Examen Parcial Análisis III

Marcelo Molinatti

1. Para el modelo de regresión líneal $y_n = \beta x_n + e_n$. Para una muestra $y_1, \ldots, y_N$ escriba los estimados de mínimos cuadrados utilizando la forma multivariada $Y = X\beta + e$ donde $X = (x_1, \ldots, x_N)^T$, sin derivar, sólo a partir del estimador escrito como proyección.

Sea E el espacio generado por las columnas de X tal que $E =< X >= \{X\beta^T : \beta^T \in \mathbb{R}^k\}$, entonces, la proyección ortogonal en E del vector y pertenece a E. Esto es, $P_E(y) \in E$, por lo que $\exists \beta \in \mathbb{R}^k$ tal que $P_E(y) = X\beta^T$, de lo que se desprende que:

$$< X^i, y >=< X^i, X\beta^T >$$

donde X^i es la i-ésima columna de X. Luego, usando la propiedad $< a, b >= a^T b$, se tiene que:

$$X^T y = X^T X \beta^T$$

Ahora, como las columnas de X generan a E, estas son linealmente independientes. Por lo que las filas de X^T son linealmente independientes, y entonces, $X^T X$ tiene rango completo. Luego, $X^T X$ es invertible, y existe la solución:

$$\beta^T = (X^T X)^{-1} X^T y$$

2. Demuestre que para el modelo $Y = X\beta + Ve$ para Y un vector $N \times 1$, X una matriz $N \times k$, V una matriz simétrica $N \times N$ y $e = (e_1, \ldots, e_N)^T \sim N(0; I_{N\times N})$. Demuestre que este modelo puede ser reducido a $Y' = X'\beta + e$ donde $e \sim N(0; IN \times N)$.

Multiplicando por la izquierda a ambos lados de la igualdad en el modelo $Y = X\beta + Ve$, por V^{-1}, se obtiene:

$$V^{-1}Y = V^{-1}X\beta + V^{-1}Ve \to Y' = X'\beta + e$$

donde $Y' = V^{-1}Y$ y $X' = V^{-1}X$.

3. Ejemplo: Utilizamos la información sobre la renta y el número de años de escolarización (formal) para estimar en qué medida la renta anual de un hombre está relacionada con sus años de escolarización. Una posibilidad sería que, para un hombre que no haya estudiado nada, su renta anual fuera de a (em dolares), y que, por cada año de escolarización que haya tenido, su renta aumentara en b (en dolares). Así, para un hombre con x años de escolaridad, esperaríamos que su ingreso anual sea de $a + bx$ dólares. Si tenemos la siguiente muestra para esta muestra podemos considerar $y_i = a + bx_i + e_i$:

y e_i es una desviación del valor medio $a + bx_i$ que puede existir entre un individuo y otro.
a) Suponga $a = 0$, utilice las fórmulas obtenidas en la pregunta 1. para los datos anteriores. Establezca conclusiones a partir de su modelo.

Dado que no se toma en cuenta el intercepto, se tiene que la matriz de diseño queda como $X = (6, 12, 10, 8, 9)^T$, por lo que $X^T X$ es la sumatoria de cuadrados de los x's:

Sujeto	Ingreso (×1000$)	Años de Escolarización
1	10	6
2	20	12
3	17	10
4	12	8
5	11	9

$$X^T X = \sum_{i=1}^{5} x_i^2 = 425$$

y $(X^T X)^{-1} = 1/425$. De forma que el estimado de la pendiente es:

$$b = (X^T X)^{-1} X^T y = \frac{1}{425} \sum_{i=1}^{5} x_i y_i = \frac{1}{425} 665 \approx 1.565$$

Esto indica que hay un incremento en los ingresos de aproximadamente 1,6 veces por cada año de escolarización que un sujeto haya realizado.

b) Supongamos en el ejemplo anterior que los datos de las edades de cada sujeto son agregados 28, 45, 40, 35, y 22. Considere el modelo de regresión $y_i = a + b_1 x_{1i} + b_2 x_{2i} + e_i$. Encuentre los estimados de mínimos cuadrados. Establezca conclusiones a partir de este modelo.

En este caso, la matriz de diseño queda (ignorando de nuevo el intercepto):

$$X = \begin{pmatrix} 6 & 28 \\ 12 & 45 \\ 10 & 40 \\ 8 & 35 \\ 9 & 22 \end{pmatrix}$$

De forma que:

$$X^T X = \begin{pmatrix} 425 & 1586 \\ 1586 & 6118 \end{pmatrix}$$

y:

$$(X^T X)^{-1} = \begin{pmatrix} 0.0721854 & -0.018713 \\ -0.018713 & 0.0050145 \end{pmatrix}$$

de tal forma que los estimados de mínimos cuadrados son:

$$\beta = \begin{pmatrix} 0.0721854 & -0.018713 \\ -0.018713 & 0.0050145 \end{pmatrix} \begin{pmatrix} 6 & 12 & 10 & 8 & 9 \\ 28 & 45 & 40 & 35 & 22 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ 17 \\ 12 \\ 11 \end{pmatrix} = \begin{pmatrix} 0.8091418 \\ 0.2024683 \end{pmatrix}$$

Se observa ahora, que el efecto de cada año de escolarización es de casi la mitad del valor encontrado en el modelo anterior, y que la edad ahora parece tener una relación positiva con el ingreso de un sujeto. Esto tiene sentido si suponemos que la edad esta relacionado con la cantidad de tiempo que una persona lleva en el campo laboral, y por tanto, con la probabilidad de que el ingreso sea mayor.

c) ¿Cuál modelo es más adecuado a utilizar? Ayuda: utilice una prueba estadística adecuada.