

**COMPARACIÓN DE ALGUNOS MÉTODOS DE REGRESIÓN ALTERNATIVA  
VS. ESTADÍSTICA BAYESIANA USANDO MCMC**

**ALFONSO SÁNCHEZ HERNÁNDEZ**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
FACULTAD DE INGENIERÍA INDUSTRIAL  
MAESTRÍA EN INVESTIGACIÓN OPERATIVA Y ESTADÍSTICA  
PEREIRA**

**2014**

**COMPARACIÓN DE ALGUNOS MÉTODOS DE REGRESIÓN ALTERNATIVA  
VS. ESTADÍSTICA BAYESIANA USANDO MCMC**

**ALFONSO SÁNCHEZ HERNÁNDEZ**

Trabajo de grado para el título de  
Magister en Investigación Operativa y Estadística

**Director:**

**JAIRO ALFONSO CLAVIJO MÉNDEZ**  
MSc. en Estadística

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
FACULTAD DE INGENIERÍA INDUSTRIAL  
MAESTRÍA EN INVESTIGACIÓN OPERATIVA Y ESTADÍSTICA  
PEREIRA  
2014**

Nota de aceptación

---

---

---

---

Presidente del jurado

---

Jurado

---

Jurado

Pereira, Marzo de 2014

## Dedicatoria

A Dios y Jesucristo por guiarme e iluminarme para seguir por el camino correcto.

A mi amada esposa, María Eugenia y mis preciosos hijos Daniel Alfonso y David Esteban por su amor, sonrisa, cariño, comprensión y por enseñarme a tener la fortaleza de continuar hacia adelante sin importar las circunstancias que la vida nos presenta.

A todos mis hermanos por brindarme siempre su apoyo y colaboración, durante este proceso.

A mis difuntos padres Gregorio Sánchez y María de los Angeles Hernández .

**Alfonso Sánchez Hernández**

## **Agradecimientos**

Jairo Alfonso Clavijo Méndez, docente de la Facultad de Ciencias de la Universidad del Tolima y director de este Trabajo de Grado, por su amistad, valioso apoyo, orientación, perseverancia y dedicación para guiarme hasta la culminación de esta tesis.

Julián Alfonso Acuña Collazos por su amistad y apoyo incondicional, ya que con sus aporte se resolvieron muchas situaciones para poder sacar este trabajo adelante.

Alex Zambrano Carbonell, por su amistad y apoyo en la parte computacional del trabajo.

Jhon Fredy Villanueva, por su acompañamiento y colaboración en el desarrollo computacional Bayesiano del presente trabajo.

A la Oficina de Investigaciones y Desarrollo Científico de la Universidad del Tolima, por su apoyo irrestricto en este proceso.

## Resumen

En el presente trabajo se pretenden abordar cuatro técnicas de regresión lineal simple y múltiple, no usuales, las cuales forman parte de una gran variedad de métodos estadísticos conocidos como métodos robustos. Estos métodos en su orden son: regresión MINMAD, MINMAXAD, MINSADBED y MINSADBAD, son contracciones de las frases en inglés con su respectivo significado y metodología. Se realizará su definición, representación geométrica cuando fuere posible, representación matricial, reducción a un problema de Programación Lineal, descripción de algunos algoritmos. Finalmente se realizará una comparación aplicada utilizando metodología Bayesiana, estimando parámetros mediante Cadenas de Markov y Métodos Montecarlo, metodología conocida como MCMC.

Palabras clave: Minmad, Minmaxad, Minsadb, Minsadbed, Mínimos Cuadrados, MCMC.

## **Abstract**

In this master's thesis we study four relatively unusual techniques of simple and multiple linear regression, which are part of the so-called robust statistical methods. These techniques are known respectively as MINMAD, MINMAXAD, MINSADBED and MINSADBAD (their English abbreviations). We present their definitions, geometric representations (when possible), matrix representations, reductions to linear programming problems and a description of their algorithms. We also perform an applied comparison by means of Bayesian methods and parameter estimation via MCMC (Markov Chains and Monte Carlo methods).

Keywords and Phrases: Minmad, Minmaxad, Minsadbad, Minsadbed, Least Squares, MCMC.

# Índice

<b>1. Antecedentes</b>	<b>15</b>
1.1. Regresión LAD según Charnes, Cooper y Ferguson . . . . .	17
1.1.1. Representación Analítica del Problema . . . . .	18
1.1.2. Transformación del Problema . . . . .	19
1.1.3. Representación Geométrica . . . . .	20
<b>2. Regresión Lineal Simple y Múltiple</b>	<b>23</b>
2.1. Regresión Lineal Simple . . . . .	23
2.1.1. Mínimos Cuadrados . . . . .	25
2.2. Regresión Lineal Múltiple . . . . .	26
2.2.1. Modelo estimado y residuos . . . . .	28
2.2.2. Algunos supuestos y resultados importantes . . . . .	28
2.2.3. Prueba de los coeficientes de regresión . . . . .	30
2.2.4. Coeficiente de determinación. . . . .	32
2.3. Influencia Local . . . . .	32
<b>3. Métodos de Regresión Alternativa</b>	<b>36</b>

3.1. Regresión MINMAD . . . . .	36
3.2. Regresión MINMAXAD . . . . .	41
3.3. Regresión MINSADBED . . . . .	42
3.4. Regresión MINSADBAD . . . . .	43
3.5. Prueba de $H_0 : \beta = 0$ en Regresión MINMAD . . . . .	43
3.6. MINMAD para Regresión Lineal Múltiple . . . . .	44
3.6.1. Procedimiento del simplex para la regresión MINMAD . . . . .	46
3.6.2. Algoritmo modificado de Barrodale y Roberts. . . . .	46
3.7. MINMAXAD para Regresión Lineal Múltiple . . . . .	47
3.8. MINSADBED para Regresión Lineal Múltiple . . . . .	49
3.9. MINSADBAD en Regresión Lineal Múltiple . . . . .	51
3.10. Prueba de $H_0 : \beta_{q+1} = \dots = \beta_p = 0$ . . . . .	51
<b>4. Discusiones y Algunos Procedimientos de Simulación</b>	<b>52</b>
4.1. Discusiones . . . . .	52
4.2. Algunos Procedimientos de Simulación . . . . .	54
4.3. Distribuciones $g$ y $h$ de Tukey . . . . .	57
4.3.1. Asimetría . . . . .	57

4.3.2. Elongación . . . . .	57
<b>5. Estadística Bayesiana y Metodología MCMC</b>	<b>59</b>
5.1. Estadística Bayesiana . . . . .	59
5.2. Teorema de Bayes . . . . .	60
5.3. Definición de un Modelo Estadístico . . . . .	61
5.4. Modelo basado en Inferencia Bayesiana . . . . .	61
5.5. Inferencia usando distribuciones apriori conjugadas . . . . .	64
5.5.1. Inferencia para la distribución de Poisson . . . . .	65
5.5.2. Inferencia para la media de una distribución normal . . . . .	66
5.6. Método de Jeffreys . . . . .	67
5.7. Estimación en Inferencia Bayesiana . . . . .	68
5.7.1. Aproximación a la teoría de la decisión . . . . .	68
5.7.2. Intervalos de credibilidad . . . . .	70
5.7.3. Pruebas de Hipótesis en Inferencia Bayesiana . . . . .	71
5.8. Simulación Montecarlo y su implementación en Inferencia Bayesiana . . . . .	72
5.9. Métodos Montecarlo y Cadenas de Markov MCMC . . . . .	73
5.9.1. Algoritmo MCMC . . . . .	74

<b>6. Simulación y Estimación de Parámetros</b>	<b>76</b>
6.1. Estimación vía MCMC . . . . .	76
6.2. Análisis de algunos resultados de la simulación. . . . .	78
<b>7. Aplicación</b>	<b>89</b>
7.1. Método Minmad . . . . .	89
<b>8. Conclusiones</b>	<b>104</b>
<b>9. Recomendaciones</b>	<b>106</b>
<b>Apéndices</b>	<b>113</b>
<b>A. Apéndice de Tablas</b>	<b>113</b>
<b>B. Programas Estadística Bayesiana(OpenBUGS)</b>	<b>149</b>
B.1. Regresión Lineal Simple . . . . .	149
B.2. Regresión Lineal Múltiple . . . . .	151
<b>C. Programas Matlab Regresión Robusta</b>	<b>155</b>
C.1. Minmad . . . . .	155
C.2. Minmaxad . . . . .	157

C.3. Minsadbed . . . . .	158
C.4. Minsadbad . . . . .	161
<b>D. Justificación cálculo de <math>\tau</math>.</b>	<b>165</b>
D.1. Justificación de (35). . . . .	165
<b>E. Programas en R para estimación GLM</b>	<b>166</b>
E.1. Programa GLM completo . . . . .	166
E.1.1. Programa GLM Reducido . . . . .	168
E.1.2. Programa Diagnóstico Regresión MINSADBAD . . . . .	168
E.1.3. Programa Diagnóstico Regresión MINSADBAD . . . . .	171

## Introducción

Uno de los métodos más extensos y discutidos entre las herramientas estadísticas para el análisis de datos es la regresión. La temática relaciona teóricamente la predicción de una o más variables llamadas variables dependientes o respuesta, utilizando como base otras variables llamadas independientes. En algunos casos las variables independientes reciben otros nombres, covariables por ejemplo, también se les denomina variables explicativas o predictoras. Tales problemas son encontrados en casi todas las áreas de la ciencia experimental y la tecnología.

Cuando el modelo usado para la explicar la variable dependiente en términos de variables independientes asume una relación lineal en los parámetros, se tiene un modelo de regresión lineal, en otro caso se tendrá un modelo de regresión no lineal. Abundante literatura referente a diferenciar los casos de linealidad y no linealidad es conocida, pero ese tópico está fuera del alcance de este trabajo, se puntualizará únicamente en el modelo de regresión lineal simple y múltiple.

En la gran mayoría de los textos de regresión lineal simple y múltiple se describen dos métodos para la estimación de parámetros, el método de los mínimos cuadrados, y el de máxima verosimilitud. Estos dos métodos coinciden teóricamente en sus propiedades cuando el modelo que relaciona las variables dependiente e independientes cumple el supuesto de normalidad en los errores, situación que en la realidad difícilmente se presenta y puede llevar a conclusiones erróneas y modelos que no describen apropiadamente la realidad de los datos poblacionales. Estos dos métodos anteriores se conocen como métodos clásicos. No obstante, existen otros métodos menos conocidos, que en algunos textos especializados se encuentran en uno o dos capítulos y son referenciados como temas especiales, algunas veces con poca profundidad teórica y pobre aplicación, generando desinterés por parte del lector.

Uno de los métodos anteriormente mencionados es el de la minimización de la desviación media absoluta entre los valores del modelo observado y el modelo teórico de la variable dependiente, conocido como

minimización de la norma  $L_1$ . Otro método poco común considerado en la literatura es el criterio de Tchebyshev, el cual consiste en minimizar la mayor de las desviaciones absolutas entre el valor observado y predicho. Un tercer método estudia diferentes transformaciones de los errores dando origen a un sinnúmero de estimadores, dentro de los que se cuentan los estimadores Ridge o robustos.

El presente trabajo pretende realizar una recopilación de algunos métodos de regresión lineal, no convencionales, basados en la norma  $L_1$  que dan origen a estimadores que son una alternativa en conjuntos de datos que no cumplen los supuestos de normalidad de los errores y describen con mayor claridad la realidad de los mismos.

Adjunto a lo anterior, el objetivo fundamental del trabajo consiste en estudiar la profundidad matemática de estos métodos, su representación matricial, la reducción a problemas de programación lineal, validación de hipótesis en los parámetros y la bondad que ofrecen estos métodos ante la presencia de valores atípicos, y finalmente realizar una comparación aplicada frente a un método de estimación Bayesiana, mediante simulación, utilizando Métodos Montecarlo y Cadenas de Markov (MCMC).

## 1. Antecedentes

Para poder hablar de Regresión se hace necesario hacer una retrospectiva histórica de lo que antecedió a su descubrimiento. Tal como lo señala Lozano (2010) en su página [prezi.com/qpxysldecq4u/regresion-lineal/](https://prezi.com/qpxysldecq4u/regresion-lineal/) “el término regresión fue introducido por Francis Galton en su libro *Natural Inheritance* (1889) y fue confirmada por su amigo Karl Pearson. Su trabajo se centró en la descripción de los rasgos físicos de los descendientes (variable A) a partir de los de sus padres (variable B). Estudiando la altura de padres e hijos a partir de más de mil registros de grupos familiares, se llegó a la conclusión de que los padres muy altos tenían una tendencia a tener hijos que heredaban parte de esta altura, pero que revelaban también una tendencia a regresar a la media. Galton generalizó esta tendencia bajo la “ley de la regresión universal”: Cada peculiaridad en un hombre es compartida por sus descendientes, pero en media, en un grado menor.

La primera forma de regresión lineal documentada fue el método de los mínimos cuadrados, el cual fue publicado por Legendre en 1805, en *principle of Least Squares*, citado por Eisenhart (1961, [2]) y en dónde se incluía una versión del teorema de Gauss-Márkov. Mínimos cuadrados es una técnica de análisis numérico encajada dentro de la optimización matemática, en la que, dados un conjunto de pares (o ternas, etc), se intenta encontrar la función que mejor se aproxime a los datos, de acuerdo con el criterio de mínimo error cuadrático.

En su forma más simple, intenta minimizar la suma de cuadrados de las diferencias ordenadas (llamadas residuos) entre los puntos generados por la función y los datos (el modelo observado y el modelo estimado). Específicamente, se llama mínimos cuadrados promedio (LMS) cuando el número de datos medidos es 1 y se usa el método de descenso por gradiente para minimizar el residuo cuadrado. Se puede demostrar que LMS minimiza el residuo cuadrado esperado, con el mínimo de operaciones (por iteración), pero requiere un gran número de iteraciones para converger”.

No obstante el descubrimiento de los mínimos cuadrados, medio siglo antes, entre 1755 y 1757, R.J. Boscovitch, una de las más inusuales figuras de la ciencia en el siglo 18, articuló un interesante criterio para ajustar una línea a  $n \geq 2$  puntos en el plano (Eisenhart (1961, [2])), Si  $(\bar{x}, \bar{y})$  es el centroide de los  $n$  puntos  $(x_i, y_i)$ , la línea propuesta por Boscovitch escoge  $c$  para minimizar:

$$\sum_{i=1}^n |y_i - \bar{y} - c(x_i - \bar{x})|$$

Esta es la línea que minimiza el criterio LAD (Least Absolute Deviations) entre todas las líneas restringidas a pasar por la media de los datos. En 1760 Boscovitch propone un algoritmo geométrico para encontrar  $c$ , teniendo muchas dificultades computacionales, sin embargo Laplace ofrece una solución algebraica y elegante, la cual se puede parafrasear de la siguiente manera:

Sin pérdida de generalidad imagínese  $\bar{x} = \bar{y} = 0$  y se observa que la línea LAD a través del origen minimiza:

$$f(c) = \sum_{i=1}^n |y_i - cx_i| = \sum_{i=1}^n |r_i(c)| \quad (1)$$

se puede asumir que  $x_i \neq 0$  a causa de que  $f(c) = \sum_{i=1}^n |y_i| + \sum_{i=1}^n |y_i - cx_i|$ , donde la primera suma es para los  $x_i = 0$  y la segunda para los  $x_i \neq 0$ , entonces  $f$  es mínimo cuando la segunda suma lo es.

Ahora imagínese que  $\frac{y_i}{x_i} \leq \frac{y_{i+1}}{x_{i+1}}$  pueden ser ordenados en forma ascendente y si  $c$  se restringe al intervalo  $(\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$ ,  $f$  llega a ser:

$$f(c) = \sum_{i=1}^p |x_i| \left( c - \frac{y_i}{x_i} \right) - \sum_{i=p+1}^n |x_i| \left( c - \frac{y_i}{x_i} \right) \quad (2)$$

diferenciar la ecuación anterior permite obtener:

$$f'(c) = \sum_{i=1}^p |x_i| - \sum_{i=p+1}^n |x_i| \quad (3)$$

lo cual genera una función lineal a trozos continua con una derivada no decreciente.

Si  $f' = 0$  para un intervalo  $J = (\frac{y_p}{x_p}, \frac{y_{p+1}}{x_{p+1}})$  cualquier  $c$  en la clausura de  $J$  minimiza (1). Esto permite proponer el siguiente lema:

**Lema 1.1**  $f$  en (1) tiene un minimizador  $\hat{c} = \frac{y_i}{x_i}$  para algún  $i = 1, \dots, n$ , llámese  $i = p$ . De esta forma la línea LAD atraviesa el origen conteniendo  $(x_p, y_p)$ , así al menos un residuo en (1),  $r_p(\hat{c})$  es cero.

El anterior lema motiva un algoritmo:

**Algoritmo 1.1** (1) Calcule  $c_i = \frac{y_i}{x_i}$  con  $i = 1, \dots, n$

(2) Evaluar  $f(c_i)$  para todo  $i = 1, \dots, n$  y encontrar la función mínima, supóngase  $f(c_p)$  y su correspondiente minimizador  $c_p$ .

Según Laplace el optimal  $c$  es la menor razón en la que la derivada derecha de  $f$  es no negativa. de (3)

se puede ver que  $\min\left(j : \sum_{i=1}^j |x_i| \geq \sum_{i=j+1}^n |x_i|\right)$ . En otras palabras:

$$p = \min\left(j : \sum_{i=1}^j |x_i| \geq \sum_{i=j+1}^n |x_i|/2\right) \quad (4)$$

El valor  $c = \frac{y_p}{x_p}$  es la mediana ponderada de los  $\frac{y_i}{x_i}$  con pesos  $|x_i|$ , y la mediana ponderada puede ser obtenida en aproximadamente un tiempo proporcional a  $n \log(n)$ .

**Lema 1.2** La línea LAD a través del origen es la mediana ponderada de los  $\frac{y_i}{x_i}$  con pesos  $|x_i|$  ( $x_i, y_i$ ) son aquellos puntos para los cuales  $x_i \neq 0$ . la complejidad esperada no es mayor de  $O(n \log(n))$ .

## 1.1. Regresión LAD según Charnes, Cooper y Ferguson

Charnes, Cooper y Ferguson (1955, [33]), dos siglos después de que R.J. Boscovitch propusiera la recta de ajuste LAD, retomaron la propuesta y plantearon un interesante problema para la asignación salarial de los ejecutivos de una empresa, situación en la que es improbable utilizar mínimos cuadrados. No sólo realizaron la aplicación, sino que también dieron los elementos teóricos fundamentales e hicieron una representación geométrica del problema. A continuación se presenta la parte técnica propuesta por estos autores.

### 1.1.1. Representación Analítica del Problema

En este punto, lo mejor es comenzar señalando las necesidades de la empresa en una forma matemática precisa. Se requiere determinar una fórmula de la siguiente manera:

$$s = \sum_{i=1}^n a_i y_i$$

Donde:

$s$  = Salario

$y_i$  = Cantidad del factor  $i$  que tiene la persona cuyo salario va a ser determinado.

$a_i$  = Peso asignado al factor  $i$ .

Sean:

1.  $x_{ik}$  : La cantidad (conocida) del factor  $i$  (según la clasificación) requerida en el nivel de trabajo  $k = 1, 2, \dots, L$ .
2. Los trabajos son rankeados en forma descendente  $1, 2, \dots, L$ , así que el  $k$ -ésimo subíndice indica la posición del trabajo en la jerarquía. Se asume que:
3. Salarios alto  $s_M$  y bajo  $s_m$ , se han establecido por adelantado, cuando se puedan usar valores intermedios. Finalmente las restricciones de no negatividad en los pesos son introducidas:
4.  $a_i \geq 0$

Los requerimientos de consistencia pueden ser incluidos en una serie de desigualdades:

$$\begin{aligned}
 \sum_{i=1}^n a_i x_{i1} &\leq s_M \\
 \sum_{i=1}^n a_i x_{i2} &\leq \sum_{i=1}^n a_i x_{i1} \\
 &\dots\dots \\
 \sum_{i=1}^n a_i x_{iL} &\leq \sum_{i=1}^n a_i x_{i(L-1)} \\
 s_m &\leq \sum_{i=1}^n a_i x_{iL}
 \end{aligned}$$

Y  $a_i \geq 0$  donde los requerimientos de nivel son dados por  $s_M$  y  $s_m$ . Es deseable para la compañía que estos niveles al ser reunidos sean lo más *cerrados* posible. Esta condición puede ser interpretada matemáticamente en términos de escoger los  $a_i$ , de tal forma que la suma de las desviaciones absolutas de estos niveles conocidos sea mínima. Esto es los  $a_i$  son escogidos de tal forma que:

$$\sum_{k \in K} \left| \sum_{i=1}^n a_i x_{ik} - s_k \right| = \text{sea mínima} \quad (5)$$

Donde  $K$  es el conjunto de índices alto - bajo y niveles de salario intermedio especificados por adelantado. Cuando solamente índices alto y bajo son especificados el problema se reduce a:

$$\left| \sum_{i=1}^n a_i x_{i1} - s_M \right| + \left| \sum_{i=1}^n a_i x_{iL} - s_m \right| = \min \quad (6)$$

### 1.1.2. Transformación del Problema

Para reducir el problema no lineal a un problema de programación lineal, se introducen nuevas variables  $u_k$  y  $v_k$  por medio de las ecuaciones:

$$\sum_{i=1}^n a_i x_{ik} - s_k = u_k - v_k = w_k$$

Es claro que  $w_k$  no es restringida, puesto que representa la diferencia entre dos variables no negativas. Sin embargo en el cálculo de una solución por el método Simplex, alguna de las variables  $u_k$  ó  $v_k$ ,

para cualquier  $k$ , en todo estado de los cálculos deberá ser distinta de cero. Esto resulta del teorema propuesto en Dantzig, de que los vectores columna seleccionados en cualquier estado Simplex son linealmente independientes. Claramente los coeficientes del vector columna de  $u_k$  es el negativo de los coeficientes del vector columna de  $v_k$ . En consecuencia los dos son linealmente dependientes y no pueden aparecer simultáneamente en el conjunto de vectores seleccionados  $P_j$ 's, donde  $P_j$  representa el vector de coeficientes,  $j = 1, 2, \dots, n$ .

### 1.1.3. Representación Geométrica

Antes de proceder a dar un ejemplo numérico se puede examinar la geometría de transformar una variable sin restricciones a la diferencia de dos variables no negativas. Considérese, por ejemplo, el conjunto convexo con puntos extremos  $E_1, E_2, E_3, E_4$  en el espacio de dos dimensiones  $x - w$  como se ilustra en la Figura I. Si aquí  $w$  es una variable de restricción como se ha indicado por la existencia de  $E_3$  a la izquierda de  $0, w = 0$ . Se desea transformar ésta por  $w$  considerada como la diferencia entre dos variables no negativas,  $u, v > 0$ . El conjunto convexo correspondiente en el espacio tridimensional  $x - u - v$  (véase la Figura II) es sólido y tiene dos puntos extremos adicionales  $F_1$  y  $F_2$ . Descriptivamente este nuevo conjunto convexo puede ser considerado como formado plegando el plano  $x - w$  a lo largo del eje  $x$  mediante la rotación del plano medio,  $w \leq 0$ , 90 grados alrededor del eje  $x$ . Por lo tanto, después de hacer este plegado de un conjunto de puntos,  $U$ , en el plano  $u$  y un conjunto de puntos,  $V$ , en el plano  $v$  se obtienen. Conectando cada par de puntos, uno en  $U$  y uno en  $V$ , por medio de un segmento de línea recta en el espacio tridimensional  $x - u - v$  y adjuntando todos los puntos de estas líneas, el sólido convexo dibujado como se indica en la figura II se obtiene. Este es el conjunto correspondiente en el espacio  $x - u - v$  el cual reemplaza el conjunto original en el espacio  $x - w$  para los propósitos de análisis. Supóngase que el conjunto original en el plano  $x - w$  representa un conjunto de soluciones de un sistema de desigualdades lineales en  $x$  y  $w$  para que el mínimo de una funcional lineal en  $x$  y

$w$  fuera buscada. Evidentemente, la nueva solución establecida en  $x$ ,  $u$ , y  $v$  la funcional transformada en  $x$ ,  $u$ , y  $v$  son diferentes del conjunto de la solución original y la funcional, sin embargo, los valores mínimos de los dos funcionales será el mismo. Además, todo punto en el espacio  $x - u - v$  será asociado con una solución en el espacio  $x - w$ . En este sentido, los dos problemas son equivalentes. Es por esta razón por la que una extensión de este enfoque se utiliza aquí. Esto es, el problema de estimación de salarios de ejecutivos fue transformado de un problema no lineal a un problema lineal equivalente. La equivalencia lineal se deriva del hecho de que el interés se centró en las soluciones que minimizan el funcional.

CHARNES, COOPER AND FERGUSON

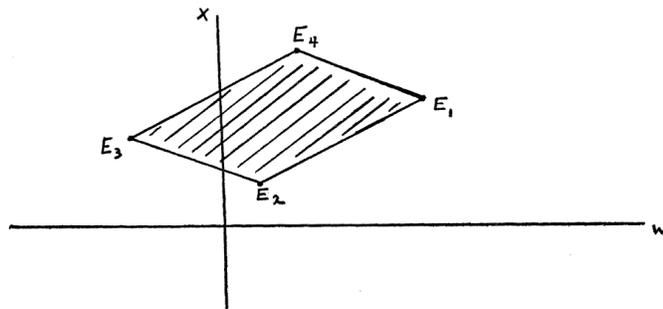


FIG. I

Teniendo esto en mente, el foco de atención ahora puede ser alterado desde la transformación de las restricciones a la transformación de la funcional. En la Figura III se dibujan dos curvas con propiedades del tipo mostrado por (1) y (2). Cualquiera de los dos puntos que se encuentran en una de estas curvas de rendimiento de la misma funcional valor,  $A_1$ , por ejemplo, para la curva primera y  $A_2$  para la segunda. Estas curvas son lineales. Sin embargo, en el problema transformado los puntos correspondientes a valores constantes de la transformada funcional por ejemplo,  $A_1$  y  $A_2$  de mentiras en planos indicados por  $f(x, u, v) = A_1$  y  $A_2$  en la Figura IV. Planos corresponden a funcionales lineales en el espacio

$x-u-v$  aunque las funciones originales en el espacio  $x-w$  son no lineales. El procedimiento de plegado descrito anteriormente transforma los segmentos de línea quebrada del  $x-w$  espacio bidimensional en intersección de los segmentos de línea en el espacio tridimensional  $x-u-v$  que generan planos por conexiones entre los pares adecuados de puntos como se ha descrito previamente. De esta manera, el problema inicial no lineal se convierte en un problema equivalente lineal en un espacio de mayor dimensión.

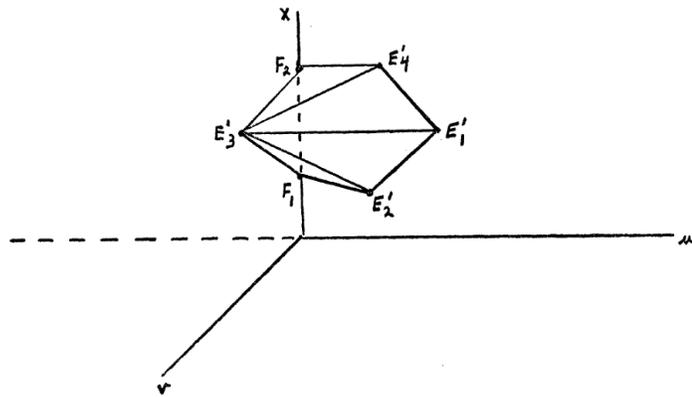


FIG. II

Fuente: Charnes, Cooper y Ferguson

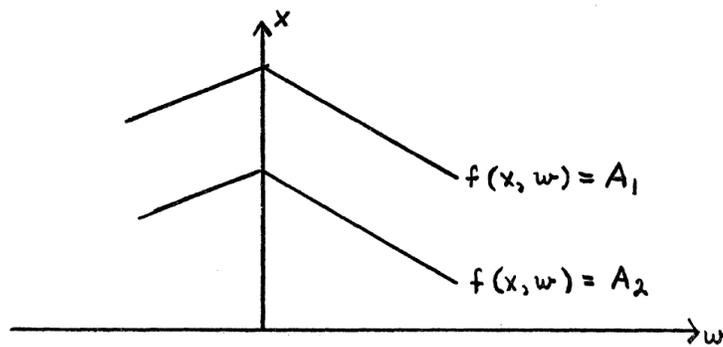


FIG. III

Fuente: Charnes, Cooper y Ferguson

Combinando las Figuras I, II, III y IV se exponen las propiedades esenciales de la geometría. Se desea encontrar el valor que cumple con las restricciones representadas en la Figura II y que minimizan  $\tilde{f}$ . Esencialmente, esto implica secuencias de localización de estos planos paralelos que cortan el conjunto convexo de la figura II hasta que se encuentra uno que se encuentra a la distancia mínima desde el origen.

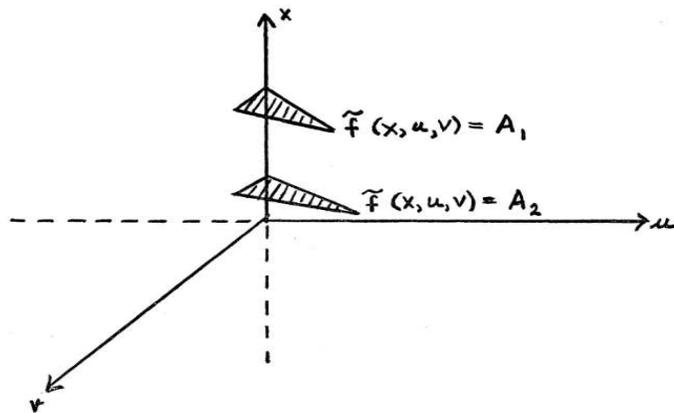


Fig. IV  
Fuente: Charnes, Cooper y Ferguson

## 2. Regresión Lineal Simple y Múltiple

### 2.1. Regresión Lineal Simple

En muchos análisis estadísticos se desea investigar cómo los cambios en una variable afectan a otra variable. Por ejemplo, altura y peso, ingreso y cantidad de alimento consumido, producto interno bruto y cambios en la moneda. En agricultura producción y fertilizante usado, etc. Supóngase que se tienen  $n$  parejas de observaciones  $(X_i, Y_i)$  con  $i = 1, \dots, n$ . Se pueden graficar estos puntos y se intenta ajustar una curva suave a través de estos. Tal diagrama es llamado scatter plot ó diagrama de puntos

(ver figura 5). Usualmente el tipo de curva ajustada es sugerida por evidencia empírica o argumentos teóricos. Cuando no se tienen bases es difícil decidir qué tipo de curva debería ser ajustada, en muchos casos y cuando los datos lo sugieren, la línea recta parece apropiada. En primer lugar se debe comenzar

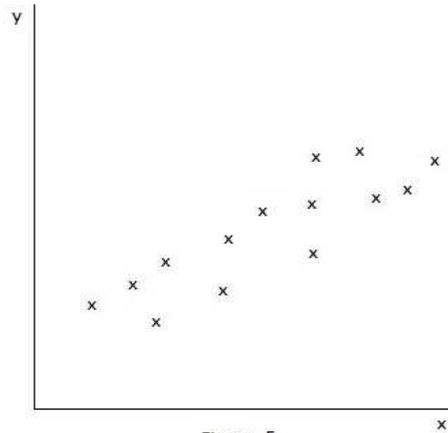


Figura 5

Fuente: Arthanari y Dodge

con un modelo, posteriormente se debe analizar cuidadosamente si éste es el adecuado. Supóngase que se asume una relación lineal entre las dos variables  $X$  e  $Y$ , esto significa que la relación funcional es de la forma:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (7)$$

el cual es conocido como un modelo de regresión lineal simple de  $Y$  en  $X$ .  $\beta_0$  y  $\beta_1$  son llamados parámetros y deben ser encontrados. La ecuación (5) significa que para un  $X_i$  dado, el correspondiente  $Y_i$  consiste de  $\beta_0 + \beta_1 X_i$  y un  $\epsilon_i$ , para el cual una observación debe caer dentro de la línea de regresión verdadera. Sobre las bases de la información disponible, se deben encontrar  $\beta_0$  y  $\beta_1$ . El término  $\epsilon$  es una variable aleatoria y se llama error. Se puede escribir.

$$Y_i - \beta_0 - \beta_1 X_i = \epsilon_i \quad (8)$$

el procedimiento para encontrar el valor de los parámetros  $\beta_0$  y  $\beta_1$  se llama estimación. existen diferentes métodos para obtener tales estimaciones.

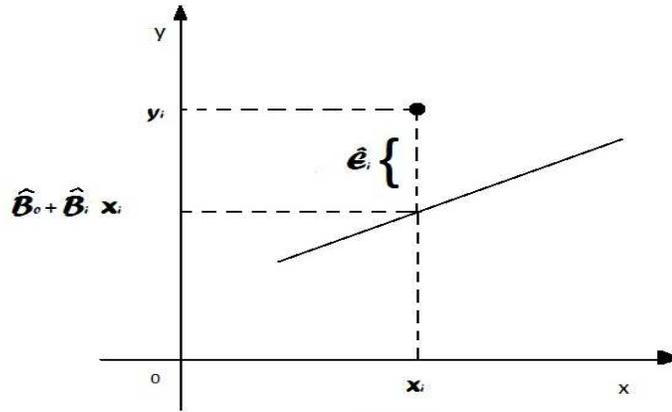


Figura 6  
Fuente: Dodge y Birkes

### 2.1.1. Mínimos Cuadrados

Este método está basado en escoger  $\beta_0$  y  $\beta_1$  como el mínimo de las sumas de cuadrados de las desviaciones verticales entre los datos y la recta ajustada. La suma de los cuadrados de las desviaciones de la línea es:

$$SSD = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (9)$$

luego de encontrar las estimaciones de  $\beta_0$  y  $\beta_1$ , se sustituyen en la ecuación anterior y la suma de cuadrados de las desviaciones es mínima. Diferenciando la ecuación (7) con respecto a  $\beta_0$  y  $\beta_1$  e igualando las derivadas parciales a cero, se tiene:

$$\begin{aligned} \frac{\partial SSD}{\partial \beta_0} &= -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \frac{\partial SSD}{\partial \beta_1} &= -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) X_i = 0 \end{aligned} \quad (10)$$

de donde se puede deducir:

$$\begin{aligned} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) &= 0 \\ \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) X_i &= 0 \end{aligned} \quad (11)$$

después de un poco de álgebra se llega a:

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ \beta_0 \sum_{i=1}^n X_i + \beta_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Y_i X_i \end{aligned} \quad (12)$$

las ecuaciones en (10) son llamadas *ecuaciones normales*. al desarrollar este sistema de ecuaciones se obtienen:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n Y_i X_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)/n}{\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i/n)^2} \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \end{aligned} \quad (13)$$

Y  $\hat{\beta}_1$  y  $\hat{\beta}_0$  son llamados los estimadores de mínimos cuadrados de  $\beta_1$  y  $\beta_0$  respectivamente. Finalmente se puede escribir la ecuación de regresión estimada:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad (14)$$

la cual es llamada *ecuación de predicción*.

## 2.2. Regresión Lineal Múltiple

Los resultados obtenidos en esta sección están basados en diferentes autores dentro de los cuales se encuentran Arthanari y Dodge (1981, [1]), Draper y Smith (1981, [13]), Seber (1977, [7]), Weisberg (1985, [8]), Dodge y Birkes (1993, [4]), Graybill (1961, [6]), Graybill (1976, [5]) y Bloomfield y Steiger (1983, [3]) entre otros. Considérese el modelo:

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \dots + \beta_p \mathbf{X}_p + \boldsymbol{\epsilon} \quad (15)$$

donde  $\mathbf{X}_1, \dots, \mathbf{X}_p$  son conocidas y los  $\beta_j$ 's son parámetros desconocidos a ser estimados y  $\boldsymbol{\epsilon}$  es el término de error. Si  $n$  valores de  $\mathbf{Y}$  son observados, se puede escribir  $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$ , y el modelo anterior escrito en forma matricial es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (16)$$

En la anterior expresión  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$  y  $\mathbf{X}_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ . También  $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots)$  y  $\boldsymbol{\epsilon}' = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ .

Tal como se propuso en la sección inmediatamente anterior, el método de mínimos cuadrados para estimar el vector de parámetros  $\boldsymbol{\beta}$ , consiste en minimizar  $\sum_{i=1}^n \epsilon_i^2$  con respecto a  $\boldsymbol{\beta}$  esto es minimizar la norma cuadrática  $\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$ . Esta forma cuadrática se puede escribir:

$$\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

realizando los productos en la anterior expresión, se tiene:

$$\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Diferenciando parcialmente esta forma cuadrática con respecto a  $\boldsymbol{\beta}$  e igualando la derivada parcial a cero, se tiene:

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0$$

la cual simplificada es:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \tag{17}$$

estas son las llamadas ecuaciones normales. Si  $\mathbf{X}$  es de rango  $k$ , en donde  $k = p + 1$ , entonces la matriz  $\mathbf{X}'\mathbf{X}$  es definida positiva y por tanto no singular, en consecuencia se tiene una solución única:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{18}$$

Entonces para cualquier  $\boldsymbol{\beta}$  se tiene:

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) &= [\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &\geq (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \end{aligned}$$

lo cual muestra claramente que el mínimo de  $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$  es  $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$  y se obtiene cuando  $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ .

### 2.2.1. Modelo estimado y residuos

Una vez obtenido el estimador de mínimos cuadrados para el modelo de regresión lineal múltiple, el modelo estimado es:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (19)$$

a la matriz  $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  se le conoce como matriz hat (sombrero), se nota  $\mathbf{H}$ . Esta a su vez satisface una serie de propiedades importantes, a saber:

1.  $\mathbf{H}$  es simétrica, esto es,  $\mathbf{H}' = \mathbf{H}$ .
2.  $\mathbf{H}$  es idempotente,  $\mathbf{H}'\mathbf{H} = \mathbf{H}\mathbf{H}' = \mathbf{H}$ .
3.  $\text{Traza}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = k$ , el número de parámetros del modelo.
4.  $(\mathbf{I} - \mathbf{H})$ , donde  $\mathbf{I}$  es la idéntica de orden  $n$  hereda las propiedades (1) y (2) de la matriz  $\mathbf{H}$ , no obstante su traza es  $n - k$ .
5. La matriz  $(\mathbf{I} - \mathbf{H})$  es ortogonal a la matriz  $\mathbf{X}$ , esto es,  $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$

Teniendo en cuenta lo anterior, el modelo estimado se puede escribir:

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y} \quad (20)$$

El vector de residuos es la diferencia entre el modelo observado y el modelo estimado:

$$\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \quad (21)$$

### 2.2.2. Algunos supuestos y resultados importantes

Según Graybill (1976, [5]) en el modelo matricial (14), si las variables independientes  $\mathbf{X}_j$  que componen la matriz  $\mathbf{X}$  son no aleatorias, además  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $\text{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$ , la relación especifica un *Modelo Lineal*

*General.* Si por su parte las variables independientes  $\mathbf{X}_j$  son aleatorias,  $\mathbf{Y}$  y  $\mathbf{X}$  tienen distribución conjunta y la estimación de los parámetros se realiza en la distribución condicional  $(\mathbf{Y}|\mathbf{X} = x)$ , se está frente a un *Modelo de Regresión* propiamente dicho. En este modelo hay dos importantes supuestos para el error:

1. **Caso 1.**  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $\text{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$ .
2. **Caso 2.**  $\boldsymbol{\epsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$ .

**Teorema 2.1** *Sea el modelo (14) con los supuestos del Caso 2, se obtienen los siguientes resultados:*

1.  $\hat{\boldsymbol{\beta}}$  es el estimador de máxima verosimilitud de  $\boldsymbol{\beta}$ .
2.  $\hat{\sigma}^2 = [1/(n - k)]\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$  es el estimador de máxima verosimilitud de  $\sigma^2$ .
3.  $\hat{\boldsymbol{\beta}} \sim \mathbf{N}_k(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1})\sigma^2$ .
4.  $(n - k)\hat{\sigma}^2/\sigma^2 \sim \chi_{(n-k)}^2$ .
5.  $\hat{\boldsymbol{\beta}}$  y  $\hat{\sigma}^2$  son independientes.
6.  $\hat{\boldsymbol{\beta}}$  y  $\hat{\sigma}^2$  son estadísticas suficientes para  $\boldsymbol{\beta}$  y  $\sigma^2$ .
7.  $\hat{\boldsymbol{\beta}}$  y  $\hat{\sigma}^2$  son estadísticas completas.
8.  $\mathbf{r} \sim \mathbf{N}_n(\mathbf{0}, (\mathbf{I} - \mathbf{H})\sigma^2)$ .
9.  $\mathbf{r}$  y  $\hat{\boldsymbol{\beta}}$  son independientes.

Es de aclarar que los resultados del teorema anterior son exactamente iguales a los resultados obtenidos mediante el método de mínimos cuadrados, para el modelo (14) bajo los supuestos Caso 2, en otras

palabras, bajo el supuesto de normalidad de los errores, la estimación de parámetros y resultados posteriores, bajo los métodos de mínimos cuadrados y máxima verosimilitud coinciden.

También el modelo de regresión lineal simple, es un caso particular del modelo de regresión lineal múltiple ó el modelo de regresión lineal múltiple es un caso general del modelo de regresión lineal simple.

### 2.2.3. Prueba de los coeficientes de regresión

Uno de los primeros problemas en regresión consiste en ensayar si las variables independientes o regresoras contienen cualquier información explicativa significativa. Se busca comparar el modelo *completo* conteniendo todas las variables explicativas, con el modelo reducido  $\mathbf{Y} = \beta_0 + \epsilon$  que no contiene variables explicativas.

**Prueba Estadística.** La idoneidad de un modelo puede ser juzgada por el tamaño de los residuos. Si  $SSR$  representa la suma de cuadrados de los residuos de un modelo. Se puede comparar el modelo completo con el modelo reducido, es decir comparando  $SSR_f$  con  $SSR_r$ . Específicamente la estadística para probar la hipótesis nula  $H_0 : \beta_1 = \beta_2 = \dots, \beta_p = 0$  es

$$F = \frac{SSR_r - SSR_f}{4 \hat{\sigma}^2} \quad (22)$$

donde  $\hat{\sigma}^2$  es una estimación de la varianza de los errores, estipulada en el numeral (2) del teorema anterior.

**Valor P.** La evidencia en contra de la hipótesis nula  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  es medida por la magnitud de la estadística  $F$ , la cual si es grande redonda en un valor pequeño de  $P$ , es decir si  $P$  es muy pequeña, por ejemplo 0,01 significa que el valor observado  $F$  es tan grande que resulta improbable que la hipótesis nula sea cierta, o es más sensato concluir que la hipótesis nula es falsa. Para calcular el valor  $P$  se hace necesario conocer la distribución de la variable  $F$  cuando la hipótesis nula es verdadera.

**Prueba de  $H_0 : \beta_{q+1} = \dots = \beta_p = 0$ .** Al considerar el modelo  $Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$ . La prueba de hipótesis anterior se puede extender, para comparar el modelo completo con un modelo reducido de cualquier subconjunto de variables explicativas. Este modelo reducido puede ser  $Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_q X_q + \epsilon$  con  $q < p$ . Se puede notar que comparar estos dos modelos equivale a probar la hipótesis nula  $H_0 : \beta_{q+1} = \dots = \beta_p = 0$ . Sea  $\hat{\sigma}^2$  una estimación insesgada de  $\sigma^2$  dada por:

$$\hat{\sigma}^2 = \frac{\text{SSR}_f}{n - p - 1} \quad (23)$$

la estadística para probar la hipótesis nula  $H_0 : \beta_{q+1} = \dots = \beta_p = 0$  es:

$$F = \frac{\text{SSR}_r - \text{SSR}_f}{(n - p)\hat{\sigma}^2} \quad (24)$$

**Prueba de  $\beta_j = 0$ .** El valor de  $\hat{\beta}_j$  debería indicar si  $\beta_j = 0$  o no. La estadística para probar  $H_0 : \beta_j = 0$  en el modelo de regresión lineal simple es:

$$|t| = \frac{|\hat{\beta}_j|}{\text{est.SD}(\hat{\beta}_j)}$$

esta ecuación representa una estadística razonable para probar la hipótesis  $H_0 : \beta_j = 0$ . Luego se hace indispensable cómo obtener  $\text{est.SD}(\hat{\beta}_j)$ . En el modelo de regresión lineal simple, se cumple que:

$$\text{Var}(\hat{\beta}_j) = \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$$

Similarmente en regresión lineal múltiple  $\text{Var}(\hat{\beta}_j)$  puede ser obtenida multiplicando  $\sigma^2$  por la  $(j+1)$ ésima entrada en  $(\mathbf{X}'\mathbf{X})^{-1}$  para cualquier  $j = 0, 1, 2, \dots, p$ . La estimación de  $\sigma^2$  está dada por  $\hat{\sigma}^2 = \frac{\text{SSR}_f}{(n-k)}$ , al sacar raíz cuadrada de esta expresión se obtiene  $\text{est.SD}(\hat{\beta}_j)$ . Cuando  $\beta_j = 0$  la estadística  $t \sim t_{(n-k)}$  asociada con  $\hat{\sigma}^2$ .

#### 2.2.4. Coeficiente de determinación.

El *coeficiente de determinación* es una medida de cuánto las variables independientes explican la variable respuesta. Se define por:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$R^2$  es la proporción de la variabilidad total en la variable respuesta, que ha sido explicada por las variables independientes. El coeficiente de determinación está estrechamente relacionado a la estadística  $F$  para probar la hipótesis  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p$ . En efecto  $R^2$  puede ser expresado como una función de  $F$ . En el caso de regresión simple,  $R^2$  es el cuadrado de la correlación muestral entre  $X$  y  $Y$ .

### 2.3. Influencia Local

Tal como lo señala Paula (2004, [15]), uno de los métodos más modernos de diagnóstico fué propuesto por Cook (1987, [16]). La idea básica consiste en estudiar el comportamiento de alguna medida particular de influencia según pequeñas perturbaciones (influencia local), en los datos de un modelo. Esto es, verificar la existencia de puntos que sobre modificaciones modestas en el modelo, causan variaciones desproporcionales en los resultados. Supóngase que el logaritmo de la función de verosimilitud para el parámetro  $\beta$  se exprese de la siguiente forma:

$$L_\delta(\beta; \mathbf{y}) = \sum_{j=1}^n \delta_j L(\beta; \mathbf{y}_j) \quad (25)$$

en donde  $L(\beta; \mathbf{y}_j)$  es el logaritmo de la función de verosimilitud correspondiente a la  $j$ -ésima observación y  $\delta_j$  es un tipo de perturbación, definida tal que  $0 \leq \delta_j \leq 1$ . Cuando  $\delta_j = 1, \forall j$  significa que no hay perturbación en el modelo y cuando  $\delta_j = 0$  significa que la  $j$ -ésima observación fué excluida. Una estimación de mínimos cuadrados, bajo la estructura (23) es dada por:

$$\hat{\beta}_\delta = (\mathbf{X}^T \Delta \mathbf{X})^{-1} \mathbf{X}^T \Delta \mathbf{y}$$

donde  $\mathbf{\Delta} = \text{diag}\{\delta_1, \delta_2, \dots, \delta_n\}$ . En particular cuando apenas la  $i$ -ésima observación es perturbada, esto es, cuando  $\delta_i = \delta$  y  $\delta_j = 1$  para  $j \neq i$  se muestra que

$$\hat{\beta}_\delta = \hat{\beta} - \frac{(1-\delta)r_i}{\{1-(1-\delta)h_{ii}\}}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i. \quad (26)$$

Para  $\delta = 0$ , o sea que el  $i$ -ésimo punto es excluido (2) queda expresada en forma simplificada

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{r_i}{(1-h_{ii})}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i \quad (27)$$

que es bastante conocida en Regresión Lineal normal (Cook y Weisberg, [17]). La medida de influencia más conocida está en la región de confianza para el parámetro  $\beta$ ,

$$(\hat{\beta} - \beta)(\mathbf{X}^T\mathbf{X})(\hat{\beta} - \beta) \leq \mathbf{ps}^2\mathbf{F}_{p,(n-p)}(\alpha)$$

que para el caso  $p = 2$  es un elipsoide de  $\mathbf{R}^2$  centrado en  $\hat{\beta}$ . Tal medida, conocida como distancia de Cook es definida por:

$$D_\delta = \frac{(\hat{\beta} - \hat{\beta}_\delta)(\mathbf{X}^T\mathbf{X})(\hat{\beta} - \hat{\beta}_\delta)}{ps^2} \quad (28)$$

y mide cuánta perturbación  $\delta = (\delta_1, \delta_2, \dots, \delta_n)^T$  aleja  $\hat{\beta}_\delta$  de  $\hat{\beta}$ , según la métrica  $(\mathbf{X}^T\mathbf{X})$ . Por ejemplo, si  $D_\delta > F_{p,(n-p)}(1-\alpha)$ , significa que una perturbación está distorsionando el contorno de la elipse a un nivel de significancia menor que  $\alpha$ . En particular, cuando el  $i$ -ésimo punto es excluido, la distancia de Cook queda expresada en la forma:

$$\begin{aligned} D_i &= \frac{(\hat{\beta} - \hat{\beta}_i)(\mathbf{X}^T\mathbf{X})(\hat{\beta} - \hat{\beta}_i)}{ps^2} \\ &= \left\{ \frac{r_i}{s(1-h_{ii})^{\frac{1}{2}}} \right\}^2 \frac{h_{ii}}{(1-h_{ii})} \frac{1}{p} \\ &= t_i^2 \frac{h_{ii}}{(1-h_{ii})} \frac{1}{p} \end{aligned}$$

Por tanto  $D_i$  será grande cuando el  $i$ -ésimo punto es aberrante ( $t_i$  grande) o cuando  $h_{ii}$  es próximo a 1. La distancia  $D_i$  podrá no ser adecuada cuando  $r_i$  sea grande y  $h_{ii}$  pequeño. En este caso,  $s^2$  podrá quedar inflado y no se tendría ninguna compensación por parte de  $h_{ii}$ ,  $D_i$  puede quedar pequeño. Una

medida supuestamente más apropiada fué propuesta por Belsley, Kuh y Welsch (1980):

$$\begin{aligned} \text{DFFITs}_i &= \frac{|r_i|}{s_{(i)}(1-h_{ii})^{\frac{1}{2}}} \left\{ \frac{h_{ii}}{(1-h_{ii})} \right\}^{\frac{1}{2}} \\ &= |t_i^*| \left\{ \frac{h_{ii}}{(1-h_{ii})} \right\}^{\frac{1}{2}} \end{aligned}$$

Como el valor esperado de  $h_{ii}$  es  $\frac{p}{n}$  parece razonable dar más atención a aquellos puntos tales que

$$\text{DFFITs}_i \geq 2 \left\{ \frac{p}{n-p} \right\}^{\frac{1}{2}}$$

Aparentemente  $D_i$  y  $\text{DFFITs}_i$  serían medidas de influencia competitivas, una vez que  $\text{DFFITs}_i$  parece ser más adecuada para validar la influencia en las estimaciones de los coeficientes de un punto aberrante con  $h_{ii}$  pequeño. En tanto, como muestran Cook, Peña y Weisberg (1988, [17])  $D_i$  y  $\text{DFFITs}_i$  miden cosas diferentes. Las dos pueden ser expresadas a partir de una medida más general de influencia denominada *alejamiento de verosimilitud ó likelihood displacement* propuesta por Cook y Weisberg (1982). La medida  $D_i$  mide esencialmente la influencia de las observaciones en los parámetros de localización, en tanto  $\text{DFFITs}_i$  tiene el propósito de medir la influencia de las observaciones en los parámetros de localización y escala. Como es poco probable que un punto con  $r_i$  alto y  $h_{ii}$  pequeño sea influyente en las estimaciones de los coeficientes, el uso de  $D_i$  no compromete una detección de observaciones influyentes. Cook, Peña y Weisberg observan también que los  $\text{DFFITs}_i$  no es una medida completa de influencia en los parámetros de localización y escala simultáneamente, pudiendo fallar en algunas situaciones. Atkinson (1985) propone otra medida de influencia que es un perfeccionamiento de los  $\text{DFFITs}_i$ ,

$$C_i = \left\{ \frac{n-p}{p} \frac{h_{ii}}{(1-h_{ii})} \right\}^{\frac{1}{2}} |t_i^*|$$

Aquí cuando el experimento es balanceado, esto es, todos los  $h_{ii}$  son iguales, se tiene  $C_i = |t_i^*|$ . Una ventaja de  $C_i$  es que la misma puede ser utilizada en gráficos de probabilidad normal. A fin de introducirnos en la metodología, supóngase que el logaritmo de verosimilitud dado como en (1) es una medida de Cook como en (4). El objetivo aquí es estudiar los cambios producidos en  $D_\delta$  cuando  $\delta_i \rightarrow 1$ ,  $\forall i$ .

Expandiendo  $D_\delta$  en series de Taylor de segundo orden alrededor de  $\delta_0 = 1$ , se obtiene:

$$\begin{aligned} D_\delta &\cong D_{\delta_0} + (\delta_0 - \delta)^T D_{\delta_0}' + \frac{1}{2}(\delta_0 - \delta)^T D_{\delta_0}''(\delta_0 - \delta) \\ &\cong \frac{1}{2}(\delta_0 - \delta)^T D_{\delta_0}''(\delta_0 - \delta) \end{aligned}$$

se muestra que para el proceso de perturbación dado en (1), que

$$D_{\delta_0}'' = \text{diag}(\mathbf{r})\mathbf{H}\text{diag}(\mathbf{r})$$

en donde  $\text{diag}(\mathbf{r}) = \text{diag}\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ . Una sugerencia de Cook, que usa conceptos de geometría diferencial, es estudiar la mayor variación de  $D_\delta$  alrededor de  $\delta_0$ . Eso es equivalente a maximizar la forma cuadrática  $d^T \mathbf{A}d$ , en donde  $d = \delta_0 - \delta$  y  $d^T d = 1$  además  $\mathbf{A} = \text{diag}(\mathbf{r})\mathbf{H}\text{diag}(\mathbf{r})$ . Note que el máximo de  $d^T \mathbf{A}d$  corresponde al mayor valor propio de la matriz  $\mathbf{A}$ , el cual se denota como  $\lambda_{max}$ . Los valores de  $\mathbf{d}_{max}$  contienen una influencia local de las observaciones en esa dirección particular. Luego un gráfico de  $|\mathbf{d}_{max}|$  contra el orden de las observaciones puede revelar aquellos puntos con mayor influencia en la vecindad de  $D_{\delta_0}$ . Tales puntos pueden ser responsables por cambios sustanciales en las estimaciones de los parámetros, sobre pequeñas perturbaciones en el modelo. Sería por tanto prudente ojear con más cuidado esos puntos, a fin de entender mejor la influencia de los mismos y consecuentemente intentar proponer una forma segura de usar el modelo ajustado. Se han publicado numeroso artículos acerca de influencia local en los últimos años. Por ejemplo, en la clase de errores normales, Lawrence (1988, [18]) investiga una aplicación de influencia local en Modelos Lineales con parámetros en la transformación de respuesta, Beckman, Nachtsheim y Cook (1987) presentan estudios de influencia en modelos de análisis de varianza con efectos mixtos, Tsai y Wu (1992, [19]) investiga la influencia local en modelos autorregresivos de primer orden y modelos heterocedásticos. Por su parte Paula (1993) aplica influencia local en modelos lineales con restricciones en los parámetros, en forma de desigualdades lineales. Saliendo de la clase de errores normales, se tiene por ejemplo, el trabajo de Petit y Bin Daud (1989, [21]) que investiga influencia local en Modelos de Cox con riesgos proporcionales. Escobar y Meeker (1992, [22]) adaptan influencia local en una clase paramétrica de modelos para análisis de sobrevivencia. O'Hara

Hines, Lawless y Carter (1992) y Pan, Fang y Von Rosen (1997) aplican métodos de influencia local en regresión multivariada. Mas recientemente Galea, Paula y Bolfarine (1997, [24]), Liu (2000), y Galea, Paula y Uribe-Opazo (2003, [25]) presentan estudios de influencia local en modelos elipticos lineales, en cuanto Kwan y Fung (1998, [26]) aplican una metodología en análisis factorial, Gu y Fung (1998, [27]) en análisis de correlación canónica, Paula (1996) en modelos propios de dispersión y Ortega, Bolfarine y Paula (2003) en modelos log-gamma generalizados con datos censurados. Svetliza y Paula (2001, 2003, [29],[30]) discuten influencia local en modelos con respuesta binomial negativa. Estos últimos modelos han sido usados para corregir problemas de sobredispersión, frecuentemente encontrados en modelos con respuesta Poisson. Una discusión bastante interesante con respecto al uso de influencia local es presentado por Fung y Kwan (1997, [31]). Los autores muestran que el desplazamiento del logaritmo de verosimilitud (*likelihood displacement*) es una medida de influencia invariante con cambios de escala en los datos, hecho que no ocurre con otras medidas de influencia propuestas.

### 3. Métodos de Regresión Alternativa

Los siguientes métodos constituyen una alternativa, para los modelos de regresión lineal simple y múltiple vistos en la sección anterior.

#### 3.1. Regresión MINMAD

Charnes, Cooper y Ferguson (1955, [33]) en un interesante artículo proponen la necesidad de utilizar la minimización de la norma  $L_1$ , para ajustar un modelo simple, en una situación donde no es posible utilizar mínimos cuadrados, con el fin de determinar el porcentaje de distintos factores aplicados, para determinar el salario de los ejecutivos en una empresa del sector industrial. Es decir, dos siglos después

de que R.J. Boscovitch propusiera la recta de ajuste LAD (Least Absolute Deviations), estos autores deciden retomar esta metodología. En este artículo los autores proponen que se realice la estimación, utilizando el *Método Simplex*. Tres años más tarde Karst (1958, [34]), propone una metodología estadística para encontrar la solución a este problema. Se trata de minimizar la media de los valores absolutos de las desviaciones de las observaciones con respecto a la línea de regresión, de ahí su nombre MINMAD (Minimizing Mean of Absolute Deviations), en un modelo de regresión lineal simple, a diferencia de mínimos cuadrados que minimiza la norma cuadrática ó euclideana. Es decir, ahora se trata de estimar  $\beta_0$  y  $\beta_1$  minimizando:

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \beta_0 - \beta_1 X_i| \quad (29)$$

lo cual equivale a minimizar:

$$\sum_{i=1}^n |Y_i - \beta_0 - \beta_1 X_i| \quad (30)$$

Para simplificar el análisis, primeramente se imponen restricciones a  $\beta_0$  y  $\beta_1$  para que satisfagan la condición  $Y_0 = \beta_0 + \beta_1 X_0$  para un par dado  $(X_0, Y_0)$ . Además dado el par  $(X_0, Y_0)$  es posible transformar los datos:

$$x_i = X_i - X_0$$

$$y_i = Y_i - Y_0$$

Luego al reemplazar los datos transformados en (28), se tiene:

$$\begin{aligned} \sum_{i=1}^n |Y_i - \beta_0 - \beta_1 X_i| &= \sum_{i=1}^n |y_i + Y_0 - \beta_0 - \beta_1(x_i + X_0)| \\ \sum_{i=1}^n |y_i + \beta_0 + \beta_1 X_0 - \beta_0 - \beta_1 x_i - \beta_1 X_0| &= \sum_{i=1}^n |y_i - \beta_1 x_i| \end{aligned}$$

Así el problema se reduce ahora a calcular un  $\beta$  que minimice la expresión

$$\sum_{i=1}^n |y_i - \beta x_i| \quad (31)$$

Arthanari y Dodge (1981, [1]) presentan un sencillo ejemplo con el fin de visualizar la interpretación geométrica del problema. Para ello utilizan los siguientes tres datos:

$i$	$x_i$	$y_i$
1	1	3
2	1	1
3	2	4

Se puede visualizar en la figura 7 que  $|y_i - \beta x_i|$  para cualquier  $i$  son dos líneas rectas con un mínimo en  $(\frac{y_i}{x_i}, 0)$  y pendientes  $-|x_i|$  y  $|x_i|$

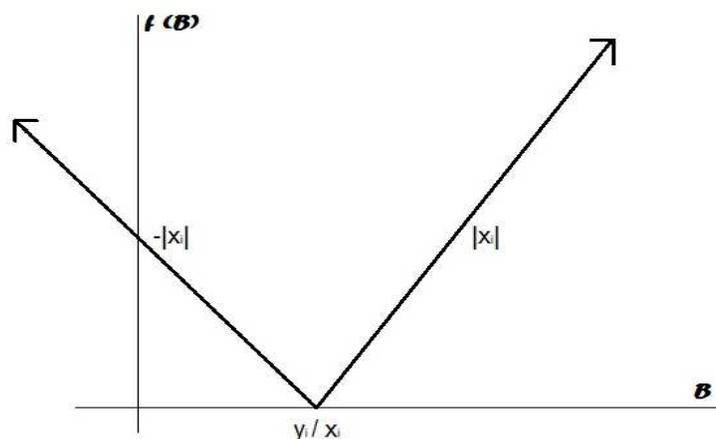


Figura 7  
Fuente: Arthanari y Dodge

La gráfica de  $|y_i - \beta x_i|$  para  $i = 1, 2, 3$  y  $\sum |y_i - \beta x_i|$  es una función lineal a trozos y convexa (ver figura 8).

**Proposición 3.1** La función  $f(\beta) = \sum_{i=1}^n |y_i - \beta x_i|$ , para los valores dados de  $(x_i, y_i)$ , con  $i = 1, \dots, n$  es una función lineal a trozos y convexa.

**Demostración:**

Se debe demostrar que para  $\beta' < \beta''$  y  $0 \leq \lambda \leq 1$ , y  $\beta = \lambda\beta' + (1 - \lambda)\beta''$ :

$$f(\beta) \leq \lambda f(\beta') + (1 - \lambda)f(\beta'')$$

En efecto, se sabe que  $f_i(\beta) = |y_i - \beta x_i|$

$$\begin{aligned}
 f_i(\beta) &= f_i(\lambda\beta' + (1-\lambda)\beta'') = |y_i - \lambda\beta'x_i - (1-\lambda)\beta''x_i| \\
 &= |\lambda(y_i - \beta'x_i) + (1-\lambda)(y_i - \beta''x_i)| \\
 &\leq \lambda|y_i - \beta'x_i| + (1-\lambda)|y_i - \beta''x_i| \\
 &= \lambda f_i(\beta') + (1-\lambda)f_i(\beta'')
 \end{aligned}$$

Sabiendo que la suma de dos funciones convexas es convexa, se tiene que  $f(\beta)$  es convexa. Además sabiendo que  $f_i(\beta)$  es lineal a trozos, la suma de funciones lineales a trozos es lineal a trozos, en consecuencia el resultado.

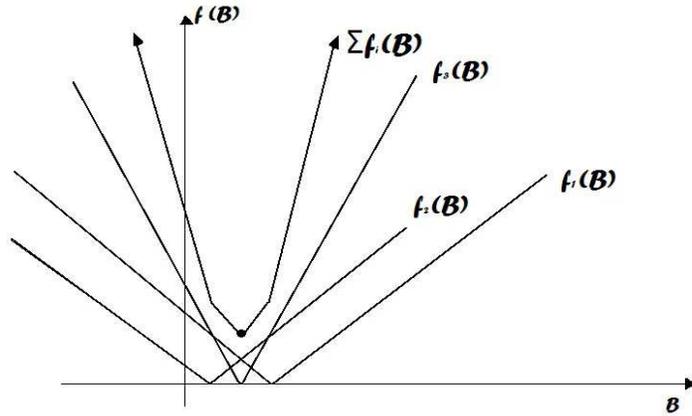


Figura 8  
Fuente: Arthanari y Dodge

**Proposición 3.2** La función  $f(\beta)$  tiene las siguientes propiedades:

1. La pendiente del segmento del extremo de la izquierda es  $-\sum_{i=1}^n |x_i|$ , y de la derecha es  $\sum_{i=1}^n |x_i|$ .
  2. Los vértices de la función poligonal  $f(\beta)$  son de la forma  $\left(\frac{y_i}{x_i}\right)$ , donde  $\frac{y_i}{x_i}$  es el mínimo de  $f_i(\beta)$ .
- Si  $(i_1, \dots, i_n)$  es un conjunto de índices tales que  $\frac{y_{i_1}}{x_{i_1}} \leq \dots \leq \frac{y_{i_n}}{x_{i_n}}$ , entonces la pendiente de  $f(\beta)$  se incrementa en  $2|x_{i_k}|$  al pasar por  $\beta_k = \frac{y_{i_k}}{x_{i_k}}$ .

Estos resultados proporcionan un método para calcular el mínimo de  $f(\beta)$ . El mínimo se alcanza en un  $\beta_r$  tal que:

$$\begin{aligned} -\sum_{i=1}^n |x_i| + 2 \sum_{k=1}^{r-1} |x_{i_k}| &< 0 \\ -\sum_{i=1}^n |x_i| + 2 \sum_{k=1}^r |x_{i_k}| &\geq 0 \end{aligned} \quad (32)$$

Si  $-\sum_{i=1}^n |x_i| + 2 \sum_{k=1}^r |x_{i_k}| = 0$  entonces  $\beta_{(r)} \leq \beta \leq \beta_{(r+1)}$  son optimales. Entonces se puede escoger  $\beta_{(r)}$  o  $\beta_{(r+1)}$  con igual probabilidad. Los parámetros estimados en este caso son:

$$\begin{aligned} \tilde{\beta}_1 &= \frac{y_r}{x_r} \\ \tilde{\beta}_0 &= Y_0 - \left(\frac{y_r}{x_r}\right) X_0 \end{aligned} \quad (33)$$

si se escoge  $\beta_{(r)}$  como solución. La teoría anterior permite generar un algoritmo:

**Algoritmo 3.1** *Dados los puntos muestrales  $(X_i, Y_i)$  para  $i = 1, \dots, n$  de  $(X, Y)$  calcular:*

1.  $(X_0, Y_0) = (\bar{X}, \bar{Y})$ .
2. Calcular las variables transformadas  $x_i = X_i - X_0$  y  $y_i = Y_i - Y_0$ .
3. Calcular los mínimos de las funciones  $f_i$  como  $\frac{y_i}{x_i}$ .
4. Asignar rangos a los mínimos anteriores en forma ascendente, dando el valor 1 al más pequeño y  $n$  al más grande.
5. Calcular  $-\sum_{i=1}^n |x_i|$  y comience a sumar  $2|x_i|$ , siguiendo el orden de los rangos.
6. Cuando la suma anterior pase de negativa a positiva escójase el mínimo  $\frac{y_i}{x_i}$  en ese paso como estimador de  $\beta_1$ . Si esto ocurre en el paso  $r$ , entonces  $\tilde{\beta}_1 = \frac{y_r}{x_r}$  y calcúlese  $\tilde{\beta}_0$ .

Dodge y Birkes (1993, [4]) proponen un algoritmo, que aunque parezca eficiente, es demasiado lento y necesita muchas tablas para ser comparadas, lo cual a su vez implica muchas iteraciones.

**Algoritmo 3.2** El objetivo de este algoritmo es encontrar la mejor línea de ajuste entre todas las líneas. Dado un punto  $(X_0, Y_0)$  de los datos, para cada punto  $(X_i, Y_i)$  calcular:

1. La pendiente  $\frac{(Y_i - Y_0)}{(X_i - X_0)}$  de la línea pasando a través de los dos puntos  $(X_0, Y_0)$  y  $(X_i, Y_i)$ . Si  $X_i = X_0$  para algún  $i$  tales puntos pueden ser ignorados.
2. Reindexar los puntos tal que:  $\frac{(Y_1 - Y_0)}{(X_1 - X_0)} \leq \frac{(Y_2 - Y_0)}{(X_2 - X_0)} \leq \dots \leq \frac{(Y_n - Y_0)}{(X_n - X_0)}$ . Sea  $T = |X_i - X_0|$ .
3. Encontrar el índice  $k$  que satisfaga las condiciones:

$$|X_i - X_0| + \dots + |X_{k-1} - X_0| < \frac{1}{2}T$$

$$|X_i - X_0| + \dots + |X_{k-1} - X_0| + |X_k - X_0| > \frac{1}{2}T$$

4. La mejor línea pasando a través de  $(X_0, Y_0)$  es la línea  $\hat{Y} = \beta_0^* + \beta_1^* X$ , donde:

$$\beta_1^* = \frac{Y_k - Y_0}{X_k - X_0}$$

$$\beta_0^* = Y_0 - \beta_1^* X_0$$

### 3.2. Regresión MINMAXAD

Este método consiste en estimar los parámetros  $\beta_0$  y  $\beta_1$  minimizando la máxima de las desviaciones absolutas (Minimizing Maximum of Absolute Deviations). Bajo este criterio, la función objetivo es:

$$\text{Min}_{(\beta_0, \beta_1)} \left[ \text{Máx} |Y_i - \beta_0 - \beta_1 X_i| \right] \quad (34)$$

De momento se discute el problema sin el término  $\beta_0$ . Si  $f_i(\beta) = |Y_i - \beta X_i|$  y  $g(\beta) = \text{máx}_i f_i(\beta)$ , se puede demostrar que  $g(\beta)$  es una función lineal a trozos convexa. Los vértices de  $g(\beta)$  no son necesariamente los puntos  $\frac{Y_i}{X_i}$  como ocurría antes. Los vértices de  $g(\beta)$  son las coordenadas  $\beta$  de las intersecciones de las líneas  $g(\beta) = Y_i + X_i \beta$  o  $g(\beta) = -(Y_i + X_i \beta)$  con las líneas  $g(\beta) = Y_j + X_j \beta$  o  $g(\beta) = -(Y_j + X_j \beta)$  para cada  $i \neq j$ . La interpretación geométrica de este método es ampliamente

discutida por Wagner (1959, [36]), quien propone un algoritmo para variables acotadas y por Stiefel (1960, [35]), quien reduce el problema al método Simplex. La figura 9 ilustra la idea geométrica, se pretende encontrar el mínimo de la función  $g(\beta)$ .

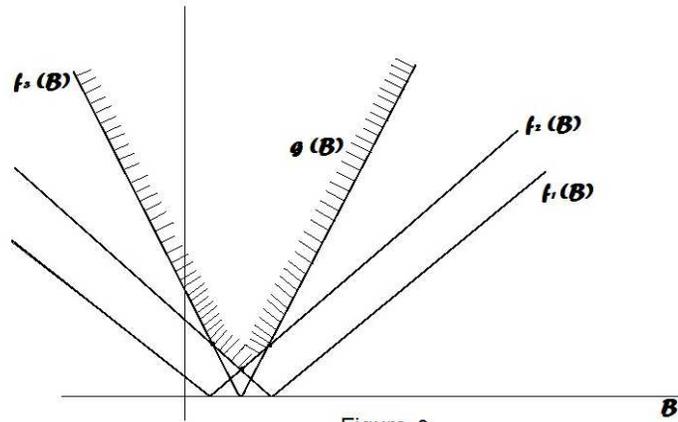


Figura 9  
Fuente: Arthanari y Dodge

### 3.3. Regresión MINSADBED

Este tipo de regresión considera la estimación de los parámetros  $\beta_0$  y  $\beta_1$  minimizando la suma de las diferencias absolutas entre desviaciones (Minimizing Sum of Absolute Difference Between Deviations), esto es:

$$\text{Minimizar } \sum_{i < j} |d_i - d_j|$$

es de aclarar que las distancias  $d_i$  y  $d_j$  representan las diferencias entre el modelo observado  $Y_i$  y el modelo a ser estimado  $\beta_0 + \beta_1 X_i$  en el primer caso y  $\beta_0 + \beta_1 X_j$  en el segundo. Así:

$$\begin{aligned} \sum_{i < j} |d_i - d_j| &= \sum_{i < j} |(Y_i - \beta_0 - \beta_1 X_i) - (Y_j - \beta_0 - \beta_1 X_j)| \\ &= \sum_{i < j} |(Y_i - Y_j) - \beta_1 (X_i - X_j)| \end{aligned}$$

Haciendo  $Y_{ij} = Y_i - Y_j$  y  $X_{ij} = X_i - X_j$ , se tiene:

$$\sum_{i < j} |d_i - d_j| = \sum_{i < j} |Y_{ij} - \beta_1 X_{ij}|$$

este problema se reduce al problema MINMAD, excepto que se deben realizar las diferencias entre  $Y_i$  y  $Y_j$  y entre  $X_i$  y  $X_j$  las cuales suman un total de  $\frac{n(n-1)}{2}$ . El parámetro  $\beta_0$  se calcula haciendo:

$$\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1$$

y otra forma alternativa de estimarlo es:

$$\hat{\beta}_0 = \text{Mediana}_{i < j} \frac{1}{2}(Y_i + Y_j)$$

### 3.4. Regresión MINSADBAD

Este método consiste en estimar los parámetros  $\beta_0$  y  $\beta_1$  minimizando la suma de las diferencias absolutas entre desviaciones absolutas (Minimizing Sum of Absolute Differences Between Absolute Deviations). Es decir:

$$\text{Minimizar}_{\beta_0, \beta_1} \sum_{i < j} ||d_i| - |d_j||$$

donde  $d_i = Y_i - (\beta_0 + \beta_1 X_i)$ . Este método resulta ser más apropiado explicarlo, cuando se considera el modelo de regresión lineal múltiple.

