

Regresión lineal

Emilio Congregado
Universidad de Huelva

1

Sesión 1

- 1.- Especificación del modelo
- 2.- Los supuestos
- 3.- OLS
- 4.- Dos propiedades básicas del estimador OLS
- 5.- Distribución muestral del estimador OLS
- 6.- Bondad del ajuste
- 7.- Test de significatividad
- 8.- Predicción

1.- Especificación del modelo lineal

De forma genérica, queremos estimar un modelo del tipo:

$$y = f \left(\underbrace{x_1, \dots, x_k}_{\text{variables explicativas}}, \underbrace{u}_{\text{v.a. no observable}} / \beta \right)$$

Para analizar empíricamente este problema hemos de recoger información muestral, esto es, recolectar valores de y, x_1, \dots, x_k

Las muestras pueden ser de dos tipos:

■ **Con datos de sección cruzada:** distintos agentes en un mismo instante de tiempo

$$y_i = f(x_{1i}, \dots, x_{ki}, u_i / \beta) \text{ con } i = 1, \dots, N$$

■ **Series temporales:** una unidad económica a lo largo del tiempo

$$y_t = f(x_{1t}, \dots, x_{kt}, u_t / \beta) \text{ con } t = 1, \dots, T$$

■ **Datos de panel:** diferentes agentes observados en periodos distintos de tiempo.

$$y_{it} = f(x_{1,it}, \dots, x_{k,it}, u_{it} / \beta) \text{ con } i = 1, \dots, N \text{ y } t = 1, \dots, T$$

Nuestra especificación habitual será una relación de dependencia lineal, el llamado modelo de regresión lineal múltiple, modelo econométrico o mínimos cuadrados generalizados:

En corte transversal:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i \quad i = 1, \dots, N$$

En el que y es la variable endógena las X_{ki} , son las variables exógenas y u_i es el término de error.

En ocasiones se añade al modelo una constante, de forma que ésta puede interpretarse como el coeficiente que acompaña a una variable explicativa x_{1i} , cuyo valor es 1, para todo i .

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i \quad i = 1, \dots, N$$

Análogamente, para **series temporales:**

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t \quad t = 1, \dots, T$$

Como recordará la función de consumo keynesiana básica nos sugería que el consumo agregado depende de la renta disponible, y que la propensión marginal a consumir era un parámetro acotado entre 0 y 1.

Si queremos estimar la propensión marginal a consumir, podemos partir de una especificación lineal del tipo:

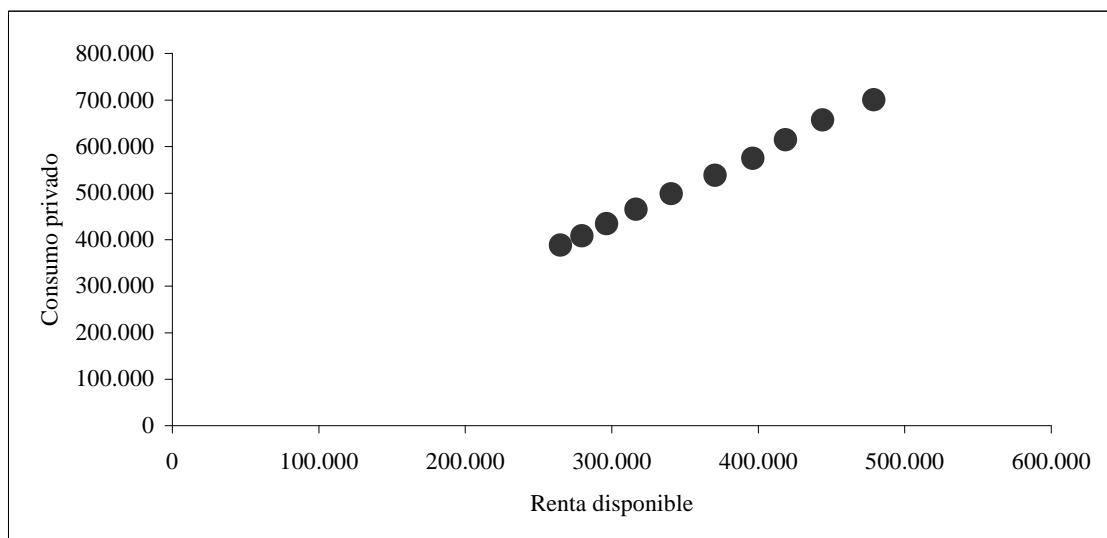
$$C = \alpha + \beta Y_d \text{ con } \alpha > 0 \text{ y } 0 < \beta < 1 \quad (1)$$

