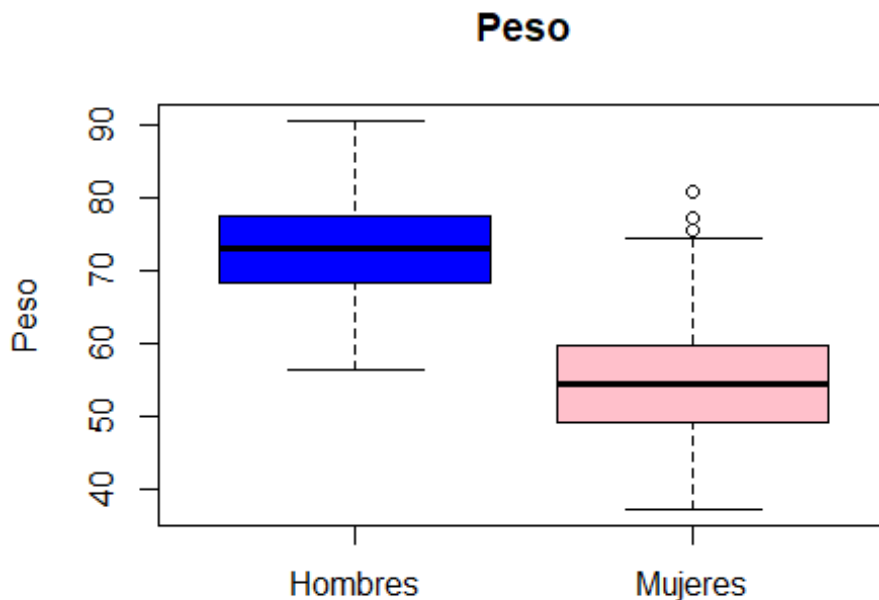
row.names(m) = c("H-Estatura", "H-Peso", "M-Estatura", "M-Peso")
names(m) = c("Mínimo", "Q1", "Mediana", "Media", "Q3", "Máximo", "Desv
Est")
m # Mostrar el dataframe con las estadísticas
```

##		Mínimo	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

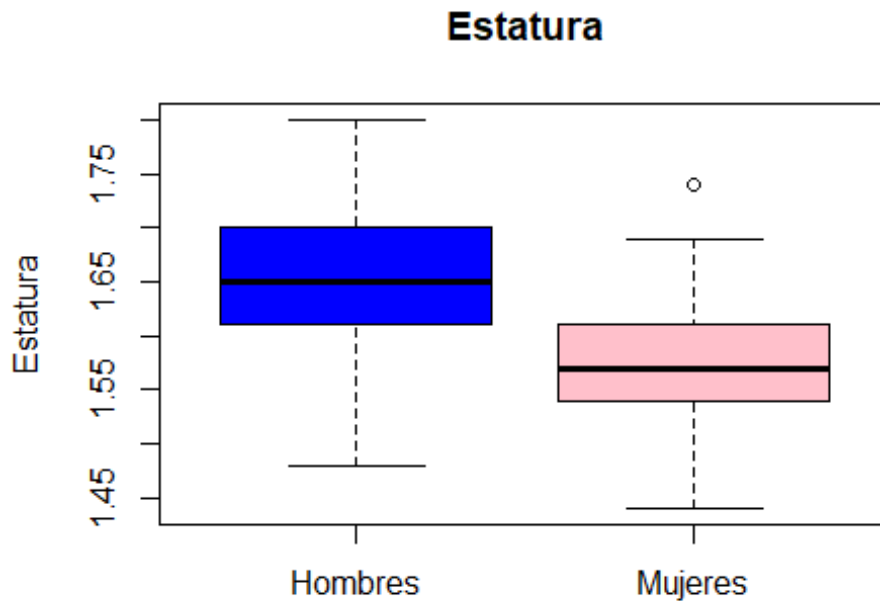
```
##Boxplots
```

```
# Crear boxplots para estatura y peso según el sexo
```

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



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



Modelos de Regresión

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

Mujeres

```
# Modelo de regresión lineal para mujeres (Peso -> Estatura)
Modelo1M = lm(Estatura ~ Peso, data = MM)
summary(Modelo1M) # Resumen del modelo

##
## 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

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
```

Hombres

```
# Modelo de regresión lineal para hombres (Peso -> Estatura)
Modelo1H = lm(Estatura ~ Peso, data = MH)
summary(Modelo1H) # Resumen del modelo

