Procedimientos robustos para el análisis de la varianza en un diseño en bloques aleatorizados completos

García Ben, Marta Susana

1988

Tesis presentada para obtener el grado de Doctor en Ciencias Matemáticas de la Universidad de Buenos Aires

UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

Tema de Tesis

PROCEDIMIENTOS ROBUSTOS PARA EL ANALISIS DE LA VARIANZA EN UN DISEÑO EN BLOQUES ALEATORIZADOS COMPLETOS

Autora

Marta Susana García Ben

Director de Tesis

Dr. Víctor J. Yohai

Lugar de Trabajo:

Facultad de Ciencias Exactas y Naturales

Tesis presentada para optar al título de
Doctor en Ciencias Matemáticas

1988

PROCEDIMIENTOS ROBUSTOS PARA EL ANALISIS DE LA VARIANZA EN UN DISEÑO EN BLOQUES ALEATORIZADOS

RESUMEN

En los diseños en bloques aleatorizados generalmente se emplean estimadores de cuadrados minimos, que no son robustos. Si se consideran I tratamientos, J bloques y una observacion por casilla y se supone el modelo lineal

$$X_{ij} = \alpha_i + \beta_j + u_{ij},$$

con u_{ij} variables aleatorias i.i.d., es natural aplicar el metodo de M-estimadores propuesto por Huber para el modelo lineal general a este caso particular con el propósito de obtener estimadores robustos. Sin embargo en este caso no se cumplen las hipótesis que aseguran la consistencia y normalidad asintótica de M-estimadores para modelos lineales (Huber (1973), Yohai y Maronna (1979), Portnoy (1984 y 1985)). En este trabajo se demuestran la consistencia y normalidad asintótica de los M-estimadores de los parámetros α_i, se compara su eficiencia asintótica con la de los estimadores de cuadrados mínimos, se estudia su punto de ruptura, se propone un test basado en estos M-estimadores y se compara por simulación la eficiencia de este test con el test del análisis de la varianza clásico y con dos tests de rangos.

1. INTRODUCCION.

Supóngase que se desean comparar I tratamientos y que para ello se observan estos tratamientos en J bloques, haciendo una observación por cada combinación tratamiento-bloque. Sea X_{ij} la observación del i-ésimo tratamiento en el j-ésimo bloque. Supóngase el siguiente modelo lineal, que es usual para este tipo de diseños:

$$\begin{cases} X_{ij} = \alpha_i + \beta_j + u_{ij}, \quad i=1,\dots,I, \quad j=1,\dots,J \\ \sum_{i=1}^{I} \alpha_i = 0 \\ u_{ij} \text{ i.i.d. con función de distribución } F_u. \end{cases}$$

Si se agrega a estas suposiciones la de que los errores aleatorios u_{ij} tienen distribución normal, el metodo de cuadrados mínimos proporciona los estimadores insesgados de mínima varianza para los parámetros α_i y β_j. Aunque la suposición de normalidad no se cumpla, los estimadores de cuadrados mínimos de los α_i siguen siendo fuertemente consistentes - cuando el número de bloques tiende a infinito - con sólo pedir que $E(u_{ij})$ sea finita y siguen teniendo distribución asintótica normal con sólo agregar la suposición de que $Var(u_{ij})$ sea finita. Pero el inconveniente que tienen estos estimadores es que no son robustos pues si la distribución de los errores aleatorios se aparta un poco de la normalidad, la distribución de los estimadores puede cambiar mucho y llegar así a ser muy ineficientes aún para distribuciones cercanas a la normal.

Para obtener estimadores robustos se puede aplicar el metodo de M-estimadores propuesto por Huber (1973) para el modelo lineal general a este modelo lineal particular.

Para estudiar la consistencia y distribución asintótica de los M-estimadores de los parámetros α_i (efectos de tratamientos) es natural suponer que el número de bloques tiende a infinito, mientras el número de tratamientos se mantiene constante. En estas condiciones el número de parámetros del modelo tiende a infinito con la misma velocidad que el número de observaciones y esto hace que no sean aplicables a este caso los resultados obtenidos sobre consistencia y distribución asintótica para el modelo lineal general. Por ello, luego de un resumen de algunos resultados conocidos sobre M-estimadores para el modelo de posición que se usan en este trabajo (sección 2) y de la definición de M-estimadores para el modelo de diseño en bloques (sección 3), se demuestra la consistencia (sección 4) y la normalidad asintótica (sección 5) de los estimadores de los efectos de los tratamientos.

Obtenida la distribución asintótica, en la sección 6 se compara la eficiencia asintótica de los M-estimadores de los efectos de tratamientos con la de los estimadores de cuadrados mínimos, para distintas distribuciones de los errores u_{ij}. En la sección 7 se estudia el punto de ruptura de los M-estimadores y se propone un M-estimador en dos pasos, con alta eficiencia y alto punto de ruptura.

No solamente los argumentos que se usan en la demostración de los resultados asintóticos para el modelo lineal general no son aplicables a este modelo, sino que la matriz de varianzas covarianzas asintotica resulta ser diferente. O sea que los resultados obtenidos en este trabajo sobre la distribucion asintotica, muestran que los estimadores de las varianzas asintóticas que proporcionan los programas para computadora de regresión mútiple robusta no son aplicables a este diseño, ni evidentemente a cualquier diseño donde el número de observaciones no sea mucho mayor que el numero de parámetros. Schrader y McKean (1979) y Schrader y Hettmansperger (1980) proponen un metodo para hacer tests robustos en el modelo lineal general, en reemplazo del análisis de la varianza clasico, que llaman "análisis de la varianza robusto", pero emplean la distribución asintotica de Huber (1973), por lo que su propuesta no es válida para muchos diseños, en particular para el que se estudia en el presente trabajo.

Dado que en la sección 5 se encontró la distribución asintótica de los M-estimadores de los efectos de los tratamientos, es natural proponer un estimador para la matriz de covarianza asintótica, que puede emplearse para hallar intervalos de confianza asintóticos y tests asintóticos para hipótesis sobre los efectos de los tratamientos: esto es lo que se hace en la sección 8. En esa misma sección se estudia el sesgo asintótico negativo que tiene el estimador de la matriz de covarianza asintótica basado en la teoría de Huber. Finalmente, en la seccion siguiente, se compara por simulación la eficiencia del test basado en los M-estimadores con respecto a los siguientes tests: el test F del analisis de la varianza, el de Friedman y el basado en rangos alineados.

2. RESULTADOS PREVIOS. Dada una función ρ y una muestra $X_1,\ldots,X_n$ se define el M-estimador de posición como la solución de

$$\sum_{i=1}^{n} \rho(X_i - \hat{\mu}) = \text{minimo!} \qquad (1)$$

Los M-estimadores pueden ser interpretados como una extensión del estimador de cuadrados mínimos ya que coinciden con éste en el caso particular en que $\rho(x)=x^2$. Si ρ tiene derivada no decreciente (o sea ρ es convexa) entonces (1) es equivalente a

$$\sum_{i=1}^{n} \psi(X_i - \hat{\mu}) = 0, \qquad (2)$$

donde $\psi=\rho'$.

Si las v.a. X_i son independientes igualmente distribuidas (i.i.d.) con distribución simétrica alrededor de μ_o, ψ es impar, no decreciente y acotada y $E(\psi(X-\mu))$ (considerada como función de μ) es estrictamente decreciente en μ_o, entonces $\hat{\mu}$ converge a μ_o con probabilidad uno. Si no se cumple la hipótesis de simetría, $\hat{\mu}$ converge con probabilidad uno al valor μ_o que satisface $E(\psi(X-\mu_o))=0$, suponiendo también que $E(\psi(X-\mu))$ es estrictamente decreciente en μ_o. Condiciones suficientes para asegurar que la función $E(\psi(X-\mu))$ se anule para un valor de μ y que sea estrictamente decreciente en ese punto son las siguientes: ψ es continua, acotada, $\psi(0)=0$, es no decreciente y estrictamente creciente en cero y el soporte de la distribución de la variable X es conexo.

El estimador $\hat{\mu}$ definido por (2) tiene el inconveniente de no ser equivariante por cambios de escala. Un estimador equivariante se obtiene resolviendo

$$\sum_{i=1}^{n} \psi((X_i-\hat{\mu})/s) = 0 \qquad (3)$$

donde s es cualquier estimador de la dispersión de la muestra. El estimador s puede ser calculado previamente a la resolución de (3) o puede también obtenerse en forma simultánea con $\hat{\mu}$, resolviendo simultáneamente las ecuaciones:

$$\Sigma\ \psi((X_i-\hat{\mu})/s) = 0$$

$$(1/n)\Sigma\ \chi((X_i-\hat{\mu})/s) = c$$

donde la función $\chi(u)$ se la elige habitualmente acotada, par, con $\chi(0)=0$ y no decreciente en $[0,+\infty)$ y c es una constante que que generalmente se la elige del siguiente modo:

$$c = E(\chi(Z)) \quad \text{para} \quad Z \sim N(0,1). \qquad (4)$$

Con esta expresión de c se logra que s sea un estimador consistente de la desviación standard, en el caso normal. Una elección posible para la función χ es $\chi=\psi^2$ (propuesta 2 de Huber (1964)).

Huber (1973) extendió la aplicación de los M-estimadores al modelo lineal general con variables independientes fijas y demostró su consistencia y normalidad asintótica. Tanto en su demostración como en la de Yohai y Maronna (1979) y en la de Portnoy (1984 y 1985) se supone que el número de observaciones tiende a infinito más rapidamente que el número de parámetros libres de la función de regresión.

3. M-ESTIMADORES PARA EL DISEÑO EN BLOQUES ALEATORIZADOS.

Sea X_{ij} la observación del i-ésimo tratamiento en el j-ésimo bloque. Supóngase el modelo lineal:

$$\begin{cases} X_{ij} = \alpha_i + \beta_j + u_{ij}, \quad i=1,\dots,I, \quad j=1,\dots,J \\ \sum_{i=1}^{I} \alpha_i = 0 \\ u_{ij} \text{ i.i.d. con función de distribución } F_u. \end{cases} \tag{5}$$

Como se comentó en la introducción, los estimadores de cuadrados mínimos son óptimos cuando F_u es normal, pero no son robustos. Para obtener estimadores robustos se puede aplicar el método de M-estimadores propuesto por Huber para el modelo lineal general a este caso particular, o sea en vez de minimizar la suma de los cuadrados de los residuos se minimiza la suma de una función ρ aplicada a los residuos, eligiendo ρ de modo que crezca menos rápidamente que la función cuadrática:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \rho(X_{ij} - \hat{\alpha}_i - \hat{\beta}_j) = \text{mínimo!} \qquad \sum \hat{\alpha}_i = 0. \tag{6}$$

Si ρ es convexa y $\psi=\rho'$ resolver (6) es equivalente a resolver:

$$\begin{cases} \sum_{i=1}^{I} \psi(X_{ij} - \hat{\alpha}_i - \hat{\beta}_j) = 0, \quad j=1,\dots,J, \\ \sum_{j=1}^{J} \psi(X_{ij} - \hat{\alpha}_i - \hat{\beta}_j) = 0, \quad i=1,\dots,I \\ \sum \hat{\alpha}_i = 0 \end{cases} \tag{7}$$

Los estimadores recién definidos no son equivariantes por cambios de escala. Para obtener esta propiedad se modifica la definición de M-estimadores, introduciendo un estimador de dispersión s del modo siguiente:

$$
\left\{
\begin{array}{ll}
\sum_{i=1}^{I} \psi((X_{ij} - \hat{\alpha}_i - \hat{\beta}_j)/s) = 0 , & j=1,\dots,J, \\
\sum_{j=1}^{J} \psi((X_{ij} - \hat{\alpha}_i - \hat{\beta}_j)/s) = 0 , & i=1,\dots,I, \\
\sum \hat{\alpha}_i = 0 &
\end{array}
\right. \tag{8}
$$

Entendemos por estimador de dispersión a un estadístico no negativo que cumple las dos propiedades siguientes:

i) si $X^*_{ij} = X_{ij} + a_i + b_j$ $i=1,\dots,I$, $j=1,\dots,J$, el estimador s calculado con la muestra $\{X^*_{ij}\}$ coincide con el calculado con $\{X_{ij}\}$;

ii) si $X^*_{ij} = \lambda.X_{ij}$, $i=1,\dots,I$, $j=1,\dots,J$, entonces

$$s(\{X^*_{ij}\}) = |\lambda|.s(\{X_{ij}\}).$$

El estimador de dispersión s puede ser calculado previamente a la resolución de las ecuaciones (8) o pueden resolverse simultáneamente las ecuaciones (8) y la ecuación

$$(1/(I-1)(J-1)) \sum_{i=1}^{I} \sum_{j=1}^{J} \chi((X_{ij} - \hat{\alpha}_i - \hat{\beta}_j)/s) = c , \tag{9}$$

donde la constante c cumple que $0<c<\mathrm{Max}\chi$. Se puede elegir c usando la expresión (22), para que s sea un estimador consistente de la desviación standard para el caso particular en que u_{ij} tenga distribución normal, pero esta elección no es necesaria para ninguno de los teoremas que se demuestran en este trabajo.

La existencia de solución del sistema simultáneo (8) y (9) para el modelo lineal general se discute en Huber (1981, sección 7.7). Considérense las siguientes hipótesis:

Suposición A: ρ es una función convexa, par, que tiene un mínimo estricto en cero, $\rho(0)=0$, ρ es derivable, $\psi=\rho'$ y $\chi(x)=x\psi(x)-\rho(x)$.

De la suposición A se deduce que ψ es no decreciente, impar, estrictamente creciente en cero y continua. χ es continua, par, no decreciente en $[0,+\infty)$ y $\chi(0)=0$.

En el apéndice A, siguiendo la idea de Huber (1981, sección 7.7), se demuestra que si se cumple la suposición A y si la función de distribución F_u del modelo (5) es continua entonces el sistema simultáneo (8) y (9) tiene solución, para todo J, con probabilidad uno. En los teoremas de la sección siguiente, en los que se demuestra la consistencia de los estimadores, no es necesario pedir que el sistema (8) y (9) tenga solución para todo J, sino solamente a partir de un J en adelante.

La siguiente propuesta de Huber (1964) cumple la suposición A:

$$\rho(x) = \rho_k(x) = (1/2).x^2 \quad \text{si } |x|\leq k$$
$$= k|x| - (1/2).k^2 \quad \text{en otro caso}$$
$$\psi(x) = \psi_k(x) = \rho_k'(x) = sg(x).\min(k,|x|)$$
$$\chi(x) = (1/2)(\psi(x))^2.$$

4. CONSISTENCIA DE LOS M-ESTIMADORES. Para estudiar el comportamiento de los M-estimadores para el diseño que estamos considerando, supondremos que el número de bloques tiende a infinito mientras el número de tratamientos se mantiene constante, lo que consideramos que es una suposición realista. En el modelo (5) el número de parámetros que varían libremente es I+J−1, el número de observaciones es I.J, luego ambos tienden a infinito con el mismo orden y por ello no pueden aplicarse los teoremas que aseguran la consistencia de los M-estimadores para el modelo lineal general (Huber(1973), Yohai y Maronna(1979), Portnoy(1984)), sino que se necesita hacer una demostración para este diseño particular.

Para poder obtener muchos de los resultados que se demuestran a continuación, se harán las siguientes suposiciones acerca de la función ψ:

Suposición B: ψ es continua, impar, no decreciente, estrictamente creciente en cero y acotada. Se nota $k_1 = \sup_{x\in R}\psi(x)$.

Cuando se estima simultáneamente la dispersión, se supondrá generalmente que se cumple la suposición A y que además:

Suposición C1: χ es acotada. Se notará $K_2 = \sup_{x\in R} \chi(x)$.

Para algunos resultados no será necesario suponer que las funciones ψ y χ están relacionadas, como en la suposición A, sino que bastará suponer que ψ cumple B y que χ cumple C1 y

Suposición C2: χ es continua y $\chi(0)=0$.

Las suposiciones A y C1 implican

$$\lim_{|x|\to\infty} \chi(x) = k_2.$$

Para demostrar el lema 6 se usarán las siguientes hipótesis:

Suposición D: ψ y χ cumplen la suposición A. ψ (y en consecuencia χ) es derivable, salvo en un conjunto finito de puntos. ψ' es estrictamente positiva en el conjunto $D = \{x: \psi(x) < \sup \psi\}$ (salvo en un conjunto finito de puntos).

El objetivo de esta sección es demostrar la consistencia fuerte de los estimadores, pero para ello va a ser necesario demostrar previamente que los estimadores están acotados con probabilidad uno. Los resultados principales de esta sección son los siguientes:

Teorema 1: Supóngase que se cumple el modelo (5).

a) Si se hallan los estimadores $\hat{\alpha}_i$ resolviendo (8), empleando cualquier estimador de dispersión, si ψ cumple la suposición B y s cumple que existen $B>0$ y $J_1(w)$ tales que

$$J \geq J_1(w) \quad \Rightarrow \quad 0 < s < B \text{ , con probabilidad uno,}$$

entonces existen $M>0$ y $J_o(w)$ tales que

$$J \geq J_o(w) \quad \Rightarrow \quad \max_i |\hat{\alpha}_i| \leq M$$

(Se denota w a un punto genérico del espacio de probabilidades donde están definidas las v.a. X_{ij}).

b) Si se hallan los estimadores $\hat{\alpha}_i$ y s resolviendo simultáneamente (8) y (9), si se supone que el sistema (8) y (9) tiene solución con probabilidad uno a partir de un J en adelante y si las funciones ψ y χ cumplen las suposiciones B, C1 y C2, entonces existen $M>0$, $B>0$ y $J_o(w)$ tales que

$$J \geq J_o(w) \quad \Rightarrow \quad \max_i |\hat{\alpha}_i| \leq M \text{ , } s \quad B$$

con probabilidad 1.

Teorema 2: Supóngase que se cumple el modelo (5) y que se calculan los estimadores resolviendo el sistema de ecuaciones simultáneo (8) y (9) (sistema que se supone tiene solución desde un J en adelante, con probabilidad uno), con funciones ψ y χ que cumplen las suposiciones A, C1 y D. Se supone que la distribución F_u cumple que

$$\sup_{x\in R} P(u=x) < (k_2-c)/(k_2\binom{I}{2})$$

Entonces existe $A>0$ y $J_o(w)$ tales que

$$J \geq J_o(w) \quad \Rightarrow \quad s>A\ .$$

Teorema 3: (consistencia de los estimadores de los efectos de tratamientos). Supóngase que se cumple el modelo (5) con F_u con soporte conexo. Se calculan los estimadores $\hat{\alpha}_i$ resolviendo (8) con cualquier estimador de dispersión s y con una función ψ continua, no decreciente, estrictamente creciente en cero, acotada y tal que $\psi(0)=0$. Se supone además que se cumplen las tesis de los teoremas 1 b) y 2.
Entonces

$$\lim_{J\to\infty} \hat{\alpha}_i = \alpha_i \quad \text{con probabilidad 1.}$$

Teorema 4: (consistencia del estimador de dispersión calculado simultáneamente). Supóngase que se cumple el modelo (5). Se obtienen los estimadores resolviendo simultáneamente (8) y (9). Las funciones ψ y χ cumplen las suposiciones C1 y D. El conjunto $\{x:\ \chi(x)<k_2\}$ es un intervalo acotado. Supóngase además que se cumple una de las dos hipótesis siguientes:

a) ψ (y por consiguiente χ) tiene derivada continua en todo punto y la distribución F_u es tal que

$$\sup_{x\in R} P(u=x) \quad (k_2-c)/\ k_2\binom{I}{2}.$$

o bien

b) ψ (y por consiguiente χ) tiene derivada continua salvo a lo sumo en un conjunto finito de puntos que no contiene al cero y F_u es continua.

Finalmente se supone que los estimadores $\hat{\alpha}_i$ son fuertemente consistentes.

Entonces existe $\sigma_o>0$ que es la única raíz de la ecuación

$$E((1/(I-1))\sum_{i=1}^{I}\chi((u_i - \beta(u,0,\sigma_o))/\sigma_o)) = c$$

y

$$\lim_{J\to\infty} s = \sigma_o \quad \text{con probabilidad 1.}$$

Para demostrar estos teoremas se necesitan algunos resultados previos.

Se usará la siguiente notación: si $X_1, \ldots, X_n$ son n números reales, $\hat{\mu}(X)$ designará a la solución de la ecuación

$$\Sigma\ \psi(X_i - \hat{\mu}) = 0. \tag{10}$$

Cuando ψ es continua, toma valores positivos y negativos y es estrictamente creciente esta ecuación tiene siempre solución única. Si ψ es no decreciente pero no es estrictamente creciente, la solución puede no ser única, pero el conjunto de soluciones será un intervalo. En este caso $\hat{\mu}(X)$ indicará el punto medio de ese intervalo.

