

# Regresión múltiple y otras técnicas multivariadas

Tarea 03

*Rivera Torres Francisco de Jesús*

*Rodríguez Maya Jorge Daniel*

*Samayoa Donado Víctor Augusto*

*Trujillo Barrios Georgina*

*Febrero 27, 2019*

## Ejercicio 1

Suponer que se ajusta un modelo RLS a las observaciones  $(x_i, y_i)$  con  $i = 1, \dots, n$ . Mostrar que

$$SCE = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}$$

Donde:

$$SCE = \sum_i^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$S_{yy} = \sum_i^n (y_i - \bar{y}_n)^2$$

## Ejercicio 2

Mostrar la desigualdad de Bonferroni. Si  $E_1, \dots, E_k$  son eventos en un espacio de probabilidad  $(\omega, A, P)$  entonces

$$Pr(\cap E_j) \geq 1 - Pr(E_j^c)$$

## Ejercicio 3

Considerar los datos de ingreso y escolaridad utilizados en los ejemplos de intervalos de confianza de las notas. Reportar intervalos simultáneos de confianza 95% para las medias del ingreso por hora para 9, 15 y 19 años de escolaridad a) con el método de Bonferroni y b) con el método de Hotelling–Scheffé

x	5	6	7	10	12	15	18	20
y	7.4	9.3	10.6	15.4	18.1	22.2	24.1	24.8

## Ejercicio 4

El conjunto de datos *airquality*, de paquete *datasets* de R contiene información sobre la calidad del aire en Nueva York registrada de Mayo a Septiembre de 1973 (se puede consultar más información con el comando `help(airquality)`). Para responder este ejercicio, descartar las observaciones con valores perdidos.

**a**

Ajustar un modelo RLS para explicar el nivel de ozono como función del  $\log_2$  de la velocidad del viento. Reportar las estimaciones de los parámetros.

**b**

Mostrar una gráfica de dispersión de los datos utilizados para ajustar el modelo del inciso anterior, la recta de regresión ajustada y bandas de confianza 95%. Anexar el código relacionado con el cómputo de las bandas de confianza.

## Ejercicio 5

(Sheater) Un estadístico colaboró en un proyecto de investigación con dos entomólogos. El análisis involucró el ajuste de modelos de regresión a grandes conjuntos de datos. Entre los tres escribieron y sometieron un manuscrito a una revista de entomología. El escrito contenía varias gráficas de dispersión mostrando la recta de regresión ajustada y las bandas de confianza 95% para la verdadera recta de regresión calculadas con los IC individuales, así como los datos observados. Uno de los revisores del manuscrito hizo la siguiente observación:

No puedo entender cómo el 95% de las observaciones cae fuera de las bandas de confianza 95% que se muestran en las figuras

## Ejercicio 6

(Ross) Suponer que se tiene el siguiente conjunto de datos donde  $x$  representa la humedad de una mezcla fresca de un determinado producto y  $y$  la densidad del producto terminado.

Ajustar un modelo RLS a los datos anteriores y responder lo siguiente.

**a**

Reportar la estimación puntual de  $\sigma^2$  e interpretar el resultado en cuanto a la utilidad del modelo RLS ajustado.

**b**

Reportar el IC 90% para  $\sigma^2$  con los cuantiles simétricos y su longitud.

**c**

Indicar cuáles son los cuantiles que proporcionan el IC 90% para  $\sigma^2$  de menor longitud.

**d**

Reportar el IC 90% para  $\sigma^2$  de menor longitud y compararlo con el intervalo del inciso *a*).