ECONOMETRIA

Tema 3: El modelo de regresión lineal múltiple

César Alonso

Universidad Carlos III de Madrid

Introducción

- En la mayoría de las relaciones económicas intervienen más de dos variables. Los factores que afectan al fenómeno económico objeto de estudio suelen ser múltiples.
- Esto supone una limitación en la aplicación del modelo de regresión lineal simple $Y = \beta_0 + \beta_1 X_1 + \varepsilon_1$ en el análisis empírico.
- Aunque nuestro interés fundamental radique en el efecto concreto de una determinada variable X_1 sobre un fenómeno económico Y, generalmente entran en juego factores adicionales $X_2, \ldots, X_K$.
  - Dichos factores adicionales están, generalmente, relacionados con X_1.
  - Recordemos que todos los factores no incorporados al modelo están contenidos en el término inobservable (la parte no explicada) del modelo, ε_1.

Introducción

- En la medida en que X_1 no sea independiente de $X_2, \ldots, X_K$, tendremos que, en el modelo simple, el término inobservable no verificará $E\left(\varepsilon_1 | X_1\right) = 0$.
- En tal caso, debemos incorporar los factores adicionales $X_2, \ldots, X_K$ al modelo para poder **aislar el efecto causal** de dicha variable
- Al **controlar por varios factores**, en el modelo de regresión lineal múltiple

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

  será más fácil que se cumpla que su término inobservable ε es independiente de dichos factores que en el modelo simple.

- En el modelo de regresión lineal múltiple, la pendiente β_j ($j = 1, \ldots, K$) se interpreta como el **efecto parcial** o **efecto ceteris paribus** de un cambio en la variable asociada X_j
  - Si los supuestos que veremos a continuación se cumplen, dicha interpretación es correcta aunque los datos no procedan de un experimento
  - En tales condiciones, el modelo de regresión múltiple permite reproducir las condiciones de un experimento controlado (mantener los restantes factores fijos) en un contexto no experimental.

Introducción

- **Ejemplo 1:** Efecto causal de la educación sobre el salario
  - Y = Salario X_1 = Educación
  - Nuestro interés fundamental radica en el efecto de la educación.
- Pero sabemos que otras variables afectan también al salario. Por ejemplo, sean X_2 = Sexo, X_3 = Experiencia (laboral), X_4 = Capacidad (que suponemos medible).

$$\text{Salario} = \beta_0 + \beta_1 \text{Educación} + \beta_2 \text{Sexo} + \beta_3 \text{Experiencia} + \beta_4 \text{Capacidad} + \varepsilon$$

- Cabe esperar que Educación no sea independiente de Experiencia, Sexo y Capacidad.

Introducción

- En ese caso, al *variar* el nivel de Educación también varían Experiencia, Sexo y Capacidad:
  - Las proporciones de hombres y mujeres puede diferir por niveles de educación
  - La Experiencia puede tener una distribución diferente por niveles de educación.
    (Si adquirir más Educación retrasa la incorporación al mercado de trabajo, esperíamos que $C(X_1, X_2) < 0$).
  - La Capacidad puede tener una distribución diferente por niveles de educación.
    (Si los individuos con mayor Capacidad tienden a adquirir más Educación, esperíamos que $C(X_1, X_4) > 0$).
- 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, Sexo y Capacidad, fijos.
  - En la regresión simple, Experiencia, Sexo y Capacidad forman parte del término inobservable ε_1, sobre el que no tenemos control.

- **Ejemplo 2** (Wooldridge)**:** Efecto del gasto en educación sobre el rendimiento escolar
  - Supongamos que disponemos de datos de distintas escuelas públicas en distintos distritos, para las que observamos, para un curso determinado, la calificación media *avgscore* y el gasto en educación por alumno *expend*
  - Nuestro interés radica en cuantificar el efecto del gasto en educación por alumno *expend* (que es una herramienta de política pública) en el rendimiento escolar *avgscore*.
  - Pero el entorno familiar también afecta al rendimiento escolar, que a su vez esta correlacionado con la renta media de las familias *avginc*.
  - ADEMÁS, el gasto per capita en educación tiende a crecer con la renta de las familias *avginc* (porque el gasto en educación depende de la capacidad fiscal del distrito).
  - En la regresión simple de *avgscore* sobre *expend*, el término de error incluirá, entre otros factores no controlados, *avginc*.
    - En tal caso, es difícil defender que el término inobservable del modelo simple no va a variar cuando varíe *expend*.

El modelo de regresión lineal múltiple

Supuestos

En el Modelo Lineal de Regresión Múltiple suponemos:

[1.]**Linealidad en los parámetros**$E\left(\varepsilon \left| X_1, X_2, \ldots, X_K \right.\right) = 0$.
(Para cualquier combinación de valores de $X_1, X_2, \ldots, X_K$, la media del término inobservable es siempre la misma e igual a 0)

Implicaciones:

- $E\left(\varepsilon\right) = 0$ (por la ley de esperanzas iteradas)
- $C\left(h\left(X_j\right), \varepsilon\right) = 0 \Rightarrow$ (en particular) $C\left(X_j, \varepsilon\right) = 0$ ($j = 1, \ldots, K$).

[3.]**Homocedasticidad condicional**:

$$
\begin{array}{rl}
 & V(\varepsilon \,|X_1, X_2, \ldots, X_K\,) = \sigma^2 \\
\Rightarrow & V(\,Y\,|X_1, X_2, \ldots, X_K\,) = \sigma^2
\end{array}
$$

(La **varianza de** ε y/o de Y **cada una de las subpoblaciones** asociadas a cualquier combinación de valores de X_1, $X_2, \ldots, X_K$ es constante)$(X_1, X_2, \ldots, X_K)$ no forman una combinación lineal exacta (**Ausencia de Multicolinealidad Exacta**)

El modelo de regresión lineal múltiple

Supuestos

Nótese que los supuestos **1.** y **2.** implican que: Tenemos una **Función Esperanza Condicional (FEC)** lineal:

$$E\left(Y \left| X_1, X_2, \ldots, X_K \right.\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$$

La FEC nos proporciona, para cada combinación de valores de $(X_1, X_2, \ldots, X_K)$, la **media de Y en la subpoblación** correspondiente dicha combinación de valores.La FEC nos proporciona la **mejor predicción posible** en el sentido de que minimiza $E(\varepsilon^2)$, siendo $\varepsilon =$ error de predicción $= Y - c(X_1, X_2, \ldots, X_K)$.La FEC, al igual que ocurría en el modelo de regresión lineal simple, coincide con el $L(Y|X_1, X_2, \ldots, X_K)$, que es el mejor predictor lineal en el sentido de que minimiza $E(\varepsilon^2)$, siendo $\varepsilon = Y - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K)$.

- Por tanto, las condiciones de primer orden que determinan los β's son:

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

- Vamos a considerar el modelo de regresión múltiple más sencillo posible (con sólo 2 variables).

