Contrastes de hipótesis en MLG

Emilio Congregado
Universidad de Huelva

2

1.- Contrastes de un solo coeficiente

Recordemos que el estimador OLS es insesgado $E(\hat{\beta}) = \beta$ y la expresión de su varianza viene dada por $\operatorname{var}(\hat{\beta}) = \sigma^2 (X^t X)^{-1}$. Además si suponíamos normalidad en el error estocástico $\varepsilon_t \approx N[0, \sigma^2 I]$. la distribución muestral del estimador viene dada por $\hat{\beta} \approx N\left(\beta, \sigma^2 (X^t X)^{-1}\right)$.

Para cada coeficiente individual podríamos escribir:

$$\hat{\beta}_i \approx N(\beta_i, \sigma^2 s_{ii})$$

Donde s_{ii} es el elemento *ii* de la matriz $(X^t X)^{-1}$.

Si conociéramos σ^2 podríamos comparar la distribución de $\hat{\beta}_i$ con la de una normal tipificada:

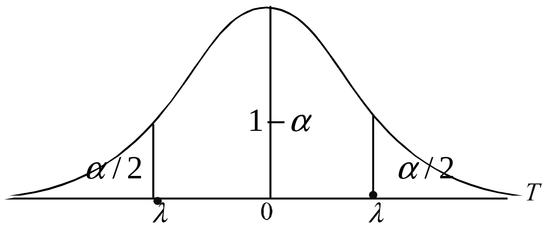
$$z_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 s_{ii}}}$$

Como no conocemos la varianza del error poblacional, debemos hacer uso de nuestro conocimiento acerca de que $\hat{\sigma}^2 = \dfrac{e^t e}{n-k}$ es un estimador insesgado de σ^2 que se distribuye según $\chi^2(n-k)$.

Por tanto, si usamos $\hat{\sigma}^2$, tendremos una t de Student:

$$t_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 s_{ii}}}$$

Recuerde que esta distribución es simétrica, por lo que la región crítica de rechazo del test viene dada por las zonas marcadas:



Los programas suelen proporcionar el error estándar de cada coeficiente estimado y el estadístico t correspondiente a la hipótesis nula de que el coeficiente es igual a cero. En este caso nuestro estadístico experimental queda como:

$$t_i[n-k] = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 s_{ii}}}$$

Este es el test de significatividad de las variables. Si $t_i[n-k]$ cae en la región crítica entonces rechazamos la hipótesis nula. En suma, $t_i[n-k]$ tiene que ser grande en valor absoluto. (A ojo: pensad siempre en mayor que $|2|$).

Los programas también suelen ofrecernos el **P-value**, esto es, el valor de Prob>$|T|$, es decir, cuánto tiene que valer α para que la región crítica fuera tal que se aceptara la hipótesis nula. En otros términos, el nivel de significatividad (100-α) al que el parámetro es significativo.

<u>Construcción de un intervalo de confianza para</u> β

Sabemos que:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 s_{ii}}} \approx t_i[n-k]$$

En la t de Student anterior, hemos visto como a derecha e izquierda dejamos el α % de la distribución (nivel de significatividad del test). El valor λ, deja $\alpha / 2$ en cada cola de forma que el área central representa un $1-\alpha$ % de la distribución. Por tanto podemos escribir que:

$$P\left[-\lambda < \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 s_{ii}}} < \lambda\right] = 1-\alpha$$

Como el estadístico se distribuye según una t de Student, λ será aquel valor que deja $\alpha / 2$ de la distribución en cada cola, ya que ésta es simétrica.
Por tanto, podemos rescribir la anterior expresión como:

$$P\left[-t_{\alpha/2} < \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 s_{ii}}} < t_{\alpha/2}\right] = 1-\alpha$$