### 3.5. Prueba de $H_0 : \beta = 0$ en Regresión MINMAD

Primero se calculan las estimaciones MINMAD  $\tilde{\beta}_0$  y  $\tilde{\beta}_1$  y los residuos  $\tilde{e}_i = y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i)$ . Sea  $m = n - 2$ , el número de residuos distintos de cero. Se ordenan los residuos en orden ascendente, de tal forma que  $\tilde{e}_{(1)}$  sea el menor,  $\tilde{e}_{(2)}$  el siguiente y  $\tilde{e}_{(m)}$  el mayor. Sea  $k_1$  el entero más cercano a  $(m + 1)/2 - \sqrt{m}$  y  $k_2$  el entero más cercano a  $(m + 1)/2 + \sqrt{m}$ . Se calcula:

$$\tilde{\tau} = \frac{\sqrt{m}[\tilde{e}_{(k_2)} - \tilde{e}_{(k_1)}]}{4} \quad (35)$$

Luego se calcula la desviación estándar de  $\tilde{\beta}$ :

$$\text{est.SD}(\tilde{\beta}) = \frac{\tilde{\tau}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (36)$$

El estadístico de prueba es:

$$|t| = \frac{|\tilde{\beta}|}{\text{est.SD}} \quad (37)$$

Esta prueba se encuentra sin demostración en Dodge y Birkes (1993, [4], p.p. 63), no obstante en el apéndice D, se presenta una justificación de la prueba, basada en el *Teorema del Límite Central*.

El valor  $P$  de la prueba es calculado como la probabilidad  $P[|T| \geq |t|]$ , donde  $T$  representa una variable aleatoria con distribución  $t$ -Student de  $n - 2$  grados de libertad.

**Importancia del Parámetro  $\tau$ .** La cantidad  $\tilde{\tau}$  es una estimación del parámetro  $\tau$  en regresión MINMAD el cual juega un papel importante análogo al de  $\sigma$  en mínimos cuadrados. La desviación estándar de  $\hat{\beta}$  es  $\sigma/\sqrt{\sum(x_i - \bar{x})}$ , mientras que la desviación estándar de  $\tilde{\beta}$  es  $\tau/\sqrt{\sum(x_i - \bar{x})}$ .  $\tau$  es grande cuando  $\sigma$  también lo es, pero la razón  $\tau/\sigma$  depende de la forma de la distribución poblacional de los errores. Si los errores tienen distribución normal  $\tau/\sigma = 1,253 > 1$ , y así, para grandes muestras, las estimaciones de regresión MINMAD son menos seguras que las estimaciones de la regresión por mínimos cuadrados. Si los errores tienen distribución de Laplace, entonces  $\tau/\sigma = 0,707 < 1$ .

### 3.6. MINMAD para Regresión Lineal Múltiple

Dado el modelo  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ . La situación ahora es presentar el siguiente problema, para la estimación del vector de parámetros  $\boldsymbol{\beta}$ :

Minimizar	$\sum_{i=1}^n  d_i $	(38)
sujeto a	$\mathbf{X}\boldsymbol{\beta} + \mathbf{d} = \mathbf{Y}$	
	$\mathbf{d}, \boldsymbol{\beta}$ cualesquiera.	

Haciendo  $|d_i| = d_{1i} + d_{2i}$  con  $d_{1i} d_{2i} \geq 0$ , y  $d_i = d_{1i} - d_{2i}$ , el problema se puede reformular como:

Minimizar	$\sum d_{1i} + \sum d_{2i}$
sujeto a	$\mathbf{X}\boldsymbol{\beta} + \mathbf{d}_1 - \mathbf{d}_2 = \mathbf{Y}$
	$\boldsymbol{\beta}$ no restringido en signo.
	$\mathbf{d}_1 \mathbf{d}_2 \geq 0$

(39)

Se define:

$\mathbf{A} = (\mathbf{X}, \mathbf{I}, -\mathbf{I})$	matriz de orden $n \times (k + 2n)$
$\mathbf{W}' = (\boldsymbol{\beta}', \mathbf{d}'_1, \mathbf{d}'_2)$	vector de orden $k + 2n$
$\mathbf{a}_j$	la $j$ -ésima columna de $\mathbf{A}$
$\mathbf{C}' = (\mathbf{0}'_{(1 \times k)}, \mathbf{e}'_{(1 \times n)}, \mathbf{e}'_{(1 \times n)})$	los coeficientes de la función objetivo

(40)

De esta manera el problema se transforma en el siguiente:

Minimizar	$\mathbf{C}'\mathbf{W}$
sujeto a	$\mathbf{A}\mathbf{W} = \mathbf{Y}$
	$\mathbf{W}_k \geq 0 \quad r = 1, 2, \dots, 2n$
	$W_1, W_2, \dots, W_k$ cualesquiera.

(41)

**Definición 3.1** *Cualquier  $(\boldsymbol{\beta}, \mathbf{d}_1, \mathbf{d}_2)$  que satisfaga  $\mathbf{X}\boldsymbol{\beta} + \mathbf{I}\mathbf{d}_1 - \mathbf{I}\mathbf{d}_2 = \mathbf{Y}$  es llamada una solución factible al problema (34).*

**Definición 3.2** *Cualquier solución  $\mathbf{W}$  para (36) si además satisface:*

$$W_j \geq 0 \quad j = k + 1, \dots, k + 2n$$

*es llamada una solución factible al problema.*

Una solución básica factible del problema anterior es:

$$W_{k+r} = \begin{cases} Y_r & \text{si } Y_r > 0 \\ 0 & \text{en otro caso} \end{cases} \quad W_{n+k+r} = \begin{cases} -Y_r & \text{si } Y_r < 0 \\ 0 & \text{en otro caso} \end{cases} \quad (42)$$

y cero en el resto de los  $W_j$ .

### 3.6.1. Procedimiento del simplex para la regresión MINMAD

A pesar de que Arthanari (1981, [1]), discute el problema de una manera razonada y detallada; el procedimiento del simplex adaptado en este trabajo se limita a resumir brevemente el algoritmo.

**Algoritmo 3.3** *Empezar con la solución básica factible dada en la sección anterior. Si  $W_{k+n+r}$  está en la base, multiplicar por  $-1$  la fila correspondiente de la tabla inicial del simplex. Como de costumbre,  $C_j - Z_j = Z_j - C_B \alpha_j$ , donde  $C_B$  es el vector de coeficientes básicos y  $\alpha_j$  es la  $j$ -ésima columna de la tabla del simplex. El valor  $Z$  de la función objetivo se calcula como de costumbre, pero cambiando de signo, y se debe tener en cuenta que en tablas sucesivas el valor de la función objetivo es  $-Z$ .*

**Criterio para escoger el vector que sale de la base.**

Si  $C_j - Z_j > 0$  escoger  $r$  tal que  $\frac{W_{B_r}}{\alpha_{rj}} = \max_{i \in R_B} \left[ \frac{W_{B_i}}{\alpha_{ij}}, \alpha_{ij} < 0 \right]$

Si  $C_j - Z_j < 0$  escoger  $r$  tal que  $\frac{W_{B_r}}{\alpha_{rj}} = \max_{i \in R_B} \left[ \frac{W_{B_i}}{\alpha_{ij}}, \alpha_{ij} > 0 \right]$

donde  $R_B$  denota el conjunto de variables básicas que están restringidas en el problema,  $W_{B_i}$  es la  $i$ -ésima variable básica, y  $\alpha_{rj}$  es el elemento de la fila  $r$  y columna  $j$  de la tabla del simplex.

Una vez alcanzado el óptimo, los parámetros  $\beta_i$  se hallan en las variables  $W_1, \dots, W_p$ , y el error absoluto en la  $r$ -ésima observación es  $W_{p+r} + W_{p+r+n}$ .

### 3.6.2. Algoritmo modificado de Barrodale y Roberts.

Barrodale y Roberts (1973, [37]) proponen un algoritmo que modifica el simplex, permitiendo encontrar más rápidamente la solución básica factible del problema.

Se consideran únicamente los  $\beta_j$  no básicos para entrar en la base en las primeras iteraciones. El valor de la variable  $\beta_j$  que entra en la base se incrementa o disminuye según sea  $C_j - Z_j$  menor ó mayor que

0.

Se puede modificar el algoritmo como sigue:

Calcular  $\hat{C}_j - \hat{Z}_j = C_j - Z_j + 2\alpha_{rj} (*)$ .

Si  $C_j - Z_j$  y  $\hat{C}_j - \hat{Z}_j$  son de distinto signo, se introduce la variable correspondiente en lugar de la  $r$ -ésima en la base.

En caso contrario se cambian todos los  $C_k - Z_k$  por  $C_k - Z_k + 2\alpha_{rk}$ , y  $Z$  por  $Z - 2W_{B_r}$ . Cambiar la variable básica  $d_{1r}$  ó  $d_{2r}$  por  $d_{2r}$  ó  $d_{1r}$ , respectivamente. Multiplicar la  $r$ -ésima fila por  $-1$ . Descartar  $B_r$  para salir de la base. Ir a (\*). La tabla Simplex para resumir este tipo de problema, se puede escribir de la siguiente forma:

$C_B$	Vectores en la base		$W_B$	$\alpha_1$	$\alpha_2$	$\dots$	$\alpha_j$	$\dots$	$\alpha_{k+2n}$
1	$a_{k+1}$	ó $a_{k+n+1}$	$ Y_1 $	$\alpha_{11}$	$\alpha_{12}$	$\dots$	$\alpha_{1j}$	$\dots$	$\alpha_{1,k+2n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	$a_{k+n}$	ó $a_{k+2n}$	$ Y_n $	$\alpha_{n1}$	$\alpha_{n2}$	$\dots$	$\alpha_{nj}$	$\dots$	$\alpha_{n,k+2n}$
		$C_k - Z_k$	$Z = \sum_{i=1}^n  Y_i $				$C_j - Z_j$		

### 3.7. MINMAXAD para Regresión Lineal Múltiple

Para la estimación del vector de parámetros, mediante este método se supone el cumplimiento de 3 supuestos:

1.  $\mathbf{Y}$  no es combinación lineal de  $\mathbf{X}_1, \dots, \mathbf{X}_p$ .
2.  $\mathbf{X}_j \neq 0$  para todo  $j = 1, \dots, p$ .
3. Cualquier submatriz de orden  $p$  de  $\mathbf{X}$  es no singular.

Si no se cumpliera el supuesto 1,  $\mathbf{X}\boldsymbol{\beta} - \mathbf{Y} = \mathbf{0}$  tendría solución  $\boldsymbol{\beta}^*$  que sería óptima. Si no se cumpliera el supuesto 2, se podría prescindir de las columnas nulas y de las variables correspondientes. El supuesto 3 es necesario en la demostración de las proposiciones. Sea:

$$d = \max_{1 \leq i \leq n} \left| Y_i - \sum_{j=1}^p X_{ij} \beta_j \right|$$

Se puede suponer que las  $\mathbf{X}_j$ 's están normalizadas, es decir:

$$\sum_{j=1}^p X_{ij}^2 = 1 \quad i = 1, \dots, n. \quad (43)$$

Así  $Y_i - \sum_{j=1}^p X_{ij} \beta_j = 0 \quad i = 1, \dots, n$  son las ecuaciones de  $n$  hiperplanos en  $\mathbf{R}^p$  respecto de la variable  $\boldsymbol{\beta}$ . Si  $\boldsymbol{\beta} \in \mathbf{R}^p$ ,  $|Y_i - \sum_{j=1}^p X_{ij} \beta_j|$  es la distancia de  $\boldsymbol{\beta}$  al  $i$ -ésimo hiperplano.  $d$  es el máximo de estas distancias. Luego el problema se puede plantear como:

Minimizar	$d$	
sujeto a	$d + \sum_{j=1}^p \beta_j X_{ij} \geq Y_i$	
	$d - \sum_{j=1}^p \beta_j X_{ij} \geq -Y_i$	(44)
	$i = 1, \dots, n$	
	$d, \beta_j$ cualesquiera.	

No hay problema en suponer  $d > 0$ . Luego se transforman las variables como sigue:

$$b_0 = \frac{1}{d} \quad b_j = \frac{\beta_j}{d} \quad j = 1, \dots, p$$

y queda el problema equivalente siguiente:

Minimizar	$b_0$	
sujeto a	$b_0 Y_i + \sum_{j=1}^p b_j X_{ij} \leq 1$	
	$-b_0 Y_i - \sum_{j=1}^p b_j X_{ij} \leq 1$	(45)
	$i = 1, \dots, n$	
	$d, b_j$ cualesquiera.	

Entonces basta con aplicar el algoritmo del simplex a este problema minimizando  $-b_0$ .

### 3.8. MINSADBED para Regresión Lineal Múltiple

Sea  $d_i$  la desviación entre los valores observados y esperados de  $Y_i$ :

$$d_i = Y_i - (\mathbf{X}\boldsymbol{\beta})_i \quad i = 1, \dots, n$$

donde  $\mathbf{X}$  es una matriz de orden  $n \times p$ ,  $\boldsymbol{\beta}$  es un vector de dimensión  $p$  y  $(\mathbf{X}\boldsymbol{\beta})_i$  es la  $i$ -ésima fila de  $\mathbf{X}\boldsymbol{\beta}$ .

Aquí se considera el problema de calcular  $\boldsymbol{\beta}$  que minimice la expresión:

$$\sum_{i < j} |d_i - d_j|$$

Al intentar transformar en un problema de programación lineal.

Si  $d_{ij} = d_i - d_j$ ,  $d_{ij}$  no restringido en signo. Sea  $d_{ij} = d_{ij}^1 - d_{ij}^2$ , donde  $d_{ij}^k \geq 0$  para  $k = 1, 2$ . De esta forma  $|d_i - d_j|$  la diferencia absoluta entre las desviaciones, está dada por:

$$\begin{aligned} |d_i - d_j| &= (d_{ij}^1)^1 + d_{ij}^2 \\ d_i - d_j - d_{ij}^1 + d_{ij}^2 &= 0 \\ d_{ij}^k &\geq 0, \quad k = 1, 2 \quad 1 \leq i < j \leq n. \end{aligned} \tag{46}$$

La función objetivo se puede escribir como:

$$\text{Minimizar} \quad \sum_{i < j} d_{ij}^1 + \sum_{i < j} d_{ij}^2$$

Rescribiendo este problema en forma matricial, y usando:

$$\mathbf{D}^k = [d_{12}^k, d_{13}^k, \dots, d_{(n-1)n}^k] \quad \text{para } k = 1, 2$$

se tiene:

Minimizar	$\mathbf{D}^1 \mathbf{e} + \mathbf{D}^2 \mathbf{e}$
sujeto a	$\mathbf{X}\boldsymbol{\beta} + \mathbf{d} = \mathbf{Y}$
	$\mathbf{d}, \mathbf{D}^1, \mathbf{D}^2$ satisfaciendo (43).
	$\boldsymbol{\beta}$ no restringido en signo.

(47)

Se encuentra así la matriz correspondiente a (44). Obsérvese que se tiene una ecuación para cada elección del par  $(i, j)$ ,  $i < j$ . Hay  $r = n(n - 1)/2$  restricciones. La variable  $d_i$  aparece en  $n - i$  restricciones donde  $d_j$  aparece con coeficiente  $-1$  para todo  $i + 1 \leq j \leq n$  y  $d_{ij}^k$  aparece sólo una vez en estas  $r$  restricciones.  $d_{ij}^1 - d_{ij}^2$  aparece en la restricción correspondiente a  $(i, j)$ . estas restricciones se pueden expresar por  $[\mathbf{0}, \mathbf{H}, \mathbf{I}_r, -\mathbf{I}_r]\mathbf{V} = \mathbf{0}$ , donde  $\mathbf{V}' = (\boldsymbol{\beta}', \mathbf{d}', \mathbf{D}^{1'}, \mathbf{D}^{2'})$  y  $\mathbf{D}^1 \geq \mathbf{0}$ ,  $\mathbf{D}^2 \geq \mathbf{0}$ , donde:

$$\mathbf{H} = \begin{bmatrix} \mathbf{e}_{n-1} & -\mathbf{I}_{n-1} & \\ \mathbf{0}_{(n-2) \times 1} & \mathbf{e}_{n-2} & -\mathbf{I}_{n-2} \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1 \times (n-2)} & \mathbf{e}_1 & -\mathbf{I}_1 \end{bmatrix}$$

Finalmente el problema (44) se escribe:

Minimizar	$\mathbf{D}^1 \mathbf{e} + \mathbf{D}^2 \mathbf{e}$
sujeto a	$\begin{pmatrix} \mathbf{X}_{n \times p} & \mathbf{I}_n & \mathbf{0}_{n \times r} & \mathbf{0}_{n \times r} \\ \mathbf{0}_{r \times p} & \mathbf{H}_{r \times n} & -\mathbf{I}_r & \mathbf{I}_r \end{pmatrix} \mathbf{V} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \end{pmatrix}$
	$\boldsymbol{\beta}, \mathbf{d}$ no restringidos en signo.
	$\mathbf{D}^1 \geq \mathbf{0}, \mathbf{D}^2 \geq \mathbf{0}$
	$\mathbf{V}' = (\boldsymbol{\beta}', \mathbf{d}', \mathbf{D}^{1'}, \mathbf{D}^{2'})$

(48)

### 3.9. MINSADBAD en Regresión Lineal Múltiple

Ahora se considera la estimación de  $\beta$  usando como función objetivo la minimización de la suma de diferencias absolutas entre desviaciones absolutas. Es decir se tiene:

$$\begin{aligned}
 &\text{Minimizar} && \sum_{i < j} |\underline{d}_i - \underline{d}_j| \\
 &\text{sujeto a} && \underline{\mathbf{d}} \leq \mathbf{Y} - \mathbf{X}\beta \leq \underline{\mathbf{d}} \\
 &&& \underline{\mathbf{d}} \geq 0. \\
 &&& \beta \text{ no restringido en signo.}
 \end{aligned} \tag{49}$$

Aquí  $\underline{\mathbf{d}}' = (\underline{d}_1, \underline{d}_2, \dots, \underline{d}_n)$  son las desviaciones absolutas en las observaciones. Este problema puede ser formulado como un problema de Programación Lineal

Minimizar	$\mathbf{D}^1 \mathbf{e} + \mathbf{D}^2 \mathbf{e}$
sujeto a	$  \begin{pmatrix}  \mathbf{X}_{n \times p} & \mathbf{I}_n & \mathbf{0} & \mathbf{0} \\  -\mathbf{X}_{n \times p} & \mathbf{I}_n & \mathbf{0} & \mathbf{0} \\  \mathbf{0}_{r \times p} & \mathbf{H}_{r \times n} & \mathbf{I}_r & -\mathbf{I}_r  \end{pmatrix}  \mathbf{v} =  \begin{pmatrix}  \beta \\  \underline{\mathbf{d}} \\  \mathbf{D}^1 \\  \mathbf{D}^2  \end{pmatrix}  \begin{pmatrix}  \geq \\  \geq \\  =  \end{pmatrix}  \begin{pmatrix}  \mathbf{Y} \\  -\mathbf{Y} \\  \mathbf{0}  \end{pmatrix}  $
	$\beta$ no restringido en signo.
	$\underline{\mathbf{d}}, \mathbf{D}^1, \mathbf{D}^2 \geq \mathbf{0}$

(50)

### 3.10. Prueba de $H_0 : \beta_{q+1} = \dots = \beta_p = 0$

Al considerar el modelo de regresión  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$ . La estadística para ensayar  $\beta_{q+1} = \dots = \beta_p = 0$  es:

$$F_{LS} = \frac{SSR_r - SSR_f}{(p - q)\hat{\sigma}^2}$$

donde SSR es la suma de cuadrados de los residuos,  $SSR = \sum \hat{e}_i^2$ . Una prueba estadística similar para regresión MINMAD es usado:

$$F_{MDA} = \frac{SAR_r - SAR_f}{(p - q)(\tilde{\tau}/2)} \quad (51)$$

donde SAR es la suma de los valores absolutos de los residuos  $SAR = \sum |\tilde{e}_i|$ . La estimación de  $\tilde{\tau}$  es dada por (35) con  $m = n - k$ . Cuando se asume que los errores tienen distribución normal y cuando  $H_0$  es cierta, la estadística  $F_{LS}$  tiene distribución  $F$ . Cuando la distribución de los errores no es especificada,  $F_{MDA}$  tiene aproximadamente distribución  $F$ , teniendo en cuenta que  $n$  el tamaño de la muestra sea grande. De esta forma, cuando  $n$  es grande se calcula el valor  $P$  aproximado como  $P[F \geq F_{MDA}]$ , donde  $F \sim F_{(p-q), (n-k)}$ . La aproximación de  $P$  se encuentra utilizando la siguiente modificación: calcular  $P[G \geq (p - q)(1 - (p - q)/n)F_{MAD}]$ , donde  $G$  denota una variable aleatoria con distribución chi-cuadrado con  $p - q$  grados de libertad. Para todo  $n$  grande, es casi lo mismo  $P[F \geq F_{MAD}]$ , a causa de que  $1 - (p - q)/n \approx 1$  y  $(p - q)F$  con  $n = \infty$  tiene la misma distribución como  $G$ . Para  $n$  moderadamente, el valor  $P$  basado en  $G$  se ha encontrado que es muy seguro.

## 4. Discusiones y Algunos Procedimientos de Simulación

### 4.1. Discusiones

Según Acuña (2011, [42], p.193) el método de los mínimos cuadrados ha sido el más popular desde sus inicios, sin embargo, es un hecho reconocido por muchos autores que ante la presencia de datos *outliers*<sup>1</sup> (verticales u horizontales), hay una gran influencia en el ajuste por mínimos cuadrados. Un *outlier* entre más exagerado sea, hará que el ajuste lineal tienda a pasar cerca de él, y el análisis de los residuos no sería muy confiable ya que estos darían la impresión de que nada malo estuviera pasando.

Ante la presencia de datos influyentes, las alternativas de mínimos cuadrados son:

---

<sup>1</sup>Según Dodge y Birkes un outlier es un dato con un residual estandarizado mayor a 2,5 en valor absoluto.

1. Utilizar mínimos cuadrados como si nada hubiera pasado.
2. Realizar un análisis exploratorio de los datos antes de realizar el ajuste del modelo, esto incluye uso de diagnósticos (dentro de los cuales se encuentran distintos métodos para la detección de valores influenciales ó outliers), transformaciones, gráficas dinámicas, etc.
3. Usar Modelos Lineales Generalizados.
4. Utilizar métodos de Regresión Robusta, los cuales son modificaciones de los mínimos cuadrados, y tienen como objetivo ajustar un modelo que resista la influencia de los outliers.

Los inicios de la regresión robusta, se remontan hacia los años sesenta, cuando Huber (1964, [40]) publicó el método de regresión M. Según el mismo Huber, existen gran variedad de métodos de regresión Robusta, los cuales se agrupan en tres clases:

1. Regresión M (M es por Máxima Verosimilitud).
2. Regresión R (R es por Rangos).
3. Regresión L (L es por combinación lineal de estadísticos de orden).

Existen también distintas variaciones de éstos, pero los que mayor popularidad han alcanzado son estimadores de Regresión M, así el método MINMAD, es una modificación del grupo 3. Ante la gran cantidad de métodos robustos existentes, se podría pensar que el uso de la regresión robusta debería ser bastante amplio, pero ese no es el caso. La mayoría de paquetes estadísticos existentes, excepto S-PLUS y GRETL los ignoran. Después de los textos de Hampel (1986, [44]) y Rousseeuw y Leroy (1987, [43]), la investigación en regresión robusta ha disminuído, y desde 1990 no ha aparecido ningún otro texto al respecto. La mayoría de investigadores se han inclinado hacia otras áreas, especialmente la no paramétrica. Vale la pena aclarar que la regresión MINMAD es más conocida en los paquetes

como *MAD* lo cual significa *Minimizing Absolute Deviation*, el nombre MINMAD fué asignado por Edgeworth (1887)<sup>2</sup>. El método MINMAXAD fué estudiado por Wagner (1959, [36]) quien sugiere la aproximación dual para su solución, mientras que Stiefel (1960, [35]) da una conexión entre la programación lineal y la eliminación de Jordan, además de presentar aspectos geométricos del problema. Por su parte los métodos MINSADBED y MINSADBAD fueron propuestos por Arthanari y colaboradores (1977, [45]), quienes en su libro presentan un ejemplo simple de simulación. La Regresión MINMAD, ha recibido distintos nombres: Regresión LAD (Least Absolute Deviation), Bloomfield y Steiger (1983, [3]), Regresión LAV (Least Absolute Value), Armstrong, Elam y Hultz (1977, [46]), Regresión MAD (Minimum Absolute Deviations), Ashar y Wallace (1963, [47]), entre muchos otros. La Regresión MINMAD ha sido restrictiva en su uso, debido a cuatro razones:

1. El vector de coeficientes estimados no es único.
2. La regresión MINMAD resiste a la presencia de datos anormales en la dirección vertical, pero es poco efectiva en la dirección horizontal.
3. La eficiencia del estimador disminuye a medida que aumenta el número de casos.
4. Para obtener las estimaciones del coeficiente de regresión, hay que resolver un problema de programación lineal, el cual es muy lento computacionalmente.

## 4.2. Algunos Procedimientos de Simulación

Desde la propuesta de Charnes, Cooper y Ferguson (1955, [33]), muchos autores dirigieron sus esfuerzos a encontrar algoritmos de aproximación cada vez más sofisticados que se pudieran aplicar en métodos alternativos de regresión, no obstante y según Bloomfield y Steiger (1983, [3]), los resultados durante 25 años, no fueron más que *Reminiscencias del redescubrimiento de la rueda*, en los que los escritos

---

<sup>2</sup>Citado por Arthanari y Dodge, 1981

ignoraban trabajos anteriores. Fueron Barrodale y Roberts (1973, [37]) quienes siguiendo los principios expuestos por Charnes, Cooper y Ferguson, propusieron un algoritmo eficiente, basado en el simplex, el cual se sigue utilizando hoy en día. Dentro de los muchos trabajos de simulación, en este trabajo solamente se presentan tres en forma cronológica y en su respectivo orden: Bloomfield y Steiger (1983), Arthanari y colaboradores (1977) y Torres (2001). En el año 1983, Bloomfield y Steiger llevaron a cabo un problema de simulación usando *Métodos Montecarlo*, para estimar los parámetros del modelo  $Y_i = \beta_0 + X_{i1} + X_{i2} + U_i$ , con tamaños de muestra  $n = 10$  (pequeñas muestras),  $n = 50$  (muestras moderadas) y  $n = 100$  (muestras grandes) utilizando ocho distribuciones de probabilidad para  $U$  (normal estándar, doble exponencial, logística, pareto(1,2), normal contaminada al 5 %, normal contaminada al 25 %, distribución de colas pesadas al 5 % y distribución de colas pesadas al 25 %). Los autores proponen como normal contaminada a la siguiente distribución:

$$f_\epsilon(t) = \begin{cases} (1 - \epsilon)e^{-\frac{t^2}{2}}/(2\pi) & \text{si } |t| \leq k \\ (1 - \epsilon)e^{-k(|t|-k/2)}/(2\pi) & \text{si } |t| \geq k \end{cases}$$

es decir normal si  $t \in [-k, k]$  y exponencial si  $t \notin [-k, k]$ ,  $k(\epsilon)$  es solución de  $2\phi(k)/k - 2\Phi(-k) = \epsilon/(1 - \epsilon)$ , donde  $\phi$  y  $\Phi$  son respectivamente la densidad y la distribución de una normal estándar,  $f_\epsilon$  tiene una derivada continua y mínima información de Fisher para localización, entre todas las distribuciones contaminadas. Los autores tomaron:

$$\epsilon = 0,05 \quad \text{y} \quad k = 1,399 = k(0,05)$$

$$\epsilon = 0,25 \quad \text{y} \quad k = 0,776 = k(0,25)$$

para distribuciones de colas más pesadas que la exponencial, los autores utilizaron una distribución de Pareto de índice  $a$ , es decir:

$$h_\epsilon(t) = \begin{cases} (1 - \epsilon)e^{-\frac{t^2}{2}}/(2\pi) & \text{sit } \in [-k(\epsilon), k(\epsilon)] \\ a/(2(1 + |t|^{1+a})) & \text{sit } \notin [-k(\epsilon), k(\epsilon)] \end{cases}$$

es de aclarar que en la distribución de Pareto si  $a > 2$  la varianza es finita, pero si  $a \leq 2$  se tiene varianza infinita. Cuando  $a = 1,2$  se tiene media finita pero varianza infinita. Finalmente los autores

realizaron 901 iteraciones de cada distribución y los resultados los muestran en 24 engorrosas tablas de números, los cuales no muestran claramente los resultados, tampoco las conclusiones de la simulación. Arthanari y colaboradores (1977, [45]) proponen simular el modelo  $Y_i = 16 + 4X_{i1} + X_{i2} + 0,25X_{i3} + \epsilon_i$  con  $i = 1, 2, \dots, 20$ . Los vectores  $X_1, X_2, X_3$  con correlación no significativa  $r_{12} = 0,03, r_{13} = -0,06, r_{23} = -0,28$  se obtuvieron usando un conjunto de números aleatorios entre  $-10$  y  $10$ . Los errores fueron seleccionados aleatoriamente e independientemente de distribuciones simétricas con media 0 y varianza 1. Se escogieron las distribuciones: Uniforme, Normal y Laplace. Finalmente y usando un programa en fortran realizaron cinco tipos de estimación alternativa: Mínimos Cuadrados, MINMAD, MINMAXAD, MINSADBED y MINSADBAD. En su texto los autores solamente se limitan a mostrar la media y la varianza para cada una de las distribuciones y los métodos de estimación sin dar más detalles.

Por último Torres (2001, [39]) presenta una comparación entre tres métodos de Regresión: Mínimos Cuadrados, MINMAD y Regresión No Paramétrica, usando procedimientos de simulación. En este trabajo el autor pretende dar más y mejores detalles de lo que Bloomfield y Steiger pretendían mostrar. En este trabajo se utilizaron dos modelos: uno de regresión lineal simple y otro de regresión lineal múltiple, el primero  $Y = 4 + 5X_1 + \epsilon$ , dejando a  $X_1 \sim U(2, 5)$  fijo para cada iteración y el segundo  $Y = 3 + 4X_1 + 5X_2 + \epsilon$  con  $X_1 \sim U(1, 4)$  y  $X_2 \sim U(2, 5)$ . El objetivo de su trabajo fué determinar cuál de los tres métodos de regresión presenta mejor ajuste, cuando la distribución de los errores tiene forma distinta a la normal en elongación y cuando hay presencia de observaciones atípicas u *outliers*. Torres realizó las simulaciones y las estimaciones con el paquete estadístico SAS.

Se utilizaron 1000 simulaciones con tamaños de muestra de 20 y 50. Para la distribución normal con porcentaje de outliers se utilizó la normal contaminada, dada por, con valores de  $\epsilon$  de 0,05, 0,10, 0,15 y 0,20, lo que equivale a contaminar la distribución normal estándar con porcentajes de outliers de 5%, 10%, 15% y 20%, para cada uno de los modelos descritos anteriormente. Para asimetría y elongación de los errores, se utilizaron las distribuciones  $g$  y  $h$  de Tukey.

### 4.3. Distribuciones $g$ y $h$ de Tukey

Es una familia de distribuciones con características muy especiales en cuanto a asimetría y elongación, utilizadas para realizar estudios de sensibilidad, bastante útiles para realizar estudios de simulación de datos provenientes de distribuciones con formas distintas a la normal.

#### 4.3.1. Asimetría

Si  $Z$  es una variable aleatoria normal estándar, y  $g$  es una constante real, la variable aleatoria  $Y_g(Z) = g^{-1}(e^{gZ} - 1)$  se dice tiene una distribución  $g$  de Tukey para un valor dado de  $g$ . El parámetro  $g$  controla la magnitud y la dirección de la asimetría.

#### 4.3.2. Elongación

Si  $Z$  es una variable aleatoria normal estándar, y  $h$  es una constante real, la variable aleatoria  $Y_h(Z) = Ze^{hZ^2/2}$  se dice tiene una distribución  $h$  de Tukey para un valor dado de  $h$ . El parámetro  $h$  controla la cantidad de la elongación de la distribución. Las distribuciones de la familia  $h$  de Tukey son simétricas y su valor esperado es 0.

Los autores introducen parámetros de localización y escala ( $A$ ) y ( $B$ ), con el fin de determinar la distribución de una variable aleatoria  $X$  y asumiendo que  $p > 0,5$  se deben estimar cuatro parámetros que satisfacen cualquiera de las siguientes relaciones:

$$X_p = A + BY_p \quad \text{y} \quad X_{1-p} = A - B \exp\{-gZ_p\}Y_p \quad (52)$$

donde  $X_q$  denota la  $q$ -ésima cuantila de la variable aleatoria  $X$  tal que:

$$X_q = \inf\{x \mid P[X \leq x] > q\} = \sup\{x \mid P[X < x] \leq q\}$$

En la tabla 4 del apéndice A, se encuentra una tabla de distribuciones de la familia de Tukey, con el valor de los parámetros, trabajo desarrollado por Jiménez (2011, [38]). Torres utilizó cinco valores de los parámetros  $g$  y  $h$  en sus simulaciones: 0,2, 0,4, 0,6, 0,8 y 1. Los resultados de las simulaciones se encuentran en las tablas 2–3 del Apéndice A. Aquí se refieren dos gráficas y las conclusiones obtenidas por Torres.

1. Cuando la distribución de los errores es simétrica, de colas pesadas, generadas de las distribuciones  $h$  de Tukey, se obtienen mejores estimaciones de los parámetros del modelo con la regresión No Paramétrica basada en rangos y la regresión MINMAD que con la regresión de Mínimos Cuadrados. Si la distribución de los errores es simétrica de colas muy pesadas, se obtienen mejores estimaciones con la regresión MINMAD.
2. Se puede observar la robustez de los métodos de regresión MINMAD y No Paramétrica, frente a la regresión de Mínimos Cuadrados, cuando la distribución de los errores presenta colas pesadas.
3. Cuando la distribución de los errores es normal con porcentaje de outliers mayor o igual al 5%, se obtienen mejores estimaciones con regresión no paramétrica y le sigue la regresión MINMAD.
4. En cuanto a la detección de outliers, cuando la distribución de los errores es normal, el método que detectó más en promedio fué el método MINMAD, seguido de la regresión no paramétrica.
5. Ninguno de los tres métodos fué uniformemente mejor, para la estimación de los parámetros. Si la distribución de los errores es simétrica de colas muy pesadas, es mejor usar regresión MINMAD. Si la distribución de los errores es normal con altos porcentajes de outliers, se recomienda usar no paramétrica.

