Act11 Regresion Lineal

Luis Maximiliano López Ramírez

2024-09-03

1. Retoma el notebook en el que realizaste el análisis de regresión que encontraste 'La recta de mejor ajuste'

```r
datos <- read.csv("Estatura-peso_HyM.csv")

datos$Sexo_numeric <- ifelse(datos$Sexo == 'H', 0, 1)

dataM = subset(datos, datos$Sexo=="M")
dataH = subset(datos, datos$Sexo == "H")
```

2. Propón un nuevo modelo. Esta vez toma en cuenta la interacción de la Estatura con el Sexo y realiza los mismos pasos que hiciste con los modelos anteriores

1. Obtén el modelo e interpreta las variables Dummy

```r
modelo_general <- lm(Peso ~ Estatura * Sexo_numeric, data = datos)
```

2. Significancia del modelo

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

Para validar la significancia del modelo, se realiza una prueba F global. Las hipótesis a probar son las siguientes:

La prueba se realiza con un nivel de significancia de $\alpha = 0.03$. Si el valor p de la prueba F global es menor que 0.03, se rechaza la hipótesis nula y se concluye que el modelo es significativo.

```r
# Resumen del modelo general
resumen_general <- summary(modelo_general)

# Verificar la significancia del modelo (Prueba F global) para modelo
general
p_valor_f_general <- pf(resumen_general$fstatistic[1],
resumen_general$fstatistic[2], resumen_general$fstatistic[3], lower.tail
= FALSE)


cat("\n*** Significancia del Modelo general ***\n")
```

```
##
## *** Significancia del Modelo general ***
```

```r
cat("Valor p del modelo (Prueba F) - Modelo general:", p_valor_f_general,
"\n")
```

```
## Valor p del modelo (Prueba F) - Modelo general: 5.93158e-145
```

```r
if (p_valor_f_general < 0.03) {
  cat("El modelo es significativo a un nivel de alfa = 0.03\n")
} else {
  cat("El modelo NO es significativo a un nivel de alfa = 0.03\n")
}
```

```
## El modelo es significativo a un nivel de alfa = 0.03
```

2. Validala significancia de β̂i con un alfa de 0.03 (incluye las hipótesis que pruebas)

Para validar la significancia de cada coeficiente $\hat{\beta}_i$ del modelo, se realizan pruebas individuales para cada coeficiente. Las hipótesis a probar para cada coeficiente son las siguientes:

La prueba se realiza con un nivel de significancia de $\alpha = 0.03$. Si el valor p asociado al coeficiente $\hat{\beta}_i$ es menor que 0.03, se rechaza la hipótesis nula, concluyendo que el coeficiente es significativo y que la variable correspondiente influye significativamente en el modelo.

```r
# Verificar la significancia de los coeficientes (betai) para modelo
general
cat("\n*** Significancia de los Coeficientes - Modelo general ***\n")
```

```
##
## *** Significancia de los Coeficientes - Modelo general ***
```

```r
coeficientes_general <- resumen_general$coefficients
print(coeficientes_general)
```

```
##                        Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)           -83.68454   9.734572 -8.5966329 1.484670e-16
## Estatura               94.66024   5.882365 16.0922078 4.524701e-46
## Sexo_numeric           11.12409  14.949668  0.7441028 4.572151e-01
## Estatura:Sexo_numeric -13.51113   9.304767 -1.4520657 1.472025e-01
```

```r
# Evaluar significancia de cada coeficiente
for (i in 1:nrow(coeficientes_general)) {
  p_valor <- coeficientes_general[i, 4]
  cat(rownames(coeficientes_general)[i], "- Valor p:", p_valor)
  if (p_valor < 0.03) {
    cat(" -> Significativo\n")
  } else {
    cat(" -> No significativo\n")
```

```
  }
}

## (Intercept) - Valor p: 1.48467e-16 -> Significativo
## Estatura - Valor p: 4.524701e-46 -> Significativo
## Sexo_numeric - Valor p: 0.4572151 -> No significativo
## Estatura:Sexo_numeric - Valor p: 0.1472025 -> No significativo
```

