

Estadística para la Economía y la Gestión

IN 3401

15 de octubre de 2009

1 Inferencia

- Hipótesis General
- Test t: una hipótesis lineal
- Test F: conjunto de hipótesis lineales
- Intervalos de Confianza

2 Predicción

- Predicción de Valor Individual
- Predicción Valor Esperado
- Medición de la precisión de la predicción

Así, uno de nuestros objetivos puede ser el testear la significancia individual de los parámetros.

Pero lo anterior es sólo una de las preguntas que como investigadores podemos estar interesados en responder.

Por ejemplo, en la estimación de la función de producción de una firma, que asumimos Cobb Douglas:

$$Y = AK^{\alpha}L^{\beta}e^{\epsilon}$$

o en logaritmo

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \epsilon$$

podemos estar interesados en descubrir si la firma presenta rendimientos constantes, crecientes o decrecientes a la escala, lo cual se reflejará en que $\alpha + \beta >, \leq 1$.

Dos notas precautorias.

1.- En esta sección nos ocuparemos de restricciones o hipótesis **lineales** sobre los coeficientes.

2.- En todo lo que se refiere a este apartado, asumiremos que los errores de nuestra regresión muestral siguen una **distribución normal** (ya veremos porqué).

Entonces, sea nuestro modelo poblacional

$$Y = X\beta + \epsilon$$

donde X es una matriz de $(n \times k)$, ϵ e Y son vectores $(n \times 1)$ y β es vector de $(k \times 1)$.

Sean entonces las siguientes hipótesis:

1. $H_0 : \beta_i = 0 \Rightarrow$ Plantea que el regresor X_i no posee influencia alguna sobre Y . Este es el test más común y nos referiremos a él como *test de significancia*.
2. $H_0 : \beta_i = \beta_{i0} \Rightarrow$ Plantea que el regresor X_i posee un impacto determinado por β_{i0} sobre Y .
3. $H_0 : \beta_i + \beta_j = 1 \Rightarrow$ Plantea que la suma de los regresores X_i y X_j poseen un impacto conjunto de magnitud 1.

4. $H_0 : \beta_i = \beta_j \Rightarrow$ Plantea que los regresores X_i y X_j poseen el mismo impacto sobre Y .
5. $H_0 : \beta_i = 0 \quad \forall i = 1, \dots, k \Rightarrow$ Plantea que todos los regresores conjuntamente, excepto la constante, son cero.
6. $H_0 : \beta_l = 0$ donde el vector β ha sido particionado en dos (β_l y β_p) con dimensiones $(k_l \times 1)$ y $(k_p \times 1)$ respectivamente, tal que $k_l + k_p = k$. Plantea entonces que un subconjunto de parámetros son estadísticamente no significativos.

Todas las hipótesis anteriores pueden ser resumidas en la siguiente expresión:

$$R\beta = r$$

donde R es una matriz de $(q \times k)$ constantes conocidas (ceros o unos), cuyo objetivo será seleccionar los parámetros a testear, cuyo número de filas, $q \times 1$, representa el número de restricciones. A su vez, r es un vector de dimensión q y contiene el real al cual es restringido cada parámetro.

Veamos como serán las matrices R y r en cada una de nuestras hipótesis...

1. $R = [0\ldots, 010\ldots, 0]$; $r = 0$; $q = 1$

donde 1 se encuentra en la i -ésima posición.

2. $R = [0\ldots, 010\ldots, 0]$; $r = \beta_{i0}$; $q = 1$

donde 1 se encuentra en la i -ésima posición.

3. $R = [0\ldots, 010\ldots, 010\ldots, 0]$; $r = 0$; $q = 1$ donde 1 se encuentra en la i -ésima observación y en la j -ésima posición.

4. $R = [0\ldots, 010\ldots, 0 - 10\ldots, 0]$; $r = 0$; $q = 1$ donde 1 se encuentra en la i -ésima observación y -1 en la j -ésima posición.

5. $R = [0_{q \times 1} I_{k-1}]$; $r = 0$; $q = k - 1$

6. $R = [0_{k_i \times k_j} I_{k-1}]$; $r = 0$; $q = k_i$

Entonces, nuestra hipótesis nula corresponde a:

$$H_0 : R\beta = r$$

con lo cual, sólo nos resta derivar el test que nos permita rechazar o no rechazar nuestra nula. La construcción del estadígrafo es como sigue.