Operando dentro del paréntesis, tenemos:

$$P\left[-t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} < \hat{\beta}_i - \beta_i < t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}}\right] = 1-\alpha$$
$$P\left[-t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} - \hat{\beta}_i < -\beta_i < t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} - \hat{\beta}_i\right] = 1-\alpha$$
$$P\left[t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} + \hat{\beta}_i > \beta_i > -t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} + \hat{\beta}_i\right] = 1-\alpha$$
$$P\left[\hat{\beta}_i - t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} < \beta_i < \hat{\beta}_i + t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}}\right] = 1-\alpha$$

Por tanto, cabe esperar que de cada 100 muestras que tomemos en 1-α de ellas el estimador ha de encontrarse entre los dos extremos del intervalo de confianza. $\left[\hat{\beta}_i - t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}} < \beta_i < \hat{\beta}_i + t_{\alpha/2}\sqrt{\hat{\sigma}^2 s_{ii}}\right]$

<u>Intervalo de confianza de la predicción</u>

Supongamos que trabajamos con un modelo univariante con constante. Dado x_0 queremos predecir el verdadero valor de y^0. Hay dos fuentes de error.

Por un lado, el verdadero valor de y^0 viene dado por:

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

Las dos fuentes de error son:

1.- El error muestral que cometemos al estimar α y β.

2.- El error que cometemos al no poder predecir ε^0.

La predicción será:

$$\hat{y}^0 = \hat{\alpha} + \hat{\beta} x^0$$

El error de predicción será:

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

Si calculamos la esperanza del error de predicción, se tiene que:

$$E(e^0) = E(\alpha - \hat{\alpha}) + E\left[(\beta - \hat{\beta})x^0\right] + E(\varepsilon^0) = 0 + 0 + 0 = 0$$

Por lo que el predictor OLS será insesgado.

La varianza del error de predicción viene dada por:

$$\mathrm{var}(e^0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(x^0 - \bar{x})^2}{S_{xx}}\right]$$

De forma que el intervalo de confianza para la predicción viene dado por:

$$\hat{\alpha} + \hat{\beta} x^0 \pm t_{\alpha/2}\sqrt{s2\left[1 + 1/n + (x^0 - \bar{x})^2 / S_{xx}\right]}$$

2.- <u>Contrastes sobre restricciones lineales</u>

Para contrastar restricciones lineales disponemos de dos métodos:

A) Estimar el modelo sin restricciones y con restricciones y comparar la suma de los cuadrados de los residuos.

Este método consiste en calcular mínimos cuadrados restringidos. La suma de los errores en el modelo restringido siempre será mayor o igual que en el modelo sin restringir, ya que un mínimo restringido siempre es mayor que un mínimo sin retringir.

$$e_r^t e_r \geq e^t e$$

El test consiste en ver si la suma de los errores en el modelo restringido es significativamente mayor que en el modelo sin restringir.

$$\frac{\left(e_r^t e_r - e^t e\right)/j}{e^t e/(n-k)} \approx F(j, n-k)$$

Si el valor muestral cae en la región crítica, rechazaremos la hipótesis nula que indica que las restricciones no son válidas.

B) Estimar el modelo y comprobar si las restricciones son estadísticamente significativas.

Supongamos que tenemos un conjunto de restricciones de la forma:

$$a_{11}\beta_1 + a_{12}\beta_2 + \ldots + a_{1k}\beta_k = q_1$$
$$a_{21}\beta_1 + a_{22}\beta_2 + \ldots + a_{2k}\beta_k = q_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{j1}\beta_1 + a_{j2}\beta_2 + \ldots + a_{jk}\beta_k = q_j$$

Con $j < k$. Note que si j y k fuesen iguales los coeficientes estarían exactamente determinados.

