eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Tema 4

El Modelo de Regresión Lineal General

Especificación

Pilar González y Susan Orbe

Dpto. Economía Aplicada III (Econometría y Estadística)

Objetivos de aprendizaje

- Realizar el análisis de la forma funcional en Gretl.
- Identificar los principales elementos del modelo econométrico.
- Conocer los supuestos básicos del modelo de regresión.
- Interpretar los coeficientes desconocidos del modelo.
- Gestionar las variables ficticias en Gretl.

Contenido

1. Planteamiento del modelo.
2. Análisis gráfico y forma funcional.
   - Supuesto de linealidad.
   - Análisis gráfico y forma funcional en Gretl.
3. Variables explicativas cualitativas.
   - Inclusión en el modelo econométrico.
   - Gestión de las variables ficticias en Gretl.
4. Especificación del MRLG.
   - Supuestos básicos del modelo.
   - Interpretación de los coeficientes.

Contenido

Planteamiento del modelo.

Modelo econométrico.

$$\text{Variable dependiente} = \text{Parte sistemática} + \text{Parte aleatoria}$$

$$\text{Parte sistemática} = f(\text{variables explicativas})$$

$$\text{Parte aleatoria} = \text{Perturbación}$$

$$Y = f(\text{variables explicativas}) + \text{perturbación}$$

$$Y = f(X_2, X_3, \ldots, X_k) + u$$

Planteamiento del modelo.

Se ha de especificar:

- Variables explicativas: cuantitativas y/o cualitativas.
- Forma funcional $f(.)$: lineal, cuadrática, logarítmica, ...
- La perturbación es aleatoria, por lo que habrá que tener información sobre su distribución: media, varianza, ...

Contenido

1. Planteamiento del modelo.
2. Análisis gráfico y forma funcional.
   - Supuesto de linealidad.
   - Análisis gráfico y forma funcional en Gretl.
3. Variables explicativas cualitativas.
   - Inclusión en el modelo econométrico.
   - Gestión de las variables ficticias en Gretl.
4. Especificación del MRLG.
   - Supuestos básicos del modelo.
   - Interpretación de los coeficientes.

Análisis gráfico y forma funcional.

Ejemplo: función de consumo familiar.

Modelo económico:

$$C = f(R) \qquad\qquad \frac{dC}{dR} > 0 \quad \text{(propensión marginal a consumir)}$$

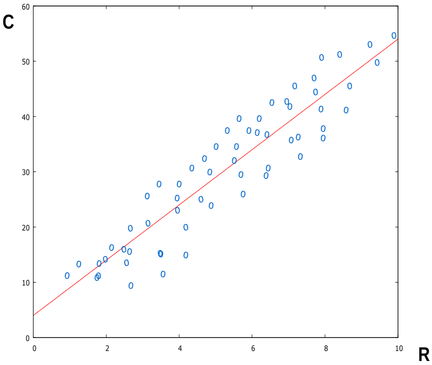
Modelo econométrico:

$$C = f(R) + u$$

La variable explicativa o regresor es la renta que es cuantitativa.

¿Y la forma funcional? ¿Qué tipo de relación existe entre el consumo y la renta?
Instrumento: gráfico del consumo frente a la renta.

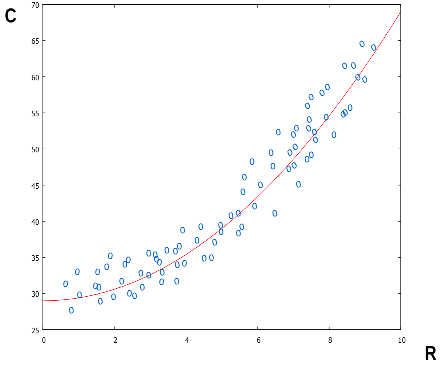
Análisis gráfico y forma funcional.