Lema 1. Sean X_i $(i=1,\ldots,n)$; $Y_i = X_i + d_i$ $(i=1,\ldots,n)$ y sean $\hat{\mu}(X)$ y $\hat{\mu}(Y)$ los M-estimadores de posición definidos como en (10) con ψ continua, no decreciente, que toma valores positivos y negativos . Sea $M = \text{Máx}|d_i|$. Entonces

$$\hat{\mu}(X) - M \leq \hat{\mu}(Y) \leq \hat{\mu}(X) + M$$

La demostración de este lema para ψ estrictamente creciente es inmediata y la generalización para el caso ψ no decreciente se basa en un argumento similar al que se usará en el lema 2.

Lema 2. Sean X_i, Y_i, d_i, como en el lema 1 y $\Sigma\ d_i = 0$. Sea $M = \text{Máx}|d_i|$ y $R = \text{Máx}\ X_i - \text{mín}\ X_i$. Sean $\hat{\mu}(X)$, $\hat{\mu}(Y)$, los M-estimadores de posición definidos por (10) con ψ no decreciente, continua, impar (no idénticamente nula). Si $M \geq 2(n+1)R$, entonces

$$\hat{\mu}(X) - c_n.M \leq \hat{\mu}(Y) \leq \hat{\mu}(X) + c_n.M$$

donde $$c_n = (n + 1/2) / (n + 1).$$

Obsérvese que este lema proporciona, con más hipótesis, cotas más finas para $\hat{\mu}(Y)$ que el lema anterior.

Demostración: Se hará primero la demostración para el caso en que ψ es estrictamente creciente.

Solamente se probará la segunda desigualdad de la tesis, pues la demostración de la primera es similar. Para ello bastará probar que

$$\sum_{i=1}^{n} \psi(Y_i - ((\hat{\mu}(X) + c_n.M)) \quad \_\_ \tag{11}$$

es menor que cero.

En primer lugar se acotarán superiormente los valores de d_i. Para ello se ordenan las coordenadas de $\mathbf{d}=(d_1,\ldots,d_n)$ de menor a mayor, notando $d_{(i)}$ al i-ésimo valor ordenado. Se aplica a los vectores $\mathbf{X}$ e $\mathbf{Y}$ la misma permutación aplicada a $\mathbf{d}$ y se notan $X^{(i)}$ e $Y^{(i)}$ las i-ésimas coordenadas de estos vectores. Sea n impar. Como $\Sigma\, d_i = 0$ es

$$\sum_{i=(n+1)/2}^{n} d_{(i)} = - \sum_{i=1}^{(n-1)/2} d_{(i)} \qquad (12)$$

luego

$$((n+1)/2)\, d_{(n+1)/2} \leq \sum_{i=1}^{(n-1)/2} |d_{(i)}| \leq ((n-1)/2)\, M \, .$$

Resulta inmediatamente que

$$d_{(i)} \leq ((n-1)/(n+1)).M \qquad \text{para} \quad i=1,\ldots,(n+1)/2 \, . \qquad (13)$$

Sea ahora n par. De

$$\sum_{i=n/2}^{n} d_{(i)} = - \sum_{i=1}^{n/2-1} d_{(i)}$$

resulta

$$(n/2 + 1)\, d_{(n/2)} \leq \sum_{i=1}^{n/2-1} |d_{(i)}| \leq (n/2 - 1)\, M \, .$$

Luego

$$d_{(n/2)} \leq ((n/2-1)/(n/2+1)).M = ((n-2)/(n+2)).M \quad ((n-1)/(n+1))M$$

o sea

$$d_{(i)} \quad ((n-1)/(n+1)).M \qquad \text{para} \quad i=1,\ldots,n/2 \, . \qquad (14)$$

La expresión (11) se puede escribir

$$\sum_{i=1}^{n} \psi(X^{(i)} + d_{(i)} - \hat{\mu}(X) - c_n.M) \, . \qquad (15)$$

Sea

$$Z_i = X^{(i)} + d_{(i)} - \hat{\mu}(X) - c_n.M \, .$$

Para el caso en que n es impar, de (13), de la definición de M y de

$$X^{(i)} - \hat{\mu}(X) \leq R \leq M \,/\, 2(n+1) \qquad (16)$$

resulta

$$Z_i \leq - M/(n+1) \qquad \text{para} \quad i=1,\ldots,(n+1)/2$$

y

$$Z_i \leq M/(n+1) \qquad \text{para} \quad i=(n+3)/2,\ldots,n \, . \qquad (17)$$

Reemplazando (17) en la expresión (15) resulta

$$\Sigma\ \psi(Z_i) \leq ((n+1)/2)\ \psi(-M/(n+1)) + ((n-1)/2)\ \psi(M/(n+1))$$

y por ser ψ impar esta expresión es igual a

$$-\ \psi(M/(n+1))$$

que es estrictamente menor que cero, como deseábamos probar.

Para el caso n par, de (14), la definición de M y (16) resulta

$$Z_i < -\ M/(n+1) \quad \text{para} \quad i=1,\ldots,n/2$$

y $$Z_i \leq \ M/(n+1) \quad \text{para} \quad i=n/2+1,\ldots,n$$

y por lo tanto (15) se acota así:

$$\Sigma\ \psi(Z_i) < (n/2)\ \psi(-M/(n+1)) + (n/2)\ \psi(M/(n+1)) = 0 \qquad (18)$$

por lo que también en el caso n par la expresión (15) (o su equivalente (11)) es negativa y queda así demostrado el lema para ψ estrictamente creciente.

Para el caso ψ no decreciente pero no estrictamente creciente la demostración anterior no es válida, pues por ejemplo para el caso n par, en (18) habría que reemplazar el símbolo menor por menor o igual. Para demostrar que el lema sigue valiendo bastará designar

$$[\hat{\mu}_-(\mathbf{X}),\ \hat{\mu}_+(\mathbf{X})]$$

al intervalo de soluciones de la ecuación (10). Reemplazando en la demostración anterior $\hat{\mu}(\mathbf{X})$ por $\hat{\mu}_-(\mathbf{X})$, se concluye que

$$\Sigma\ \psi(Z_i) = \Sigma\ \psi(Y_i - (\hat{\mu}_-(\mathbf{X})+c_n.M)) \leq 0$$

de donde resulta que

$$\hat{\mu}_-(\mathbf{Y}) \leq \hat{\mu}_-(\mathbf{X}) + c_n.M\ . \qquad (19)$$

En forma análoga resulta que

$$\hat{\mu}_+(\mathbf{Y}) \geq \hat{\mu}_+(\mathbf{X}) - c_n.M\ . \qquad (19^*)$$

Intercambiando los papeles de **X** e **Y**, (19*) se convierte en

$$\hat{\mu}_+(\mathbf{X}) \geq \hat{\mu}_+(\mathbf{Y}) - c_n.M\ .$$

o, lo que es lo mismo

$$\hat{\mu}_+(\mathbf{Y}) \leq \hat{\mu}_+(\mathbf{X}) + c_n.M\ . \qquad (20)$$

De (19) y (20) resulta

$$\hat{\mu}(Y) = (\hat{\mu}_-(Y) + \hat{\mu}_+(Y)) / 2 \le$$

$$(\hat{\mu}_-(X) + \hat{\mu}_+(X)) / 2 + c_n.M = \hat{\mu}(X) + c_n.M$$

que es lo que se deseaba demostrar. Finaliza así la demostración del lema 2.

COMENTARIOS Y NOTACIONES. Debido a la equivariancia de los M-estimadores definidos por (8) para transformaciones de la forma $X^*_{ij} = X_{ij} + a_i + b_j$, se puede suponer sin perdida de generalidad que en el modelo (5) los parámetros α_i y β_j valen cero, o sea que $X_{ij} = u_{ij}$. En lo sucesivo, en muchas demostraciones se hace esta suposición para simplificar la notación.

Para cada $a \in R^I$, cada $t > 0$ y cada vector de residuos aleatorios $\mathbf{u} = (u_1,...,u_I)$, se denotará $\beta(\mathbf{u},\mathbf{a},t)$ la solución de la ecuación

$$\sum_{i=1}^{I} \psi((u_i - a_i - \beta(\mathbf{u},\mathbf{a},t))/t) = 0 \ . \qquad (21.1)$$

Cuando ψ no es estrictamente creciente (21.1) puede no tener solución única, en ese caso el conjunto de soluciones es un intervalo y $\beta(\mathbf{u},\mathbf{a},t)$ designará al punto medio de dicho intervalo.

Sea $\mathbf{u}_j = (u_{1j},...,u_{Ij})$ el vector de los valores observados en el bloque j. Que el vector $\hat{\alpha} = (\hat{\alpha}_1,...,\hat{\alpha}_I)$ sea solución del sistema (8) es equivalente a decir que cumple

$$\sum_{j=1}^{J} \psi((u_{ij} - \hat{\alpha}_i - \beta(\mathbf{u}_j,\hat{\alpha},s))/s) = 0 \ , \text{ para } i=1,...,I; \qquad (21.2)$$

$$\Sigma\, \hat{\alpha}_i = 0 \ ; \qquad (21.3)$$

donde s es un estimador de dispersión calculado previamente o resolviendo simultáneamente (21.2), (21.3) y

$$(1/(I-1)*(J-1)) \sum_{i=1}^{I} \sum_{j=1}^{J} \chi((u_{ij} - \hat{\alpha}_i - \beta(\mathbf{u}_j,\hat{\alpha},s))/s) = c, \qquad (21.4)$$

donde c es una constante que cumple $0<c<\text{Máx}\ \chi(x)$.

Generalmente, para lograr que en el caso en que las variables aleatorias u_{ij} tengan distribución $N(0,\sigma)$, s sea un estimador consistente de σ, se elige

$$c = E((1/(I-1)) \sum_{i=1}^{I} \chi(Z_i - \beta(Z,0,1)) \qquad (22)$$

donde $Z_1, \dots, Z_I$ son independientes $N(0,1)$.

Demostraremos a continuacion el teorema 1. Subdividiremos su demostración en los lemas 3 y 4.

Lema 3. Supóngase que se cumple el modelo (5) y que se calculan los estimadores resolviendo (8) y (9) (sistema simultáneo que se supone tiene solución con probabilidad uno, a partir de un $J(w)$ en adelante). Supóngase además que ψ y χ cumplen las suposiciones B, C1 y C2. Entonces existen $B_1>0$, $B_2>0$ y $J_1(w)$ tales que

$$J \geq J_1(w) \quad \Rightarrow \quad s \quad B_1 + B_2 . \underset{i}{\text{Máx}} \, |\hat{\alpha}_i| \text{ , con probabilidad uno.}$$

Demostración: Sea $k_2 = \max \chi$. Sean $u_1, \dots, u_I$, i.i.d. con distribución F_u. Sea M_1 tal que

$$P(\underset{1 \leq i \leq I}{\text{Máx}} \, |u_i| \quad M_1) < c.(I-1)/(4k_2 I) \text{ .}$$

Como $\chi(0)=0$ y χ es continua, existe $d>0$ tal que si $|x| \leq d$ entonces $\chi(x) < c(I-1)/2I$.

Demostraremos que existe $J_1(w)$ tal que

$$J \geq J_1(w) \quad \Rightarrow \quad s < 2.(M_1 + \text{Máx}|\hat{\alpha}_i|)/d \text{ , con prob. 1} \qquad (23)$$

lo que implica, llamando $B_1 = 2M_1/d$, $B_2 = 2/d$, la tesis del lema.

Sea $T = \{(t,\mathbf{a}) : \mathbf{a} \in R^I, \ \Sigma a_i = 0, \ t \geq 2(M_1 + \text{Máx}|a_i|)/d\}$.

Demostraremos la siguiente proposición que implica (23): existe $J_1(w)$ tal que

$$J \geq J_1(w) \quad \Rightarrow \underset{(t,\mathbf{a}) \in T}{\text{Máx}} ((1/(I-1)(J-1)) \sum_i \sum_j \chi(u_{ij} - a_i - \beta(\mathbf{u}_j,\mathbf{a},t))/t)) \quad c$$

con probabilidad uno. (24)

Si $\underset{i}{\text{Máx}}|u_{ij}| \leq M_1$ entonces $|u_{ij} - a_i| \leq M_1 + \text{Máx}|a_i| \ \forall i$, luego

$$\beta(\mathbf{u}_j,\mathbf{a},t) \leq M_1 + \text{Máx}|a_i|$$

y

$$|u_{ij} - a_i - \beta(\mathbf{u}_j,\mathbf{a},t)| \leq 2.(M_1 + \text{Máx}|a_i|) \text{ .}$$

Luego si $(t,\mathbf{a}) \in T$ entonces

$$|(u_{ij} - a_i - \beta(\mathbf{u}_j,\mathbf{a},t)/t| \leq d \text{ .}$$

Luego

$$\underset{(t,\mathbf{a})\in T}{\text{Máx}} \left((1/(I-1)(J-1)) \sum_j \sum_i \chi(u_{ij}-a_i-\beta(u_j,\mathbf{a},t))/t) \right) \leq$$

$$(1/(J-1)) \sum_{j=1}^{J} (1/(I-1)) . \Big((Ic(I-1)/2I).\mathrm{Ind}_{\text{Máx}|u_{ij}|\leq M_1}$$

$$+ I.k_2.\mathrm{Ind}_{\text{Máx}|u_{ij}|>M_1} \Big)$$

que es un promedio de v.a. i.i.d. con esperanza (por la elección de M_1) estrictamente menor que c. La ley de los grandes números implica la proposicion (24), que a su vez implica la tesis.

Lema 4. Supóngase que se cumple el modelo (5). Si se hallan los estimadores $\hat{\alpha}_i$ resolviendo (8), empleando cualquier estimador de dispersión que cumpla la tesis del lema 3 y si ψ cumple la suposición B, entonces existen $M>0$ y $J_o(w)$ tales que

$$J \geq J_o(w) \quad \Rightarrow \quad \underset{i}{\text{Máx}}|\hat{\alpha}_i| \leq M \text{ , con probabilidad uno.}$$

Demostración: Sean B_1 y B_2 las constantes que aparecen en **la tesis del lema 3.**

Sea $c_I = (I+1/2)/(I+1)$ la constante del lema 2.

Sea $h=(1-c_I)/B_2 > 0$.

Como $\psi(-h/2)<0$, se puede elegir $\varepsilon>0$ tal que

$$\psi(-h/2).(1-2\varepsilon) + k_1.2\varepsilon \quad 0.$$

Sea $$R_j = \underset{i}{\text{Máx}}\, u_{ij} - \underset{i}{\min}\, u_{ij} \text{ .}$$

Sea R_o tq. $$P(R_j \geq R_o) \quad \varepsilon \text{ .}$$

Sea M_2 tq. $$P(\underset{i}{\text{Máx}}|u_{ij}| \geq M_2) \quad \varepsilon \text{ .}$$

Como

$$\lim_{a\to+\infty} (2M_2 - (1-c_I)a)/(B_1+B_2a) = -(1-c_I)/B_2 = -h \text{ ,}$$

existe $A_o>0$ tal que si $a\geq A_o$ es

$$(2M_2 - (1-c_I)a)/(B_1+B_2.a) < -h/2 \text{ .}$$

Sea

$$M = \text{Máx } (A_o, 2(I+1)R_o) \text{ .}$$

Sea $S = \{(t,\mathbf{a}): \mathbf{a}\in R^I, \Sigma a_i=0, 0<t<(B_1+B_2.\text{Máx}|a_i|)\}$.

Se sabe por el lema 3 que

$$J \geq J_1(w) \quad \Rightarrow \quad (s,\hat{\alpha}) \in S \text{ , con prob. 1.}$$

Sea $$S_1 = S \cap \{(t,a): \ \text{Máx}|a_i| \quad M \} \ .$$

Sea $(t,a) \in S_1$.

Considérese el caso en que $\text{Máx } |a_i| = a_{i*} \quad M$.

Si $R_j \quad R_o$, aplicando el lema 2 resulta

$$u_{i*j} - a_{i*} - \beta(u_j,a,t) \leq u_{i*j} - \beta(u_j,0,t) - (1-c_I)a_{i*} \tag{25}$$

y si es además $\underset{i}{\text{Máx}}|u_{ij}| \leq M_2$, la expresión (25) resulta menor o igual que

$$2M_2 - (1-c_I)a_{i*} \ . \tag{26}$$

Como $(t,a)\in S$ y (26) es un número negativo, resulta

$$(u_{i*j} - a_{i*} - \beta(u_j,a,t))/t$$
$$(2M_2 - (1-c_I)a_{i*})/(B_1 + B_2 a_{i*}) \quad -h/2 \ ,$$

donde la última desigualdad surge de la elección de M y A_o.

Luego

$$\underset{\substack{(t,a)\in S_1 \\ \text{Máx } |a_i|=a_{i*}}}{\text{Máx}} (1/J) \ \Sigma \ \psi((u_{i*j} - a_{i*} - \beta(u_j,a,t))/t) \leq \tag{26.1}$$

$$\leq (1/J) \ \Sigma \ (\psi(-h/2).\text{Ind}_{R_j < R_o \text{ y } \text{Máx}|u_{ij}| < M_2} + k_1.\text{Ind}_{R_j \geq R_o \text{ ó } \text{Máx}|u_{ij}| \geq M_2})$$

que es un promedio de v.a. i.i.d. con esperanza menor que

$$\psi(-h/2).(1-2\varepsilon) + k_1.2\varepsilon \ ,$$

que es negativo por la elección de ε. Por la ley de los grandes números resulta que (26.1) es negativo para $J \geq J_{o,i*}(w)$, con probabilidad uno. En forma análoga, considerando el subconjunto de S_1 donde $\text{Máx}|a_i| = -a_{i*}$, resulta que

$$\underset{\substack{(t,a)\in S_1 \\ \text{Máx } |a_i|=-a_{i*}}}{\text{Min}} (1/J) \ \Sigma \ \psi((u_{i*j} - a_{i*} - \beta(u_j,a,t))/t) \tag{26.2}$$

es estrictamente positivo para $J \geq J_{1,i*}(w)$, con probabilidad uno.

Como (26.1) y (26.2) son diferentes de cero para J grande con probabilidad uno, resulta que si $(t,a)\in S_1$ no se pueden verificar las ecuaciones (21.2) para $J \geq J_o(w)$, con probabilidad uno (con $J_o(w)=\underset{i*}{\mathrm{Max}}(\mathrm{Máx}(J_{o,i*}(w),J_{1,i*}(w)))$. Luego

$$J \leq J_o(w) \quad \Rightarrow \quad (s,\hat{\alpha}) \in S - S_1 \text{ , con probabilidad uno ,}$$

que implica la tesis del lema 4.

<u>Demostración del inciso a) del teorema 1</u>:

El teorema 1 a) es idéntico al lema 4, salvo que la hipótesis acerca del estimador de dispersión "$s<B_1+B_2.\mathrm{Máx}|\hat{\alpha}_1|$, para J grande" se la ha reemplazado por la hipótesis más fuerte "$s<B$, para J grande".

<u>Demostración del inciso b) del teorema 1</u>:

La tesis del lema 4 es parecida a la del teorema 1 b), salvo que le falta la parte que afirma que "$s<B$". Pero esto surge de los lemas 3 y 4, llamando $B = B_1 + B_2.M$.

Para demostrar el teorema 2, se enunciarán y demostrarán previamente los lemas 5 y 6.

Lema 5. Se supone que ψ cumple las hipótesis del lema 1 y que χ es una función no negativa tq. $\lim_{|x|\to\infty} \chi(x) = k_2$, con $c<k_2$.

Sean $u_1,\ldots, u_I$ v.a. i.i.d. tq.

$$\sup_{x\in R} P(u_1 = x) \quad (k_2 - c)/(k_2.\binom{I}{2}). \tag{27}$$

Sea $a\in R^I$. Entonces existen $A>0$ y $\varepsilon>0$ (que no dependen de a) tq.

$$E\ (\inf_{0<t\leq A} (1/(I-1)) \sum_i \chi((u_i - a_i - \beta(u,a,t)/t))) >$$

<u>Demostración</u>: Se nota $b = \sup_{x\in R} P(u_1 = x)$.

Como por (27) es

$$k_2.(1 - b.\binom{I}{2}\ )$$

existe $\varepsilon>0$ tq.

$$(k_2 - \varepsilon).(1 - b\binom{I}{2}- \varepsilon\binom{I}{2}) > c+\varepsilon\ .$$

Sea M tq. $\quad |x|\geq M \quad \Rightarrow \quad \chi(x) \geq k_2 - \varepsilon$.

