Econometría 1 curso 2009-2010

Tema 3: El modelo de regresión lineal múltiple

Genaro Sucarrat

(Departamento de Economía, UC3M)

http://www.eco.uc3m.es/sucarrat/

Recordamos:

- El modelo de regresión lineal simple esta dado por

$$\underbrace{Y}_{\text{Respuesta}} = \underbrace{\beta_0 + \beta_1 X}_{\substack{\text{Explicación} \\ \text{económica}}} + \underbrace{\varepsilon}_{\substack{\text{Error de la} \\ \text{explicación}}}$$

con los 3 supuestos:

1. Linealidad en los parámetros
2. $E(\varepsilon|X) = 0$ para todo X
3. $V(\varepsilon|X) = \sigma^2$ para todo X

- El modelo de regresión lineal múltiple esta dado por

$$\underbrace{Y}_{\text{Respuesta}} = \underbrace{\beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K}_{\text{Explicación económica}} + \underbrace{\varepsilon}_{\text{Error de la explicación}}$$

con los 4 supuestos:

1. Linealidad en los parámetros

2. $E(\varepsilon | X_1, \ldots, X_K) = 0$ para cualquier combinación de valores de $X_1, \ldots, X_K$

3. $V(\varepsilon | X_1, \ldots, X_K) = \sigma^2$ para cualquier combinación de valores de $X_1, \ldots, X_K$ (homocedasticidad condicional)

4. Ninguna combinación de los $X_1, \ldots, X_K$ forman una relación lineal exacta (ausencia de multicolinealidad exacta)

- Interpretación de los parámetros:

Las pendientes $\beta_1, \ldots, \beta_K$ se interpretan como **efectos parciales** o **efectos ceteris paribus** de un cambio en la variable asociada

Ejemplo 1: Efecto de la educación (X_1) sobre el salario (Y). Nuestro interés fundamental radica en el efecto de la educación. Pero sabemos que otras variables afectan también al salario, por ejemplo $X_2 = Sexo$ y $X_3 = Experiencia\ laboral$. Eso es la motivación de hacer una regresión múltiple:

$$Salario = \beta_0 + \beta_1 Educacion + \beta_2 Sexo + \beta_3 Experiencia + \varepsilon$$

En este caseo, el objectivo es estimar el efecto parcial de educación "quitando" los efectos de otras variables

Ejemplo 1 (cont.):

Porque al variar el nivel de Educación también varían Experiencia y Sexo:

- El efecto de educación puede diferir entre mujeres y hombres
- La experiencia puede tener una distribución diferente por niveles de educación

Si (por ejemplo) β_1 es el parámetro de mayor interés:

- En la regresión múltiple, nos aseguramos de que β_1 captura el efecto parcial de la educación manteniendo otros factores, en este caso Experiencia y Sexo, fijos
- En la regresión simple, Experiencia y Sexo forman parte del término inobservable

¿Qué pasa si estimamos un modelo simple cuando otras variables afecta el salario Y (Wooldridge, pp. 96-100)?

▶ Por ejemplo, supongamos que el Salario Y depende de dos factores, Educación (X_1) y Sexo (X_2) de manera lineal:

$$Y = \beta_0 + \beta_1 Educacion + \beta_2 Sexo + \varepsilon$$

tal que los supuestos del modelo múltiple se cumplen. ¿Qué pasa si estimamos un modelo simple?:

$$Y = \beta_0 + \beta_1 Educacion + \varepsilon'$$

▶ Si $C(Educacion, Sexo) = 0$, entonces poco importante pasa. Pero si $C(Educacion, Sexo) \neq 0$, entonces

$$E(\hat{\beta}_1 | X_1, X_2) = \beta_1 + \beta_2 \frac{\sum (x_{i1} - \overline{x}_1)(x_{i2} - \overline{x}_2)}{\sum (x_{i1} - \overline{x}_1)^2}$$

Es decir, el sesgo depende de la covarianza muestral entre Educación y Sexo, y del valor de β_2

Los supuestos del modelo múltiple tienen implicaciones similares a los del modelo simple, entre otras:

