

# Modelos Lineales con R

## Práctica 2: Análisis de Regresión Lineal

Profesor: Andrés García Medina  
andres.garcia.medina@uabc.edu.mx

Fecha de entrega: jueves 29 de febrero, 2024 (12pm).

**Instrucciones:** Subir un documento pdf a classroom con las respuestas de cada uno de los ejercicios solicitados. Adjunta el código fuente en formato `.r` o `.ipynb`. Justificar detalladamente cada una de sus respuestas.

### Ejercicio 1

Considere nuevamente el modelo del ejercicio 2 de la practica 1 dado por:

$$dist_i = \beta_0 + \beta_1 speed_i + \beta_2 speed_i^2 + \epsilon \quad (1)$$

- (a) Usar el metodo `pairs` para graficar los datos e interpretar.
- (b) Graficar la matriz de covarianza de la matriz de diseño asociada  $X$  como un mapa de calor con la función `heatmap`.
- (c) Estimar el vector  $\beta$  del modelo via la factorización QR y comparar con resultado de la función `lm`. Se puede apoyar de la función `qr`.
- (d) Historicamente la solución óptima del vector  $\beta$  está dada por la expresión:  
$$\beta = (X^T X)^{-1} X^T y.$$
  
Estimar el vector  $\beta$  via la expresion anterior (método clásico) y comparar con los resultados obtenidos via `qr` y la función `lm`.
- (e) Grafique los residuales del modelo ajustado ¿Cumplen con los supuestos distribucionales? ¿Con esta información como estimaría la desviación estándar de cada elemento de  $\beta$ ?

### Ejercicio 2

Las 10 corporaciones industriales más grandes de Estados Unidos arrojan los datos mostrados en la figura1 Resuelva los incisos (a),(b), (c), y (d) del ejercicio anterior con estos nuevos datos. En este caso considere *profits* como la variable dependiente y *sales* y *assets* como la variable independiente. Los datos se anexan a la practica con el nombre de `datos.txt`

The World's 10 Largest Companies <sup>1</sup>			
Company	$x_1$ = sales (billions)	$x_2$ = profits (billions)	$x_3$ = assets (billions)
Citigroup	108.28	17.05	1,484.10
General Electric	152.36	16.59	750.33
American Intl Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1,110.46
HSBC Group	62.97	9.52	1,031.29
ExxonMobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1,175.16
Toyota Motor	165.68	11.13	211.15

<sup>1</sup>From [www.Forbes.com](http://www.Forbes.com) partially based on *Forbes* The Forbes Global 2000, April 18, 2005.

Figure 1: Datos de ejercicio 2

## Ejercicio 3

Regresando al ejercicio 1, calcule lo siguiente:

- Para cada  $\hat{\beta}$  estime  $\hat{\sigma}_{\hat{\beta}}$ . Compare su resultado que el que arroja la función `lm`
- Para cada  $\hat{\beta}$  estime el estadístico t y el p-value asociado bajo la hipótesis nula  $H_0 : \beta_i = 0$ . Compare su resultado que el que arroja la función `lm`. ¿Cuales regresores son estadísticamente significativos?
- Estime los coeficientes  $\beta$  vía la factorización de Cholesky. Se puede apoyar de la función `chol`. Compare su resultado que el que arroja la función `lm`.
- Estime el error estándar de los residuales. Compare su resultado que el que arroja la función `lm`.
- Estime el estadístico F y el p-value asociado. Considere como modelo nulo:  $dist_i = \beta_0 + \epsilon$ . Compare su resultado que el que arroja la función `lm`. ¿Su modelo es estadísticamente significativo?
- Calcule las métricas conocidas como  $R^2$  y  $R^2$  ajustado. Las expresiones matemáticas las puede encontrar en la solución de la práctica 1. Compare su resultado que el que arroja la función `lm`.
- Grafique en una sola figura lo siguiente:
  - residuales vs. valores ajustados
  - valores absolutos de los residuales estandarizados vs. valores ajustados

Comparar sus resultados con lo que arroja el comando `plot` usando como argumento la salida del modelo ajustado. Hint: Se requiere calcular la matriz de influencia  $A$ .