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

- Para ilustrarlo, supongamos que la población de interés está compuesta por individuos de edad, experiencia laboral y capacidad similares.
- Consideramos las siguientes variables:
  - Y = Salario
  - X_1 = Educación
  - X_2 = Sexo= $\begin{cases} 1 & \text{si es mujer} \\ 0 & \text{si es hombre} \end{cases}$

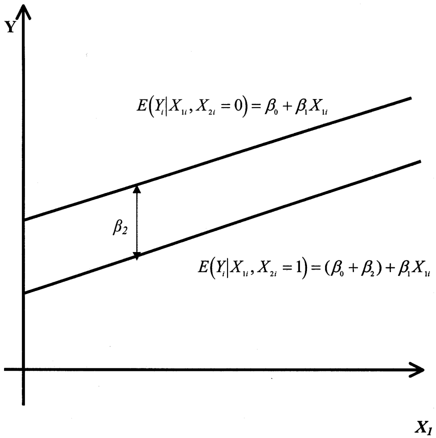
- Tenemos que:

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

de manera que

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

- Como se ve en el gráfico, si $\beta_2 < 0$, $E(Y|X_1, X_2 = 0)$ 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,

$$\Delta E(Y|X_1, X_2, \ldots, X_K) = \beta_j \Delta X_j$$

de manera que

$$\beta_j = \frac{\Delta E(Y|X_1, X_2, \ldots, X_K)}{\Delta X_j}$$

- β_j: Cuando X_j varía en una unidad (permaneciendo el resto de las variables constantes), Y varía, en promedio, en β_j unidades.
- Nótese que: la regresión múltiple $Y = E(Y|X_1, X_2, \ldots, X_K) + \varepsilon$ responde a una pregunta diferente que las (teóricas) regresiones simples $Y = E(Y|X_1) + \varepsilon_1$,..., $Y = E(Y|X_K) + \varepsilon_K$

- 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 sexo.
    - $\beta_1 =$ 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 al comparar dos distribuciones con distintos años de educación, las proporciones de hombres y mujeres pueden ser distintas).

- Continuamos con el modelo de regresión lineal múltiple más sencillo, con dos variables, que denotaremos como **Regresión larga** poblacional,

$$Y = E(Y|X_1, X_2) + \varepsilon$$

donde

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

- Los parámetros β_0, β_1 y β_2 han de verificar:

$$\begin{aligned} E(\varepsilon) &= 0 \Rightarrow \beta_0 = E(Y) - \beta_1 E(X_1) - \beta_2 E(X_2)1 && (1) \\ C(X_1, \varepsilon) &= 0 \Rightarrow \beta_1 V(X_1) + \beta_2 C(X_1, X_2) = C(X_1, Y)2 && (2) \\ C(X_2, \varepsilon) &= 0 \Rightarrow \beta_1 C(X_1, X_2) + \beta_2 V(X_2) = C(X_2, Y)3 && (3) \end{aligned}$$

- De (2) y (3) tenemos:

$$\beta_1 = \frac{V(X_2)C(X_1,Y) - C(X_1,X_2)C(X_2,Y)}{V(X_1)V(X_2) - [C(X_1,X_2)]^2}$$

$$\beta_2 = \frac{V(X_1)C(X_2,Y) - C(X_1,X_2)C(X_1,Y)}{V(X_1)V(X_2) - [C(X_1,X_2)]^2}$$

- Nótese que en el caso de que $C(X_1,X_2) = 0$:

$$\beta_1 = \frac{C(X_1,Y)}{V(X_1)} \quad \text{(pendiente de } L(Y|X_1))$$

$$\beta_2 = \frac{C(X_2,Y)}{V(X_2)} \quad \text{(pendiente de } L(Y|X_2))$$

- Supongamos que estamos particularmente interesados en el efecto de X_1 sobre Y.
  - Pero nos preocupa que X_1 y X_2 puedan estar relacionadas.
  - En tal caso, el coeficiente del modelo de regresión lineal simple NO nos proporciona el efecto de interés.
- Sea el Modelo de Regresión Lineal Simple, que denotamos como **Regresión corta** poblacional,

$$L(Y|X_1) = \gamma_0 + \gamma_1 X_1$$

Los parámetros γ_0 y γ_1 han de verificar:

$$\begin{aligned} E(\varepsilon_1) &= 0 \Rightarrow \gamma_0 = E(Y) - \gamma_1 E(X_1) 4 \quad (4) \\ C(X_1, \varepsilon_1) &= 0 \Rightarrow \gamma_1 = C(X_1, Y) \big/ V(X_1)\ 5 \quad (5) \end{aligned}$$

- (Obviamente, podríamos considerar otra regresión corta poblacional correspondiente a la proyección lineal de Y sobre X_2, con argumentos similares)

- A partir de (2) y de (5) tenemos que:

$$\gamma_1 = \frac{C(X_1, Y)}{V(X_1)} = \frac{\beta_1 V(X_1) + \beta_2 C(X_1, X_2)}{V(X_1)} = \beta_1 + \beta_2 \frac{C(X_1, X_2)}{V(X_1)}$$

  Nótese que:

  - $\gamma_1 = \beta_1$ solamente si $C(X_1, X_2) = 0$ ó si $\beta_2 = 0$.
  - $\dfrac{C(X_1, X_2)}{V(X_1)}$ es la pendiente de $L(X_2|X_1)$:

$$L(X_2|X_1) = \delta_0 + \delta_1 X_1$$

Estimación MCO en el modelo de regresión lineal múltiple

- Nuestro objetivo consiste en estimar los parámetros poblacionales, los "betas", a partir de un conjunto de datos.
- Supondremos que nuestros datos $(y_i^*, x_{1i}^*, x_{2i}^*, \ldots, x_{ki}^*)$, con $i = 1, \ldots, n$, son una realización de una **muestra aleatoria** de tamaño n de una población, $(Y_i, X_{1i}, X_{2i}, \ldots, X_{ki})$.
- Sea el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

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

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i \qquad i = 1, \ldots, n$$

donde para todo $i = 1, \ldots, n$, se cumplen los supuestos **1.** a **4.**

- Vamos a ver cómo podemos obtener estimadores de los parámetros β_0, $\beta_1, \cdots, \beta_K$ y σ^2, denotados como $\widehat{\beta}_0$, $\widehat{\beta}_1, \cdots, \widehat{\beta}_K$ y $\widehat{\sigma}^2$.
- Partiremos igualmente del principio de analogía para proponer estimadores, generalizando el resultado del modelo de regresión lineal simple al caso del modelo de regresión múltiple
- La obtención de los estimadores requiere también resolver un sistema de ecuaciones (aunque analíticamente la resolución es más compleja).
- Los parámetros poblacionales $\beta_0, \beta_1, \cdots, \beta_K$ son aquellos que resuelve el problema

$$\begin{array}{ll} \min\limits_{\beta_0,\beta_1,\cdots,\beta_K} & E(\varepsilon^2), \\ \text{donde} & \varepsilon = Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_K X_K \end{array}$$

- Para estimar dichos parámetros, en lugar del error ε_i (inobservable), podemos utilizar el **residuo** (desviación entre el valor observado y el valor predicho) como,

$$\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_K X_{Ki})$$

donde $\widehat{Y}_i$ es el **valor predicho** o **valor ajustado**, y definir el criterio MCO, que es el análogo muestral del criterio min $E(\varepsilon^2)$,

$$\min_{\beta_0, \beta_1, \cdots, \beta_K} \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2,$$

Estimación MCO en el modelo de regresión lineal múltiple

- Las condiciones de primer orden son:

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

o, de forma equivalente, definiendo $x_{ji} = \left(X_{ji} - \overline{X}_j\right)$, $j = 1, \ldots, K$.

$$\begin{array}{ll} \frac{1}{n}\sum_i \widehat{\varepsilon}_i = 0 & \text{(media muestral de los residuos 0)} \\ \left.\begin{array}{c} \frac{1}{n}\sum_i \widehat{\varepsilon}_i x_{1i} = 0 \\ \vdots \\ \frac{1}{n}\sum_i \widehat{\varepsilon}_i x_{Ki} = 0 \end{array}\right\} & \text{(covarianza muestral 0 entre } X_j \text{ y } \widehat{\varepsilon}\text{)} \end{array}$$

- Nótese que estas condiciones de primer orden son el **análogo muestral** de las condiciones de primer orden para el modelo de regresión clásico referido a los β's en la población:

$$E(\varepsilon) = 0 \qquad \text{(media poblacional de los errores 0)}$$

$$\left.\begin{array}{c} C(X_1, \varepsilon) = 0 \\ \vdots \\ C(X_K, \varepsilon) = 0 \end{array}\right\} \qquad \text{(covarianza poblacional 0 entre } X_j \text{ y } \varepsilon\text{)}$$

