Regresión múltiple y otras técnicas multivariadas

Tarea 08

Rivera Torres Francisco de Jesús
Rodríguez Maya Jorge Daniel
Samayoa Donado Víctor Augusto
Trujillo Barrios Georgina

Abril 10, 2019

Ejercicio 1

Suponer que se tienen n observaciones de tres variables x, y y z.

$$x_1, \ldots, x_n \qquad\qquad y_1, \ldots, y_n, \qquad\qquad z_1, \ldots, z_n$$

Mostrar que

$$r_{xy,z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1 - r_{xz}^2}, \sqrt{1 - r_{yz}^2}}$$

donde $r_{xy,z}$ denota el coeficiente de correlación parcial de x y y dado z y $r_{\bullet\bullet}$ denota el coeficiente de correlación de Pearson entre dos variables.

Ejercicio 2

Con el conjunto de datos `mtcars` de `R`, ajustar un modelo RLM para explicar la distribución del rendimiento (`mpg`) como función del peso (`wt`) y del número de cilindros (`cyl`).

Inciso 2.a

Reportar las estimaciones puntuales de los coeficientes del modelo e interpretar los resultados.

Considerando el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

donde y_i hace referencia rendimiento del auto i-ésimo (`mpg`), x_{1i} hace referencia al peso del auto i-ésimo (`wt`), y x_{2i} hace referencia al número de cilindros del auto i-ésimo (`cyl`):

```r
# Se genera el mdoelo de RLM
modelo <- mtcars %>%
          lm(formula = mpg ~ wt + cyl)

# Se extraen los coeficientes de regresión
coeficientes <- coefficients(modelo)

# Se imprime el resultado
summary(modelo)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2893 -1.5512 -0.4684  1.5743  6.1004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.6863     1.7150  23.141  < 2e-16 ***
## wt           -3.1910     0.7569  -4.216 0.000222 ***
## cyl          -1.5078     0.4147  -3.636 0.001064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 on 29 degrees of freedom
## Multiple R-squared:  0.8302, Adjusted R-squared:  0.8185
## F-statistic: 70.91 on 2 and 29 DF,  p-value: 6.809e-12
```

De lo anterior se concluye que los coeficientes son:

$$\hat{\beta}_0 = 39.6862615 \qquad \hat{\beta}_1 = -3.1909721 \qquad \hat{\beta}_2 = -1.507795$$

El coeficiente β_0 no tiene un interpretación en este contexto ya que no existen autos que tengan un peso de 0 libras y 0 cilindros.

El coeficinete β_1 se interpreta como el decremento del rendimiento cuando se incrementa el peso del auto en 1 libra, mientras el número de cilindros de mantiene constante. Por tal motivo, el incremento de una libra en el peso del auto implica un decremento de -3.19 unidades en la métrica del rendimiento.

El coeficinete β_2 se interpreta como el decremento del rendimiento cuando se incrementa el el número de cilindros en 1, mientras el peso del auto se mantiene constante. Por tal motivo, el incremento de cilindros en una unidad implica un decremento de -1.51 unidades en la métrica del rendimiento.

Inciso 2.b

Contrastar las hipótesis de significancia del efecto del peso. Interpretar el resultado en el contexto de los datos.

El contraste de hipótesis está dado por:

$$H_0 : \beta_1 = 0 \qquad \text{vs} \qquad H_1 : \beta_1 \neq 0$$

De la tabla mostrada en el inciso anterior, se observa que, con un nivel de significancia de $\alpha = 0.05$, $|T_{\beta_1}| = 4.2158076 > 2.0452296 = t_{29}^{1-0.05/2}$. Por lo tanto, existe evidencia para rechazar la hipótesis nula y afirmar que el peso del auto puede tener un efecto sobr el rendimiento del auto.

Inciso 2.c

Contrastar las hipótesis de significancia del efecto del número de cilindros. Interpretar el resultado en el contexto de los datos.