- $E(\varepsilon|X_1,\ldots,X_K) = 0 \Rightarrow E(\varepsilon) = 0$. Recordad la "ley de esperanzas iteradas": $E(\varepsilon) = E[E(\varepsilon|X_1,\ldots,X_K)] = 0$
- $E(\varepsilon|X_1,\ldots,X_K) = 0 \Rightarrow C(X_j,\varepsilon) = 0$ para $j = 1,\ldots,K$
- $V(\varepsilon|X_1,\ldots,X_K) = \sigma^2 \Rightarrow V(Y|X_1,\ldots,X_K) = \sigma^2$
- $V(\varepsilon|X_1,\ldots,X_K) = \sigma^2 \Rightarrow V(\varepsilon) = \sigma^2$
- $E(Y|X_1,\ldots,X_K) = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K$
- $E(Y) = \beta_0 + \beta_1 E(X_1) + \cdots + \beta_K E(X_K)$

- Vamos a considerar el modelo de regresión múltiple más sencillo posible:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- Para ilustrarlo, definimos $Y = Salario$, $X_1 = Educacion$ y $X_2 = Sexo = \begin{cases} 1 \text{ si es mujer} \\ 0 \text{ si es hombre} \end{cases}$

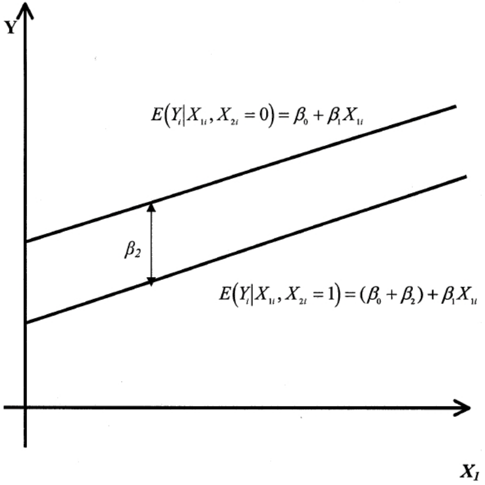
- Tenemos que:

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

de manera que

$$\begin{aligned} E(Y|X_1, X_2 = 0) &= \beta_0 + \beta_1 X_1 \\ E(Y|X_1, X_2 = 1) &= \beta_0 + \beta_1 X_1 + \beta_2 \end{aligned}$$

Graficamente, si (por ejemplo) $\beta_2 < 0$, entonces $E(Y|X_1, X_2 = 0)$ es una recta paralela a $E(Y|X_1, X_2 = 1)$ y por encima de ésta:



- Si todas las variables excepto X_j permanecen constantes, entonces

$$\Delta E(Y|X_1,\ldots,X_K) = \beta_j \Delta X_j$$

de manera que

$$\beta_j = \frac{\Delta E(Y|X_1,\ldots,X_K)}{\Delta X_j}$$

- β_j: Cuando X_j varía en una unidad (permaneciendo el resto de las variables constantes), Y varía, en promedio en β_j unidades

- Notese que: la regresión múltiple $Y = E(Y|X_1,\ldots,X_K) + \varepsilon$ responde a una pregunta diferente que las (teóricas) regresiones simples $Y = E(Y|X_1) + \varepsilon_1, \ldots, Y = E(Y|X_1) + \varepsilon_K$

En nuestro ejemplo

- $E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 =$ *valor esperado del salario para unos valores dados de educación y exo*
    - $\beta_1 =$ incremento en el salario medio asociado a un año adicional de educación manteniendo el sexo constante (es decir, para individuos del mismo sexo)

- $E(Y|X_1) = \gamma_0 + \gamma_1 X_1 =$ *valor esperado del salario para unos valores dados de educación*
    - $\gamma_1 =$ incremento en el salario medio asociado a un año adicional de educación, pero sin mantener el sexo constante (es decir, ignorando que el valor puede diferir entre mujeres y hombres)

Estimación MCO del model de regresión múltiple:

- Nuestro objetivo consiste en estimar los parámetros poblacionales $\beta_1, \ldots, \beta_K$ a partir de un conjunto de datos

- Supondremos que nuestros datos $(y_i, x_{1i}, \ldots, x_{Ki})$, con $i = 1, \ldots, n$, son una realización de una muestra aleatoria de tamaño n de una población $(Y_i, X_{1i}, \ldots, X_{Ki})$