- El sistema de ecuaciones normales es ahora:

$$\left\{\begin{array}{c} n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_i X_{1i} + \widehat{\beta}_2 \sum_i X_{2i} + \cdots + \widehat{\beta}_K \sum_i X_{Ki} = \sum_i Y_i \\ \widehat{\beta}_1 \sum_i x_{1i}^2 + \widehat{\beta}_2 \sum_i x_{2i} x_{1i} + \cdots + \widehat{\beta}_K \sum_i x_{Ki} x_{1i} = \sum_i y_i x_{1i} \\ \vdots \\ \widehat{\beta}_1 \sum_i x_{1i} x_{Ki} + \widehat{\beta}_2 \sum_i x_{2i} x_{Ki} + \cdots + \widehat{\beta}_K \sum_i x_{Ki}^2 = \sum_i y_i x_{Ki} \end{array}\right.$$

- En general, para K variables, tendremos un sistema con $K+1$ ecuaciones lineales, donde las $K+1$ incógnitas son los coeficientes $\widehat{\beta}$'s de la regresión.
- 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)

- Podemos ver el álgebra de la estimación MCO para el modelo con dos variables.
- Los parámetros poblacionales tienen la forma:

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

- Aplicando el principio de analogía, tenemos que

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{Y} - \widehat{\beta}_1 \overline{X}_1 - \widehat{\beta}_2 \overline{X}_2 \\
\widehat{\beta}_1 &= \frac{s_2^2 s_{1y} - s_{12} s_{2y}}{s_1^2 s_2^2 - s_{12}^2} \\
\widehat{\beta}_2 &= \frac{s_1^2 s_{2y} - s_{12} s_{1y}}{s_1^2 s_2^2 - s_{12}^2}
\end{aligned}$$

- donde

$$
\begin{aligned}
s_1^2 &= \tfrac{1}{n}\textstyle\sum_i \left(X_{1i} - \overline{X}_1\right)^2 \\
s_{1y} &= \tfrac{1}{n}\textstyle\sum_i \left(X_{1i} - \overline{X}_1\right)\left(Y_i - \overline{Y}\right) \\
s_{12} &= \tfrac{1}{n}\textstyle\sum_i \left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right) = s_{21} \\
s_2^2 &= \tfrac{1}{n}\textstyle\sum_i \left(X_{2i} - \overline{X}_2\right)^2 \\
s_{2y} &= \tfrac{1}{n}\textstyle\sum_i \left(X_{2i} - \overline{X}_1\right)\left(Y_i - \overline{Y}\right)
\end{aligned}
$$

- Los estimadores $\widehat{\beta}_1$ y $\widehat{\beta}_2$ de las pendientes miden los efectos parciales estimados de X_1 y X_2, respectivamente, en el valor medio de Y.

Propiedades de los estimadores MCO

- Al igual que en el modelo de regresión simple, los estimadores MCO del modelo de regresión múltiple verifican las propiedades de:
  - Linealidad en las observaciones de Y (por definición del estimador MCO)
  - Insesgadez (bajo los supuestos **1.** y **2.** y **4.**)
  - Teorema de Gauss-Markov: Bajo los supuestos **1.** a **4.**, $\widehat{\beta}_0$, $\widehat{\beta}_1, \cdots, \widehat{\beta}_K$ 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.

- Además de los supuestos **1.** y **2.**, utilizaremos el supuesto **3.** $(V(\varepsilon|X_1, X_2, \ldots, X_K) = \sigma^2$ para todo $X)$.
- $V\left(\widehat{\beta}_j\right) = \dfrac{\sigma^2}{nS_j^2\left(1 - R_j^2\right)} = \dfrac{\sigma^2}{\sum_i x_{ji}^2\left(1 - R_j^2\right)}$, $(j = 1, \ldots, K)$ donde
  - $S_j^2 = \frac{1}{n}\left(\sum_i x_{ji}^2\right) = \frac{1}{n}\left(\sum_i \left(X_{ji} - \overline{X}_j\right)^2\right)$
  - R_j^2 es el coeficiente de determinación de la proyección lineal muestral de X_j sobre las restantes variables explicativas $X_{1i}, X_{2i}, \ldots, X_{(j-1)i}, X_{(j+1)i}, \ldots X_{Ki}$:

$$\begin{aligned} X_{ji} \quad = \quad & \theta_0 + \theta_1 X_{1i}, + \theta_2 X_{2i} + \cdots + \theta_{j-1} X_{(j-1)i} \\ & + \theta_{j+1} X_{(j+1)i} + \cdots + \theta_K X_{Ki} + u_i \end{aligned}$$

- R_j^2 mide la proporción de información de la variable X_j que ya está contenida en las demás variables.
- Por tanto $\left(1 - R_j^2\right)$ 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 sería 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, $V\left(\widehat{\beta}_j\right)$ 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.

- Intuitivamente:
  - Cuanto mayor es $S_j^2 = \frac{1}{n}\sum_i x_{ji}^2$, mayor es la variabilidad muestral de X_j, y mayor es la precisión del estimador.
  - Cuanto mayor es el tamaño muestral n, mayor es la precisión del estimador (al mejorar el grado de representatividad de la muestra).
  - Cuanto mayor es R_j^2, menor es la precisión del estimador.
- Demostración: Véase Wooldridge.
  (Sin pérdida de generalidad, dado que el orden de las variables explicativas es arbitrario, la prueba es para $\widehat{\beta}_1$; la prueba para $\widehat{\beta}_2, \cdots, \widehat{\beta}_K$ es análoga).

- El problema de estimación de σ^2 es similar al caso del modelo de regresión lineal simple.
- El problema es que los errores ε_i $(i = 1, \ldots, n)$ son inobservables.
- Una vez estimado el modelo por MCO, observamos los residuos $\widehat{\varepsilon}_i$:

$$\widehat{\varepsilon}_i = Y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_K X_{Ki}\right)$$

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

$$\widetilde{\sigma}^2 = \frac{\sum_i \widehat{\varepsilon}_i^2}{n}.$$

- Este estimador sí es factible, pero es sesgado. La razón es que, los residuos verifican $K+1$ restricciones lineales,

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

  de manera que sólo hay $(n-K-1)$ residuos independientes (lo que se conoce como **grados de libertad**).
- Alternativamente, podemos obtener un estimador insesgado (que para n grande es muy similar a $\widetilde{\sigma}^2$):

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

- Tanto $\widetilde{\sigma}^2$ como $\widehat{\sigma}^2$ son estimadores consistentes de σ^2.
- En general, para tamaños muestrales moderados, es irrelevante cuál 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

$$\widehat{V}\left(\widehat{\beta}_j\right) = \frac{\widehat{\sigma}^2}{nS_j^2\left(1 - R_j^2\right)}$$

donde S_j^2 es la varianza muestral de la variable X_j,

$$S_j^2 = \frac{1}{n}\sum_i \left(X_{ji} - \overline{X}_j\right)^2 .$$

- Tal y como argumentamos en la regresión lineal simple, podemos utilizar la raíz cuadrada de $\widehat{\sigma}^2$, $\widehat{\sigma}$, 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

$$R^2 = \frac{\sum_i \widehat{y}_i^2}{\sum_i y_i^2} = 1 - \frac{\sum_i \widehat{\varepsilon}_i^2}{\sum_i y_i^2}, \qquad 0 \leq R^2 \leq 1$$

donde $y_i = Y_i - \overline{Y}_i$, $\widehat{y}_i = \widehat{Y}_i - \overline{Y}_i$, $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$.
(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. (Véase Goldberger, pp. 82 y 83).

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

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

- En cualquier caso, para tamaños muestrales grandes $\overline{R}^2 \simeq R^2$.

Interpretación de los parámetros del modelo de regresión lineal

Capítulos 7 (7.5) y 13 (13.2) de Goldberger
Capítulos 2 (2.4), 3 (3.1) y 6 (6.2) de Wooldridge

- Hasta ahora nos hemos 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.
- Con frecuencia, el modelo se expresa en términos de **transformaciones no lineales de las variables originales**.