Forma funcional: lineal

$$C_i \;=\; \beta_1 \;+\; \beta_2\, R_i \;+\; u_i \quad i = 1, 2, ..., N$$

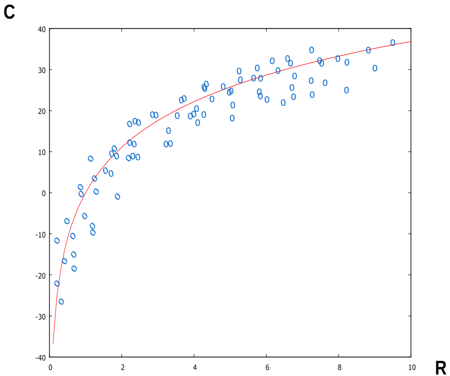
Análisis gráfico y forma funcional.



Forma funcional: cuadrática

$$C_i = \beta_1 + \beta_2 R_i + \beta_3 R_i^2 + u_i \quad i = 1, 2, ..., N$$

Análisis gráfico y forma funcional.



Forma funcional: logarítmica

$$\ln C_i = \beta_1 + \beta_2 \ln R_i + u_i \quad i = 1, 2, ..., N$$

Análisis gráfico y forma funcional.

$$
\begin{aligned}
\text{Lineal}: \quad C_i &= \beta_1 + \beta_2 R_i + u_i \\
\text{Cuadrática}: \quad C_i &= \beta_1 + \beta_2 R_i + \beta_3 R_i^2 + u_i \\
\text{Logarítmica}: \quad \ln C_i &= \beta_1 + \beta_2 \ln R_i + u_i \\
\text{Semilogarítmica}: \quad \ln C_i &= \beta_1 + \beta_2 R_i + u_i \\
C_i &= \beta_1 + \beta_2 \ln R_i + u_i
\end{aligned}
$$

¿Es válida cualquier forma funcional dentro del marco del modelo de regresión lineal general?

Supuesto. Linealidad

El modelo de regresión lineal es lineal en los coeficientes.

Análisis gráfico y forma funcional en Gretl.

EJEMPLO 4.1. Análisis gráfico y forma funcional.

1. Representar funciones en Gretl.

   Ejemplo 4.1.1.

2. Análisis gráfico de los datos y forma funcional.

   Ejemplo 4.1.2.

Contenido

1. Planteamiento del modelo.
2. Análisis gráfico y forma funcional.
   - Supuesto de linealidad.
   - Análisis gráfico y forma funcional en Gretl.
3. Variables explicativas cualitativas.
   - Inclusión en el modelo econométrico.
   - Gestión de las variables ficticias en Gretl.
4. Especificación del MRLG.
   - Supuestos básicos del modelo.
   - Interpretación de los coeficientes.

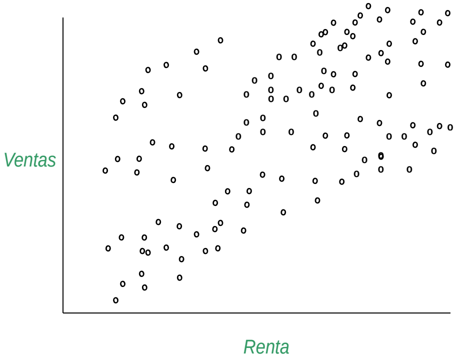
Variables explicativas cualitativas.

Algunos ejemplos:

1. Género, raza, nivel de estudios, puesto de trabajo, localización, zona de residencia...
2. Estacionalidad, cambios estructurales (crisis/no crisis, ...)
3. Variables cuantitativas recogidas por intervalo: renta, edad, ...

Ejemplo: ventas de una cadena de perfumerías.

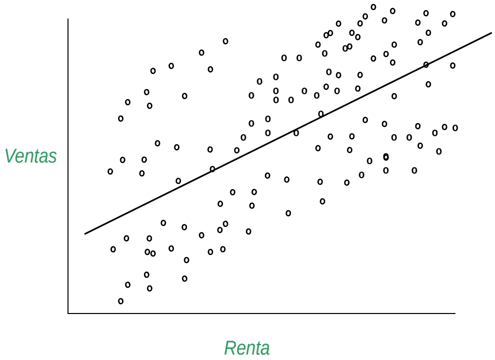
Una cadena de perfumerías que tiene tiendas en Francia, España e Italia quiere analizar la evolución de sus ventas. Para ello cuenta con datos de las ventas de 350 de sus tiendas para el año 2011, así como la renta promedio del municipio donde se encuentra la tienda.



