CAPÍTULO 3

PROCEDIMIENTOS PARA CONTRASTAR HIPÓTESIS

3.1. INTRODUCCIÓN

En los dos capítulos anteriores hemos definido estadísticos con el objetivo de aproximar numéricamente los valores de ciertos parámetros. Se han derivado resultados que garantizan que la aproximación es buena, aunque el valor concreto al que se llegue en una aplicación dependerá de la calidad de la evidencia empírica de la que se parta.

En este capítulo variaremos de óptica. En lugar de buscar aproximaciones razonablemente válidas, el objetivo ahora es dar respuesta a cuestiones del siguiente tipo: ¿Puede aceptarse que un parámetro o una combinación de los parámetros del modelo puede tomar un valor concreto a la vista de la evidencia empírica disponible?. En torno a los estadísticos derivados en los dos primeros capítulos se pueden definir regiones de valores de forma que, en principio, todo valor situado dentro de esas regiones es un valor que puede aceptarse para el parámetro en cuestión.

En este capítulo vamos a presentar desarrollos orientados a tomar la decisión de si aceptar o rechazar una hipótesis nula determinada. Comenzaremos introduciendo una serie de definiciones que nos permitirán establecer un conjunto de contrastes que sean admisibles; posteriormente, se analizará como elegir uno de los contrastes entre el conjunto de procedimientos admisibles.

En la sección 3 se prestará una atención especial a los contrastes de la Razón de Verosimilitud (LR), Wald (W) y los Multiplicadores de Lagrange (LM); estos procedimientos pertenecen al conjunto de contrastes admisibles, al menos asintóticamente, y han sido ampliamente utilizados en el trabajo aplicado.

El trabajo termina aplicando algunos de los procedimientos estudiados previamente al contraste de hipótesis lineales en el marco del Modelo Lineal General.

3.2. CONCEPTOS BÁSICOS

Supongamos, como punto de partida, que disponemos de una muestra de tamaño T que representamos por y_1, ... y_T extraída a partir de una población con una distribución de probabilidad dada por $f(y, \theta)$ en donde θ es un vector de k parámetros cuyos valores pertenecen al espacio de parámetros, Θ, es decir, $\theta \in \Theta$.

Una hipótesis es una conjetura en el sentido de que θ pertenece a un subespacio de Θ, Θ_0. Esto se indica por:

$$H_0\text{: } \theta \in \Theta_0;\ \Theta_0 \subset \Theta.$$

Según la forma que adopte Θ_0, se puede hablar de diferentes tipos de hipótesis. Si Θ_0 incluye solamente un punto, se trata de una hipótesis simple; si incluye más de un punto, se trata de una hipótesis compuesta. Si la conjetura afecta sólo a uno de los elementos de θ, se trata de una hipótesis individual; si la conjetura afecta a varios elementos de θ tenemos una hipótesis conjunta. Dentro de las hipótesis compuestas, si Θ_0 incluye los valores correspondientes a una región definida como $\theta > \theta_0$, siendo θ_0 valores concretos que toman los elementos del vector, estamos en el caso de la hipótesis compuesta unilateral; si la región se define como $\theta \neq \theta_0$, se trata de una hipótesis compuesta bilateral.

Un **contraste de hipótesis** es un procedimiento que, utilizando toda la información relevante, permite concluir si la hipótesis que se contrasta es o no aceptable. La información relevante se refiere a los valores muestrales y a cualquier otra información a priori disponible sobre los parámetros incluidos en θ.

Sea S un conjunto de todas las posibles muestras. Cuando se observa una muestra particular que pertenece a S, hay que decidir si la hipótesis que se contrasta es aceptada o rechazada. Para ello, todo método de contraste distingue, dentro de S, dos regiones que llamaremos S_0 y S_1. Si la muestra observada pertenece a S_0 entonces H_0 se rechaza y si pertenece a S_1 la hipótesis es aceptada. S_0 es la llamada **Región Crítica** del contraste. La cuestión clave que determina todo el proceso es, por tanto: ¿como se lleva a cabo la partición de S en S_0 y S_1?.

Como indica Leamer (1978) en su libro, para definir un contraste hay que resolver dos cuestiones:

Primera: definir el conjunto de procedimientos de contraste admisibles, y,

Segunda: elegir un contraste entre todos los que sean admisibles.

La primera cuestión es un problema puramente matemático-estadístico que nos dice cómo utilizar eficientemente la evidencia disponible. No todo uso de la evidencia es admisible. Para serlo debe de cumplir determinadas condiciones que son las que vamos a estudiar en esta sección.

Cuando se contrasta una hipótesis hay que tomar una decisión entre aceptarla o rechazarla. Como yo no sé si la hipótesis que estoy contrastando es o no cierta, y es por eso precisamente por lo que la contrasto, entonces puedo cometer dos errores: rechazar la hipótesis cuando es cierta y aceptarla cuando es falsa. Un contraste será bueno si minimiza, en algún sentido, la posibilidad de cometer estos dos errores.

Comenzaremos definiendo la hipótesis nula como la hipótesis definida al comienzo de la sección, es decir H_0. La hipótesis alternativa, que escribiremos H_1, será aquella que especifica que θ pertenece a la región complementaria de Θ_0 en Θ.

Definición 3.1: **Error Tipo 1 y Error Tipo 2**

El Error Tipo 1 se comete cuando se rechaza la hipótesis nula siendo cierta. Corresponde a aquellos casos en que la muestra pertenece a S_0 siendo H_0 cierta.

El Error Tipo 2 se comete cuando se acepta la hipótesis nula en aquellos casos en que es falsa. Ahora la muestra pertenece a S_1 siendo H_0 falsa.

Asociado a cada tipo de error está su tamaño.

Definición 3.2: **Tamaños de los dos Errores**

El Tamaño del Error Tipo 1, que denotaremos con ε, es la probabilidad de cometer el Error Tipo 1; es decir:

$$\varepsilon = \text{Prob}_{\theta}\left\{ y \in S_0 \, / \, H_0 \text{ cierta} \right\} \tag{3.1}$$

El <u>Tamaño del Error Tipo 2</u>, que denotaremos con δ, es la probabilidad de cometer el Error Tipo 2; es decir:

$$\delta = \text{Prob}\left\{ y \notin S_0 \, / \, H_0 \text{ falsa} \right\} \tag{3.2}$$

Para el caso en que tanto la hipótesis nula como la alternativa sean simples, es decir cuando sólo haya un tamaño del Error Tipo 1 y un solo tamaño del Error Tipo 2 entonces podemos definir un contraste admisible de la siguiente manera:

<u>Definición 3.3</u>: Entre dos contraste, A y B, decimos que A es **admisible** con respecto a B cuando si los dos tienen el mismo tamaño del Error Tipo 1, es decir, si $\varepsilon(A) = \varepsilon(B)$ entonces el contraste A tiene un tamaño del Error Tipo 2 menor o igual que el de B, es decir, $\delta(A) \leq \delta(B)$. Decimos que el contraste A es admisible cuando lo es con respecto a cualquier otro contraste.

Pero esta definición corresponde a una situación muy restringida: aquella que viene caracterizada por el contraste de una hipótesis nula que es simple frente a una hipótesis alternativa que es también simple. Cuando la hipótesis alternativa es compuesta, que es lo más frecuente, entonces para definir el conjunto de contrastes admisibles es necesario introducir nuevos conceptos.

<u>Definición 3.4</u>: **Función de Potencia (P(θ))**

La función de potencia de un contraste es una función que nos proporciona, para cada valor de θ, la probabilidad de rechazar la hipótesis nula; es decir:

$$P(\theta) = \text{Prob}_{\theta}\{y \in S_0\} \tag{3.3}$$

Otros dos conceptos relevantes son los de Tamaño y Potencia de un contraste.

<u>Definición 3.5</u>: **Tamaño y Potencia**

El <u>Tamaño</u> de un contraste es la máxima probabilidad de rechazar un punto de la hipótesis nula siendo cierta; es decir:

$$\text{TAMAÑO} = \varepsilon = \underset{\theta \in \Theta_0}{\text{máx}} \text{ Prob } (y \in S_0 \,/\, H_0 \text{ cierta})$$

La Potencia de un contraste es el valor que toma la Función de Potencia para valores de θ pertenecientes a H_1; es decir:

$$\text{POTENCIA} = P(\theta) \qquad\qquad \theta \in \Theta_1$$

Definición 3.6: **Contraste uniformemente más potente de tamaño ε. (UMP)**

Decimos que un contraste es UMP de tamaño ε cuando cumple:

(i) Tiene un tamaño igual a ε.

(ii) Su función de potencia toma siempre un valor superior a la de cualquier otro contraste que tenga el mismo tamaño.

Llegados aquí podemos dar una nueva definición de procedimiento de contraste admisible que se adecúa mejor a la nueva situación.

Definición 3.7: Cuando contrastamos hipótesis compuestas decimos que un contraste es admisible cuando es UMP cualquiera que sea el tamaño que se considere.

Pero en muchas situaciones resulta difícil encontrar un contraste que sea UMP bien porque es matemáticamente imposible o bien porque las dificultades de derivación son grandes y no puede llegarse a resultados satisfactorios.

Para resolver algunos de estos problemas se han seguido dos vías; la primera, ha consistido en restringir las categorías de contrastes que se comparan exigiendo que cumplan alguna exigencia adicional. Por ejemplo, se habla de contrastes insesgados, invariantes, etc... La otra vía ha consistido en adoptar un marco asintótico y redefinir el concepto de admisibilidad para adaptarlo a ese nuevo marco. Veamos con un poco más de detalle esta segúnda vía.