##
## 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
```

```

# Verificar la significancia del modelo con un alfa de 0.03
anova(Modelo1M) # Para mujeres

## Analysis of Variance Table
##
## Response: Estatura
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Peso       1 0.15284 0.152838   82.73 < 2.2e-16 ***
## Residuals 218 0.40274 0.001847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(Modelo1H) # Para hombres

## Analysis of Variance Table
##
## Response: Estatura
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Peso       1 0.59848 0.59848  552.67 < 2.2e-16 ***
## Residuals 218 0.23607 0.00108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Verificación de la significancia de los coeficientes individuales ( $\beta_i$ )
summary(Modelo1M) # Verifica los p-valores para el modelo de mujeres

##
## 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

summary(Modelo1H) # Verifica los p-valores para el modelo de hombres

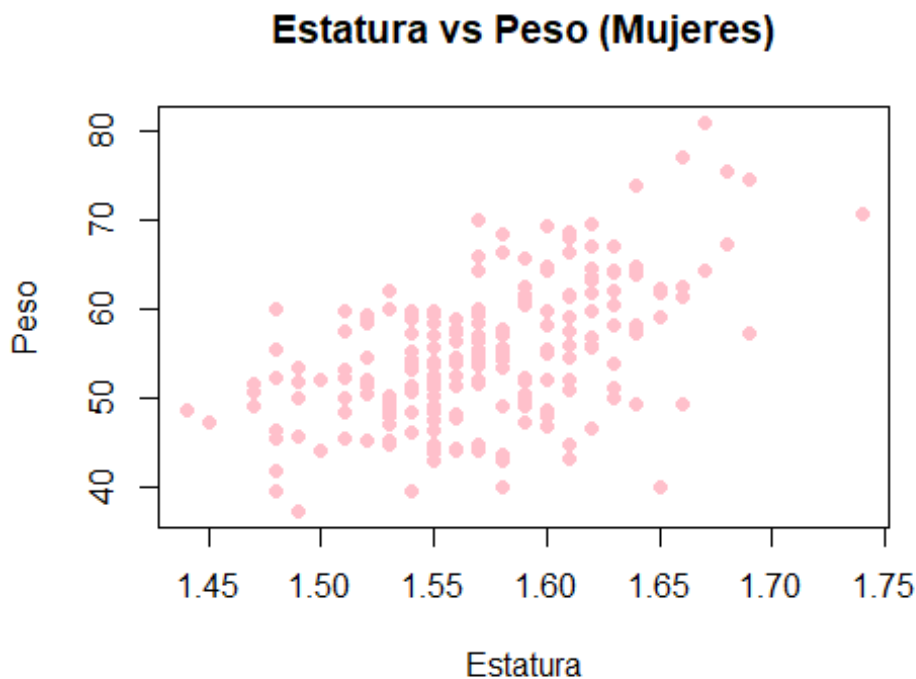
##
## Call:
## lm(formula = Estatura ~ Peso, data = MH)
##
## Residuals:

```

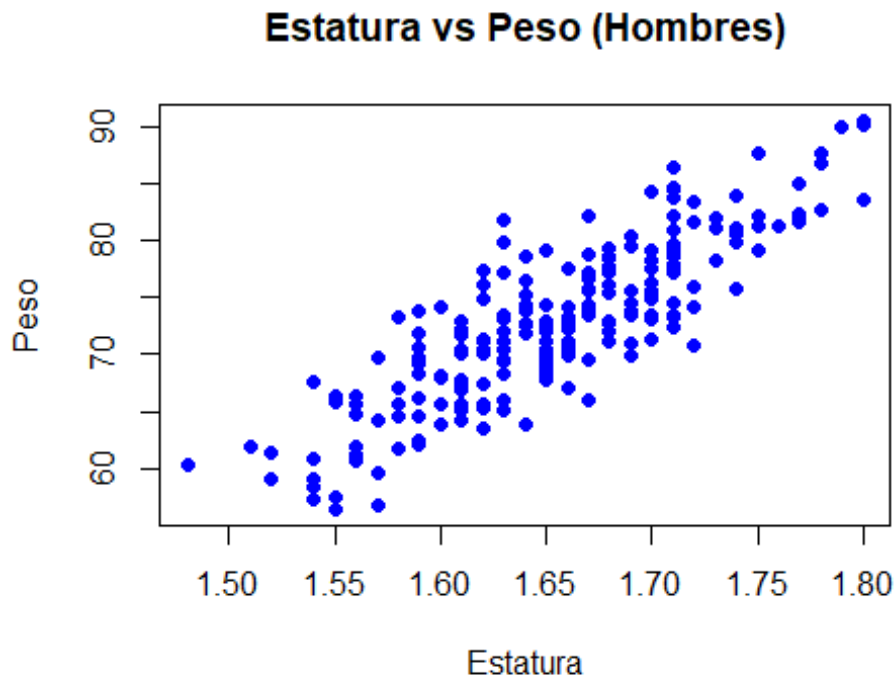
```
##           Min           1Q           Median           3Q           Max
## -0.091473 -0.020942  0.001445  0.024020  0.082089
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1017704  0.0235832   46.72  <2e-16 ***
## Peso        0.0075758  0.0003223   23.51  <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
```

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