De la definición de b resulta que, para cada $x\in R$, existe $h = h(x)>0$ tq.

$$P(|u_1-x| \quad h) \quad b+\varepsilon \tag{27*}$$

Así elegido, h depende de pero considerando un intervalo $[-m,m]$ suficientemente grande para que la probabilidad de que u_i caiga afuera de el sea menos que $b+\varepsilon$ y luego usando la compacidad de $[-m-1,m+1]$, se puede elegir un $h>0$, que no depende de x, tal que (27*) se cumple para todo $x\in R$. Sea $\mathbf{a}$ cualquier vector de R^I, también se verifica que

$$P(|u_i - u_{i*} - (a_i - a_{i*}))| \quad h) \quad b + \varepsilon \quad \forall\ i\neq i^*, \qquad (28)$$

ya que

$$P(|u_i - u_{i*} - (a_i - a_{i*}))| < h) =$$

$$E(P(|u_i-(u_{i*}+a_i-a_{i*})|<h \mid u_{i*})) \quad E(b+\varepsilon) = b+\varepsilon.$$

Sea $A = h/(2.M)$. Obsérvese que

$$|u_i - a_i - \beta(\mathbf{u},\mathbf{a},t)| \quad h/2 \ , \ |u_{i*} - a_{i*} - \beta(\mathbf{u},\mathbf{a},t)| \quad h/2$$

$$\Rightarrow \quad |(u_i-a_i) - (u_{i*} - a_{i*})| \quad h \ . \qquad (28^*)$$

Sea B_{ii*} el suceso

$$B_{ii*} = \{|u_i - a_i - \beta(\mathbf{u},\mathbf{a},t)| < h/2 \ , \ |u_{i*} - a_{i*} - \beta(\mathbf{u},\mathbf{a},t)| < h/2\}.$$

Por (28) y (28*) es

$$P(B_{ii*}) \quad b + \varepsilon \quad \forall\ i\neq i^* \ .$$

Sea

$$B = \bigcup_{i\neq i^*} B_{ii*} \ .$$

Por la desigualdad de Boole es

$$P(B) \quad \binom{I}{2}.(b + \varepsilon) \ .$$

Luego

$$E(\inf_{0<t\leq A} (1/(I-1)) \sum_i \chi((u_i-a_i-\beta(\mathbf{u},\mathbf{a},t)/t)) \geq$$

$$\geq E(\inf_{0<t\leq A} (1/(I-1)) \sum_i \chi((u_i-a_i-\beta(\mathbf{u},\mathbf{a},t)/t).\mathrm{Ind}_{B^c})$$

$$\geq (k_2 - \varepsilon).P(B^c) \geq (k_2 - \varepsilon).(1 - \binom{I}{2} b - \binom{I}{2}\varepsilon)$$

que es mayor que $c+\varepsilon$ por la elección de ε, con lo que queda demostrado el lema 5.

Lema 6. Supóngase que ψ y χ cumplen la suposición D. Sean $y_1,\ldots,y_I$; $a_1,\ldots,a_I$; números reales. Entonces

$$G(t) = \sum_{i=1}^{I} \chi((y_i - a_i - \beta(\mathbf{y},\mathbf{a},t))/t) \qquad (29)$$

es función no creciente de (para $t>0$).

<u>Demostración:</u> Obsérvese que como por la suposición A (que esta incluida en la D) es $\chi(x)=x.\psi(x)-\rho(x)$, resulta que $\chi'(x)=x.\psi'(x)$ y de aquí $\chi'(x)=0$ sii $\psi'(x)=0$ y luego $D=\{x: |\psi(x)|<\sup \psi\}=\{x: \chi(x)<\sup \chi\}$.

Se hará primero la demostracion para el caso particular en que ψ es derivable en todo punto y que $\psi'(x)>0$ para todo $x\in D$.

Llamando $x_i = y_i - a_i$, (29) se escribe

$$G(t) = \sum_{i=1}^{I} \chi((x_i - \beta(x,0,t))/t). \tag{30}$$

Por la definicion de $\beta(x,0,t)$, se cumple

$$\sum_{i=1}^{I} \psi((x_i - \beta(x,0,t))/t) = 0 . \tag{31}$$

Derivando esta expresión con respecto a t resulta

$$\sum_{i=1}^{I} \psi'((x_i-\beta(x,0,t))/t).(-\partial\beta(x,0,t)/\partial t.t-(x_i-\beta(x,0,t)))/t^2 = 0,$$

de donde notando para simplificar $r_i = (x_i-\beta(x,0,t))/t$, resulta

$$\partial\beta(x,0,t)/\partial t = - (\Sigma\ \psi'(r_i).r_i)/(\Sigma\ \psi'(r_i)) \quad \text{si } \Sigma\ \psi'(r_i) \neq 0. \tag{32}$$

Si $\Sigma\ \psi'(r_i)=0$, la función (30) alcanza su valor máximo $I.k_2$.

La derivada de la expresión (30) con respecto a t es

$$\sum_{i=1}^{I} \chi'((x_i-\beta(x,0,t))/t).(-\partial\beta(x,0,t)\partial t.t-(x_i-\beta(x,0,t)))/t^2$$

que, usando la expresión (32) resulta igual a

$$(1/t)\ ((\Sigma\ \psi'(r_i).r_i)(\Sigma\chi'(r_i))/(\Sigma\ \psi'(r_i)) - \Sigma\ \chi'(r_i).r_i) \tag{33}$$

Reemplazando ahora $\chi'(r_i)$ por $\psi'(r_i).r_i$, la derivada de (30) resulta igual a

$$((\sum_{i=1}^{I} \psi'(r_i)r_i)^2 - (\sum_{i=1}^{I} \psi'(r_i)r_i^2).(\sum_{i=1}^{I}\psi'(r_i))) \ /\ (t.\sum_{i=1}^{I} \psi'(r_i)). \tag{34}$$

Por la desigualdad de Cauchy-Schwartz resulta que el numerador de (34) es menor o igual que cero. Como para $t>0$ el denominador es mayor que cero, resulta que la derivada de (30) es menor o igual a cero salvo a los sumo en los puntos en los que $\Sigma\ \psi'(r_i)=0$, pero en estos puntos la función (30) toma su valor máximo, esto implica que (30) es no creciente y queda así demostrado el lema 6, para el caso en que ψ es derivable en todo punto y que $\psi'(x)>0$ para todo $x\in D$.

Consideremos ahora el caso en que ψ no es derivable en un conjunto finito de puntos o ψ' se anula en un conjunto finito de puntos de D. Sea $\{a_1,...,a_r\}$ la unión de ambos conjuntos. En este caso se puede encontrar una sucesión de funciones ψ_n que coinciden con ψ en el complemento del conjunto

$$\bigcup_{i=1}^{r} (a_i-1/n, a_i+1/n),$$

que son no decrecientes, derivables en todo punto, con derivada estrictamente positiva en D y tales que,

$$\lim_{n\to\infty} \psi_n(x) = \psi(x), \quad \forall\ x.$$

LLamando

$$\rho_n(x) = \int_0^x \psi_n(x)dx \quad y \quad \chi_n(x) = x\psi_n(x)-\rho_n(x),$$

resulta que

$$\lim_{n\to\infty} \rho_n(x) = \rho(x) \quad y \quad \lim_{n\to\infty} \chi_n(x) = \chi(x)$$

para todo LLamando $\beta_n(x,0,t)$ a la solución de la ecuación

$$\sum_{i=1}^{I} \psi_n((x_i - \beta_n(x,0,t))/t) = 0,$$

como, por la definición de ψ_n, para n suficientemente grande se cumple

$$\psi_n(x-2/n) \leq \psi(x) \leq \psi_n(x+2/n) \quad \forall x,$$

resulta que

$$|\beta_n(x,0,t) - \beta(x,0,t)| \leq 2t/n,$$

luego para cada t fijo

$$\lim_{n\to\infty} \beta_n(x,0,t) = \beta(x,0,t).$$

Finalmente llamando

$$G_n(t) = \sum_{i=1}^{I} \chi_n((x_i - \beta_n(x,0,t))/t)$$

resulta que para cada t es

$$\lim_{n\to\infty} G_n(t) = G(t).$$

Como ψ_n y χ_n cumplen todas las hipótesis de este lema y ψ_n es derivable en todo punto, ya hemos demostrado que G_n es no decreciente; por ser $G=\lim G_n$ también G lo es y queda así demostrado el lema 6.

<u>Demostración del teorema 2</u>: Sea A y ε las constantes del lema 5. Se sabe por dicho lema que

$$E\left(\inf_{0<t\leq A} (1/(I-1)) \Sigma\, \chi((u_i - a_i - \beta(u,a,t))/t)\right) \quad c+\varepsilon .$$

Sea M la constante del teorema 1. Consideremos el compacto $C=\{a\in R^I : \text{Máx}|a_i|\leq M\}$. Por el lema 6 es

$$\inf_{\substack{0<t\leq A\\ a\in C}} (1/(J-1)) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i - \beta(u_j,a,t))/t) \qquad (35)$$

$$= \inf_{a\in C} (1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i - \beta(u_j,a,A))/A).$$

Por el lema 1

$$\max_{1\leq i\leq I} |a_i^*-a_i| \quad h \quad \Rightarrow \quad |\beta(u,a^*,t)-\beta(u,a,t)| \quad h,$$

para todo u y todo $t>0$. Como χ es continua y tiene límite finito para $|x|\to\infty$, la función $\chi(x/A)$ resulta uniformemente continua, luego existe $h>0$ tal que

$$\text{Max } |a_i^*-a_i| < h \quad \Rightarrow$$

$$|\chi((u_i-a_i^*-\beta(u,a^*,A))/A)-\chi((u_i-a_i-\beta(u,a,A))/A)| \quad \varepsilon(I-1)/2I \quad \forall u$$

$$\Rightarrow \quad | (1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i^* - \beta(u_j,a^*,A))/A) -$$

$$(1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i - \beta(u_j,a,A))/A)| \quad \varepsilon/2.$$

Consideremos un cubrimiento finito de C con bolas de radio h. Sean $a^{(q)}$, $q=1,\ldots,Q$, los centros de esas bolas.

Por la definición de ε y la ley de los grandes números, para cada $q=1,\ldots,Q$, existe $J_q(w)$ tal que

$$J\geq J_q(w) \quad \Rightarrow$$

$$(1/J)\sum_j (1/(I-1))\sum_i \chi((u_{ij}-a_i^{(q)} - \beta(u_j,a^{(q)},A))/A) \quad c + 3\varepsilon/4 ,$$

con probabilidad 1.

Sea $J_o(w) = \text{Max } (J_1(w),\ldots,J_Q(w))$. Sea $a\in C$, existe q tal que $|a - a^{(q)}| < h$, entonces, si $J\geq J_o$

$$(1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i - \beta(u_j,a,A))/A) \geq$$

$$\geq (1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i^{(q)} - \beta(u_j,a^{(q)},A))/A)$$

$$- |(1/J) \sum_j (1/(I-1)) \sum_i \chi((u_{ij} - a_i - \beta(u_j,a,A))/A) -$$

$$(1/J)\sum_j (1/(I-1))\sum_i \chi((u_{ij}-a_i^{(q)}-\beta(\mathbf{u}_j,\mathbf{a}^{(q)},A))/A)|$$

$$\geq c+3\varepsilon/4-\varepsilon/2=c+\varepsilon/4\ .$$

Luego, si $J\geq J_o(w)$, con probabilidad uno, la expresión (35) es estrictamente mayor que c, lo que implica la tesis del teorema 2.

Demostración del teorema 3:

Sea $\beta(\mathbf{u}_j,\mathbf{a},t)$ definido en (21.1). Sea, para cada i (i=1,...,I), $\hat{\mu}_i(\mathbf{a},t)$ la solución de la ecuación

$$\sum_{j=1}^{J}\psi((u_{ij}-\beta(\mathbf{u}_j,\mathbf{a},t)-\hat{\mu}_i(\mathbf{a},t))/t)=0\ . \qquad (37)$$

Se notará $\hat{\mu}(\mathbf{a},t)=(\hat{\mu}_1(\mathbf{a},t),...,\hat{\mu}_I(\mathbf{a},t))$.

El vector de los estimadores de los efectos de los tratamientos cumple

$$\hat{\mu}(\hat{\alpha},s)=\hat{\alpha} \qquad (38)$$

$$\sum\hat{\alpha}_i=0.$$

Sea $\mathbf{u}=(u_1,...,u_I)$ con $u_1,...,u_I$ v.a. i.i.d. con distribución F_u.

Sea

$$\Psi(\mathbf{u},\mathbf{a},t,\mathbf{m})=(\psi((u_1-\beta(\mathbf{u},\mathbf{a},t)-m_1)/t),...,\psi((u_I-\beta(\mathbf{u},\mathbf{a},t)-m_I)/t))$$

y sea $$\mathbf{g}(\mathbf{a},t,\mathbf{m})=E(\Psi(\mathbf{u},\mathbf{a},t,\mathbf{m}))\ .$$

Sea $D\subset R^I$ el subespacio de los vectores que tienen todas sus coordenadas iguales:

$$D=\{\ \mathbf{d}\in R^I:\ d_1=d_2=...=d_I\}\ .$$

Una de las hipótesis del teorema es que la distribución de las v.a. u_i tiene soporte conexo. Por lo tanto $\mathbf{u}$ tiene soporte conexo y $u_i-\beta(\mathbf{u},\mathbf{a},t)$ también por ser función continua de $\mathbf{u}$. Por el comentario hecho en la sección 2, este hecho y las hipótesis sobre ψ, aseguran que para cada $(\mathbf{a},t)$ existe un único $\mathbf{m}$ que anula $\mathbf{g}$, por esta razón y por ser u_i igualmente distribuídas resulta

$$\mathbf{g}(\mathbf{a},t,\mathbf{m})=0 \quad\Rightarrow\quad \mathbf{m}\in D\ . \qquad (39)$$

Consideremos la norma

$$\|\mathbf{a}\|=\max_i |a_i|\ ,$$

y la distancia definida por esta norma

$$\mathrm{dist}(\mathbf{a},\mathbf{b}) = \|\mathbf{a} - \mathbf{b}\| .$$

Sea $\varepsilon > 0$. Sea

$$D\varepsilon = \{v \in R^I : \mathrm{dist}(v,D) <$$

Sea
$$K = \{\mathbf{a} \in R^I : \|\mathbf{a}\| \leq M\} ,$$
donde M es la constante del teorema 1. Considérese el compacto $K - D\varepsilon$ y el compacto

$$C\varepsilon = K \times [A,B] \times (K - D\varepsilon)$$

donde A y B son las constantes de los teoremas 1 y 2. Por ser $\mathbf{g}$ continua y por la propiedad (39) resulta

$$\min_{(\mathbf{a},t,\mathbf{m}) \in C\varepsilon} \|\mathbf{g}(\mathbf{a},t,\mathbf{m})\| = b \quad 0 . \tag{40}$$

Para cada $(\mathbf{a},t,\mathbf{m}) \in C\varepsilon$ considérese un entorno de radio h y el supremo

$$\sup_{\|(\mathbf{a}^*,t^*,\mathbf{m}^*)-(\mathbf{a},t,\mathbf{m})\| \leq h} \|\boldsymbol{\psi}(u,\mathbf{a}^*,t^*,\mathbf{m}^*) - \boldsymbol{\psi}(u,\mathbf{a},t,\mathbf{m})\| .$$

Por continuidad, este supremo tiende a cero cuando h tiende a cero y por ser $\boldsymbol{\psi}$ acotada también tiende a cero su esperanza. Luego existe $h(\mathbf{a},t,\mathbf{m})$ tq.

$$E\Big(\sup_{\|(\mathbf{a}^*,t^*,\mathbf{m}^*)-(\mathbf{a},t,\mathbf{m})\| \leq h(\mathbf{a},t,\mathbf{m})} \|\boldsymbol{\psi}(u,\mathbf{a}^*,t^*,\mathbf{m}^*) - \boldsymbol{\psi}(u,\mathbf{a},t,\mathbf{m})\|\Big) < b/8 .$$

Estos entornos de radio $h(\mathbf{a},t,\mathbf{m})$ forman un cubrimiento de $C\varepsilon$ del que puede extraerse un subcubrimiento finito. Sean $(\mathbf{a}_q,t_q,\mathbf{m}_q)$ para $q=1,\dots,Q$, los centros de los entornos que forman el subcubrimiento finito y $h_q = h(\mathbf{a}_q,t_q,\mathbf{m}_q)$ sus radios. Para cada q ($q=1,\dots,Q$), por la ley de los grandes números existe $J_q(w)$ tal que

$$J \geq J_q(w) \quad \Rightarrow$$

$$(1/J) \sum_j \sup_{\|(\mathbf{a}^*,t^*,\mathbf{m}^*)-(\mathbf{a}_q,t_q,\mathbf{m}_q)\| \leq h_q} \|\boldsymbol{\psi}(u_j,\mathbf{a}^*,t^*,\mathbf{m}^*) - \boldsymbol{\psi}(u_j,\mathbf{a}_q,t_q,\mathbf{m}_q)\| \quad b/4,$$

con probabilidad uno. (41)

También por la ley fuerte, para cada q existe $J_q^*(w)$ tal que

$$J \geq J_q^*(w) \quad \Rightarrow$$

$$(1/J)\|\sum_j (\boldsymbol{\psi}(u_j,\mathbf{a}_q,t_q,\mathbf{m}_q) - \mathbf{g}(\mathbf{a}_q,t_q,\mathbf{m}_q))\| \quad b/4, \text{ con prob. 1.} \tag{42}$$

Sea ahora $(\mathbf{a},t,\mathbf{m})$ un punto cualquiera de $C\varepsilon$. Entonces existe un $q \in \{1,\dots,Q\}$ tq. $\|(\mathbf{a},t,\mathbf{m}) - (\mathbf{a}_q,t_q,\mathbf{m}_q)\| \leq h_q$. Luego

$$\|(1/J)\sum_{j}\psi(u_j,a,t,m)\| \quad \|g(a_q,t_q,m_q)\| -$$

$$-(1/J)\sum_{j}\sup_{\|(a^*,t^*,m^*)-(a_q,t_q,m_q)\|\leq h_q}\|\psi(u_j,a^*,t^*,m^*)-\psi(u_j,a_q,t_q,m_q)\|$$

$$-(1/J)\|\sum_{j}(\psi(u_j,a_q,t_q,m_q)-g(a_q,t_q,m_q))\| .$$

Por (40), (41) y (42) resulta que

$$J \geq J_o(w) = \text{Máx Máx } (J_q(w), J_q^*(w)) \quad \Rightarrow$$

$$\inf_{(a,t,m)\in C\varepsilon}\|(1/J)\sum_{j}\psi(u_j,a,t,m)\| \geq b-b/4-b/4 = b/2 > 0,$$ con probabilidad uno.

O, dicho de otro modo, hemos demostrado que

$$J \geq J_o(w) \quad y \quad (a,t,m)\in C\varepsilon \quad \Rightarrow$$

$$\sum_{j=1}^{J}\psi((u_{ij}-\beta(u_j,a,t)-m_i)/t) \neq 0, \text{ para algún } i=1,\ldots,I, \qquad (43.1)$$

con probabilidad uno.

Sean s y $\hat{\alpha}$ los estimadores de dispersión y de los efectos de los tratamientos que se sabe que satisfacen (37) y (38). De (43.1) resulta que

$$J \geq J_o(w) \quad \Rightarrow \quad (\hat{\alpha},s,\hat{\alpha}) \notin C\varepsilon = K \times [A,B] \times (K-D\varepsilon), \text{ con prob. } 1$$

Pero como por los teoremas 1 y 2, con probabilidad uno, para J grande es

$$\hat{\alpha}\in K \qquad s\in[A,B]$$

resulta que, para J grande es

$$\hat{\alpha} \in D\varepsilon \text{ , con prob. 1.} \qquad (43.2)$$

Como $\sum \hat{\alpha}_i=0$, de (43.2) resulta fácilmente que

$$J \geq J_o(w) \quad \Rightarrow \quad |\hat{\alpha}_i| \leq 2\varepsilon \text{ , } \forall\, i=1,\ldots,I \text{ , con prob. 1}$$

o sea

$$\lim_{J\to\infty}\hat{\alpha}_i = 0 \quad \text{con probabilidad 1,}$$

que es la tesis del teorema 3.

Para demostrar el teorema 4 se usarán los siguientes lemas:

Lema 7. Sean $u_1,\ldots,u_I$ variables aleatorias i.i.d. con distribución F_u continua. Se supone que ψ es continua, impar, no decreciente, estrictamente creciente en $\{x: |\psi(x)|<k_1\}$. Entonces para todo número $x\neq 0$ es

$$P((u_{i*}-\beta(\mathbf{u},0,\sigma_0))/\sigma_0 = x) = 0,$$

para todo $i*=1,\ldots,I$.

Demostración:

Para simplificar la notación, sin perder generalidad, supondremos $\sigma_0=1$.

Se desea demostrar que

$$P(u_{i*}-\beta(\mathbf{u},0,1) = x) = 0, \qquad (44.1)$$

para todo $x\neq 0$, para todo $i*=1,\ldots,I$.

Vamos a considerar primero el caso en que $|\psi(x)|<k_1$.

$$P(u_{i*}-\beta(\mathbf{u},0,1) = x) = P(\beta(\mathbf{u},0,1) = u_{i*}-x) \leq$$

$$P(\sum_{i=1}^{I}\psi(u_i-(u_{i*}-x)) = 0) = P(\sum_{i\neq i^*}\psi(u_i-(u_{i*}-x)) = -\psi(x)) =$$

$$E(P(\sum_{i\neq i^*}\psi(u_i-(u_{i*}-x)) = -\psi(x) \mid u_{i*})).$$

Por la independencia de $u_1,\ldots,u_I$ es

$$P(\sum_{i\neq i^*}\psi(u_i-(u_{i*}-x)) = -\psi(x) \mid u_{i*}=a) =$$

$$P(\sum_{i\neq i^*}\psi(u_i-(a-x)) = -\psi(x)). \qquad (44.2)$$

Llamemos c a la constante $c=a-x$. Como u_i tiene función de distribución F_u continua y ψ es estrictamente creciente salvo en el conjunto

$$|\psi(x)|=k_1\},$$

es

$$P(\psi(u_i-c) = d) = 0 \quad \forall d \text{ tq. } |d|\neq k_1.$$

Luego

$$P(\sum_{i\neq i^*}\psi(u_i-c) = d) = 0 \qquad (44.3)$$

salvo para los valores de d que se pueden obtener como suma de I-1 valores iguales a k_1 ó $-k_1$ (o sea $0,k_1,-k_1,2k_1,-2k_1,\ldots$).