El contraste y la función de potencia se definen en términos de estadísticos cuyas propiedades se han derivado asintóticamente. Así, puede hablarse de la función de potencia asintótica que, siguiendo a Spanos (1986), escribiremos como:

$$\Pi(\theta) = \Pr\text{ob}\left\{y \in S_0^{\infty}\right\} \quad \forall \theta \in \Theta$$

en donde S_0^∞ es la región crítica determinada asintóticamente.

Más que hablar de un contraste habría que hablar de una secuencia de contrastes, siendo ahora el concepto relevante el de consistencia.

Definición 3.8: Se dice que una secuencia de contrastes con función de potencia $\prod(\theta)$ es **consistente** de tamaño ε si:

(i) $\max_{\theta\in\Theta_0} \prod(\theta) = \varepsilon$

(ii) $\prod(\theta) = 1 \quad \theta\in\Theta_1$

Entonces en este nuevo marco asintótico diremos que una secuencia de contrastes es admisible si es consistente sea cual sea el tamaño adoptado.

Establecidas estas definiciones pasamos a estudiar cómo se llega al conjunto de procedimientos admisibles. Vamos a comenzar prestando atención al principio de la Razón de Verosimilitud, considerando posteriormente, otros enfoques que han sido utilizados en Econometría para derivar contrastes admisibles.

La función de verosimilitud de las T observaciones la denotaremos L(θ) y, suponiendo que estas observaciones son independientes, la escribiremos como:

$$L(\theta) = \prod_{i=1}^{T} f(y_i;\theta) \tag{3.4}$$

Los estimadores máximo-verosímiles (MV) que denotaremos por $\tilde{\theta}$, son aquellos que maximizan el valor de la función de verosimilitud, pudiendose escribir:

$$L(\tilde{\theta}) = \sup_{\theta\in\Theta} L(\theta) \tag{3.5}$$

Los estimadores máximo-verosímiles restringidos, que denotaremos por $\tilde{\theta}_R$, son aquellos que maximizan el valor de la función de verosimilitud, pero cumpliendo las restricciones impuestas por la hipótesis nula. Podemos escribir:

$$L(\tilde{\theta}_R) = \sup_{\theta\in\Theta_0} L(\theta) \tag{3.6}$$

La función de verosimilitud reune toda la información contenida en los datos acerca de los parámetros y, por tanto, es el punto de referencia obligado a la hora de llevar a cabo la partición de S en S_0 y S_1.

Supongamos que C es una expresión que puede tomar cualquier valor en el intervalo (0,1). El marco general que proponemos para llevar a cabo la partición de S en S_0 y S_1 viene dado por:

S_0: incluye todas las muestras que cumplen que:

$$LR^* = \frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} \leq C$$

S_1: incluye todas las muestras que cumplen que:

$$LR^* > C$$

Cuando se toma la decisión de aceptar o rechazar una hipótesis nula, como ya hemos indicado, se pueden cometer dos tipos de errores: el error tipo 1 y el error tipo 2.

Ya hemos dicho que el error tipo 1 se comete cuando la hipótesis nula es rechazada siendo cierta. Es decir cuando:

$y \in S_0$, siendo H_0 verdadera o, bien, cuando:

$\frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} \leq C$, siendo H_0 verdadera.

El error de tipo 2 se comete cuando la hipótesis nula se acepta siendo falsa. Es decir, cuando:

$y \in S_1$, siendo H_0 falsa o bien, cuando:

$\frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} > C$, siendo H_0 falsa.

El tamaño del error tipo 1, se puede escribir ahora como:

$$\varepsilon = \text{Prob} \quad \left[\frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} \leq C / H_0 \text{ cierta} \right]$$

y el tamaño del error tipo 2 adopta la forma siguiente:

$$\delta = \text{Prob} \quad \left[\frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} > C / H_0 \text{ falsa} \right]$$

A continuación, vamos a demostrar cómo el uso del principio de la Razón de Verosimilitud nos lleva a contrastes admisibles.

Comenzando con el marco más restrictivo de hipótesis nula simple versus alternativa también simple, la respuesta está en el llamado Lema de Neyman-Pearson. El siguiente resultado da cuenta de este Lema aunque siguiendo la presentación que hace del mismo De Groot (1975).

<u>Resultado 3.1:</u> Sea A un procedimiento de contraste tal que la hipótesis H_0 se rechaza si $aL_0(y) \leq bL_1(y)$ y dicha hipótesis se acepta si $aL_0(y) > bL_1(y)$ siendo a y b dos constantes positivas y $L_0(y)$ y $L_1(y)$ los valores de la función de verosimilitud bajo H_0 y H_1, respectivamente. Entonces, para cualquier otro contraste B se cumple que:

$$a\varepsilon(A) + b\delta(A) \leq a\varepsilon(B) + b\delta(B).$$

<u>Prueba</u>: Supongamos, sin merma de generalidad, que la función de probabilidad f( ) es discreta. Podemos escribir:

$$a\varepsilon(\ ) + b\delta(\ ) = a \sum_{y \in S_0} L_0(y) + b \sum_{y \in S_1} L_1(y) =$$

$$= a \sum_{y \in S_0} L_0(y) + b \left[1 - \sum_{y \in S_0} L_1(y) \right] =$$
$$= b + \sum_{y \in S_0} [aL_0(y) - bL_1(y)]$$

Esta expresión alcanzará un mínimo cuando el segundo término de la derecha alcance su valor mínimo. Esto se logra cuando en la región crítica se incluyen aquellas muestras para las que se cumple que: a $L_0(y)$ $\leq$ b $L_1(y)$ o, equivalentemente, aquellas muestras para las que se cumple: $aL(\tilde{\theta}_R) \leq bL(\tilde{\theta})$, que es precisamente lo que define el procedimiento A. $\square$

A partir de este resultado, el Lema de Neyman-Pearson establece que si un procedimiento de contraste, A, se define en la forma indicada, entonces para cualquier otro procedimiento, B, tal que se cumple que $\varepsilon(B) \leq \varepsilon(A)$ necesariamente se cumple que $\delta(B) \geq \delta(A)$. Es decir, no puede existir otro contraste que tenga los dos tipos de error con tamaños menores que los correspondientes a A.

Este principio de que no puede haber un criterio de contraste que, teniendo el tamaño del error tipo 1 menor, tenga también el tamaño del error tipo 2 menor es seguramente el que explica el amplio uso que se ha hecho del criterio basado en la comparación de la razón de verosimilitud con una constante.

Pero la utilidad del principio de la Razón de Verosimilitud para derivar procedimientos admisibles no se limita a este marco restrictivo de hipótesis nula simple versus hipótesis alternativa simple. Es bien conocido y destacado en la literatura que el procedimiento de contraste tratado en <u>el principio de la Razón de Verosimilitud</u> es consistente en cualquier situación. Por lo tanto, asintóticamente nos proporciona métodos de contraste que son admisibles.

En consecuencia, hemos encontrado un método para definir el conjunto de procedimientos de contraste admisibles. Aunque no es el único y luego comentaremos otros, por el momento nos basta como ilustración.

Pero nos falta todavía cómo resolver la segunda cuestión que es la de elegir uno de los procedimientos entre los admisibles. Gráficamente este conjunto puede representarse tal como aparece en la Figura 3.1.

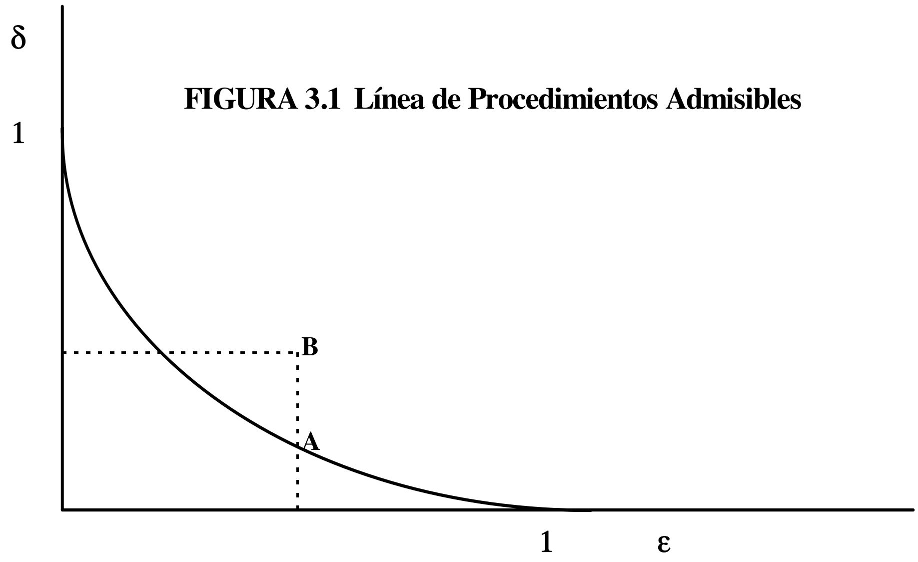


FIGURA 3.1 Línea de Procedimientos Admisibles

Esta línea nos proporciona, dado el tamaño de uno de los errores, el menor tamaño alcanzable por el otro error. Si un contraste garantiza la combinación B de los tamaños de los dos errores, la línea de procedimientos admisibles nos indica la existencia de otro contraste con el mismo tamaño del error tipo 1 pero con un tamaño del error tipo 2 menor, que es el que corresponde al punto A. Por lo tanto, cuando nos situemos en la línea de procedimientos admisibles podemos estar seguros de que no hay otra forma mejor de utilizar la evidencia disponible que la que corresponde a esta línea.

Pero como ya hemos dicho, queda una cuestión por resolver y es la de determinar el punto concreto de la línea en el que uno va a situarse. Una vez que se está seguro de que la evidencia se utiliza de forma óptima, hay que pensar en un par (ε_0, δ_0) que concrete la forma en que un contraste admisible va a ser utilizado. La primera cuestión, como ya hemos comentado, es puramente estadístico-matemática; la segunda tiene un carácter más complejo y entra dentro del esquema de preferencias y de decisión.

