Act7 Regresion Lineal

Luis Maximiliano López Ramírez

2024-08-30

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

```
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")
data1 = data.frame(dataH$Estatura, dataH$Peso, dataM$Estatura,
dataM$Peso)

# Calcular la matriz de correlación
matriz_correlacion <- cor(data1, use = "complete.obs")

# Mostrar la matriz de correlación
print(matriz_correlacion)

##                 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
```

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

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

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("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
m
```

```
##             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
```

3. Encuentra la ecuación de regresión de mejor ajuste:

1. Realiza la regresión entre las variables involucradas

Hombres

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

modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)

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

2. Verifica el modelo

1. Verifica la significancia del modelo con un alfa de 0.03.

```r
# Resumen del modelo para hombres
resumen_hombres <- summary(modelo_hombres)

# Resumen del modelo para mujeres
resumen_mujeres <- summary(modelo_mujeres)

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

# Verificar la significancia del modelo (Prueba F global) para hombres
p_valor_f_hombres <- pf(resumen_hombres$fstatistic[1],
resumen_hombres$fstatistic[2], resumen_hombres$fstatistic[3], lower.tail
= FALSE)
cat("\n*** Significancia del Modelo - Hombres ***\n")

##
## *** Significancia del Modelo - Hombres ***

cat("Valor p del modelo (Prueba F) - Hombres:", p_valor_f_hombres, "\n")

## Valor p del modelo (Prueba F) - Hombres: 1.063532e-61

if (p_valor_f_hombres < 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
```

```r
# Verificar la significancia del modelo (Prueba F global) para mujeres
p_valor_f_mujeres <- pf(resumen_mujeres$fstatistic[1],
resumen_mujeres$fstatistic[2], resumen_mujeres$fstatistic[3], lower.tail
= FALSE)
cat("\n*** Significancia del Modelo - Mujeres ***\n")
```

```
##
## *** Significancia del Modelo - Mujeres ***
```

```r
cat("Valor p del modelo (Prueba F) - Mujeres:", p_valor_f_mujeres, "\n")
```

```
## Valor p del modelo (Prueba F) - Mujeres: 5.997517e-17
```

```r
if (p_valor_f_mujeres < 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
```

```r
# 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 - Mujeres ***\n")
```

```
##
## *** Significancia del Modelo - Mujeres ***
```

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

```
## Valor p del modelo (Prueba F) - Mujeres: 5.338707e-146
```

```r
if (p_valor_f_mujeres < 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. Verifica la significancia de beta_i con un alfa de 0.03.

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

```
##
## *** Significancia de los Coeficientes - Hombres ***
```

```r
coeficientes_hombres <- resumen_hombres$coefficients
print(coeficientes_hombres)
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -83.68454   6.663457 -12.55873 1.410453e-27
## Estatura     94.66024   4.026565  23.50893 1.063532e-61
```

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

```
## (Intercept) - Valor p: 1.410453e-27 -> Significativo
## Estatura - Valor p: 1.063532e-61 -> Significativo
```

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

```
##
## *** Significancia de los Coeficientes - Mujeres ***
```

```r
coeficientes_mujeres <- resumen_mujeres$coefficients
print(coeficientes_mujeres)
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -72.56045  14.040773 -5.167839 5.342762e-07
## Estatura     81.14911   8.921817  9.095581 5.997517e-17
```

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

```
## (Intercept) - Valor p: 5.342762e-07 -> Significativo
## Estatura - Valor p: 5.997517e-17 -> Significativo
```

```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 ***
```

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

##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)  -74.75460  7.5555310  -9.894023 5.821123e-21
## Estatura      89.26035  4.5635201  19.559540 1.179664e-61
## Sexo_numeric -10.56447  0.6317092 -16.723630 6.932573e-49

