ECONOMETRIA I.
Departamento de Fundamentos del Análisis Económico
Universidad de Alicante. Curso 2011/12

GUIÓN TEMA 1. EL MODELO DE REGRESIÓN LINEAL (MRL)

1.1 INTRODUCCIÓN

1.2. FORMULACIÓN DEL MRL: especificación del modelo e hipótesis básicas

Bibliografía Apartado 1.2
Greene, 6.2 y 6.3;
A. Fernández Gallastegui, 2.3) y 3.1
J.M. Wooldridge. 2.1, 2.2, 3.1, 3.2 (hasta pag. 87), pag 102

1.2.1 Especificación del modelo

Objetivo: Cuantificar la relación existente entre una variable Y y un conjunto de k variables explicativas $X_1, ..., X_k \rightarrow Y = f(X_1, ..., X_k)$, mediante una relación lineal y usando una muestra de tamaño T. La relación entre la variable Y y las variables explicativas no es exacta, por lo que se incluye un término de error aleatorio, u. El modelo que se plantea es sólo una aproximación al verdadero modelo, por omisión de variables relevantes, errores de medida en las variables, aleatoriedad en el comportamiento humano, etc. El término de error es una variable aleatoria que recoge el efecto de diversos aspectos: la aleatoriedad en el comportamiento humano, variables no incluidas en el modelo, errores de medida, imposibilidad de capturar todas las influencias que existen sobre la variable dependiente.

Notación:

$(a)\ Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t, \qquad t = 1, 2, ..., T$

(b) Notación vectorial:

$Y_t = X_t'\beta + u_t, \qquad t = 1, 2, ..., T$; donde $X_t = (1, X_{2t}, ..., X_{kt})'$

(c) Notación matricial:

$Y = X\beta + u$, donde

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_T \end{bmatrix}, \underset{(T\times k)}{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & & X_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2T} & \cdots & X_{kT} \end{bmatrix}, \underset{(k\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_k \end{bmatrix},$$

$$\underset{(T\times 1)}{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_T \end{bmatrix}$$

1.2.2 Hipótesis básicas:

1. las $X's$ no son variables aleatorias. Esto es, X es una matriz de constantes conocida.

2. rg$(X) = k$; es decir, X tiene rango completo de columnas, por lo que $\exists$ $(X'X)^{-1}$. En otros términos, no hay relaciones lineales exactas entre las variables (esto es, no hay multicolinalidad exacta).

3. Hipótesis sobre los coeficientes: Los coeficientes son constantes a lo largo de la muestra y aparecen en el modelo de forma lineal (Hipótesis de linealidad).

4. $E(u_t) = 0$, $t = 1, 2, ..T \Leftrightarrow E(u) = 0$

5. $\left\{ \begin{array}{l} var(u_t) = E(u_t^2) = \sigma^2, \ t = 1, 2, ..T \text{ (homocedasticidad)} \\ cov(u_t, u_s) = E\left[u_t u_s\right] = 0, \forall t \neq s. \text{(ausencia de correlación serial)} \end{array} \right\}$

   $\Leftrightarrow \ Var(u) = E(uu') = \sigma^2 I_T$

6. Hipótesis adicional de normalidad: $u \sim N(0, \sigma^2 I)$; $u_t \sim N(0, \sigma^2)$.

Las hipótesis 4 a 6 se pueden escribir también en función de la variable dependiente

4. $\Leftrightarrow E(Y_t) = X_t'\beta,\ t = 1, 2, ..T \Leftrightarrow E(Y) = X\beta$

5. $\Leftrightarrow \left\{ \begin{array}{l} var(Y_t) = \sigma^2,\ t = 1, 2, ..T \text{ (homocedasticidad)} \\ cov(Y_t, Y_s) = 0, \forall t \neq s. \text{(ausencia de correlación serial)} \end{array} \right\} \Leftrightarrow\ Var(Y) = \sigma^2 I_T$

6. $\Leftrightarrow Y \sim N(X\beta, \sigma^2 I)$

NOTAS:

1. Cuando hablemos de las hipótesis básicas del MRL nos estamos refiriendo a los supuestos 1-5, y cuando hablemos del MRL con errores normales nos referimos a los supuestos 1-6.

2. A partir de la esperanza poblacional de la variable dependiente $E(Y_t) = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt}$, es fácil interpretar los coeficientes de cada una de las variables explicativas. Así pues para $j = 2, ..., k$

$$\beta_j = \frac{\partial E(Y_t)}{\partial X_{jt}}$$

   mide el efecto marginal de cada variable independiente sobre la esperanza poblacional de la variable dependiente. Por lo tanto, β_j puede interpretarse como el cambio en el valor esperado de Y ante un aumento de X_j en una unidad, manteniendo constante el resto de las variables explicativas (ceteris paribus). El término constante puede interpretarse como la media de la variable dependiente Y cuando todas las variables explicativas tomen el valor cero. Así, si $X_{2t} = ... = X_{kt} = 0$,

$$E(Y_t) = \beta_1$$

   En muchos modelos, puede no tener sentido que todas las variables explicativas sean cero, en cuyo caso, la constante carecerá de interpretación.