- Sea el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + \varepsilon$$

- Dada una muestra aleatoria de tamaño n, podemos escribir:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + \varepsilon_i$$

donde para todo $i = 1, \ldots, n$ se cumplen los 4 supuestos del modelo de regresión múltiple

▶ Los estimadores de los $K+1$ parámetros $\beta_0, \beta_1, \ldots, \beta_K$ vamos a denotar $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K$

▶ Según el criterio MCO poblacional los parámetros poblacionales $\beta_0, \beta_1, \ldots, \beta_K$ son aquellos que resuelve el problema

$$\min_{\beta_0, \beta_1, \ldots, \beta_K} \left\{ E(\varepsilon^2) \right\}$$

donde $\varepsilon = Y - \beta_0 - \beta_1 X_1 - \cdots - \beta_K X_K$

Las $K+1$ condiciones de primer orden de esta problema son:

$$E(\varepsilon) = 0, \quad C(\varepsilon X_1) = 0, \quad \ldots, \quad C(\varepsilon, X_K) = 0$$

▶ En la práctica, en lugar del error ε inobservable, podemos utilizar el residuo $\hat{\varepsilon}_i$:

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_K x_{Ki})$$

donde $\hat{y}_i$ es el valor predicho o valor ajustado

▶ En consecuencia, según el criterio MCO muestral (el análogo del criterio MCO poblacional) los estimadores $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K$ son aquellos que resuelve el problema

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K} \left\{ \frac{1}{n} \sum \hat{\varepsilon}_i^2 \right\}$$

▶ Las $K+1$ condiciones de primer orden de esta problema son:

$$\sum \hat{\varepsilon}_i = 0, \quad \sum \hat{\varepsilon}_i x_{1i} = 0, \ldots, \sum \hat{\varepsilon}_i x_{Ki} = 0$$

o de forma equivalente

$\frac{1}{n}\sum \hat{\epsilon}_i = 0$ media de los residuos igual a 0

$$\left.\begin{array}{c} \frac{1}{n}\sum \hat{\epsilon}_i (x_{1i} - \overline{x}_1) = 0 \\ \vdots \\ \frac{1}{n}\sum \hat{\epsilon}_i (x_{Ki} - \overline{x}_K) = 0 \end{array}\right\} \text{covarianzas muestrales iguales a 0}$$

- Estas condiciones constituyen un sistema lineal de $K+1$ ecuaciones en $K+1$ incognitas (las $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K$)

- El sistema tendrá solución única siempre que se cumpla el supuesto 4, es decir, que no exista multicolinealidad exacta. (Si existiera multicolinealidad exacta, el sistema tendría infinitas soluciones)

- En el caso del modelo múltiple con solo dos variables, el álgebra de la estimación MCO es relativamente fácil

- En el caso poblacional:

$$
\begin{aligned}
\beta_0 &= E(Y) - \beta_1 E(X_1) - \beta_2 E(X_2) \\
\beta_1 &= \frac{V(X_2)C(X_1, Y) - C(X_1, X_2)C(X_2, Y)}{V(X_1)V(X_2) - [C(X_1, X_2)]^2} \\
\beta_2 &= \frac{V(X_1)C(X_2, Y) - C(X_1, X_2)C(X_1, Y)}{V(X_1)V(X_2) - [C(X_1, X_2)]^2}
\end{aligned}
$$

- En el caso muestral:

$$
\begin{aligned}
\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x}_1 - \hat{\beta}_2 \overline{x}_2 \\
\hat{\beta}_1 &= \frac{S_{x_2}^2 S_{x_1 y} - S_{x_1 x_2} S_{x_2 y}}{S_{x_1}^2 S_{x_2}^2 - [S_{x_1 x_2}]^2} \\
\hat{\beta}_2 &= \frac{S_{x_1}^2 S_{x_2 y} - S_{x_1 x_2} S_{x_1 y}}{S_{x_1}^2 S_{x_2}^2 - [S_{x_1 x_2}]^2}
\end{aligned}
$$

La proyección (/predictor) lineal en el caso múltiple con dos variables independientes:

- Recordad: en el caso simple la proyección lineal es igual a

$$\begin{aligned} L(Y|X) &= E(Y) + \frac{C(X,Y)}{V(X)}[X - E(X)] \\ &= \alpha_0 + \alpha_1 X \end{aligned}$$

con los estimadores $\hat{\alpha}_0$ y $\hat{\alpha}_1$ iguales a los estimadores MCO: $\hat{\beta}_0$ y $\hat{\beta}_1$