Variables explicativas cualitativas.

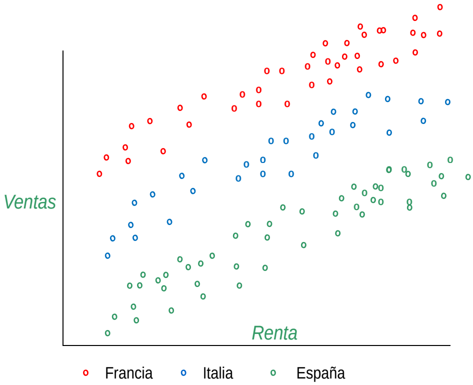
Si consideramos las 350 tiendas conjuntamente, el modelo sería:

$$V = f(R) \quad \Rightarrow \quad V_i = \beta_1 + \beta_2 \, R_i + u_i \qquad i = 1, 2, \ldots, 350$$



Variables explicativas cualitativas.

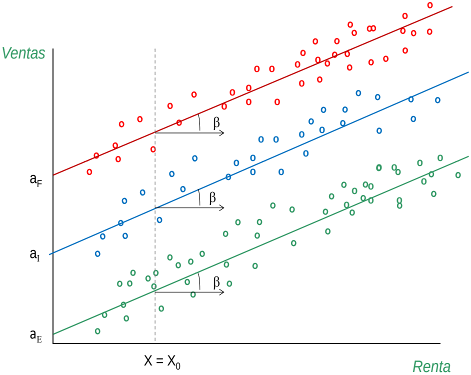
Consideremos el siguiente gráfico:



$\Rightarrow$ existen diferencias entre las ventas de los diferentes países.

$$V = f(Renta, País)$$

Variables explicativas cualitativas.



Es preciso introducir el efecto país en el modelo de regresión de forma que el efecto sobre las ventas de un incremento unitario en la renta sea el mismo en los tres países, pero, para una misma renta, el nivel de ventas sea diferente en cada país.

Variables explicativas cualitativas.

Las variables cualitativas se introducen en el modelo a través de variables ficticias.

Variable ficticia.

Una variable ficticia es una variable artificial binaria del tipo:

$$D_i = \begin{cases} 1 & \text{si la característica está presente en la observación i} \\ 0 & \text{en otro caso} \end{cases}$$

Variables explicativas cualitativas.

En principio, se construyen tantas variables ficticias como categorías tenga la variable cualitativa.

Ejemplo.

Número de categorías: 3 $\Rightarrow$ 3 Variables Ficticias

$$I_i = \begin{cases} 1 & i \in \text{ Italia} \\ 0 & \text{en otro caso} \end{cases}$$

$$F_i = \begin{cases} 1 & i \in \text{ Francia} \\ 0 & \text{en otro caso} \end{cases}$$

$$E_i = \begin{cases} 1 & i \in \text{ España} \\ 0 & \text{en otro caso} \end{cases}$$

Variables explicativas cualitativas.

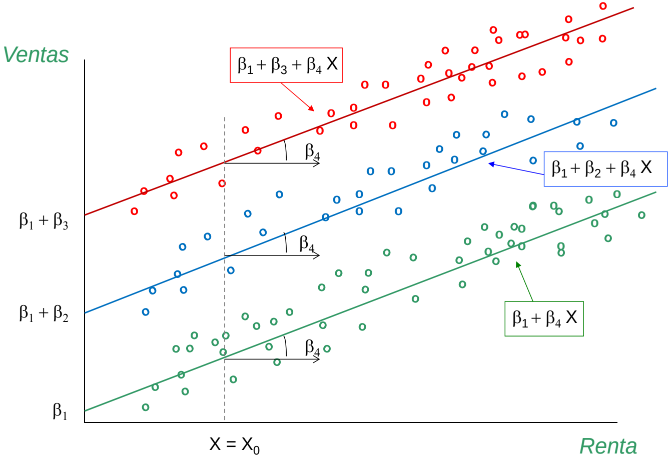
Especificación:

Se introducen tantas variables ficticias como categorías tiene la variable cualitativa menos 1, manteniendo el término constante.

$$V_i = \beta_1 + \beta_2\, I_i + \beta_3\, F_i + \beta_4\, R_i + u_i \quad i = 1, 2, ..., 350$$

$$\begin{aligned} V_i \,|\, i \,\epsilon \text{ Italia } &= \quad \beta_1 + \beta_2 \, + \beta_4\, R_i + u_i \\ V_i \,|\, i \,\epsilon \text{ Francia } &= \quad \beta_1 + \beta_3 + \beta_4\, R_i + u_i \\ V_i \,|\, i \,\epsilon \text{ España } &= \quad \beta_1 + \beta_4\, R_i + u_i \end{aligned}$$

Representación gráfica.



Variables explicativas cualitativas

¿Por qué no introducimos todas las variables ficticias?

$$V_i = \beta_1 + \beta_2\, I_i + \beta_3\, F_i + \beta_4\, E_i + \beta_5\, R_i + u_i \quad i = 1, 2, ..., 350$$

$$V_i \,|\, i \,\epsilon \;\; \text{Italia} = \quad \beta_1 + \beta_2 \; + \beta_5\, R_i + u_i$$

$$V_i \,|\, i \,\epsilon \;\; \text{Francia} = \quad \beta_1 + \beta_3 + \beta_5\, R_i + u_i$$

$$V_i \,|\, i \,\epsilon \;\; \text{España} = \quad \beta_1 + \beta_4 + \beta_5\, R_i + u_i$$

En este modelo con término independiente más las tres variables ficticias, los coeficientes $\beta_1, \beta_2, \beta_3, \beta_4$ **no están identificados** porque $F_i + I_i + E_i = 1, \forall i$.

Supuesto. Ausencia de colinealidad perfecta

No existen combinaciones lineales entre los regresores del modelo.

Gestión de las variables ficticias en Gretl.

EJEMPLO 4.2. Diseñar variables ficticias en Gretl.

1. Introducir la variable manualmente.
   Aplicación en el **Ejemplo 4.2.1.**

2. Generar la variable a partir de una variable discreta.
   Aplicación en el **Ejemplo 4.2.2.**

3. Definir la variable para un rango de observaciones.
   Aplicación en el **Ejemplo 4.2.3.**

4. Variables ficticias incluidas en Gretl. Aplicación en el **Ejemplo 4.2.4.**

5. La variable tendencia. Aplicación en el **Ejemplo 4.2.5.**

Contenido

Especificación del MRLG.

$$\boxed{Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \quad i = 1, 2, \ldots, N}$$

- Y: variable a explicar o endógena.
- $X_j \;\; j = 1, \ldots, k$: variables explicativas cuantitativas y/o variables ficticias.
- $\beta_j \;\; j = 1, \ldots, k$: coeficientes a estimar.
- u es la perturbación aleatoria (no observable).
- N es el tamaño muestral.

Especificación del MRLG.

Parte sistemática: $\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}$

- Recoge todos los factores relevantes y todos los factores que incluye son relevantes.
- Refleja el comportamiento promedio de Y condicionado a X en la población:

$$E_X(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} \quad \Rightarrow \quad E_X(u_i) = 0$$

Parte aleatoria:

u es una variable aleatoria no observable que quiere recoger:

- Efectos no incluidos en la parte sistemática del modelo.
- Comportamiento aleatorio de los agentes económicos.
- Errores de medida.

Especificación del MRLG. Supuestos básicos.

Supuesto S1.