3. En algunos casos estaremos interesados en especificar un modelo de regresión sin término constante

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t . t = 1, 2, ..., T$$

este modelo se denomina "modelo de regresión por el origen". Todo lo que vamos a ver en la sección 1.3 sobre estimación del modelo de regresión es válido tanto para el modelo de regresión por el origen como para el modelo de regresión con término constante. Sin embargo, algunas de las propiedades del ajuste que veremos en la sección 1.4 no se verifican para el modelo de regresión por el origen.

1.3 ESTIMACIÓN MCO

Bibliografía apartado 1.3:
Greene, 6.4.1, 6.4.2,
A. Fernández Gallastegui, 2.1.1, 3.2, 3.4,
J.M. Wooldridge. 2.2, 3.2, y Apéndice E.1

1.3.1 Estimación MCO en el modelo de regresión simple

- Consideremos el modelo de regresión simple

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

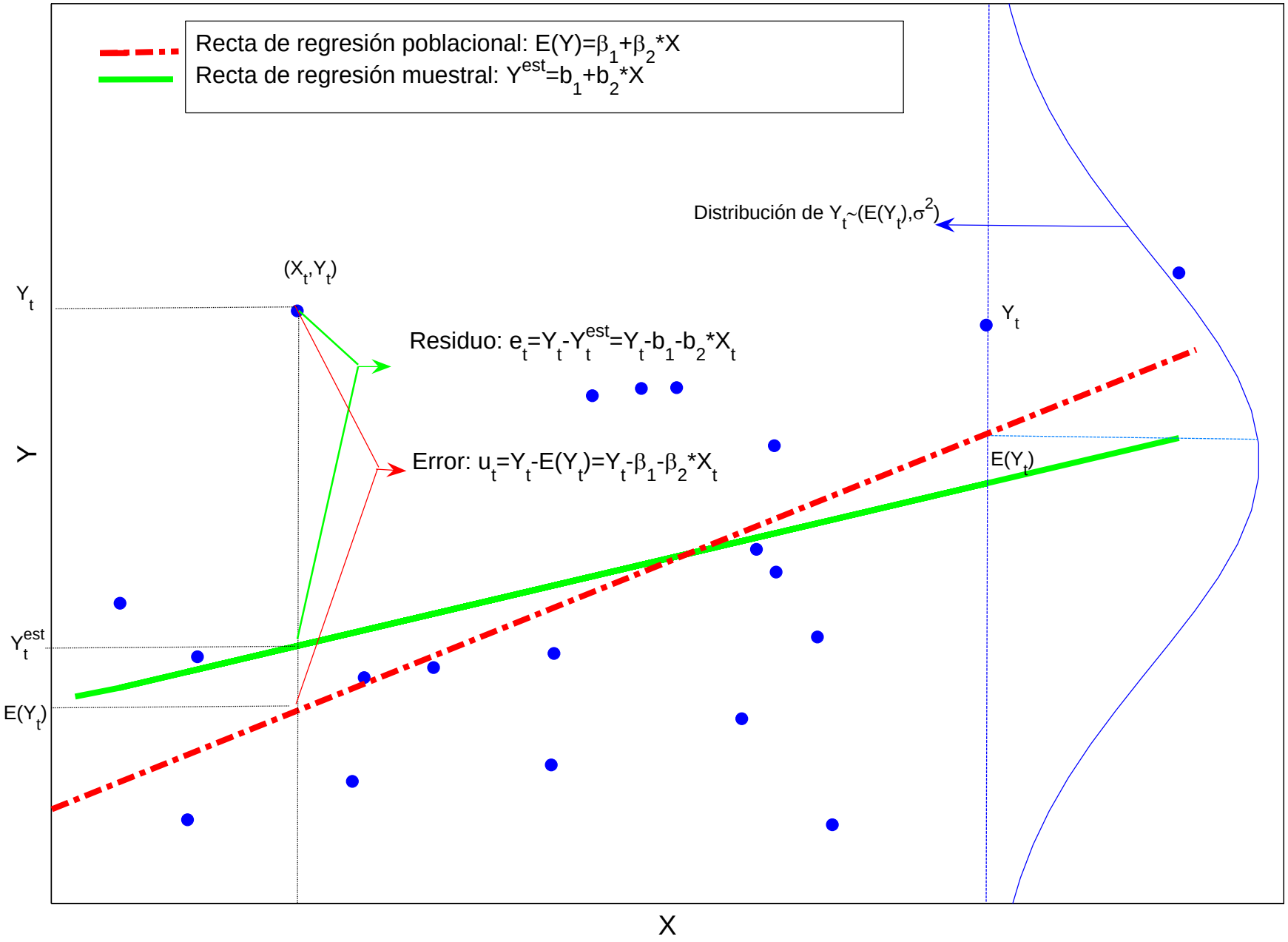
y dibujemos la nube de puntos asociada a una determinada muestra de tamaño T y una recta cualquiera

$$y = b_1 + b_2 x$$

El método de estimación por Mínimos Cuadrados Ordinarios (MCO) consiste en elegir los valores de b_1 y b_2 de forma que la recta esté lo más "próxima" posible a los puntos de la nube según un determinado criterio de proximidad. En concreto el criterio MCO consiste en minimizar la suma de los cuadrados de las distancias verticales de los puntos de la nube a la recta de regresión.

