

# Regresión múltiple y otras técnicas multivariadas

Tarea 05

*Rivera Torres Francisco de Jesús*

*Rodríguez Maya Jorge Daniel*

*Samayoa Donado Víctor Augusto*

*Trujillo Bariños Georgina*

*Marzo 13, 2019*

## Ejercicio 1

En un estudio que consistió en explorar la relación entre la longitud de la mandíbula (en milímetros) y la edad gestacional (en semanas) de 158 fetos, se obtuvieron los siguientes resultados:

$$\bar{x} = 20.1, \quad \bar{y} = 23.6, \quad S_{xx} = 2473.8, \quad S_{yy} = 8652.4, \quad S_{xy} = 4385.4$$

Responder lo siguiente:

### Inciso 1.a

Ajustar un modelo RLS para explicar la distribución de la longitud de la mandíbula de los fetos como función de la edad gestacional. Reportar las estimaciones de los parámetros. Interpretar los resultados en el contexto de los datos.

Sabemos que los estimadores de los coeficientes del modelo RLS, se pueden obtener de la siguiente manera:

$$\hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x} \qquad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

por lo tanto, se tiene que:

$$\hat{\beta}_0 = 23.6 - \frac{4385.4}{2473.8} 20.1 = -12.03204 \quad \text{y} \quad \hat{\beta}_1 = \frac{4385.4}{2473.8} = 1.772738$$

teniendo así el modelo RLS,  $\hat{y} = -12.03204 + 1.772738x$ .

De acuerdo con el contexto de los datos, el parámetro  $\hat{\beta}_0$  no tiene una interpretación dentro del problema.

Por otro lado, el estimador del parámetro  $\hat{\beta}_1$  nos indica que por cada semana que transcurre en la gestación del feto, la longitud de la mandíbula crece, en promedio, 1.77 milímetros.

**Inciso 1.b**

Construir la tabla ANOVA y contrastar la significancia del modelo del inciso anterior. Usar un tamaño de prueba  $\alpha = 0.05$ .

Sabemos que, para el modelo RLS, la tabla ANOVA se define como sigue:

Tabla 1: Tabla ANOVA					
FV	GL	SC	CM	F	p-value
Regresión	1	SCR	SCR	$\frac{CMR}{CME}$	$P(F > F_{(1,n-2)})$
Error	n - 2	SCE	$\frac{SCE}{n-2}$		
Total	n - 1	SCT			

donde  $n = 158$  y

$$\begin{aligned}
 SCR &= S_{xx}\hat{\beta}_1^2, & SCE &= \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}, & SCT &= SCR + SCE \\
 SCR &= 2473.8(1.772738)^2, & SCE &= \frac{(2473.8)(8652.4) - (4385.4)^2}{2473.8}, & SCT &= SCR + SCE \\
 SCR &= 7774.164 & SCE &= 878.2335 & SCT &= 8652.397
 \end{aligned}$$

Con estos valores se procede a calcular los campos restantes de la tabla ANOVA:

$$\begin{aligned}
 CMR &= SCR & CME &= \frac{SCE}{n-2} & F &= \frac{CMR}{CME} \\
 CMR &= 7774.164 & CME &= 5.629702 & F &= 1380.919
 \end{aligned}$$

Utilizando un  $\alpha = 0.05$  se tiene que el cuantil de la distribución de referencia es

```
alpha <- 0.05
f.a <- qf(1 - alpha, df1 = 1, df2 = 156)
```

Es decir, el cuantil de la distribución de referencia es 3.9017607. El cual es un valor menor a  $F = 1380.919$ . Esto implica que se debe rechazar  $H_0$  con un nivel de significancia del 0.05. Esto implica que la variable explicativa **edad gestacional** no tiene algún efecto en la distribución de la variable objetivo **longitud de la mandíbula**.

Usando la información anterior, la tabla ANOVA queda como sigue:

Tabla 2: Tabla ANOVA					
FV	GL	SC	CM	F	p-value
Regresión	1	7774.164	7774.164	1380.919	3.9017607
Error	156	878.2335	5.629702		
Total	157	8652.397			

**Inciso 1.c**

Calcular un intervalo de predicción para la media de la longitud de la mandíbula de un feto con 23 semanas de gestación.

Sabemos que la media de la longitud de la mandíbula de un feto con  $x_0$  semanas de gestación está dada por

$$\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad \text{entonces} \quad \hat{\mu}_0 = -12.03204 + 1.772738(23) = 28.74093$$

Además, sabemos que el intervalo de confianza está dado por:

$$\hat{\mu}_0 \pm t_{n-1}^{1-\alpha/2} \hat{\sigma}_{\text{MCO}} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

donde podemos estimar la varianza mediante la igualdad  $\hat{\sigma}_{\text{MCO}} = \frac{SCE}{n-2} = CME$ .

Calculando el estadístico  $t_{n-2}^{1-\alpha/2}$  se obtiene

```
alpha <- 0.05
```

```
n <- 158
```

```
ta <- qt(1 - alpha/2, n - 2)
```

$$t_{n-2}^{1-\alpha/2} = 1.9752875$$

Entonces tenemos que el intervalo de confianza está dado por:

$$\begin{aligned} & \hat{\mu}_0 \pm t_{n-1}^{1-\alpha/2} \hat{\sigma}_{\text{MCO}} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \\ & 28.74093 \pm (1.975288) \sqrt{5.629702} \sqrt{\frac{1}{158} + \frac{(23 - 20.1)^2}{2473.8}} \\ & 28.74093 \pm (1.975288)(2.372699) \sqrt{0.006329114 + 0.003399628} \\ & 28.74093 \pm (1.975288)(2.372699)(0.09863439) \\ & 28.74093 \pm 0.4622761 \end{aligned}$$

Por lo tanto, el intervalo de predicción para la media de la longitud de la mandíbula de un feto con 23 semanas de gestación está dado por:

$$[28.27865, 29.20321]$$

esto es, la media de la longitud de la mandíbula de un feto con 23 semanas de gestación está entre 28.27865 y 29.20321 milímetros con un 95% de confianza.

### Inciso 1.d

Calcular el  $R^2$  del modelo e interpretar el resultado.

Recordemos que

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{878.2335}{8652.397} = 0.8984982$$

Esto nos indica que el ajuste del modelo es adecuado, ya que el modelo logra explicar un 89.84% de la variación total de la longitud de la mandíbula de los fetos.

## Ejercicio 2

En un estudio sobre los 67 condados del estado de Florida (EUA), se obtuvieron los siguientes resultados sobre

Variable	Promedio	Desv. est.
Ingreso	24.51	4.69
Educación	69.49	8.86

Se ajustó un modelo RLS a los datos, para explicar la distribución del ingreso como función de la educación, y se obtuvieron los siguientes resultados:

$$\hat{\beta}_0 = -4.63, \quad \hat{\beta}_1 = 0.42$$

### Inciso 2.a

Construir intervalos simultáneos de confianza 95% para  $\beta_0$  y  $\beta_1$ .

### Inciso 2.b

Construir la tabla ANOVA y contrastar la significancia del modelo ajustado. Utilizar  $\alpha = 0.05$ .

**Inciso 2.c**

Reportar la estimación de  $\sigma$  y calcular un intervalo de confianza 95%.

**Inciso 2.d**

Calcular el  $R^2$  del modelo e interpretar el resultado.