# Clase auxiliar # 1

## ¡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 $\varepsilon$.

### 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.

### Formas funcionales:

| Forma funcional | Ejemplo de forma funcional | Cambios de Y respecto a X |
|---|---|---|
| Lin-lin | $Y_i = \beta_0 + \beta_1 X_i + U_i$ | Si $X$ aumenta en una unidad, $Y$ cambia en $\beta_1$ unidades. |
| Log-log | $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + U_i$ | Si $X$ aumenta en un 1 %, $Y$ cambia en un 1 % |
| Lin-log | $Y_i = \beta_0 + \beta_1 \ln(X_i) + U_i$ | Si $X$ aumenta en un 1 %, $Y$ cambia en $\frac{\beta_1}{100}$ unidades. |
| Log-lin | $\ln(Y_i) = \beta_0 + \beta_1 X_i + U_i$ | Si $X$ incrementa en una unidad, $Y$ cambia en un 100·$\beta_1$ % |
| Cuadrática | $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + U_i$ | Si $X$ incrementa en una unidad, $Y$ cambia en $\beta_1 + 2\beta_2 X$ unidades |

# 1. Comente

1. **[P1, Control 1, 2017]** Explique si la acumulación de evidencia empírica que no contradice las predicciones de una teoría fortalece a esta última.

   ***Respuesta:***

   La acumulación de evidencia empírica que no contradice a una teoría la fortalecerá siempre que:

   - No sea explicada de mejor manera por otra teoría (siguiendo el pensamiento de Thomas Kuhn).
   - Dicha evidencia proporcione nueva información que permita generalizar resultados, en lugar de ser acotada y redundante.

2. **[P2, Control 1, 2017]** En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*

   Explique qué dificultades podrían haber para probar que salud mental de una persona **es causada** por la de sus amigos.

   ***Respuesta:***

   A continuación se enuncian posibles dificultades:

   *a*) Defender que se trata de una causalidad y no únicamente correlación. Se deben encontrar argumentos teóricos que respalden la relación causal entre tener amigos saludables mentalmente y estar sanos en ese sentido, al igual que ellos.

   *b*) Encontrar o desarrollar alguna métrica cuantitativa con sentido que mida la salud mental de las personas.

3. **[P1.c, Control 1, 2016]** Un médico le indica a usted que "la correlación negativa entre número de cigarrillos fumados al día durante el embarazo de una mujer y el peso de su bebé al nacer es evidencia clara del efecto negativo del tabaquismo en la salud del recién nacido". Comente.

   ***Respuesta:***

   Aunque, correlación no es causalidad, hay tres posibilidades consistentes con una correlación negativa descrita:

   - El mayor número de cigarrillos causa reducción del peso del bebé. $(X \rightarrow Y)$
   - Los bebés de menor peso causan un mayor número de cigarrillos fumados $(Y \rightarrow X)$
   - Existen otras variables Z tales como bajo nivel de educación de salud sobre efectos del tabaquismo en el bebé y su madre, así como pobreza. Estos factores pueden causar, a su vez, a mala nutrición de la madre, insuficientes controles médicos, etc que pueden causar, al mismo tiempo, mayor tabaquismo y menor peso del bebé $(Y \leftarrow Z \rightarrow X)$.

4. **[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

   ***Respuesta:***

   - Datos de corte transversal: datos de distinta índole en un momentos del tiempo. Ejemplo: PIB de países latinoamericanos en 2013.
   - Series de tiempo: datos para una unidad (individuo, empresa, país, etc) a través del tiempo. Ejemplo: PIB anual para Chile 1990-2013.
   - Datos de panel: datos de distinta índole a través de diferentes momentos del tiempo. Ejemplo: PIB anual de países latinoamericanos 1980-2013.

# 2. Formas funcionales (lin-lin, log-log, lin-log, log-lin)

Suponiendo causalidad, demuestre que:

1. En un modelo lin-lin ($Y_i = \beta_0 + \beta_1 X_i + U_i$), si $X$ aumenta en una unidad, $Y$ aumenta en $\beta_1$ unidades.