Alternativamente, en forma matricial las restricciones quedan representadas por:

$$H_0 : A\beta = q$$

Si las restricciones fueran ciertas, los parámetros poblacionales deberían verificar que:

$$A\beta - q = 0$$

Sin embargo, con las estimaciones, la expresión no tiene porque ser igual a 0.

$$A\hat{\beta} - q = d$$

Si llamamos d, a esta discrepancia, lo que tenemos es que obtener su distribución y saber si es significativamente distinto de 0.
La varianza de d será:

$$\operatorname{var}(d) = \operatorname{var}\left(A\hat{\beta} - q\right) = A \operatorname{var}\left(\hat{\beta}\right) A^t$$

Para construir el estadístico se utiliza el **test de Wald:**

$$\frac{\left(A\hat{\beta} - q\right)^t \left[A\left(X^t X\right)^{-1} A^t\right]\left(A\hat{\beta} - q\right)/j}{\left(e^t e\right)/(n-k)} \approx F[j, n-k]$$

Si la F experimental cae en la región crítica, rechazamos la hipótesis nula, por lo que las restricciones no serían válidas.

3.- Contrastes de cambio estructural o test de Chow[1]

[1] Este test sólo es aplicable si las varianzas de las submuestras son idénticas. Si éstas varían, pero sólo para grandes muestras, se utiliza la siguiente versión del test de Wald $\left(\hat{\beta}_1 - \hat{\beta}_2\right)^t (V_1 + V_2))^{-1}\left(\hat{\beta}_1 - \hat{\beta}_2\right) \approx \chi^2(k)$

Supongamos que tenemos un modelo de la forma:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}}_{\varepsilon}$$

El modelo sin restringir lo podríamos especificar como:

$$Y = X\beta + \varepsilon$$

Que podríamos estimar por OLS.

$$\hat{\beta} = \left(X^t X\right)^{-1} X^t y$$

¿Cómo podríamos estimar el modelo restringido?

La restricción es $\beta_1 = \beta_2$ es decir que los coeficientes sean los mismos entre los dos grupos de observaciones o los dos períodos de tiempo. Si incluyésemos esta restricción en el modelo inicial tendríamos ahora:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

Los residuos de este nuevo modelo serían los correspondientes al modelo restringido, es decir, $e_r^t e_r$. A partir de aquí aplicaremos el método de tests con restricciones, es decir:

$$\frac{\left(e_r^t e_r - e^t e\right)/j}{e^t e/(n-k)} \approx F(j, n-k)$$

4.- Contrastes de restricciones no lineales

Supongamos que una restricción adopta una forma no lineal:

$$H_0 : f(\beta) = q$$

Por analogía, nosotros desearemos construir un estadístico para el contraste de la forma:

$$z = \frac{f\left(\hat{\beta}\right) - q}{\textit{estimación error estándar}}$$

En la que el numerador refleja si $f\left(\hat{\beta}\right)$ se aleja de q, significativamente.
Para calcular el error estándar, se usa una aproximación de Taylor:

$$f\left(\hat{\beta}\right) = f(\beta) + \left(\frac{\partial f(\beta)}{\partial \beta}\right)^t \left(\hat{\beta} - \beta\right)$$

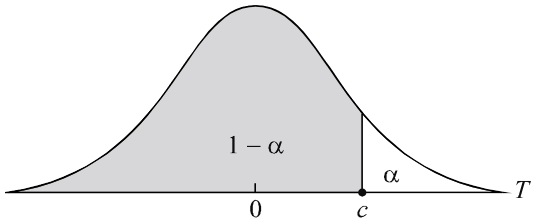
Cuando la muestra sea significativamente grande, ha de verificarse que:

$$\operatorname{var}\left[f\left(\hat{\beta}\right)\right] \approx \left(\frac{\partial f}{\partial \beta}\right)^t \operatorname{var}\left(\hat{\beta}\right)\left(\frac{\partial f}{\partial \beta}\right)$$

Donde: $\operatorname{var}\left(\hat{\beta}\right) = S^2 \left(X^t X\right)^{-1}$

TABLA DE LA DISTRIBUCION *t*–Student

La tabla da áreas $1-\alpha$ y valores $c = t_{1-\alpha, r}$, donde, $P[T \le c] = 1-\alpha$, y donde T tiene distribución t-Student con r grados de libertad.



1 – α

r	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576