Gráficamente podemos ver que la distancia vertical del punto (X_t, Y_t) a la recta $y = b_1 + b_2 x$ viene dada por

$$Y_t - b_1 - b_2 X_t$$


Recta de regresión poblacional: E(Y)=β1+β2*X
Recta de regresión muestral: Y^est=b1+b2*X
Distribución de Yt~(E(Yt),σ2)
(Xt,Yt)
Yt
Residuo: et=Yt-Yt^est=Yt-b1-b2*Xt
Error: ut=Yt-E(Yt)=Yt-β1-β2*Xt
Yt
E(Yt)
Yt^est
E(Yt)
Y
X

y por tanto la función objetivo que tenemos que minimizar es

$$s(b_1, b_2) = \sum_{t=1}^{T} (Y_t - b_1 - b_2 X_t)^2$$

Los coeficientes estimados se obtienen igualando a cero las derivadas parciales de la función objetivo. Las derivadas parciales vienen dadas por:

$$\begin{aligned} \frac{\partial s(b_1, b_2)}{\partial b_1} &= -2 \sum_{t=1}^{T} (Y_t - b_1 - b_2 X_t) \\ \frac{\partial s(b_1, b_2)}{\partial b_2} &= -2 \sum_{t=1}^{T} (Y_t - b_1 - b_2 X_t) X_t \end{aligned}$$

Igualando a cero y simplificando obtenemos las condiciones de primer orden que se denominan ecuaciones normales

$$\begin{aligned} T\widehat{\beta}_1 + \widehat{\beta}_2 \sum_{t=1}^{T} X_t &= \sum_{t=1}^{T} Y_t \\ \widehat{\beta}_1 \sum_{t=1}^{T} X_t + \widehat{\beta}_2 \sum_{t=1}^{T} X_t^2 &= \sum_{t=1}^{T} Y_t X_t \end{aligned}$$

y despejando obtenemos las expresiones de los estimadores MCO de β_1 y β_2

$$\begin{aligned} \widehat{\beta}_2 &= \frac{S_{XY}}{S_X^2} \\ \widehat{\beta}_1 &= \overline{Y} - \widehat{\beta}_2 \overline{X} \end{aligned}$$

donde $\overline{X}$ es la media muestral de las observaciones para X, $\overline{Y}$ es la media muestral de las observaciones para Y, S_{XY} es la covarianza muestral entre X e Y, y S_X es la varianza muestral de X, es decir

$$\overline{X} = \tfrac{1}{T} \sum_{t=1}^{T} X_t \qquad \overline{Y} = \tfrac{1}{T} \sum_{t=1}^{T} Y_t$$

$$S_{XY} = \tfrac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})(Y_t - \overline{Y}) \quad S_X^2 = \tfrac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})^2$$

- Las distancias verticales de los puntos a la recta de regresión

$$e_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t, \quad t = 1, 2, .., T$$

se denominan residuos MCO.

- Los valores estimados, valores ajustados o predicciones para la variable dependiente en función del modelo de regresión los denotaremos por:

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_t, \quad t = 1, 2, .., T$$

- Notése que podemos calcular la predicción para la variable dependiente para cualquier valor de X aunque dicho valor no corresponda con ninguno de los valores observados en la muestra.

- Nótese que las distancias verticales de los puntos de la nube a la recta definidas por

$$Y_t - b_1 - b_2 X_t$$

pueden ser positivas o negativas, y por tanto un criterio que consistiera en minimizar la suma de las distancias no sería apropiado.

- Podríamos considerar otros criterios alternativos como por ejemplo minimizar la suma de los valores absolutos de las distancias verticales

$$\min_{b_1, b_2} \sum_{t=1}^{T} |Y_t - b_1 - b_2 X_t|$$

El problema de utilizar este criterio es que no es diferenciable y por tanto es más complicado calcular el mínimo.