- De misma manera, en el caso múltiple (dos variables independientes) la proyección lineal es igual a

$$L(Y|X_1, X_2) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2$$

con los estimadores $\hat{\alpha}_0$, $\hat{\alpha}_1$ y $\hat{\alpha}_2$ iguales a los estimadores MCO: $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$

▶ Al igual que en el modelo de regresión simple, los estimadores MCO del modelo de regresión múltiple verifican las propiedades de:

- Linealidad
- Insesgadez (bajo los supuestos 1. 2. y 4.)
- Teorema de Gauss-Markov: bajo los supuestos 1. a 4., los estimadores $\hat{\beta}_1, \ldots, \hat{\beta}_K$ son los de menor varianza entre los estimadores lineales e insesgados
- Consistencia

▶ La justificación de estas propiedades es similar a la del caso del modelo de regresión lineal simple

Propiedades de los $\hat{\beta}_j$:

- Para $j = 1, \ldots, K$:

$$V(\hat{\beta}_j | X_1, \ldots, X_K) = \frac{\sigma^2}{nS^2_{X_j}(1 - R^2_j)}$$

- Demonstración: veasé Wooldridge. (Sin pérdida de generalidad, dado que el orden de las variables explicativas es arbitrario, la prueba es para $\hat{\beta}_1$; la prueba para $\hat{\beta}_2, \ldots, \hat{\beta}_K$ es análoga)

- Intuitivamente, R^2_j es una medida de la correlación entre X_j y las otras variables explicativas

- En terminos precisos, R^2_j es el coeficiente de determinación de la regresión de X_j sobre las otras variables explicativas:

$$X_j = \theta_0 + \theta_1 X_1 + \cdots + \theta_{j-1} X_{j-1} + \theta_{j+1} X_{j+1} + \cdots + \theta_K X_K + U$$

Es decir, la variable X_j no aparece en el lado derecho

Propiedades de los $\hat{\beta}_j$ [cont.]:

- R_j^2 mide la proporción de información de la variable X_j que ya está contenida en las demas variables

- Por tanto $(1 - R_j^2)$ mide la proporción de información distinta de la proporcionada por las restantes variables que aporta la variable X_j:

  - No es posible que $R_j^2 = 1$, porque en ese caso X_j será una combinación lineal exacta de las restantes variables (lo que se descarta por el supuesto 4.). Pero si R_j^2 estuviera cercano a 1, entonces $V(\hat{\beta}_j | X_1, \ldots, X_K)$ se dispararía
  - Por el contrario, si $R_j^2 = 0$, lo que ocurre si la correlación de X_j con las restantes variables es 0, entonces la varianza sería la mínima posible

Propiedades de los $\hat{\beta}_j$ [cont.]:

- En resumen:
  - Cuanto mayor es $S^2_{X_j}$, mayor es la precisión del estimador (por causa de mayor variabilidad muestral de X_j)
  - Cuanto mayor es el tamaño muestral n, mayor es la precisión del estimador (al mejorar el grado de representividad de la muestra)
  - Cuanto mayor es R^2_j, menor es la precisión del estimador
  - Cuanto mayor es σ^2, es decir, $V(\varepsilon|X_1, \ldots, X_K)$, menor es la precisión del estimador

Estimación de σ^2:

- El problema de estimación de σ^2 es similar al caso del modelo de regresión lineal simplelos errores ε_i son inobservables

- Pero una vez estimado el modelo por MCO, observamos los residuos $\hat{\varepsilon}_i$:

$$\begin{aligned} \hat{\epsilon}_i &= Y_i - \hat{Y}_i \\ &= Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_K XK) \end{aligned}$$

- Utilizando los residuos como análogos muestrales de los errores (inobservables), podemos calcular como estimadores de σ^2:

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_i \hat{\varepsilon}_i^2$$

Este estimador sí es factible y consistente, pero es sesgado

Estimación de σ^2 [cont.]:

- La fuente del sesgo es que los residuos $\hat{\varepsilon}_i$ verifican $K+1$ restricciones lineales (las condiciones de primera orden),

$$\sum_i \hat{\varepsilon}_i = 0, \sum_i \hat{\varepsilon}_i X_{1i} = 0, \ldots, \sum_i \hat{\varepsilon}_i X_{Ki} = 0$$

de manera que sólo hay $(n-K-1)$ residuos independientes (lo que se conoce como *grados de libertad* ("degrees of freedom"; df)

- De hecho, un estimador insesgado (que para n grande es muy similar a $\widetilde{\sigma}^2$) es:

$$\hat{\sigma}^2 = \frac{1}{n-K-1} \sum_i \hat{\varepsilon}_i^2$$

- Nótase, tanto $\hat{\sigma}^2$ como $\widetilde{\sigma}^2$ son estimadores consistentes de σ^2

Estimación de σ^2 [cont.]:

- En general, para tamaños muestrales moderados ($n > 50$?; es que depende *mucho* de contexto), es irrelevante cúal de los dos estimadores utilizar, porque siempre que n no sea muy pequeño, proporcionan estimaciones numéricas muy parecidas

- De igual modo, hemos de emplear un estimador consistente de σ^2 y sustituir en la expresión de la varianza, obteniendo

$$\hat{V}(\hat{\beta}_j|X_1,\ldots,X_K) = \frac{\hat{\sigma}^2}{nS^2_{X_j}(1-R^2_j)}$$

Medidas de bondad de ajuste:

- Tal y como argumentamos en la regresión lineal simple, podemos utilizar la raíz cuadrada de $\hat{\sigma}^2$, es decir $\hat{\sigma}$, que se denomina *error estándar de la regresión*, como medida de la bondad del ajuste

- Al igual que en el modelo de regresión simple, el R^2 o *coeficiente de determinación*, se define como

$$R^2 = \frac{\sum_i (\hat{Y}_i - \overline{Y})^2}{\sum_i (Y_i - \overline{Y})^2} = 1 - \frac{\sum_i \hat{\varepsilon}_i^2}{\sum_i (Y_i - \overline{Y})^2} \quad 0 \leq R^2 \leq 1$$

(La segunda igualdad es cierta siempre que el modelo tenga término constante)

- La interpretación del R^2 es similar a la del modelo de regresión lineal simple: se interpreta como la proporción de la variación muestral de Y explicada por el modelo

Medidas de bondad de ajuste [cont.]:

- El R^2 puede ser útil (sobre todo) para comparar distintos modelos para la misma variable dependiente Y
- Desventaja: el R^2 aumenta siempre de valor al aumentar de número de regresores, sean éstos relevantes o no
- Para tratar este problema se define el $\overline{R}^2$, también llamado R^2 *corregido* ó *ajustado de grados de libertad*:

$$\overline{R}^2 = 1 - \left[(1 - R^2) \frac{n-1}{n-K-1} \right] = 1 - \frac{\sum_i \hat{\varepsilon}_i^2/(n-K-1)}{\sum_i (Y_i - \overline{Y})^2/(n-1)}$$

Nótase, para tamaos muestrales grandes $\overline{R}^2 \approx R^2$

Interpretación de los parámetros del modelo de regresión lineal: cápitulos 2 (2.4), 3 (3.1) y 6 (6.2) de Wooldridge

- Hasta ahora nos hemos (principalmente) centrado en relaciones lineales (tanto en parámetros como en variables) entre la variable dependiente y las variables explicativas
- Sin embargo, en economía muchas relaciones no son lineales
- Es fácil incorporar relaciones no lineales en el análisis lineal de regresión (manteniendo la linealidad en parámetros) definiendo adecuadamente la variable dependiente y las explicativas
- Muy importante: con carácter general, cuando decimos que el modelo de regresión es lineal, queremos decir que es *lineal en los parámetros*, pudiendo ser no lineal en las variables. De hecho, con frecuencia, el modelo se expresa en términos de *transformaciones no lineales de las variables originales*

- En este contexto, un concepto muy importante es el de *elasticidad*: la variación porcentual que experimenta una variable (Y) en respuesta a la variación porcentual de otra (X)

- En la mayoría de las especificaciones, la elasticidad no es constante, dependiendo de los valores concretos de la variable explicativa (X) y la respuesta (Y). Además, las transformaciones que se apliquen a las variables afectan a la expresión que adopta la elasticidad