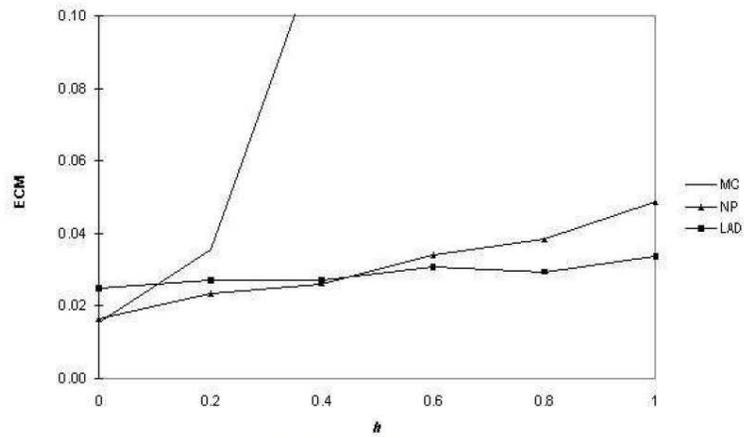


Figura 10. Error cuadrático medio. Estimador de  $\beta$ . Errores de familia h de Tukey  
Fuente: Torres(2001)

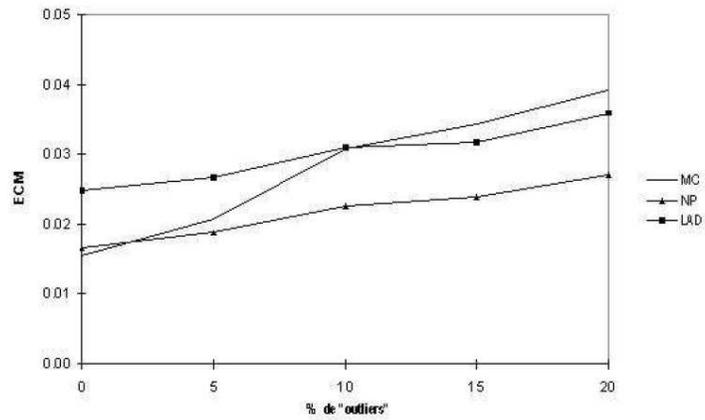


Figura 11. Error cuadrático medio para  $\beta$ . Errores de porcentaje de "outliers"  
Fuente: Torres (2001)

## 5. Estadística Bayesiana y Metodología MCMC

### 5.1. Estadística Bayesiana

La Estadística Bayesiana se atribuye a Thomas Bayes (1702-1761), propuso el teorema que lleva su nombre y fué uno de los primeros en estudiar la probabilidad en forma inductiva, además de fundamen-

tar los principios de la inferencia probabilística. Su teoría está siendo actualmente aplicada en modelos complejos, que hace algún tiempo eran difíciles de resolver. Según Ntzoufras (2009, [48]) desde mediados de los años 80's, el desarrollo en potencia y capacidad de los computadores y la implementación de las Cadenas de Markov y métodos Monte Carlo (MCMC) han llevado a una explosión de interés en Estadística Bayesiana y Modelamiento.

## 5.2. Teorema de Bayes

Considérense dos posibles salidas de un experimento  $A$  y  $B$ . Se puede asumir que  $A = A_1 \cup \dots \cup A_n$ , para los cuales  $A_i \cap A_j = 0$ , para todo  $i \neq j$ . Entonces, el teorema de Bayes provee una expresión para la probabilidad condicional de  $A_i$ , dado  $B$ , la cual es igual a:

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{P(B)} = \frac{P(B | A_i) P(A_i)}{\sum_{i=1}^n P(B | A_i) P(A_i)} \quad (53)$$

En una forma simple y más general, para cualquier salida  $A$  y  $B$ , se puede escribir

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)} \propto P(B | A) P(A) \quad (54)$$

Esta ecuación es también llamada regla de Bayes, aunque fué originalmente encontrada por Piere-Simon de Laplace (Hoffmann-Jorgensen, (1994, [49], p. 102)).

La anterior regla puede ser usada para inferencia inversa. Si se asume que  $B$  es la salida finalmente observada y que  $A$ , denota posibles causas que provocan  $B$ . Entonces  $P(B | A_i)$  puede ser interpretada como la probabilidad que  $B$  aparecerá cuando  $A_i$ , cause su presencia mientras  $P(A_i | B)$  es la probabilidad que  $A_i$ , es responsable por la ocurrencia de que  $B$  haya sido observado. La inferencia Bayesiana está basada en esta racionalidad. La ecuación anterior, ofrece un mecanismo para aprender de los datos (Bernardo y Smith (1994, [51], p. 2)). En consecuencia, después de observar datos  $(y_1, y_2, \dots, y_n)$  se calcula la distribución posterior  $f(\theta | y_1, y_2, \dots, y_n)$ , la cual combina distribución apriori e información muestral. Esta distribución posterior es el elemento fundamental en inferencia Bayesiana.

### 5.3. Definición de un Modelo Estadístico

Una de las más importantes cuestiones en ciencia estadística es la construcción de modelos probabilísticos que representen, o aproximen suficientemente, el verdadero mecanismo generador de un fenómeno bajo estudio. La construcción de tales modelos está usualmente basada en argumentos lógicos y probabilísticos concernientes a la naturaleza y función del fenómeno dado. Se asume una variable aleatoria  $Y$ , llamada respuesta, la cual sigue una regla probabilística con densidad ó función de probabilidad  $f(y | \boldsymbol{\theta})$ , donde  $\boldsymbol{\theta}$  es el vector de parámetros. Al considerar una muestra independiente e idénticamente distribuída  $\mathbf{y} = (y_1, \dots, y_n)^T$  de tamaño  $n$  de ésta. La distribución conjunta

$$f(\mathbf{y} | \boldsymbol{\theta}) = \prod_{i=1}^n f(y_i | \boldsymbol{\theta})$$

es llamada la *verosimilitud* del modelo y contiene toda la información provista por la muestra.

Usualmente los modelos son construídos con el fin de evaluar ó interpretar relaciones causales entre la variable respuesta  $Y$  y algunas características expresadas como variables  $X_j \in \nu$ , llamadas *covariables* o *variables explicativas*;  $X_j$  indica una covariable o término del modelo,  $\nu$  el conjunto de todos los términos bajo consideración. En tales casos, las variables explicativas están enlazadas con la variable respuesta mediante una función determinística y parte del vector de parámetros es sustituido por un conjunto alternativo de parámetros (denotado por  $\boldsymbol{\beta}$ ) que usualmente encapsula el efecto de cada covariable en la variable respuesta.

### 5.4. Modelo basado en Inferencia Bayesiana

La Estadística Bayesiana difiere de la teoría estadística clásica en que todos los parámetros desconocidos son considerados como variables aleatorias. Por esta razón, la distribución a priori debe ser definida inicialmente. Esta distribución a priori expresa la información disponible por el investigador antes de que cualquier “dato” sea involucrado en el análisis estadístico. El interés se centra en el cálculo de la

distribución posterior ó  $f(\boldsymbol{\theta} | \mathbf{y})$  del vector de parámetros  $\boldsymbol{\theta}$  dados los datos observados  $\mathbf{y}$ . De acuerdo al teorema de Bayes, la distribución posterior puede ser escrita como

$$f(\boldsymbol{\theta} | \mathbf{y}) = \frac{f(\mathbf{y} | \boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y})} \propto f(\mathbf{y} | \boldsymbol{\theta})f(\boldsymbol{\theta}) \quad (55)$$

Esta distribución posterior involucra, tanto la distribución a priori ó anterior, como los datos observados en la muestra, los cuales se representan mediante  $f(\boldsymbol{\theta})$  y la verosimilitud

$$f(\mathbf{y} | \boldsymbol{\theta}) = \prod_{i=1}^n f(y_i | \boldsymbol{\theta})$$

respectivamente.

La especificación de la distribuciónn apriori es importante en inferencia Bayesiana puesto que ésta influye en la inferencia posterior. Usualmente, se enfatiza en la especificación de la media y la varianza apriori. La media apriori provee una estimación puntual para los parámetros de interés, mientras que la varianza expresa la incertidumbre concerniente a esta estimación. Usualmente la información apriori no está disponible. En este caso se necesita una apriori que no influya en la distribución posterior y dejar que los datos “hablen por si mismos”. Tales distribuciones son llamadas *no informativas* ó *distribuciones apriori vagas*. Una de tales distribuciones es  $f(\boldsymbol{\theta}) \propto 1$ , la cual es la uniforme apriori sobre el espacio paramétrico. Se puede usar la regla de Bayes para inferir acerca de cualquier parámetro de interés  $\boldsymbol{\theta}$  siempre y cuando los datos observados hayan sido recolectados secuencialmente en diferentes puntos del tiempo (por ejemplo, en estudios prospectivos). Antes de que cualquier dato esté disponible, se usa sólo la distribución apriori  $f(\boldsymbol{\theta})$  para inferencia. Cuando un conjunto datos  $\mathbf{y}^{(1)}$  es observado, se puede usar la distribución posterior  $f(\boldsymbol{\theta} | \mathbf{y}^{(1)}) \propto f(\mathbf{y}^{(1)} | \boldsymbol{\theta})f(\boldsymbol{\theta})$ . Cuando un segundo conjunto de datos está disponible, se puede usar la posterior de la primera instancia como una apriori e incorporar los nuevos datos en una nueva distribución posterior. En consecuencia la distribución posterior será dada por

$$\begin{aligned} f(\boldsymbol{\theta} | \mathbf{y}^{(1)}, \mathbf{y}^{(2)}) &\propto f(\mathbf{y}^{(2)} | \boldsymbol{\theta})f(\boldsymbol{\theta} | \mathbf{y}^{(1)}) \\ &\propto f(\mathbf{y}^{(2)} | \boldsymbol{\theta})f(\mathbf{y}^{(1)} | \boldsymbol{\theta})f(\boldsymbol{\theta}) \end{aligned} \quad (56)$$

Esta ecuación puede ser generalizada para datos recolectados en  $t$  diferentes instantes de tiempo usando la ecuación

$$\begin{aligned}
 f(\boldsymbol{\theta} | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(t)}) &\propto f(\mathbf{y}^{(t)} | \boldsymbol{\theta}) f(\boldsymbol{\theta} | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(t-1)}) \\
 &\propto \prod_{k=1}^t f(\mathbf{y}^{(k)} | \boldsymbol{\theta}) f(\boldsymbol{\theta})
 \end{aligned}
 \tag{57}$$

donde es obvio que la teoría Bayesiana suministra un medio fácil que permite actualizar el conocimiento del parámetro de interés  $\boldsymbol{\theta}$ .

Con el fin de describir más apropiadamente la construcción de un *modelo estadístico*, realizamos los siguientes pasos:

1. Identificar la variable principal del problema (llamada variable respuesta  $\mathbf{Y}$ ) y sus correspondientes datos  $\mathbf{y}$ .
2. Encontrar una distribución que describa adecuadamente a  $\mathbf{Y}$ .
3. Identificar otras variables que puedan influenciar la variable respuesta  $\mathbf{Y}$  (llamadas *covariables* o *variables explicativas*).
4. Construir una estructura para los parámetros de la distribución (usando funciones determinísticas).
5. Especificar la distribución apriori (seleccionar la familia distribucional y especificar la apriori de los parámetros; seleccionar entre una apriori no informativa ó incorporar información apriori ó someter a opinión de expertos).
6. Escribir la verosimilitud del modelo.

En un segundo estado se puede identificar primero el método de cálculo de la distribución posterior (analíticamente, asintóticamente, o usando técnicas de simulación) y luego implementar el método seleccionado para estimar la distribución posterior.

Concerniente al análisis de la distribución posterior, se puede proceder con (todas o algunas) de las medidas que se proponen:

1. Inspeccionar visualmente las distribuciones marginales posteriores de interés. Posibles gráficas que pueden ser obtenidas como sigue
  - a) Densidades marginales posteriores ó gráficos de probabilidad si el método es analítico ó asintótico.
  - b) Histogramas marginales posteriores (ó densidades estimadas) para variables continuas.
  - c) Diagramas de cajas de las distribuciones marginales posteriores.
  - d) Gráficos bivariados posteriores (por ejemplo gráficos de contorno) para identificar y estudiar correlaciones.
2. Calcular resúmenes posteriores (medias, medianas, desviaciones estándar, correlaciones, odds ratios y gráficos de barras para variables discretas o categóricas, cuantiles e intervalos de confianza posteriores)
3. Calcular la moda posterior y el área de la densidad posterior (cuando fuere posible).

### **5.5. Inferencia usando distribuciones apriori conjugadas**

Usualmente la distribución objetivo posterior no es analíticamente tratable. En el pasado, este inconveniente fué evitado mediante el uso de distribuciones apriori conjugadas. Estas distribuciones apriori tienen la propiedad de resultar en posteriores de la misma familia distribucional. Una distribución apriori que es miembro de la familia distribucional  $D$  con parámetro  $\alpha$  es conjugada a la distribución  $f(\mathbf{y}|\theta)$  si la distribución resultante posterior  $f(\theta|\mathbf{y})$  es también un miembro de la misma familia

distribucional. De esta manera

$$\text{Si } \boldsymbol{\theta} \in D(\boldsymbol{\alpha}) \text{ entonces } \boldsymbol{\theta} | \mathbf{y} \in D(\tilde{\boldsymbol{\alpha}})$$

donde  $\boldsymbol{\alpha}$  y  $\tilde{\boldsymbol{\alpha}}$  son los parámetros apriori y posterior de  $D$ . En muchos casos los parámetros posteriores son medias ponderadas de los parámetros apriori y la verosimilitud. En la siguiente sección se ilustra un ejemplo de una distribución combinada y su respectiva inferencia.

### 5.5.1. Inferencia para la distribución de Poisson

Considérense un conjunto de datos de conteo  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  en los que se desea estimar la media  $\lambda$ . Se asume la distribución de Poisson para cada uno de los datos y se escribe:

$$y_i \sim \text{Poisson}(\lambda) \quad \text{para } i = 1, 2, \dots, n$$

en este caso el parámetro de interés es  $\lambda$ , por tanto  $\boldsymbol{\theta} = \lambda$ , mientras que la verosimilitud está dada por

$$f(\mathbf{y} | \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}$$

Se puede ahora considerar la densidad gamma con parámetros  $a$  y  $b$  como apriori para  $\lambda$ , cuya densidad es:

$$f(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

El vector de parámetros apriori está dado por  $\mathbf{a} = (a, b)$  y la distribución posterior resultante es

$$\begin{aligned} f(\lambda | \mathbf{y}) &\propto f(\mathbf{y} | \lambda) f(\lambda) \propto \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\ &\propto e^{-(n+b)\lambda} \lambda^{n\bar{y}+a} \end{aligned}$$

donde  $\bar{y}$  es la media muestral. Sabiendo que la distribución gamma de parámetros  $a$  y  $b$ , notada  $\Gamma(a, b)$  es proporcional a  $\lambda^{(a-1)} e^{-b\lambda}$ , se llega a la conclusión

$$\lambda | \mathbf{y} \sim \text{gamma}(n\bar{y} + a, n + b) \tag{58}$$

lo cual quiere decir que la distribución posterior es una gamma con parámetros  $\tilde{\mathbf{a}} = (n\bar{y} + a, n + b)'$ .

De esta forma la distribución gamma es conjugada para la distribución de Poisson.

La media posterior es:

$$E(\lambda | \mathbf{y}) = \tilde{\mu}_\lambda = \frac{n\bar{y} + a}{n + b}$$

Y la varianza posterior:

$$V(\lambda | \mathbf{y}) = \tilde{\sigma}_\lambda^2 = \frac{n\bar{y} + a}{(n + b)^2}$$

Las cantidades anteriores pueden ser reescritas de la siguiente forma

$$E(\lambda | \mathbf{y}) = \left(\frac{n}{n + b}\right)\bar{y} + \left(\frac{b}{n + b}\right)\left(\frac{a}{b}\right) = w\bar{y} + (1 - w)\left(\frac{a}{b}\right) = w\bar{y} + (1 - w)E(\lambda)$$

$$V(\lambda | \mathbf{y}) = \frac{n^2}{(n + b)^2}\left(\frac{\bar{y}}{n}\right) + \frac{b^2}{(n + b)^2}\left(\frac{a}{b^2}\right) = w^2\left(\frac{\bar{y}}{n}\right) + (1 - w)^2V(\lambda)$$

donde  $w = n/(n + b)$ . De esta manera la media posterior es expresada como un promedio ponderado entre la media muestral y la media apriori, la cual a su vez es un estimador de *máxima verosimilitud*.

### 5.5.2. Inferencia para la media de una distribución normal

No es el propósito de este trabajo dar un compendio acerca de distribuciones conjugadas, pero es importante resaltar algunas de las distribuciones conjugadas más utilizadas, por ende se menciona el caso normal, sin entrar en detalles teóricos. En el apéndice de tablas se encuentran resumidas las distribuciones conjugadas más importantes. También parece relevante mencionar que las distribuciones: Normal, Binomial, Poisson, Gamma e Inversa Gaussiana, entre otras pertenecen a una gran familia conocida como la *Familia Exponencial*.

Cuando un conjunto de datos  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  proviene de una población normal, con media desco-

nocida  $\mu$  y varianza conocida  $\sigma^2$ , la verosimilitud es:

$$\begin{aligned} f(\mathbf{y}) &= \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right) \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left( \sum_{i=1}^n y_i^2 + n\mu^2 - 2n\mu\bar{y} \right)\right) \end{aligned}$$

usando como apriori para la media una distribución normal  $N(\mu_0, \sigma_0^2)$ , es decir  $f(\mu | \sigma^2) = f_N(\mu; \mu_0, \sigma_0^2)$ ,

la distribución posterior está dada por

$$\begin{aligned} f(\mu | \sigma^2, \mathbf{y}) &\propto f(\mathbf{y} | \mu, \sigma^2) f(\mu | \sigma^2) \\ &\propto f(\mathbf{y} | \mu, \sigma^2) f_N(\mu; \mu_0, \sigma_0^2) \\ &\propto \exp\left(-\frac{1}{2} \left[ \frac{n\sigma_0^2 + \sigma^2}{\sigma^2 \sigma_0^2} \right] \left\{ \mu^2 - 2\mu \left[ \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \right] \right\}\right) \end{aligned}$$

La distribución posterior es una normal con media y varianza dadas por

$$E(\mu | \mathbf{y}) = \tilde{\mu} = \frac{n\bar{y}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{y} \quad V(\mu | \mathbf{y}) = \tilde{\sigma}^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2}$$

Las anteriores expresiones pueden ser reescritas como

$$\tilde{\mu} = w\bar{y} + (1-w)\mu_0 \quad \text{y} \quad \tilde{\sigma}^2 = w \frac{\sigma^2}{n}$$

donde  $w = n\sigma_0^2 / (n\sigma_0^2 + \sigma^2)$  y consecuentemente la media posterior es expresada como un promedio ponderado de la media muestral y la media de la distribución apriori.

## 5.6. Método de Jeffreys

Cuando se trata un problema en donde el investigador no tiene conocimiento previo de la situación, es decir desconoce por completo la distribución apriori del conjunto de parámetros  $\boldsymbol{\theta}$ , así conozca la evidencia de la muestra  $\mathbf{y}$ , Jeffreys (1961) propone como apriori invariante a:

$$\text{Pr}(\boldsymbol{\theta}) \propto \sqrt{|\mathbf{I}(\boldsymbol{\theta})|}$$

donde  $\mathbf{I}(\boldsymbol{\theta})$  es la matriz de información de Fisher:

$$\mathbf{I}(\boldsymbol{\theta}) = -E_{\theta} \left[ \frac{\partial^2 \text{Ln } f(\mathbf{y} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right]$$

Si  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)'$  es un vector se tiene

$$\text{Pr}(\boldsymbol{\theta}) \propto \sqrt{\det \mathbf{I}(\boldsymbol{\theta})} \quad (59)$$

donde  $\mathbf{I}(\boldsymbol{\theta})$  es la matriz de información de Fisher de orden  $p \times p$ . El elemento  $(i, j)$  de esta matriz es

$$\mathbf{I}_{ij} = -E_{\theta} \left[ \frac{\partial^2 \text{Ln } f(\mathbf{y} | \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]$$

Por transformación de variables la densidad apriori  $f(\boldsymbol{\theta}) = \text{Pr}(\boldsymbol{\theta})$  es equivalente a la siguiente densidad apriori para  $\phi$ :

$$\text{Pr}(\phi) = \text{Pr}(\boldsymbol{\theta} = h^{-1}(\phi)) \left| \frac{d\boldsymbol{\theta}}{d\phi} \right| \quad (60)$$

El principio general de Jeffreys consiste en que al aplicar el método para calcular la densidad apriori  $\text{Pr}(\boldsymbol{\theta})$ , se debe obtener un resultado equivalente en  $\text{Pr}(\phi)$  si se aplica la transformación del parámetro para calcular  $\text{Pr}(\phi)$  a partir de  $\text{Pr}(\boldsymbol{\theta})$  en la ecuación (60) o si se obtiene  $\text{Pr}(\phi)$  directamente a partir del método inicial. Es decir, debe cumplirse la siguiente igualdad:

$$\sqrt{\mathbf{I}(\phi)} = \sqrt{\mathbf{I}(\boldsymbol{\theta})} \left| \frac{d\boldsymbol{\theta}}{d\phi} \right|$$

## 5.7. Estimación en Inferencia Bayesiana

### 5.7.1. Aproximación a la teoría de la decisión

En estadística Bayesiana, la estimación es un problema de decisión. Es posible asociar con cada estimador  $a$  una función de pérdida  $L(\theta, a)$ , la cual establece la diferencia entre  $\theta$  y  $a$ .

La función de pérdida permite cuantificar las posibles penalidades, al escoger  $a$  como estimador de  $\theta$ .

Existen muchas funciones de pérdida que se pueden utilizar, dependiendo del contexto ó del problema.

Las más utilizadas son:

1. Pérdida cuadrática:  $L(\theta, a) = (\theta - a)^2$

2. Pérdida lineal absoluta ó error absoluto:  $L(\theta, a) = |\theta - a|$

3. Pérdida 0, 1:

$$L(\theta, a) = \begin{cases} 0 & \text{si } |\theta - a| \leq \epsilon \\ 1 & \text{si } |\theta - a| > \epsilon \end{cases}$$

4. Pérdida lineal para  $g, h \geq 0$

$$L(\theta, a) = \begin{cases} g(a - \theta) & \text{si } a > \theta \\ h(\theta - a) & \text{si } a < \theta \end{cases}$$

Se debe aclarar que la función de pérdida  $L(\theta, a)$ , significa la pérdida incurrida al adoptar la acción  $a$ , cuando el verdadero estado de la naturaleza es  $\theta$ .  $\Pr(a, y)$  es la pérdida esperada posterior. Con base en esto se tiene:

$$R_a(\theta) = E(L(\theta, a)) = \Pr(a, y) = \int L(\theta, a) \Pr(a, y) d\theta$$

La *Regla de Decisión*  $d(y)$  es la acción que minimiza  $\Pr(a, y)$  y el *Riesgo de Bayes* es  $RB(d) = \int \rho(d(y), y) \rho(y) dy$

Un estimador puntual en el contexto Bayesiano, es una simple estadística descriptiva de la distribución posterior  $f(\theta | \mathbf{y})$ . Usando la calidad del estimador a través de una función de pérdida, la teoría de la decisión conduce a elecciones óptimas de estimadores puntuales. Lo más natural, es que la elección de funciones de pérdida llevan a estimadores como la media posterior, mediana y moda como estimadores óptimos.

### 5.7.2. Intervalos de credibilidad

Un intervalo de credibilidad es el análogo a un intervalo de confianza en estadística clásica. Los estimadores puntuales no proporcionan una medida de la precisión de la estimación. Esta situación ocasiona problemas en la estadística clásica, puesto que los parámetros no son aleatorios, por tanto no es posible dar un intervalo, para que con cierta probabilidad el parámetro se encuentre dentro de ese intervalo, no obstante en estadística Bayesiana no hay dificultad puesto que los parámetros se consideran aleatorios.

**Definición 5.1** Una región de credibilidad  $100(1 - \alpha)$  de  $\theta$  es un subconjunto  $C$  de  $\theta$  tal que:

$$1 - \alpha \leq Pr(C | \boldsymbol{\theta}) \int_C dF(\boldsymbol{\theta} | \mathbf{y})(\boldsymbol{\theta}) = \begin{cases} \int_C f(\boldsymbol{\theta} | \mathbf{y}) d(\boldsymbol{\theta}) & \text{caso continuo} \\ \sum_{\boldsymbol{\theta} \in C} f(\boldsymbol{\theta} | \mathbf{y}) & \text{caso discreto} \end{cases}$$

Las regiones de credibilidad (llamadas también conjuntos veraces, por algunos autores) no son definidos de manera única (lo mismo ocurre con los intervalos de confianza en estadística clásica). Cualquier región con probabilidad  $(1 - \alpha)$  cumple con la definición (5,1). La idea es encontrar el intervalo que contiene los valores “más posibles” del parámetro, por tanto se hace imprescindible imponer una restricción adicional la cual indica que el ancho de la región debe ser tan pequeño como sea posible. Para hacer esto se deben sólo considerar aquellos puntos con  $f(\boldsymbol{\theta} | \mathbf{y})$  más grandes, esto da lugar a una región de la forma

$$C = C_\alpha(\mathbf{y}) = \{\boldsymbol{\theta} | f(\boldsymbol{\theta} | \mathbf{y}) \geq \gamma\}$$

donde  $\gamma$  es elegido de tal manera que  $\int_C f(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = 1 - \alpha$

La región que cumple las anteriores condiciones es llamada *Región de densidad posterior más grande* y sólo puede ser resuelta por métodos numéricos.

### 5.7.3. Pruebas de Hipótesis en Inferencia Bayesiana

Una prueba de hipótesis relaciona una decisión en la que se debe elegir entre dos conjeturas, relacionadas con el (los) parámetro (os) de interés.

$$\begin{cases} H_0 : \theta = \Omega_0 \\ H_1 : \theta = \Omega_1 \end{cases}$$

Se puede también tener el caso en el que los conjuntos  $\Omega_0$  y  $\Omega_1$  son valores simples de  $\theta$ , es decir:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

#### Aproximación clásica

Se ejecuta la prueba usando la razón de verosimilitud:

$$\lambda = \frac{f(y | \theta_1)}{f(y | \theta_0)}$$

Cuando  $\lambda$  toma valores grandes, significa que los datos observados  $Y$  fueron tomados si el verdadero valor del parámetro es  $\theta$  en lugar de  $\theta_0$ .

#### Aproximación Bayesiana

Para la estadística Bayesiana, parece natural realizar la prueba de los valores en la hipótesis, con las distribuciones posteriores relativas, esto es:

$$\lambda_B = \frac{f(y | \theta_1)}{f(y | \theta_0)} = \frac{f(y | \theta_1)}{f(y | \theta_0)} = \frac{f(\theta_1) f(y | \theta_1)}{f(\theta_0) f(y | \theta_0)}$$

Lo anterior en palabras significa: *Razón de apuestas a priori* es igual a *Razón de apuestas posteriores* por *Razón de verosimilitud*. Si  $\lambda_B$  toma valores grandes significa que hay preferencia por  $H_1$ .

**Definición 5.2** El cociente  $\frac{\alpha_1}{\alpha_0} = \frac{f(\theta_1 | y)}{f(\theta_0 | y)}$  es la razón de apuestas posterior de  $H_1$  a  $H_0$ , y  $\frac{\Pi_1}{\Pi_0}$  es la razón de apuestas a priori. La cantidad

$$B = \frac{\text{Razón de apuestas posterior}}{\text{Razón de apuestas a priori}} = \frac{\frac{f(\theta_1 | y)}{f(\theta_0 | y)}}{\frac{f(\theta_1)}{\theta_0}}$$

$$B = \frac{f(\theta_1 | y) f(\theta_0)}{f(\theta_0 | y) f(\theta_1)} = \frac{\alpha_1 \prod_0}{\alpha_0 \prod_1} = \lambda = \frac{f(y | \theta_1)}{f(y | \theta_0)}$$

se denomina factor de Bayes en favor de  $\Theta_1$ .

$B$  es una medida del peso que tienen los datos en favor de  $H_1$  sobre  $H_0$ . Si  $B$  es grande, se anula cualquier preferencia apriori por  $H_0$ . La preferencia posterior es  $H_1$ .

## 5.8. Simulación Montecarlo y su implementación en Inferencia Bayesiana

En muchos campos de la ciencia y la tecnología es importante determinar integrales de la forma:

$$I = \int_x g(x) dx$$

Muchísimas soluciones han sido propuestas a este tipo de problemas, comenzando desde los métodos teóricos, hasta soluciones computacionales sofisticadas. Una de tales técnicas está basada en la generación de muestras aleatorias de una densidad, y luego obtener la integral mediante la obtención de estadísticas insesgadas, como por ejemplo la media muestral. Se puede asumir que la función de densidad  $f(x)$  de una variable aleatoria  $X$ , puede generar valores aleatorios. esto se puede escribir

$$I = \int_x \left[ \frac{g(x)}{f(x)} \right] dx = \int_x g^*(x) dx$$

donde  $g^* = g(x)/f(x)$  y la integral  $I$  puede ser estimada eficientemente por:

1. Generar  $(x^{(1)}, x^{(2)}, \dots, x^{(T)})$  de la distribución objetivo con función de densidad de probabilidad (p.d.f.)  $f(x)$ .
2. Calcular la media muestral:

$$\hat{I} = \frac{1}{T} \sum_{t=1}^T \left[ \frac{g(x^{(t)})}{f(x^{(t)})} \right] \quad (61)$$

Este concepto fué utilizado en los inicios del primer computador electrónico y adoptado por Metrópolis y su equipo de investigación (1949 y 1986, [50], [52]).

El método descrito anteriormente es directamente aplicable a muchos problemas en inferencia Bayesiana. Por tanto para toda función del parámetro de interés  $G(\boldsymbol{\theta})$ , se pueden calcular la media y la varianza posteriores por simplicidad.

1. Generar una muestra  $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(T)}$  de la distribución posterior  $f(\boldsymbol{\theta} | \mathbf{y})$ .
2. Calcular la media muestral de  $G(\boldsymbol{\theta})$  por simplicidad utilizando

$$\hat{I} = \frac{1}{T} \sum_{t=1}^T G(\boldsymbol{\theta}^{(t)})$$

La simulación puede también ser utilizada para estimar y visualizar la distribución posterior de  $G(\boldsymbol{\theta})$ , la cual puede ser obtenida con la asistencia de las estimaciones *kernel* de los valores

$$G(\boldsymbol{\theta}^{(1)}), G(\boldsymbol{\theta}^{(2)}), \dots, G(\boldsymbol{\theta}^{(T)})$$

El principal problema es cómo encontrar la distribución posterior, lo cual muchas veces no es sencillo. La distribución posterior de aprioris conjugadas parece ser el camino más sencillo, cuando la posterior de  $G(\boldsymbol{\theta})$  se pretende encontrar cuando el conjunto de parámetros  $\boldsymbol{\theta}$ , no sea analíticamente tratable.

## 5.9. Métodos Montecarlo y Cadenas de Markov MCMC

Los métodos de simulación descritos en la sección inmediatamente anterior (también llamados métodos de simulación “directa”), no pueden ser aplicados en todos los casos. Ellos se refieren principalmente a distribuciones unidimensionales. Sin embargo, algunos de ellos están enfocados en el cálculo efectivo de integrales específicas (Givens and Hoeting, (2005, [53] p. 183)) y no pueden ser usadas para obtener muestras de alguna distribución posterior de interés. Las técnicas de simulación basadas superan

tales problemas, a causa de su flexibilidad y generalidad. Estas características, junto con el desarrollo masivo de facilidades de cómputo han hecho de las Cadenas de Markov y Métodos Montecarlo (MCMC), técnicas populares desde comienzos de los años 90. No obstante, los métodos MCMC no son nuevos pues fueron introducidos en la física en 1953 mediante una versión simplificada de Metrópolis y colaboradores (1953, [54]). Publicaciones intermedias incluyeron generalizaciones del algoritmo de Metrópolis como Hastings (1970, [56]) y el desarrollo de los *Muestreadores de Gibbs* por Geman and Geman (1984). Sinembargo, este pensamiento duró cerca de 35 años hasta que los métodos MCMC fueron redescubiertos por científicos Bayesianos tales como Tanner and Wong (1987, [58]); Gelfand y colaboradores (1990, [57]); Gelfand and Smith (1990, [55]) hasta llegar a ser una de las principales herramientas computacionales en inferencia estadística moderna.

Las Cadenas de Markov y Métodos Montecarlo motivaron a investigadores cuantitativos a utilizar modelos complicados y estimar las correspondientes distribuciones posteriores con mucha seguridad. En este sentido, los métodos MCMC han contribuido al desarrollo y propagación de la teoría Bayesiana. Extensivos detalles del uso del uso de los métodos MCMC pueden ser encontrados en Gilks y colaboradores (1996, [59]). BUGS (Spiegelhalter y colaboradores 1996, [60]) y el software WinBUGS (Spiegelhalter y colaboradores 2003, [61]). Además el uso de técnicas MCMC para generar muestras de distribuciones posteriores de modelos complicados, ha provisto una forma efectiva para evaluar modelos Bayesianos.

### 5.9.1. Algoritmo MCMC

Una *Cadena de Markov* es un proceso estocástico  $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(T)}\}$  tal que

$$f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t-1)}, \dots, \boldsymbol{\theta}^{(1)}) = f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)})$$

esto es, la distribución de  $\boldsymbol{\theta}$  en la secuencia  $t+1$ , dados todos los valores  $\boldsymbol{\theta}$  precedentes (para tiempos  $t, t-1, \dots, 1$ ), depende sólo del valor precedente inmediatamente anterior  $\boldsymbol{\theta}^{(t)}$ . Sinembargo,  $f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)})$  es

independiente del tiempo  $t$ , Finalmente cuando la cadena es irreducible, aperiódica y positiva recurrente, cuando  $t \rightarrow \infty$  la distribución de  $\boldsymbol{\theta}^{(t)}$  converge a su estado de equilibrio, el cual no depende de valores iniciales de la cadena  $\boldsymbol{\theta}^{(0)}$ , para detalles se puede ver Gilks y colaboradores (1996, [59]).

Con el fin de generar una muestra de la distribución posterior  $f(\boldsymbol{\theta} | \mathbf{y})$ , se debe construir una cadena de Markov con dos propiedades deseables:

1.  $f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)})$  debería ser fácil de generar, y
2. La distribución de equilibrio de la cadena de Markov seleccionada, debería ser la distribución posterior de interés  $f(\boldsymbol{\theta} | \mathbf{y})$ .

Asumiendo que se ha generado una cadena de Markov con estos requerimientos, entonces

1. Seleccionar un valor inicial  $\boldsymbol{\theta}^{(0)}$ .
2. Generar  $T$  valores hasta que la distribución de equilibrio es alcanzada.
3. Monitorear la convergencia del algoritmo usando diagnósticos de convergencia. Si el diagnóstico de convergencia falla, generar más observaciones.
4. Cortar las primeras  $B$  observaciones.
5. Considerar  $\{\boldsymbol{\theta}^{(B+1)}, \boldsymbol{\theta}^{(B+2)}, \dots, \boldsymbol{\theta}^{(T)}\}$  como la muestra para el análisis posterior.
6. Graficar la distribución posterior (usualmente debe enfocarse en las distribuciones marginales univariadas).
7. Finalmente, obtener resúmenes de la distribución posterior (media, mediana, desviación estándar, cuantiles y correlaciones).

## 6. Simulación y Estimación de Parámetros

### 6.1. Estimación vía MCMC

Para utilizar este proceso de estimación en el presente trabajo se utilizaron los dos modelos propuestos por Torres (2001, [39]), uno de Regresión Lineal Simple y otro de Regresión Lineal Múltiple, el primero  $Y = \beta_0 + \beta_1 X + \epsilon$ , donde  $\beta_0 = 4$ ,  $\beta_1 = 5$ ,  $X$  se estableció fija con distribución uniforme en el intervalo  $(2, 5)$ , los errores se consideraron con distribución normal de media cero y desviación estándar 3. El segundo modelo  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$  con  $\beta_0 = 3$ ,  $\beta_1 = 4$ ,  $\beta_2 = 5$ ,  $X_1$  y  $X_2$  se pusieron fijas con distribuciones uniformes en  $(1, 4)$  y  $(2, 5)$  respectivamente, de igual forma el error se consideró con distribución normal de media cero y desviación estándar 3.

Una vez propuestos estos modelos, los errores fueron contaminados para generar outliers, con porcentajes de 5, 10, 15 y 20 por ciento. También se utilizó la distribución  $g$  de Tukey para asimetría, con valores para  $g$  de 0,2, 0,4, 0,6, 0,8 y 1. Luego se procedió a generar elongación en los errores usando la distribución  $h$  de Tukey con valores para  $h$  de 0,2, 0,4, 0,6, 0,8 y 1. Posteriormente muestras de tamaños 10, 20, 50 y 100 fueron generadas mediante el paquete estadístico R, el error normal sin contaminación también se tuvo en cuenta.