Analizamos esta relación con ayuda de datos españoles procedentes de la Contabilidad Nacional, para el período 1995-2006. Si observamos, la relación es “aproximadamente” lineal, pero no “exactamente”. Si aplicásemos una relación determinista, como la reflejada en la ecuación (1), no estaríamos dejando ningún margen a la aleatoriedad inherente a cualquier tipo de datos.

Por ello es preferible, describir la relación entre el consumo y la renta como:

$$C = \alpha + \beta Y_d + \varepsilon$$

Donde ε es una perturbación aleatoria o estocástica, introducida de forma aditiva.



¿Por qué incluir este término?

- ⚠ En primer lugar, porque el modelo es sólo una aproximación a la verdadera relación entre la independiente y la dependiente.
- ⚠ Por otra parte, siempre existirán errores de medida que estarán incluidos en la perturbación aleatoria.
- ⚠ Por la posible existencia de otros factores determinantes del comportamiento de la variable endógena

2.- Supuestos del modelo lineal

Consideremos una relación del tipo:

$$y_t = \alpha + \beta x_t + \varepsilon_t \text{ con } t = 1, \dots, T$$

Donde

y_t es la variable dependiente, explicada o endógena.

x_t es la variable independiente, explicativa o exógena.

A este modelo le llamaremos modelo de regresión lineal. Sus supuestos serán los siguientes:

1.- La forma funcional es lineal.

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Aunque parezca, a priori, demasiado restrictivo, con este supuesto se pueden acomodar otras posibles variantes.

Por ejemplo, las siguientes formas funcionales pueden transformarse en lineales con los cambios de variables sugeridos en la segunda columna:

$y = \alpha + \beta \frac{1}{x}$	$z = \frac{1}{x}$
$y = \alpha + \beta e^x$	$z = e^x$
$y = \alpha + \beta \log x$	$z = \log x$
$y = Ax^\beta$ (Modelo log-lineal de elasticidad constante)	$\log y = \log A + \beta \log x$

También es posible transformar modelos en lineales a través de aproximación de Taylor.

En definitiva, se permite no linealidad en las variables pero no en los parámetros.

Ⓞ Demuestre que el último modelo tiene elasticidad constante igual a beta.

2.- El error esperado es 0.

$E(\varepsilon_t) = 0$ (nota: si el error fuera igual a una constante, distinta de cero, el modelo sería determinista)

¿Qué significa este supuesto?

Que la media de los errores respecto a nuestro modelo sea cero y que los errores se promedien entre sí, es decir, que se compensen de forma que no distorsionen sistemáticamente nuestra relación a estimar.

3.- Homoscedasticidad (la varianza del error es constante).

$$\text{var}(\varepsilon_t) = \sigma^2, \quad \forall t$$

4.- Ausencia de correlación de los residuos.

$$\text{cov}(\varepsilon_t, \varepsilon_k) = 0, \quad \forall t \neq k$$

Los supuestos 3 y 4 se pueden resumir diciendo que la matriz de varianzas y covarianzas es una matriz del tipo:

$\text{Var}(u) = \sigma_u^2 \mathbf{I}$, lo que resume que los términos de error están incorrelados entre sí y que la matriz es escalar o esférica, siendo los elementos de la diagonal principal constantes. (Dicho de otra forma: no existe heteroscedasticidad ni autocorrelación).

5.- Ausencia de correlación entre el regresor y la perturbación.

$$\text{cov}(x_t, \varepsilon_k) = 0, \quad \forall t, k$$

6.- Los coeficientes son constantes en el tiempo.

7.- Las variables explicativas no son linealmente dependientes. (no existe multicolinealidad exacta)

8.- Las variables explicativas son deterministas.