Interpretación de los parámetros del modelo de regresión lineal

- Un concepto muy importante en economía es el concepto de **elasticidad**, que es 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 explicativas (X) y la variable respuesta (Y).
    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.
- Por simplicidad, ilustraremos las especificaciones con una o dos variables explicativas.
- Utilizaremos el modelo lineal en parámetros y en variables como referencia.

- El modelo considerado es simplemente:

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

donde $E(\varepsilon|X) = 0 \Rightarrow E(Y|X) = \beta_0 + \beta_1 X$.

- Interpretación de β_1:

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X} \Rightarrow$$ Si X varía 1 unidad, Y varía en promedio β_1 unidades de Y.

- Elasticidad de $E(Y|X)$ con respecto a X:

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

- Nótese que la elasticidad depende de los valores concretos de X y de Y, y por lo tanto no es constante.
    - Es habitual aproximar elasticidades para individuos concretos (usando sus valores observados de X, Y) como
    $$\beta_1 \frac{X}{Y}.$$
    - En otras ocasiones, se evalúan las elasticidades para los valores medios de X e Y
    $$\beta_1 \frac{E(X)}{E(Y)}.$$

- En algunas situaciones queremos modelizar que variaciones en términos porcentuales en X producen variaciones constantes en términos absolutos en Y.
- El modelo considerado sería:

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

  donde $E(\varepsilon|X) = 0 \Rightarrow E(Y|X) = \beta_0 + \beta_1 \ln X$.
- Interpretación de β_1:

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

  (nótese que si $h(X) = \ln X$, como $h'(X) = \dfrac{dh(X)}{dX} = \dfrac{1}{X}$, entonces diferenciando tenemos que $dh(X) = d \ln X \equiv \dfrac{dX}{X}$).

- β_1 es una **semielasticidad**.
- 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 $E(Y|X)$.

- En todo caso, es habitual aproximar elasticidades para individuos concretos (usando sus valores de X, Y) como

$$\frac{\beta_1}{Y}.$$

- Multiplicando y dividiendo por 100 para expresar la variación de X **en términos porcentuales:**

$$\beta_1/100 \simeq \frac{\Delta E(Y|X)}{100 \times \Delta X/X}$$

$\Rightarrow$ Si X varía en un 1%, Y varía en promedio en $\beta_1/100$ unidades de Y.

- **Ejemplo:** Sean Y = Consumo (en euros) y X = Renta.
  Definimos la Propensión Marginal a Consumir (PMC) como dY/dX.

Modelo 1	**Modelo 2**
$Y = \beta_0^* + \beta_1^* X + \varepsilon^*$	$Y = \beta_0 + \beta_1 \ln X + \varepsilon$
$\beta_1^* = \dfrac{\Delta E(Y\|X)}{\Delta X}$	$\dfrac{\beta_1}{100} \simeq \dfrac{\Delta E(Y\|X)}{100 \times \Delta X / X}$
Si la renta $\uparrow$ 1 unidad, el consumo $\uparrow$ en media en β_1^* euros	Si la renta $\uparrow$ 1%, el consumo $\uparrow$ en media en $\beta_1/100$ euros

- La Propensión Marginal a Consumir dY/dX es:
  - β_1^* (constante), en el Modelo 1.
  - es β_1/X (decreciente con la renta), en el Modelo 2.

- En algunas situaciones, queremos modelizar que variaciones en términos absolutos en X producen variaciones constantes en términos porcentuales en Y.
- El modelo considerado sería:

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

donde $E(\varepsilon|X) = 0 \Rightarrow E(\ln Y|X) = \beta_0 + \beta_1 X$.

- Este modelo se expresaría en términos de las variables originales como:

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

- Interpretación de β_1:

$$\beta_1 = \frac{\Delta E(\ln Y|X)}{\Delta X} \simeq \frac{E\left[(\Delta Y/Y) \mid X\right]}{\Delta X}$$

  - Si multiplicamos por 100 para expresar la variación de Y en términos porcentuales, tendremos que

$$\beta_1 \times 100 \simeq \frac{E\left[(100 \times \Delta Y/Y) \mid X\right]}{\Delta X}$$

$\Rightarrow$ Cuando X varía en un 1 unidad, Y varía en promedio en un $(\beta_1 \times 100)$ %.

- β_1 es una **semielasticidad**.
- La elasticidad de $E(Y|X)$ con respecto a X es igual a $\beta_1 X$ (y depende por tanto de los valores concretos que tome X).
- Esta especificación es de gran utilidad para describir curvas de crecimiento exponencial.
- En particular, supongamos que $X = t$ (tiempo). Entonces, $Y = \exp(\beta_0 + \beta_1 t + \varepsilon)$ y como $\beta_1 = \dfrac{\Delta E(\ln Y|X)}{\Delta t}$, β_1 recoge la tasa de crecimiento medio de Y a lo largo del tiempo.

Interpretación. Especificaciones más usuales

Modelos semilogarítmicos: *logaritmo en la variable endógena* (cont.)

- **Ejemplo:** Sean Y = Salario-hora (euros) y X = Educación (años).

Modelo 1	**Modelo 2**
$Y = \beta_0^* + \beta_1^* slX + \varepsilon^*$	$\ln Y = \beta_0 + \beta_1 X + \varepsilon$
$\beta_1^* = \dfrac{\Delta E(Y\|X)}{\Delta X}$	$\beta_1 \times 100 \simeq \dfrac{E\left[(100 \times \Delta Y / Y1) \mid X\right]}{\Delta X}$
Si la educ. $\uparrow$ 1 unidad, el salario $\uparrow$ en media en β_1^* euros	Si la educ. $\uparrow$ 1 ud., el consumo $\uparrow$ en media en $(\beta_1 \times 100)$ %

- El incremento medio del salario-hora por año adicional de educación es:
  - β_1^* euros en el Modelo 1 (constante, no depende del nivel de educación considerado).
  - $(\beta_1 \times Y)$ euros en el Modelo 2 (no es constante: depende del nivel salarial, que a su vez es creciente con la educación).

- En algunas situaciones queremos modelizar que variaciones % en X producen variaciones % constantes en Y $\Rightarrow$ **Elasticidad constante**.
- De gran utilidad en estudios de demanda, producción, costes, etc..
- El modelo considerado sería:

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

donde $E(\varepsilon|X) = 0 \Rightarrow E(\ln Y|X) = \beta_0 + \beta_1 \ln X$.

- Interpretación de β_1:

$$\beta_1 = \frac{\Delta E(\ln Y|X)}{\Delta \ln X} \simeq \frac{E\left[(\Delta Y/Y) \mid X\right]}{\Delta X/X} \Rightarrow$$

Cuando X varía en un 1%, Y varía en promedio un β_1%.

(β_1 es una elasticidad)

Interpretación de los parámetros del modelo de regresión lineal

Modelo doble logarítmico (cont.)

- **Ejemplo:** Sean Y = Producción, X_1 = Input 1 (Trabajo) y X_2 = Input 2 (Capital).

$Y = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \varepsilon^*$	$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \varepsilon$
$\beta_j^* = \dfrac{\Delta E(Y \mid X_1, X_2)}{\Delta X_j}$	$\beta_j \simeq \dfrac{E\left[(\Delta Y / Y) \mid X_1, X_2\right]}{\Delta X_j / X_j}$
Si el input $j \uparrow 1$ unidad, permaneciendo el otro input constante la prod. $\uparrow$ en media en:	
β_j^* ud. de producto	β_j %
elasticidad no constante: $\dfrac{\Delta Y / Y}{\Delta X_j / X_j} = \beta_j^* \dfrac{X_j}{Y}$	elasticidad constante: β_j

- El Modelo 2 tiene la siguiente representación en términos de las variables originales:

$$Y = B_0 X_1^{\beta_1} X_2^{\beta_2} \exp(\varepsilon) \quad \text{(Cobb-Douglas)}$$

Interpretación. Especificaciones más usuales

Modelo con términos cuadráticos

- En algunas situaciones, queremos modelizar la relación entre X e Y considerando la existencia de efectos marginales crecientes o decrecientes.
- Es útil en la especificación de tecnologías o de funciones de gasto o de costes, etc.
- El modelo considerado sería:

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

donde $E(\varepsilon|X) = 0 \Rightarrow E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2$.