Las muestras anteriores se procesaron en el lenguaje de programación WinBUGS, utilizando la versión experimental OpenBUGS, de la Universidad de Helsinki (Finlandia), disponible en la website <http://www.mrc-bsu.cam.ac.uk/bugs>, la cual puede correrse dentro de otro Software tal como R, Matlab ó Excel. Con los programas descritos en el apéndice B, utilizando dos cadenas de mil iteraciones cada una, se obtuvieron los estimadores Bayesianos para  $\beta_0, \beta_1$  en el modelo de regresión lineal simple y  $\beta_0, \beta_1, \beta_2$  para el modelo de regresión lineal múltiple. Los estimadores puntuales para cada tamaño de muestra y modelo se encuentran en las tablas 5 y 6 del apéndice A. De igual manera las *regiones de credibilidad* entre el 2,5% y el 97,5% se encuentran escritas para cada modelo y tamaño

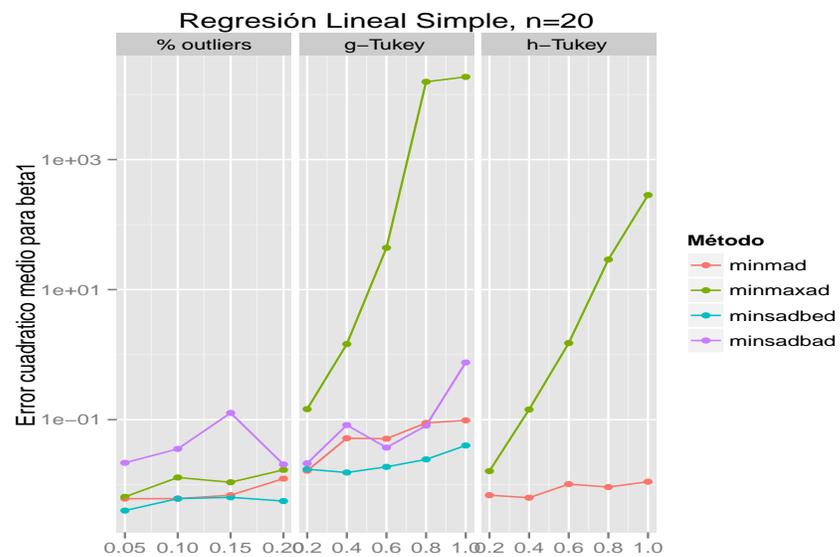
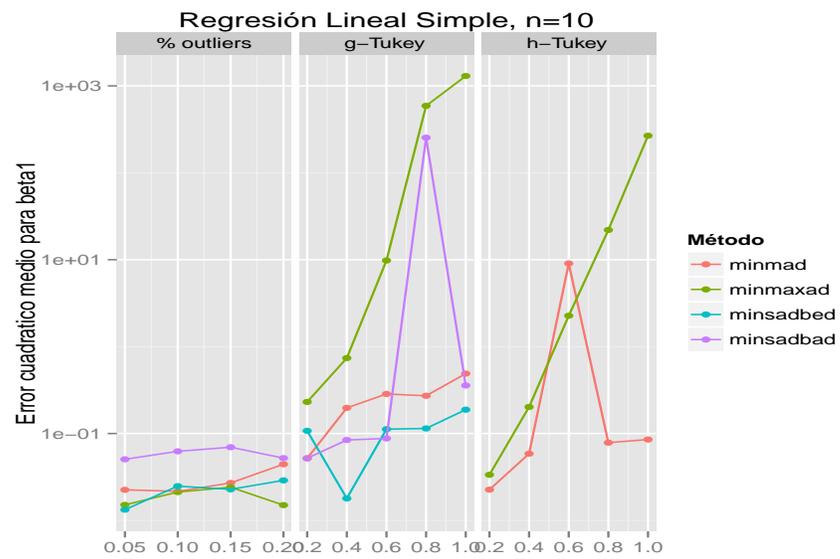
de muestra en las tablas 7 a la 12 del mismo apéndice.

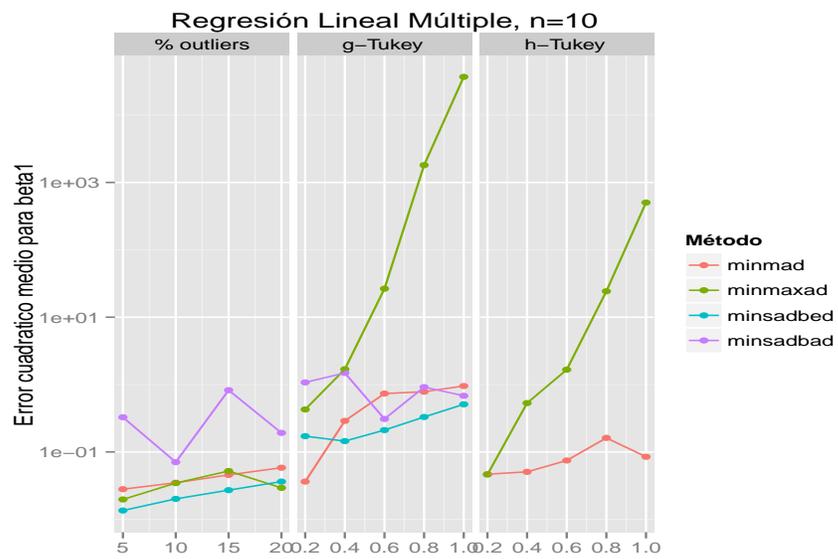
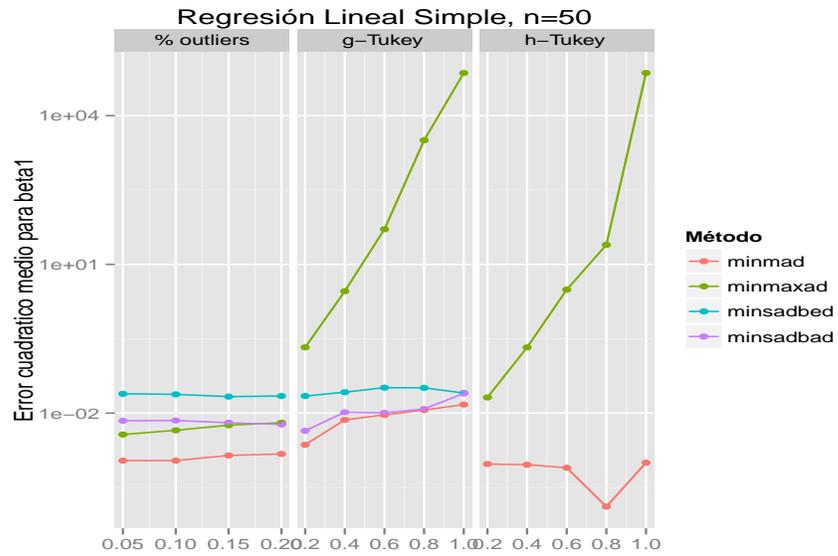
## **Estimación Robusta**

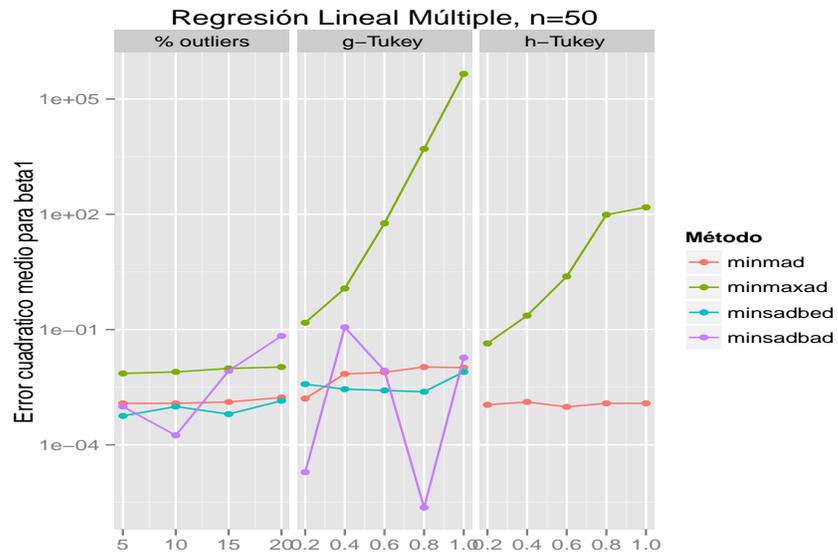
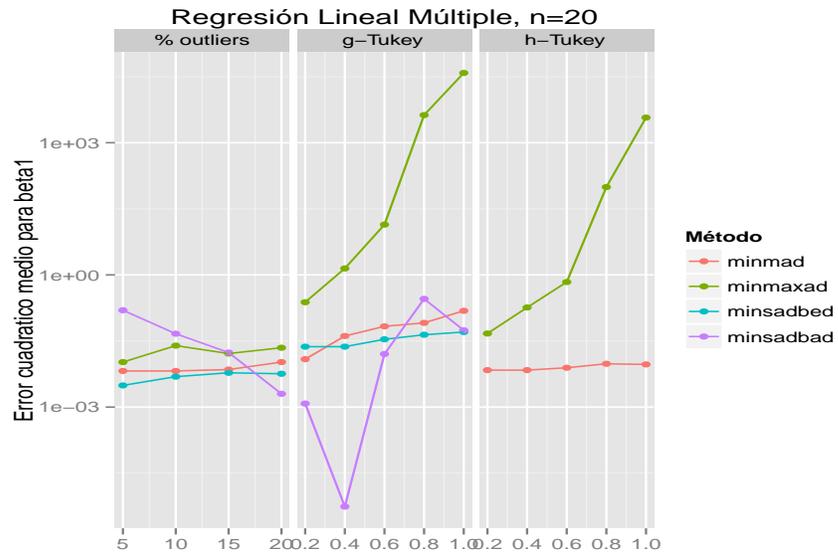
El objetivo principal del presente trabajo es comparar las estimaciones obtenidas en el procedimiento descrito anteriormente con las estimaciones de los parámetros ajustadas mediante los cuatro métodos de regresión robusta MINMAD, MINMAXAD, MINSADBED y MINSADBAD, los cuales fueron desarrollados con los programas en Matlab del apéndice C. El criterio de comparación utilizado en este trabajo es el *Error Cuadrático Medio*, también utilizado por Torres y colaboradores. Es bueno anotar que las estimaciones procesadas con estos métodos no son *únicas*, y por esta razón fue necesario realizar 1000 iteraciones. Esto significa que se generaron 1000 modelos con tamaños de muestra de 10, 20 y 50, utilizando errores normales, normales contaminados al 5 %, 10 %, 15 % y 20 %, normales con asimetrías utilizando la distribución  $g$  de Tukey con valores de  $g = 0,2,0,4,0,6,0,8$  y 1, normales con elongación mediante la distribución  $h$  de Tukey y valores de  $h = 0,2,0,4,0,6,0,8$  y 1 respectivamente. A cada modelo se le implementó el algoritmo simplex, para la estimación de los parámetros mediante Programación Lineal. Además en cada algoritmo se utilizó la Región de Credibilidad Bayesiana correspondiente en cada caso, con el fin de calcular el porcentaje de cobertura de cada uno de los parámetros, es decir contar cuántas veces de las mil, la estimación cae dentro de la Región en porcentaje. Los programas en Matlab propuestos para tal fin se encuentran en el Apéndice C, para cada uno de los cuatro métodos MINMAD, MINMAXAD, MINSADBED y MINSADBAD. Los algoritmos para cada uno de los métodos anteriores, se encuentran descritos en los problemas: (41), (44), (48) y (50), de la sección 3. Los resultados obtenidos están en las tablas 13 a la 36 del Apéndice A.

## 6.2. Análisis de algunos resultados de la simulación.

Es importante resaltar graficamente algunos resultados de las simulaciones y realizar una breve pero clara descripción de algunos detalles, los cuales permiten visualizar las ventajas y desventajas de los cuatro métodos de regresión robusta que se proponen en esta tesis.

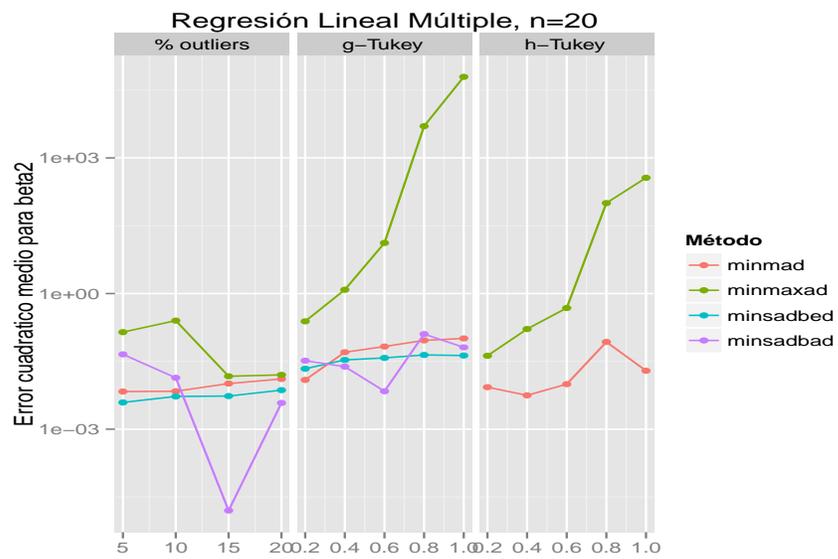
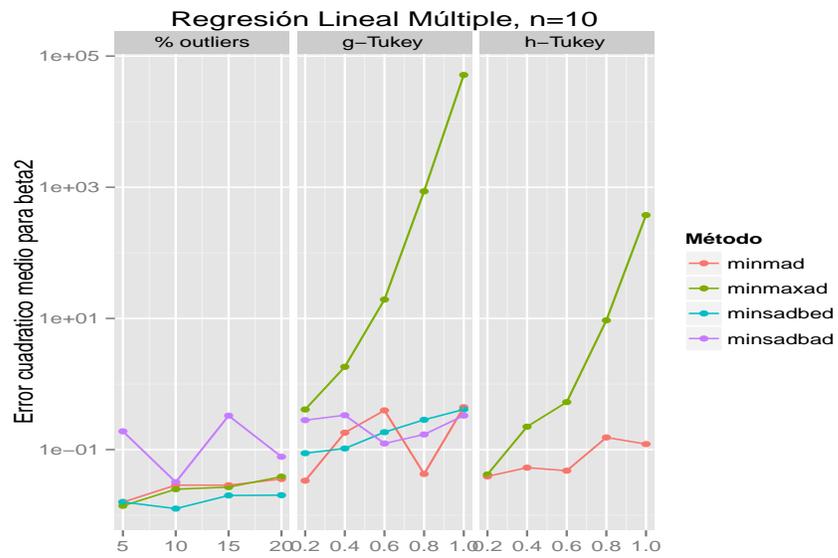


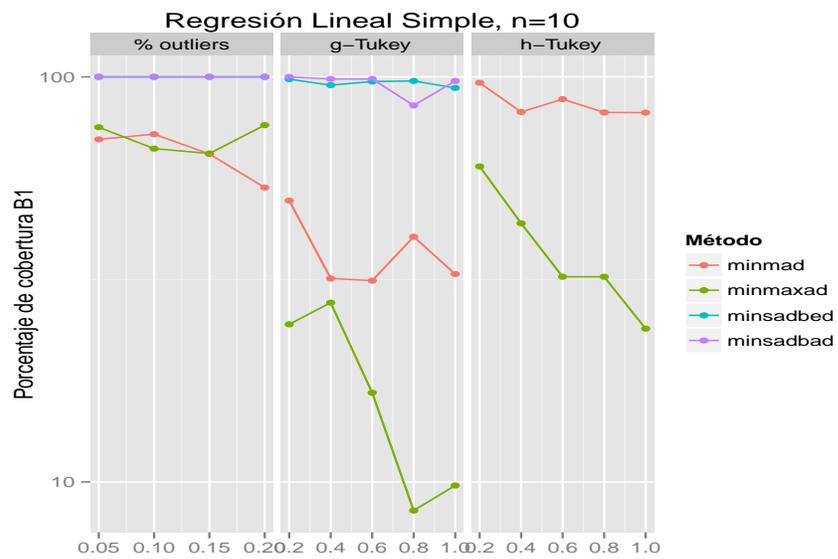
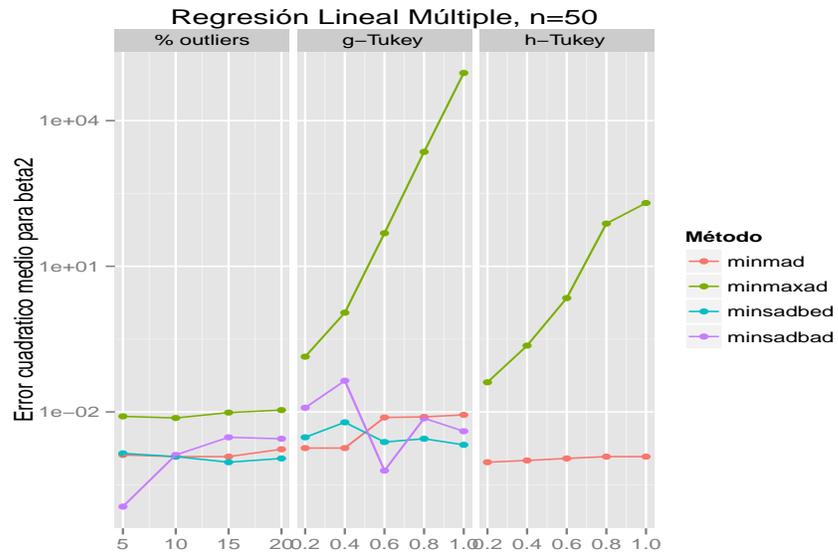


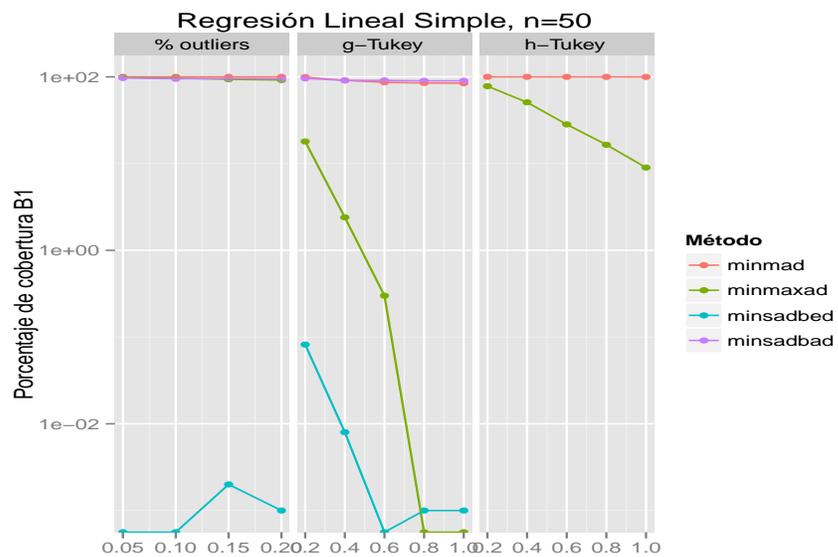
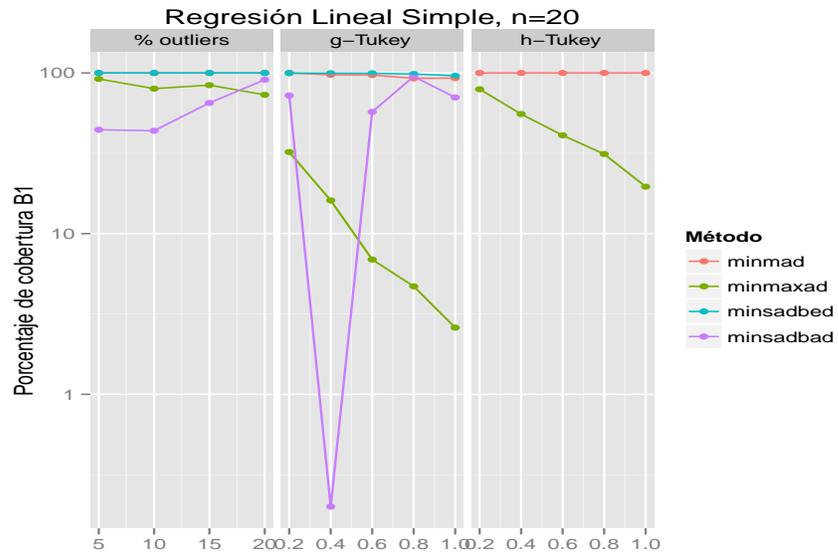


En las gráficas anteriores se puede apreciar la estabilidad en *Error Cuadrático Medio* del método MINSADBED (línea azul clara), tanto para porcentaje de outliers, como para asimetría (g de Tukey). Es importante aclarar que el método de elongación (h de Tukey) se realizó únicamente para los métodos MINMAD (color rojo) y MINMAXAD (color verde), puesto que para los métodos MINSADBED y MINSADBAD (color violeta), el programa que realiza las simulaciones aborta. Las estimaciones de los anteriores gráficos corresponden a la estimación del parámetro  $\beta_1$  en los modelos de regresión lineal simple y múltiple, con tamaños de muestra 10, 20 y 50.

Un buen observador diría que el método MINMAD también tiene una muy buena estabilidad al igual o aún mejor que el método MINSADBED, pero esto ya fué indicado por Torres y colaboradores. Se puede incluso percibir que MINMAD tiene un excelente comportamiento cuando la distribución de los errores tiene colas pesadas ó una alta elongación (h-Tukey). Se puede también evidenciar la poca estabilidad y los valores grandes del método MINMAXAD. Por su parte, el método MINSADBAD, aunque no presenta tanta inestabilidad como el método MINMAXAD, si evidencia serias variaciones. No se puede pasar por alto que MINMAXAD en porcentaje de outliers tiene un comportamiento aceptable.



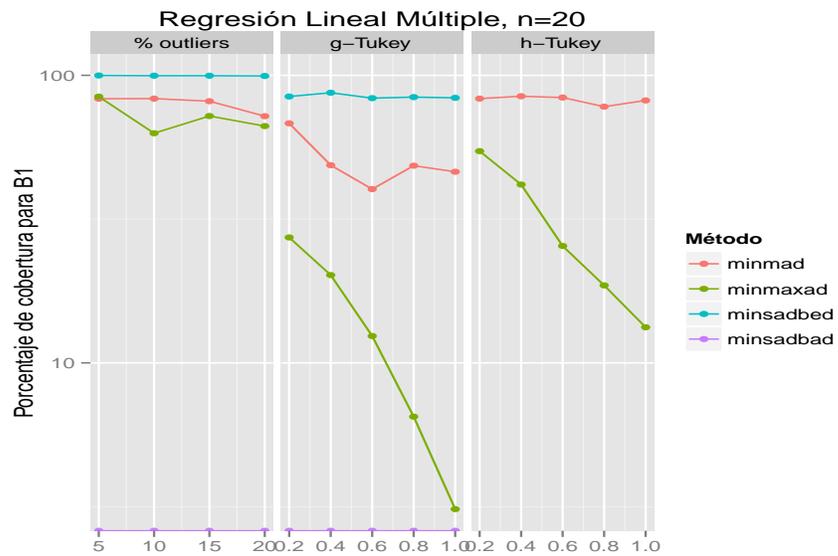
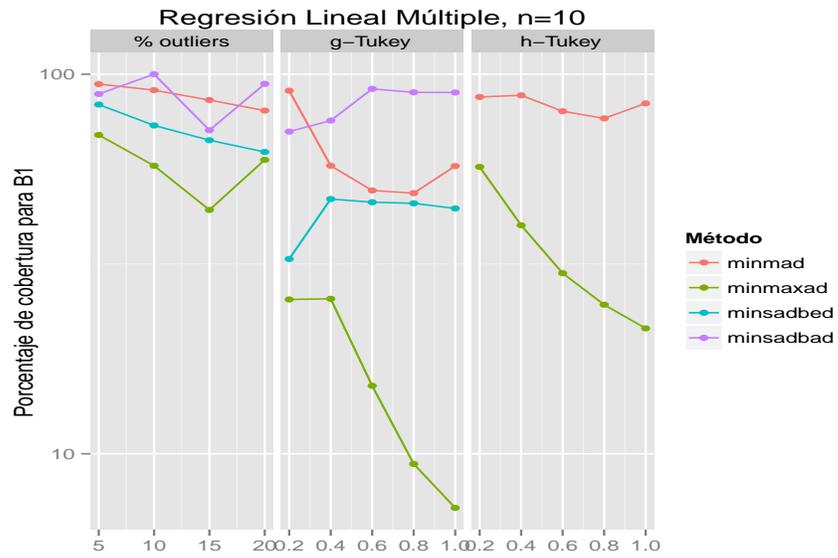


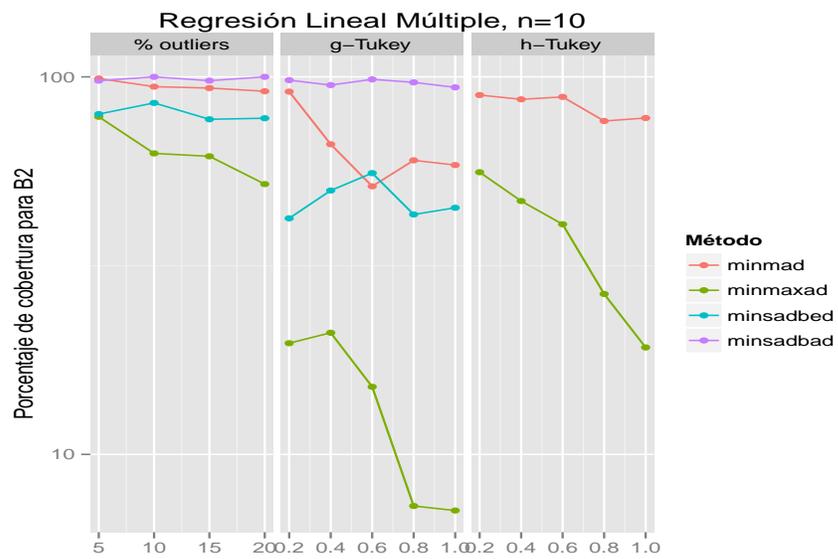
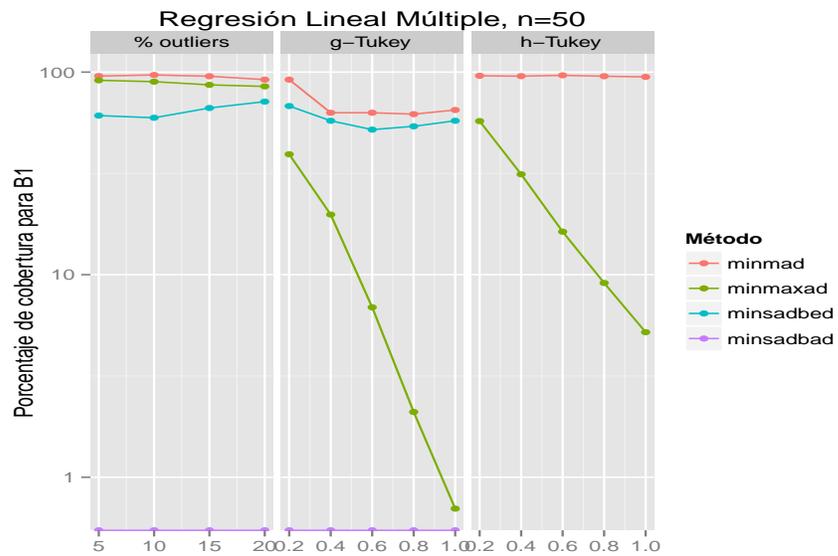


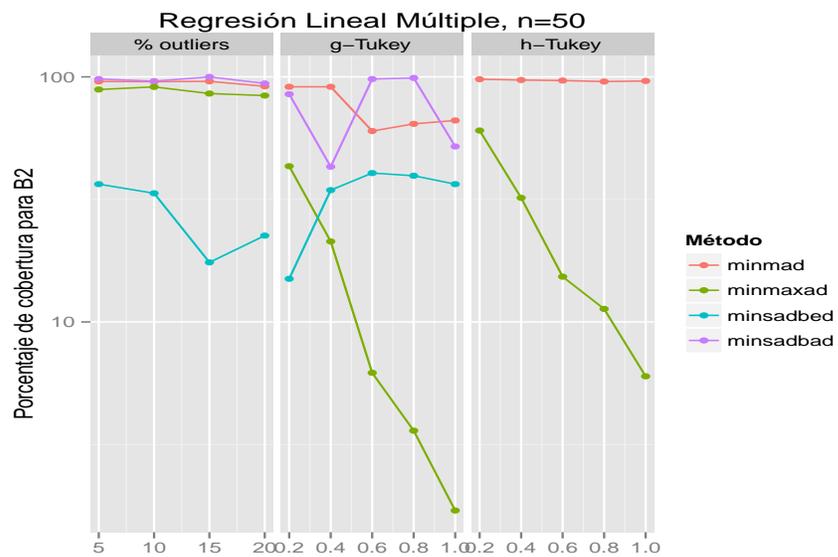
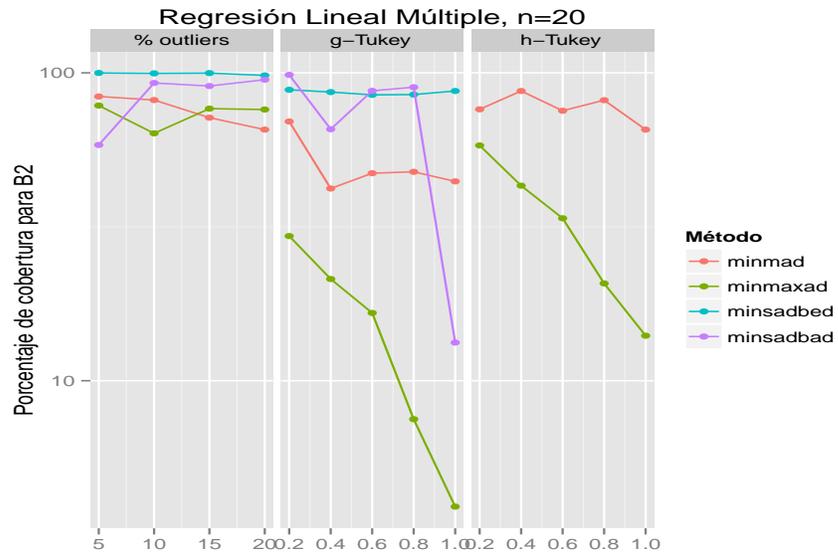
En los anteriores gráficos, se continúa evidenciando el buen comportamiento y estabilidad en *Error Cuadrático Medio* del método MINSADBED, para porcentaje de Outliers y Asimetría (g-Tukey), para la estimación del parámetro  $\beta_2$  en el modelo de regresión lineal múltiple.

En cuanto al porcentaje de cobertura para el parámetro  $\beta_1$  en el modelo de regresión lineal simple, se percibe un valor bajo para el método de regresión robusta MINSADBED cuando el tamaño de muestra

es 50, sin embargo presenta un excelente porcentaje de cobertura en muestras de tamaño 10 y 20.







Por último en los gráficos anteriores se exhibe el buen porcentaje de cobertura para la estimación del parámetro  $\beta_1$  en el modelo de regresión lineal múltiple, con el método MINSADBED. Para el parámetro  $\beta_2$ , se nota una pequeña disminución cuando el tamaño de muestra es 50, pero no ocurre lo mismo con tamaños de muestra 10 y 20.

No se puede pasar por alto la complejidad computacional del método MINSADBED, el cual presenta

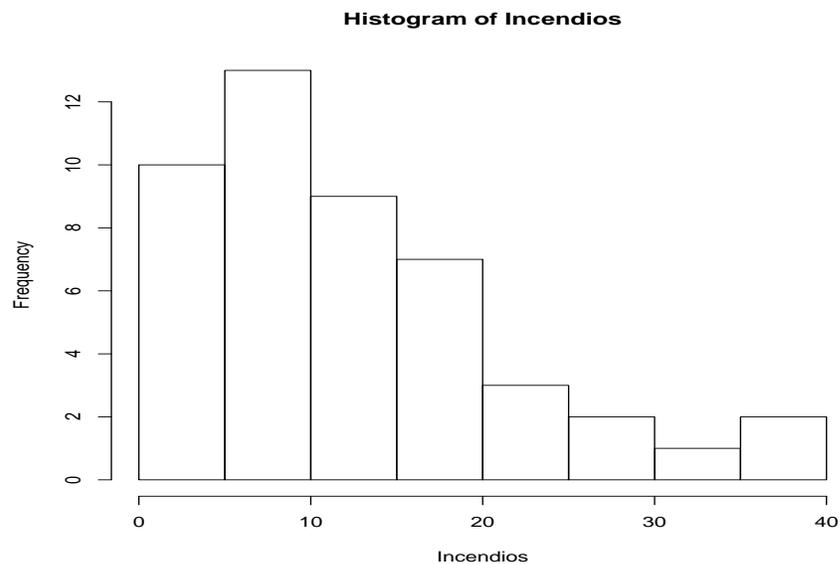
la matriz  $H$ , incluida la fórmula (48) de la sección 3. Esta matriz contiene 10 filas y 5 columnas en el de caso de manejar un tamaño de muestra igual a 5. En el caso de tener una muestra tamaño 10, la matriz  $H$  posee 45 filas y 10 columnas. Si se maneja una muestra tamaño 50,  $H$  tendría 1225 filas y 50 columnas. Esta matriz forma parte de las restricciones  $A$  en el algoritmo Simplex.

## 7. Aplicación

### 7.1. Método Minmad

Con el propósito de ilustrar estos métodos, se utilizaron los datos expuestos en el libro de Dodge y Birkes (1993, [4], p. 67). Los datos representan 47 áreas residenciales en la ciudad de Chicago, para el año 1975. La primera variable (la cual representa la variable dependiente ó respuesta  $Y$ ) son los incendios presentados por cada 1000 casas en el área. La segunda variable, corresponde a la primera variable explicativa ó independiente es *Edad*, representa la proporción de casas construídas antes de 1940. La tercera variable representa el número de robos por cada 1000 residentes, por último la variable *Ingresos*, representa la mediana del ingreso familiar como múltiplo de 1000.

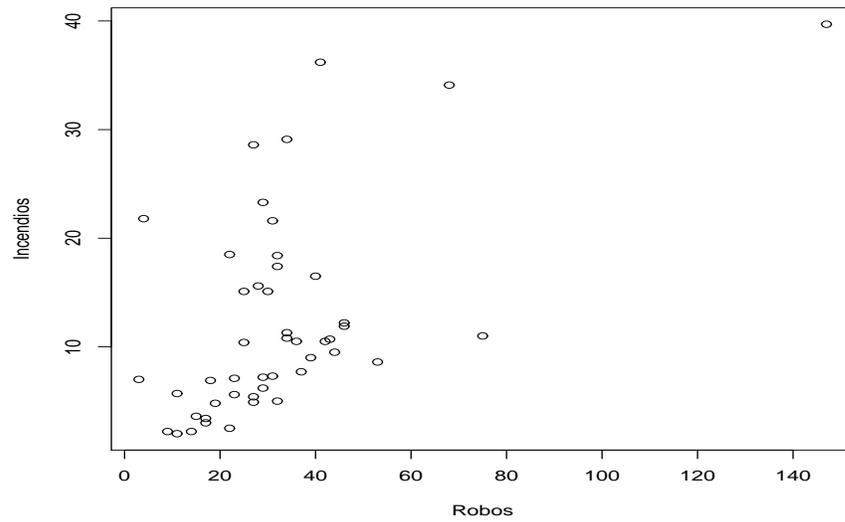
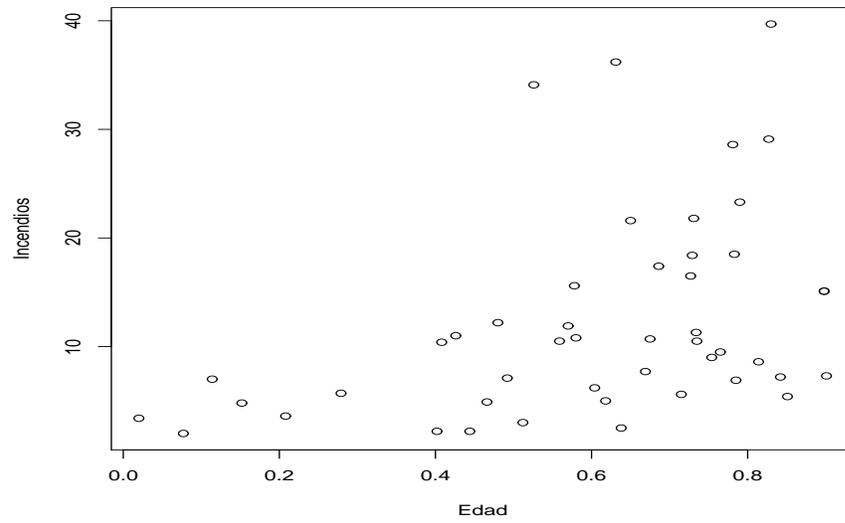
El siguiente gráfico muestra la distribución aproximada de la variable respuesta  $Y$  (Incendios):



En el gráfico anterior se puede apreciar una moderada asimetría positiva en los datos.