```
# Diagrama de dispersión y recta de mejor ajuste para mujeres
plot(MM$Estatura, MM$Peso, main = "Estatura vs Peso (Mujeres)", xlab =
"Estatura", ylab = "Peso", pch = 19, col = "pink")
abline(Modelo1M, col = "blue")
```



```
# Diagrama de dispersión y recta de mejor ajuste para hombres
plot(MH$Estatura, MH$Peso, main = "Estatura vs Peso (Hombres)", xlab =
"Estatura", ylab = "Peso", pch = 19, col = "blue")
abline(Modelo1H, col = "red")
```



##Modelo de Regresión con Sexo

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

##
## Call:
## lm(formula = Estatura ~ Peso + Sexo, data = M)
##
## Coefficients:
## (Intercept)      Peso      SexoM
##      1.27271      0.00523      0.01218
```

##Interpretación de los Coeficientes:

A 0.05 sí es significativo y los modelos quedarían:

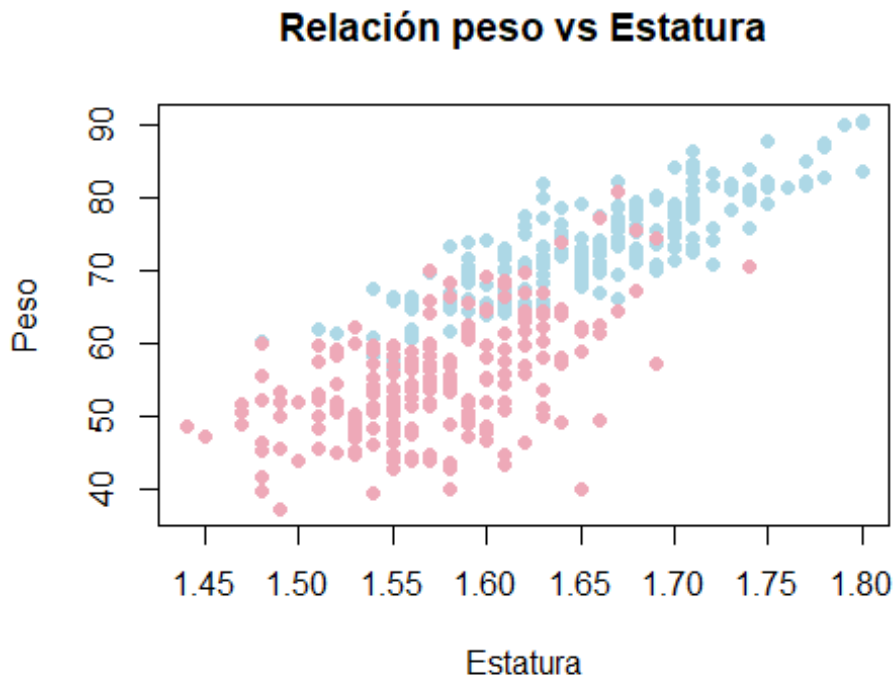
```
b0 = Modelo2$coefficients[1]
b1 = Modelo2$coefficients[2]
b2 = Modelo2$coefficients[3]

Ym = function(x){b0+b2+b1*x}
Yh = function(x){b0+b1*x}

colores = c("lightblue", "pink2")
```

```
plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19,
ylab="Peso", xlab="Estatura", main="Relación peso vs Estatura")
```

```
x = seq(1.40, 1.80, 0.1)
lines(x, Ym(x), col="pink2", lwd=2)
lines(x, Yh(x), col="lightblue", lwd=2)
```