Ya se sabe por hipótesis que $\psi(x)\neq 0$ y como estamos considerando el caso $|\psi(x)|<k_1$, $\psi(x)$ no puede ser ninguno de esos valores, luego (44.2) vale cero y ya está demostrado (44.1).

Consideremos ahora el caso en que $|\psi(x)|=k_1$. Supongamos que $\psi(x)=k_1$ (el caso $\psi(x)=-k_1$ es similar). En este caso la expresión (44.2) vale

$$P(\sum_{i\neq i^*}\psi(u_i - c) = -k_1),$$

que vale cero si I-1 es par, con lo que también está demostrado (44.1) para este caso.

Sólo resta considerar el caso $\psi(x)=k_1$ y I-1 impar. Sea A_i el suceso

$$A_i = \{\psi(u_i-\beta(\mathbf{u},0,1))=k_1 \text{ ó } \psi(u_i-\beta(\mathbf{u},0,1))=-k_1\}.$$

$$P(u_{i*}-\beta(\mathbf{u},0,1)=x) = P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1) =$$

$$P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1 \text{ y } \bigcap_{i\neq i^*} A_i) +$$

$$P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1 \text{ y } \bigcup_{i\neq i^*} A_i^c) \quad (44.4)$$

Consideremos el último término de (44.4):

$$P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1 \text{ y } \bigcup_{i\neq i^*} A_i^c)$$

$$\leq P(\sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1 \text{ y } \bigcup_{i\neq i^*} A_i^c),$$

y esta última expresión vale cero por los mismos argumentos que se emplean para justificar (44.3).

El primer término del último miembro de (44.4) es

$$P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \sum_{i\neq i^*}\psi(u_i-\beta(\mathbf{u},0,1))=-k_1 \text{ y } \bigcap_{i\neq i^*} A_i)$$

$$= P(u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \bigcap_{i=1}^{I} A_i).$$

Llamemos $(-x_o,x_o)$ al intervalo $D=\{x: |\psi(x)|<k_1\}$. O lo que es lo mismo

$$\{x: |\psi(x)|=k_1\} = (-\infty,-x_o] \cup [x_o,+\infty).$$

Sean $u_{(1)},\ldots,u_{(I)}$ los valores $u_1,\ldots,u_I$ ordenados de menor a mayor. Entonces

$$\bigcap_{i=1}^{I} A_i \quad \Rightarrow \quad \begin{array}{l} u_{(i)}-\beta(\mathbf{u},0,1) \in (-\infty,-x_o] \text{ para } i=1,\ldots,I/2; \\ u_{(i)}-\beta(\mathbf{u},0,1) \in [x_o,+\infty) \text{ para } i=I/2+1,\ldots,I. \quad (44.5) \end{array}$$

Si ocurre (44.5) entonces

$$\sum_{i=1}^{I}\psi(u_i-b)=0 \quad \Leftrightarrow \quad b\in [u_{(I/2)}+x_o;u_{(I/2+1)}-x_o]$$

y luego el punto medio es

$$\beta(\mathbf{u},0,1) = (u_{(I/2)}+u_{(I/2+1)})/2.$$

Luego

$$u_{i*}-\beta(\mathbf{u},0,1)=x \text{ y } \bigcap A_i \quad \Rightarrow$$

$$u_{i*} - (u_{(I/2)}+u_{(I/2+1)})/2 = x.$$

Por ser F_u continua

$$P(u_{i*} - (u_{(I/2)}+u_{(I/2+1)})/2 = x) = 0$$

y luego el primer término del último miembro de (44.4) es cero, que es lo que faltaba demostrar.

Lema 8. Supóngase que se cumplen las hipótesis de los lemas y 6. Sea $D=\{x: \psi(x)<k_1\}$ un intervalo acotado. Supóngase además que se cumple uno de las dos hipótesis siguientes:

a) ψ (y por consiguiente χ) tiene derivada continua en todo punto;

o

b) ψ (y por consiguiente χ) tiene derivada continua salvo a lo sumo en un conjunto finito de puntos que no contiene al cero y se cumple el lema 7.

Sean $u_1,...u_I$ v.a. i.i.d. con distribución F_u. Entonces

$$G(t) = E((1/(I-1)) \sum_{i=1}^{I} \chi((u_i - \beta(\mathbf{u},0,t))/t)) ,$$

(que es no creciente, por el lema 6 para $t>0$), es estrictamente decreciente en $\{t: t>0,\ G(t)\leq c\}$.

<u>Demostración</u>:

Sea $g(\mathbf{u},t) = (1/(I-1)) \sum \chi((u_i - \beta(\mathbf{u},0,t))/t)$

Entonces la función G a la que se refiere la tesis del lema es

$$G(t) = E(g(\mathbf{u},t)).$$

Sea $t>0$ fijo. Por ser $\partial g(\mathbf{u},t)/\partial t$ continua (para todo punto o salvo en un conjunto de probabilidad cero, por el lema 7) y acotada es

$$G'(t) = E(\partial g(\mathbf{u},t)/\partial t).$$

Se vió en el lema 6 que

$\partial g(\mathbf{u},t)/\partial t =$

$$\left(\sum_{i=1}^{I} \psi'(r_i)r_i\right)^2 - \left(\sum_{i=1}^{I} \psi'(r_i)r_i^2\right).\left(\sum_{i=1}^{I}\psi'(r_i)\right) \Big/ \left(t.\left(\sum_{i=1}^{I} \psi'(r_i)\right)\right).$$

donde se nota $r_i=(u_i - \beta(\mathbf{u},0,t))/t$.

Puede verse facilmente que el numerador de esta derivada es igual a

$$-\sum_{i<j} \psi'(r_i)\psi'(r_j)(r_i-r_j)^2.$$

Sea $B = \{\exists\ i\neq j \text{ tq. } r_i=r_j\} = \{\exists\ i\neq j \text{ tq. } u_i=u_j\}$

Por la hipótesis sobre F_u del lema 5 es

$$P(B) < (k_2-c)/k_2. \tag{45}$$

Sea $C^* = \{ \#\{i: \psi'(r_i)=0\} \geq (I-1) \}$. Obsérvese que si ocurre

C^* también ocurre

$$C = \{ \#\{i: \ \chi(x_i)=k_2\} \geq (I-1) \ \},$$

lo que implica que $g(\mathbf{u},t) \geq k_2$.

Entonces

si ocurre $C \ \Rightarrow \ \partial g(\mathbf{u},t)/\partial t \leq 0$ y $g(\mathbf{u},t) \geq k_2$,

si ocurre $B \ \Rightarrow \ \partial g(\mathbf{u},t)/\partial t \leq 0$,

si ocurre $B^c \cap C^c \ \Rightarrow \ \partial g(\mathbf{u},t)/\partial t \quad 0$.

Del párrafo anterior se deduce que

$$P(B^c \cap C^c) \qquad \Rightarrow \quad G'(t) \quad 0.$$

Si, por el contrario, $P(B^c \cap C^c)=0$, entonces $P(B \cup C)=1$ y de (45) se deduce que

$$P(C) \quad 1 - (k_2-c)/k_2,$$

luego

$$G(t) \geq E(g(\mathbf{u},t).\mathrm{Ind}_C) \quad k_2.(1-(k_2-c)/k_2) = c.$$

O sea que está demostrado que $G'(t)<0$ salvo a lo sumo cuando $G(t)>c$, lo que implica la tesis del lema.

<u>Demostración del teorema 4</u>. Del lema anterior resulta que si $u_1,\dots,u_I$ son v.a. i.i.d. con distribución F_u

$$G(t) = E((1/(I-1)) \sum_{i=1}^{I} \chi((u_i - \beta(\mathbf{u},0,t))/t))$$

es función estrictamente decreciente de t si $G(t) \leq c$. Como además el límite de $G(t)$ para $t \to \infty$ es cero y el límite para $t \to 0$ es mayor que c (por el lema 5), existe un único $\sigma_0 > 0$ que satisface la ecuación

$$G(\sigma_0) = E((1/(I-1)) \ \Sigma \ \chi((u_i - \beta(\mathbf{u},0,\sigma_0))/\sigma_0)) = c \ .$$

Sea $\varepsilon > 0$. Por la ley de los grandes números es, para J grande,

$$(1/(J-1))\sum_j (1/(I-1)) \sum_i \chi((u_{ij} - \beta(\mathbf{u}_j,0,(\sigma_0-\varepsilon))/(\sigma_0-\varepsilon)) > c \qquad (46)$$

$$\text{y} \ (1/(J-1))\sum_j (1/(I-1)) \sum_i \chi((u_{ij} - \beta(\mathbf{u}_j,0,(\sigma_0+\varepsilon))/(\sigma_0+\varepsilon)) < c \qquad (47)$$

Sean $\hat{\alpha}_i$ los estimadores de los efectos de los tratamientos. Como estos estimadores son fuertemente consistentes, dado $h>0$, a partir de un J en adelante es $|\hat{\alpha}_i|<h$ con probabilidad 1. Luego también es

$$|\beta(\mathbf{u}_j,\hat{\alpha},\sigma_0-\varepsilon) - \beta(\mathbf{u}_j,0,\sigma_0-\varepsilon)| \quad h$$

para todo j, por el lema 1. Dado que $\chi(x/(\sigma_0-\varepsilon))$ es función uniformemente continua de x, por ser continua y tener límite finito para $|x| \to \infty$, eligiendo h suficientemente pequeño es

$$\left| (1/(J-1))\sum_j (1/(I-1)) \sum_i (\chi((u_{ij}-\hat{\alpha}_i-\beta(u_j,\hat{\alpha},(\sigma_0-\varepsilon))/(\sigma_0-\varepsilon)) - \chi((u_{ij}-\beta(u_j,0,(\sigma_0-\varepsilon))/(\sigma_0-\varepsilon))) \right|$$

arbitrariamente cercano a cero para J grande, con probabilidad uno. De ello y (46) resulta que si J es grande es

$$(1/(J-1))\sum_j (1/(I-1)) \sum_i \chi((u_{ij} - \hat{\alpha}_i - \beta(u_j,\hat{\alpha},(\sigma_0-\varepsilon))/(\sigma_0-\varepsilon))>c$$

con probabilidad 1 y en forma análoga de (47) resulta que

$$(1/(J-1))\sum_j (1/(I-1)) \sum_i \chi((u_{ij} - \hat{\alpha}_i - \beta(u_j,\hat{\alpha},(\sigma_0+\varepsilon))/(\sigma_0+\varepsilon))<c.$$

Como $\hat{\alpha}_1,\ldots,\hat{\alpha}_I$ y s satisfacen (21.4) y por el lema 6 resulta que

$J \geq J_0(w) \quad \Rightarrow \quad \sigma_0 - \varepsilon \quad s < \sigma_0 + \varepsilon$, con probabilidad 1, o sea, la tesis del teorema 4.

5. DISTRIBUCION ASINTOTICA DE LOS M-ESTIMADORES. Los resultados obtenidos sobre distribución asintótica son los teoremas 5.1 y 5.2.

Teorema 5.1: (distribución asintótica). Supóngase que se cumple el modelo (5). Sea $\hat{\alpha}$ el vector de los M-estimadores, que resuelven el sistema de ecuaciones (8), con cualquier estimador de dispersión s. Se supone que ψ cumple la suposición (B), tiene derivada ψ' continua y positiva y $\psi'(x).x$ es acotada

Se supone además que cuando el número de bloques J tiende a infinito, $\hat{\alpha}_i$ converge a α_i y s tiende a $\sigma_0>0$, con probabilidad uno.

Como $\Sigma\hat{\alpha}_i=0$, la distribucion del vector aleatorio $\hat{\alpha}$ es degenerada, para evitar esto excluimos una de sus coordenadas, por ejemplo la última y llamamos

$$\hat{\alpha}^* = (\hat{\alpha}_1,\ldots,\hat{\alpha}_{I-1})^t$$

y

$$\alpha^* = (\alpha_1,\ldots,\alpha_{I-1})^t .$$

Entonces

$$J^{1/2}.(\hat{\alpha}^* - \alpha^*)$$

converge en distribución a una ley normal multivariada con media 0 y matriz de covarianza

$$\sigma_0^2.V(\psi,F_u).\mathbf{C} \tag{48}$$

donde

$$V(\psi, F_u) = v_i/a^2 .$$

Para simplificar se nota

$$r_i = (u_i - \beta(\mathbf{u}, 0, \sigma_0))/\sigma_0,$$

con esa notación es

$$v_i = E(\psi^2(r_i)), \tag{49}$$

$$a = E(\psi'(r_i).(-1 + (\psi'(r_i) - \psi'(r_I)) / \sum_{l=1}^{I} \psi'(r_l))) \tag{50}$$

y **C** es la matriz de $(I-1)\times(I-1)$ con $c_{ii}=1$ en la diagonal y $c_{ii^*} = -1/(I-1)$ para $i\neq i^*$.

<u>Demostración</u>: Al igual que en la sección anterior se supondrá, para facilitar la notación que en el modelo (5) $\alpha_i=0$ y $\beta_j=0$.

Para calcular la distribución asintótica de los estimadores de máxima verosimilitud y de los M-estimadores para distintos modelos, lo que se hace habitualmente es desarrollar por Taylor alrededor del punto al que convergen los estimadores, y justificar, bajo ciertas hipótesis, el poder despreciar el término complementario cuando el tamaño de la muestra es grande (por ejemplo Huber (1967), Huber(1973), Yohai(1987)). Para justificar el desprecio del término complementario, se usará en este trabajo un argumento similar al de Yohai(1987), que sirve para el caso que nos interesa: poder estudiar la distribución asintótica no solamente cuando el estimador de dispersión s se obtiene en forma simultánea, sino para cualquier estimador de dispersión fuertemente consistente.

En lo que sigue **a** es un vector de I componentes tal que $\Sigma a_i = 0$ y $\mathbf{a}^*$ es el vector formado por las (I-1) primeras componentes de **a**. Para cada $\mathbf{a}^* \in R^{(I-1)}$, cada $t>0$ y cada $\mathbf{u} \in R^I$ se nota

$\boldsymbol{\psi}^*(\mathbf{u}, \mathbf{a}^*, t) =$

$(\psi((u_1-a_1-\beta(\mathbf{u},\mathbf{a},t))/t), \ldots, \psi((u_{I-1}-a_{I-1}-\beta(\mathbf{u},\mathbf{a},t))/t))^t.$

Luego el sistema de ecuaciones (8) o su equivalente (21.2) y (21.3), con esta nueva notación equivale a

$$\sum_{j=1}^{J} \boldsymbol{\psi}^*(\mathbf{u}_j, \hat{\alpha}^*, s) = 0. \tag{51}$$

Seguidamente se buscarán expresiones para las componentes de la matriz $\partial\boldsymbol{\psi}^*/\partial \mathbf{a}^*$. Previamente se calculará el vector $\partial\beta/\partial \mathbf{a}^*$. Para simplificar las expresiones que se obtendrán, se usará la siguiente notación:

$$r_i(u,a^*,t) = (u_i-a_i-\beta(u,a^*,t))/t.$$

Derivando la ecuación (21.1) y despejando se obtiene

$$[\partial\beta(u,a^*,t)/\partial a^*]_i = \partial\beta(u,a^*,t)/\partial a_i =$$

$$(\psi'(r_I(u,a^*,t)) - \psi'(r_i(u,a^*,t))) \Big/ \sum_{l=1}^{I} \psi'(r_l(u,a^*,t)). \qquad (52)$$

De la definición de ψ^* resulta

$$[\partial\psi^*(u,a^*,t)/\partial a^*]_{ii} =$$

$$(1/t).\psi'(r_i(u,a^*,t)).(-1-\partial\beta(u,a^*.t)/\partial a_i)$$

que, usando (52) resulta igual a

$$(1/t).\psi'(r_i(u,a^*,t)).(-1 + (\psi'(r_i(u,a^*,t)) - \psi'(r_I(u,a^*,t)) \Big/$$

$$\sum_{l=1}^{I} \psi'(r_l(u,a^*,t))). \qquad (53)$$

En forma similar resulta para $i \neq i^*$

$$[\partial\psi^*(u,a^*,t)/\partial a^*]_{ii^*} =$$

$$(1/t).\psi'(r_i(u,a^*,t)) \;.\; (\psi'(r_{i^*}(u,a^*,t)) - \psi'(r_I(u,a^*,t)) \Big/$$

$$\sum_{l=1}^{I} \psi'(r_l(u,a^*,t)). \qquad (54)$$

Desarrollando el primer miembro de (51) por la fórmula de Taylor alrededor de 0 es

$$\sum_{j=1}^{J} \psi^*(u_j,\hat{\alpha}^*,s) = \sum_{j=1}^{J} \psi^*(u_j,0,s) + A \;.\; \hat{\alpha}^* = 0 \qquad (54.1)$$

donde se ha designado $A=A(u_1,\dots,u_J)$ a la matriz de $(I-1)\times(I-1)$ cuya i-ésima fila es igual al vector fila

$$\sum_{j=1}^{J} \partial[\psi^*(u_j,\tau_i,s)]_i/\partial a^*,$$

donde τ_i es un punto intermedio entre 0 y $\hat{\alpha}^*$. De la última ecuación se despeja $\hat{\alpha}^*$ y resulta

$$J^{1/2}.\hat{\alpha}^* = - \;[(1/J)\; A\;]^{-1}.[(1/J^{1/2}) \sum_j \psi^*(u_j,0,s)] \qquad (55)$$

La ley de los grandes números asegura que con probabilidad 1

$$\lim_{J\to\infty} (1/J) \sum_{j=1}^{J} \partial\psi^*(u_j,0,\sigma_0)/\partial a^* = E(\partial\psi^*(u,0,\sigma_0)/\partial a^*).$$

Como, por la consistencia de los estimadores, $\tau_i \to 0$ y $s \to \sigma_0$ cuando $J\to\infty$, con probabilidad uno, y $\partial\psi^*/\partial a^*$ es continua, parece natural y se deduce del lema 4.2 de Yohai(1987) que también

$$\lim_{J\to\infty} (1/J)\; A(u_1,\dots,u_J) = E(\partial\psi^*(u,0,\sigma_0)/\partial a^*) \qquad (56)$$

con probabilidad 1.

(La hipótesis de dicho lema que pide que exista $\delta>0$ tal que

$$E\left(\sup_{\|a^*\|\leq\delta,\ |t-\sigma_o|\leq\delta} |[\partial\psi^*(u,a^*,t)/\partial a^*]_{ii^*}|\right) < \infty ,$$

se cumple debido a que $[\partial\psi^*(u,a^*,t)/\partial a^*]_{ii^*}$ son acotadas si se mantiene t alejado de cero).

El teorema central del límite asegura que

$$(1/J^{1/2}) \sum_{j=1}^{J} \psi^*(u_j,0,\sigma_o) \overset{D}{\rightarrow} N(0,E(\psi^*(u,0,\sigma_o).\psi^*(u,0,\sigma_o)^t) \quad (57)$$

En el apéndice B se demuestra el lema 9 que afirma que

$$p\lim (1/J^{1/2})\|\sum_j \psi^*(u_j,0,s) - \sum_j \psi^*(u_j,0,\sigma_o)\| = 0.$$

Este resultado y (57) aseguran que

$$(1/J^{1/2}) \sum\psi^*(u_j,0,s) \overset{D}{\rightarrow} N(0,E(\psi^*(u,0,\sigma_o).\psi^*(u,0,\sigma_o)^t). \quad (58)$$

Notemos

$$\mathbf{A_o} = E(\partial\psi^*(u,0,\sigma_o)/\partial \mathbf{a}^*),$$

y

$$V = E(\psi^*(u,0,\sigma_o).\psi^*(u,0,\sigma_o)^t).$$

De (55), (56) y (58) resulta que, si $\mathbf{A_o}$ es no singular

$$J^{1/2}.\hat{\alpha}^* \overset{D}{\rightarrow} N(0,\mathbf{A_o}^{-1}.V.(\mathbf{A_o}^{-1})^t)). \quad (59)$$

Las componentes de la matriz $\mathbf{A_o}$ son las esperanzas de las expresiones (53) y (54), especificando $a=0$, $t=\sigma_o$. Usando quelas v.a. u_i son igualmente distribuidas resulta que la esperanza de (54) es cero, o sea la matriz $\mathbf{A_o}$ es diagonal. El elemento i-ésimo de su diagonal es

$$[\mathbf{A_o}]_{ii} = (1/\sigma_o).a$$

donde a está definido en (50) y es un número estrictamente menor que cero, por lo que $\mathbf{A_o}$ es no singular.