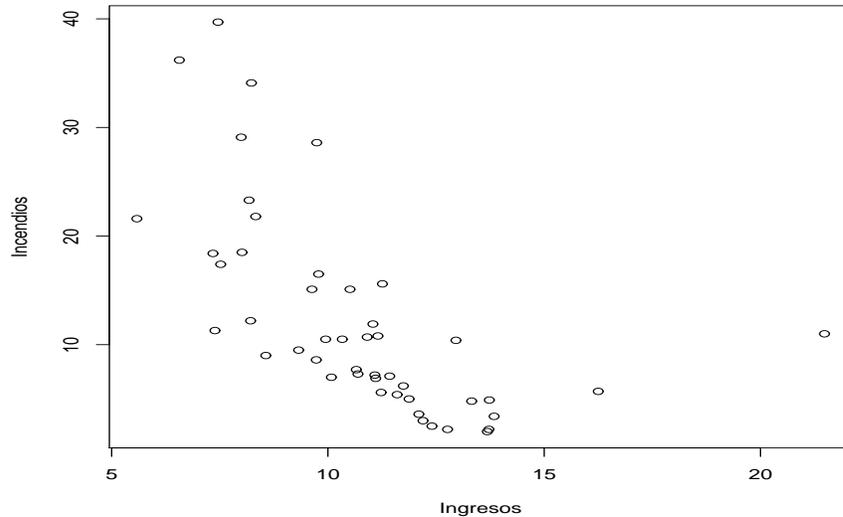
Los siguientes gráficos representan la variable respuesta  $Y$ : Incidencia de incendios, versus cada una de

las tres variables independientes ó explicativas  $X_1$ : Proporción de casas viejas,  $X_2$ : Incidencia de robos y  $X_3$ : Mediana del ingreso familiar.<sup>3</sup>



---

<sup>3</sup>Ver Dodge y Birkes ([4], p. 20-24).



En la gráfica de  $Y$  versus  $X_2$  un punto está más hacia la derecha que los demás, esto significa que la incidencia de robos en el área 24 es 147, mientras que en las demás áreas no supera los 75. Un punto que tiene un valor extremo para cualquiera de las variables explicativas es llamado *leverage point*, que traducido es *punto de apalancamiento*. El término *outlier* está reservado para puntos que tienen valores extremos en la variable respuesta. Los puntos de apalancamiento tienen una influencia desproporcional en las estimaciones de los coeficientes de regresión, puesto que éstos no siguen la misma distribución de los demás datos. Por tanto se sugiere borrar el área 24 de los datos y continuar el procedimiento con áreas que presentan robos no superiores a 75. También en la gráfica de  $Y$  versus  $X_3$  el área 7 es también un punto de apalancamiento pues presenta ingresos de 21,480, el siguiente mayor valor es 16,250, mientras que los demás ingresos no superan 14,000, también se sugiere borrar el área 7 de los datos.

En la gráfica de  $Y$  versus  $X_1$  la nube de puntos tiene una forma de abanico, lo cual indica varianzas desiguales. La gráfica de  $Y$  versus  $X_3$  también tiene forma de abanico con la diferencia que con respecto a la anterior abre en dirección opuesta. Este problema se puede corregir utilizando la transformación

logarítmica de la variable  $Y$ . Por esta razón se propone el modelo:

$$\log Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Dodge y Birkes proponen utilizar el método de estimación MINMAD, para lo cual se pueden procesar los datos en el paquete estadístico GRETL<sup>4</sup>, mediante la rutina de regresión robusta. En el siguiente cuadro se presentan los estadísticos básicos:

Cuadro 1: Estadísticas básicas MINMAD

Mediana	Media	$\hat{\sigma}_Y$	SAR <sub>f</sub>	SAR <sub>r</sub>	SSR
2.251	2.19953	0.754192	15.78	27.29	9.12609

Fuente: autor(2013)

Cuadro 2: Modelo estimado MINMAD

Parámetro	Estimación	Desv.Tip	Est.T	P	Significancia
$\beta_0$	4.36214	0.731929	5.960	0.00001	***
$\beta_1$	-0.0909789	0.420823	-0.216	0.82991	
$\beta_2$	0.0129898	0.00697726	1.862	0.06982	*
$\beta_3$	-0.242533	0.0479510	-5.058	0.00001	***
$\tau$	0.6371				

Fuente: autor(2013)

En la tabla anterior se estima el modelo de regresión MINMAD  $\hat{Y} = 4,36214 - 0,0909789 X_1 + 0,0129898 X_2 - 0,242533 X_3$ . La estimación de la desviación estándar de los errores estimados ó re-

<sup>4</sup>La versión libre de este paquete se encuentra en <http://es.kioskea.net/download/descargar-16584-gretl>.

residuos es 0,5100. Los  $n - k = n - (p + 1) = 45 - 4 = 41$  residuos debidamente ordenados en forma ascendente se encuentran en Dodge y Birkes (1993, [4], p. 77). Se propone entonces construir la estadística  $F_{\text{MAD}}$  con el fin de probar la hipótesis nula  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , la cual fué ya discutida y expuesta en la fórmula (51) de la sección 3. Para ello se hace necesario estimar el parámetro  $\tau$ , también discutido en la fórmula (35) de la sección 3, el cual emula la desviación estándar en regresión robusta. Se hace también necesario aclarar que el valor  $q$  incluido en la hipótesis  $H_0 : \beta_{q+1} = \dots = \beta_p = 0$  expresada en la sección (3,10) tiene un valor  $q = 0$ . Los  $m = n - k = n - (p + 1) = 45 - (3 + 1) = 45 - 4 = 41$  residuos distintos de cero, son usados para calcular el parámetro  $\hat{\tau}$ .

$$k_1 = (m + 1)/2 - \sqrt{m} = (41 + 1)/2 - \sqrt{41} = 14,60 \approx 15.$$

$$k_2 = (m + 1)/2 + \sqrt{m} = (41 + 1)/2 + \sqrt{41} = 27,40 \approx 27.$$

En la tabla de residuos expuesta en la página 77 del texto en mención, se puede apreciar que  $\hat{\epsilon}_{k_1} = \hat{\epsilon}_{15} = -0,1492$  y  $\hat{\epsilon}_{k_2} = \hat{\epsilon}_{27} = 0,2488$ , de donde se puede calcular:

$$\hat{\tau} = \frac{\sqrt{m}[\hat{\epsilon}_{k_2} - \hat{\epsilon}_{k_1}]}{4} = \frac{\sqrt{41}[0,2488 + 0,1492]}{4} = 0,6371$$

$$F_{\text{MAD}} = \frac{\text{SAR}_r - \text{SAR}_f}{(p - q)(\hat{\tau})/2} = \frac{27,29 - 15,78}{3(0,6371)/2} = 12,04$$

Para obtener la probabilidad de significancia P, se calcula primero:

$$(p - q)(1 - (p - q)/n)F_{\text{MAD}} = (3 - 0)(1 - (3 - 0)/45) \times 0,0464 = (2)(1 - 0,02) \times 12,04 = 33,71$$

El valor P es aproximadamente  $\text{Prob}[G \geq 33,71]$ , donde G tiene distribución chi-cuadrado con  $p = 3$  grados de libertad, se obtiene  $P = 0,001$ . Con este valor se puede afirmar cualquiera de las variables independientes tiene significancia alguna sobre la variable incidencia de incendios.

Probando secuencialmente cada una de las tres hipótesis  $H_{01} : \beta_1 = 0$ ,  $H_{02} : \beta_2 = 0$  y  $H_{03} : \beta_3 = 0$ . En la primera hipótesis se ve la necesidad de eliminar la variable  $X_1$ , con lo que se propone el modelo:

$$\log Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Al realizar nuevamente la estimación MINMAD de los parámetros se obtiene el modelo estimado:  $\hat{Y} = 4,252 + 0,01156X_2 - 0,2330X_3$ , el cual produce una estimación de la varianza residual de  $\hat{\sigma}_e = 0,5232$ . Posteriormente al probar las hipótesis  $H_{02}$  y  $H_{03}$ , se evidencia que se debe borrar la variable  $X_2$  del modelo, con lo que finalmente se considera el nuevo modelo:

$$\log Y = \beta_0 + \beta_3 X_3 + \epsilon$$

Realizando una vez más la estimación MINMAD de los parámetros, se obtiene finalmente el modelo estimado:  $\hat{Y} = 4,988 - 0,2718X_3$  que produce una desviación estándar residual de  $\hat{\sigma}_e = 0,5225$ . Al realizar los gráficos entre los residuos estandarizados y el valor estimado de  $Y$  y las variables independientes, no se evidencia no linealidad ni varianza no constante (estos gráficos se obvian), pues se encuentran descritos en Dodge y Birkes (1993, [4], p. 23).

Finalmente con el fin de probar los eficientes resultados de esta regresión robusta y siguiendo a Paula (2013, [15]), éste autor propone, teniendo en cuenta la asimetría de la variable respuesta, utilizar un modelo multiplicativo en lugar de un modelo aditivo y aplicar un *Modelo Lineal Generalizado*, para el ajuste de los parámetros, utilizando una densidad *Gamma* para la componente aleatoria del modelo (variable respuesta  $Y$ ), con función de enlace *log*. Paula propone el modelo:

$$\mu = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3} \cdot e^\epsilon$$

Al aplicar logaritmo, se obtiene:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Empleando el programa en R el cual se encuentra en el apéndice E1, se obtiene:

Cuadro 3: Estimación GLM

Parámetro	Estimación	Desv.Tip	Est.T	P	Significancia
$\beta_0$	3.996344	0.661223	6.044	3.7e-07	***
$\beta_1$	0.352791	0.409176	0.862	0.394	
$\beta_2$	0.008529	0.006761	1.262	0.214	
$\beta_3$	-0.204680	0.042529	-4.813	2.04e-05	***

Fuente: autor(2013)

Es decir el modelo estimado es:  $\hat{Y} = 3,996344 + 0,352791X_1 + 0,008529X_2 - 0,204680X_3$ . En la tabla se visualiza que las variables  $X_1$  y  $X_2$  no son significativas, lo cual permite proponer el modelo:

$$\log(\mu) = \beta_0 + \beta_3 X_3 + \epsilon$$

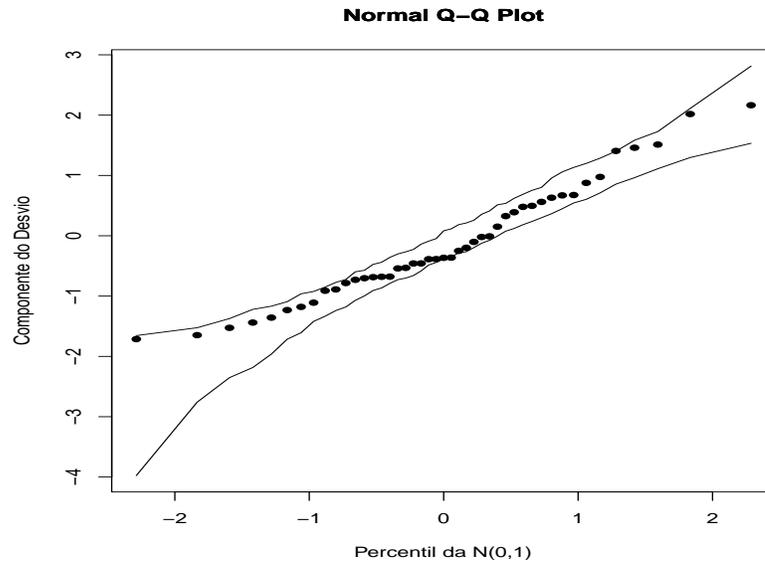
Utilizando el programa en R, del apéndice E2, se obtiene la estimación del modelo anterior:  $\hat{Y} = 4,9040 - 0,2466X_3$ . La tabla siguiente muestra los resultados:

Cuadro 4: Estimación GLM

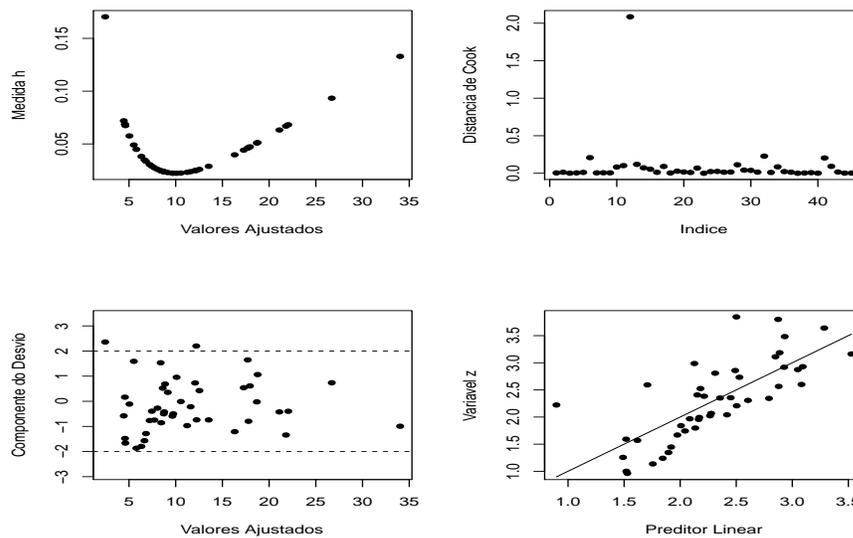
Parámetro	Estimación	Desv.Tip	Est.T	P	Significancia
$\beta_0$	4.9040	0.3604	13.607	2e-16	***
$\beta_3$	-0.2466	0.0335	-7.362	3.8e-09	***

Fuente: autor(2013)

Se puede evidenciar claramente la similitud entre el modelo robusto reducido y el modelo lineal generalizado. La gráfica siguiente muestra la bondad del ajuste:



Por último la siguiente gráfica presenta el diagnóstico del modelo ajustado.



En el gráfico que representa la distancia de *Cook*, se observa que sólo dos observaciones tienen una distancia de *Cook* considerable, la cual no supera 2,5 y en general ésta distancia se considera incoherente

para el conjunto de datos.

No se puede pasar por alto la aplicación de los métodos MINSADBAD y MINSADBED a éste conjunto de datos, se realizaron las estimaciones de los parámetros, obteniendo los siguientes resultados:

Cuadro 5: Estimación MINSADBAD

Parámetro	Estimación	Desv.Tip	Est.T	P	Significancia
$\beta_0$	2.2062	0.4121	5.3535	1e-6	***
$\beta_1$	0	0.7681	0	1	
$\beta_2$	0.0047	0.0134	0.3507	0.7275	
$\beta_3$	0.2564	0.07772	3.2990	0.0019	**

Fuente: autor(2013)

Se observa que se obtiene el modelo estimado  $\hat{Y} = 2,2062 + 0,25 X_3$ , pues las variables independientes  $X_1$  y  $X_2$  no son significativas. La desviación estándar residual es  $\hat{\sigma}_e = 1,2271$ , la cual es mucho mayor que la obtenida en la regresión MINMAD. Los estadísticos para la regresión robusta MINSADBAD se resumen en la siguiente tabla:

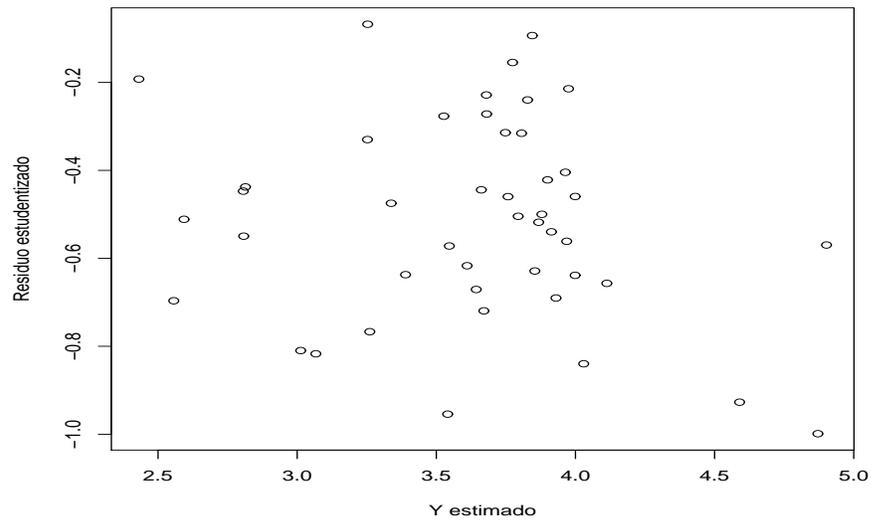
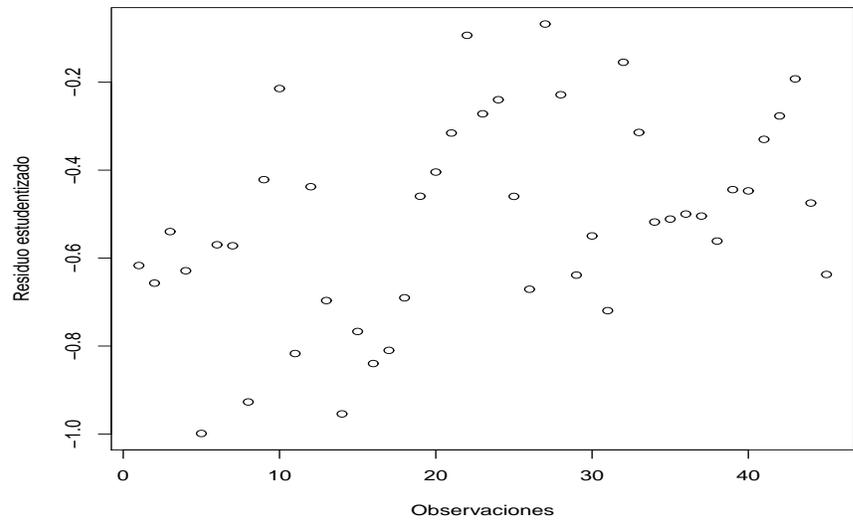
Cuadro 6: Estadísticas básicas MINSADBAD

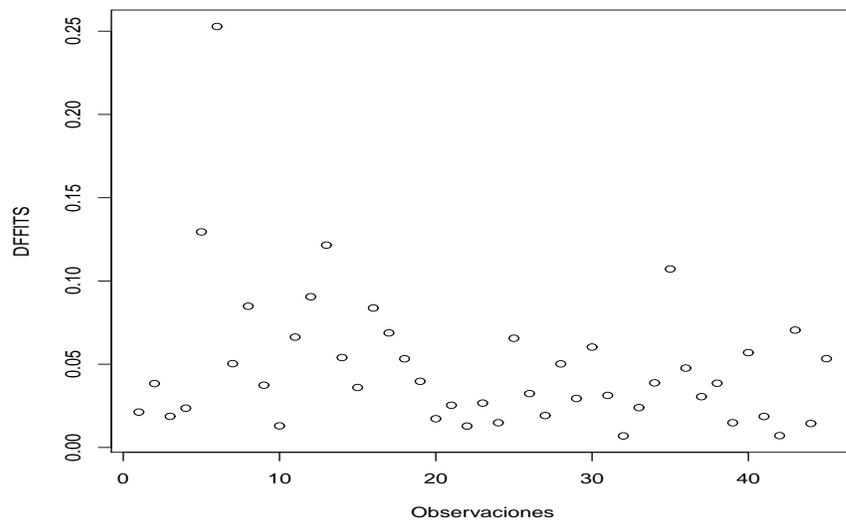
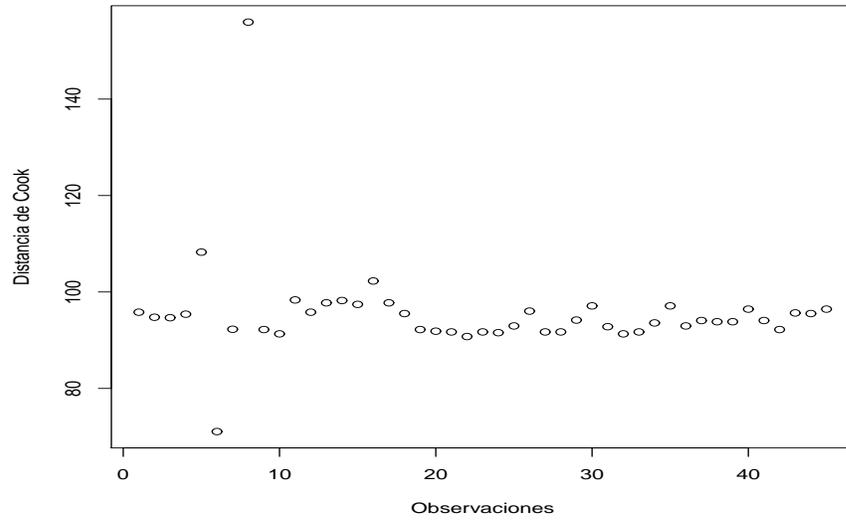
$\hat{\sigma}_e$	SAR <sub>r</sub>	SAR <sub>f</sub>	$\tau$	F <sub>MSBAD</sub>	$P(G > F_1)$
1.2271	27.2880	127.8938	1.16824	57.14	0.0000

Fuente: autor(2013)

En esta tabla se observa que  $SAR_r > SAR_f$  lo cual significaría que el modelo estimado explica más del 100% de los datos, lo cual es imposible y por tanto MINSADBAD no sería un método apropiado para

este conjunto de datos. No obstante se realiza el diagnóstico gráfico de este procedimiento utilizando el programa en Matlab, expuesto en el apéndice E3.





En estas gráficas se observan varios aspectos a describir:

- Todos los residuos son negativos.
- Persiste la heterocedasticidad.

- Existen tres puntos influyentes.
- Se presentan tres valores extremos de la variable respuesta que pueden ser outliers.

Se realizó el análisis para el mismo conjunto de datos, usando el método MINSADBED, obteniendo los resultados de la siguiente tabla:

Cuadro 7: Estimación MINSADBED

Parámetro	Estimación	Desv.Tip	Est.T	P	Significancia
$\beta_0$	2.2062	0.2270	9.7189	0.00000	***
$\beta_1$	0.6860	0.4231	1.6213	0.1122	
$\beta_2$	0.0217	0.007424	2.9230	0.0055	*
$\beta_3$	0	0.0428	0	1	

Fuente: autor(2013)

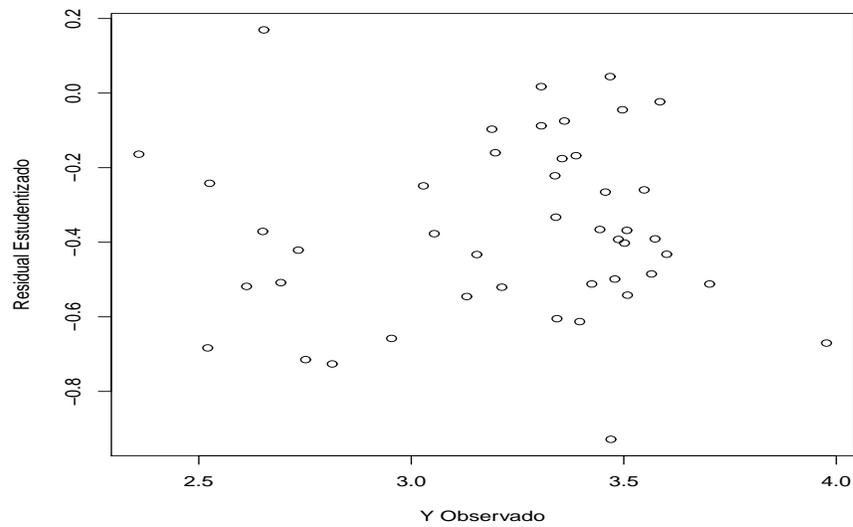
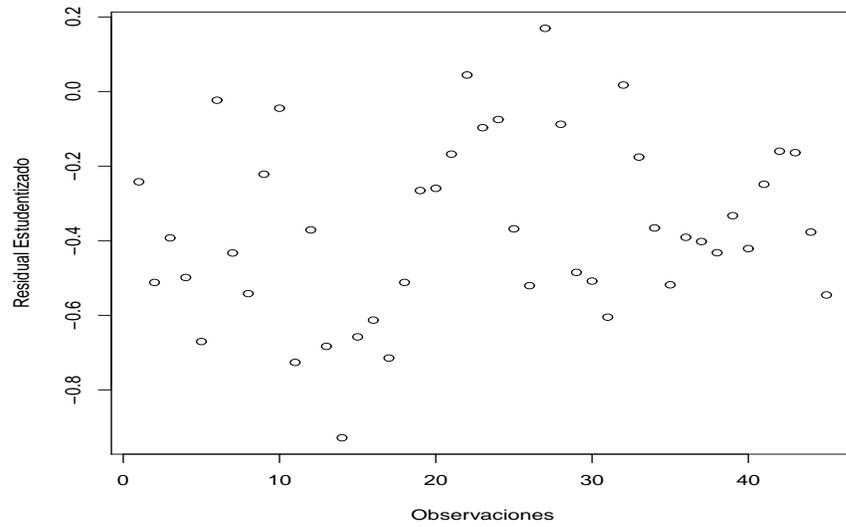
Se observa el modelo estimado  $\hat{Y} = 2,2062 + 0,0217 X_2$ , las variables independientes  $X_1$  y  $X_3$  se excuyen del modelo pues no son significantes. La desviación típica residual es  $\hat{\sigma}_e = 0,6056$ , la cual sigue siendo mayor que la obtenida con regresión MINMAD. Los estadísticos básicos para este tipo de estimación son:

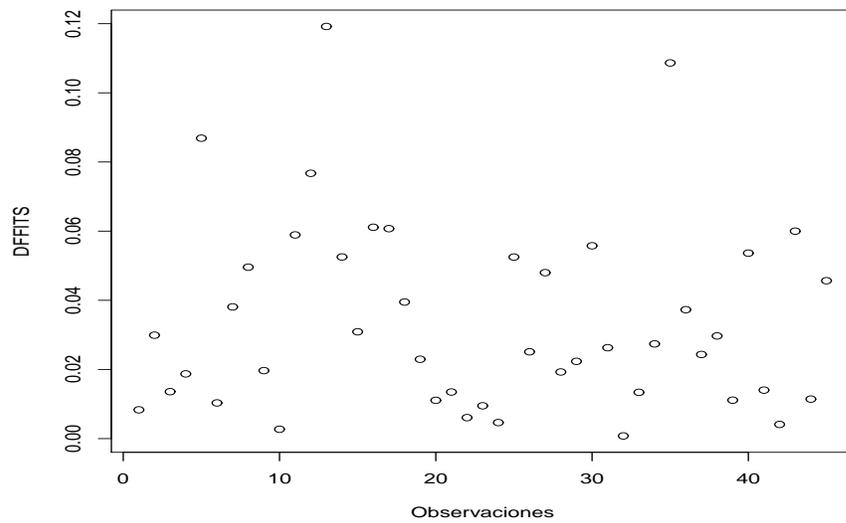
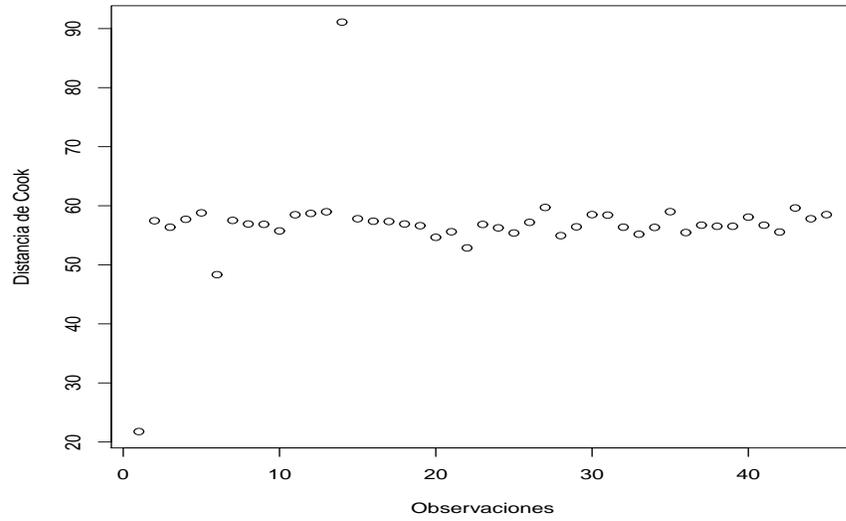
Cuadro 8: Estadísticas básicas MINSADBED

$\hat{\sigma}_e$	SAR <sub>r</sub>	SAR <sub>f</sub>	$\tau$	F <sub>MSBAD</sub>	P(G > F <sub>1</sub> )
0.6056	27.2880	47.8745	0.6436	21.3243	0.0000

Fuente: autor(2013)

Al igual que el método MINSADBAD  $SAR_r > SAR_f$ , lo cual evidencia sobreestimación del modelo y por tanto el método no es recomendable. Los gráficos para el diagnóstico de este método fueron obtenidos con base en los resultados del programa en Matlab del apéndice E4.





A diferencia del método anterior se puede describir:

- Hay más residuos negativos que positivos.
- No se evidencia heterocedasticidad.

- Existen dos valores influyentes.
- Se presentan dos valores extremos de la variable respuesta que pueden ser outliers.

## 8. Conclusiones

1. El método de regresión MINMAD es bastante robusto, estable en error cuadrático medio, presenta un alto grado de cobertura frente a los estimadores bayesianos del modelo, obtenidos mediante el procedimiento MCMC (Métodos Montecarlo y Cadenas de Markov). Además es bastante robusto cuando la distribución de los errores es simétrica de colas pesadas. En la aplicación propuesta en este trabajo, se evidencia la similitud de los estimadores a los de un Modelo Lineal Generalizado.
2. El método de regresión MINMAXAD, aunque parece flexible ante el manejo de pocas variables, resultó tener demasiada inestabilidad en error cuadrático medio y en porcentaje de cobertura. Estas características lo hacen inaplicable a cualquier conjunto de datos a ser procesados mediante regresión robusta.
3. El método de regresión MINMAD presenta una excelente flexibilidad en su manejo computacional, cuando la distribución de los errores presenta una alta elongación e incluso cuando la varianza es infinita, no obstante cuando la distribución de los errores es altamente asimétrica, el programa que permite calcular la estimación de los parámetros aborta, es decir no converge a ningún resultado.
4. Los métodos de regresión robusta MINSADBAD y MINSADBED, pueden considerarse como modificaciones alternas del método MINMAD, esto permite utilizar las mismas fórmulas para la aplicación inferencial de los parámetros, es decir estos dos métodos se pueden considerar casos particulares con su propia metodología de aplicación del método MINMAD.
5. Los métodos MINSADBAD y MINSADBED a diferencia del método MINMAD, presentan una buena flexibilidad computacional en su aplicación, cuando la distribución de los errores es altamente asimétrica, pero cuando la distribución de los errores tiene una alta elongación, no convergen a ningún resultado.
6. Cuando la distribución de los errores es simétrica de colas pesadas, la calidad de la estimación

de los parámetros del modelo mediante MINMAD, es tan buena como la estimación obtenida mediante la aplicación de un Modelo Lineal Generalizado, método bastante robusto que permite obtener estimaciones de máxima verosimilitud, usando el algoritmo de *Newton-Raphson-Fisher*.

7. La Estadística Bayesiana, utilizada en Modelos de Alta Complejidad, mediante la utilización de Métodos Montecarlo y Cadenas de Markov (MCMC), es la alternativa más moderna y poderosamente robusta, que permite obtener no sólo la estimación de los parámetros, sino su distribución aproximada y permite calcular las regiones de confiabilidad de cada parámetro, incluso en pequeñas muestras.

## 9. Recomendaciones

1. Se propone recomendar el uso de métodos alternativos, especialmente el método MINMAD, cuando existan datos cuyos residuos sean simétricos de colas pesadas, y no exagerar en el uso y abuso de *Mínimos Cuadrados*.
2. Es indispensable realizar pruebas de ensayo y error con los métodos MINSADBAD y MINSADBED en conjuntos de datos, los cuales representan un modelo cuyos residuos revelen demasiada asimetría en su distribución. No obstante es indispensable realizar el respectivo diagnóstico del modelo estimado.
3. Sería importante desarrollar muchas aplicaciones simuladas del método MINMAXAD, el cual demostró demasiadas falencias, y examinar en detalle las posibles ventajas en la utilización de este método..
4. El presente estudio se limitó a estudiar sólo los modelos de regresión lineal simple y múltiple con sólo dos covariables, se recomienda estudiar los mismos métodos con un mayor número de covariables y por que no en modelos que no sean lineales.
5. Proponer el estudio de estos mismos métodos mediante procedimientos matemáticos que conlleven a otros resultados (por que no más relevantes), por ejemplo las Metaheurísticas y métodos de Programación Dinámica.

## Referencias

- [1] Arthanari, T.S. and Dodge, Y. *Mathematical Programming in Statistics*. John Wiley & Sons Inc, New York, 1981.
- [2] Eisenhart, C. *Roger Joseph Boscovitch and the Combination of Observation in Whyte L.L.* Fordham University Press, New York, 1961.
- [3] Bloomfield, C. and Steiger, W.L. *Least Absolute Deviation. Theory, Applications and Algorithms*. Edited by Peter Huber, Birkhäus, Boston, 1983.
- [4] Dodge, Y. and Birkes, D. *Alternative Methods of Regression*. John Wiley & Sons, Inc, New York, 1993.
- [5] Graybill, F.A. *Theory and Application of the Linear Models*. Duxbury/Brooks Cole, Toronto, 1976.
- [6] Graybill, F.A. *An Introduction to Linear Statistical Models*. McGraw-Hill, New York, 1961.
- [7] Seber, G.A.F. *Linear Regression Analysis*. Wiley, New York, 1977.
- [8] Weisberg, S. *Applied Linear Regression*, 2nd. ed. Wiley, New York, 1985.
- [9] Seber, G.A.F., and Wild, C.J. *Nonlinear Regression*. Wiley, New York, 1989.
- [10] McCullagh, P. and Nelder, J.A. *Generalized Linear Models*, 2nd.ed. Chapman & Hall, London, 1989.
- [11] Huber, P.J. *Robust Statistics*. Wiley, New York, 1981.
- [12] Hastie, T.J. and Tibshirani, R.J. *Generalized Additive Models*. Chapman & Hall, London, 1990.
- [13] Draper, N.R. and Smith, H. *Applied Regression Analysis*, 2nd. ed. Wiley, New York, 1981.

- [14] Dodge, Y. (1984). Robust estimation of regression coefficients by minimizing a convex combination of least squares and least absolute deviations. *Computational Statistics Quarterly*, vol. 1, pp. 139-153.
- [15] Paula, G.A. *Modelos de Regresión com apoio computacional*. Instituto de Matemática e Estatística. Universidade de Sao Paulo. Versión preliminar. Sao Paulo, 2004.
- [16] Cook, R.D.(1987) Influence assesment. *Journal of Applied Statistics*. **14**, 117-131.
- [17] Cook, R.D; Peña, D. and Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics, Theory and Methods*. **17**, 623-640.
- [18] Lawrence, A.J. (1988). Regression transformation diagnostics using local influence, *Journal of the American Statistical Asociation*, **84**, 125-141.
- [19] Tsai, C.H. and Wu, X. (1992). Assesing local influence in linear regression models with first-order autorregresive or heteroscedastic error structure. *Statistics and Probability Letters*, **14**, 247-252.
- [20] Beckman, R.J.; Nachtsheim, C.J. and Cook, R.D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413-426.
- [21] Petit, A.N. and Bin Daud, I. (1989). Case-weight measures of influence for proportional hazards regression. *Applied Statistics*, **38**, 51-67.
- [22] Escobar, L.A. and Meeker, W, Q. (1992). Assesing influence in regression analysis with censored data. *Biometrics*, **48**, 507-528.
- [23] Kim, M.G. (1995). Local influence in multivariate regression. *Communications in Statistics, Theory Methods*, **20**, 1271-1278.
- [24] Galea, M; Paula, G.A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statician*, **46**, 71-79.