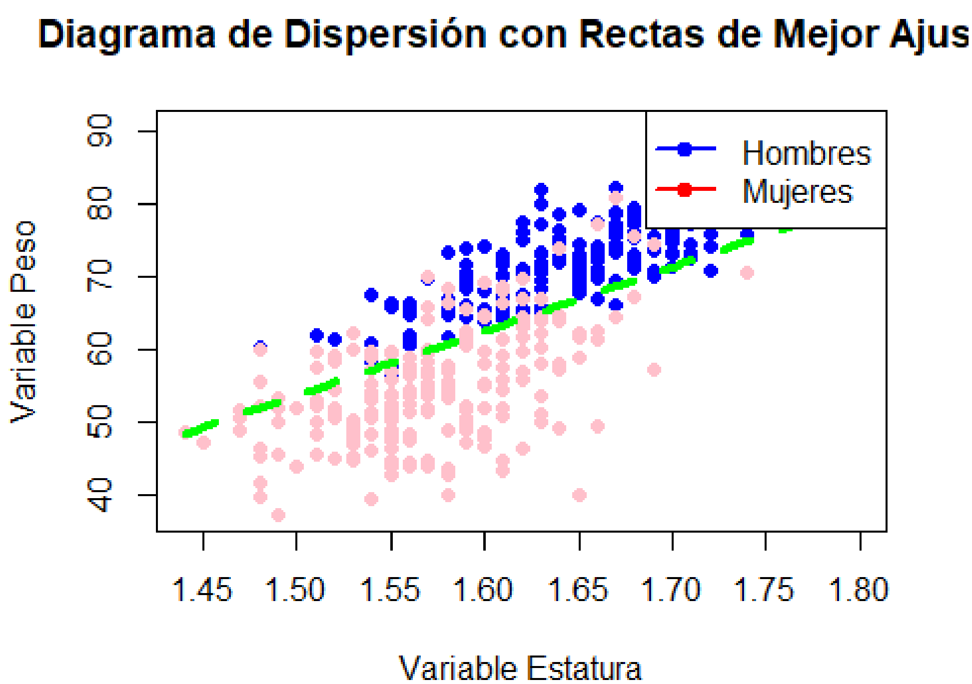
3. Indica cuál es el porcentaje de variación explicada por el modelo

```
# Verificar el porcentaje de variación explicada (R2) para modelo general
r_squared_general <- resumen_general$r.squared
cat("\n*** Porcentaje de Variación Explicada - Modelo general ***\n")

##
## *** Porcentaje de Variación Explicada - Modelo general ***

cat("R² - Modelo general:", r_squared_general, "\n")

## R² - Modelo general: 0.7847011

cat("El modelo explica el", round(r_squared_general * 100, 2), "% de la
variación de la variable dependiente.\n")

## El modelo explica el 78.47 % de la variación de la variable
dependiente.
```