Para resolver esta cuestión de cómo determinar un punto concreto de la línea existen dos grandes enfoques en Econometría: Verificacionista y Preferencialista.

En el Enfoque Verificacionista se especifica, a priori, el tamaño del Error Tipo 1 quedando el tamaño del Error Tipo 2 como una consecuencia de esa especificación a priori. Hay que destacar que esa determinación a priori es independiente de la evidencia muestral. Lo más normal ha sido especificar un tamaño del Error Tipo 1 muy pequeño, digamos el 1% o el 5%, sin prestar atención al valor de δ resultante. Es como si la hipótesis nula fuera privilegiada en algún sentido proporcionándole un contrato blindado según el cual sólo será rechazada si la evidencia es claramente contraria a dicha hipótesis nula.

Este enfoque verificacionista tiene, al menos, dos inconvenientes. El primero es que, en muchas ocasiones, no hay ningún tipo de justificación para privilegiar una hipótesis frente a otra por lo que no hay razón para que se especifique, a priori, un valor de ε muy pequeño. Pero, si no se acepta que sea pequeño, entonces ¿cómo elegir su valor entre 0 y 1?. A falta de otras especificaciones sólo queda la arbitrariedad más absoluta. En segundo lugar, si la evidencia muestral crece (por ejemplo, si crece el tamaño muestral) entonces el tamaño del Error Tipo 1 se mantiene fijo mientras que el tamaño del Error Tipo 2 se hace más pequeño pudiendo, incluso, ser menor que el valor de ε fijado a priori, lo cuál sería contradictorio con el carácter privilegiado que parecía asignarse a la hipótesis nula.

La crítica a esta fijación a priori del valor de ε tiene una larga tradición dentro del enfoque bayesiano. Ver, por ejemplo, el libro de Leamer (1978). Los bayesianos siempre han defendido que el contraste de una hipótesis debe equipararse a la toma de una decisión y, para tomar correctamente una decisión debe adoptarse un marco de decisión en el que no falte ninguno de sus elementos básicos: información a priori sobre los estados de la naturaleza, función de pérdida y función de verosimilitud. La consideración conjunta de todos estos elementos nos conducirá a la adopción de la decisión asociada con el contraste de una hipótesis. Puede verse en el Capítulo 2 de Aznar (1989) las diferentes opciones abiertas en éste sentido.

En el segundo de los enfoques que hemos comentado, el Preferencialista, la determinación del par $(\varepsilon_0, \delta_0)$ se hace depender de la evidencia muestral. Se adopta una función de pérdida y, a partir de ella, se deriva la correspondiente función de riesgo para cada una de las hipótesis que se comparan. El contraste de hipótesis se hará a partir de la comparación de las estimaciones de estas funciones de riesgo. A posteriori, se demuestra que la regla resultante equivale a determinar un par concreto de los dos tamaños de los errores. Pero este par es una consecuencia de la evidencia muestral disponible. Ninguno de los elementos del par queda fijo al variar la evidencia. Ejemplos de las diferentes funciones de pérdida utilizadas en la literatura pueden verse en Aznar (1989) y en Aznar y Trívez (1993).

Examinemos ahora dos ejemplos que sirven como ilustración de algunos de los puntos comentados.

Ejemplo 3.1. Supongamos que disponemos de una muestra que se ha obtenido a partir de una población cuyos elementos siguen una distribución Normal con media μ y varianza conocida igual a σ^2.

La hipótesis nula que pretendemos contrastar viene dada por:

$$H_0: \mu = 0$$

La hipótesis alternativa es que $\mu \neq 0$.

El estimador MV sin restricciones puede escribirse como:

$$\tilde{\theta} = \tilde{\mu} = \bar{y}$$

El estimador MV con restricciones viene dado por:

$$\tilde{\theta}_R = \tilde{\mu}_R = 0$$

La función de verosimilitud correspondiente a la muestra es:

$$L(\theta) = (2\pi)^{-T/2} \left(\sigma^2\right)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(y_t - \mu)^2\right\}$$

Los valores que esta función toma introduciendo los estimadores MV sin y con restricciones son:

$$L(\tilde{\theta}) = (2\pi)^{-T/2} \left(\sigma^2\right)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(y_t - \bar{y})^2\right\}$$

$$L(\tilde{\theta}_R) = (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma y_t^2\right\}$$

Como ya hemos indicado, la regla de decisión asociada con el contraste puede formularse como: la hipótesis nula se rechaza cuando:

$$LR^* \leq C$$

o, equivalentemente, cuando:

$$\exp\left\{-\frac{1}{2\sigma^2}\Sigma y_t^2 + \frac{1}{2\sigma^2}\Sigma(y_t - \bar{y})^2\right\} = \exp\left\{-\frac{1}{2\sigma^2}T\bar{y}^2\right\} \leq C$$

Tomando logaritmos esta desigualdad es equivalente a:

$$-\frac{1}{2\sigma^2}T\bar{y}^2 \leq 1n\ C = C^*$$

o bien:

$$T\ \frac{\bar{y}}{\sigma^2}^2 \geq -\ 2C^*$$

Teniendo en cuenta que, bajo H_0, $\bar{y} \sim N\left[0, \frac{\sigma^2}{T}\right]$, la anterior expresión puede escribirse como:

$$\chi^2(1) \geq\ -\ 2C^* \tag{3.7}$$

Dentro del Enfoque Verificacionista, la constante C* y, como consecuencia, C, se determinan fijando previamente un nivel de significación $\varepsilon = \varepsilon_0$ y calculando C* de forma que:

$$\text{Prob}\ \{\chi^2(1) \geq -\ 2C^*\} = \varepsilon_0$$

En última instancia, el principio que subyace es que la hipótesis nula se rechaza siempre que la media muestral se distancie "sustancialmente" de 0.

Quizás sea ésta la razón que explica que en la literatura se haya seguido una vía alternativa consistente en adoptar como región crítica la siguiente:

$$|\bar{y}| > C^{**}$$

Esta región crítica puede escribirse también como:

$$\left|\frac{\sqrt{T}\bar{y}}{\sigma}\right| > \frac{\sqrt{T}C^{**}}{\sigma}$$

Por ser $\bar{y} \sim N\left(\mu, \frac{\sigma}{\sqrt{T}}\right)$, para determinar C**, especificamos el tamaño del error tipo I, $\varepsilon = \varepsilon_0$, y obtenemos:

$$\text{prob}\left\{\left|\frac{\sqrt{T}\bar{y}}{\sigma}\right| > \frac{\sqrt{T}C^{**}}{\sigma}\right\} = \varepsilon_0$$

que también puede escribirse:

$$\text{prob}\left\{|N(0,1)| > N\varepsilon_{0/2}\right\} = \varepsilon_0$$

El valor de $N\varepsilon_{0/2}$ está disponible en cualquier tabla de la Normal tipificada, por lo que:

$$C^{**} = \frac{\sigma\, N\varepsilon_{0/2}}{\sqrt{T}}$$

Se ve como los extremos de la Región Crítica dependen de tres factores:

1) El nivel de significación adoptado

2) El valor de la varianza

3) El tamaño muestral

Hay que destacar que cuando el único parámetro que se estima es el que es objeto de contraste entonces tanto el tamaño del Error Tipo 1 como la expresión C no dependen de la evidencia muestral. En el momento en el que se estime otro parámetro, en nuestro caso σ^2, entonces C deja de ser independiente de la evidencia muestral.

Ejemplo 3.2. Supongamos que se desea comparar entre los dos modelos siguientes:

$$M1: y_t = \beta_0 + \beta_1 X_{1t} + u_{1t}$$

$$M2: y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_{2t}$$

La comparación entre estos dos modelos equivale al contraste de la hipótesis nula:

$$H_0: \beta_2 = 0.$$

Suponemos que u_{2t} se distribuye como $N(0, \sigma^2)$ con σ^2 conocida.

La función de verosimilitud toma la forma siguiente:

$$L(\beta_0, \beta_1, \beta_2, \sigma^2) = (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma u_{2t}^2\right\}$$

Los estimadores MV sin restricciones pueden escribirse como:

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

en donde X es la matriz T x 3 de observaciones de las tres variables incluidas en M2. El vector de residuos sin restricciones definido como:

$$\tilde{u} = y - X\tilde{\beta}$$

Los estimadores MV con restricciones toman la forma siguiente:

$$\tilde{\theta}_R = \tilde{\beta}_R = \begin{bmatrix} \tilde{\beta}_{0R} \\ \tilde{\beta}_{1R} \\ \tilde{\beta}_{2R} \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ 0 \end{bmatrix}$$

en donde X_1 es la matriz T x 2 de observaciones de las variables incluidas en M1. El vector de residuos restringidos lo definimos como:

$$\tilde{u}_R = y - X\tilde{\beta}_R$$

Los valores que toma la función de verosimilitud después de sustituir los parámetros por, respectivamente, los estimadores MV sin y con restricciones son:

$$L(\tilde{\theta}) = (2\pi)^{-T/2} (\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\tilde{u}'\tilde{u}\right\}$$

$$L(\tilde{\theta}_R) = (2\pi)^{-T/2} (\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\tilde{u}_R'\tilde{u}_R\right\}$$

La Región Crítica del contraste de la hipótesis nula formulada anteriormente será:

$$\frac{L(\tilde{\theta}_R)}{L(\tilde{\theta})} = \exp\left\{-\frac{1}{2\sigma^2}\left(\tilde{u}_R'\tilde{u}_R - \tilde{u}'\tilde{u}\right)\right\} \leq C$$

y tomando logaritmos la hipótesis nula se rechaza cuando:

$$\frac{\tilde{u}_R'\tilde{u}_R - \tilde{u}'\tilde{u}}{\sigma^2} > -2 \ln C \qquad (3.8)$$

El término de la izquierda, bajo la hipótesis nula, sigue una distribución χ^2 con $k - k_1$ grados de libertad. Si se fija un ε_0, la expresión C se obtiene a partir de:

$$\chi^2_{\varepsilon_0}(k - k_1) = -2 \ln C$$

se ve como en este caso en que se conoce σ^2, la expresión C es independiente de la evidencia muestral. Si σ^2 no fuera conocido, entonces la expresión C pasaría a ser dependiente de la evidencia muestral.