- Vamos a contemplar los ejemplos más frecuentes en los trabajos aplicados, y por simplicidad ilustraremos las especificaciones con una o dos variables explicativas

En resumen vamos a distinguir entre 7 casos:

1. Modelos lineales en variables
2. Modelo doble logarítmico
3. Modelos semilogarítmicos 1: logaritmo en la variable exógena
4. Modelos semilogarítmicos 2: logaritmo en la variable endógena
5. Modelo con termino cuadrático
6. Modelo recíproco
7. Modelo con interacciones

1. Modelo lineal en variables:

▶ El modelo considerado es simplemente

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Si los supuestos 1. a 3 se cumplen: $E(Y|X) = \beta_0 + \beta_1 X$

▶ Interpretación de β_1: si X varía 1 unidad, entonces Y varía en promedio β_1 unidades (de Y):

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X}$$

▶ La elasticidad de $E(Y|X)$ con respecto a X:

$$\frac{\frac{\Delta E(Y|X)}{E(Y|X)}}{\frac{\Delta X}{X}} = \frac{\Delta E(Y|X)}{\Delta X} \cdot \frac{X}{E(Y|X)} = \beta_1 \frac{X}{E(Y|X)}$$

1. Modelo lineal en variables [cont.]:

- Nótese que la elasticidad depende de los valores concretos de X y de Y, y por lo tanto no es siempre constante

- Es habitual aproximar elasticidades para individuos concretos (usando sus valores observados de X, Y como

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

- En otras ocaciones, se evalúan las elasticidades para los valores medios de X e Y:

$$\beta_1 \frac{E(X)}{E(Y)}$$

2. Modelo doble logarítmico:

- En algunas situaciones queremos modelizar que variaciones % en X producen variaciones % constantes en Y. Eso es la motivación por la noción de *elasticidad constante*, que es de gran utilidad en estudios de demanda, producción, costes, etc.

- El modelo considerado sería

$$\ln Y = \beta_0 + \beta_1 \ln X + \varepsilon$$

Si los supuestos 1. a 4. se cumplen: $E(\ln Y|X) = \beta_0 + \beta_1 \ln X$

- Interpretación de β_1: una elasticidad de Y con respecto a X. Es decir, cuando X varía 1%, Y varía en promedio un β_1%:

$$\beta_1 = \frac{\Delta E(\ln Y|X)}{\Delta \ln X} \approx \frac{E[(\Delta Y/Y)|X]}{\Delta X/X}$$

3. Modelos semilogarítmicos 1: logarítmo en la variable exógena:

- En algunas situaciones queremos modelizar que variaciones en términos porcentuales en X producen variaciones constantes en términos absolutos en Y

- Entonces, el modelo considerado sería

$$Y = \beta_0 + \beta_1 \ln X + \varepsilon$$

Si los supuestos 1. a 3 se cumplen: $E(Y|X) = \beta_0 + \beta_1 \ln X$

- Interpretación de β_1: si X varía 1%, entonces Y varía en promedio $\beta_1/100$ unidades de Y:

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta \ln X} \approx \frac{\Delta E(Y|X)}{\Delta X/X} \Leftrightarrow \frac{\beta_1}{100} = \frac{\Delta E(Y|X)}{100 \cdot (\Delta X/X)}$$

(Nótese que si $\frac{d \ln X}{dX} = \frac{1}{X}$, entonces $\Delta \ln X \approx \frac{1}{X} \cdot \Delta X$)

3. Modelos semilogarítmicos 1: logarítmo en la variable exógena [cont.]:

- A veces se dice se dice que β_1 es una *semielasticidad* ("casi una elasticidad"), porque en este caso la elasticidad de $E(Y|X)$ con respecto a X es igual a

$$\frac{\beta_1}{E(Y|X)}$$

que depende por tanto del valor concreto que tome $E(Y|X)$

- También en este caso es habitual aproximar elasticidades para individuos concretos (usando valores de X, Y) como

$$\frac{\beta_1}{Y_i}$$

4. Modelos semilogarítmicos 2: logarítmo en la variable endógena:

- En algunas situaciones queremos modelizar que variaciones en términos absolutos en X producen variaciones porcentuales en Y