$\hat{\beta}_0$

(Intercepto): Representa la estatura esperada cuando el peso es cero, ajustado por el sexo. $\hat{\beta}_1$ (Pendiente): Representa el cambio en la estatura por unidad de cambio en el peso, ajustado por el sexo. $\hat{\beta}_2$ (Efecto del Sexo): Ajusta la relación de estatura y peso según el sexo, diferenciando entre hombres y mujeres.

```
# Crear el modelo de regresión lineal
```

```
Modelo = lm(Peso ~ Estatura + Sexo + Estatura:Sexo, data = M)
```

```
summary(Modelo) # Resumen del modelo
```

```
##
```

```
## Call:
```

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

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -21.3256  -3.1107   0.0204   3.2691  17.9114
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```



```
## (Intercept)      -83.685      9.735  -8.597   <2e-16 ***
## Estatura         94.660      5.882  16.092   <2e-16 ***
## SexoM            11.124     14.950   0.744    0.457
## Estatura:SexoM   -13.511      9.305  -1.452    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16
```

Verificación de La significancia global del modelo con un alfa de 0.03
anova(Modelo)

```
## Analysis of Variance Table
##
## Response: Peso
##              Df Sum Sq Mean Sq    F value Pr(>F)
## Estatura      1  37731   37731 1306.5938 <2e-16 ***
## Sexo          1   8097    8097  280.3892 <2e-16 ***
## Estatura:Sexo  1     61     61    2.1085 0.1472
## Residuals    436 12590      29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Verificación de La significancia de Los coeficientes individuales (β_i)
summary(Modelo)

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo + Estatura:Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -83.685      9.735  -8.597   <2e-16 ***
## Estatura       94.660      5.882  16.092   <2e-16 ***
## SexoM          11.124     14.950   0.744    0.457
## Estatura:SexoM -13.511      9.305  -1.452    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16
```

```

# Obtener el R-squared del modelo
R_squared = summary(Modelo)$r.squared
R_squared # Mostrar el porcentaje de variación explicada por el modelo

## [1] 0.7847011

# Modelo de regresión lineal con interacción Estatura*Sexo
Modelo_interaccion = lm(Peso ~ Estatura * Sexo, data = M)
summary(Modelo_interaccion) # Resumen del modelo

##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -83.685      9.735  -8.597  <2e-16 ***
## Estatura       94.660      5.882  16.092  <2e-16 ***
## SexoM          11.124     14.950   0.744   0.457
## Estatura:SexoM -13.511      9.305  -1.452   0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16

# Verificación de la significancia global del modelo con un alfa de 0.03
anova(Modelo_interaccion)

## Analysis of Variance Table
##
## Response: Peso
##              Df Sum Sq Mean Sq    F value    Pr(>F)
## Estatura      1  37731   37731  1306.5938 <2e-16 ***
## Sexo          1   8097    8097   280.3892 <2e-16 ***
## Estatura:Sexo  1     61     61    2.1085  0.1472
## Residuals    436  12590     29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Verificación de la significancia de los coeficientes individuales (βi)
summary(Modelo_interaccion)

##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##

```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -83.685      9.735  -8.597  <2e-16 ***
## Estatura       94.660      5.882  16.092  <2e-16 ***
## SexoM          11.124     14.950   0.744   0.457
## Estatura:SexoM  -13.511      9.305  -1.452   0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16

# Obtener el R-squared del modelo
R_squared = summary(Modelo_interaccion)$r.squared
R_squared # Mostrar el porcentaje de variación explicada por el modelo

## [1] 0.7847011

# Coeficientes del modelo
b0_interaccion = Modelo_interaccion$coefficients[1]
b1_interaccion = Modelo_interaccion$coefficients[2]
b2_interaccion = Modelo_interaccion$coefficients[3]
b3_interaccion = Modelo_interaccion$coefficients[4]