El modelo de regresión lineal general (MRLG) poblacional se expresa como:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, .., N$$

donde:

- Las variables explicativas son estocásticas.
- $\beta_1, \beta_2, \ldots, \beta_k$ son parámetros desconocidos constantes.
- El modelo es lineal en los coeficientes.
- El modelo está correctamente especificado, es decir, todos los factores relevantes están incluidos en el modelo y todos los factores incluidos en el modelo son relevantes.

Especificación del MRLG. Supuestos básicos.

S2. Ausencia de colinealidad.

En la muestra, ningún regresor es constante ni puede haber combinaciones lineales exactas entre los regresores.

Sobre la perturbación aleatoria

S3. Media condicionada cero. $E(u_i|X_2, X_3, \ldots, X_k) = 0 \quad \forall i = 1, 2, \ldots, N$

S4. Homocedasticidad. La varianza de la perturbación es constante.

$$\text{Var}(u_i|X_2, X_3, \ldots, X_k) = \sigma^2 \quad \forall i = 1, 2, \ldots, N$$

S5. Ausencia de Autocorrelación: $\text{Cov}(u_i, u_j|X_2, X_3, \ldots, X_k) = 0 \quad \forall i \neq j$.

S6. Normalidad: Las pertubaciones u_j son independientes de las variables explicativas y están idénticamente y normalmente distribuidas.

$$\boxed{u_i \sim NID(0, \sigma^2) \quad \forall i = 1, \ldots, N}$$

Interpretación de los coeficientes.

Dados los supuestos del MRLG, tenemos que:

$$\begin{aligned} E_X(Y_i) &= E_X(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i) \\ &= \beta_1 + \beta_2\, X_{2i} + \ldots + \beta_k\, X_{ki} + \underbrace{E_X(u_i)}_{=0} \end{aligned}$$

$$E_X(Y_i) = \beta_1 + \beta_2\, X_{2i} + \ldots + \beta_k\, X_{ki}$$

$E_X(Y_i)$ es la *Función de Regresión Poblacional* (FRP).

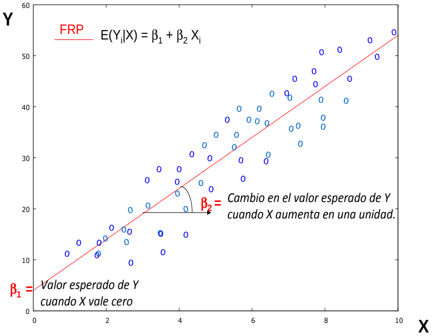
Modelo de regresión lineal simple.

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \text{FRP:} \;\; E(Y_i|X) = \beta_1 + \beta_2 X_i$$

donde:

β_1 : ordenada de la recta de regresión poblacional.

β_2 : pendiente de la recta de regresión poblacional.



Interpretación de los coeficientes

Modelo de regresión lineal general.

Supongamos que todos los regresores son variables explicativas cuantitativas.

- $\beta_j, j = 2, 3, \ldots, k$: cambio (incremento o decremento) en el valor esperado de Y cuando X_j aumenta en una unidad, manteniendo el resto de las variables explicativas constantes:

$$\beta_j = \frac{\Delta E_X(Y)}{\Delta X_j = 1} \quad \text{Efecto marginal de } X_j \text{ sobre } Y.$$

- $\beta_1 = E[Y_i | X_{2i} = 0, \ldots, X_{ki} = 0]$

  Es el valor esperado de Y cuando todas las variables explicativas toman el valor cero.

Interpretación de los coeficientes.

1. Para distintas formas funcionales.

Forma funcional	Efecto marginal	Elasticidad $= \frac{\Delta E_X(Y)/Y}{\Delta X/X}$
Lineal $Y_i = \beta_1 + \beta_2 X_i + u_i$	β_2	$\beta_2 \frac{X}{Y}$
Cuadrática $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$	$\beta_2 + 2\beta_3 X$	$(\beta_2 + 2\beta_3 X)\dfrac{X}{Y}$
log-log $\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$	$\beta_2 \dfrac{Y}{X}$	β_2
log-lin $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$	$\beta_2\, Y$	$\beta_2\, X$
lin-log $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$	$\beta_2 \dfrac{1}{X}$	$\beta_2 \dfrac{1}{Y}$

Interpretación de los coeficientes.