- Entonces, el modelo considerado sería

$$\ln Y = \beta_0 + \beta_1 X + \varepsilon$$

Si los supuestos 1. a 3. se cumplen: $E(\ln Y|X) = \beta_0 + \beta_1 X$

- Nótese que este modelo se expresaría en términos de las variables originales como

$$Y = \exp(\beta_0 + \beta_1 X + \varepsilon)$$

4. Modelos semilogarítmicos 2: logarítmo en la variable endógena [cont.]:

- Interpretación de β_1: cuando X varía en 1 unidad, entonces Y varía en promedio en un $(\beta_1 \cdot 100)\%$:

$$\beta_1 = \frac{\Delta E(\ln Y|X)}{\Delta X} \approx \frac{E[(\Delta Y/Y)|X]}{\Delta X} \Rightarrow \beta_1 \cdot 100 \approx \frac{E[100 \cdot (\Delta Y/Y)|X]}{\Delta X/X}$$

- Se dice que β_1 es una *semielasticidad*, y en este caso la elasticidad de $E(Y|X)$ con respecto a X es igual a $\beta_1 X$, que depende por tanto del valor concreto que tome X

- Esta especificación es de gran utilidad para describir curvas de crecimiento exponencial

- En particular, supongamos que $X = t$, es decir, el "tiempo", entonces

$$Y = \exp(\beta_0 + \beta_1 t + \varepsilon),$$

y como $\beta_1 = \frac{\Delta E(\ln Y|X)}{\Delta t}$, β_1 recoge la tasa de crecimiento medio de Y a lo largo del tiempo

5. Modelo con términos cuadráticos:

- En algunas situaciones queremos modelizar la relación entre X e Y considerando la existencia de efectos marginales crecientes o decrecientes. Esto es útil en la especificación de tecnologías o de funciones de gasto o de costes, etc.

- El modelo considerado sería

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

Si los supuestos 1. a 4. se cumplen: $E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2$

- En este contexto

$$\frac{\Delta E(Y|X)}{\Delta X} \approx \beta_1 + 2\beta_2 X$$

- Es decir, la interpretación es que cuando X varía en 1 unidad, Y varía en media en $(\beta_1 + \beta_2 X)$ unidades

5. Modelo con términos cuadráticos [cont.]:

- Nótese que β_1 y β_2 *no tienen interpretación por separado*:
  - Dependiendo del signo de los efectos marginales serán crecientes ($\beta_2 > 0$) o decrecientes ($\beta_2 < 0$)
  - Existe un valor crítico de X en el que el efecto de X sobre $E(Y|X)$ cambia signo. Dicho punto es $X^* = -\beta_1/2\beta_2$

6. Modelo recíproco:

- El modelo considerado sería

$$Y = \beta_0 + \beta_1 \frac{1}{X} + \varepsilon$$

Si los supuestos 1. a 3. se cumplen: $E(Y|X) = \beta_0 + \beta_1 \frac{1}{X}$

- Permite una formulación con curvatura hiperbólica
- Se emplea, por ejemplo, para la curva de Phillips (una teoría sobre la relación entre inflación y desempleo)
- Al variar X en una unidad, Y varía en media en $-\beta_1 \frac{1}{X^2}$ unidades

7. Modelos con interacciones:

- Intuitivamente, una interacción entre dos variables explicativas X_1 y X_2 significa que (por ejemplo) el efecto de X_1 depende del nivel de X_2

- En el modelo lineal sería

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon,$$

Si los supuestos 1. a 4. se cumplen:
$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

  - Al variar (por ejemplo) X_1 en una unidad, Y varía en media en $\beta_1 + \beta_3 X_2$ unidades
  - Nótese que β_1 y β_3 *no tienen interpretación por separado*

Comentarios finales sobre interpretación:

- Las características de los modelos anteriores pueden combinarse, de manera que podemos tener modelos logarítmicos o semilogarítmicos con interacciones, potencias, etc.