Dado que MCO (bajo los supuestos relevantes) es insesgado, tenemos que $E(\hat{\beta}) = \beta$, por lo tanto, $E(R\hat{\beta}) = R\beta$, mientras que la varianza de $R\hat{\beta}$ corresponde a

$$\begin{aligned} V[R\hat{\beta}] &= E[R(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'] \\ &= RV(\hat{\beta})R' \\ &= \sigma^2 R(X'X)^{-1}R' \end{aligned}$$

Necesitamos aún un supuesto más para determinar la distribución muestral de nuestra nula. Dado que $\hat{\beta}$ es función de ϵ y $\epsilon \sim N(0; \sigma^2)$, entonces $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ y por lo tanto $R\hat{\beta} \sim N(r, R\sigma^2(X'X)^{-1}R')$, entonces:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

y

$$R\hat{\beta} \sim N[R\beta, \sigma^2 R(X'X)^{-1}R']$$

y si la nula $R\beta = r$ es cierta:

$$\therefore (R\hat{\beta} - r) \sim N[0, \sigma^2 R(X'X)^{-1}R']$$

luego estandarizamos, con lo cual:

$$\frac{R\hat{\beta} - r}{\sqrt{\sigma^2 R(X'X)^{-1}R'}} \sim N[0, 1]$$

Además, se puede demostrar que (hacerlo):

$$\frac{\hat{\epsilon}'\hat{\epsilon}}{\sigma^2} \sim \chi^2_{(n-k)}$$

Luego se puede demostrar que (hacerlo):

$$(R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \sim \chi^2_q$$

luego, combinando los dos resultados anteriores, se puede demostrar que (hacerlo):

$$\frac{[(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)]/q}{\hat{\epsilon}'\hat{\epsilon}/(n-k)} \sim F_{(q,n-k)}$$

El test expuesto corresponde a la forma general del test F . Dicho test es de utilidad para testear cualquier hipótesis de la forma expuesta. A continuación veremos subcasos de dicho test general.

Reescribiendo el test F como:

$$[(R\hat{\beta} - r)'[RV(\hat{\beta})R']^{-1}(R\hat{\beta} - r)] \sim F_{(q,n-k)}$$

y haciendo el reemplazo respectivo de R y r correspondientes a las hipótesis 1 o 2 ($H_0 : \beta_i = 0 = \beta_{i0}$), llegaremos a:

$$F = \frac{(\hat{\beta} - \beta_{i0})^2}{V(\beta_i)} \sim F_{(1,n-k)}$$

Recordando que t^2 es una caso particular de una F con un grado de libertad en el numerador, tenemos que:

$$t = \frac{\hat{\beta} - \beta_{i0}}{\sqrt{V(\hat{\beta}_i)}} \sim t_{n-k}$$

Lo anterior es conocido como el **test t** (test de significancia) y en su versión más utilizada corresponde a $t = \frac{\hat{\beta}}{\sqrt{V(\hat{\beta}_i)}}$, donde se busca testear la hipótesis nula de que el parámetro es cero.

El test t también cubre los casos 3. y 4.. En el caso 3. por ejemplo ($H_0 : \beta_i + \beta_j = 1$), el estadígrafo corresponderá a:

$$t = \frac{\hat{\beta}_i + \hat{\beta}_j - 1}{V(\hat{\beta}_i) + 2Cov(\hat{\beta}_i, \hat{\beta}_j) + V(\hat{\beta}_j)} \sim t_{n-k}$$

Los casos 6. y 5. corresponden a un conjunto de hipótesis a testear. En el caso 5. correspondía a un subconjunto particular de parámetros, mientras que el caso 6. correspondía a la nula de que todos ellos eran cero, menos la constante.

En dichos casos se aplica la fórmula del test F según la ecuación general y los criterios de rechazo siguen lo expuesto en la sección anterior.

Sin embargo, en ambos casos podemos derivar expresiones alternativas para nuestro test.

- **Todas las pendientes del modelo son cero:** En este caso, se puede demostrar que el test F puede expresarse como:

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} \sim F_{(k-1, n-k)}$$

o alternativamente, utilizando la definición del R^2 :

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

- **Un subconjunto de las pendientes del modelo son cero:**

En este caso, se puede demostrar que el test F puede expresarse como:

$$F = \frac{(\hat{\epsilon}'_* \hat{\epsilon}_* - \hat{\epsilon}' \hat{\epsilon})/k_2}{\hat{\epsilon}' \hat{\epsilon}/(n-k)} \sim F_{(k_2, n-k)}$$

donde $\hat{\epsilon}_*$ denotan los residuos MCO restringidos (donde k_2 representa el número de regresores que han sido restringidos a cero), mientras que $\hat{\epsilon}$ representan los residuos del modelo MCO original.