- [25] Galea, M.; Paula, G.A. and Uribe-Opazo, M. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers*, **44**, 23-45.
- [26] Kwan, C.W. and Fung, W.K. (1998). Assessing local influence for specific restricted likelihood: Applications to factor analysis. *Psychometrika*, **63**, 35-46.
- [27] Gu, H. and Fung, W.K. (1998). Assessing local influence in canonical correlation analysis. *Annals of the Institute of Statistical Mathematics* **50**, 755-772.
- [28] Paula, G.A. (1996). Influence diagnostics in proper dispersion models. *Australian Journal of Statistics*, **38**, 307-316.
- [29] Svetliza, C.F. and Paula, G.A. (2001). On diagnostics in log-linear negative binomial models. *Journal of Statistical Computation and Simulation*, **71**, 231-244.
- [30] Svetliza, C.F. and Paula, G.A. (2003). Diagnostics in nonlinear negative binomial models. *Communications in Statistics, Theory Methods*, **32**, 1227-1250.
- [31] Fung, W.K. and Kwan, C.W. (1997). A note on local influence based on normal curvature. *Journal of the Royal Statistical Society, B* **59**, 839-843.
- [32] Belsley, D.A., Welsch, R.E. and Kuh, E. *Regression diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York, 1980.
- [33] Charnes, A., Cooper, W.W. and Ferguson, R.O. (1955). Optimal Estimation of Executive Compensation by Linear Programming. *Management Science*, Vol. 1, **2**, 138-151.
- [34] Karst, O.J. (1958). Linear Curve Fitting Using Linear Programming. *J. Am. Stat. Assoc.*, **53**, 118.
- [35] Stiefel, E. (1960). Note on Jordan elimination, linear Programming and Tschebyscheff. *Numerische Mathematik*, **2**, 1-17.

- [36] Wagner, H.M. (1959). Linear Programming Techniques for Regression Analysis. *J. A. Stat. Assoc.*, **54**, 206.
- [37] Barrodale, I. and Roberts, F.D.K. (1973). An Improved Algorithm for Discrete  $l_1$  Linear Approximation. *SIAM J. Numer. Anal.*, **10**, 839.
- [38] Jimenez, J.A. and Anurachalam, W. (2011). Using Tukey's  $g$  and  $h$  family of distributions to calculate value-at-risk and conditional value-at-risk. *The Journal of Risk*, **13**, 95-116.
- [39] Torres, J.C. y Pulido, L. (2001). Comparación de tres métodos de Regresión Lineal usando procedimientos de simulación. *Revista Colombiana de Estadística*, **24**, 33-43.
- [40] Huber, P. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*. **35**, 73-101.
- [41] Huber, P. *Robust Statistics*. Wiley, New York, 1981.
- [42] Acuña, E. (2011). Análisis de Regresión. Departamento de Matemáticas. *Universidad de Puerto Rico*. Recinto Universitario de Mayagüez, 1-245.
- [43] Rousseeuw, P.J. and Leroy, W.A. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.
- [44] Hampel, F.R., Ronchetti, M., Rousseeuw, P.J. and Stahel, W.A. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.
- [45] Arthanari, T.S., Dodge, Y., Nazemi, I. and Panosian, E. (1977). *On Performance of Five Methods of Estimation: A Simulation Study*. School of Planning and Computer Applications, Tehran.
- [46] Armstrong, R.D., Elam, J.J., and Hultz, J.W. (1977). Obtaining Least Absolute Value Estimates for a Two-Way Classification Model. *Commun. Stat. Simul. Comput.* **6**(4), 365.
- [47] Ashar, V.G., and Wallace, T.D. (1963). A Sampling Study of Minimum Absolute Deviations Estimator. *Oper. Res.* **11**, 477.

- [48] Ntxoufras, I. *Bayesian Modeling using WinBUGS*. Wiley Series in Computational Statistics, John Wiley & Sons, Inc., New York, 2009.
- [49] Hoffmann-Jorgensen, J. *Probability with a View Towards Statistics*, Vol. 1, Probability Series, Chapman & Hall, New York, 1994.
- [50] Anderson, H. (1986), “Metropolis, Monte Carlo and the MANIAC”, *Los Alamos Science* pp. 96-107.
- [51] Bernardo, J. and Smith, A. *Bayesian Theory*, Wiley, Chichester, UK. 1994.
- [52] Metropolis, N. and Ulam, S. (1949), “The Monte Carlo method”, *Journal of the American Statistical Association*, **44**, 335-341.
- [53] Givens, G. and Hoeting, J. *Computational Statistics*, Wiley Series in Probability and Statistics, Wiley, Hoboken, N.J., 2005.
- [54] Metrópolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953), “Equations of state calculation by fast computing machine”, *Journal of Chemical Physics*, **21**, 1087-1092.
- [55] Gelfand, A. and Smith, A. (1990), “Sampling-based approaches to calculating marginal densities”, *Journal of the American Statistical Association*, **85**, 398-409.
- [56] Hastings, W. (1970), “Monte Carlo sampling methods using Markov chains and their applications”, *Biometrika*, **57**, 97-109.
- [57] Gelfand, A., Hills, S., Racine-Poon, A. and Smith, A. (1990), “Illustration of Bayesian inference in normal data models using Gibbs sampling”, *Journal of the American Statistical Association*, **85**, 972-985.
- [58] Tanner, M. and Wong, W. (1987), “The calculation of the posterior distributions by data augmentation”, *Journal of the American Statistical Association*, **82**, 528-549.

- [59] Gilks, W., Richardson, S. and Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, Chapman & Hall, Suffolk, UK., 1996.
- [60] Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. *BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual*, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK., 1996.
- [61] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. *WinBUGS User Manual, Version 1.4*, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK, available at [ac.uk/bugs.http://www.mrc-bsu.cam.ac.uk/bugs](http://www.mrc-bsu.cam.ac.uk/bugs), 2003.
- [62] Zambrano, A.J.y Orjuela, A. *Una Aplicación en el Monitoreo de Perfiles para el bajo peso en recién nacidos en el Departamento del Tolima 2003-2007*, Tesis de Especialista en Estadística, Fundación Universitaria Los Libertadores, Bogotá, D.C., 2012.

# Apéndices

## A. Apéndice de Tablas

Cuadro 9: Distribuciones  $h$  de Tukey simuladas con su varianza teórica

h	V(e)	h	V(e)
0	1	0.6	Infinita
0.2	2.1517	0.8	Infinita
0.4	11.1803	1	Infinita

Fuente: Torres (2001, p. 36)

Cuadro 10: Estimaciones Regresión Lineal Simple

Distribución		$\beta_0$			$\beta_1$		
		MC	NP	LAD	MC	NP	LAD
normal		1.1716	1.2739	1.8169	0.0448	0.0472	0.0688
h	0.2	2.8079	1.8567	2.2316	0.1047	0.0708	0.0869
	0.4	12.2780	2.4435	2.6841	0.3529	0.0972	0.1051
	0.6	60.3540	3.3256	3.2183	2.6535	0.1332	0.1269
	0.8	828.1600	4.2417	3.2424	27.5080	0.1738	0.1318
	1	5432.562	5.4006	3.6525	174.5601	0.2148	0.1440
% outliers	5	1.7075	1.5531	2.1669	0.0652	0.0588	0.0827
	10	2.1565	1.8126	2.4468	0.0835	0.0687	0.0938
	15	2.7169	1.9623	2.4529	0.1058	0.0755	0.0965
	20	3.4506	2.5211	3.0297	0.1345	0.0963	0.1164

Fuente: Torres (2001, p. 38)

Cuadro 11: Resultados de regresión simple  $n = 50$

Distribución		$\beta_0$			$\beta_1$		
		MC	NP	LAD	MC	NP	LAD
normal		0.4407	0.483	0.7151	0.0154	0.0166	0.0248
h	0.2	1.0208	0.676	0.7780	0.0356	0.0235	0.0271
	0.4	3.4151	0.7573	0.7885	0.1207	0.0260	0.0271
	0.6	17.2182	0.9823	0.8817	0.7438	0.0341	0.0306
	0.8	236.1500	1.1016	0.8631	7.7681	0.0382	0.0295
	1	1546.770	1.3799	0.9745	49.4998	0.0485	0.0338
% outliers	5	0.5962	0.5575	0.7769	0.0207	0.0187	0.0267
	10	0.8799	0.6575	0.8821	0.0307	0.0226	0.0309
	15	0.99689	0.6949	0.9055	0.0342	0.0238	0.0316
	20	1.1307	0.7779	1.0061	0.0391	0.0271	0.0359

Fuente: Torres (2001, p. 38)

Cuadro 12: Distribuciones Familia de Tukey

Name of distribution	Parameters	Estimate values			
		A	B	g	h
Cauchy	$\mu, \sigma > 0$	$\mu$	$\sigma$	0	1
Double exponential	$\alpha, \beta > 0$	$\mu$	$\sigma$	0	1
Exponential	$\lambda > 0$	$(1/\lambda) \ln 2$	$g/\lambda$	0,773	-0,09445
Logistic	$\alpha, \beta > 0$	$\alpha$	$\beta$	0	0,0017771
Lognormal	$\mu, \sigma^2, C > 0$	$C^\mu$	$gC^\mu$	$\sigma \ln C$	0
Normal	$\mu, \sigma^2$	$\mu$	$\sigma$	0	0
$t_{10}$	$v = 10$	0	1	0	0,057624

Fuente: Jiménez(2011, p. 100)

Cuadro 13: Modelo de Regresión Lineal Simple

Coef.		$\beta_0$				$\beta_1$			
n		10	20	50	100	10	20	50	100
normal		3.49	3.18	3.03	3.41	5.20	4.89	5.21	5.16
h	0.2	3.62	3.78	3.61	3.67	5.20	4.89	5.18	5.06
	0.4	3.95	3.13	4.14	3.96	5.66	5.16	4.89	4.98
	0.6	3.78	4.01	3.00	3.97	5.17	4.59	4.70	5.12
	0.8	5.54	3.20	4.20	3.86	4.61	5.09	5.18	5.29
	1	3.08	5.45	5.93	7.58	5.18	4.56	4.18	4.25
g	0.2	4.04	4.00	4.28	4.77	4.94	4.99	4.92	4.88
	0.4	2.63	5.82	4.69	4.09	5.68	4.47	4.82	4.97
	0.6	3.77	3.87	4.59	3.77	5.11	5.10	4.93	5.10
	0.8	3.88	3.83	3.92	4.10	5.10	5.16	5.17	5.08
	1	6.19	3.67	3.50	4.50	4.52	5.17	5.33	4.96
% NC	5	2.58	4.51	4.04	4.02	5.43	4.88	4.97	4.97
	10	3.53	3.78	3.87	3.89	4.98	4.99	4.98	4.97
	15	2.64	3.49	3.59	3.56	5.45	5.11	5.08	5.14
	20	3.81	4.23	4.43	4.41	5.08	5.05	4.85	4.91

Cuadro 14: Modelo de Regresión Lineal Múltiple

Coef.		$\beta_0$				$\beta_1$				$\beta_2$			
n		10	20	50	100	10	20	50	100	10	20	50	100
normal		3.20	3.34	2.88	2.72	4.00	4.05	3.88	4.25	4.52	4.87	5.07	4.89
h	0.2	3.19	2.50	3.09	2.94	3.85	4.40	3.79	4.02	5.05	5.15	5.11	4.95
	0.4	2.60	1.51	2.80	3.20	4.12	4.72	4.30	4.08	5.00	4.99	4.85	5.29
	0.6	2.81	2.63	2.81	2.10	3.87	3.70	4.38	4.60	5.14	5.42	4.79	5.10
	0.8	2.54	2.79	2.33	2.58	3.89	3.88	4.03	3.60	5.21	5.08	5.11	5.40
	1	3.20	2.64	2.78	2.40	4.01	4.30	4.55	3.81	5.02	4.99	4.58	5.80
g	0.2	2.44	3.50	2.86	2.36	4.04	4.02	3.91	4.08	5.08	4.79	5.14	5.11
	0.4	3.25	3.03	2.79	3.50	3.84	4.18	4.22	3.95	5.07	4.90	4.95	4.91
	0.6	2.83	2.50	2.53	3.25	4.31	3.82	3.75	3.71	4.82	5.43	5.34	5.20
	0.8	3.40	2.85	2.63	3.37	4.03	4.37	4.20	3.88	4.77	5.07	5.06	5.13
	1	3.50	3.50	6.72	4.84	4.04	2.47	4.71	3.92	4.79	6.34	3.20	4.69
% NC	5	3.20	2.46	3.16	2.28	3.54	3.45	4.05	4.11	5.30	5.70	4.79	5.14
	10	3.46	3.32	2.45	3.09	3.66	4.50	4.37	4.01	5.16	5.90	5.11	4.94
	15	2.32	2.81	2.87	2.88	3.68	3.80	3.78	4.13	5.55	5.23	5.35	4.95
	20	2.34	2.47	2.63	2.61	4.17	4.18	3.93	4.26	4.06	5.01	5.11	4.98

Cuadro 15: Modelo de Regresión Lineal Simple  $n = 10$

Coef.		$\beta_0$		$\beta_1$	
		IC 2.5	IC 97.5	IC 2.5	IC 97.5
normal		-6.37	13.6	2.10	8.33
h	0.2	0.48	6.78	4.25	6.15
	0.4	-3.48	7.708	4.21	7.11
	0.6	-1.91	9.85	3.40	6.86
	0.8	-3.82	15.0	1.47	7.77
	1	-4.6	11.0	2.74	7.63
g	0.2	0.65	7.50	3.86	6.00
	0.4	0.24	5.07	4.85	6.47
	0.6	-0.78	8.61	3.78	6.37
	0.8	0.09	7.90	3.96	6.24
	1	2.07	10.2	3.27	5.80
% NC	5	-0.50	5.87	4.60	6.22
	10	0.14	6.92	4.06	5.93
	15	-1.5	7.13	4.32	6.54
	20	2.62	5.06	4.74	5.42

Cuadro 16: Regresión Lineal Simple  $n = 20$

Distribución		$\beta_1$
		R.C.B
normal		[2.50;7.22]
% outliers	5	[4.13;5.60]
	10	[4.23;5.75]
	15	[4.59;5.66]
	20	[4.17;5.90]
g	0.2	[4.38;5.60]
	0.4	[3.68;5.62]
	0.6	[4.44;5.98]
	0.8	[4.22;6.10]
	1	[4.44;5.85]
h	0.2	[3.84;5.92]
	0.4	[4.02;5.15]
	0.6	[2.49;6.66]
	0.8	[3.45;6.72]
	1	[3.50;5.64]

Fuente: autor(2013)

Cuadro 17: Regresión Lineal Simple  $n = 50$

Distribución		$\beta_1$
		R.C.B
normal		[3.99;4.99]
% outliers	5	[4.69;5.42]
	10	[4.53;5.75]
	15	[4.70;5.47]
	20	[4.48;5.16]
g	0.2	[4.46;5.34]
	0.4	[4.92;5.23]
	0.6	[4.55;5.32]
	0.8	[4.61;5.73]
	1	[4.65;5.32]
h	0.2	[4.65;5.84]
	0.4	[3.72;6.06]
	0.6	[4.12;5.27]
	0.8	[4.14;6.21]
	1	[1.85;6.60]

Fuente: autor(2013)

Cuadro 18: Modelo de Regresión Lineal Múltiple  $n = 10$

Coef.		$\beta_0$		$\beta_1$		$\beta_2$	
		IC 2.5	IC 97.5	IC 2.5	IC 97.5	IC 2.5	IC 97.5
normal		-6.3	13.6	2.10	7.26	0.97	9.32
h	0.2	0.60	7.52	3.12	4.60	3.86	6.12
	0.4	3.70	8.16	2.78	5.54	3.55	6.46
	0.6	2.08	8.53	2.91	4.87	3.60	6.56
	0.8	3.27	9.19	2.78	5.07	3.42	6.85
	1	2.34	4.06	2.89	4.90	3.89	7.02
g	0.2	1.23	6.57	2.94	5.20	3.93	6.19
	0.4	0.31	7.33	3.15	4.55	3.98	6.08
	0.6	0.67	6.63	3.44	5.23	3.88	5.76
	0.8	1.02	9.88	3.15	4.90	3.33	5.84
	1	0.69	7.63	3.05	5.07	3.76	5.77
% NC	5	1.92	8.12	2.14	4.96	4.08	6.49
	10	0.33	7.06	2.88	4.46	4.40	5.87
	15	0.61	4.90	2.93	4.28	4.73	6.33
	20	0.63	4.29	4.34	5.22	4.04	5.15

Cuadro 19: Regresión Lineal Múltiple  $n = 20$

Distribución		$\beta_1$	$\beta_2$
		R.C.B	R.C.B
normal		[2.84;4.16]	[4.44;6.00]
% outliers	5	[2.03;4.83]	[3.88;7.27]
	10	[2.51;6.00]	[-1.8;15.00]
	15	[2.63;6.46]	[4.62;5.75]
	20	[3.78;4.59]	[4.49;5.47]
g	0.2	[3.58;4.46]	[4.00;5.47]
	0.4	[3.67;4.74]	[4.25;5.48]
	0.6	[3.23;4.44]	[4.69;6.09]
	0.8	[2.42;10.25]	[4.23;5.84]
	1	[1.06;3.89]	[4.34;8.14]
h	0.2	[3.92;5.05]	[4.30;5.92]
	0.4	[1.96;7.63]	[1.99;8.18]
	0.6	[2.50;4.84]	[3.93;6.71]
	0.8	[1.98;5.04]	[3.56;7.20]
	1	[3.72;5.09]	[4.28;5.64]

Fuente: autor(2013)

Cuadro 20: Regresión Lineal Múltiple  $n = 50$

Distribución		$\beta_1$	$\beta_2$
		R.C.B	R.C.B
normal		[3.93;4.58]	[4.46;5.26]
% outliers	5	[3.23;4.85]	[3.69;5.69]
	10	[3.43;5.34]	[3.78;6.28]
	15	[2.89;4.66]	[4.18;6.32]
	20	[3.63;4.22]	[4.72;5.44]
g	0.2	[3.59;4.21]	[4.77;5.45]
	0.4	[3.88;4.56]	[4.52;5.29]
	0.6	[3.41;4.09]	[4.93;5.72]
	0.8	[3.85;4.56]	[4.58;5.49]
	1	[2.52;4.01]	[3.66;5.50]
h	0.2	[3.37;4.20]	[4.59;5.55]
	0.4	[3.83;4.78]	[4.34;5.29]
	0.6	[3.80;4.99]	[4.14;5.34]
	0.8	[3.64;4.96]	[4.36;5.74]
	1	[3.68;5.69]	[3.49;5.51]

Fuente: autor (2013)

Cuadro 21: Regresión simple Minmad,  $n = 10$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		0.1343	6.4272	0.1974	23.80
% outliers	5	4.7927	4.7142	0.0225	70.10
	10	2.0876	5.7038	0.0216	72.20
	15	3.7632	5.1393	0.0270	64.50
	20	3.0759	5.0321	0.0443	53.30
g	0.2	0.00	6.3876	0.0518	49.50
	0.4	2.3871	5.4861	0.1972	31.80
	0.6	7.0738	4.0804	0.2851	31.40
	0.8	1.664	5.810	0.2717	40.30
	1	3.2793	5.0230	0.4891	32.60
h	0.2	4.8653	4.7887	0.0226	96.70
	0.4	5.0494	5.0647	0.0586	81.90
	0.6	4.6813	4.9132	9.0438	88.10
	0.8	2.7553	5.2365	0.0785	81.70
	1	6.8514	4.0608	0.0853	81.60

Fuente: autor (2013)

Cuadro 22: Regresión simple Minmad,  $n = 20$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		5.0747	4.5106	0.0403	99.40
% outliers	5	3.8526	4.8894	0.0061	100
	10	5.3016	4.6785	0.0061	100
	15	4.1989	5.0539	0.0069	100
	20	4.7560	4.7444	0.0124	100
g	0.2	3.8639	5.0225	0.0164	100
	0.4	7.5875	3.9177	0.0518	97.40
	0.6	7.3142	4.0644	0.0509	97.10
	0.8	0.3823	6.673	0.0890	92.60
	1	7.2309	4.2069	0.0976	92.80
h	0.2	3.1736	5.3187	0.0069	100
	0.4	4.5886	4.9020	0.0063	100
	0.6	4.7356	4.7181	0.0102	100
	0.8	4.5376	4.7537	0.0092	100
	1	2.6593	5.3180	0.0111	100

Fuente: autor (2013)

Cuadro 23: Regresión simple Minmad,  $n = 50$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		1.5096	5.7380	0.0056	93.40
% outliers	5	3.5922	5.0418	0.0011	100
	10	4.5650	4.9416	0.0011	100
	15	5.1485	4.6152	0.0014	100
	20	3.0617	5.3297	0.0015	100
g	0.2	6.7722	54.2720	0.0023	99.50
	0.4	4.2571	4.7621	0.0073	91.10
	0.6	3.2886	5.0746	0.0092	86.50
	0.8	4.3878	4.6995	0.0115	85
	1	5.6256	4.8969	0.0149	84.50
h	0.2	4.8276	4.7756	0.00094	100
	0.4	3.3374	5.1838	0.00091	100
	0.6	3.0402	5.2704	0.00079	100
	0.8	3.3266	5.2123	0.00013	100
	1	5.6914	4.5592	0.0010	99.80

Fuente: autor (2013)

Cuadro 24: Minmad Regresión Múltiple  $n = 10$

Distribución		$\beta_0$	$\beta_1$			$\beta_2$		
		Estimación	Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		4.1602	4.5350	0.1367	61.30	3.9911	0.1535	58.10
% outliers	5	4.4929	3.90325	0.0280	94.20	4.4694	0.0158	99.00
	10	6.9473	3.3054	0.0349	90.80	4.2222	0.02877	94.20
	15	3.2345	4.1963	0.0456	85.50	4.8276	0.0287	93.40
	20	2.6440	3.6338	0.0583	80.20	5.2743	0.0356	91.60
g	0.2	3.1511	4.0360	0.0363	90.50	4.8221	0.0336	91.40
	0.4	2.0202	3.7736	0.2891	57.40	5.1047	0.1811	66.30
	0.6	21.8592	5.1777	0.7354	49.40	0	0.3965	51.30
	0.8	30.3905	0	0.7836	48.60	0.7577	0.04253	60.10
	1	0	3.7814	0.9517	57.30	6.1222	0.4432	58.40
h	0.2	5.0512	4.0290	0.0468	87.10	4.4920	0.0390	89.50
	0.4	4.0254	4.3032	0.0506	88.00	4.3127	0.0531	87.20
	0.6	4.1724	3.3359	0.0747	79.90	5.0198	0.0477	88.50
	0.8	3.1918	4.0254	0.1615	76.50	4.9892	0.1533	76.40
	1	5.3875	3.9829	0.0847	83.80	4.3867	0.1214	77.80

Fuente: autor (2013)

Cuadro 25: Minmad Regresión Múltiple  $n = 20$

Distribución		$\beta_0$	$\beta_1$			$\beta_2$		
		Estimación	Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		1.7149	4.2704	0.0439	42.10	5.0487	0.0474	39.30
% outliers	5	3.1623	4.4580	0.0066	82.90	4.7144	0.0068	83.80
	10	1.5405	3.9669	0.0066	83.00	5.4753	0.0069	81.60
	15	2.7954	4.4078	0.0071	81.30	4.8039	0.0102	71.50
	20	2.4003	4.1262	0.0105	72.10	5.0227	0.0129	65.40
g	0.2	3.2494	4.6404	0.0122	68.10	4.6014	0.0123	69.50
	0.4	0	4.0458	0.0410	48.70	5.5728	0.0503	42.10
	0.6	2.0046	4.0009	0.0680	40.20	5.0955	0.0673	47.20
	0.8	1.8128	5.1176	0.0815	48.50	4.5968	0.0920	47.70
	1	6.1836	5.5449	0.1530	46.20	3.2767	0.1016	44.40
h	0.2	5.1539	4.0086	0.0069	83.00	4.3450	0.0085	76.10
	0.4	2.8038	3.9575	0.0069	84.60	5.0564	0.0056	87.30
	0.6	1.8233	4.2545	0.0078	83.70	5.0629	0.0099	75.30
	0.8	2.9232	4.1927	0.0096	77.80	4.7984	0.085	81.50
	1	7.2113	3.5569	0.0093	81.80	3.9221	0.0196	65.40

Fuente: autor (2013)

Cuadro 26: Minmad Regresión Múltiple  $n = 50$

Distribución		$\beta_0$	$\beta_1$			$\beta_2$		
		Estimación	Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		5.8583	3.4157	0.0076	58.70	4.7602	0.0080	57.60
% outliers	5	3.0449	4.0431	0.0012	95.60	4.9669	0.0013	95.80
	10	3.0304	3.9037	0.0012	96.80	5.1064	0.0012	95.60
	15	3.0566	3.7226	0.0013	95.40	5.1022	0.0012	95.90
	20	2.1473	3.7226	0.0017	91.80	5.0829	0.0017	91.50
g	0.2	0.0850	4.4270	0.0016	91.80	5.7273	0.0018	91.10
	0.4	0.5879	3.8962	0.0070	63.00	5.8632	0.0018	91.10
	0.6	3.4971	3.6641	0.0077	63.00	5.0361	0.0077	60.10
	0.8	1.8055	3.5478	0.0106	62.00	5.7266	0.0079	64.30
	1	3.8314	5.8410	0.0102	65.00	4.0128	0.0087	66.40
h	0.2	3.611	4.0171	0.0011	95.90	4.8435	0.00092	97.80
	0.4	2.3649	4.3617	0.0013	95.50	4.9007	0.0010	97.10
	0.6	4.8990	3.8278	0.00097	96.40	4.5877	0.0011	96.60
	0.8	0.7374	4.4661	0.0012	95.40	5.2332	0.0012	95.70
	1	3.5189	4.0127	0.0012	94.70	4.8759	0.0012	96.20

Fuente: autor(2013)

Cuadro 27: Regresión simple Minmaxad,  $n = 10$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		6.1886	5.0147	0.1818	29.50
% outliers	5	4.1058	4.9940	0.0151	75.10
	10	3.7345	5.0283	0.0212	66.50
	15	3.2015	5.0155	0.0241	64.70
	20	3.9539	5.0113	0.0150	76.00
g	0.2	3.2503	5.3957	0.2309	24.50
	0.4	2.0231	6.1437	0.7382	27.70
	0.6	0.3765	8.3614	9.7925	16.60
	0.8	37.9885	16.7126	589.3245	8.50
	1	671.9371	70.6021	1299	9.80
h	0.2	4.1929	4.9705	0.0335	60.10
	0.4	2.7919	5.0520	0.2021	43.50
	0.6	3.7193	5.0465	2.2664	32.10
	0.8	7.5801	4.0425	21.9893	32.30
	1	0.8486	6.7176	267.0801	23.90

Fuente: autor (2013)

Cuadro 28: Regresión simple Minmaxad,  $n = 20$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		3.8641	4.9954	0.0565	45.60
% outliers	5	4.3214	5.0092	0.0065	91.80
	10	4.1626	5.0146	0.0129	79.90
	15	4.0357	4.9996	0.0109	84.00
	20	4.7207	4.9951	0.0169	73.10
g	0.2	2.2872	5.7511	0.1453	32.20
	0.4	-2.3887	7.2206	1.4536	16.10
	0.6	-16.8448	12.9901	43.9115	6.90
	0.8	-87.2444	36.5878	15694	4.70
	1	-305.8155	115.8893	18655	2.60
h	0.2	3.7336	5.0167	0.0162	79.20
	0.4	3.8959	4.9698	0.1432	55.50
	0.6	4.9182	5.1519	1.5058	40.90
	0.8	8.1978	3.8027	28.99	31.30
	1	-3.0450	7.5286	284.4700	19.60

Fuente: autor (2013)

Cuadro 29: Regresión simple Minmaxad,  $n = 50$

Distribución		$\beta_0$	$\beta_1$		
		Estimación	Estimación	MSE	%Cob
normal		2.9442	4.9929	0.0254	64.10
% outliers	5	3.7854	4.9974	0.0037	98.40
	10	4.1260	4.9955	0.0045	96.50
	15	4.1079	4.9995	0.0057	93.70
	20	3.9788	4.9972	0.0064	91.90
g	0.2	1.3879	6.2399	0.2118	18.00
	0.4	-7.9272	9.3018	2.8773	2.40
	0.6	-42.2778	19.9133	51.2307	0.30
	0.8	-229.1764	76.9280	3184.6	0
	1	-1519.5	431.1459	72424	0
h	0.2	4.3354	4.9787	0.0207	78.10
	0.4	3.7716	5.0244	0.2115	50.90
	0.6	3.7846	5.0156	3.0893	28.30
	0.8	0.6242	4.7736	24.8024	16.50
	1	-35.4045	716.7339	72497	9.00

Fuente: autor (2013)

Cuadro 30: Minmaxad Regresión Múltiple  $n = 10$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		3.9646	0.1233	30.60	5.0042	0.1190	33.30
% outliers	5	4.0054	0.0197	69.20	5.0202	0.0139	78.40
	10	4.0010	0.0345	57.40	4.9757	0.0249	62.70
	15	3.9895	0.0521	43.90	4.9793	0.0268	61.60
	20	3.9847	0.0293	59.50	4.9958	0.0387	52.00
g	0.2	3.9201	0.4270	25.50	5.0622	0.4089	19.70
	0.4	3.3862	1.6880	25.60	5.5247	1.8283	21.00
	0.6	3.1919	26.4785	15.10	6.2104	19.3981	15.10
	0.8	8.6164	1807.6	9.40	0.9430	866.0063	7.30
	1	39.2911	36922	7.20	50.6893	51502	7.10
h	0.2	4.0300	0.0462	57.00	4.9971	0.0418	55.90
	0.4	4.0368	0.5309	40.00	5.0389	0.2238	46.90
	0.6	4.2253	1.6646	29.90	5.0472	0.5269	40.70
	0.8	3.8239	24.2723	24.70	4.9427	9.3491	26.60
	1	7.0291	503.0509	21.40	5.3604	376.0979	19.20

Fuente:autor(2013)

Cuadro 31: Minmaxad Regresión Múltiple  $n = 20$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		4.0546	0.1001	35.50	4.9452	0.0737	40.40
% outliers	5	3.9974	0.0105	84.30	4.9930	0.0140	78.30
	10	3.9929	0.0248	62.90	5.0110	0.0252	63.60
	15	4.0095	0.0164	72.20	4.9777	0.0148	76.50
	20	3.9969	0.0222	66.60	5.0202	0.0159	76.00
g	0.2	3.8356	0.2398	27.30	5.1676	0.2427	29.50
	0.4	4.1263	1.4011	20.20	4.9642	1.2126	21.40
	0.6	2.9866	13.8478	12.40	5.8076	13.1560	16.60
	0.8	11.7993	4244	6.50	5.4105	5038	7.50
	1	-39.8040	38341	3.10	69.8668	62005	3.90
h	0.2	3.9981	0.0467	54.50	5.0428	0.0420	58.10
	0.4	3.9657	0.1818	41.70	5.0624	0.1640	43.00
	0.6	3.9200	0.6884	25.50	5.0611	0.4793	33.70
	0.8	3.0173	98.5491	18.60	4.9513	100.2075	20.70
	1	6.5516	3716.70	13.30	3.7171	363.20	14.00

Fuente: autor(2013)

Cuadro 32: Minmaxad Regresión Múltiple  $n = 50$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		4.0082	0.0651	44.80	4.9856	0.0676	38.50
% outliers	5	3.9992	0.0072	91.10	4.9821	0.0081	88.80
	10	3.9921	0.0079	89.70	5.0050	0.0075	91.00
	15	3.9989	0.0097	86.50	4.9994	0.0097	85.50
	20	4.0118	0.0106	85.00	4.9762	0.0109	83.90
g	0.2	4.0152	0.1508	39.30	4.9798	0.1373	43.20
	0.4	3.5677	1.1718	19.80	5.4063	1.1068	21.30
	0.6	2.3756	58.1669	6.90	6.3168	48.0263	6.20
	0.8	-4.1886	5029.0	2.10	12.4359	2279.0	3.60
	1	-81.8915	451860	0.70	40.3480	92067	1.70
h	0.2	3.9663	0.0435	57.30	5.0296	0.0408	60.40
	0.4	3.9911	0.2303	31.30	5.0003	0.2322	32.10
	0.6	3.7952	2.4056	16.30	5.1378	2.2164	15.30
	0.8	4.1442	97.2468	9.10	4.8477	75.7355	11.30
	1	5.3863	152.0984	5.20	4.6177	201.1546	6.00

Fuente: autor(2013)

Cuadro 33: Regresión simple Minsadbed,  $n = 10$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		4.6980	0.0928	100
% outliers	5	5.0336	0.0133	100
	10	5.3557	0.0248	100
	15	5.2055	0.0228	100
	20	4.9669	0.0289	100
g	0.2	4.5875	0.1074	98.70
	0.4	4.7659	0.01792	95.40
	0.6	4.8112	0.1121	97.40
	0.8	4.8370	0.1144	97.70
	1	5.4602	0.1888	93.90

Fuente: autor (2013)

Cuadro 34: Regresión simple Minsadbed,  $n = 20$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		5.2999	0.0399	99.40
% outliers	5	5.1411	0.0040	100
	10	4.8666	0.0061	100
	15	5.4287	0.0064	100
	20	4.8044	0.0056	100
g	0.2	4.7426	0.0173	99.80
	0.4	4.7102	0.0154	99.60
	0.6	5.1138	0.0188	99.50
	0.8	4.9811	0.0245	98.50
	1	4.9905	0.0401	96.10

Fuente: autor (2013)

Cuadro 35: Regresión simple Minsadbed,  $n = 50$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		6.123	0.0246	5.60
% outliers	5	6.1063	0.0245	0
	10	6.0243	0.0239	0
	15	5.9995	0.0215	0.002
	20	6.0607	0.0221	0.001
g	0.2	6.0366	0.0221	0.082
	0.4	6.4032	0.0264	0.008
	0.6	6.3162	0.0326	0
	0.8	6.3256	0.0323	0.001
	1	6.0832	0.0254	0.001

Fuente: autor (2013)

Cuadro 36: Minsadbed Regresión Múltiple  $n = 10$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		4.4886	0.2380	24.10	4.5837	0.1042	38.70
% outliers	5	4.2970	0.0135	83.20	4.9677	0.0159	79.70
	10	3.8944	0.0201	73.30	4.4211	0.0126	85.30
	15	3.2346	0.0271	67.00	4.5038	0.0200	77.20
	20	43.6314	0.0363	62.40	5.0360	0.0202	77.70
g	0.2	3.6095	0.1714	32.60	4.5359	0.0881	42.20
	0.4	5.1974	0.1448	46.90	3.6636	0.1040	50.00
	0.6	3.8755	0.211	46.00	4.9639	0.1845	55.60
	0.8	3.9314	0.3303	45.70	5.5696	0.2847	43.20
	1	2.1781	0.5096	44.30	2.4972	0.4095	45.00

Fuente: autor (2013)

Cuadro 37: Minsadbed Regresión Múltiple  $n = 20$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		3.2442	0.0271	82.90	5.5399	0.0277	83.30
% outliers	5	4.2945	0.0031	99.90	4.9390	0.0039	99.90
	10	3.7647	0.0049	99.70	4.8543	0.0053	99.60
	15	3.9975	0.0060	99.70	4.9620	0.0054	99.80
	20	3.6131	0.0057	99.50	5.3437	0.0073	98.10
g	0.2	3.37446	0.0235	84.40	5.3362	0.0217	88.10
	0.4	4.2812	0.0235	87.00	4.7436	0.0340	86.60
	0.6	5.2743	0.0345	83.30	3.2792	0.0377	84.80
	0.8	5.3745	0.0438	84.00	1.9752	0.0438	85.00
	1	3.9127	0.0505	83.50	5.0659	0.0426	87.30

Fuente: autor (2013)

Cuadro 38: Regresión Múltiple Minsadbed,  $n = 50$ 

Distribución		$\beta_0$	$\beta_1$			$\beta_2$		
		Est	Est	ECM1	C	Est	ECM2	C
normal		7.1069	3.5951	0.0089	81.5	3.9727	0.0124	1.5
% outliers	5	3.5887	3.7972	0.00056474	61	5.0178	0.0014	36.5
	10	2.6458	4.1123	0.00099356	59.5	5.104	0.0012	33.5
	15	5.0556	3.7469	0.00063725	66.5	4.5738	0.00092162	17.5
	20	5.0143	3.5274	0.0014	71.5	4.7775	0.0011	22.5
g	0.2	5.2992	3.7533	0.0038	68	4.7949	0.003	15
	0.4	5.2744	3.8744	0.0028	57.5	5.1876	0.0061	34.5
	0.6	6.9693	4.3121	0.0026	52	4.9507	0.0024	40.5
	0.8	11.7571	3.9938	0.0024	54	5.0089	0.0028	39.5
	1	14.2402	3.8634	0.0079	57.5	5.0733	0.0021	36.5

