Clase auxiliar # 1

Estimación de Mínimos Cuadrados Ordinarios y densidades de probabilidad

¡Recordar!

Variables y parámetros:

- Número de observaciones: $N \in \mathbb{R}$.
- Número de variables explicativas : $K \in \mathbb{R}$.
- Vector de observaciones de la variable dependiente: $Y = (Y_1, ..., Y_N) \in \mathbb{R}^{N \times 1}$.
- Matriz de observaciones de las variables independientes: $X \in \mathbb{R}^{N \times K}$
- Vector de parámetros poblacionales: $\beta = (\beta_1, ..., \beta_N) \in \mathbb{R}^{K \times 1}$
- Vector de errores aleatorios: $U = (U_1, ..., U_N) \in \mathbb{R}^{N \times 1}$. También se suele notar como ε.

Supuestos de MCO:

1. **Linealidad de parámetros:** $Y = X\beta + U$.

   Notar que la linealidad es sobre los parámetros, no sobre las variables. Podemos tener perfectamente un modelo de la forma: $\tanh(Y_i) = \beta_1 \ln(X_{i1}) + \beta_2 \exp(X_{i2}) + \beta_3 X_{i3}^{0,7} + \beta_4 \arctan(X_{i3}) + U_i$.
2. **Matriz $X^T X$ invertible:** $rg(X^T X) = K$. Se puede verificar.
3. **Errores nulos en esperanza:** $\mathbb{E}(U) = 0$. Supuesto aceptable.
4. **Independencia y homocedasticidad:** $\mathbb{V}(U) = \sigma^2 I_N$. Se puede testear.
5. **Exogeneidad:** $\mathbb{E}(X^T U) = 0$. Supuesto fuerte y problemático. En ciencias sociales siempre hay correlación entre variables que observamos y que no observamos.

1. Estimación teórica de MCO

Considere el modelo $Y = X\beta + U$. Llamaremos $\hat{\beta}$ a un estimador genérico de β, $\hat{Y} = X\hat{\beta}$ a la predicción de la variable dependiente Y según el estimador $\hat{\beta}$ y $\hat{U} = Y - \hat{Y}$ a los residuos del modelo –diferencia entre el valor real de la variable dependiente y su predicción–.

1. Demuestre que si $X^T X$ es invertible y definida positiva, entonces $\hat{\beta}_{MCO} = (X^T X)^{-1} X^T Y$ es minimizador global de la suma de los residuos al cuadrado.

2. Suponiendo que se cumplen las hipótesis anteriores, obtenga el estimador β_{MCO} para un modelo de la forma $Y_i = \beta_0 + \beta_1 X_{i,1} + U_i$.

2. Estimación computacional de MCO

En este ejercicion analizaremos cómo afecta el gasto per cápita en salud (US$) a la tasa de mortalidad de menores de 5 años (por cada 1.000) a nivel global. Para lo anterior, ocuparemos datos proporcionados por el Banco Mundial correspondientes al año 2014 y donde cada muestra corresponde a un país. La base de datos se encuentra en U-Cursos bajo el nombre de `mortalidad_gastosalud.xls`.

1. **Stata:**
   - Importe los datos y realice un gráfico *scatter* (de disperción) que muestre cómo se comportan conjuntamente ambas variables.
   - Regresione la mortalidad en función del gasto per cápita y analice los resultados. Si le parece conveniente, puede realizar transformaciones en las variables con tal de mejorar el ajuste.

2. **Matlab:** importe los datos y obtenga el estimador MCO para la misma regresión de la parte anterior. Compare los resultados con los de Stata.

3. **Solver de Excel:** proponga un estimador inicial $\hat{\beta}$ y obtenga el estimador MCO para la misma regresión de las partes anteriores. Recordar que queremos minimizar la suma de los residuos al cuadrado. Compare los resultados con los de Stata y Matlab.

3. Función de densidad conjunta, marginal y condicional

Considere la siguiente densidad de probabilidad conjunta entre las variables aleatorias X e Y [1].

$$f(x,y) = \begin{cases} 8xy & \text{si } 0 < x < y < 1 \\ \\ 0 & \text{en caso contrario} \end{cases}$$

1. Muestre que la función f está bien definida. Es decir: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = 1$ y $f(x,y) = 0 \forall (x,y) \in \mathbb{R}^2$.

2. Obtenga las densidades marginales de X e Y: $f_X(x)$, $f_Y(y)$.

3. Obtenga las densidades de probabilidad condicionales $f_{X|Y}(x|y), f_{Y|X}(y|x)$.

[1] Referencia: Luis Rincón, Universidad Nacional Autónoma de México (`http://lya.fciencias.unam.mx/lars/0625/`)