Podría pensarse en sustituir σ^2 por el estimador MV resultando la siguiente Región Crítica:

$$\frac{\tilde{u}_R'\tilde{u}_R - \tilde{u}'\tilde{u}}{\tilde{u}'\tilde{u}/T} > -2 \ln C$$

y mediante algunas transformaciones elementales se llegaría a:

$$\frac{\tilde{u}_R'\tilde{u}_R}{\tilde{u}'\tilde{u}} > 1 - \frac{2}{T}\ln C. \tag{3.9}$$

Esta fórmula general que determina la Región Crítica nos servirá para derivar, en el Capítulo 8, la mayor parte de los criterios de selección de modelos conocidos en la literatura.

3.3 CONTRASTES LR, W y LM

Hemos visto en la Sección anterior que la Región Crítica de un contraste se determinaba a partir del cociente de los valores tomados por la función de verosimilitud sustituyendo los parámetros por sus estimaciones con y sin restricción. En particular, se definía una distancia entre los valores que toma la función de verosimilitud, y si esta distancia era grande se rechazaba la hipótesis nula.

Pero no hay un procedimiento único para dar cuenta de esta distancia. En la literatura se han desarrollado otras formas de tratar esa distancia y nosotros vamos a prestar atención a tres de ellas asociadas con los llamados contrastes de la Razón de Verosimilitud (LR), Wald (W) y Multiplicadores de Lagrange (LM).

En principio, la argumentación la llevaremos a cabo utilizando un caso muy simple en el que la muestra de tamaño T se ha obtenido a partir de una población con media μ y varianza σ^2 conocida.

En aquellos casos en que sea de interés para el contenido de otras secciones posteriores, los resultados se extenderán a un marco más general en el que la función de verosimilitud depende de un vector de k parámetros , θ , y se pretende contrastar las r restricciones lineales que escribimos como:

$$R\,\theta = q \tag{3.10}$$

En donde R es una matriz r.k y q es un vector de r elementos.

La función de verosimilitud para el caso simple viene dada por:

$$L(\mu, y, \sigma^2) = (2\pi)^{-T/2} (\sigma^2)^{-T/2} \exp\left\{ -\frac{1}{2\sigma^2} \Sigma (y_t - \mu)^2 \right\}$$

Su logaritmo puede escribirse como:

$$\ell(\mu) = \log L(\ ) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma (y_t - \mu)^2$$

Estimación MV sin restricciones ($\tilde{\mu}$)

El estimador MV sin restricciones es aquél que cumple que:

$$\ell(\tilde{\mu}) = \sup_{\forall \mu} \ell(\mu)$$

La condición necesaria para alcanzar el máximo es que el gradiente se iguale a cero. El gradiente se define como:

$$d(\mu) = \frac{\partial \ell(\mu)}{\partial \mu} = \frac{\Sigma\ (y_t - \mu)}{\sigma^2}$$

Por lo tanto, la condición necesaria puede escribirse:

$$d(\tilde{\mu}) = 0 \Rightarrow \tilde{\mu} = \bar{y}$$

El valor que toma el logaritmo de la función de verosimilitud sustituyendo los parámetros por los estimadores MV sin restricciones es:

$$\ell(\tilde{\mu}) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma (y_t - \tilde{\mu})^2$$

Estimación MV con restricciones ($\tilde{\mu}_R$)

Supongamos que queremos estimar μ cumpliéndose la restricción: $\mu=\mu_0$, siendo

μ_0 un valor conocido.

El estimador MV restringido de μ es aquel que cumple que:

$$\ell(\tilde{\mu}_R) = \sup_{\mu=\mu_0} \ell(\mu)$$

Para obtener la condición necesaria se define la función lagrangiana como:

$$L = \ell(\mu) + \lambda(\mu - \mu_0)$$

Las primeras derivadas de esta función respecto a μ y λ vienen dadas por:

$$\frac{\partial L}{\partial \mu} = d(\mu) + \lambda$$

$$\frac{\partial L}{\partial \lambda} = \mu - \mu_0$$

Las condiciones necesarias vienen dadas, por tanto, por:

$$d(\tilde{\mu}_R) + \tilde{\lambda} = 0 \Rightarrow d(\tilde{\mu}_R) = -\tilde{\lambda}$$

$$\tilde{\mu}_R = \mu_0$$

Podemos resumir los resultados obtenidos en el siguiente Cuadro:

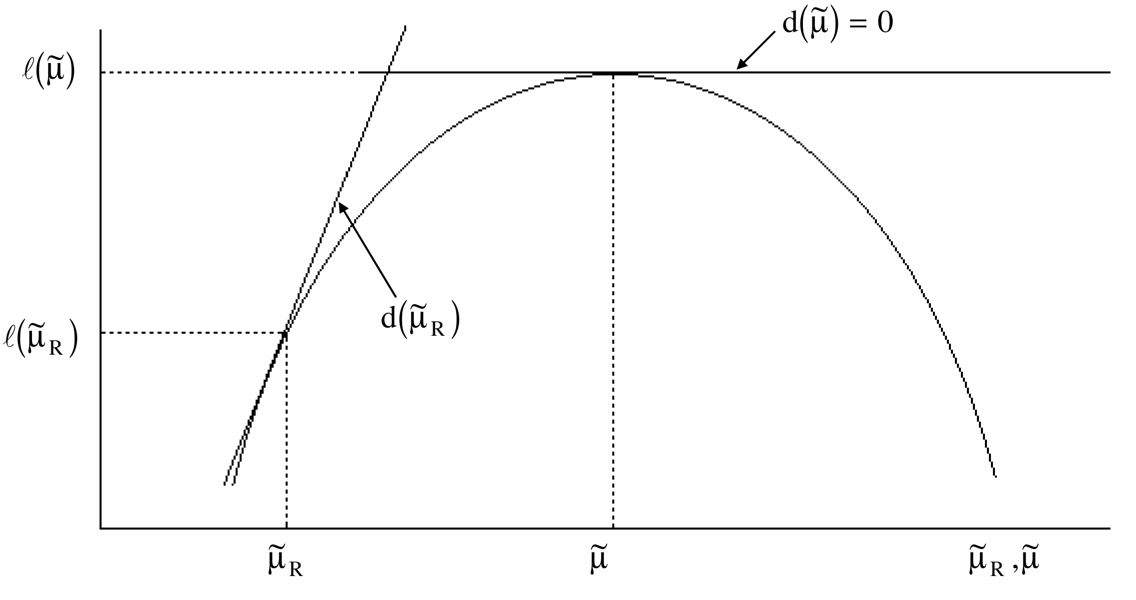
	SIN RESTRICCIONES	CON RESTRICCIONES
Logaritmo de la Func. de Verosimilitud	$\ell(\tilde{\mu})$	$\ell(\tilde{\mu}_R)$
Estimador MV	$\tilde{\mu}$	$\tilde{\mu}_R$
Gradiente	$d(\tilde{\mu}) = 0$	$d(\tilde{\mu}_R)$

Hemos comentado en la sección anterior que la Región Crítica se determina prestando atención a la distancia entre los valores tomados por la función de verosimilitud o, equivalentemente, por los tomados por sus logaritmos, es decir, $\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R)$.

Pero a similares resultados puede llegarse si prestamos atención a la distancia entre los dos estimadores o entre los valores tomados por el gradiente.

El siguiente gráfico puede ayudarnos a entender lo que subyace a estos tres planteamientos:

FIGURA 3.2. Ilustración de los Contrastes LR,W y LM.



Puede apreciarse cómo existe una relación monótona entre las tres medidas alternativas de la distancia. Conforme $\ell(\tilde{\mu}_R)$ se aproxima a $\ell(\tilde{\mu})$, $\tilde{\mu}_R$ se aproxima a $\tilde{\mu}$ y $d(\tilde{\mu}_R)$ se aproxima a $d(\tilde{\mu})$.

El contraste de cualquier hipótesis está basado en el siguiente principio: **"sólo se puede aceptar una restricción cuando su toma en consideración no distorsiona de forma relevante la evidencia contenida en los datos".**

El que una distorsión sea o no relevante puede ponerse en relación con las tres medidas de distancia comentadas.

Podemos decir que la distorsión es relevante cuando:

1) La distancia entre $\ell(\tilde{\mu})$ y $\ell(\tilde{\mu}_R)$ es grande.

2) La distancia entre $\tilde{\mu}$ y $\tilde{\mu}_R$ es grande.

3) La distancia entre $d(\tilde{\mu})$ y $d(\tilde{\mu}_R)$ es grande o, equivalentemente, si la distancia

entre $d(\tilde{\mu}_R)$ y cero es grande.

Ahora bien, la distancia entre estadísticos hay que establecerla prestando atención a la distribución estocástica que siguen dichos estadísticos.