1.3.2 Estimación MCO en el modelo de regresión múltiple

- La idea intuitiva de la estimación MCO del modelo de regresión múltiple es análoga al caso de regresión simple. La función objetivo que tenemos que minimizar es

$$s(b_1, b_2, ..b_k) = \sum_{t=1}^{T} (Y_t - b_1 - b_2 X_{2t} - .. - b_K X_{kt})^2$$

y los coeficientes estimados se obtienen igualando a cero las derivadas parciales de la función objetivo. Las derivadas parciales vienen dadas por:

$$\begin{aligned}
\frac{\partial s(b_1, b_2, ..b_k)}{\partial b_1} &= -2\sum_{t=1}^{T}(Y_t - b_1 - b_2X_{2t} - .. - b_KX_{kt}) \\
\frac{\partial s(b_1, b_2, ..b_k)}{\partial b_2} &= -2\sum_{t=1}^{T}(Y_t - b_1 - b_2X_{2t} - .. - b_KX_{kt})X_{2t} \\
&\vdots \\
\frac{\partial s(b_1, b_2, ..b_k)}{\partial b_k} &= -2\sum_{t=1}^{T}(Y_t - b_1 - b_2X_{2t} - .. - b_KX_{kt})X_{Kt}
\end{aligned}$$

Igualando a cero y simplificando obtenemos las condiciones de primer orden (ecuaciones normales)

$$\begin{aligned}
\widehat{\beta}_1 T + \widehat{\beta}_2\sum_{t=1}^{T}X_{2t} + \widehat{\beta}_3\sum_{t=1}^{T}X_{3t} + ... + \widehat{\beta}_k\sum_{t=1}^{T}X_{kt} &= \sum_{t=1}^{T}Y_t \\
\widehat{\beta}_1\sum_{t=1}^{T}X_{2t} + \widehat{\beta}_2\sum_{t=1}^{T}X_{2t}^2 + \widehat{\beta}_3\sum_{t=1}^{T}X_{2t}X_{3t} + ... + \widehat{\beta}_k\sum_{t=1}^{T}X_{2t}X_{kt} &= \sum_{t=1}^{T}X_{2t}Y_t \\
&\cdots \\
\widehat{\beta}_1\sum_{t=1}^{T}X_{kt} + \widehat{\beta}_2\sum_{t=1}^{T}X_{2t}X_{kt} + \widehat{\beta}_3\sum_{t=1}^{T}X_{3t}X_{kt} + ... + \widehat{\beta}_k\sum_{t=1}^{T}X_{kt}^2 &= \sum_{t=1}^{T}X_{kt}Y_t
\end{aligned}$$

A la hora de obtener la expresión del estimador MCO de los parámetros $\beta_1, \beta_2, .., \beta_k$ es más sencillo reescribir el sistema en notación matricial. Se puede demostrar que el sistema de ecuaciones normales se puede escribir como

$$X'X\widehat{\beta} = X'Y$$

y como X es de rango completo:

$$\widehat{\beta} = (X'X)^{-1}X'\dot{Y}.$$

donde

$$\underset{(k\times 1)}{\widehat{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \cdots \\ \widehat{\beta}_k \end{bmatrix}$$

X es la matriz de observaciones de las variables explicativas

$$\underset{(T\times k)}{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & & X_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2T} & \cdots & X_{kT} \end{bmatrix}$$

e Y es el vector de observaciones para la variable dependiente

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_T \end{bmatrix}$$

- Análogamente al caso del modelo de regresión simple los residuos MCO se definen como

  $$e_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - .. - \widehat{\beta}_K X_{kt}, \quad t = 1, 2, .., T$$

  y el vector de residuos MCO es

  $$\underset{(T\times 1)}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_T \end{bmatrix}$$

- Análogamente al caso del modelo de regresión simple, los valores estimados, valores ajustados o predicciones para la variable dependiente en función del modelo de regresión los denotaremos por:

  $$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + .. + \widehat{\beta}_K X_{kt}, \quad t = 1, 2, .., T$$

- También, igual que en el modelo de regresión simple, podemos calcular la predicción para la variable dependiente para cualquier valor de $X_2, .., X_k$ aunque dichos valores no correspondan con ninguno de los valores observados en la muestra.

1.3.3 Interpretación de los parámetros estimados, unidades de medida y forma funcional (Wooldridge, 2.4)

- Los valores estimados para las pendientes miden los efectos marginales estimados de cada variable sobre la variable dependiente ya que

$$\frac{\partial \widehat{Y}}{\partial X_j} = \widehat{\beta}_j, \quad j = 2, ..k$$

Por tanto $\widehat{\beta}_j$ mide el efecto que tendría sobre la variable dependiente un aumento en una unidad en X_j, manteniendo constante las restantes variables explicativas del modelo. La característica fundamental del modelo de regresión lineal es que los efectos marginales son constantes.