Veamos ahora el valor de las componentes de la matriz V. Llamando, como en la tesis del teorema

$$r_i = r_i(u,0,\sigma_o) = (u_i - \beta(u,0,\sigma_o))/\sigma_o,$$

es

$$V_{ii} = E(\psi^2(r_i)),$$

que hemos llamado v_i en (49) y

$$V_{ii^*} = E(\psi(r_i).\psi(r_{i^*})), \text{ para } i\neq i^*.$$

Obsérvese que, por ser u_i igualmente distribuidas, V_{ii^*} no depende de (i,i^*) para $i\neq i^*$, o sea, son todas iguales, digamos a v^*. Como por la definición de $\beta(u,0,\sigma_o)$ es

$$\Sigma\ \psi(r_i) = 0,$$

igualando las varianzas de ambos miembros de esta resulta

$$I.v_1 + I.(I-1).v^* = 0$$

y luego

$$v^* = -v_1/(I-1).$$

Luego la matriz **V** es

$$\mathbf{V} = v_1.\mathbf{C}\ ,$$

donde **C** es la matriz que aparece en (48).

De (59) y las expresiones halladas para las matrices **A** y **V**, resulta (48), o sea la tesis del teorema 5.1.

El teorema anterior no contempla el caso de la familia de funciones ψ_k de Huber, ya que éstas no son derivables en dos puntos. Por eso se demuestra el teorema siguiente.

Teorema 5.2: Con las mismas hipótesis del teorema 5.1 salvo que se cambian las hipótesis de que ψ tiene derivada ψ'continua y positiva y $\psi'(x).x$ es acotada por las siguientes hipótesis: la función de distribución F_u es continua, ψ tiene derivada ψ' continua salvo a lo sumo en un conjunto finito de puntos que no incluye al cero, ψ' es estrictamente positiva en el conjunto $\{x: |\psi(x)| < k_1\}$ (salvo en un conjunto finito de puntos que no incluye al cero), existe una constante b tal que

$$|(\psi(x+h) - \psi(x))/h| \leq b \quad \forall\ x,\ \forall\ h.$$

y existe una constante c tal que si $\sigma^* > \sigma > 0$ es

$$\psi((x-c(\sigma^*-\sigma))/\sigma^*) \leq \psi(x/\sigma) \leq \psi((x+c(\sigma^*-\sigma))/\sigma^*) \quad \forall x.$$

Entonces la distribución asintótica de $\hat{\alpha}$ está dada por (48). (Obsérvese que las hipótesis sobre ψ se satisfacen para las funciones ψ_k de Huber, en este caso b=1, c=k).

<u>Demostración</u>: Sea

$\psi^*(\mathbf{u},\mathbf{a}^*,t) =$

$$(\psi((u_1-a_1-\beta(\mathbf{u},\mathbf{a},t))/t),\ldots,\psi((u_{I-1}-a_{I-1}-\beta(\mathbf{u},\mathbf{a},t))/t))^t.$$

como en el teorema anterior. Las fórmulas (53) y (54) que dan expresiones para las derivadas de ψ^* con respecto a $\mathbf{a}^*$ siguen siendo válidas siempre que ψ sea derivable en $r_i(\mathbf{u},\mathbf{a}^*,t)$ para $i=1,\ldots,I$ y que $\Sigma\psi'(r_i(\mathbf{u},\mathbf{a}^*,t))\neq 0$. En el caso $\Sigma\psi'(r_i(\mathbf{u},\mathbf{a}^*,t))=0$, es $\psi'(r_i(\mathbf{u},\mathbf{a}^*,t))$ igual a cero para todo

$i=1,\ldots,I$, y de aquí se deduce fácilmente (tomando el cociente incremental y usando el hecho de que, según el lema 1, el incremento de $\beta(\mathbf{u},\mathbf{a}^*,t)$ es menor o igual que el de a_{i*}) que

$$[\partial\psi^*(\mathbf{u},\mathbf{a}^*,t)/\partial\mathbf{a}^*]_{ii*} = 0 \quad \text{si } \Sigma\psi'(r_i(\mathbf{u},\mathbf{a}^*,t))=0.$$

En el teorema anterior se hizo un desarrollo de Taylor. Considérese para cualquier función f: $R^n \to R$ un "pseudodesarrollo" de Taylor del modo siguiente. Llamemos

$$D_i(f)(x,\mathbf{h}) = (f(x_1+h_1,\ldots,x_i+h_i,x_{i+1},\ldots,x_n) - f(x_1+h_1,\ldots,x_{i-1}+h_{i-1},x_i,\ldots,x_n))/h_i$$

para $h_i \neq 0$ y $D_i(f)(x,\mathbf{h})=0$ si $h_i=0$.

Entonces el "pseudodesarrollo" de f alrededor de x es:

$$f(x+\mathbf{h}) = f(x) + \sum_i D_i(f)(x,\mathbf{h}).h_i.$$

Por la definición de $\hat{\alpha}$ se cumple (21.2) o sea

$$\sum_{j=1}^{J} \psi((u_{ij} - \hat{\alpha}_i - \beta(u_j,\hat{\alpha},s))/s) = 0 \text{ , para } i=1,\ldots,I-1. \quad (60)$$

Si se llama $f_i(\hat{\alpha}^*)$ al primer miembro de (60) considerado como función de $\hat{\alpha}^*$ y se hace el pseudodesarrollo de Taylor alrededor de cero se obtienen expresiones análogas a (54.1) y (55) con la única diferencia que el elemento (i,i^*) de la matriz $\mathbf{A}$ es ahora

$$A_{ii*} = D_{i*}(f_i)(0,\hat{\alpha}^*) = \sum_{j=1}^{J} (\psi((u_{ij} - a_i^{(i*)} - \beta(\mathbf{u}_j,\mathbf{a}^{(i*)},s))/s) - \psi((u_{ij} - a_i^{(i*-1)} - \beta(\mathbf{u}_j,\mathbf{a}^{(i*-1)},s))/s)) \;/\; \hat{\alpha}_{i*}, \quad (61)$$

donde se designó

$$\mathbf{a}^{(i*)} = (\hat{\alpha}_1,\ldots,\hat{\alpha}_{i*},0,\ldots,0).$$

Llamemos g_{ii*} al término genérico de (61) o sea:

$$g_{ii*}(\mathbf{u}_j,\hat{\alpha}^*,s) = \psi((u_{ij} - a_i^{(i*)} - \beta(\mathbf{u}_j,\mathbf{a}^{(i*)},s))/s) - \psi((u_{ij} - a_i^{(i*-1)} - \beta(\mathbf{u}_j,\mathbf{a}^{(i*-1)},s))/s)) \;/\; \hat{\alpha}_{i*}.$$

Por hipótesis la función ψ es derivable salvo en un conjunto finito de puntos $\{x_1,\ldots,x_k\}$ pero se puede definir ψ' arbitrariamente en esos puntos para que la función ψ' esté formalmente definida en todo punto (aunque obviamente no es la derivada de ψ en $x_1,\ldots,x_k$) y así las expresiones (53) y (54) están definidas para todo punto, y se las designará formalmente como $[\partial\psi^*(\mathbf{u},\mathbf{a}^*,t)/\partial\mathbf{a}^*]_{ii*}$, aunque en realidad esta derivada no exista si, para algún i, $r_i(\mathbf{u},\mathbf{a}^*,t)$ pertenece al conjunto donde ψ no es derivable. Luego

$$(1/J).A_{ii*} = (1/J).\sum_{j=1}^{J} q_{ii*}(u_j,\hat{\alpha}^*,s) =$$

$$(1/J)\ \Sigma\ [\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*} \quad +$$

$$(1/J)\ \Sigma\ (q_{ii*}(u_j,\hat{\alpha}^*,s) \quad [\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}). \qquad (62)$$

Por la ley de los grandes números

$$\lim_{J\to\infty} (1/J) \sum_{j=1}^{J} [\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*} = E([\partial\psi^*(u,0,\sigma_o)/\partial a^*]_{ii*}). \qquad (63)$$

donde la esperanza no depende de la forma arbitraria en que se definió la derivada en los puntos en los que ψ no era derivable, ya que se demostró en el lema que la continuidad de F_u implica que

$$P(r_i(u,0,\sigma_o) \in \{x_1,\ldots,x_k\}) = 0.$$

Para cada $\delta>0$ se denota $A(\delta)$ al conjunto

$$A(\delta) = \bigcup_{r=1}^{k} (x_r-\delta,x_r+\delta)$$

y $E(\delta)$ al suceso

$$E(\delta) = \bigcap_{i=1}^{I} \{r_i(u,0,\sigma_o) \notin A(\delta)\}.$$

Sea $\varepsilon>0$ arbitrario. Se elige $\delta>0$ tal que

$$P(E(\delta)) \geq 1-\varepsilon.$$

Sea M tal que

$$P(\bigcap_{i=1}^{I} |u_i|\leq M) \geq 1-\varepsilon.$$

Por la continuidad uniforme de r_i en un compacto existe $h_1>0$ tal que

$$E(\delta),\ \bigcap|u_i|\leq M,\ \text{Máx}|a_i|<h_1,\ |t-\sigma_o|<h_1$$

$$\Rightarrow \quad r_i(u,a,t) \quad \notin \quad A(\delta/2) \quad \forall i=1,\ldots,I. \qquad (63.1)$$

Como

$$\lim_{h\to 0} \sup_{\substack{\text{Máx}|a_i|<h \\ |t-\sigma_o|<h}} |\partial\psi^*(u_j,a^*,t)/\partial a^*]_{ii*}-\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\text{Ind}_{E(\delta)}=0$$

existe $h_2>0$ tal que

$$E(\sup_{\substack{\text{Máx}|a_i|<h_2 \\ |t-\sigma_o|<h_2}} |\partial\psi^*(u_j,a^*,t)/\partial a^*]_{ii*}-\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\text{Ind}_{E(\delta)})<\varepsilon.$$

Sea $$h = \min(h_1,h_2,\sigma_o/2).$$

Para $J\geq J_o$ es

$$P(|s-\sigma_o|<h \text{ y } \text{Máx}|\hat{\alpha}_i|<h) \quad 1-\varepsilon.$$

Supóngase que $|s-\sigma_o|<h$ y $\text{Máx}|\hat{\alpha}_i|<h$. Entonces por (63.1)

$\bigcap_{i=1}^{I} |u_{ij}| \leq M$ y $u_j \in E(\delta)$

$$g_{ii*}(u_j,\hat{\alpha}^*,s) = [\partial\psi^*(u_j,\lambda_{ii*j},s)/\partial a^*]_{ii*}. \qquad (63.2)$$

donde λ_{ii*j} es punto intermedio entre $\hat{\alpha}^*$ y 0.

Queremos demostrar ahora que el último término de (62) tiende a cero en probabilidad.

Notemos B al suceso

$$B = (\bigcap |u_i| \leq M) \cap E(\delta).$$

Entonces

$$(1/J)\ \Sigma\ |g_{ii*}(u_j,\hat{\alpha}^*,s) - [\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}| =$$

$$(1/J)\ \Sigma\ |g_{ii*}(u_j,\hat{\alpha}^*,s)-[\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\mathrm{Ind}_{u_j\in B} \ + \qquad (64)$$

$$(1/J)\ \Sigma\ |g_{ii*}(u_j,\hat{\alpha}^*,s)-[\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\mathrm{Ind}_{u_j\in B^c}\ .$$

Como por hipótesis es

$$|(\psi(x+d) - \psi(x))/d| \leq b,$$

resulta

$$|g_{ii*}(u_j,\hat{\alpha}^*,s) \leq 2b/(\sigma_o-h)$$

y en consecuencia

$$[\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*} \leq 2b/(\sigma_o-h).$$

Luego el último término de (64) es menor o igual que

$$4b/(\sigma_o-h).(1/J).\sum_{j=1}^{J} \mathrm{Ind}_{u_j\in B^c}.$$

Como la esperanza del indicador de B^c es menor que 2ε resulta que para $J \geq J_1$ es

$$P(|\text{último término de }(64)| \leq 9b\varepsilon/(\sigma_o-h) \quad 1-\varepsilon. \qquad (65)$$

Se acotará ahora el primer término del segundo miembro de (64). Por (63.2) es

$$(1/J)\ \Sigma\ |g_{ii*}(u_j,\hat{\alpha}^*,s)-[\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\mathrm{Ind}_{u_j\in B} =$$

$$(1/J)\Sigma|[\partial\psi^*(u_j,\lambda_{ii*j},s)/\partial a^*]_{ii*} - [\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\mathrm{Ind}_{u_j\in B}$$

$$\leq \Sigma \sup_{\substack{\mathrm{Máx}|\lambda_i|<h \\ |t-\sigma_o|<h}} |\partial\psi^*(u_j,a^*,t)/\partial a^*]_{ii*}-\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}|.\mathrm{Ind}_{u_j\in B}.$$

Por la elección de h_2 resulta que si $J \geq J_2$

$$P(|\text{1er. término del 2do miembro de }(64)|) \quad 2\varepsilon) \quad 1-\varepsilon. \qquad (66)$$

De (64), (65) y (66) resulta que si $J \geq \mathrm{Max}(J_o,J_1,J_2)$ entonces

$$P((1/J)\Sigma|g_{ii*}(u_j,\hat{\alpha}^*,s)-[\partial\psi^*(u_j,0,\sigma_o)/\partial a^*]_{ii*}| < 2\varepsilon+9b\varepsilon/(\sigma_o-h))$$

$1-2\varepsilon$.

Como ε es un numero positivo arbitrario, esto implica que

$$p\lim\ (1/J)\Sigma|g_{ii*}(u_j,\hat{\alpha}^*,s)-[\partial\psi^*(u_j,0,\sigma_o)/\partial \mathbf{a}^*]_{ii*}| \;=0$$

y por (62) y (63) esto implica que los elementos de la matriz **A** tienden en probabilidad al mismo límite que bajo las hipótesis del teorema 5.1 y por lo tanto $\hat{\alpha}^*$ tiene la misma distribución que se demostró en el teorema 5.1.

6. EFICIENCIA ASINTOTICA DE LOS M-ESTIMADORES CON RESPECTO A LOS ESTIMADORES DE CUADRADOS MINIMOS. Los estimadores de cuadrados mínimos (ECM) son un caso particular de M-estimadores considerando $\psi(x)=x$. En este caso particular, el factor $V(\psi,F_u)$ de la varianza asintótica (ver tesis del teorema 5) es

$$V(\psi,F_u)\ =(Var(u)/\sigma_o^2).((I-1)/I).$$

Se desea comparar aquí esta varianza asintótica con la de los M-estimadores obtenidos usando la función ψ_k de Huber, ya mencionada en la sección 3. Debido a la dificultad para calcular las expresiones (49) y (50) se han calculado algunos valores aproximados de $V(\psi_k,F_u)$ por simulación. Se eligió k=1,345 que es el valor que hace que la eficiencia relativa del estimador de cuadrados mínimos con respecto al M-estimador sea 1,05 para el problema de posición, cuando la distribución es normal. Para el modelo del diseño en bloques los resultados obtenidos cuando F_u es la distribucion $N(0,\sigma_o^2)$ son los siguientes:

I	$V(\psi_k,F_u)$	V para ECM	Efic. asint. relativa
3	$0,685\leq V\leq 0,689$	2/3	$1,027\leq EAR\leq 1,033$
4	$0,777\leq V\leq 0,782$	3/4	$1,036\leq EAR\leq 1,043$
8	$0,909\leq V\leq 0,917$	7/8	$1,039\leq EAR\leq 1,048$

Para cada I, el intervalo calculado para $V(\psi_k,F_u)$ es un intervalo de confianza con nivel asintótico 95%. Se puede observar que en todos los casos la eficiencia asintótica del estimador de cuadrados mínimos es apenas superior a la del M-estimador, siendo la eficiencia relativa algo menor que para el caso de posición (se acerca al valor 1,05 cuando I crece).

La ventaja de emplear M-estimadores se aprecia al considerar distribuciones F_u con colas más pesadas que la normal. Por ejemplo para la distribución N(0,1) con un 10% de contaminación con errores N(0,9) se obtuvo, también por simulación, que el varianza asintótica $\sigma_0^2.V(\psi_k,F_u)$ está comprendida entre 0,934 y 0,957 (con un nivel de confianza del 95%) para I=3 tratamientos. Como la varianza asintótica es 1,2 para el estimador de cuadrados mínimos, resulta que la eficiencia asintótica relativa está comprendida entre 0.78 y 0.80. Para la distribución de Cauchy, la varianza asintótica del M-estimador está comprendida entre 5.22 y 5.38 (también con un nivel del 95%). Para esta distribución la varianza asintótica del estimador de cuadrados mínimos es infinita, por lo que la eficiencia relativa es cero. O sea que, al igual que ocurre para posición, los M-estimadores son un poco menos eficientes que los estimadores de cuadrados mínimos cuando los errores aleatorios u_{ij} tienen distribución normal pero pueden llegar a ser mucho más eficientes cuando hay una pequeña proporción de valores atípicos.

7. **PUNTO DE RUPTURA.** Consideraremos primero el caso más simple, aunque de poca utilidad práctica, en que se conoce un parámetro de dispersión. Demostraremos que para este caso los M-estimadores de los efectos de los tratamientos permanecen acotados aunque haya observaciones muy alejadas, . siempre que estas observaciones atípicas se hayan producido en menos de la mitad de los bloques. Luego (en 7.2) se considera el caso más útil en que para el calculo de los M-estimadores se emplea un estimador de dispersión. En este caso se verá que el punto de ruptura de los estimadores de los efectos de los tratamientos, depende esencialmente del punto de ruptura del estimador de dispersión. Luego se considera el caso particular en que $\hat{\alpha}_i$ y s se calculan simultáneamente, en este caso el punto de ruptura de los estimadores de los efectos de los tratamientos coincide con el del estimador de dispersión y están dados por la expresión (71.2). Finalmente en 7.4, se propone un método para obtener estimadores con alta eficiencia y alto punto de ruptura.

7.1. **DISPERSION CONOCIDA.** Se nota σ el parámetro de dispersión que se supone conocido. Los M-estimadores se obtienen resolviendo el sistema de ecuaciones (21.2) (21.3), reemplazando en estas ecuaciones s por σ, o con otra notación resolviendo el sistema

$$E_F(\psi((X_i - \hat{\alpha}_i(F) - B(\mathbf{X},\hat{\alpha}(F),\sigma))/\sigma)) = 0, \text{ para } i=1,\ldots,I \quad (67.1)$$

$$\Sigma\, \hat{\alpha}_i(F) = 0 \quad (67.2)$$

para $F=F_J$ la distribución I-variada empírica que asigna probabilidad $1/J$ al vector $x_j=(X_{1j},\ldots,X_{Ij})$ observado en el j-ésimo bloque. Se asume que ψ cumple la suposición B.

Para estudiar el punto de ruptura de los estimadores $\hat{\alpha}_i$ consideraremos modelos de contaminación, o sea supondremos que una proporción $(1-\varepsilon)$ de los bloques tienen una distribución fija F_0, llamada distribución central, mientras que una proporción ε tienen una distribución cualquiera, desconocida. Esto es lo mismo que decir que la distribución del vector I-dimensional $\mathbf{X}$ pertenece a la familia de distribuciones

$$D\varepsilon = \{F:\ F=(1-\varepsilon).F_0+\varepsilon.H,\ H \text{ cualquier distribución}\}$$

con $0\leq\varepsilon\leq1$.

Se define punto de ruptura del estimador $\hat{\alpha}$ así:

$$\varepsilon_0 = \sup\{\varepsilon:\ \sup_{F\in D\varepsilon}\ \max_{1\leq i\leq I} |\hat{\alpha}_i(F)|<+\infty\}.$$

Se desea ver que el punto de ruptura es 0,5. Por una parte es evidente que si $\varepsilon>0,5$ los estimadores no están acotados. Por ejemplo si $\varepsilon>0,5$ y se considera una sucesión de distribuciones $F_n=(1-\varepsilon).F_0+\varepsilon.H_n$, donde H_n es tal que la primera coordenada del vector aleatorio $(X_1,\ldots,X_I)$ toma con probabilidad uno un valor $x_n\to+\infty$ mientras que las otras coordenas se mantienen acotadas, será

$$\lim\ \max_{1\leq i\leq I} |\hat{\alpha}_i(F_n)| = +\infty.$$

Luego el punto de ruptura es a lo sumo 0.5.