Los dos resultados que se precisan para calibrar esas distancias son:

$$\sqrt{T}(\tilde{\mu}-\mu_0) \overset{H_0 \text{cierta}}{\rightarrow} N[0, \lim T\, I(\mu)^{-1}]$$

$$\frac{1}{\sqrt{T}} d(\tilde{\mu}_R) \overset{H_0 \text{cierta}}{\rightarrow} N[0, \lim \frac{I(\mu)}{T}]$$

(la demostración de estos resultados puede verse en Godfrey (1988))

en donde $I(\mu)$ es la matriz de información -en este caso en que μ es un escalar, elemento de información- definida como:

$$I(\mu) = - E\ \frac{\partial^2 \ell(\mu)}{\partial \mu^2} = \frac{T}{\sigma^2}$$

Para llegar a los estadísticos que sirven de base para el contraste necesitamos un resultado adicional que podemos formular como:

<u>Resultado 3.2</u>: Sea X un vector de p variables aleatorias distribuidas conjuntamente como: $N(0, \Sigma_p)$ siendo Σ_p una matriz no singular. Entonces se tiene que $X' \Sigma^{-1} X \sim \chi^2(p)$.

Veamos ahora la forma que toman los tres contrastes comentados.

<u>Contraste de Wald</u>: Está basado en la distancia $(\tilde{\mu}-\tilde{\mu}_R)$ o bien en la distancia $(\tilde{\mu}-\mu_0)$. Por los resultados comentados, podemos escribir:

$$W = \sqrt{T}(\tilde{\mu}-\mu_0)\, [\widehat{Var}\, \sqrt{T}(\tilde{\mu}-\mu_0)]^{-1}\ \sqrt{T}(\tilde{\mu}-\mu_0) =$$

$$= \sqrt{T}(\tilde{\mu}-\mu_0)\, [\, I(\tilde{\mu})/T\,]\ \sqrt{T}(\tilde{\mu}-\mu_0) = (\tilde{\mu}-\mu_0)\ I(\tilde{\mu})(\tilde{\mu}-\mu_0)$$

Asintóticamente y bajo la hipótesis nula este estadístico se distribuye como una χ^2

con un grado de libertad.

La forma de operar con este estadístico es la siguiente: se fija un tamaño de error tipo 1, $\varepsilon = \varepsilon_0$, se busca en las tablas el correspondiente punto crítico $\chi^2_{\varepsilon_0}(1)$ y la Región Crítica del contraste se determina como:

$$W > \chi^2_{\varepsilon_0}(1)$$

Para el caso general en el que lo que se pretende es contrastar (3.10) entonces el contraste toma la forma siguiente:

$$W = (R\tilde{\theta} - q)'[RI(\tilde{\theta})^{-1}R']^{-1}(R\tilde{\theta} - q) \tag{3.11}$$

y bajo H_0, W sigue una distribución χ^2 con r grados de libertad.

<u>Contraste de Multiplicadores de Lagrange (LM)</u>

Este contraste se basa en la distancia $d(\tilde{\mu}_R)$ - $d(\tilde{\mu})$, y teniendo en cuenta que $d(\tilde{\mu}) = 0$, simplemente en lo distante que $d(\tilde{\mu}_R)$ está de cero. Por los resultados comentados se tiene que:

$$LM = T^{-1/2} d(\tilde{\mu}_R)\left[V\hat{a}r\left(T^{-1/2} d(\tilde{\mu}_R)\right)\right]^{-1} T^{-1/2} d(\tilde{\mu}_R) = d(\tilde{\mu}_R) I(\tilde{\mu}_R)^{-1} d(\tilde{\mu}_R)$$

Asintóticamente y bajo la hipótesis nula este estadístico se distribuye como una χ^2 con un grado de libertad.

La Región Crítica del contraste es:

$$LM > \chi^2_{\varepsilon_0}(1)$$

Para el caso general asociado con el contraste de (3.10), el contraste de los Multiplicadores de Lagrange es:

$$LM = d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R) \tag{3.12}$$

LM bajo H_0 sigue una distribución χ^2 con r grados de libertad.

Contraste de la Razón de Verosimilitud

El contraste está basado en la distancia $(\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R))$. La distribución de este estadístico puede obtenerse fácilmente tras tomar una expansión de Taylor a partir de la de $\sqrt{T}(\tilde{\mu} - \mu_0)$ pudiéndose escribir:

$$LR = 2[\ell(\tilde{\mu}) - \ell(\tilde{\mu}_R)] \sim \chi^2(1) \tag{3.13}$$

la Región Crítica del contraste se determina como:

$$LR > \chi^2_{\varepsilon_0}(1)$$

Para el caso general, el contraste de la Razón de Verosimilitud es:

$$LR = 2[\ell(\tilde{\theta}) - \ell(\tilde{\theta}_R)] \tag{3.14}$$

LR bajo H_0 sigue una distribución χ^2 con r grados de libertad.

Hemos presentado una justificación intuitiva de los tres contrastes. Pero ahora la pregunta relevante es: ¿son admisibles estos tres contrastes?. Para muestras finitas se puede demostrar la admisibilidad en situaciones concretas como se verá en la sección siguiente. La admisibilidad con carácter general se ha demostrado asintóticamente; como puede verse en Godfrey (1988), los tres son consistentes garantizando el mejor uso de la evidencia disponible cuando ésta es grande.

Veamos en la sección siguiente como aplicar los diferentes procedimientos de contraste en el marco del Modelo Lineal General.

3.4. CONTRASTE DE HIPOTESIS EN EL MARCO DEL MODELO LINEAL GENERAL

Partiremos del modelo escrito en (1.3), suponiendo que se cumplen las ocho hipótesis introducidas en la Sección 1.2.

Nuestro objetivo es contrastar las r restricciones lineales escritas en (1.52). Para tomar la decisión de aceptar o rechazar esta hipótesis nula utilizaremos el siguiente estadístico:

$$F = \frac{(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/r}{\hat{\sigma}^2} \tag{3.15}$$

<u>Resultado 3.3</u>: Cumplidas las hipótesis de la Sección 1.2, el estadístico escrito en (3.15), bajo la hipótesis nula (1.52), sigue una distribución de probabilidad F central con r grados de libertad en el numerador y T-k en el denominador.

<u>Prueba</u>: Una forma equivalente de escribir (3.15) es:

$$F = \frac{(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/r\sigma^2}{\frac{1}{T-k} x \left(\frac{(T-k)\hat{\sigma}^2}{\sigma^2} \right)}$$

El numerador sigue una distribución χ^2 con r grados de libertad tomando el Resultado 1.8 y aplicando el Resultado 3.2 a $(R\hat{\beta} - q)$.

A partir de (1.32) se ve que el denominador es una distribución χ^2 con T-K grados de libertad dividida por esos grados de libertad. Por último el Resultado 1.17 establece la independencia necesaria entre numerador y denominador, probándose así el resultado. □

<u>Resultado 3.4</u>: Una forma equivalente de escribir el estadístico F escrito en (3.15) es la siguiente:

$$F = \frac{(\hat{u}_R'\hat{u}_R - \hat{u}'\hat{u})/r}{\hat{\sigma}^2} \tag{3.16}$$

en donde $\hat{u}$ es el vector de residuos MC0 definido en (1.13) y $\hat{u}_R$ es el vector de residuos MCR definido en (1.60).

<u>Prueba</u>: Por coincidir los denominadores, basta probar la equivalencia de los numeradores.

A partir de (1.53) se tiene que:

$$\hat{\beta} - \hat{\beta}_R = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \tag{3.17}$$

Por lo tanto, podemos escribir:

$$\left(\hat{\beta}-\hat{\beta}_R\right)' X'X(\hat{\beta}-\hat{\beta}_R) = (R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X$$

$$(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) =$$

$$= (R\hat{\beta}-q)'[R(X'X)^{-1}R]^{-1}(R\hat{\beta}-q)$$

que es precisamente el numerador de (3.15). Por otra parte, utilizando lo que hemos llamado la Identidad Fundamental en el Resultado 1.3 podemos escribir:

$$(\hat{\beta}-\hat{\beta}_R)'X'X(\hat{\beta}-\hat{\beta}_R) = \hat{u}_R'\hat{u}_R - \hat{u}'\hat{u} \tag{3.18}$$

llegándose así al resultado buscado tras corregir con los grados de libertad. □

La forma de determinar la región crítica utilizando el estadístico F se hace siguiendo lo que hemos llamado un enfoque verificacionista. Se especifica a priori un nivel del tamaño del Error Tipo 1, $\varepsilon = \varepsilon_0$, y la hipótesis nula se rechaza cuando:

$$F > F_{\varepsilon_0}\ (r, T\text{-}k) \tag{3.19}$$

Para estudiar la forma que toman los estadísticos LR, W y LM cuando se contrastan las r restricciones vamos a recopilar los resultados relevantes reescribiendo algunos de ellos. Primero, hay que tener en cuenta que:

$$\theta' = (\beta', \sigma^2)$$

El logaritmo de la función de verosimilitud y los estimadores MV de β y σ^2 los reescribimos como:

$$\ell(\theta) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \tag{1.39}$$

$$\tilde{\beta} = (X'X)^{-1}X'y \tag{1.43}$$

$$\tilde{\sigma}^2 = \frac{\tilde{u}'\tilde{u}}{T} \tag{1.44}$$

El gradiente puede verse en (1.40) y (1.41) y la matriz de información en (1.50).

Para obtener los estimadores MV restringidos la función Lagrange la escribimos como:

$$L = \ell(\theta) + \lambda'(R\beta - q) \tag{3.20}$$

en donde λ es el vector de r multiplicadores de Lagrange.

