

1. Responda las siguientes preguntas.
 - a) Suponga que se realiza escalamiento de longitud unitaria en las predictoras pero no en la variable respuesta, ¿qué unidades tienen los coeficientes de la regresión una vez esta es ajustada?
 - b) ¿Por qué hay problemas de multicolinealidad cuando se tienen más covariables que observaciones en los datos?
 - c) Si la traza de la matriz $\mathbf{X}'\mathbf{X}$ es muy grande, ¿mayor es la distancia entre el vector de parámetros estimados y el verdadero vector de parámetros?
 - d) Si la correlación entre las variables X_j y X_k es pequeña, ¿se puede descartar la presencia de multicolinealidad?
2. Se genera un modelo de regresión lineal múltiple $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ con vector de parámetros $\Theta' = (\beta_0 = -3, \beta_1 = 2, \beta_2 = -4, \sigma^2 = 4)$. Cree dos bases de datos usando las siguientes instrucciones.

```
gen_dat <- function(n) {  
  x1 <- runif(n=n, min=0, max=10)  
  x2 <- x1 * 2 + rnorm(n=n, sd=0.01) # x2 es el doble de x1 + ruido  
  y <- rnorm(n=n, mean= - 3 + 2 * x1 - 4 * x2, sd=2)  
  data.frame(y, x1, x2)  
}  
set.seed(12345)  
datos <- gen_dat(n=40)  
datos1 <- datos[1:20, ]  
datos2 <- datos[21:40, ]
```

Luego de ajustar el modelo, obtenga los coeficientes estimados y compárelos con los reales, ¿qué sucede? Además, calcule los VIF y haga análisis del espectro de la matriz $\mathbf{X}'\mathbf{X}$.

3. Considere la base de datos `earthquake` del paquete `MPV`, seleccione el mejor modelo usando como criterios el MSE_p y el C_p de Mallows al emplear el método de selección de todas las regresiones posibles.