ANALISIS DE REGRESION

MULTICOLINEALIDAD

INTRODUCCIÓN

Uno de los supuestos básicos del modelo lineal general $Y = X\beta + \varepsilon$ establece que las variables explicativas son linealmente independientes.

La multicolinealidad en regresión es una condición que ocurre cuando algunas variables predictoras incluidas en el modelo están correlacionadas con otras variables predictoras.

1. TIPOS DE MULTICOLINEALIDAD

Se distinguen dos tipos de multicolinealidad: exacta y aproximada.

Multicolinealidad Exacta:

Multicolinealidad Aproximada:

2.- FUENTES DE MULTICOLINEALIDAD

1º Método empleado en la recolección de los datos, cuando el analista muestrea sólo un subespacio de la región definida para los regresores.

2º Restricciones en el modelo o en la población que se muestrea, se presentan restricciones en casos que se refieren a procesos de producción o procesos químicos, cuando las variables regresoras son las componentes de un producto y ésos suman una constante.

3º Especificación del modelo, cuando se incluyen variables regresoras irrelevantes o términos polinomiales en el modelo, o cuando el rango de los valores de la variable regresora es pequeño.

4º Modelo sobredefinido, cuando en el estudio se considera más variables regresoras que observaciones, es muy frecuente en las investigaciones médicas.

3.- CONSECUENCIAS DE LA MULTICOLINEALIDAD

1. ***Las varianzas y covarianzas de las estimaciones son muy grandes.***

2. ***Al efectuar contrastes de significación individual no se rechazará la hipótesis nula, mientras que al realizar contrastes conjuntos sí.***

3. ***Los intervalos de confianza, ($\hat{\beta}_i \pm t\sqrt{V(\hat{\beta}_i)}$), son muy amplios, incluyendo valores negativos y positivos.***

Consecuencias prácticas de la multicolinealidad

- **Efectos en las varianzas:**
- **Imprecisión en la estimación por intervalo:**
- **Consecuencias en la toma de decisiones:**
- **Efecto en las Estadísticas de relevancia global:**

4.-DIAGNÓSTICO DE LA MULTICOLINEALIDAD

Un gran número de técnicas estadísticas han sido propuestas para detectar multicolinealidad, presentaremos algunas medidas para diagnosticarla:

4.1. Matriz de Correlación:

Es de gran utilidad para detectar dependencia lineal entre pares de variables regresoras, si las correlaciones $r_{ij} \rightarrow \pm 1$, existe colinealidad perfecta entre la i-ésima y j-ésima variable regresora,
Algunos autores consideran que $r_{ij} \geq 0.60$ ya es una situación problemática.

4.2. Tolerancia de una variable:

Se define como: $\mathbf{TOL_j = 1- R_j^2}$,

la cantidad de variabilidad de X_j que no es explicada por las otras variables regresoras. $\mathbf{0 \leq TOLj \leq 1}$

Valores cercanos a uno indican ausencia de multicolinealidad.
Valores inferiores a 0.20, indica grado elevado de multicolinealidad.
Valores inferiores a 0.10 indica severa multicolinealidad y la necesidad de tomar medidas para reducirla.

4.3. Factor de Inflación de la varianza: FIV,

Sabemos que: $V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

Si existe colinealidad la varianza tiende al infinito. Para el j-ésimo parámetro tenemos que

$$V(\hat{\beta}_j) = \sigma^2 c_{jj}$$

c_{jj} es el elemento de X´X esto es:

$$\mathbf{FIV_j = C_{jj} = 1 / (1- R_j^2)}$$

Es el recíproco de la tolerancia de una variable.

El FIV para cada término en el modelo mide el efecto combinado de la dependencia entre los regresores sobre la varianza de dicho término.
Si FIV_j es un valor cercano a uno, indica ausencia de multicolinealidad.

Si FIVj >5 ó, indica que los parámetros del modelo de regresión asociados son pobremente estimados, debido a la presencia de multicolinealidad, se recomienda tomar medidas correctivas.
Si FIVj >10, indica multicolinealidad severa, y es necesario tomar medidas para eliminarla.
Si el FIV excede a 5 o 10 entonces indica que los coeficientes de regresión asociados son estimados pobremente debido a la multicolinealidad.

4.4. Análisis del sistema propio de X´X:

Las raíces características o autovalores de X´X, denotados por λ_1 , λ_2 ,.....,λ_p nos permite evaluar el alcance de la multicolinealidad en las variables regresoras.

Uno o más autovalores pequeños (cercanos a cero) implica que existe una cercana dependencia lineal entre las columnas de X, es decir severa multicolinealidad.

4.4.1.- Número de condición de X´X:

Los valores propios de X´X se pueden usar para medir el grado de multicolinealidad en los datos. El número condición se denota por "K":

Definida como: $K = \sqrt{\lambda máx \ / \ \lambda mín}$

Es una medida de dispersión en el espectro de los autovalores de X´X .

Si:

K < 100, no existe problema serio de multicolinealidad.
100≤ K ≤1000, existe multicolinealidad de moderada a fuerte.
K > 1000 existe severa multicolinealidad.

4.4.2.- Índice de condición de X

Un análisis de los autovalores de X´X puede ser utilizado para identificar la naturaleza de los datos.

El índice de condición K_j está definida como:

$$K_j = \frac{\sqrt{\lambda_{máx}}}{\sqrt{\lambda_j}} \qquad j = 1,2,....,\ p$$

Si:

$K_j < 100$ no hay problema grave de multicolinealidad

$100 \leq K_j \leq 1000$ implican multicolinealidad de moderada a fuerte.

$K_j > 1000$ indicio de una multicolinealidad fuerte.

4.4.3.- Proporción de la varianza;

La matriz de covarianza del estimador de los parámetros del modelo de regresión es:

$$\mathbf{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\sigma}^2(\mathbf{X}^{*T}\ \mathbf{X}^*)^{-1} = \boldsymbol{\sigma}^2\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^{T}$$

Los autores Belsley y Welsch, sugieren usar proporciones de descomposición de varianza, definidas como sigue:

$\pi_{ij} = (t_{ji}^2 / \mu_i^2) \ /\ VIF_j \qquad j = 1,2,.....,p$

como medidas de multicolinealidad.

Si $\boldsymbol{\pi}_{ij} > 0.5$ indica presencia de multicolinealidad.

Si una alta proporción de la varianza de dos o más parámetros estimados es asociada con un autovalor pequeño, decimos que existe evidencia de multicolinealidad.

4.4.4.- El determinante de X´X ;

Puede ser utilizado como un indicador de multicolinealidad.
Si $|X´X| = 1$ los regresores son ortogonales pero
Si $|X´X| = 0$ existe una exacta dependencia lineal exacta entre las variables regresoras.

5.-MÉTODOS PARA TRATAMIENTO DE LA MULTICOLINEALIDAD

5.1.-Recolección de datos adicionales, puede atenuar el grado de multicolinealidad, en el caso de series temporales (pasar de datos anuales a mensuales, trimestrales, etc.) y de corte transversal. Otra opción es incrementar el tamaño de la muestra.

5.2.- Especificación del modelo, incluir nuevas variables en el modelo o eliminar variables que se encuentren altamente correlacionadas.

5.3.- Regresión ridge(cresta), es un procedimiento de ajuste de los estimadores MCO con el objetivo de reducir su variación. Tema de exposición.

5.4.-Componentes principales (transformación lineal de los estimadores mínimo cuadráticos) Proporciona estimadores sesgados.
Tema de exposición.