2. Variables explicativas cualitativas.

   Ejemplo de las tiendas de perfumería:

$$V_i = \beta_1 + \beta_2\, I_i + \beta_3\, F_i + \beta_4\, E_i + \beta_5\, R_i + u_i \quad i = 1, 2, ..., 350$$

Función de regresión poblacional.

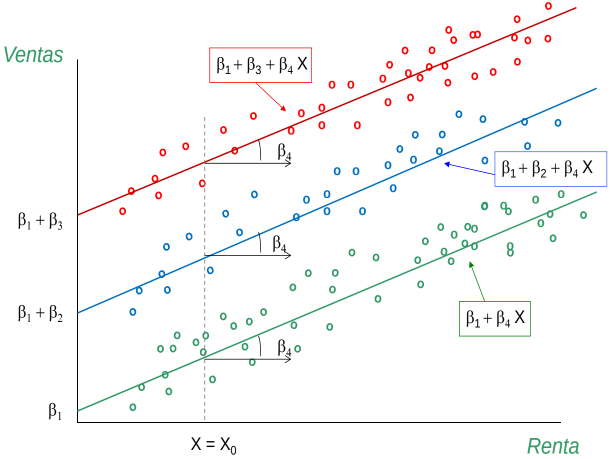
$$E(V_i|X_i) = E_X(\beta_1 + \beta_2\, I_i + \beta_3\, F_i + \beta_4\, R_i + u_i) = \beta_1 + \beta_2\, I_i + \beta_3\, F_i + \beta_4\, I_i$$

España: $E(V_i|R_i, I_i = 0, F_i = 0) = \beta_1 + \beta_4\, R_i.$

Italia: $E(V_i|R_i, I_i = 1, F_i = 0) = (\beta_1 + \beta_2) + \beta_4\, R_i$

Francia: $E(V_i|R_i, I_i = 0, F_i = 1) = (\beta_1 + \beta_3) + \beta_4\, R_i$

Representación gráfica.



Interpretación de los coeficientes.

Coeficiente de la variable renta.

▷ $\beta_4 =$ Cambio en el valor esperado de las ventas cuando la renta aumenta en una unidad, manteniendo la variable país constante.

Término independiente.

▷ $\beta_1 = E(V_i | R_i = 0, I_i = 0, F_i = 0)$

Valor esperado de las ventas en las tiendas españolas cuando el valor de la variable renta es cero.

Interpretación de los coeficientes.

Coeficientes asociados a las variables ficticias.

▷ β_2, β_3 no tienen una interpretación de pendiente porque las variables ficticias no son variables continuas, sino que son variables discretas.

Para España: $E(V_i|R_i, F_i = 0, I_i = 0) = \beta_1 + \beta_4\, R_i$

Para Italia: $E(V_i|R_i, I_i = 1, F_i = 0) = (\beta_1 + \beta_2) + \beta_4\, R_i$

$$\Longrightarrow \quad \beta_2 = E(V_i|R_i, I_i = 1, F_i = 0) - E(V_i|R_i, I_i = 0, F_i = 0)$$

$\beta_2 =$ Diferencia en el valor esperado de las ventas entre las tiendas italianas y las españolas, manteniendo la variable renta constante.

Interpretación de los coeficientes

Para España:

$$E(V_i|R_i, F_i = 0, I_i = 0) = \beta_1 + \beta_4\, R_i$$

Para Francia:

$$E(V_i|R_i, I_i = 0, F_i = 1) = (\beta_1 + \beta_3) + \beta_4\, R_i$$

$$\Longrightarrow \quad \beta_3 = E(V_i|R_i, I_i = 0, F_i = 1) - E(S_i|R_i, I_i = 0, F_i = 0)$$

$\beta_3 =$ Diferencia en el valor esperado de las ventas entre las tiendas francesas y las españolas, manteniendo la variable renta constante.