Demostraremos ahora que $\varepsilon_0\geq0.5$. Para ello consideremos un $\varepsilon>\varepsilon_0$. Sea H_n una sucesión de distribuciones tal que

$$\max_{1\leq i\leq I} |\hat{\alpha}_i(F_n)| \to +\infty$$

(donde $F_n=(1-\varepsilon)F_0+\varepsilon H_n$). Pasando en caso necesario a una subsucesión, se puede suponer que existe un i^* fijo tal que o bien

$$\max |\hat{\alpha}_i(F_n)| = \hat{\alpha}_{i^*}(F_n)\to+\infty$$

o bien

$$\max |\hat{\alpha}_i(F_n)| = -\hat{\alpha}_{i^*}(F_n)\to+\infty.$$

Consideremos por ejemplo el primer caso. De (7.1) resulta

$$(1-\varepsilon).E_{F_0}(\psi((X_{i*}-\hat{\alpha}_{i*}(F_n)-\beta(X,\hat{\alpha}(F_n),\sigma))/\sigma)) + \varepsilon.k_i \geq 0. \quad (68)$$

Como para "n" grande es

$$\underset{1\leq i\leq I}{\text{Máx}}\ X_i - \underset{1\leq i\leq I}{\min}\ X_i \quad \hat{\alpha}_{i*}(F_n)/(2.(I+1)),$$

aplicando el lema 2 resulta

$$X_{i*}-\hat{\alpha}_{i*}(F_n)-\beta(X,\hat{\alpha}(F_n),\sigma) \leq X_{i*}-\hat{\alpha}_{i*}(F_n)-(\beta(X,0,\sigma)-c_I.\hat{\alpha}_{i*}(F_n)) =$$

$$= X_{i*} - \beta(X,0,\sigma) - \hat{\alpha}_{i*}(F_n).(1-c_I) \to -\infty .$$

Luego, tomando límite para $n\to\infty$ en (68) resulta

$$(1-\varepsilon).(-k_i) + \varepsilon.k_i \geq 0,$$

de donde resulta $\varepsilon\geq 0,5$. Queda así demostrado que el punto de ruptura es 0,5.

7.2 PARA ESTIMAR LOS EFECTOS DE LOS TRATAMIENTOS SE ESTIMA LA DISPERSION. Se considera ahora el caso más realista en el que se resuelven las ecuaciones (21.2) y (21.3) empleando un estimador de dispersión s, o lo que es lo mismo resolviendo el sistema (67.1), (67.2), donde se reemplaza σ por un estimador de dispersión s(F). (El estimador de dispersión puede ser calculado previamente o en forma simultánea).

Sea ε_1 el punto de ruptura a infinito del estimador de dispersión o sea

$$\varepsilon_1 = \sup \{\varepsilon: \sup_{F\in D\varepsilon} s(F)<\infty\}.$$

Considérese un ε tq. $\varepsilon_0<\varepsilon<\varepsilon_1$. Como para ese ε el estimador de dispersión se mantiene acotado, notemos

$$B = \sup_{F\in D\varepsilon} s(F).$$

Con una demostración similar a la hecha en la sección 7.1 se ve que $\varepsilon\geq 0,5$. De esto se deduce que

$$\text{si} \quad \varepsilon_1\geq 0,5 \quad \Rightarrow \quad \varepsilon_0\geq 0,5 \quad (69.1)$$

$$\text{si} \quad \varepsilon_1\leq 0,5 \quad \Rightarrow \quad \varepsilon_0\geq \varepsilon_1. \quad (69.2)$$

Claro que evidentemente cualquier estimador razonable tiene punto de ruptura a lo sumo 0.5, por lo que (69.1) no tiene interés.

7.3 ESTIMACION SIMULTANEA DE LOS ESTIMADORES DE LOS EFECTOS DE LOS TRATAMIENTOS Y DE LA DISPERSION. Estos estimadores, que se obtienen resolviendo simultáneamente (21.2) a (21.4), pueden escribirse también como la solución del sistema

$$E_F(\psi((X_i - \hat{\alpha}_i(F) - \beta(X,\hat{\alpha}(F),s(F)))/s(F))) = 0,\ i=1,\dots,I,(70.1)$$

$$\sum_i \hat{\alpha}_i(F) =$$

$$E_F(1/(I-1))\sum_i \chi((X_i-\hat{\alpha}_i(F)-B(X,\hat{\alpha}(F),s(F)))/s(F))) = c, \qquad (70.3)$$

considerando $F=F_J$ la función de distribución empírica.

En esta sección se considera sólo el caso particular en que

$$\chi(x) = \psi^2(x),$$

que es la propuesta 2 de Huber (1964) y se supone que ψ cumple B. Demostraremos que el punto de ruptura del estimador de dispersión es

$$\varepsilon_1 \geq c/(d_I.k_1^2+c) \qquad (71.1)$$

donde

$$d_I = 1 \qquad \text{si } I \text{ es impar}$$

$$d_I = I/(I-1) \quad \text{si } I \text{ es par.}$$

Demostraremos también que si ψ es estrictamente creciente en $\{x: |\psi(x)|<k_1\}$ y la distribución central F_0 es no degenerada entonces se verifica

$$\varepsilon_0 = \varepsilon_1 = c/(d_I.k_1^2+c). \qquad (71.2)$$

Vamos a demostrar primero la desigualdad (71.1). Sea $\varepsilon>\varepsilon_1$, entonces existe una sucesión de distribuciones H_n, de modo que si $F_n=(1-\varepsilon).F_0+\varepsilon.H_n$ es

$$s_n = s(F_n) \to \infty.$$

Se puede suponer sin perder generalidad (pasando de ser necesario a una subsucesión) que

$$\lim_{n\to\infty} (\hat{\alpha}_i(F_n)/s_n) = a_i, \; i=1,...,I,$$

donde $-\infty\leq a_i\leq+\infty$. El sistema de ecuaciones (70.1) a (70.3) puede escribirse

$$(1-\varepsilon).E_{F_0}(\psi((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n))$$
$$+ \varepsilon.E_{H_n}(\psi((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n)) = 0, \; i=1,...,I, \qquad (72.1)$$

$$\sum \hat{\alpha}_i(F_n) = 0 \qquad (72.2)$$

$$(1-\varepsilon).E_{F_0}((1/(I-1))\sum\psi^2((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n))$$
$$+\varepsilon.E_{H_n}((1/(I-1))\sum\psi^2((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n)) = c. \qquad (72.3)$$

Por ser $\psi\leq k_1$ y por el lema 10, de (72.1) y (72.3) resulta

$$|(1-\varepsilon).E_{F_0}(\psi((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n))| \leq \varepsilon.k_1 \quad \forall i, \qquad (73)$$

$$(1-\varepsilon).E_{F_0}((1/(I-1))\sum\psi^2((X_i-\hat{\alpha}_i(F_n)-B(X,\hat{\alpha}(F_n),s_n))/s_n))$$
$$+ \varepsilon.d_I.k_1^2 \geq c \qquad (74)$$

La ecuación que define a $\beta(X,\hat{\alpha}(F_n),s_n)$, o sea

$$\sum_{i=1}^{I} \psi((X_i-\hat{\alpha}_i(F_n)-\beta(X,\hat{\alpha}(F_n),s_n))/s_n) = 0,$$

es para n grande aproximadamente igual a

$$\sum_{i=1}^{I} \psi((-a_i - \beta(X,\hat{\alpha}(F_n),s_n))/s_n) = 0. \qquad (75)$$

Consideremos el caso en que los a_i son todos finitos. Entonces está definido $\beta(0,a,1)$ que cumple

$$\sum_{i=1}^{I} \psi(-a_i-\beta(0,a,1))= 0. \qquad (76)$$

De (75) y (76) resulta que

$$\lim_{n\to\infty} (\beta(X,\hat{\alpha}(F_n),s_n)/s_n) = \beta(0,a.1).$$

Tomando límite para $n\to\infty$ en (73) y (74) resulta

$$|\psi(-a_i-\beta(0,a,1))| \leq \varepsilon.k_i/(1-\varepsilon) \qquad (77)$$

$$(1-\varepsilon).(1/(I-1))\Sigma\psi^2(-a_i-\beta(0,a,1)) + \varepsilon.d_I.k_i^2 \geq c \qquad (78)$$

De (77) y el lema 10 surge que

$$(1/(I-1))\Sigma\psi^2(-a_i-\beta(0,a,1)) \leq d_I.\varepsilon^2.k_i^2/(1-\varepsilon)^2. \qquad (79)$$

De (78) y (79) resulta

$$(c-\varepsilon.d_I.k_i^2)/(1-\varepsilon) \leq d_I.\varepsilon^2.k_i^2/(1-\varepsilon)^2$$

y de aquí resulta

$$\varepsilon \geq c/(d_I.k_i^2+c). \qquad (80)$$

Como ε es cualquier número mayor que el punto de ruptura ε_i de (80) resulta (71.1).

Falta considerar el caso en que algún a_i fuese infinito. Si $\varepsilon\geq 0,5$ (80) se cumpliría trivialmente, así que supongamos que $\varepsilon<0,5$ y vamos a demostrar que el caso a_i infinito no puede ocurrir. Supongamos, pasando en caso necesario a una subsucesión, que $a_{i*}=+\infty$ y que Máx $|\hat{\alpha}_i(F_n)|=\hat{\alpha}_{i*}(F_n)$. Usando el lema 2 resulta

$$(X_{i*}-\hat{\alpha}_{i*}(F_n)-\beta(X,\hat{\alpha}(F_n),s_n))/s_n \leq$$
$$(X_{i*}-\beta(X,0,s_n))-(1-c_I)\hat{\alpha}_{i*}(F_n))/s_n \to -\infty.$$

Tomando límite en (73) resultaría

$$(1-\varepsilon).k_i \leq \varepsilon.k_i,$$

lo que se contradice con el hecho de que $\varepsilon<0.5$. Luego el caso a_i infinito no puede ocurrir y está demostrado (71.1).

Se desea ahora demostrar que en (71.1) no puede verificarse el "mayor" sino que se cumple el "igual". Para ello consideremos

$$\varepsilon = c/(d_I.k_1^2+c)$$

y veamos que para ese valor de ε es

$$\sup_{F\in D\varepsilon} s(F) = +\infty. \qquad (82)$$

Supongamos que no fuese cierto (82). Consideremos el caso I par (en ese caso $d_I=I/(I-1)$). Sea H_n la distribución que asigna peso uno al punto $x^{(n)}\in R^I$ cuyas primeras $I/2$ coordenadas valen $-x_n$ y cuyas $I/2$ coordenadas restantes valen x_n, con $x_n\to+\infty$. Sea $F_n=(1-\varepsilon).F_0+\varepsilon.H_n$. Sea $s_n=s(F_n)$ que hemos supuesto que permanece acotado. También $\hat{\alpha}(F_n)$ permanece acotado por ser $\varepsilon_0\geq\min(\varepsilon_1,0.5)$ (sección 7.2). Podemos suponer sin perder generalidad, considerando en caso necesario una subsucesión conveniente, que

$$\lim_{n\to\infty} s_n = s\ , \qquad \lim_{n\to\infty} \hat{\alpha}_i(F_n) = a_i,\ i=1,\dots,I.$$

Por la definición de los estimadores se cumple (70.1) para cada F_n, o sea que se verifica

$$(1-\varepsilon).E_{F_0}(\psi((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n))$$
$$+\ \varepsilon.E_{H_n}(\psi((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n)) = 0,\ i=1,\dots,I. \qquad (83)$$

Como $\beta(x^{(n)},0,s_n)=0$ entonces $\beta(x^{(n)},\hat{\alpha}(F_n),s_n)$ permanece acotado, luego

$$\lim_{n\to\infty} E_{H_n}(\psi((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n)) = -k_1,\ \text{para } i\leq I/2,$$
$$= k_1,\ \text{para } i>I/2.$$

Tomando límite en (83) resulta

$$E_{F_0}(\psi((X_i-a_i-\beta(\mathbf{X},a,s))/s)) = \varepsilon k_1/(1-\varepsilon) = c(I-1)/(Ik_1),\ \text{si } i\leq I/2$$
$$= -c(I-1)/(Ik_1),\ \text{si } i>I/2. \qquad (84)$$

Por (70.3) es

$$(1-\varepsilon)E_{F_0}((1/(I-1))\Sigma\psi^2((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n))$$
$$+\varepsilon E_{H_n}((1/(I-1))\Sigma\psi^2((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n)) = c. \qquad (85)$$

Como

$$\lim_{n\to\infty} E_{H_n}((1/(I-1))\Sigma\psi^2((X_i-\hat{\alpha}_i(F_n)-\beta(\mathbf{X},\hat{\alpha}(F_n),s_n))/s_n))= Ik_1^2/(I-1)$$

tomando límite en (85) resulta

$$E_{F_0}((1/(I-1))\Sigma\psi^2((X_i-a_i-\beta(\mathbf{X},\mathbf{a},s))/s)) = (c-\varepsilon Ik_1^2/(I-1))/(1-\varepsilon) =$$
$$= (I-1)c^2/(Ik_1^2) \qquad (86)$$

De (84) y (86) resulta que

$$\sum_{i=1}^{I} Var_{F_0}(\psi((X_i-a_i-\beta(\mathbf{X},a,s))/s)) = 0. \qquad (87)$$

De (84) y (87)

$P_{F_0}(\Psi((X_i-a_i-\beta(X,a,s))/s) = c(I-1)/(Ik_1)) = 1$, si $i\leq I/2$,

$P_{F_0}(\Psi((X_i-a_i-\beta(X,a,s))/s) =-c(I-1)/(Ik_1)) = 1$, si $i>I/2$.

Por ser Ψ estrictamente creciente en $\{x: |\Psi(x)|<k_1\}$ es, bajo F_0,

$$(X_i-a_i-\beta(X,a,s))/s = \text{constante, con probabilidad uno, } \forall i,$$

de donde por ejemplo resulta que

$$(X_1-a_1) - (X_I-a_I) = \text{constante, con probabilidad uno,}$$

lo que contradice la suposición de que la distribución central F_0 no es degenerada. Esta contradicción provino de suponer que no se cumplía (82), luego (82) es verdadero. El caso I impar se trata en forma análoga y de este modo está demostrado que

$$\varepsilon_1 = c/(d_I.k_1^2+c). \qquad (88)$$

Sabemos por la sección 7.2 que $\varepsilon_0\geq\min(\varepsilon_1,0.5)$, luego por (88) es

$$\varepsilon_0 \geq c/(d_I.k_1^2+c) \qquad (89)$$

Queremos demostrar que en (89) se cumple la igualdad. Razonando por el absurdo supongamos que se cumpliese la desigualdad estricta. Sea

$$\varepsilon = c/(d_I.k_1^2+c).$$

Consideremos el caso I par y sean H_n y F_n como en la demostración anterior. Ya sabemos por la demostración anterior que $s_n\to+\infty$ y como estamos suponiendo que se cumple la desigualdad estricta en (89), $\hat{\alpha}(F_n)$ se mantendría acotado y también $\beta(x^{(n)},\hat{\alpha}(F_n),s_n)$ sería acotado. Luego tanto el primer término de (83) como el primer término de (85) tenderían a cero al crecer "n" y tomando límite en (83) resultaría

$$\lim_{n\to\infty} \Psi(x_n/s_n) = 0 \quad \text{y luego} \quad \lim_{n\to\infty} x_n/s_n = 0.$$

Pero entonces el segundo término de (85) también tendería a cero y (85) no podría verificarse. Este absurdo provino de suponer la desigualdad estricta en (89), luego queda demostrado (71.2) que era el objetivo de esta sección.

7.4. CALCULO DE UN ESTIMADOR DE DISPERSION CON ALTO PUNTO DE RUPTURA. Vimos en la sección anterior que el punto de ruptura de los estimadores obtenidos resolviendo simultáneamente (21.2) a (21.4) con $\chi=\Psi^2$ está dado por

$$\varepsilon_0 = \varepsilon_1 = c/(d_I.k_1^2+c) \qquad (90)$$

Veremos ahora que el último miembro de (90) se puede hacer tan cercano a 0,5 como se desee. Para ello bastará considerar la familia de funciones ψ_k de Huber o sea

$$\psi_k(x) = sg(x).\min(|x|,k)$$

y tomar k suficientemente pequeño.

La constante c se la elige como se indicó en (22), o sea que para la función ψ_k de Huber dicha constante vale

$$c(\psi_k) = E((1/(I-1)).\Sigma\ \psi_k^2(Z_i-\beta(Z,0,1)))$$

para $Z_1,\ldots,Z_I$ independientes $N(0,1)$. Sean $Z_{(1)}\leq Z_{(2)}\leq\ldots\leq Z_{(I)}$, los valores $Z_1,\ldots,Z_I$ ordenados. Para el caso I impar y para la función ψ_k de Huber será

$$Z_{((I+1)/2)} - k \leq \beta(Z,0,1) \leq Z_{((I+1)/2)} + k.$$

Sea k_n una sucesión de números positivos que tienden a cero. Se nota

$$c_n = c(\psi_{k_n})$$

y $\beta_n(Z,0,1)$ al $\beta(Z,0,1)$ correspondiente a ψ_{k_n}.

Sea $\nu>0$ arbitrario. Como

$$P(Z_{((I+1)/2)} - Z_{((I-1)/2)} > 0) = 1$$

es

$$\lim_{n\to\infty} P(Z_{((I+1)/2)} - Z_{((I-1)/2)} \quad 2k_n) = 1.$$

Luego si $n \geq n_0(\nu)$ es

$$P(Z_{((I+1)/2)} - Z_{((I-1)/2)} \quad 2k_n) \quad 1-\nu/2$$

y análogamente

$$P(Z_{((I+3)/2)} - Z_{((I+1)/2)} \quad 2k_n) \quad 1-\nu/2.$$

Si $Z_{((I+3)/2)}-Z_{((I+1)/2)} \quad 2k_n$ resulta

$$Z_{((I+3)/2)} - \beta_n(Z,0,1) =$$

$$(Z_{((I+3)/2)} - Z_{((I+1)/2)}) + (Z_{((I+1)/2)} - \beta_n(Z,0,1))$$

$$2k_n - k_n = k_n.$$

En forma análoga, si $Z_{((I+1)/2)}-Z_{((I-1)/2)} \quad 2k_n$ resulta

$$Z_{((I-1)/2)} - \beta_n(Z,0,1) \quad - k_n.$$

Luego, si $n \geq n_0(\nu)$ es

$$c_n \geq E((1/(I-1)).(\sum_{i=1}^{(I-1)/2}\psi^2_{k_n}(Z_{(i)}-\beta_n(Z,0,1)) + \sum_{(I+3)/2}^{I}\psi^2_{k_n}(Z_{(i)}-\beta_n(Z,0,1)).$$

$$\mathrm{Ind}_{Z_{((I+3)/2)}-Z_{((I+1)/2)}>2k_n \text{ y } Z_{((I+1)/2)}-Z_{((I-1)/2)}>2k_n}$$

$$\geq k_n^2.(1-\nu).$$

Luego para $n\geq n_0(\nu)$ el punto de ruptura es

$$c_n/(d_I.k_n^2+c_n) \geq k_n^2(1-\nu)/(k_n^2+k_n^2(1-\nu)) = (1-\nu)/(2-\nu). \quad (91)$$

Como eligiendo ν cercano a cero el segundo miembro de (91) puede hacerse tan próximo a 1/2 como se desee, se ha demostrado que usando la funcion ψ_k con cercano a cero puede hallarse un estimador con punto de ruptura tan cerca de 1/2 como se desee.

Para el caso I par, considérese nuevamente $k_n \to 0$, $v>0$ y $n_o(v)$ tq.

$$n \geq n_o(v) \quad \Rightarrow \quad P(Z_{(I/2+1)} - Z_{(I/2)} \qquad\qquad 1-v.$$

Si $Z_{(I/2+1)} - Z_{(I/2)}$ $2.k_n$, el conjunto de soluciones de la ecuación

$$\Sigma\ \psi(Z_i - \beta) = 0$$

es el intervalo

$$[Z_{(I/2)}+k_n;\ Z_{(I/2+1)}-k_n],$$

luego su punto medio es

$$\beta_n(Z,0,1) = (Z_{(I/2)} + Z_{(I/2+1)})/2,$$

o sea la mediana de los Z_i.

Luego

$$c_n \geq E((1/(I-1)).\Sigma\ \psi^2_{\varepsilon_n}(Z_i - \beta(Z,0,1)).\mathrm{Ind}_{Z_{(I/2+1)}-Z_{(I/2)}\geq 2.k_n}$$

$$\geq (I/(I-1)).k_n^2.(1-v) = d_I.k_n^2(1-v).$$

Luego el punto de ruptura es

$c_n/(d_I.k_n^2+c_n) \geq d_I.k_n^2(1-v)/(d_I.k_n^2+d_I.k_n^2(1-v)) = (1-v)/(2-v)$

y la demostración sigue igual al caso I impar.