El contraste de hipótesis está dado por:

$$H_0 : \beta_2 = 0 \qquad \text{vs} \qquad H_1 : \beta_2 \neq 0$$

De la tabla mostrada en el inciso 2.a), se observa que, con un nivel de significancia de $\alpha = 0.05$, $|T_{\beta_2}| = 3.6359719 > 2.0452296 = t_{29}^{1-0.05/2}$. Por lo tanto, existe evidencia para rechazar la hipótesis nula y afirmar que el número de cilindros puede tener un efecto sobr el rendimiento del auto.

Inciso 2.d

Estimar puntualmente y por intervalo de confianza 90% del rendimiento medio de un auto que pesa 2500 libras y tiene 3 cilindros.

```r
x0 <- tibble(wt = 2500,
             cyl = 3)

predict(modelo, newdata = x0, interval = "confidence", level = 0.9)
```

```
##         fit       lwr       upr
## 1 -7942.267 -11155.09 -4729.447
```

Inciso 2.e

Calcular un intervalo de predicción del rendimiento de un auto que pesa 2500 libras y tiene 3 cilindros.

```
predict(modelo, newdata = x0, interval = "prediction", level = 0.9)
```

```
##         fit       lwr       upr
## 1 -7942.267 -11155.09 -4729.444
```

Inciso 2.f

Escribir las tres rectas de regresión ajustadas del rendimiento medio como función del peso condicional a 4, 6 y 8 cilindros, respectivamente.

Se busca ajustar modelos

$$\begin{aligned}
\mu_{i|4} &= \mathrm{E}\left(y_i|cyl = 4\right) = \beta_{0|4} + \beta_{1|4}x_{i|4} \\
\mu_{i|6} &= \mathrm{E}\left(y_i|cyl = 6\right) = \beta_{0|6} + \beta_{1|6}x_{i|6} \\
\mu_{i|8} &= \mathrm{E}\left(y_i|cyl = 8\right) = \beta_{0|8} + \beta_{1|8}x_{i|8}
\end{aligned}$$

donde $\mu_{i|n}$ significa el valor promedio del rendimiento del auto i-ésimo con n número de cilindros, $\beta_{0|n}$ es el coeficiente de intersección del modelo condicionado a n número de cilindros, $\beta_{1|n}$ es el coeficiente del peso del modelo condicionado a n número de cilindros y $x_{i|n}$ es el peso del auto i-ésimo con n número de cilindros.

```
# Se generan los modelos
modelo_cyl4 <- mtcars %>%
               filter(cyl == 4) %>%
               lm(formula = mpg ~ wt)

modelo_cyl6 <- mtcars %>%
               filter(cyl == 6) %>%
               lm(formula = mpg ~ wt)

modelo_cyl8 <- mtcars %>%
               filter(cyl == 8) %>%
               lm(formula = mpg ~ wt)

# Se extraen los coeficientes de regresión
coeff_cyl4 <- coefficients(modelo_cyl4)
coeff_cyl6 <- coefficients(modelo_cyl6)
coeff_cyl8 <- coefficients(modelo_cyl8)
```

Por lo tanto los modelos ajustados son:

$$\begin{aligned}
\mu_{i|4} &= 39.57 - 5.65x_{i|4}, && \text{condicional a 4 cilindros} \\
\mu_{i|6} &= 28.41 - 2.78x_{i|6}, && \text{condicional a 6 cilindros} \\
\mu_{i|8} &= 23.87 - 2.19x_{i|8}, && \text{condicional a 8 cilindros}
\end{aligned}$$

Inciso 2.g

Graficar un diagrama de dispersión de `mpg` contra `wt` y sobreponer las tres rectas de regresión ajustadas. Colorear los puntos y las rectas según el número de cilindros y agregar la leyenda correspondiente.

```
mtcars %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = wt, y = mpg, color = cyl)) +
  geom_smooth(method = "lm", formula = y ~ x, fullrange = TRUE,
              se = FALSE) +
  geom_point()
```