Una forma alternativa de examinar la significancia estadística de un parámetro es a través de intervalos de confianza (IC).

Un IC nos indica, dado un nivel de confianza, el rango de valores admisibles del coeficiente que se estima.

Los niveles de confianza generalmente utilizados son 99 %, 95 % y 90 % (al igual que en los test de hipótesis), donde el tamaño de los mismos es necesariamente decreciente.

Una manera natural de obtener el IC asociado a $\hat{\beta}_i$ es a través del test t asociado. Vimos entonces que él corresponde a:

$$t = \frac{\hat{\beta} - \beta_{i0}}{\sqrt{V(\hat{\beta}_i)}} \sim t_{n-k}$$

entonces, si deseamos un IC del $(1 - \alpha) \%$ de confianza (es decir, de $\alpha \%$ de significancia) para el parámetro $\hat{\beta}_i$ tenemos que:

$$\begin{aligned} 1 - \alpha &= \left[-Z_{\alpha/2} \leq \frac{\hat{\beta} - \beta_{i0}}{\sqrt{V(\hat{\beta}_i)}} \leq Z_{\alpha/2} \right] \\ &= P \left[\hat{\beta}_i - Z_{\alpha/2} \sqrt{V(\hat{\beta}_i)} \leq \beta_{i0} \leq \hat{\beta}_i + Z_{\alpha/2} \sqrt{V(\hat{\beta}_i)} \right] \end{aligned}$$

Finalmente derivaremos el intervalo de confianza para la varianza de los errores. Sabemos que:

$$\frac{\hat{\epsilon}'\hat{\epsilon}}{\sigma^2} \sim \chi_{n-k}^2$$

por lo tanto

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-k}^2$$

Utilizando la misma lógica que utilizamos para el IC de un parámetro $\hat{\beta}_i$, tenemos que el IC para $\hat{\sigma}^2$ corresponde a:

$$\left[\frac{(n-k)\hat{\sigma}^2}{\chi_{n-k,1-\alpha}^2} \leq \sigma^2 \leq \frac{(n-k)\hat{\sigma}^2}{\chi_{n-k,\alpha}^2} \right]$$

Note que los valores críticos utilizados corresponden a $\chi_{n-k,1-\alpha}^2$ y $\chi_{n-k,\alpha}^2$, ya que la distribución χ^2 es una distribución asimétrica.

La predicción es una de las herramientas más atractivas y utilizadas en Econometría.

Si el modelo que hemos escogido confirma la teoría en consideración, es decir, a sobrevivido a las pruebas de hipótesis, podemos utilizar el modelo estimado $\hat{Y} = X\hat{\beta}$ para predecir.

La predicción se puede efectuar para un valor puntual de la variable dependiente, y^0 , correspondiente a un valor dado de los regresores, x^0 , o predecir el valor esperado $E[y^0|x^0]$ condicional a las variables explicativas.

Supongamos primero que queremos predecir un valor individual de Y , y^0 , asociado a un vector de regresores x_j^0 con $j = 1, 2, \dots, k$ de dimensión $1 \times k$.

De acuerdo con el modelo econométrico se tiene que $y^0 = \beta_1 + x_2^0 \beta_2 + \dots + x_k^0 \beta_k = \epsilon^0$. Para predecir el valor de y^0 podemos utilizar la estimación MCO del modelo, $\hat{y}^0 = x^0 \hat{\beta}$.

De esta forma, el error de predicción estará dado por:

$$e^0 = y^0 - \hat{y}^0 = x^0(\beta - \hat{\beta}) + \epsilon^0$$

En donde se distinguen dos fuentes del error de predicción

- El error en la estimación del vector β
- El error estocástico inherente al modelo ϵ^0

Sin embargo, si consideramos que el estimador MCO es insesgado y mantenemos los supuestos de nuestro modelo de regresión lineal, es trivial mostrar que el valor esperado del error de predicción será cero.