# 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: 5.821123e-21 -> Significativo
## Estatura - Valor p: 1.179664e-61 -> Significativo
## Sexo_numeric - Valor p: 6.932573e-49 -> Significativo
```

3. Verifica el porcentaje de variación explicada por el modelo

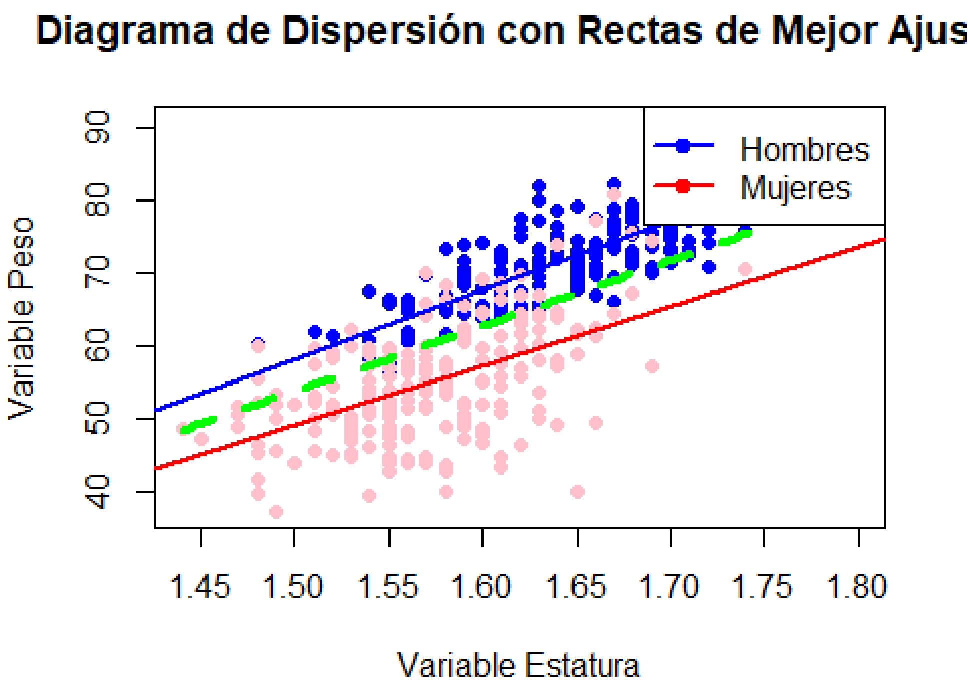
```
# Verificar el porcentaje de variación explicada (R2) para hombres
r_squared_hombres <- resumen_hombres$r.squared
cat("\n*** Porcentaje de Variación Explicada - Hombres ***\n")

##
## *** Porcentaje de Variación Explicada - Hombres ***

cat("R² - Hombres:", r_squared_hombres, "\n")

## R² - Hombres: 0.7171292

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

## El modelo explica el 71.71 % de la variación de la variable
dependiente.

# Verificar el porcentaje de variación explicada (R2) para mujeres
r_squared_mujeres <- resumen_mujeres$r.squared
cat("\n*** Porcentaje de Variación Explicada - Mujeres ***\n")

##
## *** Porcentaje de Variación Explicada - Mujeres ***

cat("R² - Mujeres:", r_squared_mujeres, "\n")

## R² - Mujeres: 0.2750963
```

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

## El modelo explica el 27.51 % de la variación de la variable dependiente.

# 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: 0.7836599

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

## El modelo explica el 78.37 % de la variación de la variable dependiente.
```

4. 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

# Añadir la recta de mejor ajuste para hombres
abline(modelo_hombres, col = "blue", lwd = 2)

# Añadir la recta de mejor ajuste para mujeres
abline(modelo_mujeres, col = "red", lwd = 2)

# 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
```

```r
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)
```



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

El análisis de los modelos de regresión y la gráfica revela diferencias significativas en cómo la Estatura predice el Peso para hombres y mujeres. El modelo para hombres tiene un R^2 de 0.7171, lo que indica que la Estatura explica el 71.71 % de la variación del Peso en los hombres. Esto se refleja en la gráfica, donde la recta azul de mejor ajuste muestra una pendiente pronunciada, indicando una fuerte relación positiva entre estatura y peso. En contraste, el modelo para mujeres tiene un R^2 de solo 0.2751, lo que significa que la estatura explica solo el 27.51 % de la variación del peso en las mujeres. En la gráfica, la recta roja tiene una pendiente más suave, evidenciando una relación más débil entre estas variables, lo que sugiere que otros factores no capturados por el modelo influyen significativamente en el peso de las mujeres.

El modelo general, que incluye tanto Estatura como Sexo_numeric, presenta un R^2 de 0.7837, explicando el 78.37 % de la variación en el peso. Esto muestra que incorporar el sexo como variable explicativa mejora sustancialmente el ajuste del modelo, capturando las diferencias entre hombres y mujeres en la relación entre estatura y peso. En la gráfica, la recta verde discontinua del modelo general se sitúa entre las rectas de los modelos específicos para cada sexo, reflejando un ajuste más balanceado. Esto destaca la importancia de considerar el género en el modelo, ya que permite una predicción más precisa del peso al tener en cuenta cómo varía la influencia de la estatura entre hombres y mujeres. En resumen, el R^2 y la gráfica subrayan que la estatura es un predictor sólido del peso en hombres, menos efectivo en mujeres, y que la inclusión del sexo en el modelo mejora significativamente la capacidad de predicción para ambos grupos.

 # 6. Interpreta en el contexto del problema:

1. ¿Qué información proporciona Beta0 sobre la relación entre la estatura y el peso de hombres y mujeres?

- **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.

2. ¿Cómo interpretas Beta1 en la relación entre la estatura y el peso de hombres y mujeres?

- **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.

Interpretación de $\hat{\beta}$ para Sexo_numeric en el Modelo General:

El coeficiente de Sexo_numeric es -10.56, indicando que, al ajustar por estatura, las mujeres (sexo numérico = 1) pesan en promedio 10.56 kg menos que los hombres (sexo numérico = 0). Este coeficiente significativo resalta la diferencia inherente entre los dos géneros, capturando el impacto del sexo en la relación entre estatura y peso.