Tarea 11 - Regresión Lineal con Interacción

Héctor Hibran Tapia Fernández - A01661114

2024-09-03

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

- Revisar entrega "Tarea 10 - Regresión Lineal".

```
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 tenine una relación (lineal) negativa.

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

```
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
```

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:

Hipótesis:

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

```
modelo_con_interaccion = lm(Peso ~ Estatura * Sexo, data = df)
```

2.1 Obtén el modelo e interpreta las variables Dummy:

```
resumen_modelo = summary(modelo_con_interaccion)
print(resumen_modelo)
```

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

Fórmula

Estos son los resultados de nuestro modelo de regresión lineal múltiple, esto analiza la relación entre el peso (variable dependiente/respuesta) y las dos variables predictoras, estatura y sexo. (variables independientes).

Al poner “Estatura * Sexo” se consideran ambos por separado y la interacción entre estos.

Residuos

Es la diferencia entre los valores observados vs los valores predichos por el modelo, la más significativa sería la mediana, en nuestro caso tenemos que es de “0.0204”. Una mediana cercana a 0 nos dice que el modelo no tiene un sesgo en las predicciones.

Coeficientes

- Intercept: Este es el valor del peso promedio cuando la estatura es 0. *** muy significativo
- Estatura: Cada que incrementa una unidad la estatura, el peso se incrementa en promedio 94.66 unidades. *** muy significativo
- SexoM: Representa la diferencia en el peso entre los hombres y las mujeres cuando la estatura es 0.
- Estatura:SexoM: Interacción entre estatura y sexo. Nos indica cómo cambia la relación entre estatura y peso dependiendo del sexo.

SexoM y Estatura:SexoM no son significativos ($p > 0.05$), lo que significa que no hay evidencia suficiente para decir que estos coeficientes son diferentes de cero.

2.2 Significancia del modelo:

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

```
p_valor_estatura = resumen_modelo$coefficients["Estatura", "Pr(>|t|)"]
p_valor_interaccion = resumen_modelo$coefficients["Estatura:SexoM", "Pr(>|t|)"]

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

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

```
if (p_valor_interaccion < 0.03) {
  cat("Se rechaza la hipótesis nula (H0) para la interacción Estatura:Sexo. Hay evidenci
a suficiente para afirmar que la relación entre Estatura y Peso depende del Sexo.\n")
} else {
  cat("No se rechaza la hipótesis nula (H0) para la interacción Estatura:Sexo. No hay ev
idencia suficiente para afirmar que la relación entre Estatura y Peso depende del Sex
o.\n")
}
```

```
## No se rechaza la hipótesis nula (H0) para la interacción Estatura:Sexo. No hay eviden
cia suficiente para afirmar que la relación entre Estatura y Peso depende del Sexo.
```

- Validala significancia de $\hat{\beta}$i con un alfa de 0.03 (incluye las hipótesis que pruebas)

```
coeficientes = resumen_modelo$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 significativ
o.\n\n")
  } else {
    cat("  Resultado: No se rechaza la hipótesis nula (H0). El coeficiente no es signifi
cativo.\n\n")
  }
}
```

```
## Coeficiente: (Intercept)
##   Valor del coeficiente: -83.6845
##   Error estándar: 9.7346
##   Valor t: -8.5966
##   Valor p: 0
##   Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.
##
## Coeficiente: Estatura
##   Valor del coeficiente: 94.6602
##   Error estándar: 5.8824
##   Valor t: 16.0922
##   Valor p: 0
##   Resultado: Se rechaza la hipótesis nula (H0). El coeficiente es significativo.
##
## Coeficiente: SexoM
##   Valor del coeficiente: 11.1241
##   Error estándar: 14.9497
##   Valor t: 0.7441
##   Valor p: 0.4572
##   Resultado: No se rechaza la hipótesis nula (H0). El coeficiente no es significativ
o.
##
## Coeficiente: Estatura:SexoM
##   Valor del coeficiente: -13.5111
##   Error estándar: 9.3048
##   Valor t: -1.4521
##   Valor p: 0.1472
##   Resultado: No se rechaza la hipótesis nula (H0). El coeficiente no es significativ
o.
```