Además, podemos calcular la varianza del error de predicción:

$$V(\hat{\epsilon}^0) = E[x^0(\beta - \hat{\beta})(\beta - \hat{\beta})'x^{0'} + 2x^0(\beta - \hat{\beta})\epsilon^0 + \epsilon^0\epsilon^{0'}]$$

$$V(\hat{\epsilon}^0) = \sigma_{\epsilon}^2 + \sigma_{\epsilon}^2 x^0 (X'X)^{-1} x^{0'}$$

Bajo supuestos de normalidad del término de error, el error de predicción es una combinación lineal de dos variables normales por lo tanto tiene una distribución Normal $(0; \sigma_\epsilon^2)$. Por lo tanto, por una razonamiento análogo al de las secciones anteriores se tiene que:

$$\frac{y^0 - \hat{y}^0}{\sqrt{\sigma_\epsilon^2(1 + x^0(X'X)^{-1}x^{0'})}} \sim N(0, 1) \Rightarrow \frac{y^0 - \hat{y}^0}{\sqrt{\hat{\sigma}_\epsilon^2(1 + x^0(X'X)^{-1}x^{0'})}} \sim t_{n-k}$$

Por lo tanto, dada una predicción puntual $\hat{y}^0$ y una estimación de la desviación estándar del error de predicción podemos construir un intervalo de confianza para el valor de y^0 :

$$P[\hat{y}^0 - t_{1-\alpha/2, n-k} \sqrt{V(\hat{e}^0)} \leq y^0 \leq \hat{y}^0 + t_{1-\alpha/2, n-k} \sqrt{V(\hat{e}^0)}] = 1 - \alpha$$

Consideremos ahora que el investigador no está interesado en predecir el valor de la variable endógena y^0 , si no tan solo su valor esperado $E(y^0) = x^0\beta$.

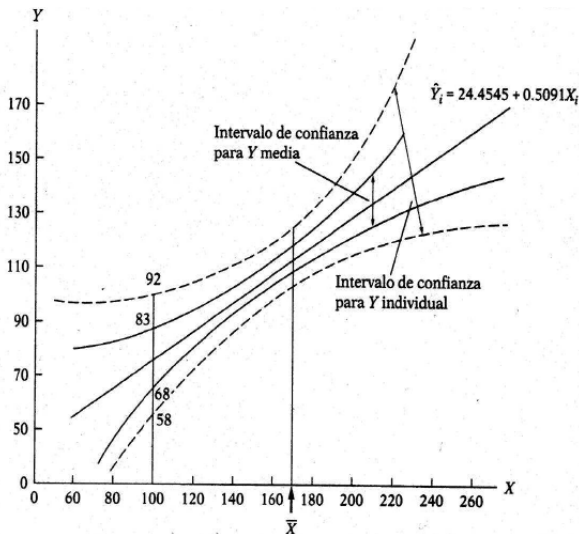
La predicción, al igual que en el caso anterior, será $x^0\hat{\beta}$. La diferencia es que el error de predicción en este caso estará definido por $\tilde{e} = E(y^0) - x^0\hat{\beta} = x^0\beta - x^0\hat{\beta}$.

Calculando entonces la varianza de este nuevo error de predicción podemos construir ahora un intervalo de confianza para $E(y^0)$ de la misma forma que antes.

$$\frac{E(y^0) - \hat{y}^0}{\sqrt{\sigma_\epsilon^2(x^0(X'X)^{-1}x^{0'})}} \sim N(0, 1) \Rightarrow \frac{E(y^0) - \hat{y}^0}{\sqrt{\hat{\sigma}_\epsilon^2(x^0(X'X)^{-1}x^{0'})}} \sim t_{n-k}$$

$$P[\hat{y}^0 - t_{1-\alpha/2, n-k} \sqrt{V(\hat{y}^0)} \leq E(y^0) \leq \hat{y}^0 + t_{1-\alpha/2, n-k} \sqrt{V(\hat{y}^0)}] = 1 - \alpha$$

La siguiente figura ejemplifica las predicciones de y^0 y $E[y^0|x^0]$ en un modelo de 2 variables independientes.



Se han propuesto varias medidas para valorar la precisión de los modelos de predicción. Muchas de estas medidas están para evaluar la predicción *ex post*, es decir, predicciones para las que las variables exógenas no tienen que ser predichas.

Dos de estas medidas que se basan en los residuos de la predicción, son la raíz cuadrada del error cuadrado medio y el error absoluto medio.

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n^0}}$$
$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n^0}$$

Donde n^0 es el número de períodos que hay que predecir.

Los métodos anteriores presentan un problema obvio de escala. Algunas medidas que no presentan este problema se basan en el estadístico U de Theil.

$$U = \sqrt{\frac{(1/n^0) \sum (y_i - \hat{y}_i)^2}{(1/n^0) \sum y_i^2}}$$