Ejemplo:

Consideremos el siguiente modelo para el gasto en vestido y calzado estimado en base a una muestra de 7427 hogares españoles

$$\widehat{gvest}_t = 1.2 + 0.064renta_t + 0.132nad_t + 0.159nhijos_t$$

donde $gvest$ es el gasto anual del hogar en vestido y calzado (en miles Euros), $renta$ es la renta anual del hogar (en miles de Euros), nad es el número de adultos en el hogar y $nhijos$ es el número de hijos menores de 18 años. Según este modelo, un aumento en 1000 Euros en la renta anual del hogar produciria un aumento estimado de 64 Euros (0.064 miles de Euros) al año en vestido y calzado, manteniendo constante el número de adultos y el número de hijos en el hogar. Un adulto adicional en el hogar supone un aumento estimado de 132 Euros (0.132 miles de Euros) en el gasto en vestido y calzado, si la renta anual no ha variado y no ha cambiado el número de hijos en el hogar. Un hijo más supone un aumento de 159 Euros (0.159 miles de Euros) en el gasto en vestido y calzado, manteniendo constante la renta anual y el número de adultos en el hogar.

- Utilizando el modelo estimado también podemos calcular las diferencias estimadas en la variable dependiente entre "individuos" con distintos valores para las variables explicativas. Siguiendo con el ejemplo anterior, según el modelo estimado, la diferencia en el gasto en vestido y calzado entre dos

familias con la misma renta, la familia A formada por una pareja con un hijo menor de 18 años y la familia B formada por una pareja con un hijo adulto es:

Predicción para la familia A

$$\widehat{gvest}_A = 1.2 + 0.064 renta_A + 0.132 * 2 + 0.159$$

Predicción para la familia B

$$\widehat{gvest}_B = 1.2 + 0.064 renta_B + 0.132 * 3$$

puesto que la renta de las dos familias coincide, la diferencia estimada en el gasto en vestido y calzado es

$$0.132 * 2 + 0.159 - 0.132 * 3 = 0.027$$

Es decir la familia A gastaría 27 Euros más que la B

- Es importante destacar que la interpretación de los coeficientes estimados depende de las variables explicativas incluidas en la regresión, porque los efectos se miden manteniendo constantes el resto de las variables incluidas en la regresión. Así pues, no es de extrañar que si en el ejemplo anterior no incluimos el número de adultos en la regresión, la estimación cambia de la siguiente forma

$$\widehat{gvest}_t = 1.56 + 0.067 renta_t + 0.121 nhijos_t$$

puesto que los coeficientes estimados están midiendo cosas distintas. Así pues, en el modelo anterior, el coeficiente estimado para $nhijos$ mide la disminución en el gasto en vestido si el número de hijos disminuye en una unidad (por ejemplo, si los hijos se mudan a vivir fuera del hogar), permaneciendo constante la renta anual y el número de adultos. Esta disminución en el modelo anterior se estima que es de 159 euros. Sin embargo, en esta ultima estimación, el efecto marginal por un hijo menos es una disminución en el gasto en vestido y calzado de 121 euros. La diferencia es que este último efecto marginal sólo mantiene constante la renta anual mientras que el número de adultos puede cambiar cuando varía el número de hijos. En estos datos, cuando un hijo cumple 18 años se considera que es un adulto y por tanto para esta familia se verá que $-\Delta nhijos_t = \Delta nad_t = 1$. Así pues,

el efecto estimado de 121 euros mide el impacto sobre el gasto en vestidos de tener un hijo menos, tanto si se queda en la familia como adulto o si se muda a vivir fuera del hogar. Ello explica por qué el efecto es menor que en la estimación anterior, porque si se queda como adulto la disminución en el gasto es obviamente menor. La estimación de este efecto está estimando el promedio del efecto para los dos tipos de casos.

- En este modelo, la estimación de la constante carece de una interpretación útil, pues indica que el gasto anual en vestido y calzado predicho para una familia sin ningún adulto, ningún hijo y con renta anual igual a cero es de 1200 euros. No es posible que exista un hogar con tales valores de las variables explicativas. Un ejemplo en el cuál la constante tiene interpretación es un modelo de regresión del salario sobre la experiencia laboral del individuo. Sea $salario_t$ el salario por hora del individuo t y exp_t los años de experiencia laboral del individuo t. Consideremos el siguiente modelo estimado

$$\widehat{salario}_t = 6.2 + 1.1exp_t$$

  La interpretación del término constante nos dice que los trabajadores sin ningún año de experiencia laboral, ganan en promedio un salario por hora de 6.2 euros. Por cada año adicional de experiencia en el mercado laboral, su salario por hora aumenta en 1.1 euros.

- Si cambiamos las unidades de medida de alguna o algunas de las variables explicativas y/o de la variable dependiente, en general, variarán los valores estimados de los parámetros. Sin embargo, podemos calcular los nuevos valores de los parámetros estimados sin tener que volver a estimar el modelo. Consideremos el modelo estimado

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + .. + \widehat{\beta}_K X_{kt}$$

  Si ahora medimos la variable X_2 en otras unidades distintas $X_2^* = dX_2$, y substituimos en el modelo estimado $X_2 = \frac{X_2^*}{d}$ tenemos

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 \frac{X_{2t}^*}{d} + .. + \widehat{\beta}_K X_{kt} = \widehat{\beta}_1 + \widehat{\beta}_2^* X_{2t}^* + .. + \widehat{\beta}_K X_{kt}$$

  donde $\widehat{\beta}_2^* = \frac{\widehat{\beta}_2}{d}$. Por tanto el coeficiente estimado de X_2^* será igual al coeficiente estimado de X_2 dividido por d, mientras que la constante del modelo

estimado y los coeficientes estimados de las restantes variables no cambian cuando cambiamos las unidades de medida de X_2.