Cuadro 39: Regresión simple Minsadbad,  $n = 10$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		4.6051	0.0384	100
% outliers	5	5.4252	0.0505	100
	10	61374	0.0624	100
	15	5.6767	0.0696	100
	20	5.6336	0.0521	100
g	0.2	5.3174	0.0519	100
	0.4	5.3927	0.0842	98.80
	0.6	5.5369	0.0878	98.80
	0.8	5.6330	253.6022	85.00
	1	5.7765	0.3569	97.70

Fuente: autor(2013))

Cuadro 40: Regresión simple Minsadbad,  $n = 20$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		6.2846	6.2846	100
% outliers	5	5.6303	0.0217	44.30
	10	5.7569	0.0355	43.70
	15	5.6750	0.127	65.10
	20	5.3988	0.0205	90.70
g	0.2	5.2259	0.0213	72.40
	0.4	5.5681	0.08272	0.20
	0.6	6.5051	0.0372	57.30
	0.8	6.9436	0.081	95.00
	1	5.7875	0.7578	70.40

Fuente: autor (2013)

Cuadro 41: Regresión simple Minsadbad,  $n = 50$

Distribución		$\beta_1$		
		Estimación	MSE	%Cob
normal		5,6113	0,0068	88.00
% outliers	5	5.6122	0.007	97.00
	10	5.4993	0.0071	95.00
	15	5.7962	0.0064	96.00
	20	5.5226	0.0059	95.00
g	0.2	5.4537	0.0044	96.00
	0.4	5.8002	0.0104	91.00
	0.6	5.4694	0.0101	91.00
	0.8	5.7195	0.0121	90.00
	1	5.7033	0.0249	90.00

Fuente: autor (2013)

Cuadro 42: Minsabbad Regresión Múltiple  $n = 10$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		6.6357	0.7950	67.40	2.9779	0.4389	97.70
% outliers	5	7.8738	0.3274	88.70	1.9423	0.1901	99.70
	10	4.7383	0.0706	100	5.1823	0.0319	100
	15	8.0878	0.8258	71.20	2.1004	0.3294	97.70
	20	4.0020	0.1916	94.30	5.4130	0.0778	100
g	0.2	6.4957	1.0762	70.60	3.7793	0.2814	98.00
	0.4	4.4750	1.4876	75.50	4.9747	0.3343	95.10
	0.6	3.8407	0.3085	91.50	5.5205	0.1240	98.50
	0.8	4.3644	0.9171	89.60	5.2542	0.1701	96.70
	1	3.6717	0.6818	89.50	5.7405	0.3301	93.80

Fuente: autor (2013)

Cuadro 43: Minsabbad Regresión Múltiple  $n = 20$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		5.1077	0.0614	0	3.3995	0.1281	53.10
% outliers	5	2.2233	0.1578	0	5.9512	0.0452	58.30
	10	4.961	0.0462	0	4.4759	0.0137	92.70
	15	4.5883	0.0173	0	4.9822	0.00001587	90.60
	20	3.8015	0.002	0	5.276	0.0038	95.00
g	0.2	4.1543	0.0012	0	5.8086	0.0327	98.50
	0.4	4.0105	0.0000055	0	5.6957	0.0242	65.60
	0.6	4.5659	0.016	0	5.3716	0.0069	87.40
	0.8	6.3907	0.2858	0	3.4065	0.127	89.80
	1	2.948	0.0553	0	6.1358	0.0645	13.30

Fuente: autor(2013)

Cuadro 44: Minsabbad Regresión Múltiple  $n = 50$

Distribución		$\beta_1$			$\beta_2$		
		Estimación	MSE1	%Cob	Estimación	MSE2	%cob
normal		2.9337	0.0227	0	4.7106	0.0017	80.00
% outliers	5	4.2223	0.00098869	0	4.9252	0.00011198	98.00
	10	3.9062	0.00017581	0	5.2515	0.0013	96.00
	15	4.6478	0.0084	0	4.6153	0.003	100
	20	4.5847	0.0068	0	4.6257	0.0028	94.00
g	0.2	3.9689	0.000019291	0	5.7826	0.0122	85.00
	0.4	6.3953	0.1148	0	3.5188	0.0439	43.00
	0.6	4.6527	0.0085	0	5.1756	0.000617	98.00
	0.8	4.0108	0.00000233	0	5.6097	0.0074	99.00
	1	4.9611	0.0185	0	5.4479	0.004	52.00

Fuente: autor (2013)

## B. Programas Estadística Bayesiana(OpenBUGS)

### B.1. Regresión Lineal Simple

```
model{

  # verosimilitud del modelo

  for (i in 1:n) {

    y[i]~ dnorm( mu[i], tau) # componente estocastica

    # link and linear predictor

    mu[i] <- beta0 + beta1 * x1[i]

  }

  # distribuciones apriori

  tau ~ dgamma ( 0.01 , 0.01)

  beta0 ~ dnorm (0,3)

  beta1 ~ dunif (2,5)

  # definicion de varianza

  s2<-1/tau

  s <-sqrt(s2)

  # calculo de la varianza muestral para la variable respuesta

  for (i in 1:n) { c.y[i]<-y[i]-mean(y[]) }

  sy2 <- inprod( c.y[], c.y[] )/(n-1)

  # calculo de la version Bayesiana R^2

  R2B <- 1 - s2/sy2

  # valor esperado de y

  typical.y <- beta0 + beta1 * mean(x1[])

}
```

```

}

INITS

list( tau=1, beta0=1, beta1=2.5)

DATA (LIST)

list(n=100,

x1=c(3.369036,2.638369,4.978520,4.390894,3.960217,2.594490,3.724941,2.512857,
4.779123,4.946744,3.960334,3.157394,2.139286,2.909687,2.092179,3.481447,
3.235954,3.785947,2.909188,2.649335,3.068273,2.977080,4.059418,3.831005,
3.820610,2.338611,3.755091,3.166892,4.938693,2.157227,3.460418,3.023992,
3.468771,4.302082,4.549101,2.661688,2.494754,4.439532,3.322084,3.844718,
3.858668,2.841569,4.288095,2.418747,4.122178,4.987347,3.458187,4.023094,
2.995630,4.779492,4.032437,3.869618,3.298186,2.977520,2.508340,2.233123,
2.638009,2.297929,4.522603,3.477912,3.469292,2.469749,2.744943,2.714848,
2.552606,4.917489,4.802364,4.498290,3.138148,2.424305,3.507171,3.217334,
4.221438,4.390198,4.885068,3.131517,2.952914,2.102326,4.780119,3.565400,
3.880853,2.327184,3.590379,2.097006,2.083146,2.523031,2.986532,3.395459,
3.627606,3.994757,4.464868,3.838983,4.066209,3.524064,2.518466,3.854188,
3.415106,2.589722,2.054058,4.745748),

y=c(20.911438,13.926528,29.557449,24.856243,25.474648,18.024376,23.616320,
16.827445,28.665728,33.742677,23.788233,19.206917,17.940784,20.376047,
10.961215,19.604431,19.232208,24.657263,18.455379,19.529388,19.407485,
17.466545,23.370772,23.661729,23.625807,19.837095,24.048531,26.649293,
23.325541,19.314997,21.806615,17.804434,19.442232,25.933845,30.462297,
15.254909,17.069147,30.216531,20.603790,23.162432,24.180963,21.104522,
29.418725,17.587800,21.610071,30.826588,21.863643,18.922138,16.259928,

```

28.480222,24.099171,21.623381,25.157509,18.128161,18.348143,11.357449,  
 12.584765,14.608542,28.744271,20.648025,17.360502,16.557327,20.843631,  
 19.666850,17.897292,27.259942,26.396685,27.942339,18.612465,22.698387,  
 23.736809,19.144932,24.673344,27.570602,24.716213,21.800097,16.046213,  
 17.508671,31.308398,22.889749,17.914193,22.040649,18.846272,16.263107,  
 6.067392,16.395247,21.526002,17.138402,25.009921,22.576589,29.709809,  
 24.671807,19.593384,25.191925,14.512331,22.440637,21.866986,13.739254,  
 13.645954,22.364616))

## B.2. Regresión Lineal Múltiple

```

model{
  # verosimilitud del modelo
  for (i in 1:n) {
    y[i]~ dnorm( mu[i], tau) # componente estocastica
    # link and linear predictor
    mu[i] <- beta0 + beta1 * x1[i] + beta2 * x2[i]
  }
  # distribuciones apriori
  tau ~ dgamma ( 0.01 , 0.01)
  beta0 ~ dnorm (0,2)
  beta1 ~ dunif (1,4)
  beta2 ~ dunif (2,5)
  # definicion de varianza
  s2<-1/tau

```

```

s <-sqrt(s2)

# calculo de la varianza muestral para la variable respuesta
for (i in 1:n) { c.y[i]<-y[i]-mean(y[]) }

sy2 <- inprod( c.y[], c.y[] )/(n-1)

# calculo de la version Bayesiana R^2

R2B <- 1 - s2/sy2

# valor esperado de y

typical.y <- beta0 + beta1 * mean(x1[]) + beta2 * mean(x2[])

}

INITS

list( tau=1, beta0=0, beta1=2, beta2=3)

DATA (LIST)

list(n=100,

x1=c(2.726786,1.210701,2.191515,3.418909,3.471436,3.367960,1.652081,2.598071,

2.790578,2.897141,3.216606,2.905180,2.047608,3.185786,1.011556,1.680415,

1.237320,3.363733,3.526320,1.244242,2.190324,1.237819,1.375986,1.808491,

2.865313,2.380941,1.068823,1.991547,3.555827,3.662744,2.724804,1.734620,

2.580241,2.200802,1.208111,1.036582,3.830443,3.224681,1.064499,3.741742,

3.919919,1.194220,3.573584,1.060416,2.432723,1.672227,2.806749,2.698222,

1.773456,1.957861,1.035968,2.600530,1.028772,1.200448,3.010708,1.072707,

3.949594,2.718692,3.307578,3.697287,2.151372,3.353857,3.739878,1.704622,

2.168416,3.414069,1.955535,3.298062,2.490829,1.498211,2.704712,2.718397,

1.615395,2.351690,3.802050,2.332879,3.680059,1.576711,2.283289,2.357411,

2.289802,2.851911,1.824298,2.799798,2.502588,1.093990,1.314889,1.064768,

2.750444,1.046763,1.293708,1.241020,3.489780,3.725710,2.490593,2.583232,

```

3.310457,3.408742,3.887595,2.227674),  
 $x_2=c$ (3.208643,4.686160,4.680358,2.561109,4.760384,2.871363,3.017710,4.927725,  
 4.502121,3.071495,4.654251,3.618172,3.562965,2.500819,4.611072,3.510341,  
 3.588345,2.884058,3.868227,2.736974,3.479042,4.428796,3.915051,2.681092,  
 2.098110,3.946128,2.117732,2.630932,2.802104,3.703967,3.343721,4.383313,  
 2.840064,3.338914,2.082062,2.178019,4.567134,3.506668,4.708884,4.956110,  
 3.854867,4.751532,4.437755,4.154916,2.323515,4.467364,4.092077,2.160920,  
 3.533430,4.501731,4.639138,4.366032,2.782976,3.134635,4.114264,3.392308,  
 4.974000,3.634704,2.984813,4.944685,2.365931,2.052924,2.863046,4.064647,  
 2.547703,2.629479,2.020121,2.786361,2.687341,2.743242,3.933449,3.633729,  
 3.788332,2.859656,3.633781,3.515543,2.032822,4.672569,2.460005,2.991986,  
 4.786734,2.755090,4.029458,2.191689,2.553415,3.159769,3.308293,4.118657,  
 3.534685,3.199253,4.494359,2.966876,2.532181,2.002699,2.261798,3.104813,  
 2.744403,4.809484,2.438697,2.001708),  
 $y=c$ (30.74408,22.30531,34.91612,27.23840,43.54390,34.54778,22.47031,32.10621,  
 38.57775,29.66503,41.04908,31.10935,27.50321,28.18103,34.10262,23.40192,  
 28.05114,31.78894,35.03924,20.21550,29.09439,26.31261,27.56980,25.92392,  
 26.69487,33.32311,21.36979,25.32200,30.93489,36.27814,36.39591,32.20101,  
 24.49631,25.02108,17.48424,15.90272,45.09695,33.06235,36.35443,45.95841,  
 34.51487,28.05995,35.54635,34.38546,23.36452,31.52497,39.17928,23.02236,  
 29.00015,35.94789,33.57187,35.83769,22.11580,29.91158,33.76747,29.58973,  
 47.11690,30.44498,32.41734,45.00516,22.73006,30.11377,37.25608,28.34880,  
 24.77123,34.34650,20.40647,35.04580,25.76625,20.65363,33.01909,32.78392,  
 30.08374,24.16115,35.75988,32.71895,24.16254,27.36703,30.95254,27.63749,  
 32.96217,25.96401,34.64889,27.60795,26.19792,22.93901,21.73065,27.73126,

29.01448,24.14208,36.07354,19.33274,32.00929,33.07673,24.86188,24.40409,  
28.77298,44.08836,32.08025,16.70128))

## C. Programas Matlab Regresión Robusta

### C.1. Minmad

```
n=10;

L1=[3,5];

L2=[4,6];

r1=0;

r2=0;

X1=unifrnd(1,4,n,1);

X2=unifrnd(2,5,n,1);

e=normrnd(0,3,n,1000);

%e1=0.05*e+0.95*(e/3);

%e2=0.10*e+0.90*(e/3);

%e3=0.15*e+0.85*(e/3);

%e4=0.20*e+0.80*(e/3);

%e5=((0.2).^(-1))*(exp(e*(0.1))-1);

%e6=((0.4).^(-1))*(exp(e*(0.4))-1);

%e7=((0.6).^(-1))*(exp(e*(0.6))-1);

%e8=((0.8).^(-1))*(exp(e*(0.8))-1);

%e9=((1).^(-1))*(exp(e*(1))-1);

%e10=(e/3).*exp((0.2).*(e/3).^2/2);

%e11=(e/3).*exp((0.4).*(e/3).^2/2);

%e12=(e/3).*exp((0.6).*(e/3).^2/2);

%e13=(e/3).*exp((0.8).*(e/3).^2/2);
```

```

e14=(e/3).*exp((1).*(e/3).^2/2);

for i=1:length(e)

    y=3+4*X1+5*X2+e14(:,i);

    A=[ones(n,1) X1 X2 eye(n) -eye(n)];

    f=[0;0;0;ones(n,1);ones(n,1)];

    b=y;

    Aeq=[];

    beq=[];

    lb=[zeros(1,23)];

    options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'Iter');

    [x fval exitflag output]=linprog(f,[],[],A,b,lb,[],[],options)

    if x(2)<L1(2) & x(2)>L1(1)

        r1=r1+1;

    end

    if x(3)<L2(2) & x(3)>L2(1)

        r2=r2+1

    end

    MSE1(:,i)=(x(2)-4).^2/n;

    MSE2(:,i)=(x(3)-5).^2/n;

end

Beta0=mean(x(1))

Beta1=mean(x(2))

Beta2=mean(x(3))

ECM1=mean(MSE1)

ECM2=mean(MSE2)

```

```
recub1=r1/1000
```

```
recub2=r2/1000
```

## C.2. Minmaxad

```
n=10;
```

```
r=0;
```

```
X1=unifrnd(1,4,n,1);
```

```
X2=unifrnd(2,5,n,1);
```

```
e=normrnd(0,3,n,1000);
```

```
%e1=0.05*e+0.95*(e/3);
```

```
%e2=0.10*e+0.90*(e/3);
```

```
%e3=0.15*e+0.85*(e/3);
```

```
%e4=0.20*e+0.80*(e/3);
```

```
%e5=((0.2).^(-1))*(exp(e*(0.1))-1);
```

```
%e6=((0.4).^(-1))*(exp(e*(0.4))-1);
```

```
%e7=((0.6).^(-1))*(exp(e*(0.6))-1);
```

```
%e8=((0.8).^(-1))*(exp(e*(0.8))-1);
```

```
%e9=((1).^(-1))*(exp(e*(1))-1);
```

```
%e10=(e/3).*exp((0.2).*(e/3).^2/2);
```

```
%e11=(e/3).*exp((0.4).*(e/3).^2/2);
```

```
%e12=(e/3).*exp((0.6).*(e/3).^2/2);
```

```
%e13=(e/3).*exp((0.8).*(e/3).^2/2);
```

```
%e14=(e/3).*exp((1).*(e/3).^2/2);
```

```

for i=1:length(e)
    y=3+4*X1+5*X2+e(:,i);
    X=[X1 X2];
    A1=[y' -y'];
    A2=[X' -X'];
    A=[A1;A2];
    f=[ones(n,1);ones(n,1)];
    b=[ones(n,1);zeros(n,1)];
    Aeq=[];
    beq=[];
    lb=[zeros(1,20)];
    options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'Iter');
    [x fval exitflag output]=linprog(f,[],[],A,b,lb,[],[],options)
end
mean(X(1))
mean(x(2))

```

### C.3. Minsadbed

```

n=20;
L1=[2.504,7.219];
L2=[3.8,5.41];
r1=0;
r2=0;
r=n*(n-1)/2;

```

```

X1=unifrnd(1,4,n,1);
X2=unifrnd(2,5,n,1);
e=normrnd(0,3,n,1000);
%e1=(0.05*e)+(1-0.05)*(e/3);
%e2=(0.10*e)+(1-0.10)*(e/3);
%e3=(0.15*e)+(1-0.15)*(e/3);
%e4=(0.20*e)+(1-0.20)*(e/3);
%e5=((0.2).^(-1)).*(exp((0.2).*e)-1);
%e6=((0.4).^(-1)).*(exp((0.4).*e)-1);
%e7=((0.6).^(-1)).*(exp((0.6).*e)-1);
%e8=((0.8).^(-1)).*(exp((0.8).*e)-1);
%e9=((1).^(-1)).*(exp((1).*e)-1);
%e10=(e/3).*exp((0.2).*((e/3).^2)/2);
%e11=(e/3).*exp((0.4).*((e/3).^2)/2);
%e12=(e/3).*exp((0.6).*((e/3).^2)/2);
%e13=(e/3).*exp((0.8).*((e/3).^2)/2);
%e14=(e/3).*exp((1).*((e/3).^2)/2);
for i=1:length(e)
    Y=3+4*X1+5*X2+e(:,i);
    X=[ones(n,1) X1];
    B=inv(X'*X)*X'*Y;
    H1=[ones(n-1,1) -eye(n-1)];
    H2=[zeros(n-2,1) ones(n-2,1) -eye(n-2)];
    H3=[zeros(n-3,2) ones(n-3,1) -eye(n-3)];
    H4=[zeros(n-4,3) ones(n-4,1) -eye(n-4)];

```

```

H5=[zeros(n-5,4) ones(n-5,1) -eye(n-5)];
H6=[zeros(n-6,5) ones(n-6,1) -eye(n-6)];
H7=[zeros(n-7,6) ones(n-7,1) -eye(n-7)];
H8=[zeros(n-8,7) ones(n-8,1) -eye(n-8)];
H9=[zeros(n-9,8) ones(n-9,1) -eye(n-9)];
H10=[zeros(n-10,9) ones(n-10,1) -eye(n-10)];
H11=[zeros(n-11,10) ones(n-11,1) -eye(n-11)];
H12=[zeros(n-12,11) ones(n-12,1) -eye(n-12)];
H13=[zeros(n-13,12) ones(n-13,1) -eye(n-13)];
H14=[zeros(n-14,13) ones(n-14,1) -eye(n-14)];
H15=[zeros(n-15,14) ones(n-15,1) -eye(n-15)];
H16=[zeros(n-16,15) ones(n-16,1) -eye(n-16)];
H17=[zeros(n-17,16) ones(n-17,1) -eye(n-17)];
H18=[zeros(n-18,17) ones(n-18,1) -eye(n-18)];
H19=[zeros(n-19,18) ones(1,1) -1];

H=[H1;H2;H3;H4;H5;H6;H7;H8;H9;H10;H11;H12;H13;H14;H15;H16;H17;H18;H19];

A1=[X1 X2 eye(n) zeros(n,r) zeros(n,r)]

A2=[zeros(r,2) H -eye(r) eye(r)];

AEje1=[A1;A2];

fEje1=[0;0;zeros(n,1);ones(r,1);ones(r,1)];

b1=Y;

b2=zeros(r,1);

bEje1=[b1;b2];

lb=[zeros(1,402)];

options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'Iter');

```

```

[x fval exitflag output]=linprog(fEje1,[],[],AEje1,bEje1,lb,[],[],options)

MSE1(:,i)=((x(1)-4).^2)/n;

MSE2(:,i)=(x(2)-5).^2)/n;

if x(1)<L1(2) & x(1)>L1(1)

    r1=r1+1;

end

if x(2)<L2(2) & x(2)>L2(1)

    r2=r2+1;

end

end

Beta1=mean(x(1))

Beta2=mean(x(2))

ECM1=mean(MSE1)

ECM2=mean(MSE2)

recub1=r1/1000

recub2=r2/1000

```

#### C.4. Minsadbad

```

n=20;

L1=[3,5];

L2=[4,6];

r1=0;

r2=0;

```

```

r=n*(n-1)/2;

X1=unifrnd(1,4,n,1);

X2=unifrnd(2,5,n,1);

e=normrnd(0,3,n,1000);

%e1=(0.05*e)+(1-0.05)*(e/3);
%e2=(0.10*e)+(1-0.10)*(e/3);
%e3=(0.15*e)+(1-0.15)*(e/3);
%e4=(0.20*e)+(1-0.20)*(e/3);
%e5=((0.2).^(-1)).*exp((0.2.*e)-1);
%e6=((0.4).^(-1)).*exp((0.4.*e)-1);
%e7=((0.6).^(-1)).*exp((0.6.*e)-1);
%e8=((0.8).^(-1)).*exp((0.8.*e)-1);
%e9=((1).^(-1)).*exp((1.*e)-1);
%e10=(e/3).*exp((0.2).*((e/3).^2)/2);
%e11=(e/3).*exp((0.4).*((e/3).^2)/2);
%e12=(e/3).*exp((0.6).*((e/3).^2)/2);
%e13=(e/3).*exp((0.8).*((e/3).^2)/2);
%e14=(e/3).*exp((1).*((e/3).^2)/2);

for i=1:length(e)

    Y=3+4*X1+5*X2+e(:,i);

    X=[ones(n,1) X1 X2];

    X3=[X1 X2];

    B=inv(X'*X)*X'*Y;

    YE=X*B;

    R=Y-YE;

```

```

H1=[ones(n-1,1) -eye(n-1)];
H2=[zeros(n-2,1) ones(n-2,1) -eye(n-2)];
H3=[zeros(n-3,2) ones(n-3,1) -eye(n-3)];
H4=[zeros(n-4,3) ones(n-4,1) -eye(n-4)];
H5=[zeros(n-5,4) ones(n-5,1) -eye(n-5)];
H6=[zeros(n-6,5) ones(n-6,1) -eye(n-6)];
H7=[zeros(n-7,6) ones(n-7,1) -eye(n-7)];
H8=[zeros(n-8,7) ones(n-8,1) -eye(n-8)];
H9=[zeros(n-9,8) ones(n-9,1) -eye(n-9)];
H10=[zeros(n-10,9) ones(n-10,1) -eye(n-10)];
H11=[zeros(n-11,10) ones(n-11,1) -eye(n-11)];
H12=[zeros(n-12,11) ones(n-12,1) -eye(n-12)];
H13=[zeros(n-13,12) ones(n-13,1) -eye(n-13)];
H14=[zeros(n-14,13) ones(n-14,1) -eye(n-14)];
H15=[zeros(n-15,14) ones(n-15,1) -eye(n-15)];
H16=[zeros(n-16,15) ones(n-16,1) -eye(n-16)];
H17=[zeros(n-17,16) ones(n-17,1) -eye(n-17)];
H18=[zeros(n-18,17) ones(n-18,1) -eye(n-18)];
H19=[zeros(n-19,18) ones(1,1) -1];
H=[H1;H2;H3;H4;H5;H6;H7;H8;H9;H10;H11;H12;H13;H14;H15;H16;H17;H18;H19];
A1=[-X3 -eye(n) zeros(n,r) zeros(n,r)];
A2=[X3 -eye(n) zeros(n,r) zeros(n,r)];
A3=[zeros(r,2) H eye(r) -eye(r)];
A4=[zeros(r,2) -H -eye(r) eye(r)];
AEje1=[A1;A2;A3;A4];

```

```

fEje1=[0;0;zeros(n,1);ones(r,1);ones(r,1)];

b1=-Y;

b2=Y;

b3=[zeros(r,1)];

b4=[zeros(r,1)];

bEje1=[b1;b2;b3;b4];

lb=zeros(1,402);

options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'Iter');

[x fval exitflag output]=linprog(fEje1,[],[],AEje1,bEje1,lb,[],[],options)

MSE1=(x(1)-4).^2/n;

MSE2=(x(2)-5).^2/n;

if x(1)<L1(2) & X(1)>L1(1)

    r1=r1+1

end

if x(2)<L2(2) & x(2)>L2(1)

    r2=r2+1

end

end

Beta1=mean(x(1))

Beta2=mean(x(2))

ECM1=mean(MSE1)

ECM2=mean(MSE2)

recub1=r1/1000

recub2=r2/1000

```

## D. Justificación cálculo de $\tau$ .

### D.1. Justificación de (35).

El parámetro  $\tau$  en regresión MINMAD es análogo al parámetro  $\sigma$  en estimación de mínimos cuadrados. La desviación estándar en una regresión por mínimos cuadrados tiene la forma  $\sigma c$ , mientras que la desviación estándar de la correspondiente regresión MINMAD es aproximadamente  $\tau c$  para la misma cantidad  $c$ . En particular cuando se considera el caso de una regresión con  $p = 0$ , es decir, no hay variables independientes y simplemente se tiene una muestra aleatoria de  $n$  observaciones independientes de la misma población  $e_1, e_2, \dots, e_n$ . La regresión por mínimos cuadrados estima la media muestral  $\bar{e}$ , a causa de que  $a = \bar{e}$  es el valor de  $a$  que da la mínima suma de cuadrados de las desviaciones  $\sum (e_i - a)^2$ . La desviación estándar es  $\sigma/\sqrt{n}$ . La regresión MINMAD estima la mediana muestral  $\tilde{e}$ , a causa de que  $a = \tilde{e}$  da la mínima suma de desviaciones absolutas  $\sum |e_i - a|$ . Su desviación estándar es aproximadamente  $\tau/\sqrt{n}$ .

El Teorema del Límite Central dice que para grandes tamaños de muestra  $n$ , la distribución de la media muestral es aproximadamente normal. De igual forma para grandes tamaños de muestra  $n$ , la mediana muestral es también aproximadamente normal. Por su parte, el valor esperado de la media muestral es igual a la media poblacional. También el valor esperado de la mediana muestral es aproximadamente igual a la mediana poblacional. Si  $\nu$  denota la mediana poblacional, se puede afirmar que para  $n$  grande,  $\tilde{e}$  tiene aproximadamente una distribución normal con media  $\nu$  y desviación estándar  $\tau/\sqrt{n}$ . En consecuencia si  $\hat{\tau}$  es una estimación de  $\tau$ , se puede construir un intervalo de confianza aproximado del 95% para  $\nu$  como  $\tilde{e} \pm 2\hat{\tau}/\sqrt{n}$ . Al estar interesado en  $\nu$  y no en su intervalo de confianza, si se puede considerar su longitud  $4\hat{\tau}/\sqrt{n}$ .

Se puede mostrar que el intervalo desde  $e_{(k_1)}$  hasta  $e_{(k_2)}$  es también un intervalo de confianza aproximado al 95% para  $\nu$ , así que los dos intervalos son similares, al menos, para grandes muestras. En

particular, sus longitudes son similares:  $4\hat{\tau}/\sqrt{n} \approx e_{(k_2)} - e_{(k_1)}$ . Escogiendo  $\hat{\tau}$ , para hacer de la aproximación una igualdad y reemplazando  $e$  por  $\tilde{e}$ ,  $n$  por  $m = n - 2$ , permite obtener la fórmula (35). Se ha encontrado que usando sólo  $m$  residuos distintos de cero, mejora el rendimiento de la prueba para pequeñas muestras.

## E. Programas en R para estimación GLM

### E.1. Programa GLM completo

```

Birkes1.dat=scan("G:/AYUDA/Tareas/ALFONSO/Birkes1.txt", what=list(Incendios=0,
lincendios=0, edad=0, robos=0, ingresos=0))

attach(Birkes1.dat)

fit1.Birkes=glm(Incendios~edad+robos+ingresos, family=Gamma(link=log))

summary(fit1.Birkes)

X <- model.matrix(fit1.Birkes)

n <- nrow(X)

p <- ncol(X)

w <- fit1.Birkes$weights

W <- diag(w)

H <- solve(t(X)%*%W%*%X)

H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)

h <- diag(H)

library(MASS)

fi <- gamma.shape(fit1.Birkes)$alpha

```

```

ts <- resid(fit1.Birkes,type="pearson")*sqrt(fi/(1-h))
td <- resid(fit1.Birkes,type="deviance")*sqrt(fi/(1-h))
di <- (h/(1-h))*(ts^2)
par(mfrow=c(2,2))
a <- max(td)
b <- min(td)
plot(fitted(fit1.Birkes),h,xlab="Valores Ajustados", ylab="Medida h", pch=16)
#title(sub="(a)")
#identify(fitted(fit1.Birkes), h, n=1)
#
plot(di,xlab="Indice", ylab="Distancia de Cook", pch=16)
#title(sub="(b)")
#identify(di, n=2)
#
plot(fitted(fit1.Birkes),td,xlab="Valores Ajustados", ylab="Componente do Desvio",
ylim=c(b-1,a+1),pch=16)
#title(sub="(c)")
abline(2,0,lty=2)
abline(-2,0,lty=2)
#identify(fitted(fit1.Birkes),td, n=1)
#
w <- fit1.Birkes$weights
eta <- predict(fit1.Birkes)
z <- eta + resid(fit1.Birkes, type="pearson")/sqrt(w)
plot(predict(fit1.Birkes),z,xlab="Preditor Linear",

```

```
ylab="Variavel z", pch=16)
lines(smooth.spline(predict(fit1.Birkes), z, df=2))
#title(sub="(d)")
```

### E.1.1. Programa GLM Reducido

```
Birkes1.dat=scan("G:/AYUDA/Tareas/ALFONSO/Birkes1.txt", what=list(Incendios=0,
lincendios=0, edad=0, robos=0, ingresos=0))
attach(Birkes1.dat)
fit1.Birkes=glm(Incendios~ingresos, family=Gamma(link=log))
summary(fit1.Birkes)
```

### E.1.2. Programa Diagnóstico Regresión MINSADBAD

```
load G:\AYUDA\Tareas\ALFONSO\Birkes1.txt
L=Birkes1;
n=length(L);
Y=L(:,2);
X1=L(:,3);
X2=L(:,4);
X3=L(:,5);
U=ones(n,1);
X=[U X1 X2 X3];
B=(inv(X'*X))*X'*Y;
I=eye(n);
```

```

H=X*(inv(X'*X))*X';
DH=diag(H);
R=I-H;
D=diag(R);
YE=[];
SSE=[];
t=[];
DC=[];
DFIT=[];
for i=1:n
    Li=L;
    Li(i,:)=[];
    ni=length(Li);
    ri=ni*(ni-1)/2;
    Yi=Li(:,2);
    X1i=Li(:,3);
    X2i=Li(:,4);
    X3i=Li(:,5);
    X4i=[X1i X2i X3i];
    Ui=ones(ni,1);
    Xi=[Ui X4i];
    Hi=[]; Fi=[];
    for j=1:n-1
        Hi=[Hi' [zeros(ni-j,j-1) ones(ni-j,1) -eye(ni-j)]']';
        Fi=[Fi' [zeros(ni-j,j-1) ones(ni-j,1) eye(ni-j)]']';
    end
end

```

```

end

Q=(1/2)*Fi;

C=Q*Yi;

A1i=[-X4i -eye(ni) zeros(ni,ri) zeros(ni,ri)];
A2i=[X4i -eye(ni) zeros(ni,ri) zeros(ni,ri)];
A3i=[zeros(ri,3) Hi eye(ri) -eye(ri)];
A4i=[zeros(ri,3) -Hi -eye(ri) eye(ri)];
AEje1i=[A1i;A2i;A3i;A4i];
fEje1i=[0;0;0;zeros(ni,1);ones(ri,1);ones(ri,1)];
b1i=-Yi;
b2i=Yi;
b3i=[zeros(ri,1)];
b4i=[zeros(ri,1)];
bEje1i=[b1i;b2i;b3i;b4i];
lb=zeros(1,length(fEje1i));

options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'off');

[xi fval exitflag output]=linprog(fEje1i,[],[],AEje1i,bEje1i,lb,[],[],options);

B1i=xi(1);
B2i=xi(2);
B3i=xi(3);

B0=median(C);

Bi=[B0;B1i;B2i;B3i];

YEi=X(i,:)*Bi;

Ii=eye(ni);

Hi=Xi*(inv(Xi'*Xi))*Xi';

```

```

SSEi=Yi'*(Ii-Hi)*Yi;

Si=sqrt(SSEi);

Ei=Y(i)-YEi;

ti=Ei/(Si*sqrt(D(i)));

DCi=(B-Bi)'*X'*X*(B-Bi);

DFITi=abs(ti)*(DH(i)/D(i));

YE(i,1)=YEi;

SSE(i,1)=SSEi;

t(i,1)=ti;

DC(i,1)=DCi;

DFIT(i,1)=DFITi;

end

YE, SSE, t, DC, DFIT

RES=[L YE t SSE DC DFIT];

```

### E.1.3. Programa Diagnóstico Regresión MINSADBAD

```

load G:\AYUDA\Tareas\ALFONSO\Birkes1.txt

L=Birkes1;

n=length(L);

Y=L(:,2);

X1=L(:,3);

X2=L(:,4);

X3=L(:,5);

U=ones(n,1);

```

```

X=[U X1 X2 X3];
B=(inv(X'*X))*X'*Y;
I=eye(n);
H=X*(inv(X'*X))*X';
DH=diag(H);
R=I-H;
D=diag(R);
YE=[];
SSE=[];
t=[];
DC=[];
DFIT=[];
for i=1:n
    Li=L;
    Li(i,:)=[];
    ni=length(Li);
    ri=ni*(ni-1)/2;
    Yi=Li(:,2);
    X1i=Li(:,3);
    X2i=Li(:,4);
    X3i=Li(:,5);
    X4i=[X1i X2i X3i];
    Ui=ones(ni,1);
    Xi=[Ui X4i];
    Hi=[]; Fi=[];

```

```

for j=1:n-1
    Hi=[Hi' [zeros(ni-j,j-1) ones(ni-j,1) -eye(ni-j)]']';
    Fi=[Fi' [zeros(ni-j,j-1) ones(ni-j,1) eye(ni-j)]']';
end
Q=(1/2)*Fi;
C=Q*Yi;
A1i=[X4i eye(ni) zeros(ni,ri) zeros(ni,ri)];
A2i=[zeros(ri,3) Hi -eye(ri) eye(ri)];
AEje1i=[A1i;A2i];
fEje1i=[0;0;0;zeros(ni,1);ones(ri,1);ones(ri,1)];
b1i=Yi;
b2i=[zeros(ri,1)];
bEje1i=[b1i;b2i];
lb=zeros(1,length(fEje1i));
options = optimset('LargeScale', 'off', 'Simplex', 'on', 'Display', 'off');
[xi fval exitflag output]=linprog(fEje1i,[],[],AEje1i,bEje1i,lb,[],[],options);
B1i=xi(1);
B2i=xi(2);
B3i=xi(3);
B0=median(C);
Bi=[B0;B1i;B2i;B3i];
YEi=X(i,:)*Bi;
Ii=eye(ni);
Hi=Xi*(inv(Xi'*Xi))*Xi';
SSEi=Yi'*(Ii-Hi)*Yi;

```

```

Si=sqrt(SSEi);
Ei=Y(i)-YEi;
ti=Ei/(Si*sqrt(D(i)));
DCi=(B-Bi)'*X'*X*(B-Bi);
DFITi=abs(ti)*(DH(i)/D(i));
YE(i,1)=YEi;
SSE(i,1)=SSEi;
t(i,1)=ti;
DC(i,1)=DCi;
DFIT(i,1)=DFITi;
end
YE, SSE, t, DC, DFIT
RES=[L YE t SSE DC DFIT];

```