Las condiciones necesarias de máximo son:

$$\frac{\partial L}{\partial \beta} = 0; \qquad \frac{\partial L}{\partial \sigma^2} = 0; \qquad \text{y} \quad \frac{\partial L}{\partial \lambda} = 0$$

Teniendo en cuenta los k primeros elementos del gradiente en (1.40) se tiene:

$$\frac{1}{\tilde{\sigma}_R^2} X'(y - X\hat{\beta}_R) + R'\tilde{\lambda} = 0 \tag{3.21}$$

y teniendo en cuenta el correspondiente a σ^2 en (1.41):

$$-\frac{T}{\tilde{\sigma}_R^2} + \frac{\Sigma \tilde{u}_{Rt}^2}{\tilde{\sigma}_R^4} = 0 \tag{3.22}$$

A partir de (3.21) y (3.22) se obtienen los estimadores MV restringidos. El correspondiente a β coincide con el MC0 escrito en (1.53) y el correspondiente a σ^2 a partir de (3.22) es:

$$\tilde{\sigma}_R^2 = \frac{\tilde{u}_R' \tilde{u}_R}{T} \tag{3.23}$$

Hay que destacar que los elementos del gradiente correspondientes a β evaluados con los estimadores restringidos son diferentes a cero tal como se ve a partir de (3.21). El elemento del gradiente correspondiente a σ^2 evaluado con estimadores restringidos si que es cero como se deriva de (3.22).

<u>Resultado 3.5</u>: En el marco del Modelo Lineal General que cumple las hipótesis de la Sección 1.2 el estadístico que se utiliza para el contraste de la Razón de Verosimilitud

puede escribirse como:

$$LR = T \log \left(\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2} \right) \tag{3.24}$$

<u>Prueba</u>: Los valores que toma el logaritmo de la función de verosimilitud sustituyendo los parámetros de la misma escrito en (1.39) por sus estimadores sin restringir y restringidos son, respectivamente:

$$\ell(\tilde{\theta}) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \tilde{\sigma}^2 - \frac{\tilde{u}'\tilde{u}}{2\tilde{\sigma}^2} = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \tilde{\sigma}^2 - \frac{T}{2}$$

$$\ell(\tilde{\theta}_R) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \tilde{\sigma}_R^2 - \frac{\tilde{u}_R'\tilde{u}_R}{2\tilde{\sigma}_R^2} = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \tilde{\sigma}_R^2 - \frac{T}{2}$$

sustituyendo ahora en (3.13) se tiene:

$$LR = 2\left[-\frac{T}{2}\log \tilde{\sigma}^2 + \frac{T}{2}\log \tilde{\sigma}_R^2 \right] = T \log \frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2} \qquad \square$$

<u>Resultado 3.6</u>: En el mismo marco supuesto en el Resultado 3.5, el estadístico que se utiliza para el contraste de Wald puede escribirse como:

$$W = T \frac{(\tilde{\sigma}_R^2 - \tilde{\sigma}^2)}{\tilde{\sigma}^2} \tag{3.25}$$

<u>Prueba</u>: Teniendo en cuenta que $R^*\tilde{\theta} = R\tilde{\beta}$ y recordando la forma que toma la matriz de información escrita en (1.50), podemos escribir (3.11) como:

$$W = \frac{1}{\tilde{\sigma}^2}(R\tilde{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta} - q) \tag{3.26}$$

Notar que la expresión (3.11) debería escribirse utilizando la matriz R* en lugar de R, en donde R* es r x (k+1).

$$R^* = [R\ 0]$$

Los desarrollos contenidos en el Resultado 3.4 nos permiten escribir:

$$(R\tilde{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta}-q)=\tilde{u}_R'\tilde{u}_R-\tilde{u}'\tilde{u}$$

Utilizando esta identidad en (3.26) y multiplicando y dividiendo por T nos permite demostrar el resultado. □

Resultado 3.7: En el mismo marco asumido en el Resultado 3.5, el estadístico que se utiliza para el contraste de los Multiplicadores de Lagrange puede escribirse como:

$$LM=T\,\frac{(\tilde{\sigma}_R^2-\tilde{\sigma}^2)}{\tilde{\sigma}_R^2} \tag{3.27}$$

Prueba: Si evaluamos (1.40) y (1.41) con los estimadores restringidos se tiene que:

$$d(\tilde{\theta}_R)=\begin{bmatrix}\dfrac{1}{\tilde{\sigma}_R^2}X'\tilde{u}_R\\ 0\end{bmatrix}$$

Por lo tanto, (3.12) es ahora:

$$LM=d(\tilde{\theta}_R)'I(\tilde{\theta}_R)^{-1}d(\tilde{\theta}_R)=$$

$$=\begin{bmatrix}\dfrac{1}{\tilde{\sigma}_R^2}\tilde{u}_R'X & 0\end{bmatrix}\tilde{\sigma}_R^2\begin{bmatrix}(X'X)^{-1} & 0\\ 0 & \dfrac{2\tilde{\sigma}_R^2}{T}\end{bmatrix}\begin{bmatrix}\dfrac{1}{\tilde{\sigma}_R^2}X'\tilde{u}_R\\ 0\end{bmatrix}=$$ □

$$=\frac{1}{\tilde{\sigma}_R^2}\tilde{u}_R'X(X'X)^{-1}X'\tilde{u}_R$$

y como:

$$X'\tilde{u}_R=X'(y-X\tilde{\beta}_R)=X'y-X'X\tilde{\beta}_R=X'X(\tilde{\beta}-\tilde{\beta}_R)$$

podemos escribir:

$$\tilde{u}_R'X(X'X)^{-1}X'\tilde{u}_R=(\tilde{\beta}-\tilde{\beta}_R)'X'X(\tilde{\beta}-\tilde{\beta}_R)$$

y teniendo en cuenta (3.18) basta dividir y multiplicar por T para llegar al resultado. □

Un último resultado que tiene interés es el que relaciona estos tres estadísticos con el que corresponde al contraste de la F.

Resultado 3.8: Los contrastes LR, W y LM pueden escribirse como:

$$\mathrm{LR} = \mathrm{T}\log\left(1+\frac{\mathrm{rF}}{\mathrm{T}-\mathrm{k}}\right) \tag{3.28}$$

$$\mathrm{W} = \frac{\mathrm{T\,r\,F}}{\mathrm{T}-\mathrm{k}} \tag{3.29}$$

$$\mathrm{LM} = \frac{\mathrm{Tr}}{(\mathrm{T}-\mathrm{k})\mathrm{F}^{-1}+\mathrm{r}} \tag{3.30}$$

Prueba: La demostración es inmediata teniendo en cuenta que a partir de (3.16) podemos escribir:

$$\mathrm{F} = \frac{\mathrm{T}}{\mathrm{r}\,\hat{\sigma}^2}(\tilde{\sigma}_R^2 - \tilde{\sigma}^2) = \frac{\mathrm{T}-\mathrm{k}}{\mathrm{r}}\,\mathrm{x}\,\frac{\tilde{\sigma}_R^2-\tilde{\sigma}^2}{\tilde{\sigma}^2}$$

de donde:

$$\frac{\tilde{\sigma}_R^2-\tilde{\sigma}^2}{\tilde{\sigma}^2} = \frac{\mathrm{F\,r}}{\mathrm{T}-\mathrm{k}}$$

o equivalentemente:

$$\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}^2} = 1+\frac{\mathrm{F\,r}}{\mathrm{T}-\mathrm{k}}$$

Utilizando estas expresiones y sustituyéndolas en (3.24), (3.25) y (3.27) haciendo ligeras transformaciones se llega al resultado. □

Este resultado nos informa de que la región crítica es la misma para los cuatro contrastes pero siempre que se utilicen los puntos críticos correctos que se derivan de las expresiones (3.28), (3.29) y (3.30). Es decir, si se está utilizando el contraste de Wald, el punto crítico a utilizar sería el que se obtendría sustituyendo en (3.29) el punto crítico de la F, $\mathrm{F}_{\varepsilon_0}$, resultando:

$$\frac{T\, r\, F_{\varepsilon_0}}{T-k}$$

Ejemplo 3.3: Ilustración de cómo aplicar los contrastes F, LR,W y LM.

En este ejemplo se van a utilizar los cuatro contrastes para determinar si se acepta o no la hipótesis nula de rendimientos constantes formulada en el Ejemplo 1.1. Sabemos que el modelo es:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

en donde se cumplen las hipótesis ya comentadas en el Capitulo 1.

La hipótesis nula asociada con rendimientos constantes ya hemos dicho que es:

$$H_0 : \beta_2 + \beta_3 = 1$$

que también puede escribirse como:

$$R\beta = q$$

en donde: $R = (0, 1, 1)$ y $q = 1$

Función de Verosimilitud

$$L(\beta,\sigma^2; y, x) = (2\pi)^{-T/2} (\sigma^2)^{-T/2}$$

$$\exp\left\{-\frac{1}{2\sigma^2} \Sigma (y_t - \beta_1 - \beta_2 X_{2t} - \beta_3 X_{3t})^2\right\}$$

El logaritmo neperiano de esta función puede escribirse como:

$$\ell(\beta,\sigma^2) = -\frac{T}{2} \ln\ 2\pi - \frac{T}{2} \ln\ \sigma^2 -$$

$$-\frac{1}{2\sigma^2} \Sigma\ (y_t - \beta_1 - \beta_2 X_{2t} - \beta_3 X_{3t})^2$$

Gradiente

$$d(\theta)=\begin{bmatrix} d_1(\beta,\sigma^2) \\ d_2(\beta,\sigma^2) \\ d_3(\beta,\sigma^2) \\ d_4(\beta,\sigma^2) \end{bmatrix}=\begin{bmatrix} \dfrac{\partial l(\ )}{\partial \beta_1} \\ \dfrac{\partial l(\ )}{\partial \beta_2} \\ \dfrac{\partial l(\ )}{\partial \beta_3} \\ \dfrac{\partial l(\ )}{\partial \sigma^2} \end{bmatrix}=\begin{bmatrix} \dfrac{1}{\sigma^2}\Sigma u_t \\ \dfrac{1}{\sigma^2}\Sigma X_{2t}u_t \\ \dfrac{1}{\sigma^2}\Sigma X_{3t}u_t \\ -\dfrac{T}{2\sigma^2}+\dfrac{\Sigma u_t^2}{2\sigma^4} \end{bmatrix}$$