9.- Normalidad de las perturbaciones.

$$\varepsilon \approx N(\cdot) \text{ (no necesario)}$$

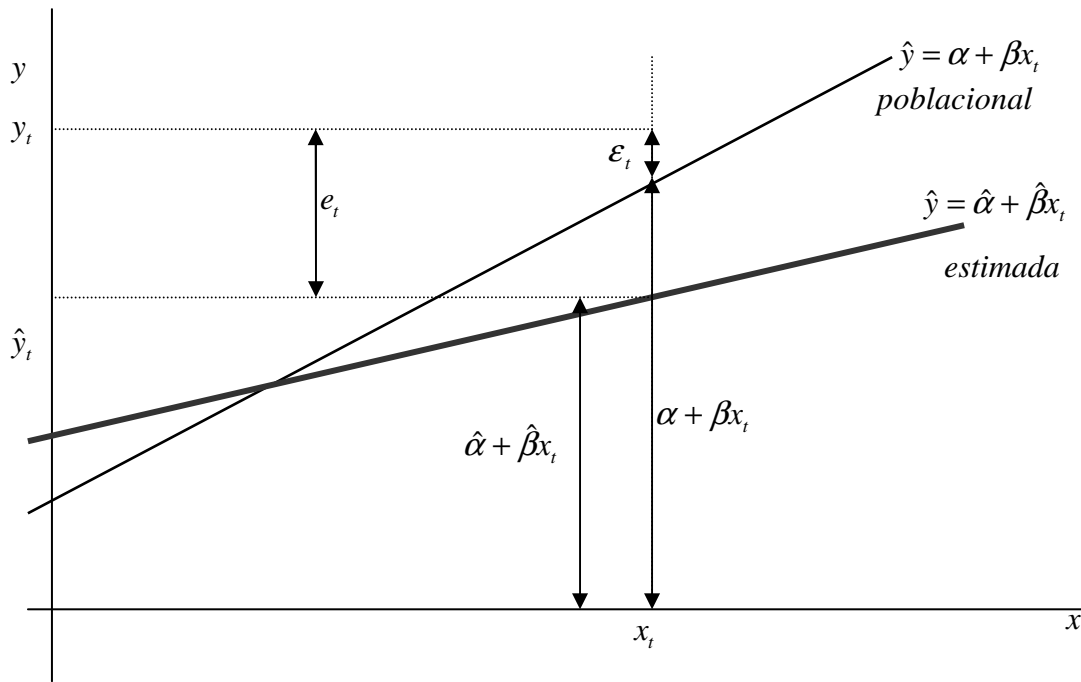
3.- Mínimos cuadrados ordinarios (OLS)

Como ya hemos señalado, nuestro objetivo no es otro que el estimar unos parámetros poblacionales desconocidos α y β a partir de unos estimadores muestrales que denotaremos por $\hat{\alpha}$ y $\hat{\beta}$, respectivamente.

De esta forma nuestra estimación quedará denotada por $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$

Es importante distinguir entre:

El **error poblacional**, que hemos denotado por ε_t y que se define como: $\varepsilon_t = y_t - \alpha - \beta x_t$, y el **error muestral**, que vendrá dado por: $e_t = y_t - \hat{\alpha} - \hat{\beta}x_t$. Esta discrepancia entre el verdadero valor y el estimado, nos permitirá obtener la estimación de la matriz de varianzas y covarianzas.



En el gráfico, la línea poblacional, la negra, no se observa. Lo que se observan, son puntos en el plano (x,y) . Mediante estos puntos, estimamos la recta estimada $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$. Igualmente, el error ε_t no se observa, pero sí podemos estimar el residuo, $e_t = y_t - \hat{\alpha} - \hat{\beta}x_t$, porque es simplemente la distancia vertical entre el punto observado y el punto estimado.

Cualquiera que sea la línea recta elegida para estimar el modelo, se cometerá un cierto error. El estimador de mínimos cuadrados ordinarios, tratará justamente de encontrar aquella recta que minimice las distancias verticales entre las observaciones y la recta estimada. Por tanto, el método consiste en minimizar la suma de los errores al cuadrado:

$$\text{Min}_{\alpha, \beta} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \approx \text{Min}_{\alpha, \beta} \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2 \approx \text{Min}_{\alpha, \beta} e' e \approx \text{Min}_{\alpha, \beta} (y - X\beta)' (y - X\beta)$$

La solución viene dada por las condiciones de primer orden de resolución de este problema de minimización:

$$\frac{\partial \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2}{\partial \hat{\alpha}} = 0$$

$$\frac{\partial \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2}{\partial \hat{\beta}} = 0$$

Cuya solución se puede expresar como:

$$\hat{\beta} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

$$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$$

O en formulación matricial como:

$$\hat{\beta} = (X'X)^{-1} X'y$$

¿Qué ocurriría si $(X'X)$, fuera una matriz singular (es decir, si $rg < k$), es decir si $|X'X| = 0$?