- En este contexto,

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

es decir: Cuando X varía en 1 unidad, Y varía en media en $(\beta_1 + 2\beta_2 X)$ unidades de Y..

- 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 de signo. Dicho punto es $X^* = -\beta_1/2\beta_2$

- **Ejemplo:** Sean $Y =$ Salario-hora (euros), $X_1 =$ Educación (años) y $X_2 =$ Experiencia (años).

Modelo 1	**Modelo 2**
$\ln Y = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \varepsilon^*$	$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 + \varepsilon$
$\beta_2^* = \dfrac{\Delta E((\Delta Y/Y) \vert X_1, X_2)}{\Delta X_2}$	$(\beta_2 + 2\beta_3 X_2) \simeq \dfrac{E\left[(\Delta Y/Y) \mid X_1, X_2\right]}{\Delta X_2}$
Si la experiencia $\uparrow$ 1 año, permaneciendo constante la educación, el salario-hora $\uparrow$ en media en:	
$(\beta_2^* \times 100)$ %	$100 \times (\beta_2 + 2\beta_3 X_2)$ %
El rendimiento de 1 año adicional de experiencia es:	
constante	no constante (depende de Experiencia)

- El modelo considerado sería:

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

donde $E(\varepsilon|X) = 0 \Rightarrow E(Y|X) = \beta_0 + \beta_1 \frac{1}{X}$.

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

- El modelo considerado sería:

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

donde:

$$E(\varepsilon | X_1, X_2) = 0 \Rightarrow E(Y | X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2.$$

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

- 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).

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

- En este modelo, las elasticidades del output con respecto al trabajo o al capital no son constantes, a pesar de estar expresadas las variables en logaritmos. En particular:

$$\frac{E\left[(\Delta Y / Y) \mid X_1, X_2\right]}{\Delta X_1 / X_1} \simeq \beta_1 + 2\beta_3 \ln X_1 + \beta_5 \ln X_2,$$

  que depende de los logaritmos de los inputs.
- La especificación translogarítmica se utiliza también para representar funciones de gasto y funciones de costes.

Inferencia en el modelo de regresión lineal: Contraste de hipótesis

Goldberger: Capítulos 7, 10 (10.3), 11 y 12 (12.5 y 12.6).
Wooldridge: Capítulos 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** sobre la población objeto de estudio.
- Las hipótesis estadísticas pueden ser:
  - **No paramétricas**: sobre propiedades de la distribución poblacional (observaciones independientes, normalidad, simetría, etc.)
  - **Paramétricas**: condiciones o restricciones sobre los valores de los parámetros poblacionales.

Inferencia en el modelo de regresión lineal: Contraste de hipótesis

- 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)
- El enfoque clásico de contraste, basado en Neyman-Pearson, consiste en dividir el espacio muestral, dada H_0, en dos regiones, una de región de aceptación y otra región de rechazo (o región crítica). Si los datos observados caen en la región de rechazo, se rechazará la hipótesis nula.

Contraste de hipótesis: Estrategia clásica

1. Definir la hipótesis nula (H_0) y la hipótesis alternativa (H_1).
2. Determinar un **estadístico de contraste**, que mide la discrepancia entre la información muestral y la hipótesis H_0.
3. Determinar qué discrepancias se consideran grandes (y por tanto, qué valores del estadístico se consideran no plausibles bajo H_0), es decir, determinar la región crítica o de rechazo. Dicha región viene dada por un **valor crítico**, dada la distribución del estadístico.
4. Dada la muestra, calcular el valor muestral del estadístico y ver si se encuentra en la región de aceptación o de rechazo.

Contraste de hipótesis: Estrategia clásica

- Dicho estadístico de contraste:
  - es función de H_0 y de los datos muestrales, con lo que es una variable aleatoria (que tomará valores distintos para cada muestra)
  - debe tener una distribución conocida (exacta o aproximada) bajo H_0 (cuando H_0 sea cierta).
    - Si la discrepancia entre la información muestral y H_0 es notable, el valor del estadístico estará dentro de un rango de valores poco probable cuando H_0 es cierta, lo que sería evidencia contra H_0.
    - Si la discrepancia entre la información muestral y H_0 es pequeña, el valor del estadístico estará dentro de un rango de valores muy probable cuando H_0 es cierta, lo que sería evidencia a favor de H_0.

Contraste de hipótesis: Estrategia clásica

- Puesto que la muestra en que se basa el contraste es aleatoria, el estadístico de contraste (que es una función de la muestra) es también una variable aleatoria.
    - Por tanto, el estadístico de contraste puede conducir a conclusiones diferentes para distintas muestras.
    - Una vez realizado el contraste, habremos optado bien por H_0, bien por H_1, y estaremos en alguna de las cuatro situaciones siguientes:

		Realidad	
		H_0 cierta	H_0 falsa
Conclusión del contraste	Aceptamos H_0	✓	Error tipo II
	Rechazamos H_0	Error tipo I	✓

- Se define **tamaño del contraste** o **nivel de significación** $=$ $\alpha = \Pr\left(\text{error de tipo I}\right) = \Pr\left(\text{Rechazar } H_0 | H_0\right)$

Contraste de hipótesis: Estrategia clásica

- En la práctica, no es posible minimizar la probabilidad de ambos tipos de error para un tamaño muestral dado
  - La única forma de reducir la probabilidad de ambos errores es incrementando el tamaño muestral
- El procedimiento clásico consiste en:
  - fijar α (tamaño o nivel de significación del contraste), es decir, establecer la probabilidad máxima de rechazar H_0 cuando es cierta.
    - Se establece un valor tan pequeño como se desee (10% o menor).
  - minimizar $\Pr(\text{error de tipo II}) = \Pr(\text{No rechazar } H_0 | H_1)$ o, de forma equivalente, maximizar

$$1 - \Pr(\text{error de tipo II}) = \textbf{ potencia} \text{ del contraste.}$$

Como la potencia está definida bajo la hipótesis alternativa, su valor dependerá del verdadero valor del parámetro (que es desconocido).

Contraste de hipótesis: Estrategia clásica

- Un **contraste es consistente** si $\lim_{n \to \infty} [1 - \Pr(\text{error de tipo } \mathbf{II})] = 1$.
- Por tanto, dada H_0:
  - se define el estadístico de contraste C,
  - se determina la región crítica (a partir de la distribución de C bajo H_0) para un nivel de significación α prefijado,
  - se evalúa el estadístico para los datos disponibles, $\widehat{C}$.
  - Si el valor del estadístico $\widehat{C}$ cae en la región crítica, se rechaza H_0 al nivel de significación $100\alpha\%$; en caso contrario, no se rechaza H_0.