Matriz Hessiana de segundas derivadas:

$$D(\theta)=\frac{1}{\sigma^2}\begin{bmatrix} -T & -\Sigma X_{2t} & -\Sigma X_{3t} & -\dfrac{\sum u_t}{\sigma^2} \\ -\Sigma X_{2t} & -\Sigma X_{2t}^2 & -\Sigma X_{2t}X_{3t} & -\dfrac{\sum X_{2t}u_t}{\sigma^2} \\ -\Sigma X_{3t} & -\Sigma X_{3t}X_{2t} & -\Sigma X_{3t}^2 & -\dfrac{\sum X_{3t}u_t}{\sigma^2} \\ -\dfrac{\Sigma u_t}{\sigma^2} & -\dfrac{\Sigma X_{2t}u_t}{\sigma^2} & -\dfrac{\Sigma X_{3t}u_t}{\sigma^2} & \dfrac{T}{2\sigma^2}-\dfrac{\Sigma u_t^2}{\sigma^4} \end{bmatrix}$$

Matriz de información definida como la esperanza matemática de la matriz Hessiana después de cambiarla de signo:

$$I(\beta,\sigma^2)=-ED(\beta,\sigma^2)=\frac{1}{\sigma^2}\begin{bmatrix} T & \Sigma X_{2t} & \Sigma X_{3t} & 0 \\ \Sigma X_{2t} & \Sigma X_{2t}^2 & \Sigma X_{2t}X_{3t} & 0 \\ \Sigma X_{3t} & \Sigma X_{2t}X_{3t} & \Sigma X_{3t}^2 & 0 \\ 0 & 0 & 0 & \dfrac{T}{2\sigma^2} \end{bmatrix}=$$

$$=\frac{1}{\sigma^2}\begin{bmatrix} X'X & 0 \\ 0 & \dfrac{T}{2\sigma^2} \end{bmatrix} \tag{3.32}$$

La inversa de esta matriz viene dada por:

$$I(\beta,\sigma^2)^{-1} = \sigma^2 \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^2}{T} \end{bmatrix} = \sigma^2 \begin{bmatrix} a^{11} & a^{12} & a^{13} & 0 \\ a^{21} & a^{22} & a^{23} & 0 \\ a^{31} & a^{32} & a^{33} & 0 \\ 0 & 0 & 0 & \dfrac{2\sigma^2}{T} \end{bmatrix}$$

<u>Estimadores Maximo-Verosímiles sin restricciones</u> $(\tilde{\beta},\tilde{\sigma}^2)$

Estos estimadores se obtienen igualando a cero los cuatro elementos del gradiente y despejando:

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{\sigma}^2 \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'y \\ \dfrac{\Sigma\, \tilde{u}_t^2}{T} \end{bmatrix}$$

en donde ahora:

$$\tilde{u}_t = y_t - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2t} - \tilde{\beta}_3 X_{3t}$$

<u>Estimadores Maximo-Verosímiles Restringidos</u> $(\tilde{\beta}_R, \tilde{\sigma}_R^2)$

Para obtener estos estimadores se forma la función Lagrangiana así:

$$L = \ell(\beta,\sigma^2) - \lambda(\beta_2 + \beta_3 - 1)$$

El vector de primeras derivadas viende dado por:

$$\begin{bmatrix} \dfrac{\partial L}{\partial \beta_1} \\ \dfrac{\partial L}{\partial \beta_2} \\ \dfrac{\partial L}{\partial \beta_3} \\ \dfrac{\partial L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \ell(\beta,\sigma^2)}{\partial \beta_1} \\ \dfrac{\partial \ell(\beta,\sigma^2)}{\partial \beta_2} \\ \dfrac{\partial \ell(\beta,\sigma^2)}{\partial \beta_3} \\ \dfrac{\partial \ell(\beta,\sigma^2)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} d_1(\beta,\sigma^2) \\ d_2(\beta,\sigma^2) - \lambda \\ d_3(\beta,\sigma^2) - \lambda \\ d_4(\beta,\sigma^2) \end{bmatrix}$$

Los estimadores restringidos se obtienen igualando estos cuatro elementos a cero:

$$d_1(\tilde{\beta}_R, \tilde{\sigma}_R^2) = 0$$
$$d_2(\tilde{\beta}_R, \tilde{\sigma}_R^2) = \tilde{\lambda}$$
$$d_3(\tilde{\beta}_R, \tilde{\sigma}_R^2) = \tilde{\lambda}$$
$$d_4(\tilde{\beta}_R, \tilde{\sigma}_R^2) = 0$$

Despejando, y tal como se ha visto en el Resultado 1.25 del Capítulo 1, se obtienen los estimadores Máximo-Verosímiles Restringidos:

$$\begin{bmatrix} \tilde{\beta}_R \\ \tilde{\sigma}_R^2 \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_{1R} \\ \tilde{\beta}_{2R} \\ \tilde{\beta}_{3R} \\ \tilde{\sigma}_R^2 \end{bmatrix} = \begin{bmatrix} \tilde{\beta} - (X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} (R'\tilde{\beta} - 1) \\ \dfrac{\tilde{u}_R' \tilde{u}_R}{T} \end{bmatrix}$$

en donde:

$$\tilde{u}_{Rt} = y_t - \tilde{\beta}_{1R} - \tilde{\beta}_{2R} X_{2t} - \tilde{\beta}_{3R} X_{3t}$$

El estimador de $\tilde{\lambda}$ se obtiene utilizando la expresión $d_2(\cdot)$ y $d_3(\cdot)$ escritas en (3.31):

$$\tilde{\lambda} = \frac{1}{\tilde{\sigma}_R^2} \Sigma X_{2t} \tilde{u}_{Rt} = \frac{1}{\tilde{\sigma}_R^2} \Sigma X_{3t} \tilde{u}_{Rt}$$

Hay que destacar que los estimadores restringidos cumplen que:

$$\tilde{\beta}_{2R} + \tilde{\beta}_{3R} = 1 \text{ o bien}: \tilde{\beta}_{2R} = 1 - \tilde{\beta}_{3R}$$

Antes de utilizar estos estadísticos veamos algunos resultados previos. A partir de los resultados presentados en el Capítulo 1 se tiene que:

$$\hat{\beta} = \begin{pmatrix} -0{,}1773 \\ 0{,}2305 \\ 0{,}8073 \end{pmatrix} \quad ; \quad \hat{\sigma}^2 = \frac{0{,}0709}{24-3} = 0{,}003380 \quad ; \quad \tilde{\sigma}^2 = \frac{0{,}0709}{24} = 0{,}002954$$

$$R^2 = 1 - \frac{\hat{u}'\hat{u}}{\Sigma(y_t - \bar{y})^2} = 0{,}9574 \quad ; \quad \bar{R}^2 = 1 - \frac{T-1}{T-k}\left(\frac{\hat{u}'\hat{u}}{\Sigma(y_t - \bar{y})^2}\right) = 0{,}9533$$

$$(X'X)^{-1} = \begin{pmatrix} a^{11} & a^{12} & a^{13} \\ & a^{22} & a^{23} \\ & & a^{33} \end{pmatrix} = \begin{pmatrix} 55{,}7988 & 3{,}5207 & -17{,}6035 \\ & 1{,}1834 & -2{,}4556 \\ & & 6{,}213 \end{pmatrix}$$

y los restringidos:

$$\hat{\beta}_R = \begin{pmatrix} 0{,}0145 \\ 0{,}25413 \\ 0{,}74587 \end{pmatrix} \quad ; \quad \hat{\sigma}_R^2 = \frac{0{,}071643}{24-2} = 0{,}003256$$

$$\tilde{\sigma}_R^2 = \frac{0{,}071643}{24} = 0{,}002985$$

Apliquemos ahora los cuatro contrastes a la hipótesis nula $H_0 : \beta_2 + \beta_3 = 1$

Utilizando (3.16) el contraste F toma el siguiente valor:

$$F = \frac{(\hat{u}_R'\hat{u}_R - \hat{u}'\hat{u})/r}{\hat{\sigma}^2} = \frac{(0{,}071643 - 0{,}0709)/1}{0{,}003380} = 0{,}21982$$

Utilizando la expresión (3.15) se tiene que:

$$(R\hat{\beta} - q) = \hat{\beta}_2 + \hat{\beta}_3 - 1 = 0{,}0378$$

$$R(X'X)^{-1}R' = a^{22} + a^{32} + a^{23} + a^{33} = 2{,}4852$$

Por lo tanto:

$$F = \frac{(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/r}{\hat{\sigma}^2} =$$

$$= \frac{(0{,}0378)^2}{0{,}00338 \text{ x } 2{,}4852} = 0{,}18$$

La diferencia con el anterior se debe a las cifras decimales tomadas en el cálculo manual en el segundo caso.

Los grados de libertad son 1 en el numerador y 21 en el denominador. Si fijamos un nivel de significación del 5% el punto crítico es: $F_{0,05}(1,21) = 4{,}32$ por lo que la región crítica viene dada por:

$$F > 4{,}32$$

Dado el valor calculado para F, la hipótesis nula de rendimientos constantes no se rechaza con la evidencia disponible.