Si ahora medimos la variable dependiente en otras unidades distintas $Y^* = cY$, si substituimos en el modelo estimado $Y = \frac{Y^*}{c}$ tendremos

$$\widehat{Y}_t^* = c\widehat{\beta}_1 + c\widehat{\beta}_2 X_{2t} + .. + c\widehat{\beta}_K X_{kt} = \widehat{\beta}_1^* + \widehat{\beta}_2^* X_{2t} + .. + \widehat{\beta}_K^* X_{kt}$$

donde $\widehat{\beta}_1^* = c\widehat{\beta}_1, \widehat{\beta}_2^* = c\widehat{\beta}_2, .., \widehat{\beta}_k^* = c\widehat{\beta}_k$, y por tanto todos los nuevos coeficientes estimados serán iguales a los coeficientes estimados que teniamos anteriormente multiplicados por c.

Siguiendo con el ejemplo anterior, si ahora medimos la renta y el gasto en Euros (en lugar de en miles de Euros como antes) el modelo estimado será

$$\widehat{gvest}_t = 1200 + 0.064 renta_t + 132 nad_t + 159 nhijos_t$$

- Nótese que la interpretación de los coeficientes no cambia cuando hacemos un cambio de unidades.
- Dentro del contexto del modelo de regresión lineal (modelo lineal en parámetros) podemos considerar relaciones no lineales entre las variables de interés. Los ejemplos de relaciones no lineales que aparecen con más frecuencia en Economía son:
  - Modelo lineal en logarítmos:

    $$\widehat{\log(Y_t)} = \widehat{\beta}_1 + \widehat{\beta}_2 \log(X_{2t}) + .. + \widehat{\beta}_K \log(X_{kt})$$

    ahora los $\widehat{\beta}_j$, $j = 2, .., k$ son las elasticidades estimadas, es decir $\widehat{\beta}_j$ mide la variación en tanto por ciento estimada para la variable dependiente ante un aumento de un 1% en la variable explicativa X_j. La característica de este modelo es que las elasticidades son constantes.
  - Modelo semilogarítmico log-nivel:

    $$\widehat{\log(Y_t)} = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + .. + \widehat{\beta}_K X_{kt}$$

    ahora, para $j = 2, ..K$, $100 * \widehat{\beta}_j$ mide la variación en tanto por ciento estimada para la variable dependiente ante un aumento en una unidad en la variable explicativa X_j.

– Ejemplo

Consideremos ahora el siguiente modelo para el gasto en vestido y calzado estimado en base a la misma muestra del ejemplo anterior

$$\widehat{\log(gvest_t)} = -1.06 + 0.49\log(renta_t) + 0.042nad_t + 0.088nhijos_t$$

donde $gvest$ es el gasto anual del hogar en vestido y calzado (en miles Euros), $renta$ es la renta anual del hogar (en miles de Euros), nad es el número de adultos en el hogar y $nhijos$ es el número de hijos menores de 18 años. Según este modelo, un aumento de un 1% en la renta anual del hogar produciría un aumento estimado de un 0.49% en el gasto en vestido y calzado. Un adulto adicional en el hogar supone un aumento estimado de 4.2% en el gasto en vestido y calzado, mientras que un hijo más supone un aumento del 8.8% en el gasto en vestido y calzado.

– Modelo semilogarítmico nivel-log:

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2\log(X_{2t}) + .. + \widehat{\beta}_K\log(X_{kt})$$

ahora, para $j = 2, ..K$, $\frac{\widehat{\beta}_j}{100}$ mide la variación en unidades estimada para la variable dependiente ante un aumento en un 1% en la variable explicativa X_j.

– Modelo polinomial.