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

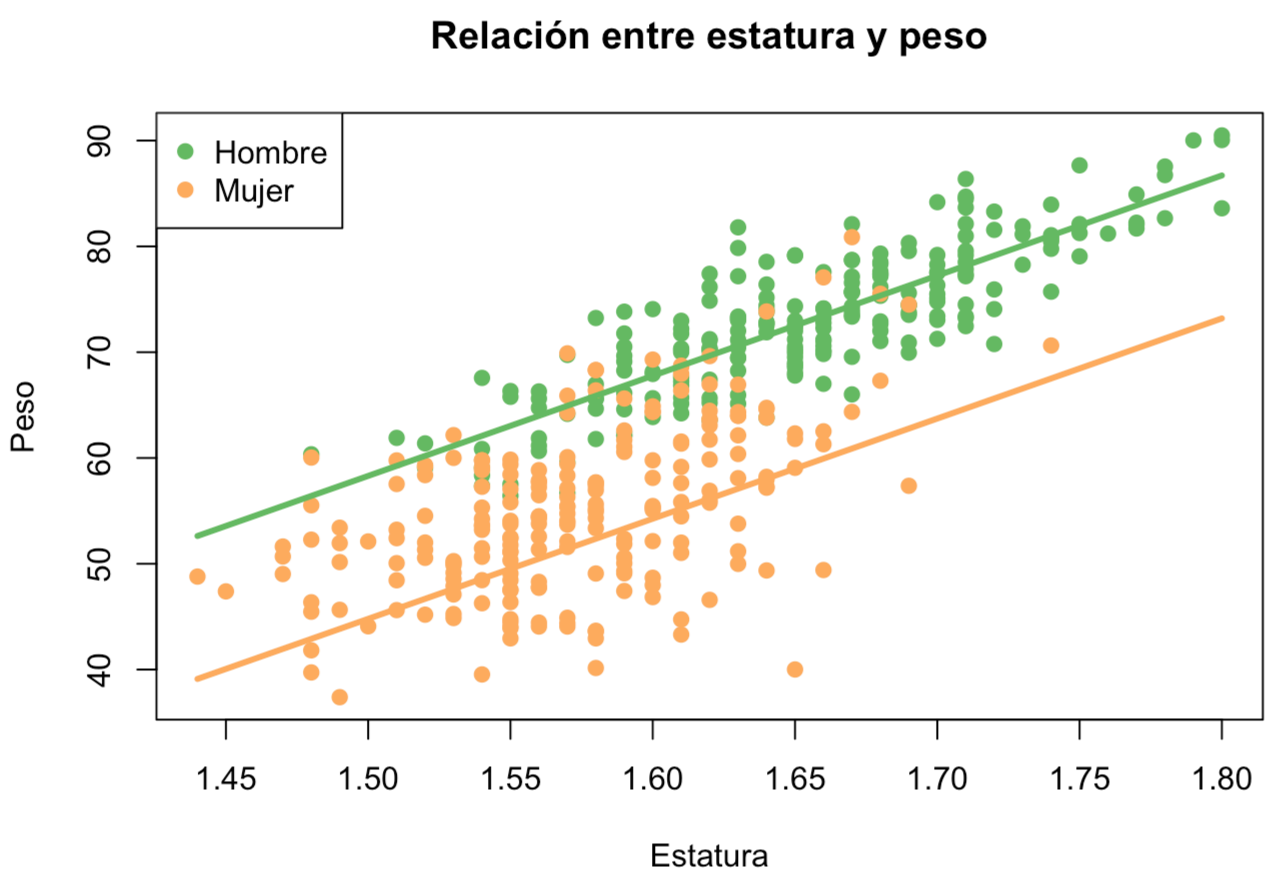
```
r2 = resumen_modelo$r.squared
porcentaje_variacion_explicada <- r2 * 100
cat("El porcentaje de variación explicada por el modelo es:", round(porcentaje_variacion
_explicada, 2), "%\n")
```

```
## El porcentaje de variación explicada por el modelo es: 78.47 %
```

2.3 Dibuja el diagrama de dispersión de los datos y la recta de

mejor ajuste:

```
b0_A = modelo_con_interaccion$coefficients[1]  # Intercepto
b1_A = modelo_con_interaccion$coefficients[2]  # Coeficiente para Estatura
b2_A = modelo_con_interaccion$coefficients[4]  # Coeficiente para la interacción Estatur
a:SexoM

Ym = function(x) { b0_A + b2_A + b1_A * x }
Yh = function(x) { b0_A + b1_A * x }

colores = c("#66BD63", "#FDAE61")

plot(df$Estatura, df$Peso, col=colores[factor(df$Sexo)], pch=19, ylab="Peso", xlab="Esta
tura", main="Relación entre estatura y peso")

x = seq(min(df$Estatura), max(df$Estatura), length.out=100)

lines(x, Ym(x), col="#FDAE61", lwd=3)
lines(x, Yh(x), col="#66BD63", lwd=3)

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



2.4 Interpreta en el contexto del problema cada uno de los

análisis que hiciste:

Vale con este modelo de regresión lineal estamos tratando de predecir el peso de una persona usando su estatura y su sexo. Incluímos una interacción que es la de Estatura y Sexo, lo que significa que el efecto de la Estatura sobre el Peso puede variar según el sexo de la persona.

Lo que más nos interesa en este caso es el coeficiente de interacción "(Estatura:SexoM):-13.5111":

Como se mencionó anteriormente este coeficiente mide el efecto que tiene la combinación de Estatura y Sexo. En este caso, el coeficiente es negativo, lo que nos dice que el efecto de la estatura sobre el peso es menor para los hombres en comparación con las mujeres, además el valor p = 0.1472, **lo que significa que no hay evidencia suficiente para concluir que la relación entre estatura y peso difiere según el sexo.**

3. Interpreta en el contexto del problema

- ¿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.

Valor del coeficiente: -83.6845

$\hat{\beta}0$ representa el peso estimado para una persona cuando la estatura es 0 y para el sexo que se considera como categoría de referencia en la variable dummy esto dependerá de como se codificó cada sexo.

En sí no nos dá información específica, nos da un punto de referencia nada más. Otros coeficientes nos dan más información.

- ¿Cómo interpretas $\hat{\beta}i$ en la relación entre la estatura y el peso de hombres y mujeres? Interpreta y compara entre este modelo con los 3 modelos anteriores.

Las diferentes B_is tienen diferentes significados:

- Valor del coeficiente: 94.6602 (Coeficiente: Estatura) - indica que por cada unidad de incremento en la estatura, el peso de las mujeres aumenta en promedio 94 unidades.
- Valor del coeficiente: 11.1241 (Coeficiente: SexoM) - indica que en promedio los hombres pesan 11.1241 unidades más que las mujeres cuando la estatura es 0
- Valor del coeficiente: -13.5111 (Coeficiente: Estatura:SexoM) - indica que el efecto de la estatura sobre el peso es 13.5111 unidades menor para los hombres en comparación con las mujeres.
- 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.

Nos quedamos con el modelo con interacción, o combinado ya que tiene una R2 de '0.7832' este explica mejor la variablidad de en los datos.