Contraste de hipótesis: Estrategia clásica

- Como procedimiento alternativo, podemos utilizar el **nivel crítico** o **p-valor**, que se define como

$$\textbf{p-valor} \;= p = \Pr\left(\widehat{C} \in \{\text{región crítica}\} \middle| H_0\right)$$

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 **p-valor** no proporciona una decisión 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

- Dada una variable aleatoria Y, queremos contrastar si $E(Y) \equiv \mu = \mu_0$, donde μ_0 es cierto valor.
- La hipótesis nula es $H_0 : \mu = \mu_0$. La hipótesis alternativa es $H_1 : \mu \neq \mu_0$.
- Para evaluar esta hipótesis, disponemos de una muestra de tamaño n, $\{Y_i\}_{i=1}^{n}$, donde las Y_i's son independientes, con la que podemos estimar la media muestral de Y:

$$\widehat{\mu} \equiv \overline{Y} = \frac{1}{n} {}_{i=1}^{n} Y_i$$

- donde la esperanza y la varianza de $\overline{Y}$ son:

$$
\begin{aligned}
E\left(\overline{Y}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(Y_i\right) = \frac{1}{n} n\mu = \mu, \\
V\left(\overline{Y}\right) &= V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(Y_i\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}
$$

Ejemplo: Contraste sobre la media poblacional (iii)

- Supongamos que $Y \sim N\left(\mu, \sigma^2\right)$, donde σ^2 es conocida.
  - Si Y tiene una distribución normal, las observaciones muestrales de Y, Y_i, que son realizaciones de Y, seguirán la misma distribución.
  - Como $\overline{Y}$ es una combinación lineal de variables aleatorias normales, tendremos que $\overline{Y}$ tendrá también una distribución normal,

$$\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

  - En particular, bajo $H_0 : \mu = \mu_0$, tenemos que $\overline{Y} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$.
  - Podemos definir el estadístico $\widehat{C} = \frac{\overline{Y} - \mu_0}{\sigma / \sqrt{n}}$. Este estadístico se distribuye bajo H_0 como

$$\widehat{C} = \frac{\overline{Y} - \mu_0}{\sigma / \sqrt{n}} = \sqrt{n}\frac{\overline{Y} - \mu_0}{\sigma} \sim N(0, 1)$$

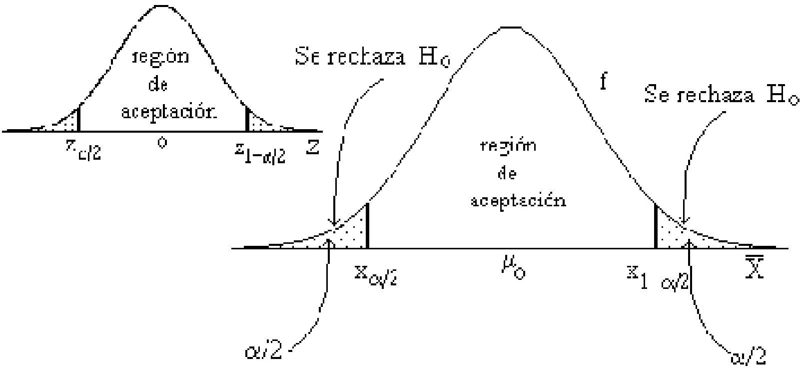
Ejemplo: Contraste sobre la media poblacional (iv)

- Por tanto, la distribución de nuestro estadístico bajo H_0 es una normal estándar.
- El contraste consiste en evaluar si la discrepancia, medida por el estadístico, es estadísticamente grande en valor absoluto.
  (nos interesa el tamaño, no la dirección, de dicha discrepancia).
- Si aplicamos el procedimiento clásico, determinamos un nivel de significación α.
- En este caso, se trata de un contraste de dos colas, con probabilidades acumuladas respectivas iguales a $\frac{\alpha}{2}$, en el que la región de aceptación corresponde a discrepancias extremas, sean positivas o negativas.

Ejemplo: Contraste sobre la media poblacional (v)

- Denotando Z_p como el valor de la $N(0,1)$ que deja una probabilidad acumulada de p a su izquierda, la región crítica es

$$\left\{\widehat{C} : \widehat{C} < Z_{\frac{\alpha}{2}}\right\} \cup \left\{\widehat{C} : \widehat{C} > Z_{1-\frac{\alpha}{2}}\right\} = \left\{\widehat{C} : \left|\widehat{C}\right| > Z_{1-\frac{\alpha}{2}}\right\}$$



Ejemplo: Contraste sobre la media poblacional (vi)

- Si $Y \sim N\left(\mu, \sigma^2\right)$, pero σ^2 es desconocida, tendremos que estimar σ^2.
  - En ese caso, el estadístico factible es

    $$\widehat{C} = \sqrt{n}\frac{\overline{Y} - \mu_0}{\widehat{\sigma}} \sim t_{n-1}$$

    donde $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$, siendo $\widehat{\sigma}^2 = \frac{1}{n-1}{}_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ un estimador consistente de σ^2.
  - La distribución de este estadístico es una t de Student con $n-1$ grados de libertad.
  - Si n es grande, el resultado del contraste será similar, tanto si suponemos que σ^2 es conocida como si no, porque la distribución t de Student se aproxima a la normal estándar a medida que n aumenta.

Ejemplo: Contraste sobre la media poblacional (vii)

- Si desconocemos tanto la distribución de Y como σ^2, podemos seguir utilizando el estadístico

$$\widehat{C} = \sqrt{n}\frac{\overline{Y} - \mu_0}{\widehat{\sigma}}$$

  - PERO ahora no tendremos una distribución conocida (exacta) para el estadístico $\widehat{C}$.
  - Sin embargo, teniendo en cuenta que las observaciones de la muestra $\{Y_i\}_{i=1}^n$ son independientes y suponiendo que tienen idéntica distribución, podemos aplicar el Teorema Central del Límite, por el que la distribución asintótica del estadístico es una normal estándar

$$\widehat{C} = \sqrt{n}\frac{\overline{Y} - \mu_0}{\widehat{\sigma}} \ \stackrel{\sim}{\cdot} \ N(0,1),$$

  de manera que podemos utilizar la normal para realizar el contraste de forma aproximada.

Contraste de hipótesis

- En la práctica:
  - No conocemos habitualmente la distibución de la(s) variable(s) consideradas.
  - Si el tamaño muestral es grande, la aproximación asintótica proporciona conclusiones similares.

- Vamos a centrarnos en el contraste de hipótesis caracterizadas por restricciones lineales sobre los parámetros $\beta_0, \beta_1, \cdots, \beta_K$.
- Para ello, construiremos estadísticos de contraste, de los que derivaremos sus distribuciones.
- Hemos derivado los estimadores MCO de los parámetros del modelo de regresión lineal múltiple y sus propiedades, a partir de los supuestos **1.** a **4.**
- Sin embargo, para hacer inferencia debemos caracterizar la distribución muestral de los estimadores MCO, para poder construir estadísticos de contraste de los que podamos derivar sus distribuciones.

- Para poder derivar distribuciones exactas, necesitaríamos suponer que:

$$
\begin{aligned}
Y_i \mid X_{1i}, \ldots, X_{Ki} &\sim N(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki},\ \sigma^2), \\
&\Leftrightarrow \varepsilon_i \mid X_{1i}, \ldots, X_{Ki} \sim N(0,\ \sigma^2)
\end{aligned}
$$

  en cuyo caso es posible demostrar que los estimadores MCO $\widehat{\beta}_j$ $(j = 1, \ldots, K)$ siguen una distribución $N\left(\beta_j, V\left(\widehat{\beta}_j\right)\right)$.
- Sin embargo, este supuesto no es en general verificable, y es difícil que se cumpla.
  - En ese caso, la distribución de los estimadores $\widehat{\beta}_j$ será desconocida.