Pues que la matriz $(X'X)$ no sería invertible. El problema no es que no tengamos solución, sino que habría infinitas soluciones (multicolinealidad exacta).

Por tanto: Para aplicar MCO, es necesario:

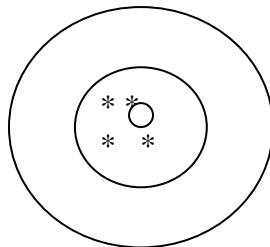
- K columnas linealmente independientes;
- $N > k$ al menos tantas observaciones como variables explicativas. Piense que si $N = k$ el error sería el vector nulo y la estimación no sería robusta.
- $N - k$ son los grados de libertad de la estimación.

4. Propiedades básicas del estimador OLS

1. **Insesgadez:** es decir, $E(\hat{\beta}) = \beta$

Es decir, si estimásemos beta con diferentes muestras, en promedio obtendríamos el “verdadero” valor de beta. Sin embargo, en economía sólo tenemos una muestra, por lo que tan sólo tendremos una realización del vector aleatorio beta. Por ello, es fundamental el conocimiento de la matriz de varianzas y covarianzas, ya que debemos minimizar la distancia $\beta - E(\hat{\beta})$.

“Hablando en plata”:



Beta es el objetivo, y las estrellas representan los valores que toma el estimador en las diferentes muestras. Lo que esperamos es que la esperanza de los diferentes beta's estimados (las estrellas) sean el punto, que es el objetivo.

2. **Eficiencia:** es el que tiene menor varianza de entre todos los estimadores lineales. El OLS es el BLUE (*best linear unbiased estimator*).

$$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$$

Cuanto mayor σ_u^2 (error) mayor varianza. Nos interesa que sea pequeña.

Cuanto mayor $(X'X)^{-1}$ más varían las variables explicativas en torno a su media, por lo que mejor se van a estimar los coeficientes.

5. Distribución muestral de los estimadores OLS

Trabajemos a partir de ahora con la notación matricial. Recordemos que la expresión matricial del estimador OLS venía dada por:

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

Que se puede describir como:

$$\hat{\beta} = (X'X)^{-1} X' \underbrace{(X\beta + \varepsilon)}_Y = \beta + (X'X)^{-1} X' \varepsilon$$

Leamos esta última expresión. Nos indica que el estimador OLS será igual al verdadero valor del parámetro β más una matriz de constantes que multiplica al vector ε . Por tanto:

- 1) Siempre que se cumpla el supuesto de que $E(X' \varepsilon) = 0$ podemos afirmar que $E(\hat{\beta}) = \beta$
- 2) También puede comprobar que: $\text{var}(\hat{\beta}) = \sigma^2 \underbrace{(X'X)^{-1}}_{\text{variación de la var explicativa}}$ (Piense en la inversa como

en una división, de forma que la varianza de la variable explicativa es muy pequeña, el estimador sería inestable numéricamente).

- 3) Teorema de Gauss-Markov: el estimador OLS es el estimador insesgado lineal de varianza mínima de beta tanto si X, es una matriz de números fijos o estocásticos.
- 4) Consecuencias del supuesto de normalidad en el error estocástico. Este supuesto nos resulta necesario para hacer inferencia estadística, es decir, para hacer contrastes de hipótesis o construir intervalos de confianza para los valores estimados. Hasta ahora sabemos que la distribución muestral de $\hat{\beta}|x \approx N(\beta, \sigma^2(X'X)^{-1})$, pero sin conocer el valor de σ^2 no estaríamos en disposición de hacer contrastes o construir intervalos de confianza, ya que en la distribución muestral de beta nos aparece este parámetro poblacional desconocido. Sin embargo como $E(\sigma^2) = \varepsilon_t^2$ y como el residuo e_t es una estimación de ε_t , entonces podemos tomar la

media de e_t^2 como estimación de σ^2 . Es decir: $\hat{\sigma}^2 = \frac{\sum_{t=1}^T e_t^2}{n}$. Sin embargo, este estimador no

es insesgado respecto de σ^2 . Un estimador insesgado viene dado por: $S^2 = \frac{e'e}{n-k}$ donde k,

indica el número de parámetros a estimar, y el numerador representa la suma de los cuadrados de los residuos. Por tanto, la estimación de la matriz de varianzas y covarianzas de beta, viene dada por: $\hat{\text{var}}(\beta) = S^2(X'X)^{-1}$.