8. ESTIMACION DE LA MATRIZ DE COVARIANZA ASINTOTICA. INTERVALOS DE CONFIANZA Y TEST DE HIPOTESIS.

8.1 ESTIMACION DE LA MATRIZ DE COVARIANZA ASINTOTICA. Se demostró en la sección 5 que

$$J^{1/2}.(\hat{\alpha}^* - \alpha^*) \xrightarrow{D} N(0,\ \sigma_o^2.V(\psi,F_u).C),$$

donde se usa la misma notación que en la expresión (48). La matriz C es conocida y s es un estimador consistente de σ_o, así que para estimar la matriz de covarianza asintótica, sólo falta estimar $V(\psi,F_u)$. Recordando que

$$V(\psi,F_u) = v_1/a^2\ . \tag{92}$$

donde

$$v_1 = E(\psi^2(r_1)), \tag{93.1}$$

$$a = E(\psi'(r_1).(-1 + (\psi'(r_1) - \psi'(r_I)) \ / \sum_{l=1}^{I} \psi'(r_l))) \tag{93.2}$$

y

$$r_i = (u_i - \beta(u,0,\sigma_o))/\sigma_o, \tag{93.3}$$

el estimador de $V(\psi,F_u)$ se puede calcular así:

$$\widehat{V(\psi,F_u)} = \hat{v}_1/\hat{a}^2 \tag{94.1}$$

$$\hat{\gamma}_i = \sum_{i=1}^{I} \sum_{j=1}^{J} \psi^2(\hat{r}_{ij})/(I.(J-1)) \qquad (94.2)$$

y

$$\hat{a} = \sum_{j=1}^{J} \left(\sum_{i=1}^{I-1} \psi'(\hat{r}_{ij}).(-1 + (\psi'(\hat{r}_{ij})-\psi'(\hat{r}_{Ij}))/\sum_{i=1}^{I} \psi'(\hat{r}_{ij})) \right)/((I-1).J)$$

(94.3)

donde $\hat{r}_{ij}$ son los "residuos standarizados"

$$\hat{r}_{ij} = (X_{ij} - \hat{\alpha}_i - \hat{\beta}_j)/s. \qquad (94.4)$$

NOTA: Para ver intuitivamente por qué $\hat{r}_{ij}$ estima a

$$r_{ij} = (u_{ij} - \beta(u_j,0,\sigma_0))/\sigma_0,$$

obsérvese que como $\hat{\beta}_j$ cumple

$$\sum_{i=1}^{I} \psi((X_{ij}-\hat{\alpha}_i-\hat{\beta}_j)/s) = 0$$

que reemplazando X_{ij} por lo que vale según el modelo es

$$\sum_{i=1}^{I} \psi((u_{ij}+\alpha_i-\hat{\alpha}_i+\beta_j-\hat{\beta}_j)/s) = 0. \qquad (95)$$

De (95) resulta que

$$\hat{\beta}_j = \beta(u_j,\hat{\alpha}-\alpha,s) + \beta_j.$$

Luego

$$\hat{r}_{ij} = (u_{ij} + \alpha_i - \hat{\alpha}_i - \beta(u_j,\hat{\alpha}-\alpha,s))/s$$

estima a r_{ij}.

Comentario: Para el modelo lineal general, la varianza asintótica de los estimadores de los parámetros es (por ejemplo Huber (1981), sección 7.6):

$$\sigma_0^2 . \; v(\psi,F_u). \; (X^t.X)^{-1}$$

donde X es la matriz de diseño y

$$v(\psi,F_u) = E(\psi^2(u_1/\sigma_0))/E^2(\psi'(u_1/\sigma_0)).$$

Para el diseño en bloques, la matriz $(X^t.X)^{-1}$ es de $(I-1+J)*(I-1+J)$, las filas y columnas de esta matriz que corresponden a la matriz de covarianza de $\hat{\alpha}^*$ valen

$$(1/J).((I-1)/I).C.$$

La propuesta de Huber para estimar la matriz de covarianza asintótica, aplicada al caso particular del diseño en bloques, daría el siguiente estimador de la matriz de covarianza asintótica de $\hat{\alpha}^*$:

$$\frac{s^2 . \sum_{i=1}^{I} \sum_{j=1}^{J} \psi^2(\hat{r}_{ij})/(I-1).(J-1)}{\left[\sum_{i=1}^{I} \sum_{j=1}^{J} \psi'(\hat{r}_{ij})/I.J \right]^2} . (1/J).((I-1)/I).C. \qquad (96)$$

Si se compara este estimador de la matriz de covarianza, con el propuesto en esta sección se observa que la diferencia es que en el estimador de Huber, en el denominador aparece la expresión

$$[\sum_{i=1}^{I} \sum_{j=1}^{J} \psi'(\hat{r}_{ij})/I.J\ ]^2 \tag{97}$$

mientras que en el estimador propuesto en este trabajo aparece la expresión $\hat{a}^2$ (con $\hat{a}$ de (94.3)).

Dicho de otro modo, si se aplica un programa de regresión múltiple general a los datos de un diseño en bloques, el estimador de la matriz de covarianza asintótica que el programa calcularía no sería correcto.

Cuando el número de bloques J tiende a infinito $\hat{a}^2$ converge a a^2 (de la expresión (93.2)) , mientras que la expresión (97) tiende a

$$E^2(\psi'(r_1)). \tag{98}$$

La expresión (93.2) es negativa pues

$$\psi'(r_1) - \psi'(r_I)) \ /\ \Sigma\ \psi'(r_i)$$

toma valores entre -1 y 1. Además, si designamos

$$X = \psi'(r_1)\ /\ (\Sigma\ \psi'(r_i))^{(1/2)}, \quad Y = \psi'(r_I)\ /\ (\Sigma\ \psi'(r_i))^{(1/2)},$$

es

$$a = -\ E(\psi'(r_1) + (E(X^2).E(Y^2))^{(1/2)} - E(X.Y),$$

y por la desigualdad de Cauchy-Schwarz es

$$-\ E(\psi'(r_1) \quad a\ . \tag{99}$$

Como a es negativo, de (99) resulta que la expresión (98) es mayor que a^2, y por consiguiente (96) tiende a un valor menor que la verdadera matriz de covarianza asintótica.

Para tener idea de la magnitud del error que se comete al estimar la matriz de covarianza asintótica con la expresión (96) se calculó (por simulación con 50.000 casos para cada distribución) el valor de a y el valor de (98). Se obtuvo el siguiente resultado: el valor límite al que tiende el estimador basado en la teoría asintótica de Huber subestima a la verdadera matriz de covarianza asintótica en un 10% para la distribución normal, en un 12% para una distribución normal contaminada $(0.9.N(0,\sigma^2)+0.1.N(0,9\sigma^2))$ y en un 19% para la distribución de Cauchy.

8.2. TEST DE HIPOTESIS. Supongamos que se desea estudiar la hipótesis nula

$$H_0:\ \alpha^* = (\alpha_1,\ldots,\alpha_{I-1}) = \mathbf{a},$$

donde $\mathbf{a}=(a_1,\ldots,a_{I-1})$ es un vector dado. el caso más usado es $\mathbf{a}=0$. Se sabe que bajo H_0, cuando $J\to\infty$ es

$$J^{1/2}.(\hat{\alpha}^* - \mathbf{a}) \xrightarrow{D} N(0,\ \sigma_0^2.V(\psi,F_u).\mathbf{C}), \qquad (100)$$

donde se usa la misma notación que en la expresión (48) de la sección 5. Esto sugiere usar como estadístico para el test

$$U = J.(\hat{\alpha}^* - \mathbf{a})^t.\mathbf{C}^{-1}.(\hat{\alpha}^* - \mathbf{a}) \ /\ s^2.\hat{V}(\psi,F_u).$$

donde $\hat{V}(\psi,F_u)$ es el estimador dado en (94.1). Observando la forma particularmente simple de la matriz $\mathbf{C}$, puede calcularse $\mathbf{C}^{-1}$ y resulta que

$$U = J.((I-1)/I).\sum_{i=1}^{I}(\hat{\alpha}_i-a_i)^2 \ /\ (s^2.\hat{V}(\psi,F_u)). \qquad (101)$$

(En esta expresión se entiende que a_I se define de modo que $\Sigma a_i=0$)

Se propone como test rechazar H_0 si

$$U \quad \chi^2(I-1,\alpha),$$

que es un test de nivel asintótico α.

En el caso particular en que ψ es la función identidad, el estadístico U recién definido coincide con el estadístico F del análisis de la varianza clásico multiplicado por el factor (I-1).

Potencia asintótica del test propuesto: Considérese que si el número de bloques es J se cumple la hipótesis alternativa $\alpha=\mathbf{a}(J)$ donde la sucesión $\mathbf{a}(J)$ cumple que

$$\lim_{J\to\infty} J^{1/2}.(\mathbf{a}(J) - \mathbf{a}) = \mathbf{d}.$$

Por la equivariancia de los M-estimadores, el estimador de α bajo la hipótesis alternativa $\alpha=\mathbf{a}(J)$ será

$$\hat{\alpha}^*(\mathbf{a}(J)) = \hat{\alpha}^* + \mathbf{a}(J) - \mathbf{a}$$

donde $\hat{\alpha}^*$ es el estimador bajo H_0 (se omite como hasta ahora en la notación su dependencia del número de bloques J) que se sabe que cumple (100). Luego

$$J^{1/2}.(\hat{\alpha}^*(\mathbf{a}(J))-\mathbf{a}) = J^{1/2}.(\hat{\alpha}^*-\mathbf{a}+\mathbf{a}(J)-\mathbf{a}) \xrightarrow{D} N(\mathbf{d},\sigma_0^2.V(\psi,F_u).\mathbf{C})$$

y el estadístico del test

$$U = J.(\hat{\alpha}^*(\mathbf{a}(J))-\mathbf{a})^t.\mathbf{C}^{-1}.(\hat{\alpha}^*(\mathbf{a}(J))-\mathbf{a}) \ /\ s^2.\hat{V}(\psi,F_u)$$

converge en distribución a una χ^2 no central con I-1 grados de libertad y parámetro de no centralidad

$$\delta^2 = \mathbf{d}^t.\mathbf{C}^{-1}.\mathbf{d} \ /\ (\sigma_0^2.V(\psi,F_u)).$$

La eficiencia asintótica relativa de dos tests de este tipo es el cociente entre los parámetros de no centralidad, o sea el cociente entre los valores de $\sigma_0^2.V(\psi,F_u)$ para ambos tests, que coincide con la eficiencia asintótica relativa entre los estimadores.

9. RESULTADOS DE UNA SIMULACION. Se hizo un estudio por simulación para comparar (para un tamaño de muestra relativamente pequeño) la eficiencia del test propuesto en la sección anterior (usando la función ψ_k de Huber con parámetro k=1.345), con la del test F del análisis de la varianza clásico y con las de dos tests de rangos que se usan para el diseño en bloques: el test de Friedman y el test de rangos alineados (por ejemplo Lehmann (1975)). La simulación se hizo con I=3 tratamientos y J=20 bloques. La hipótesis nula que se estudió es

$$H_0:\ \alpha_i=0 \quad \forall i=1,\ldots,3.$$

En la sección anterior se vió que el estadístico dado por (101), o sea

$$U = J.((I-1)/I).\sum_{i=1}^{I} \widehat{\alpha}_i^2 \ /\ (s^2.\widehat{V}(\psi,F_u))$$

tiene distribución asintótica χ^2 bajo H_0. En primer lugar se hizo una simulación con 26000 muestras bajo H_0 y distribución normal para verificar si ya para J=20 el estadístico U tenía una distribución próxima a la distribución asintótica. Se estimó la probabilidad de que el estadístico U superara a los valores críticos de la distribución χ^2 con 2 grados de libertad al 10%, 5% y 1% y resultó:

$$\widehat{P}(U>\chi^2_{2,0.10}=4.61) = 0.121,$$

con un intervalo de confianza al 95%: [0.118; 0.123];

$$\widehat{P}(U>\chi^2_{2,0.05}=5.99) = 0.067,$$

con un intervalo de confianza al 95%: [0.065; 0.069];

$$\widehat{P}(U>\chi^2_{2,0.01}=9.21) = 0.019,$$

con un intervalo de confianza al 95%: [0.018; 0.020].

Puede apreciarse que el ajuste por la distribución asintótica no es enteramente satisfactorio. Por ello el paso siguiente consistió en estimar los percentiles 90, 95 y 99 de la distribución de U (siempre bajo normalidad y H_0).

Para estimar los percentiles de U, no se utilizaron directamente los percentiles muestrales, pues para cada muestra generada, además de U se observa otro estadístico fuertemente correlacionado con U: el estadístico F, del análisis de la varianza clásico y esto permite obtener mejores estimadores de los percentiles de U. Supóngase que se desea estimar el percentil 1-α de la distribución de U. Se usa la misma idea que en el muestreo estratificado considerando dos estratos: un estrato lo forman las muestras para las que $F \leq F(\alpha)$ (donde se nota $F(\alpha)$ al percentil 1-α de la distribucion F con I-1 y (I-1)*(J-1) grados de libertad) y el otro estrato las muestras para las que $F > F(\alpha)$. Como

$$P(U \leq x) = (1-\alpha).P(U \leq x | F \leq F(\alpha)) + \alpha.P(U \leq x | F > F(\alpha)), \quad (102)$$

se estiman las probabilidades condicionales que aparecen en la expresión (102) con las proporciones muestrales correspondientes y reemplazando en (102) se obtiene un estimador para $P(U \leq x)$ para cualquier valor de x. Finalmente el percentil 1-α de la distribución de U se estima como el valor de x que resuelve la ecuación

$$\hat{P}(U \leq x) = 1-\alpha.$$

Los estimadores de los percentiles así obtenidos se muestran en la tabla siguiente.

Tabla 1.1: Percentiles de la distribución del estadístico bajo H_0 y distribución normal (estimados con 26000 muestras).

PERCENTIL	ESTIMADOR PUNTUAL	INT. DE CONF. AL 95%
P_{90}	5.02	[4.97 ; 5.08]
P_{95}	6.68	[6.61 ; 6.78]
P_{99}	10.90	[10.70 ; 11.12]

Para comparar las potencias de los cuatro tests mencionados, se simularon primero muestras con distribución normal, bajo la hipótesis alternativa

$$H_1: \alpha_1=-0.36 \;,\; \alpha_2=0 \;,\; \alpha_3=0.36,$$

que fue elegida pues bajo esta alternativa el test F, con un nivel de significación del 5%, tiene una potencia de alrededor del 50%. Se obtuvieron los siguientes resultados.

Tabla 1.2: Porcentaje de veces que se rechazo H_0 en 5300 muestras generadas bajo H_1 y distribución normal.

Test	Nivel de significación		
	10%	5%	1%
F	61.7%	48.5%	24.6%
U (robusto)	60.0%	47.1%	22.6%
Friedman	51.6%	38.3%	14.8%
Rangos alineados	58.2%	44.6%	20.8%

Se aprecia en la tabla anterior que en el caso normal se pierde muy poca eficiencia al usar el test robusto en lugar del test óptimo (el test F), se pierde un poco más de eficiencia con el test de rangos alineados y, como es sabido, el test de Friedman pierde bastante eficiencia para la distribución normal cuando el número de tratamientos es pequeño.

A continuación se generaron muestras con la siguiente distribución normal contaminada: 90% de los errores aleatorios u_{ij} tenían distribución $N(0,\sigma^2)$ y 10% distribución $N(0,25.\sigma^2)$. Los resultados obtenidos bajo H_0 y H_1 se presentan a continuación en las tablas 1.3 y 1.4. Hemos usado dos versiones del test robusto, una empleando como valores críticos los percentiles estimados bajo la hipótesis nula y normalidad, que están en tabla 1.1 y otra usando los valores críticos de la distribución asintótica χ^2.

Tabla 1.3: Porcentaje de veces que se rechazó H_0 en 3600 muestras generadas bajo H_0 y distribución normal contaminada (NC(0.10; 5))

T e s t	N i v e l d e s i g n i f i c a c i ó n		
	10%	5%	1%
F	9.8%	4.4%	0.47%
U (robusto) (1) versión 1	9.9%	4.6%	0.81%
versión 2	11.6%	6.6%	1.75%
Friedman	11.0%	5.6%	1.06%
Rangos alineados	10.9%	5.4%	1.00%

(1) versión 1: usando como valores críticos los percentiles de la tabla 1.1; versión 2: usando como valores críticos los percentiles de la distribución asintótica χ^2.

Tabla 1.4: Porcentaje de veces que se rechazó H_0 en 3600 muestras generadas bajo H_1 y distribución normal contaminada (NC(0.10; 5))

T e s t	N i v e l d e s i g n i f i c a c i ó n		
	10%	5%	1%
F	30.1%	20.4%	7.1%
U (robusto)(1) versión 1	41.6%	29.5%	11.9%
versión 2	45.0%	34.4%	16.7%
Friedman	39.6%	27.5%	9.6%
Rangos alineados	42.5%	30.5%	11.4%

(1) ver nota al pié de tabla 1.3.

En la tabla 1.3 se aprecian pocas diferencias entre los tests en cuanto a la probabilidad de error tipo I y en la 1.4 se observa el franco desmejoramiento de la potencia del test F ante la presencia de 10% de observaciones atípicas. De los otros tests considerados sigue siendo el de Friedman el menos eficiente.

Finalmente se simularon muestras con una distribución con colas muy pesadas: la distribución de Cauchy. En la tabla siguiente se muestran los resultados bajo H_0 y en la los obtenidos bajo la hipótesis alternativa

$$H_2: \alpha_1=-0.534 \ , \ \alpha_2=0 \ , \ \alpha_3=0.534 \ .$$

Tabla 1.5: Porcentaje de veces que se rechazó H_0 en 3600 muestras generadas bajo H_0 y distribución Cauchy.

Test	Nivel de significación		
	10%	5%	1%
F	3.7%	1.4%	0.17%
U (robusto) (1) versión 1	8.8%	3.8%	0.56%
versión 2	10.6%	5.3%	1.11%
Friedman	10.6%	4.9%	0.86%
Rangos alineados	10.7%	5.0%	0.67%

(1) ver nota al pié de tabla 1.3.

Tabla 1.6: Porcentaje de veces que se rechazó H_0 en 3600 muestras generadas bajo H_2 y distribución Cauchy.

Test	Nivel de significación		
	10%	5%	1%
F	6.6%	3.1%	0.6%
U (robusto) (1) versión 1	24.6%	15.2%	3.9%
versión 2	27.8%	18.1%	7.1%
Friedman	37.2%	25.6%	8.4%
Rangos alineados	34.9%	22.6%	8.2%

(1) ver nota al pié de tabla 1.3.

Como la distribución de Cauchy no tiene esperanza, los estimadores de cuadrados mínimos no son consistentes y era de esperar que el test F tuviese una potencia bajísima, como se aprecia en la tabla 1.6. En esta misma tabla se aprecia que para esta distribución tan poco parecida a la normal, el ordenamiento de los test en cuanto a eficiencia es inverso al que ocurre en el caso normal (tabla 1.2): el más potente es el test de Friedman, seguido por el de rangos alineados, luego el test robusto y finalmente, como ya dijimos, el comportamiento del test F, que no permite distiguir la hipótesis nula de la alternativa.

La conclusión que se deduce de la simulación es que el test F, que sabemos óptimo para el caso normal, se desmejora francamente ni bien nos apartamos un poco de esa distribución. El test de Friedman tampoco es aconsejable, pues su comportamiento no es bueno para la distribución normal. Los tests más satisfactorios son sin duda el test de rangos alineados y el test robusto. El primero, apenas menos eficiente que el robusto para la distribución normal, tiene la ventaja de ser más potente para distribuciones muy alejadas de la normal (Cauchy).

Para saber si la mayor potencia del test de rangos alineados para la distribución de Cauchy se debía al método de M-estimadores o a la función ψ elegida, se hizo luego otra simulación, incluyendo ahora otro test robusto: el basado en los M-estimadores calculados con una función ψ de la familia de funciones bicuadradas propuestas por Tukey, dicha familia está dada por:

$$\begin{aligned} \psi_k(x) &= x.(1-(x/k)^2)^2 \qquad && \text{si} \quad |x| \leq k, \\ &= 0 && \text{si} \quad |x| > k \end{aligned}$$

Es importante destacar que las funciones ψ de esta familia no son monótonas, así que no se cumplen las suposiciones bajo las cuales se ha demostrado en este trabajo la consistencia de los M-estimadores, esta segunda simulación se basa en la presunción de que los resultados obtenidos en este trabajo siguen valiendo para una función ψ de esta forma.

Los resultados de la simulación que se presentan a continuación se obtuvieron continuando la generación de números aleatorios a partir de los obtenidos en las simulaciones anteriores, o sea que los resultados son independientes de los que se muestran en las tablas 1.2 a 1.6. A cada muestra simulada se le aplicaron dos de los tests que se estudiaron en la simulación anterior: el test de rangos alineados y el test robusto basado en la función ψ_k de Huber con k=1.345 (usando como valor crítico el percentil correspondiente de la distribución asintótica χ^2) y un tercer test calculado en forma enteramente análoga al robusto anterior pero usando la función ψ_k bicuadrada de Tukey con k=4.685 (valor usualmente empleado para la función bicuadrada, que tiene la misma propiedad que el valor k=1.345 para la función ψ_k de Huber: permite lograr que la eficiencia relativa del estimador de cuadrados mínimos con respecto al M-estimador para el problema de posición y distribución normal sea 1.05). No se aplicaron ni el test F ni el de Friedman, pues ya se observó en la simulación anterior la poca conveniencia de emplear estos tests. En esta nueva simulación se usaron las mismas distribuciones y las mismas hipótesis alternativas que en la anterior. Los resultados obtenidos se presentan en las tablas 2.1 a 2.6.

Tabla 2.1: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_0 y distribución normal.