- **Habitualmente, trabajaremos bajo el supuesto de que desconocemos la distribución de Y condicional a las X's, de manera que nuestra inferencia se basará en la aproximación asintótica.**
- Por ello, trabajaremos con la distribución asintótica de los estimadores $\widehat{\beta}_j$ $(j = 1, \ldots, K)$.
- Puede probarse que

$$\widehat{\beta}_j \overset{a}{\sim} N(\beta_j, V\left(\widehat{\beta}_j\right))$$

y por tanto

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{V\left(\widehat{\beta}_j\right)}} \overset{a}{\sim} N(0,1)$$

- Si sustituimos $V\left(\widehat{\beta}_j\right)$ por un estimador consistente, $V\left(\widehat{\beta}_j\right) = \dfrac{\widehat{\sigma}^2}{nS_j^2\left(1 - R_j^2\right)}$, este resultado se mantiene

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{V}\left(\widehat{\beta}_j\right)}} \overset{\sim}{\cdot} N(0,1)$$

- Sea la hipótesis nula $H_0 : \beta_j = \beta_j^0$ vs. la hipótesis alternativa, $H_1 : \beta_j \neq \beta_j^0$.
- Para evaluar esta hipótesis, construimos el estadístico t,

$$t = \frac{\widehat{\beta}_j - \beta_j^0}{\sqrt{\widehat{V}\left(\widehat{\beta}_j\right)}} \equiv \frac{\widehat{\beta}_j - \beta_j^0}{s_{\widehat{\beta}_j}}$$

donde $s_{\widehat{\beta}_j} = \sqrt{\widehat{V}\left(\widehat{\beta}_j\right)}$ se conoce como error estándar de $\widehat{\beta}_j$.

- Dicho estadístico de distribuye aproximadamente bajo H_0 como:

$$t = \frac{\widehat{\beta}_j - \beta_j^0}{s_{\widehat{\beta}_j}} \overset{\cdot}{\sim} N(0,1)$$

- La mayoría de los programas econométricos presentan el coeficiente estimado seguido de su error estándar, del estadístico t y del p-valor asociado al valor del estadístico t.
  - Habitualmente, dichos programas calculan el p-valor en base a la distribución t de Student con $n-K-1$ grados de libertad.
  - Si n es moderadamente grande, utilizar la distribución aproximada o la t de Student no tiene consecuencias en la inferencia.
    Por ejemplo, si en un contraste de dos colas consideramos el valor crítico para un nivel de significación α, que determina una región crítica $|t| > t_{1-\frac{\alpha}{2}}$, para $n-K-1=120$:
    - para $\alpha = 0.05$, dicho valor crítico es 1.98 si consideramos la t de Student y 1.96 si consideramos la $N(0,1)$.
    - para $\alpha = 0.10$, dicho valor crítico es 1.658 si consideramos la t de Student y 1.645 si consideramos la $N(0,1)$.

- Por tanto, podemos tomar los p-valores que aparecen en las tablas de resultados de los programas econométricos siempre que n no sea muy pequeño.
- Por supuesto, a partir de estos resultados pueden también construirse intervalos de confianza aproximados
- **Ejemplo**: Demanda de dinero y actividad económica (Goldberger, p. 107).
  Utilizando los datos para EE.UU. del archivo TIM1.GDT, hemos estimado la siguiente regresión lineal:

- Model 1: OLS, using observations 1959–1996 ($T = 38$)
  Dependent variable: Y

	Coefficient	Std. Error	t-ratio	p-value
const	2.3296	0.2054	11.3397	0.0000
X1	0.5573	0.0264	21.1227	0.0000
X2	-0.2032	0.0210	-9.6719	0.0000

Mean dependent var	6.628638	S.D. dependent var	0.172887
Sum squared resid	0.080411	S.E. of regression	0.047932
R^2	0.927291	Adjusted R^2	0.923137
$F(2, 35)$	223.1866	P-value(F)	1.20e–20

donde:

- $Y = \ln(100 \times V4/V3)$ = logaritmo de la cantidad real de dinero (M1)
- $X_1 = \ln(V2)$ = logaritmo del PIB real
- $X_2 = \ln(V6)$ = logaritmo del tipo de interés de las Letras del Tesoro
- **Interpretación:**
  - $\widehat{\beta}_1$: estimación de la elasticidad de la demanda de dinero con respecto a la producción (manteniendo el tipo de interés constante).
    Si el PIB aumenta en un 1% (y el tipo de interés no varía) se estima que la demanda de dinero aumenta en promedio un 0.6%.
  - $\widehat{\beta}_2$: estimación de la elasticidad de la demanda de dinero con respecto al tipo de interés (manteniendo la producción constante).
    Si el tipo de interés aumenta en un 1% (y el PIB no varía) se estima que la demanda de dinero cae en promedio un 0.2%.

Inferencia en el modelo de regresión lineal: Contraste de hipótesis

Contrastes sobre el valor de un parámetro

- Queremos contrastar dos hipótesis:
  - Que la demanda de dinero es inelástica al tipo de interés.
  - Que la elasticidad de la demanda de dinero resp. a la producción es 1.
- $H_0 : \beta_2 = 0$ (la demanda de dinero es inelástica al tipo de interés) frente a $H_1 : \beta_2 \neq 0$

  Entonces, bajo H_0, $t = \dfrac{\widehat{\beta}_2 - 0}{s_{\widehat{\beta}_2}} \overset{\sim}{\cdot} N(0,1)$ y

  $|t| = \left|\frac{-0.2032}{0.021}\right| = 9.676 > Z^* = 1.96$
  - Nótese que, dado que $|t| \simeq 9.7$, el p-valor es prácticamente 0 (para un valor $Z = 3.09$, el p-valor en la tabla de la normal es $0.001 = 0.1\%$).
    - $\Rightarrow$ se rechaza H_0 al nivel de significación del 1%.
    - Intervalo de confianza al 95% para β_2:

      $\widehat{\beta}_2 \pm s_{\widehat{\beta}_2} \times 1.96 : \quad \Rightarrow -0.2032 \pm 0.0210 \times 1.96 \quad \Rightarrow [-0.244 \; ; \; -0.162]$

- $H_0 : \beta_1 = 1$ (elasticidad unitaria de la demanda de dinero con respecto a la producción)
  frente a $H_1 : \beta_1 \neq 1$
  Entonces, bajo H_0, $t = \dfrac{\widehat{\beta}_1 - 1}{s_{\widehat{\beta}_1}} \,\tilde{\cdot}\, N(0,1)$ y
  $|t| = \left|\frac{0.5573-1}{0.0264}\right| = 16.769 > Z^* = 1.96$
  - El p-valor asociado es prácticamente 0.
  - $\Rightarrow$ se rechaza H_0 al nivel de significación del 1%.
  - Intervalo de confianza al 95% para β_1:

$$\widehat{\beta}_1 \pm s_{\widehat{\beta}_1} \times 1.96 : \quad \Rightarrow 0.5573 \pm 0.0264 \times 1.96 \quad \Rightarrow [0.505 \;;\; 0.609]$$

- Sea la hipótesis nula $H_0 : \lambda_0\beta_0 + \lambda_1\beta_1 + \cdots + \lambda_K\beta_K = \mu^0$.
- Para evaluar esta hipótesis, construimos el estadístico t,

$$t = \frac{\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K - \mu^0}{\sqrt{\widehat{V}\left(\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K\right)}} \equiv \frac{\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K - \mu^0}{s_{\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K}}$$

- Empleando la aproximación asintótica tenemos que bajo H_0:

$$t = \frac{\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K - \mu^0}{s_{\lambda_0\widehat{\beta}_0 + \lambda_1\widehat{\beta}_1 + \cdots + \lambda_K\widehat{\beta}_K}} \overset{\cdot}{\sim} N(0, 1).$$

- **Ejemplo:** Tecnología de producción.

$$\begin{array}{llll} \widehat{Y} = & 2.37 & +0.632X_1 & +0.452X_2 \qquad n = 31 \\ & & (0.257) & (0.219) \end{array}$$

$$s_{\widehat{\beta}_1 \widehat{\beta}_2} = \widehat{C}(\widehat{\beta}_1, \widehat{\beta}_2) = 0.055$$

donde

- Y = logaritmo de la producción
- X_1 = logaritmo del trabajo
- X_2 = logaritmo del capital

- **Interpretación**:
  - $\widehat{\beta}_1$: estimación de la elasticidad de la producción respecto al trabajo (manteniendo el capital constante)
    Si el trabajo aumenta en un 1% (y el capital no varía) se estima que la producción aumenta en promedio en un 0.63%.
  - $\widehat{\beta}_2$: estimación de la elasticidad de la producción respecto al capital (manteniendo el trabajo constante)
    Si el capital aumenta en un 1% (y el trabajo no varía) se estima que la producción aumenta en promedio en un 0.45%.