- Ejemplo (función de producción translogarítmica): Sean $Y =$ Producción, $X_1 =$ Input 1 (Trabajo) y $X_2 =$ Input 2 (Capital):

$$\begin{aligned}\ln Y &= \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 \\ &\quad + \beta_3 (\ln X_1)^2 + \beta_4 (\ln X_2)^2 + \beta_5 (\ln X_1)(\ln X_2) + \varepsilon\end{aligned}$$

- En este modelo las elasticidades de la producción con respecto al trabajo o al capital no son constantes, a pesar de estar expresadas las variables en logaritmos

Comentarios finales sobre interpretación [cont.]:

- En particular:

$$\frac{E[(\Delta Y/Y)|X_1, X_2]}{\Delta X_1/X_1} \approx \beta_1 + 2\beta_3 \ln X_1 + \beta_5 \ln X_2$$

que depende de los logaritmos de los inputs

- La especificación translogarítmica se utiliza también para representar funciones de gasto y funciones de costes

Inferencia: Wooldridge cápitulos 4 y 5 (5.2)

- Un *contraste de hipótesis* es una técnica de inferencia estadística que permite evaluar si la información que proporcionan los datos (la muestra) avala o no una determinada *conjetura* o *hipótesis*
- Las hipótesis estadísticas pueden ser:
  - *Paramétricas*: condiciones o restricciones sobre los valores de los parámetros poblacionales (por ejemplo si $\beta_j = 0$ o no)
  - *No paramétricas*: sobre propiedades de la distribución poblacional (observaciones independientes, normalidad, simetría, etc.)

El enfoque clásico:

- La hipótesis a contrastar se denomina *hipótesis nula* (H_0)

- La negación o el complementario de la hipótesis nula se denomina *hipótesis alternativa* (H_1)

- En resumen la estrategia clásica para contrastar una hipótesis consiste en:

1. Definir la hipótesis nula (H_0) y la hipótesis alternativa (H_1)
2. Definir un *estadístico (C) de contraste*
3. Determinar la región crítica (la región de rechazo) a partir de la distribución de C bajo H_0 para un nivel de significación (α) prefijado
4. Calcular el valor del estadístico ($\widehat{C}$) y ver si se encuentra en la región de aceptación o de rechazo

El enfoque clásico [cont.]:

5. Interpretar el resultado. Si el valor del estadístico está en la región de aceptación, entonces el resultado avala H_0. Si el valor del estadístico está en la región de rechazo, entonces el resultado avala H_1

- Dado el resultado del contraste, estaremos en alguna de las cuatro situaciones siguientes:

	H_0 cierta	H_0 falsa
Aceptamos H_0	Correcto	Error (tipo II)
Rechazamos H_0	Error (tipo I)	Correcto

- El *nivel de significación* (o tamaño del contraste) se define como $Pr(\text{Rechazar } H_0 | H_0 \text{ cierta}) = Pr(\text{error tipo I}) = \alpha$

- La *potencia* del contraste se define como $Pr(\text{No rechazar } H_0 | H_1 \text{ cierta}) = 1 - Pr(\text{error tipo II})$. En palabras: la probabilidad de no rechazar la nula cuando la nula es falsa

El enfoque clásico [cont.]:

- Se dice que un *contraste es consistente* si

$$\lim_{n\to\infty} Pr(\text{error de tipo II}) = 0$$
$$\Updownarrow$$
$$\lim_{n\to\infty} [1 - Pr(\text{error de tipo II})] = 1$$

- El *p-valor* (o nivel crítico) se define como

$$\text{p-valor} = p = Pr(\widehat{C} \in \text{region critica} | H_0 \text{ cierta})$$

Es decir, el p-valor es la probabilidad de obtener una discrepancia mayor o igual que la obtenida cuando H_0 es cierta (dada la distribución de H_0):

- Cuanto menor sea el p-valor, menos probable es que la distribución obtenida para el estadístico bajo H_0 sea correcta, y por tanto es menos probable que estemos bajo H_0

El enfoque clásico [cont.]:

- El p-valor no proporciona una decición entre H_0 y H_1: nos indica cómo de probable es que estemos bajo H_0
- Una vez calculado el p-valor p para el valor muestral del estadístico, sabemos que rechazaríamos H_0 a cualquier nivel de significación igual o mayor que p

▶ Ejemplo: contraste sobre la media poblacional (pp. 69-76 de las transparencias del coordenador)

▶ Distribuciones muestrales de los estimadores MCO (pp. 77-80)

▶ Contrastes sobre el valor de un parámetro (pp. 81-87)

▶ Contrastes sobre una hipótesis lineal (pp. 88-91)

▶ Contrastes de varias restricciones lineales (pp. 92-97)

▶ Contraste de significación conjunta o global (pp. 98-101)