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

```
# Dibujar el diagrama de dispersión para ambos grupos
plot(dataH$Estatura, dataH$Peso,
     main = "Diagrama de Dispersión con Rectas de Mejor Ajuste",
     xlab = "Variable Estatura",
     ylab = "Variable Peso",
     pch = 19,                 # Tipo de punto
     col = "blue",             # Color de los puntos para hombres
     xlim = range(c(dataH$Estatura, dataM$Estatura)),  # Ajuste del rango
del eje x
     ylim = range(c(dataH$Peso, dataM$Peso)))  # Ajuste del rango del eje
y

# Añadir los puntos de dispersión para mujeres
points(dataM$Estatura, dataM$Peso,
       pch = 19,
       col = "pink")         # Color de los puntos para mujeres

# Predecir valores del modelo general y agregar la línea
estatura_vals <- seq(min(datos$Estatura), max(datos$Estatura), length.out
= 100)
sexo_numeric_vals <- rep(mean(datos$Sexo_numeric), 100)  # Media para
representar el efecto combinado
peso_pred_general <- predict(modelo_general, newdata =
```

```
data.frame(Estatura = estatura_vals, Sexo_numeric = sexo_numeric_vals))

# Dibujar la recta de mejor ajuste del modelo general
lines(estatura_vals, peso_pred_general, col = "green", lwd = 4, lty = 2)

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



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

El análisis de la significancia del modelo general revela que el valor p de la prueba F es extremadamente pequeño (5.93158×10^ −145), indicando que el modelo es significativamente mejor que uno sin predictores, es decir, un modelo que solo incluye el intercepto. Esto sugiere que las variables consideradas en el modelo (estatura, sexo y la interacción entre ambos) son relevantes en conjunto para explicar la variación en la variable dependiente. En el contexto del problema, esto implica que las variables seleccionadas aportan información valiosa para la predicción o explicación del fenómeno estudiado.

Al analizar la significancia de los coeficientes, se observa que el intercepto es significativo, lo que implica que hay un valor base de la variable dependiente incluso en ausencia de otros efectos. Esto refleja un punto de partida que es importante en el contexto del modelo. La estatura, por otro lado, se muestra como un predictor altamente significativo ($p<0.03$), lo que indica que tiene un impacto notable en la variable dependiente. Esto sugiere que la estatura es un factor clave que influye considerablemente en el fenómeno que se está modelando.

En contraste, el coeficiente asociado a la variable sexo, codificado numéricamente, no resulta significativo, lo que sugiere que esta variable no tiene un efecto importante en la variable dependiente en este contexto. Esto indica que el sexo, por sí solo, no es un buen predictor en el modelo. Asimismo, la interacción entre estatura y sexo tampoco es significativa, lo que implica que el efecto de la estatura sobre la variable dependiente no varía de manera sustancial entre diferentes sexos. En otras palabras, la combinación de estatura y sexo no proporciona información adicional significativa más allá de los efectos individuales.

El análisis del porcentaje de variación explicada (R^2) muestra que el modelo general explica el 78.47% de la variación de la variable dependiente, lo cual indica un buen nivel de ajuste. Esto significa que el modelo es capaz de capturar la mayor parte de la variabilidad observada en los datos, lo cual refuerza la importancia de las variables incluidas, particularmente la estatura. Sin embargo, todavía hay un 21.53% de la variabilidad que no es explicada, lo que podría deberse a factores no considerados en el modelo o a variaciones aleatorias inherentes a los datos.

3. Interpreta en el contexto del problema:

1. ¿Qué información proporciona $\hat{\beta}_0$ sobre la relación entre la estatura y el peso de hombres y mujeres? Interpreta y compara entre este modelo con los 3 modelos anteriores.

- **Hombres:** El valor de $\hat{\beta}_0$ para los hombres es -83.68, lo que indica que si la Estatura fuera hipotéticamente cero, el modelo predeciría un peso de -83.68 kg, lo cual no tiene un sentido práctico. Sin embargo, en el contexto del modelo, este intercepto sirve principalmente como un ajuste vertical para la línea de regresión. Su valor negativo refleja que la recta está ajustada de manera que, a medida que la Estatura aumenta, el peso aumenta rápidamente, compensando este intercepto bajo.

- **Mujeres:** El $\hat{\beta}_0$ para las mujeres es -72.56, similarmente indicando un peso negativo no realista si la estatura fuera cero. Como en el caso de los hombres, el valor del intercepto no se interpreta literalmente sino que ajusta la recta para capturar la tendencia del peso en función de la estatura en el grupo de mujeres.

- **Modelo General:** El intercepto de -74.75 en el modelo general ajusta la línea combinando la influencia de Estatura y Sexo_numeric. Aunque no tiene un valor directo práctico para interpretarse de manera aislada, refleja cómo se posiciona la recta inicial en la gráfica de los datos combinados, capturando un punto de partida antes de ajustar por estatura y sexo.

- **Comparación del 4to modelo con interacción con los 3 modelos anteriores:** Al comparar los interceptos de los cuatro modelos, se observa que el cuarto modelo se alinea más con el intercepto del modelo masculino con el mismo valor de -83.68. Esto podría indicar que, en el cuarto modelo, la influencia del sexo no está ajustando suficientemente para reflejar la contribución diferenciada de las mujeres como lo observamos en la prueba del valor p con la significancia de los coeficientes.

2. ¿Cómo interpretas beta_i en la relación entre la estatura y el peso de hombres y mujeres? Interpreta y compara entre este modelo con los 3 modelos anteriores.

- **Hombres:** $\hat{\beta}1$ es 94.66, lo que significa que por cada metro adicional de estatura, se espera que el peso de los hombres aumente en aproximadamente 94.66 kg. Este coeficiente positivo y significativo confirma una fuerte y directa relación entre estatura y peso en hombres, mostrando que la estatura es un predictor poderoso del peso en este grupo.

- **Mujeres:** En mujeres, $\hat{\beta}1$ es 81.15, lo que indica que por cada metro adicional de estatura, se espera que el peso aumente en 81.15 kg. Aunque también positivo y significativo, este valor es menor que el de los hombres, lo que sugiere que la relación entre estatura y peso es más débil en mujeres, alineándose con el menor R^2 del modelo femenino.

- **Modelo General:** En el modelo general, $\hat{\beta}1$ es 89.26, lo que significa que la estatura sigue siendo un predictor positivo y fuerte del peso cuando se consideran ambos sexos conjuntamente. Este valor refleja una combinación ponderada de la relación entre estatura y peso en hombres y mujeres, mostrando que la estatura sigue siendo una variable relevante para explicar el peso en la población completa.

- **Comparación del 4to modelo con interacción con los 3 modelos anteriores:** El coeficiente Beta1 = 94.66 en el cuarto modelo indica que por cada metro adicional de estatura, se espera que el peso aumente en aproximadamente 94.66 kg, reflejando una fuerte relación entre estatura y peso similar a la observada en el modelo de hombres. Comparado con los modelos anteriores, donde Beta1 es 94.66 para hombres, 81.15 para mujeres y 89.26 en el modelo general, se evidencia que el cuarto modelo está más alineado con la relación observada en hombres. Esto sugiere que la influencia de la estatura es mayor en hombres y que el cuarto modelo puede estar dominado por esta relación, mientras que el modelo para mujeres muestra una relación más débil y el modelo general intenta capturar un punto intermedio, equilibrando las diferencias entre ambos sexos.

3. Indica cuál(es) de los modelos probados para la relación entre peso y estatura entre hombres y mujeres consideras que es más apropiado y explica por qué.

El mejor modelo es el tercer modelo o llamado modelo general el cuál es el que tiene ambas variables pero sin interacción a diferencia del 4to modelo, su R^2 es mayor al modelo de solamente hombres y mujeres y es casi idéntico al modelo con interacción sin embargo, el modelo sin interacción tiene todos sus betas o coeficientes de manera significativa mientras que el modelo con interacción el único coeficiente significativo es el de ‘Estatura’. Por ello el mejor modelo es el modelo sin inetracción entre Estatura y Sexo.