   ***Respuesta:***

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 \Rightarrow \Delta Y_i \approx \beta_1 \Delta X_i$$

$$\therefore \Delta X_i = 1 \Rightarrow \Delta Y_i \approx \beta_1$$

2. En un modelo log-log ($\ln Y_i = \beta_0 + \beta_1 \ln X_i + U_i$), si $X$ cambia en un 1 %, $Y$ cambia en $\beta_1$ %.

   ***Respuesta:***

$$\frac{\partial(\ln Y_i)}{\partial X_i}\beta_1 \frac{1}{X_i}$$

$$\frac{\partial(\ln Y_i)}{\partial X_i}\beta_1 \frac{1}{X_i}$$

$$\beta_1 = \frac{\partial Y_i}{\partial X_i}\frac{X_i}{Y_i}$$

   Aquí se puede evidenciar que en un modelo log-log, el coeficiente corresponde a la elasticidad de la variable dependiente en función de la independiente.

$$\Rightarrow \frac{\Delta X_i}{X_i}\beta_1 \approx \frac{\Delta Y_i}{Y_i}$$

$$\therefore \frac{\Delta X_i}{X_i} = 1\,\% \Rightarrow \frac{\Delta Y_i}{Y_i} \approx \beta_1\,\%$$

3. En un modelo lin-log ($Y_i = \beta_0 + \beta_1 \ln X_i + U_i$), si $X$ cambia en un 1 %, $Y$ cambia en $\frac{\beta_1}{100}$ unidades.

   ***Respuesta:***

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 \frac{1}{X_i}$$

$$\beta_1 = X_i \frac{\partial Y_i}{\partial X_i}$$

$$\frac{\Delta X_i}{X_i}\beta_1 \approx \Delta Y_i$$

$$\therefore \frac{\Delta X_i}{X_i} = 1\,\% \Rightarrow \Delta Y_i \approx \frac{\beta_1}{100}$$

4. En un modelo log-lin ($\ln Y_i = \beta_0 + \beta_1 X_i + U_i$), si $X$ cambia en una unidad, $Y$ cambia en $100 \cdot \beta_1$ %.

   ***Respuesta:***

$$\frac{1}{Y_i}\frac{\partial Y_i}{\partial X_i} = \beta_1$$

$$\Rightarrow \Delta X_i \beta_1 \approx \frac{\Delta Y_i}{Y_i}$$

$$\therefore \Delta X_i = 1 \Rightarrow \frac{\Delta Y_i}{Y_i} \approx \beta_1 = 100 \cdot \beta_1\,\%$$

## 3. Efectos porcentuales en los precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso, dentro de Chile en un sitio web. En cada punto, formule un modelo que permita estimar:

1. Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la diferencia porcentual aproximada de dos días de la semana.

   ***Respuesta:***

Si aparece “efecto porcentual”, es necesario la variable dependiente como logaritmo natural. dado que se pide el impacto del día de la semana en el presente pasaje, definimos variables dummies.

$$D_1 = 1_{\text{lunes}}, D_2 = 1_{\text{martes}}, ..., D_7 = 1_{\text{domingo}}$$

por cada día de la semana. Podemos escoger incluir todas ellas, y eliminar el intercepto, o bien mantener el intercepto y eliminar una de las variables de día de la semana, la que se transforma en la categoría base. La primera estrategia se describe a través de la siguiente ecuación:

$$\log P = \sum_{d=1}^{7} \beta_d D_d + U$$

con $\mathbb{E}(U|D) = 0$.

La esperanza condicional del logaritmo del precio se calcula así:

$$\mathbb{E}(\log P | D_d = 1) = \beta_d$$

La diferencia porcentual aproximada de dos días de la semana es por lo tanto:

$$\mathbb{E}(\log P | D_d = 1) - \mathbb{E}(\log P | D_k = 1) = \beta_d - \beta_k$$

2. Efecto porcentual del tiempo de estadía (tiempo entre el viaje de ida y de vuelta), considerando la posible importancia del día de la semana en que se viaje. Por ejemplo: no es lo mismo un viaje viernes-domingo, que uno martes-jueves.

   ***Respuesta:***

   $$\log P = \sum_{d=1}^{7} \beta_d D_d \times T + U$$

   con $\mathbb{E}(U|D) = 0$.