6. Bondad del ajuste

$$VT = VE + VNE$$

El coeficiente de determinación se define como el porcentaje de varianza que es explicado por el ajuste.

Recordemos las siguientes expresiones:

Variación de total de la variable dependiente o explicada

$$\text{Var}(y) = \sum_{t=1}^T (y_t - \bar{y})^2$$

Variación explicada por el ajuste

$$\text{Var}(\hat{y}) = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2$$

Variación no explicada por el ajuste

$$\text{Var}(\hat{e}) = \sum_{t=1}^T (e_t)^2$$

Coefficiente de determinación

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \frac{\text{var}(y) - \text{var}(\hat{e})}{\text{var}(y)} = 1 - \frac{\text{var}(\hat{e})}{\text{var}(y)}$$

Notas:

- Todas las expresiones anteriores son válidas sólo si existe término independiente.
- Si no existe término independiente el coeficiente puede ser negativo.
- El coeficiente puede servirnos para discriminar entre modelos

Si la regresión consigue explicar perfectamente la variación de la variable dependiente, la suma de los cuadrados de los residuos ha de ser cero. Es decir, la varianza no explicada es cero y por tanto el coeficiente de determinación es igual a la unidad. Recuerde igualmente, que el menor valor que pueden tomar los cuadrados de los residuos es $e^t e = 0$, por lo que el coeficiente de determinación será igual a la unidad.

Sin embargo, si la regresión no consigue explicar nada, la varianza no explicada será igual a la varianza de la variable, por lo que el coeficiente de determinación será igual a cero.

Un problema asociado al el coeficiente de determinación surge debido a que éste crece al agregar variables explicativas. Por ello es frecuente **corregir el coeficiente de determinación** usando alternativamente, el **coeficiente de determinación ajustado**

$$\bar{R}^2 = 1 - \frac{\sum_{t=1}^T e_t^2 / (n - k)}{\sum_{t=1}^T (y_t - \bar{y})^2 / (n - 1)}$$

Ahora, el coeficiente puede ser negativo y no siempre crece cuando se agregan variables.

7. Tests de significatividad conjunta de la regresión

Este es un test con el que se contrasta la hipótesis de que todos los coeficientes son cero. Si fuera así, el coeficiente de determinación sería cero. El estadístico a utilizar es:

$$F[k - 1, n - k] = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

Donde k, son las variables explicativas (sin contar la constante). Si el estadístico F, cae en la región crítica rechazamos la hipótesis nula $H_0 : \beta_1 = \beta_2 = 0$.

Ej. Suponga que en una estimación de un modelo con siete variables explicativas y constante y 162 observaciones, el F experimental sale 50.000. $F[k - 1, n - k] = F(7 - 1, 162 - 7)$ El valor tabulado de

una $F(6,155)$ con un grado de significatividad del 95% es 2,01, por lo que $F(0,05) < F_{exp}$, por lo que el estadístico cae claramente en la región de rechazo y podemos rechazar la hipótesis nula.

8. Predicción

Supongamos que queremos predecir el valor de y asociado al valor de un regresor x_0 . El valor poblacional de esta observación vendrá dado por:

$$y^0 = \beta^t x^0 + \varepsilon^0$$

Si se cumplen los supuestos del modelo clásico de regresión, la predicción vendrá dada por:

$$\hat{y}^0 = \hat{\beta}^t x^0$$

El error de predicción será

$$y^0 - \hat{y}^0 = (\beta - \hat{\beta})^t x^0 + \varepsilon^0$$

Y la varianza de la predicción será:

$$\text{var}(e^0) = \sigma^2 \text{var}[(\beta - \hat{\beta})^t x^0]$$

Por tanto, cuanto más alejada de la media de la variable explicativa, peor será la predicción.