T e s t	N i v e l d e s i g n i f i c a c i ó n		
	10%	5%	1%
Robusto (ψ de Huber)	12.1%	6.4%	1.5%
Robusto (ψ bicuadrada)	12.4%	7.4%	2.6%
Rangos alineados	9.8	4.6	1.1%

Tabla 2.2: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_1 y distribución normal.

T e s t	N i v e l d e s i g n i f i c a c i ó n		
	10%	5%	1%
Robusto (ψ de Huber)	63.9%	52.6%	30.7%
Robusto (ψ bicuadrada)	63.8%	51.4%	30.4%
Rangos alineados	57.5%	44.6%	20.1%

Tabla 2.3: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_0 y distribución normal contaminada (NC(0.10,5)).

Test	Nivel de significación		
	10%	5%	1%
Robusto (ψ de Huber)	12.3%	6.3%	1.8%
Robusto (ψ bicuadrada)	13.1%	7.9%	3.0%
Rangos alineados	11.1%	5.3%	1.1%

Tabla 2.4: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_1 y distribución normal contaminada (NC(0.10,5)).

Test	Nivel de significación		
	10%	5%	1%
Robusto (ψ de Huber)	47.9%	35.4%	15.9%
Robusto (ψ bicuadrada)	51.7%	38.3%	20.9%
Rangos alineados	45.0%	30.5%	11.7%

Tabla 2.5: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_0 y distribución Cauchy.

Test	Nivel de significación		
	10%	5%	1%
Robusto (ψ de Huber)	10.4%	5.1%	0.9%
Robusto (ψ bicuadrada)	12.2%	6.6%	2.2%
Rangos alineados	10.2%	5.1%	0.7%

Tabla 2.6: Porcentaje de veces que se rechazó H_0 en 1600 muestras generadas bajo H_2 y distribución Cauchy.

Test	Nivel de significación		
	10%	5%	1%
Robusto (ψ de Huber)	27.1%	17.5%	6.1%
Robusto (ψ bicuadrada)	38.4%	28.7%	14.4%
Rangos alineados	34.1%	22.6%	8.0%

En la tabla 1.6 se había observado que para la distribución de Cauchy el test basado en los M-estimadores era menos eficiente que el de rangos alineados. En la tabla 2.6 se aprecia que este desventaja de los M-estimadores se corrige al usar la función ψ bicuadrada, que se comporta mucho mejor que la de Huber para el caso Cauchy y que llega a superar un poco al test de rangos alineados.

Se observa en las tablas 2.1, 2.3 y 2.5 que el test robusto usando la ψ bicuadrada converge más lentamente a su distribución asintótica que cuando se usa la ψ de Huber y esto hace que el nivel de significación sea un poco mayor que el nominal. Este problema podría corregirse usando como valor crítico del test el percentil estimado bajo distribución normal y H_0, como se hizo en las tablas 1.2 a 1.6.

APENDICE A

En este apéndice se demuestra que si se supone que se cumple el modelo (5) con F_u continua y que ψ y χ cumplen la suposición A, entonces el sistema de ecuaciones (8) y (9) tiene solución con probabilidad uno. Este problema de la existencia de soluciones para el sistema de ecuaciones, cuando se estiman simultáneamente los parámetros de la función de regresión y la dispersión, surge también para el modelo lineal general, no es un problema específico del modelo de diseño en bloques y ha sido tratado por Huber (1981, sección 7.7). Aquí se sigue su idea, aunque nos restringimos al caso del diseño en bloques.

Considérese la función

$$Q(\sigma,\alpha,\beta) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sigma.\rho((X_{ij}-\alpha_i-\beta_j)/\sigma) + (I-1).(J-1).c$$

(en realidad puede considerarse a Q como función de ($\beta_1,\ldots,\beta_J$) pues $\alpha_I=-\Sigma\alpha_i$).

Puede verse que un punto satisface el sistema de ecuaciones (8) si y solo si anula las derivadas parciales de esta función con respecto a α_i y β_j y que la derivada con respecto a σ es (usando que $\chi(x)=x.\psi(x)-\rho(x)$):

$$-\sum_{i=1}^{I}\sum_{j=1}^{J}\chi((X_{ij}-\alpha_i-\beta j)/\sigma) \quad + \quad (I-1).(J-1).c \tag{A1}$$

que, si se iguala a cero es la ecuación (9).

Se demuestra a continuación que la función Q tiene algún minimo local:

i) Consideremos la funcion Q en el conjunto $\{\sigma\geq 0,\alpha,\beta\}$. Como la definimos no estaría definida para $\sigma=0$, pero puede extenderse su definición por continuidad. En efecto, existe

$$\lim_{x\to\infty} \rho(x)/|x| = a \quad \text{con} \quad <a\leq+\infty \tag{A2}$$

ya que

$$(\rho(x)/x)' = \chi(x)/x^2,$$

luego $\rho(x)/x$ es estrictamente creciente para y por lo tanto tiene límite para $x\to+\infty$, que coincide con el límite para $x\to-\infty$ pues ρ es par. Ademas, si x_0 es cualquier número positivo, ese límite es mayor que $\rho(x_0)/x_0$, asi que es $a>0$. Se ve facilmente que de (A2) resulta

$$\lim_{\sigma\to 0} Q(\sigma,\alpha,\beta) = a.\sum_{i=1}^{I}\sum_{j=1}^{J}|X_{ij}-\alpha_i-\beta_j|$$

Luego la función Q puede extenderse por continuidad al conjunto $\{\sigma\geq 0,\alpha,\beta\}$, siempre que a sea finito.

ii) Como por ser $\rho\geq 0$ es $Q(\sigma,\alpha,\beta) \geq \sigma.(I-1).(J-1).c$, resulta que

$$\lim_{\sigma\to+\infty} Q(\sigma,\alpha,\beta) = +\infty \qquad \text{uniformemente en } (\alpha,\beta) \tag{A3}$$

iii) Si uno de los $\alpha_i\to+\infty$ toma valores tan grandes como se desee, otro $\alpha_{i*}\to$ toma también valores tan grandes como se desee en valor absoluto, pero con signo contrario, pues los α_i suman cero. Luego o bien $|\alpha_i+\beta_j|$, o bien $|\alpha_{i*}+\beta_j|$, toman valores tan grandes como se quiera. Como ρ es convexa y tiene un minimo estricto en cero, resulta que $\rho(x)\to+\infty$, cuando $x\to\pm\infty$ y de aquí $\rho(X_{ij}-\alpha_i-\beta_j)/\sigma)$ o $\rho(X_{i*j}-\alpha_{i*}-\beta_j)/\sigma)$ toman valores tan grandes como se quiera y luego la función Q tiende a infinito cuando $\text{Max}|\alpha_i|$ tiende a infinito, y la convergencia es uniforme si σ se mantiene en un intervalo $<m\leq\sigma\leq M$. Lo mismo ocurre cuando $\text{Máx}|\beta_j|\to\infty$, o sea en resumen

$$\lim_{\text{Max}(|\alpha_i|,|\beta_j|)\to\infty} Q(\sigma,\alpha,\beta) = +\infty \text{ , uniformemente si } \sigma \text{ se mantiene en un intervalo } 0<m\leq\sigma\leq M. \qquad (A4)$$

iv) Veremos ahora que pasa cuando $\sigma\to 0$ y $\text{Max}(|\alpha_i|,|\beta_j|)\to\infty$. Como se vio en iii) si algun $|\alpha_i|$ o $|\beta_j|$ es "grande" entonces algun $|\alpha_i+\beta_j|$ tambien. Luego

$$\sigma.\rho((X_{ij}-\alpha_i-\beta_j)/\sigma) = \rho((X_{ij}-\alpha_i-\beta_j)/\sigma)/(|X_{ij}-\alpha_i-\beta_j|/\sigma).|X_{ij}-\alpha_i-\beta_j|$$
$$\geq (a-\varepsilon).|X_{ij}-\alpha_i-\beta_j| \qquad (a-\varepsilon).(|\alpha_i+\beta_j|-\text{Máx}|X_{ij}|)$$

Luego

$$\lim_{\substack{\text{Max}(|\alpha_i|,|\beta_j|)\to\infty \\ \sigma\to 0}} Q(\sigma,\alpha,\beta) = +\infty \qquad (A5)$$

v) Consideremos un punto cualquiera $(\sigma_o,\alpha_o,\beta_o)$ y llamemos

$$q_o = Q(\sigma_o,\alpha_o,\beta_o) \text{ .}$$

Por (A3), $\exists$ S tq. si $\sigma>S$ entonces $Q(\sigma,\alpha,\beta) \quad q_o$.

Por (A5) $\exists$ $s>0$ y M_o tq. si $0<\sigma<s$ y $\text{Max}(|\alpha_i|,|\beta_j|) \quad M_o$ entonces $Q(\sigma,\alpha,\beta) > q_o$.

Como, por (A4),

$$\lim_{\text{Max}(|\alpha_i|,|\beta_j|)\to\infty} Q(\sigma,\alpha,\beta) = +\infty \text{ , uniformemente si } \sigma \text{ se mantiene en el intervalo } s\leq\sigma\leq S,$$

luego $\exists$ M_1 tq. si $s\leq\sigma\leq S$ y $\text{Max}(|\alpha_i|,|\beta_j|) \quad M_1$ entonces $Q(\sigma,\alpha,\beta) > q_o$.

Sea $M=\text{Max}(M_o,M_1)$. De lo dicho hasta ahora en este punto v), si $\sigma>S$ o $\text{Max}(|\alpha_i|,|\beta_j|) \quad M$ entonces $Q(\sigma,\alpha,\beta) \quad q_o$. Luego el infimo de la funcion Q se toma en el compacto

$$\{0\leq\sigma\leq S,\ \text{Max}(|\alpha_i|,|\beta_j|) \leq M\}.$$

Queremos ver que dicho infimo no se toma en un punto con $\sigma=0$. Si a es infinito, esto es evidente pues en dicho caso el límite de la función para $\sigma\to 0$ es $+\infty$. Si a es finito, la funcion Q es continua en el compacto, luego alcanza el infimo. Observando la expresión (A1) para la derivada de Q con respecto a σ, resulta que (si fijamos (α,β))

$$\lim_{\sigma\to 0} \partial Q(\sigma,\alpha,\beta)/\partial\sigma = (I-1).(J-1).c - (I.J-N).\chi(\infty)$$

donde $N=\#\{(i,j) \text{ tq. } X_{ij}-\alpha_i-\beta_j=0\}$. Pero $N\leq I+J-1$ con probabilidad uno, pues $I+J-1=p$ es el rango de la matriz de variables explicativas del modelo lineal que estamos considerando y si en un modelo lineal cualquiera el numero de residuos fuese mayor que dicho rango, es porque p+1 componentes del vector Y (que contiene a las observaciones de la variable dependiente) estarían en un subespacio de dimension p, lo que tiene prob cero si las componentes del vector **Y** son v.a. independientes, con función de distribución continua, como hemos supuesto. De $N \leq I+J-1$, resulta $I.J-N \geq (I-1).(J-1)$ y como $\chi(\infty)>c$, resulta que, para cada (α,β) fijo es

$$\lim_{\sigma \to 0} \partial Q(\sigma,\alpha,\beta)/\partial\sigma \qquad (A6)$$

Queremos ver que el mínimo de Q no puede tomarse en un punto $(0,\alpha_0,\beta_0)$. Si lo tomara, tomaría un mínimo en cero la funcion de una variable $Q^*(\sigma)=Q(\sigma,\alpha_0,\beta_0)$, pero por (A6) sería $\lim Q^{*\prime}(\sigma)<0$, para $\sigma \to 0$, lo que es una evidente contradicción.

Por todo lo dicho, el ínfimo (mínimo) de la función Q se alcanza en

$$\{0<\sigma\leq S,\ \mathrm{Max}(|\alpha_i|,|\beta_j|)\ \leq M\},$$

este mínimo absoluto es también un mínimo local, por lo tanto Q tiene algún punto crítico, que resuelve el sistema de ecuaciones (8) y (9), como queríamos demostrar.

Para demostrar la consistencia de los estimadores en la sección 4, no se necesita suponer que el sistema (8) y (9) tenga solución con probabilidad uno, cualquiera sea J, como se acaba de demostrar, sino sólo que dicho sistema tenga solución, a partir de un J en adelante, con probabilidad uno. Es intuitivamente razonable que para que esto sea cierto, no es necesario suponer que F_u es continua, sino alguna hipótesis más débil que permita que la distribución de los errores tenga puntos pesados, pero con una cota para el peso de esos puntos (como se ha hecho por ejemplo en el teorema 4). En este trabajo no se demuestra un resultado de este tipo.

APENDICE B

Lema 9. Bajo las hipótesis del teorema 5.1 o las del teorema 5.2 resulta:

$$p\lim_{J\to\infty} (1/J^{1/2})\ \left\|\sum_{j=1}^{J} \psi^*(\mathbf{u}_j,0,s) - \sum_{j=1}^{J} \psi^*(\mathbf{u}_j,0,\sigma_0)\ \right\| = 0.$$

<u>Demostración</u>: Por la definición de ψ^*, la tesis es equivalente a que

$$(1/J^{1/2})\ \left|\sum_j \psi((u_{ij}-\beta(\mathbf{u}_j,0,s))/s) - \sum_j \psi((u_{ij}-\beta(\mathbf{u}_j,0,\sigma_0))/\sigma_0)\right|$$

tiende a cero en probabilidad, para todo $i=1,\ldots,I$.

Sea

$$y_{Ji}(t) = (1/J^{1/2}) \sum_{j=1}^{J} \psi((u_{ij}-\beta(u_j,0,0.5\sigma_o+t\sigma_o)/(0.5\sigma_o+t\sigma_o)) ,$$

y_{Ji} es un elemento de C, el espacio de las funciones continuas en [0,1].

Queremos demostrar que, para cada i fijo, la sucesión $\{y_{Ji}\}$ de distribuciones sobre C es "tight". Para ello usaremos el teorema 12.3 de Billingsley(1968) que afirma que dicha propiedad es equivalente a las dos condiciones siguientes

i) La sucesion $\{y_{Ji}(0)\}$ es "tight".

ii) Existen constantes $b\geq 0$ y $d>1$ y una función g continua, no decreciente, que toma valores en [0,1] tal que

$$P\{|y_{Ji}(t^*) - y_{Ji}(t)| \geq \lambda\} \leq (1/\lambda^b).|g(t^*)-g(t)|^d$$

para todo t, t*, J y todo $\lambda>0$.

La propiedad i) es cierta gracias al teorema central del limite.

Para demostrar ii) observemos que

$$Var\ (y_{Ji}(t^*) - y_{Ji}(t)) =$$

$$E(\psi((u_i - \beta(u,0,0.5\sigma_o+t^*\sigma_o)/(0.5\sigma_o+t^*\sigma_o)) -$$

$$\psi((u_i - \beta(u,0,0.5\sigma_o+t\sigma_o)/(0.5\sigma_o+t\sigma_o)))^2 \qquad (B1)$$

Si se cumplen las hipótesis del teorema 5.1, consideremos la derivada $\partial\psi((u_i-\beta(u,0,\sigma)/\sigma)/\partial\sigma$, procediendo en forma similar a como se calculó (53) resulta igual a

$$(1/\sigma).\psi'(r_i).(\sum_{l=1}^{I}\psi'(r_l).r_l/\sum_{l=1}^{I}\psi'(r_l) - r_i)$$

donde, para simplificar la notación se llamó $r_i=(u_i-\beta(u,0,\sigma))/\sigma$. Esta derivada es acotada, siempre que se mantenga σ alejado de cero, por ser $x.\psi'(x)$ acotada. Si llamamos M a la cota superior del módulo de esta derivada, para $\sigma\geq 0.5\sigma_o$, aplicando el teorema del valor medio al incremento que está dentro del signo esperanza en (B1) resulta que (B1) es menor o igual que

$$(M.(t^*\sigma_o - t\sigma_o))^2 , \qquad (B1^*)$$

lo que, usando la desigualdad de Tchebycheff implica ii), con lo que queda probado que la sucesión $\{y_{Ji}(t)\}$ es "tight".

Si no se cumplen las hipótesis del teorema 5.1, pero sí las del teorema 5.2, de

$$\psi((x-c(\sigma^*-\sigma))/\sigma^*) \leq \psi(x/\sigma) \leq \psi((x+c(\sigma^*-\sigma))/\sigma^*)$$

resulta

$$|\beta(u,0,\sigma^*) - \beta(u,0,\sigma)| \leq c.|\sigma^*-\sigma|$$

y de aquí resulta que si $\sigma^*\geq\sigma_o/2$ y $\sigma\geq\sigma_o/2$ entonces

$$\psi((u_1-\beta(u,0,\sigma^*)/\sigma^*) - \psi((u_1-\beta(u,0,\sigma)/\sigma) \leq M.|\sigma^*-\sigma|$$

donde $M=4.b.c/\sigma_o$. De aquí se deduce (B1*) y el mismo razonamiento anterior permite demostrar que $\{y_{J1}(t)\}$ es "tight".

El teorema 8.2 de Billingsley(1968) afirma que si la sucesión $\{y_{J1}\}$ es "tight" entonces para cada $\varepsilon>0$ y $v>0$ existe un $\delta>0$ y un entero J_o tq.

$$P\{\sup_{|t^*-t|<\delta}|y_{J1}(t^*) - y_{J1}(t)| \geq \varepsilon\} \leq v, \text{ para } J\geq J_o. \quad (B2)$$

Sea $t_o=0.5$, $t_J=s/\sigma_o-0.5$. Como $s\to\sigma_o$ con probabilidad uno y en consecuencia también en probabilidad, $t_J\to t_o$ en probabilidad, luego existe J_1 tq.

$$P\{|t_J-t_o|>\delta\} \quad v, \text{ para } J\geq J_1.$$

Luego de (B2) y la definición de y_{IJ} resulta que

$$P\{(1/J^{1/2})|\Sigma\psi((u_{ij})-\beta(u,0,s))/s)-\Sigma\psi((u_{ij})-\beta(u,0,\sigma_o))/\sigma_o)|$$

$$2.v \text{ , para } J\geq \text{máx}(J_o,J_1)$$

lo que implica la tesis del lema 9.

Lema 10. Sean x_i $(i=1,...,I)$ números reales tales que $|x_i|\leq d$ y $\Sigma x_i=0$ entonces

$$\Sigma x_i^2 \leq I.d^2 \quad \text{si I es par,}$$
$$\leq (I-1).d^2 \quad \text{si I es impar.}$$

Demostración: Este lema es trivial para el caso I par y para el caso I=1. Consideremos entonces el caso I impar, mayor o igual que 3. Como la función Σx_i^2 es continua, tiene máximo en el conjunto $S = \{\mathbf{x}\in R^n : |x_i|\leq d \text{ y } \Sigma x_i=0\}$. Sea $(a_1,...,a_n)$ un punto donde se alcanza el máximo. Veremos primero que todos los a_i, salvo a lo sumo uno, cumplen que $|a_i|=d$. En efecto, supongamos que dos de los a_i, por ejemplo a_1 y a_2 tuviesen módulo menor que d, y supongamos, sin perder generalidad, que $a_1\leq a_2$. Sea $\delta>0$ tq. $a_1-\delta$ y $a_2+\delta$ tienen módulo menor que d. Pero entonces el punto $(a_1-\delta,a_2+\delta,a_3,...,a_n)$ también pertenecería a S y en ese punto la función sería estrictamente mayor que en **a**, lo que sería absurdo. Todos los a_i no pueden tener módulo igual a d, pues no se podría cumplir la condición $\Sigma d_i=0$. Luego (I-1) de los a_i son tq. $|a_i|=d$ y el restante cumple $|a_i|<d$, lo que junto a la restricción $\Sigma a_i=0$ implica que $\Sigma a_i^2=(I-1).d^2$, como queríamos demostrar.

Víctor J. Yohai

García Ben

REFERENCIAS

Billingsley, A. (1968) Convergence of Probability Measures. Wiley, New York.

Huber, P.J. (1964) Robust estimator of a location parameter. Ann. Math. Statist., vol 35, 73-101.

Huber, P.J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. V Berkeley Symposium..

Huber, P.J. (1973) Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Statist., vol. 1, 799-821.

Huber, P.J. (1981) Robust statistics. Wiley, New York.

Lehmann, E.L. (1975) Nonparametrics: statistical methods based on ranks. Holden-Day, San Francisco.

Portnoy, S. (1984) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. Ann. Statist., vol.12, nro.4, 1298-1309.

Portnoy, S. (1985) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Asymptotic normality. Ann., Statist., vol.13, nro.4, 1403-1417.

Schrader, R.M. y McKean, J.W. (1979) Robust analysis of variance. Communications in Statistics, vol A6, nro.9, 879-894.

Schrader, R.M. y Hettmansperger, T.P. (1980) Robust analysis of variance based upon a likelihood ratio criterion. Biometrika, vol. 67, nro.1, 93-101.

Yohai, V.J. y Maronna, R.A. (1979) Asymptotic behavior of M-estimators for the linear model. Ann. Statist., vol 7, 258-268

Yohai, V.J. (1987) High breakdown-point and high efficiency robust estimates for regression. Ann. Statist., vol.15.