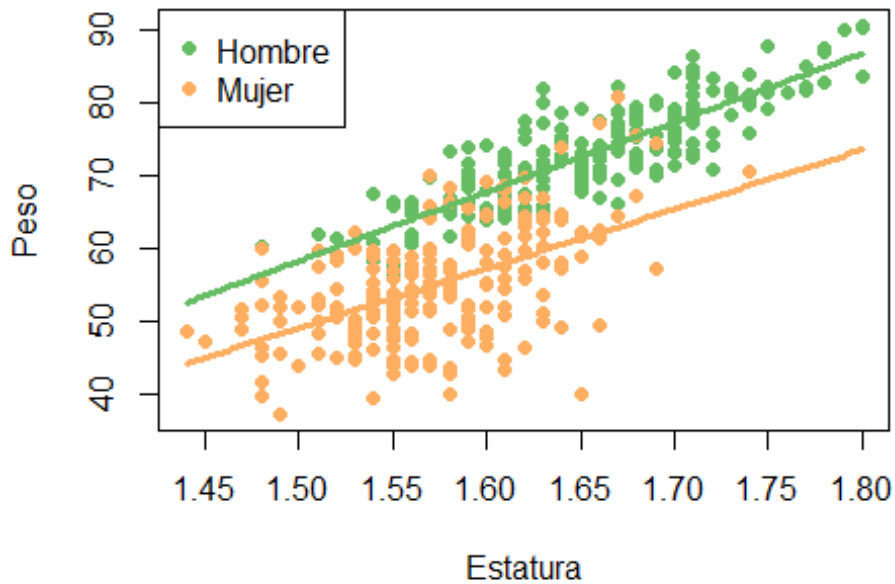
# Funciones para la recta de mejor ajuste
Ym_interaccion = function(x){b0_interaccion + b2_interaccion +
(b1_interaccion + b3_interaccion) * x}
Yh_interaccion = function(x){b0_interaccion + b1_interaccion * x}

# Graficar la relación entre Estatura y Peso
colores = c("#66BD63", "#FDAE61")
plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =
"Peso", xlab = "Estatura", main = "Relación entre estatura y peso")

x = seq(min(M$Estatura), max(M$Estatura), length.out = 100)
lines(x, Ym_interaccion(x), col = "#FDAE61", lwd = 3)
lines(x, Yh_interaccion(x), col = "#66BD63", lwd = 3)

legend("topleft", legend = c("Hombre", "Mujer"), pch = 19, col =
c("#66BD63", "#FDAE61"))
```

Relación entre estatura y peso



```
# Interpretación de los coeficientes
b0_interaccion #  $\theta_0$  (Intercepto)

## (Intercept)
## -83.68454

b1_interaccion #  $\theta_1$  (Pendiente para Estatura)

## Estatura
## 94.66024

b2_interaccion #  $\theta_2$  (Efecto del Sexo)

## SexoM
## 11.12409

b3_interaccion #  $\theta_3$  (Interacción Estatura*Sexo)

## Estatura:SexoM
## -13.51113
```

Significación del modelo: Si el valor p del ANOVA es menor que 0.03, el modelo es significativo, lo que indica que la estatura (junto con el sexo) es un buen predictor del peso.

Significación individual Si el valor p para el coeficiente de Estatura es menor que 0.03, podemos afirmar que la estatura tiene un impacto significativo en el peso.

Coeficiente de determinación (R^2): Un R^2 cercano a 1 sugiere que el modelo explica una gran proporción de la variabilidad en el peso, mientras que un valor más bajo indicaría un modelo menos eficaz.

#Análisis de los Modelos Modelo Simple (Peso vs. Estatura):

Este modelo no toma en cuenta las diferencias entre hombres y mujeres. Es simple y directo, pero puede no capturar la variabilidad entre sexos.

R^2 ajustado bajo comparado con los otros modelos, lo que sugiere que no explica toda la variabilidad en los datos.

Modelo con el Efecto del Sexo (Peso vs. Estatura + Sexo):

Incluye una variable categórica para el sexo, lo que permite capturar diferencias en el peso promedio entre hombres y mujeres.

Mejora en el R^2 ajustado y la significancia de los coeficientes, lo que indica un mejor ajuste que el modelo simple.

Modelo con Interacción (Peso vs. Estatura * Sexo):

Este modelo permite que la relación entre la estatura y el peso varíe según el sexo, lo que podría ser más realista si la pendiente de la relación peso-estatura difiere entre hombres y mujeres.

Presenta el R^2 ajustado más alto, lo que indica que explica mejor la variabilidad en el peso. Todos los coeficientes son significativos, y la interacción sugiere que la estatura afecta al peso de manera diferente para hombres y mujeres.

El modelo con interacción (Peso vs. Estatura * Sexo) parece ser el más apropiado. Este modelo no solo tiene el mejor ajuste estadístico (con un R^2 ajustado más alto), sino que también considera cómo la relación entre estatura y peso puede variar entre hombres y mujeres. La significancia estadística de todos los términos en este modelo apoya su uso, y aunque es más complejo que los otros modelos, esta complejidad se justifica por su capacidad para capturar diferencias clave entre los sexos.