INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Matriz aniquiladora y teorema de Frisch-Waugh

¿Qué debería saber antes de leer este documento?

1. Modelo real: $Y = X\beta + U$.
2. Modelo estimado: $Y = X\hat{\beta} + \hat{U}$
3. Predicción: $\hat{Y} = X\hat{\beta}$

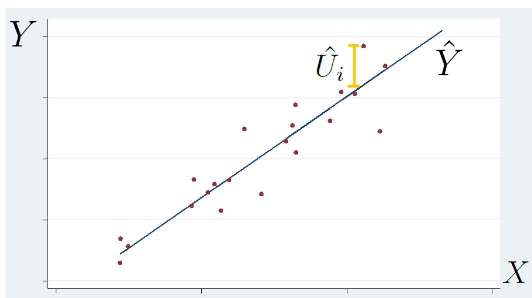


Figura 1: Gráfico genérico de muestras y el modelo que mejor se ajusta a ellas.

Álgebra e interpretación

En primer lugar, notemos que si combinamos las ecuaciones 2 y 3, obtenemos lo siguiente:

$$Y = \hat{Y} + \hat{U}$$

Es decir, Y es la suma de dos componentes:

- La que se puede explicar con las variables X que observamos en nuestra base de datos ($\hat{Y}$).
- La que se puede explicar con aquellas variables que no hemos bautizado y que no observamos ($\hat{U}$).

Ahora presentaremos a dos matrices con las que debemos familiarizarnos...

- **Matriz de proyección:** ($P_X = X(X^TX)^{-1}X^T$) al pre-multiplicar un vector por esta matriz, obtenemos aquella componente que puede ser explicada por las columnas de X. Es decir, estamos "proyectando" dicho vector en el mundo de X.

Lógicamente, la proyección de X en el mundo de las X es X...

$$P_X X = X \underbrace{(X^T X)^{-1} X^T X}_{I_{K\times K}} = X$$

Como dijimos anteriormente, $\hat{Y}$ es la parte de Y que "vive" en el mundo de las X...

$$P_X Y = X \underbrace{(X^T X)^{-1} X^T Y}_{\hat{\beta}} = \hat{Y}$$

- **Matriz aniquiladora:** ($M_X = I_{K\times K} - P_X$) al pre-multiplicar un vector por esta matriz, obtenemos aquella componente que <u>no</u> puede ser explicada por las columnas de X. Es decir, estamos "proyectando" dicho vector en el mundo de las variables no observadas.

  Como bien dice su nombre, esta matriz "aniquila" a X...

  $$M_X X = (I_{K\times K} - P_X)X = X - P_X X = X - X = 0_{N\times K}$$

  Como dijimos anteriormente, $\hat{U}$ es la parte de Y que "vive" fuera del mundo de las X...

  $$M_X Y = (I_{K\times K} - P_X)Y = Y - P_X Y = Y - \hat{Y} = \hat{U}$$

  Donde $\hat{U}$ es el residuo de regresionar Y en función de X, por lo que lo notaremos como $\hat{U}_{Y,X}$. Recuerden esto último...

Teorema de Frisch-Waugh: metodología

Consideremos el siguiente modelo y premultipliquémoslo por $M_{X_A} \in \mathbb{R}^{N\times N}$

$$\underbrace{Y}_{N\times 1} = \underbrace{\vec{\beta_0}}_{N\times 1} + \underbrace{X_A}_{N\times K_A} \underbrace{\beta_A}_{K_A\times 1} + \underbrace{X_B}_{N\times 1} \underbrace{\beta_B}_{1\times 1} + \underbrace{U}_{N\times 1}$$

$$\Rightarrow \underbrace{M_{X_A} Y}_{\hat{U}_{Y,X_A}} = \underbrace{M_{X_A}\vec{\beta_0}}_{\beta_0'} + \underbrace{M_{X_A} X_A}_{0_{N\times K}} \beta_A + \underbrace{M_{X_A} X_B}_{\hat{U}_{X_B,X_A}} \beta_B + \underbrace{M_{X_A} U}_{U'}$$

$$\Rightarrow \hat{U}_{Y,X_A} = \beta_0' + \hat{U}_{X_B,X_A}\beta_B + U'$$

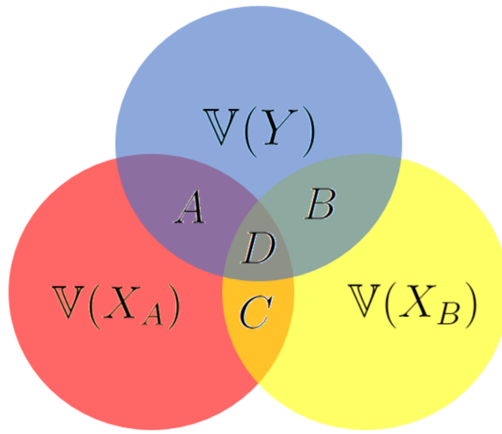
En pocas palabras, este teorema nos dice que la estimación por mínimos cuadrados ordinarios de β_B entrega el mismo valor para el modelo con y sin pre-multiplicar por M_{X_A}.

Lo anterior nos permite obtener $\hat{\beta}_B$ mediante el siguiente procedimiento.

1. Regresionar Y en función de X_A y guardar los residuos U_{Y,X_A}.
2. Regresionar X_B en función de X_A y guardar los residuos U_{X_B,X_A}.
3. Regresionar $\hat{U}_{Y,X_A}$ en función de $\hat{U}_{X_B,X_A}$. El coeficiente a estimar, asociado a $\hat{U}_{X_B,X_A}$ es nada más ni nada menos que $\hat{\beta}_B$.

Teorema de Frisch-Waugh: interpretación

Cuando planteamos un modelo de regresión lineal, nos interesa entender en qué medida explican las variables independientes X_A y X_B la disperción (o varianza) observada en la variable dependiente Y. Consideremos la siguiente imagen.



- Al regresionar Y en función de X_A y X_B, el efecto de la región A será capturado por $\hat{\beta}_A$ y el de la región B será capturado por $\hat{\beta}_B$. Por su parte, el área D representa una región conflictiva, donde no sabemos a qué variable atribuirle el efecto, pues hay colinealidad (correlación) entre X_A y X_B. Este conflicto se conoce como "problema de identificación".

- Supongamos que ahora sólo regresionamos Y en función de X_B. La estimación de mínimos cuadrados ordinarios le atribuirá a β_B el efecto de la región D, lo cual es engañoso porque parte de dicho efecto puede deberse a X_A y no a X_B. Ahora caeríamos en el "problema de endogeneidad" (una de las variables con las cuales hay problema de identificación no está siendo observada). Éste es un gran problema en estadísticas y lo estudiaremos más adelante.

- El teorema de Frisch-Waugh permite "limpiar" el efecto de X_A (lo cual se hace al momento de pre-multiplicar por M_{X_A}). Al aplicarlo, ya no tenemos que preocuparnos de la región D y podemos estimar β_B sin ningún problema.