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_t + \widehat{\beta}_3 X_t^2 + .. + \widehat{\beta}_k X_t^{k-1}$$

En este modelo los efectos marginales vendrían dados por

$$\frac{\partial\widehat{Y}}{\partial X} = \widehat{\beta}_2 + 2\widehat{\beta}_3 X + .. + (k-1)\widehat{\beta}_k X^{k-2}$$

Ejemplo

La relación entre el salario de los trabajadores y la edad no es generalmente lineal, ya que aunque el salario aumenta con la edad (al menos hasta una cierta edad) ese aumento no es constante. Muchos estudios consideran una relación cuadrática entre el salario y la edad. Consideremos el siguiente modelo para los salarios estimado en base a una muestra de 935 individuos

$$\widehat{salario_t} = -7.92 + 0.605educ_t + 0.357edad_t - 0.0022edad2_t$$

donde *salario* es el salario mensual en cientos de dolares, *educ* es el nivel de educación en años, *edad* es la edad en años y *edad*2 es la edad al cuadrado. Según el modelo estimado, el efecto marginal es

$$\frac{\partial \widehat{salario}}{\partial edad} = 0.357 - 2 * 0.0022 * edad$$

Asi, un año más de edad supone para un trabajador de 30 años un aumento en el salario mensual estimado de $0.357 - 2 * 0.0022 * 30 = 0.225$ cientos de dólares, es decir de 22.5 dólares, mientras que para un trabajador de 60 años supone un aumento de $0.357 - 2 * 0.0022 * 60 = 0.093$ cientos de dólares, es decir de 9.3 dólares. El efecto marginal de la edad sobre el salario disminuye con la edad.

- La forma de estimar el modelo no depende de cómo esten definidas las variables (en niveles, en logaritmos, etc.), pero si es muy importante tener en cuenta como están definidas para poder interpretar correctamente los resultados de la estimación.

- No todos los modelos se pueden tratar como modelos de regresión lineal. Por ejemplo el modelo $y = \frac{1}{\beta_1 + \beta_2 x} + u$ es intrínsicamente no lineal y para estimar los parámetros hay que utilizar técnicas econométricas más complejas que no se van a estudiar en este curso.

1.4 Propiedades del ajuste MCO

Bibliografía apartado 1.4:
Greene, 6.4.2 y 6.5
A. Fernández Gallastegui, 3.3 y 3.6. Apéndice 3.C
Wooldridge, 2.3 y páginas 87-91

En este apartado veremos propiedades algebraicas del ajuste MCO, que se cumplen siempre con independencia de que se verifiquen o no las hipótesis básicas.

- (a.1) $X'e = 0$. Los residuos son ortogonales a las variables explicativas.

  Demos

  $X'e = X'(Y - X\widehat{\beta}) = 0$

- (a.2) $\widehat{Y}'e = 0$. Los residuos son ortogonales a los valores ajustados.

  Demos

  $\widehat{Y} = X\widehat{\beta}$ y por tanto utilizando la propiedad (a.1) $\widehat{Y}'e = \widehat{\beta}'X'e = 0$

- (a.3) $Y'Y = \widehat{Y}'\widehat{Y} + e'e$

  Demos

  $$Y'Y = (\widehat{Y}' + e')(\widehat{Y} + e) = \widehat{Y}'\widehat{Y} + e'e + 2\widehat{Y}'e = \widehat{Y}'\widehat{Y} + e'e$$

  ya que por la propiedad (a.2) $\widehat{Y}'e = 0$.

- (a.4) $\sum_{t=1}^{T} e_t = 0$.

  Demos

  Puesto que la primera columna de la matriz X es una columna de unos, el primer elemento del vector $X'e$ es $\sum_{t=1}^{T} e_t$, y por tanto utilizando la propiedad (a.1) $\sum_{t=1}^{T} e_t = 0$.

- (a.5) El hiperplano estimado pasa por $(\overline{Y}, \overline{X}_2, ..., \overline{X}_k)$; es decir, $\overline{Y} = \widehat{\beta}_1 + \widehat{\beta}_2\overline{X}_2 + ... + \widehat{\beta}_k\overline{X}_k$.

  Demos

  Este resultado se obtiene dividiendo por T la primera ecuación normal

  $$\widehat{\beta}_1 T + \widehat{\beta}_2 \sum_{t=1}^{T} X_{2t} + \widehat{\beta}_3 \sum_{t=1}^{T} X_{3t} + ... + \widehat{\beta}_k \sum_{t=1}^{T} X_{kt} = \sum_{t=1}^{T} Y_t$$

  (a.6) $\overline{Y} = \overline{\widehat{Y}}$. Es decir, la media de la variable dependiente coincide con la media de los valores ajustados.

  Demos

  $$\left\{ \begin{array}{rc} \widehat{Y}_t = & \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + ... + \widehat{\beta}_k X_{kt} \\ \overline{Y} = & \widehat{\beta}_1 + \widehat{\beta}_2 \overline{X}_2 + \widehat{\beta}_3 \overline{X}_3 + ... + \widehat{\beta}_k \overline{X}_k \end{array} \right\} \rightarrow$$ Sumando y dividiendo por T en la primera de estas expresiones es inmediato ver que $\overline{\widehat{Y}} = \overline{Y}$

- Medidas de bondad de ajuste

  Buscamos ahora una medida que nos indique si el ajuste MCO de la nube de puntos es o no un buen ajuste, es decir, una medida que nos indique el grado de importancia de las discrepancias entre los valores de la variable dependiente observados y los ajustados.

