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?

Escalamiento de longitud unitaria

$$X_{iJ}^{*} = \frac{X_{iJ} - \bar{X}_J}{\sqrt{\sum_{k=1}^{n} (X_{kJ} - \bar{X}_J)^2}}$$

Se llama así porque la columna X_J^* tiene magnitud 1.

- La variable X_J^* es adimensional

Si se ajusta el modelo

$$Y_i = \beta_0 + \beta_1 X_{1i}^* + \cdots + \beta_K X_K^*$$

¿cuáles son las unidades de los betas?

R/ Los coeficientes de regresión tienen las unidades de Y.

b) ¿Por qué hay problemas de multicolinealidad cuando se tienen más covariables que observaciones en los datos?

Consideremos la matriz de diseño X, $X \in \mathbb{R}^{n \times p}$

Como hay más covariables que observaciones $K > n$ y $p > K$, entonces $p > n$

$$0 \leq \text{Rango}(X) \leq \min\{p, n\} = n$$

$$0 \leq \text{Rango}(X) \leq n \quad (1)$$

Multiplicando por -1 en (1)

$-n \leq -\text{Rango}(x) \leq 0$

$p-n \leq p-\text{Rango}(x) \leq p$ (Sumar p en todas partes)

$$0 < p-n \leq \underbrace{p-\text{Rango}(x)}_{\#\text{Variables libres}}$$

#Variables libres $> 0 \Rightarrow$ Hay dependencia lineal entre las columnas, es decir, multicolinealidad

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?

Recordemos que la traza de una matriz es igual a la suma de los eigenvalores.

R/ Falsa, porque se considera la traza de $(X'X)^{-1}$

Si hay multicolinealidad $X'X$ es no invertible

Llamemos $\lambda_1, \lambda_2, \ldots, \lambda_p$ a los eigenvalores de $X'X$, entonces los eigenvalores de $(X'X)^{-1}$ son $1/\lambda_1, \ldots, 1/\lambda_p$

La traza de $(X'X)^{-1} = \sum_{i=1}^{p} \frac{1}{\lambda_i}$ Si esto es grande

es porque hay λ_i pequeños y si hay λ_i pequeños $\prod_{i=1} \lambda_i$ va a ser pequeño $\Rightarrow$ va a haber problemas invirtiendo $X'X$.

d) Si la correlación entre las variables X_j y X_k es pequeña, ¿se puede descartar la presencia de multicolinealidad?

Supongamos que $X_1 = 2X_2 + 3X_3$, pero puede ser que $Cor(X_1, X_2)$ y la $Cor(X_1, X_3)$ sea pequeña.

Comparar por pares puede sesgar la información y no considera combinaciones lineales.