   El efecto marginal del tiempo $T$ sobre la esperanza condicional del logaritmo del precio se calcula así:

   $$\frac{\partial \mathbb{E}(\log P | D_d = 1)}{\partial T} = \beta_d$$

3. La respuesta porcentual del precio a la distancia entre origen y destino, considerando diferencias en dicha respuesta al número de escalas del viaje.

   ***Respuesta:***

   Sean $X$ la distancia entre el origen y el destino y $S_j$ una variable dummy que indica si es que hay $j$ escalas en el viaje.

$$\log P = \gamma_0 + \gamma_1 X + \sum_{j=0}^{J} \gamma_{2j} X \times S_j + U$$

con $\mathbb{E}(U|X,S) = 0$.

$$\frac{\partial \mathbb{E}(\log P | X, S_j = 1)}{\partial X} = \gamma_1 + \gamma_{2j}$$

4. Efecto porcentual en el precio del tiempo con anticipación de venta del pasaje, permitiendo un reacción nula para viajes comprados con 6 meses o más de anticipación.

   ***Respuesta:***

   Aquí es necesario incorporar una nueva variable que se puede definir como el tiempo de anticipación de venta del pasaje $A$ y una variable dummy $F = 1_{A>6\text{meses}}$. Para capturar el efecto descrito basta especificar:

   $$\log P = \delta_1 A + \delta_2 A \times F + U$$

   con $\mathbb{E}(U|A,F) = 0$.

   Así, el efecto marginal esperado del tiempo de compra anticipado sobre logaritmo del precio se calcula así, si la compra tiene anticipación de 6 meses o menos.

   $$\frac{\partial \mathbb{E}(\log P | A, F = 0)}{\partial X} = \delta_1$$

   El efecto marginal esperado del tiempo de compra anticipado sobre logaritmo del precio en una compra con anticipación mayor a 6 meses se calcula:

   $$\frac{\partial \mathbb{E}(\log P | A, F = 1)}{\partial X} = \delta_0 + \delta_1$$

   Para que el efecto sea numo para compras anteriores a 6 meses, necesitaríamos $\delta_0 + \delta_1 = 0$. En estricto rigor, no es necesario incorporar las variables binarias del punto anterior, pero no es incorrecto dejarlas.

# 4. Estimación computacional de MCO

En este ejercicio 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.

***Código:***

```
clear all
set more off

//Importar datos de Excel
import excel "C:\Users\Toshiba\Desktop\Ronald\Universidad\Ramos (
    Equipo Docente)\Aplicaciones de probabilidades y estad sticas
    para la gesti n 2017-1\Auxiliares\Auxiliar 01\Auxiliar 01 Pauta\
    mortalidad_gastosalud.xls", sheet("Data") firstrow

//Eliminar muestras con datos perdidos
drop if mort==. | gast==.

//Agregar etiquetas
label var mort "Tasa de mortalidad, menores de 5 a os (por cada
    1.000)"
label var gast "Gasto en salud per c pita (US$ a precios actuales)"

//Graficar mortalidad y gasto en salud
scatter mort gast

//Crear nuevas variables
gen log_mort = ln(mort)
gen log_gast = ln(gast)

//Graficar log-mortalidad y log-gasto en salud
scatter log_mort log_gast

//Regresionar log-mortalidad y log-gasto en salud
reg log_mort log_gast
```

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.

***Respuesta:***

```
bd = xlsread('mortalidad_gastosalud_edit.xls');

%% La primera variable corresponde a la mortalidad y la segunda al
    gasto
mortalidad = bd(:,1);
mortalidad = mortalidad(2:237,1);
gasto = bd(:,2);
```

```
gasto = gasto(2:237,1);

Y = log(mortalidad);
X_1 = log(gasto);

%% Creamos un vector de unos, para incorporar intercepto a nuestro
    modelo
X_0 = ones(236,1);

%% Juntamos X_0 y X_1 en una  nica matriz
X = [X_0 X_1];

%% Calculamos estimador MCO
beta_MCO = (X'*X)^(-1)*X'*Y
```

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.

   Ver material docente.