- Consideremos la hipótesis nula $H_0 : \beta_1 + \beta_2 = 1$ (Rendimientos constantes a escala)
  frente a la hipótesis alternativa $H_1 : \beta_1 + \beta_2 \neq 1$.
  - Bajo H_0:
  $$t = \frac{\widehat{\beta}_1 + \widehat{\beta}_2 - 1}{s_{\widehat{\beta}_1+\widehat{\beta}_2}} \,\tilde{\cdot}\, N(0,1).$$
  donde $s_{\widehat{\beta}_1+\widehat{\beta}_2} = \sqrt{\widehat{V}(\widehat{\beta}_1 + \widehat{\beta}_2)} = \sqrt{\widehat{V}(\widehat{\beta}_1) + \widehat{V}(\widehat{\beta}_2) + 2\widehat{C}(\widehat{\beta}_1, \widehat{\beta}_2)}$, y
  $$|t| = \left| \frac{0.632 + 0.452 - 1}{\sqrt{(0.257)^2 + (0.219)^2 + 2 \times 0.055}} \right| = 0.177 < Z^* = 1.96$$
  Por tanto, no se rechaza la hipótesis nula de rendimientos constantes a escala.

- ¿Cómo se puede contrastar más de una hipótesis sobre los parámetros del modelo?
- Por ejemplo. ¿Cómo puedo contrastar q hipótesis lineales como las siguientes:
  - $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K = 0$
  - $H_0 : \begin{array}{l} \beta_1 = 0 \\ \beta_2 = 1 \end{array}$
  - $H_0 : \begin{array}{l} \beta_1 + \beta_3 = 0 \\ \beta_2 = -1 \\ \beta_4 - 2\beta_5 = 1 \end{array}$

- **Conceptos previos**:
  - Modelo "no restringido": es aquel modelo sobre el que se desea efectuar un contraste de hipótesis (bajo H_1).
  - Modelo "restringido": es aquel modelo en que se ha impuesto la hipótesis que se desea contrastar (bajo H_0).
  - **Ejemplo 1**: $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1 : \beta_1 \neq 0$ y/o $\beta_2 \neq 0$.

Modelo no restringido	Modelo restringido
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	$Y = \beta_0 + \varepsilon$

- **Ejemplo 2**: $H_0 : \beta_2 + 2\beta_1 = 1$ vs. $H_1 : \beta_2 + 2\beta_1 \neq 1$

Modelo no restringido	Modelo restringido
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	$Y^* = \beta_0 + \beta_1 X^* + \varepsilon$ donde: $Y^* = Y - X_2$ $X^* = X_1 - 2X_2$

porque $\beta_2 + 2\beta_1 = 1 \Rightarrow \beta_2 = 1 - 2\beta_1 \Rightarrow$
$Y = \beta_0 + \beta_1 X_1 + (1 - 2\beta_1) X_2 + \varepsilon \Rightarrow$
$(Y - X_2) = \beta_0 + \beta_1 (X_1 - 2X_2) + \varepsilon$

- **Definiciones:**

	Modelo no restringido	Modelo restringido
Coeficientes de determinación:	R_S^2	R_R^2
Suma de cuadrados de los residuos:	$\sum_i \widehat{\varepsilon}_{iS}^2 = SRS$	$\sum_i \widehat{\varepsilon}_{iR}^2 = SRR$

- Empleando la aproximación asintótica tenemos que bajo H_0 [q hipótesis lineales],

$$W^0 = n\frac{SRR - SRS}{SRS} \stackrel{\sim}{\cdot} \chi^2_q$$

o de modo equivalente (siempre que en el modelo restringido la variable dependiente sea idéntica a la del modelo sin restringir),

$$W^0 = n\frac{R^2_S - R^2_R}{1 - R^2_S} \stackrel{\sim}{\cdot} \chi^2_q$$

Nótese que utilizamos n en vez de $(n - K - 1)$, lo que para tamaños muestrales grandes no supone diferencias sustanciales.

- La mayoría de los programas econométricos permiten realizar el contraste de forma automática si se escribe la hipótesis nula..
  - Habitualmente, dichos programas proporcionan el estadístico F, que supone normalidad condicional de las observaciones.
  - Dicho estadístico tendría la forma:

    $$F = \frac{SRR - SRS}{SRS} \times \frac{n - K - 1}{q} \sim F_{q,n-K-1}$$

    y se relaciona por tanto con el estadístico asintótico W^0 como $W^0 \simeq qF$.
  - De nuevo, si n es moderadamente grande, cuál de los dos se utilice no altera las conclusiones del contraste.
  - Pero nosotros aplicaremos contrastes asintóticos.
- Nótese que es fácil demostrar que $(SRR - SRS) \geq 0$ y que $(R_S^2 - R_R^2) \geq 0$.
- Todos los casos vistos anteriormente son casos particulares de este

- Este contraste se conoce también como "contraste de regresión".
- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K = 0$
  frente a
  $H_1 : \beta_j \neq 0$ para algún $j = 1, \ldots, K$.
  Modelo sin restringir

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

  Modelo restringido

$$Y = \beta_0 + \varepsilon \Rightarrow R_R^2 = 0$$

- Empleando la aproximación asintótica tenemos que bajo H_0,

$$W^0 = n \frac{R_S^2}{1 - R_S^2} \,\tilde{\cdot}\, \chi_K^2$$

- Habitualmente, los programas econométricos proporcionan el estadístico F del contraste de regresión,

$$F = \frac{R_S^2}{1 - R_S^2} \times \frac{n - K - 1}{K} \sim F_{K\,,n-K-1}$$

que se relaciona con el estadístico asintótico W^0 en la forma $W^0 \simeq KF$ (para n moderadamente grande).

- De nuevo, utilizaremos el contraste asintótico.
- **Ejemplo:** Demanda de dinero y actividad económica (Goldberger, p. 107).

  Utilizando los datos para EE.UU. del archivo TIM1.GDT, hemos estimado la siguiente regresión lineal:

- Model 1: OLS, using observations 1959–1996 ($T = 38$)
  Dependent variable: Y

	Coefficient	Std. Error	t-ratio	p-value
const	2.3296	0.2054	11.3397	0.0000
X1	0.5573	0.0264	21.1227	0.0000
X2	-0.2032	0.0210	-9.6719	0.0000

Mean dependent var	6.628638	S.D. dependent var	0.172887
Sum squared resid	0.080411	S.E. of regression	0.047932
R^2	0.927291	Adjusted R^2	0.923137
$F(2, 35)$	223.1866	P-value(F)	1.20e–20

donde:

- $Y = \ln(100 \times V4/V3)$ = logaritmo de la cantidad real de dinero (M1)
- $X_1 = \ln(V2)$ = logaritmo del PIB real
- $X_2 = \ln(V6)$ = logaritmo del tipo de interés de las Letras del Tesoro
- Consideramos la hipótesis nula $H_0 : \beta_1 = \beta_2 = 0$ (la demanda de dinero es inelástica a la producción y a los tipos de interés) frente a la alternativa $H_1 : \beta_1 \neq 0$ y/o $\beta_2 \neq 0$.
- El contraste asintótico es

$$W^0 = \frac{0.927}{1 - 0.927} \times 38 = 482.55 > \chi^2_2 = 5.99$$

- La salida de Gretl proporciona el estadístico F del contraste de regresión: $F(2, 35) = 223.1866$, que está bastante cercano a $W^0/2$.
- Claramente, se rechaza H_0 (el p-valor es prácticamente 0).