Segundo Corte: Diagnóstico e Inferencia de una regresiín lineal Múltiple

Andrés Martínez

2019

Índice

1. Regresión Lineal Múltiple

En un modelo lineal múltiple se asume qye Y es una función lineal de sus predictores mas un ruido blanco:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_p X_p + \epsilon \tag{1}$$

1.1. Supuestos

1. Todos los parámetros son lienales
2. La muestra de n observaciones sigue el modelo poblacional del primer supuesto
3. No existe perfecta colinearidad
4. EL error tiene un valor esperado de cero dado por variables independientes.
5. Homocedasticidad: El error tiene la misma varianza que los valores de las variables explicativas.

No hacemos suposiciones sobre las distribuciones (marginales o conjuntas) de X_i, pero suponemos que $E\epsilon|X = 0$, $Var\epsilon|X = \sigma^2$, y ese ϵ no está correlacionado entre las mediciones. La forma matricial del modelo es:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{2}$$

Donde $\mathbf{X}$ incluye una columna inicial de 1.

Cuando se incluye se asume que este debe tener un comportamiento normal como se ve en la siguiente ecuación:

$$\epsilon \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}) \tag{3}$$

independiente de $\mathbf{X}$.

Los coeficientes se estiman de la siguiente forma:

$$\widehat{\beta} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{Y} \tag{4}$$

Bajo el supuesto de ruido blanco, este es el mismo resultado que la estimación por máxima verosimilitud

$$\widehat{\mathbf{Y}} = \mathbf{x}\widehat{\beta} = \mathbf{H}\mathbf{Y} \tag{5}$$

Que es simetrico e idempotente. Los residuales se obteienen así:

$$SSE = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{6}$$

donde $\mathbf{I} - \mathbf{H}$ es simetrico e idempotente

1.2. Derivación de los Estimadores

$$\widehat{\beta}_{OLS} = argmin_{\beta\in\mathbb{R}^2}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) \tag{7}$$

$$\frac{\partial}{\partial\beta} = (\mathbf{Y}^T\mathbf{Y} - 2\mathbf{Y}^T(\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta)) \tag{8}$$

$$\frac{\partial}{\partial\beta} = -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\hat{\beta} = 0 \tag{9}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{10}$$

1.3. Propiedades de los Estimadores

Insesgamiento del estimador mínimo cuadrado

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{11}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}^T(\mathbf{X}\beta + \epsilon) \tag{12}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\beta \tag{13}$$

$$+(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon) \tag{14}$$

$$= \beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon) \tag{15}$$

$$E[\hat{\beta}|x] = \beta \tag{16}$$

Varianza de los errores estandar: Suponiendo Homocedasticidad

$$V[\hat{\beta}|\mathbf{X}] = V[\beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)/\mathbf{X}] \tag{17}$$

$$V[\hat{\beta}|\mathbf{X}] = V[(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)/\mathbf{X}] \tag{18}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}V[\mathbf{X}^T\epsilon/\mathbf{X}](\mathbf{X}^T\mathbf{X})^{-1} \tag{19}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V[\epsilon/\mathbf{X}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{20}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{21}$$

$$\hat{\beta} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{22}$$

$$\hat{\beta} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{23}$$

Consistencia

- Ley de los grandes números
- Teorema Límite Central

Tenemos que $\hat{\beta} = \beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)$ y se desea ver si $\hat{\beta}$ es consistente. $\lim_{n\to p}(\hat{\beta} - \beta) = 0$

Eficiencia: Bajo los supuestos de linealidad, muestreo aleatorio, no colinealidad perfecta, medica condicional cero de los errores y homocedasticidad, $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ es el mejor estimador lineal insesgado.

Además $\hat{\beta}_1 = \beta_1 + (\mathbf{X}_1^T\mathbf{X}_1)^{-1}(\mathbf{X}_1^T\epsilon_2)$.

1.4. Colinealidad

Asumiendo que $((X)^T(X))^{-1}$ existe y es invertible o no singular, existen un número equivalente de condiciones para una matriz que debe ser invertible:

- Su determinante no es cero
- Su rango de columnas es completo, lo que quiere decir que todas las columnas son lienalmente independientes.
- Su rango de filas es completo, lo que quiere decir que las filas son linealmente independientes.

Esto implica que las variables deben ser linealmente independientes El objetivo más importante de estos modelos es obtener los errores, pues permiten establecer la eficiencia del modelo y las posibles falencias de especificación que este pueda tener. Para esto definimos rapidamente los tres errores que siempre se buscan en una regresión lineal múltiple o simple:

$$SCT = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \mathbf{Y}^T(\mathbf{I} - \mathbf{I}_{n\times n})\mathbf{Y}. \tag{24}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}. \tag{25}$$

$$SSR = \sum_{i=1}^{n}(\hat{y} - \overline{y})^2 = \mathbf{Y}^T(\mathbf{H} - \mathbf{I}_{n\times n})\mathbf{Y}. \tag{26}$$

1.5. Predicciones

Diciendo que $\mathbf{x}'$ es $m \times (p+1)$ matriz dimensional guardando valores de t m puntos, donde se desean hacer las predicciones. Entonces m puede ser más grande, menor o igual a n. Por lo tanto, $\mathbf{Y}^T$ es $m \times 1$ una matriz de variables aleatorias de Y en cada punto

$$E\mathbf{Y}^T|\mathbf{X}' = \mathbf{x}^T = \mathbf{m}(\mathbf{x}^T) = \mathbf{x}^T\beta \tag{27}$$

y se estima a través de:

$$\widehat{\mathbf{m}}(\mathbf{x}^T) = \mathbf{x}^T\widehat{\beta} \tag{28}$$

Que es lo mismo que decir

$$\widehat{\mathbf{m}}(\mathbf{x}^T) = \mathbf{x}'(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} \tag{29}$$