  Definiciones:

  Suma Cuadrática Total:

  $SCT = \sum_{t=1}^{T}(Y_t - \overline{Y})^2 = \sum Y_t^2 - T\overline{Y}^2 = Y'Y - T\overline{Y}^2$;

  Suma Cuadrática Explicada:

  $SCE = \sum_{t=1}^{T}(\widehat{Y}_t - \overline{\widehat{Y}})^2 \underset{\text{como } \overline{Y}=\overline{\widehat{Y}}}{=} \sum_{t=1}^{T}(\widehat{Y}_t - \overline{Y})^2 = \sum \widehat{Y}_t^2 - T\overline{Y}^2 = \widehat{Y}'\widehat{Y} - T\overline{Y}^2$;

  Suma Cuadrática Residual:

  $SCR = \sum_{t=1}^{T} e_t^2 = e'e$.

  Los tres valores que acabamos de definir son no negativos, pues son sumas de cuadrados. SCT, SCE y SCR son medidas del grado de variabilidad de la variable dependiente, de los valores ajustados y de los residuos, respectivamente, pues son el numerador de la varianza muestral de cada una de estas variables (recuérdese que los valores ajustados tienen media $\overline{Y}$ y los residuos tienen media 0). Estas tres medidas están relacionadas entre sí, puesto que

  $$SCT = SCE + SCR$$

  como puede comprobarse fácilmente:

  $$SCT = Y'Y - T\overline{Y}^2 \underset{\text{Utilizando (a.3)}}{=} \widehat{Y}'\widehat{Y} - T\overline{Y}^2 + e'e = SCE + SCE$$

  Supondremos ahora que SCT no es nula, lo que equivale a decir que las observaciones de la variable dependiente no son todas iguales, dividiendo los tres sumandos de la igualdad anterior por SCT nos queda que:

  $$1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$$

Se define el coeficiente de determinación del modelo como

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

R^2 mide la proporción de la variabilidad de la variable dependiente que viene explicada por el modelo

El R^2 siempre cumple la siguiente condición:

$$0 \leq R^2 \leq 1$$

Es no negativo por serlo SCE y SCT, y es menor o igual a 1 porque SCR es no negativo. Es lógico que el coeficiente de determinación satisfaga estas desigualdades porque, como se ha reseñado, este valor indica una proporción; en ocasiones se expresa también en porcentaje, multiplicando el valor obtenido por 100.

Para entender mejor el papel del coeficiente de determinación, es útil examinar los casos extremos:

1. El coeficiente de determinación es 1 si y sólo si $SCR = 0$; en este caso todos los residuos tienen que ser exactamente igual a 0, luego $Y_t = \widehat{Y}_t$. Por lo tanto, todas las observaciones están exactamente sobre la recta de regresión MCO: el ajuste es perfecto.
2. El coeficiente de determinación es 0 si y sólo si $SCE = 0$; en este caso todos los valores ajustados tienen que ser exactamente igual a $\overline{Y}$, es decir, los valores ajustados no dependen de cuál sea el valor de la variable independiente, luego la recta de regresión MCO es una recta horizontal e igual a $\overline{Y}$. Por lo tanto, conocer el valor de la variable independiente no aporta ninguna información sobre la variable dependiente.

El valor de R^2 no puede disminuir cuando introducimos una variable explicativa adicional en el modelo. El motivo es que al incluir una variable adicional la SCT no varia (ya que sólo depende de la variable dependiente) mientras que la SCE aumenta (o permanece constante). Para poder analizar si el añadir una variable adicional en el modelo realmente mejora la bondad del ajuste se define el coeficiente de determinación ajustado que no sufre este problema

$$\overline{R}^2 = 1 - \frac{SCR/(T-k)}{SCT/(T-1)} = 1 - \frac{T-1}{T-k}\left(1 - R^2\right)$$

Nótese que el valor de $\overline{R}^2$ puede aumentar o diminuir al añadir una variable explicativa adicional en el modelo, ya que por una parte SCR disminuye (o permanece constante) pero por otro lado $T-k$ también disminuye.

Nota: Se puede demostrar que R^2 y $\overline{R}^2$ no dependen de las unidades de medida.

- Propiedades del ajuste que no se verifican en el modelo de regresión por el origen:
  - La suma de los residuos no es necesariamente cero. El motivo es que cuando el modelo no tiene constante ninguna de las columnas de X es una columna de unos y por tanto no se verifica la propiedad (a.4).
  - La media de la variable dependiente puede no coincidir con la media de los valores ajustados.
  - $$SCT \neq SCE + SCR$$

    ya que como $\overline{\widehat{Y}} \neq \overline{Y}$, $SCE \neq \widehat{Y}'\widehat{Y} - T\overline{Y}^2$
  - La interpretación del R^2 no está clara. De hecho Si definimos

    $$R^2 = 1 - \frac{SCR}{SCT}$$

    R^2 puede tomar valores negativos.