En lo que respecta al contraste de Wald en este caso, como se ha visto en la demostración del Resultado 3.6, coincide con el estadístico F, excepto en el denominador; bastaría para su cálculo considerar:

$$W = \frac{T}{T-k} F = \frac{24}{21} \text{ x } 0{,}18 = 0{,}2057$$

Para determinar la región crítica hay que utilizar el punto crítico que se deriva de (3.29) es decir:

$$W_{\varepsilon 0} = \frac{24 \times 4{,}32}{21} = 4{,}93$$

Entonces la Región Crítica sería:

$$W > 4{,}93$$

También en este caso se mantendría la hipótesis nula.

Puede seguirse un segundo camino consistente en aprovechar la distribución asintótica de W, que bajo la hipótesis nula sabemos que es $\chi^2(1)$, y tomar el punto crítico de esta distribución que es 3,82.

De acuerdo con (3.24) el contraste LR se puede calcular como:

$$LR = T\ \log\left(\frac{\tilde{\sigma}^2_R}{\tilde{\sigma}^2}\right) = 24\ x\ \log\left(\frac{0{,}002985}{0{,}002954}\right) = 0{,}2505$$

En este caso el punto crítico sería:

$$LR_\varepsilon = T\ \log\left(1 + \frac{r\,F_\varepsilon}{T-k}\right) = 24\ x\ \log\left(1 + \frac{4{,}32}{21}\right) = 4{,}48$$

Por lo tanto, también en este caso se mantendría la hipótesis nula.

Por último, en lo que respecta al contraste LM, a partir de (3.27) se tiene que:

$$LM = T\,x\,\frac{(\tilde{\sigma}^2_R - \tilde{\sigma}^2)}{\tilde{\sigma}^2_R} = 0{,}2492$$

y como a partir de (3.30) el punto crítico es 4,09, también en este caso se mantendría la hipótesis nula.

EJERCICIOS

3.1). Considerar una varible X distribuida como $N(\mu,\sigma^2)$ y suponer que la varianza es conocida. Se obtiene una muestra de tamaño T y se va a contrastar $H_0:\mu = 1$.

1). Derivar el gradiente y la matriz de informacion de la muestra y evaluar ambos utilizando los estimadores MV con y sin restricciones.

2). Contrastar la hipótesis nula utilizando los contrastes LR y LM. Derivar sus regiones críticas utilizando un estadístico definido en terminos de los datos eligiendo previamente el tamaño del error tipo 1.

3). Suponer que el contraste se lleva a cabo utilizando la siguiente region crítica: $LR^* \leq c$ para c=.5 y c=1.Derivar las regiones críticas correspondientes en términos de estadísticos definidos a partir de las observaciones muestrales.

4). Suponer que se adopta la siguiente región crítica: $|\bar{x}| \geq 2$. Derivar, aproximadamente, el tamaño del error tipo 1 que le corresponde en el caso en que $\sigma^2 = 4$ y T=16. Derivar tambien el tamaño del error tipo 2 para $\mu = 1{,}5$.

Obtener ambos tamaños en el caso en que T=36.

3.2). Considerar el siguiente modelo:

$$y_t = \beta + u_t$$

en donde suponemos que u es N (0, σ^2) y β es un parametro.Suponiendo que σ^2 es conocida y que se dispone de T observaciones se quiere contrastar $H_0 : \beta = 1$.

1). Obtener el gradiente y la matriz de información y evaluarlos utilizando los estimadores MV con y sin restricciones.

2). Utilizando los procedimientos de Wald y de los Multiplicadores de Lagrange contrastar la hipótesis nula especificando sus regiones críticas.

3). Suponer que se especifica la región crítica: $LR^* \leq$c para c=0,5 y 1. Obtener el tamaño del error tipo 1 en ambos casos.

4). Suponer que $\tilde{\beta}$ es el estimador MV. Se especifica la siguiente región crítica $|\tilde{\beta}| \geq 2$. Obtener el tamaño del error tipo 1 suponiendo que σ^2 es igual a 4 y T=16. Para esos mismos valores obtener el tamaño del error tipo 2 para un valor de β=1,5. Indicar cuales serían los tamaños en el caso en que $\sigma^2 = 0{,}64$.

3.3). Utilizando una muestra de tamaño T obtenida a partir de una población con media μ y varianza $\sigma^2=1$, se trata de contrastar la hipótesis nula $H_0:\mu = 0$, frente a la alternativa de que la media es diferente de cero. La región crítica que se propone para un nivel de significación del 5% es:

$$|\overline{X}| > 0{,}5$$

en donde $\overline{X}$ es la media muestral.Calcular el tamaño muestral e indicar cual sería la región crítica si el tamaño muestral fuera T=100.Tener en cuenta que:

$$\Pr ob[|N(0,1)| > 1{,}96] = 0{,}05$$

3.4). Considerar el siguiente modelo:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

en donde se supone que las variables estan medidas en desviaciones respecto a sus medias y la perturbación u cumple las hipótesis ideales.

a). Expresar las varianzas de los estimadores MCO de β en función de las varianzas muestrales de x_1 y x_2 (S_1^2 y S_2^2, respectivamente)y del coeficiente de correlación entre ambas variables (r_{12}).

b). Conociendo que $\sigma^2=1$, $S_1^2=2$, T=8, $\hat{\beta}_1$(MCO)=2, se pide determinar la región crítica para contrastar la hipótesis nula $H_0:\beta_1 = 0$ frente a la alternativa de que es diferente de cero para aquellos casos en que $r_{12}^2 = 0{,}02$ y $r_{12}^2 = 0{,}99$. Utilizar un tamaño del error tipo 1 igual a 0,05.

c). Obtener el tamaño del error tipo 2 para los dos casos contemplados en el apartado anterior suponiendo $\beta_1=1{,}8$.

3.5). Sea $x_1,.......,x_T$ una muestra aleatoria simple obtenida a partir de una población que sigue una distribución normal con media θ, desconocida, y varianza conocida igual a 1. Se va a contrastar la hipótesis nula $H_0:\theta = 0$ frente a la alternativa $H_1:\theta = 1$ utilizando un procedimiento que minimiza $\varepsilon(A)+b\,\delta(A)$, en donde $\varepsilon(A)$ y $\delta(A)$ son, respectivamente, los tamaños de los errores tipo 1 y tipo 2 de un contraste cualquiera A, y b es una constante positiva.

1). Se especifica $\varepsilon(A)=0.01$ y se pide determinar la región crítica del contraste, b y $\delta(A)$ para tres tamaños muestrales, T=1, 25, 100.

2). Se especifica b=0,16 y se pide determinar la región crítica del contraste, $\varepsilon(A)$ y $\delta(A)$ para tres tamaños muestrales, T=1, 25, 100.

3). Repetir 2) suponiendo que b=1.

3.6). Sea $y_1, y_2, ...y_T$ una muestra aleatoria simple obtenida a partir de una población normal con media μ y varianza σ^2. Sea $\bar{y}$ la media muestral. Se pide:

1). Demostrar que $\frac{\bar{y}-\mu}{\sigma}$ se puede escribir como $a'x$ siendo a y x vectores de T elementos con x distribuida como:N (0, I).

2). Demostrar que $\sum_{t=1}^{T}(y_t-\bar{y})^2$ se puede escribir como $\sigma^2(x'Ax)$ en donde A es una matriz idempotente. Obtener su rango.

3). Escribir $\frac{T(\bar{y}-\mu)^2}{\sigma^2}$ en la forma $x'Bx$ en donde B es una matriz idempotente. Obtener su rango.

4). Demostrar que Aa=0 y que AB=0 y comentar como se utilizarían todos estos resultados para contrastar una hipótesis acerca de μ.

3.7). Considerar los dos siguientes modelos lineales y no anidados:

M1: $y = X\beta + u_1$

M2: $y = Z\gamma + u_2$

M1 tiene k regresores y M2 tiene p regresores. El número de observaciones es T y cuando decimos que genera los datos M1 lo que se quiere decir es que $u_i \sim N(0, \sigma_i^2 I_T)$. Suponemos también que las matrices $T^{-1}X'X, T^{-1}Z'Z, T^{-1}X'Z$ convergen a matrices de constantes finitas.

1). Analizar el comportamiento asintótico de la relación de verosimilitud suponiendo que genera los datos M1.

2). A la vista de los problemas que se plantean en el punto anterior Cox propone utilizar el siguiente estadístico:

$$T_0 = \frac{T}{2}\log\frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_{21}^2}$$

en donde:

$\tilde{\sigma}_i^2$ es el estimador máximo-verosimil de σ_i^2 i=1,2.

$\tilde{\beta}$ es el estimador MCO de β definido a partir de M1.

$$\tilde{\sigma}_{21}^2 = [\text{plim}_1 \tilde{\sigma}_2^2]_{\theta=\tilde{\theta}_1} = \tilde{\sigma}_1^2 + \frac{1}{T}\tilde{\beta}' X' M_2 X \tilde{\beta}$$

plim_1 indica que la operación se lleva a cabo con respecto a M1.

y $M_2 = I - Z(Z'Z)^{-1}Z'$.

Se pide estudiar el comportamiento asintótico de este estadístico generando los datos cada uno de los dos modelos.

3.8) Cosiderar el modelo siguiente:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \qquad u \sim N(0, \sigma^2 I_T)$$

con:

$$X'X = \frac{1}{3}\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \qquad X'y = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \sigma^2 = 1$$

Se pide:

1). Utilizando los procedimientos LR, W, LM contrastar $H_0 : \beta_2 = 0$.

2). Calcular el estimador previo contraste de β_1 bajo la restricción introducida en 1) y adoptando un nivel de significación del 5%.

3). Comparar los tres estimadores de β_1: MCO, restringidos y previo contraste. Ordenarlos de mejor a peor e indicar como variaria la ordenación si los regresores fueran ortogonales.

Suponer que el verdadero valor de β_2 coincide con la estimación MCO del parámetro.
