Estudio Comparativo por Simulación de Métodos de Estimación para Modelos Lineales.

Autor: Francisco Vera **Coautor**: Gaudencio Zurita

Escuela Superior Politécnica del Litoral.

En el presente trabajo, se realiza un estudio comparativo por simulación de diferentes métodos de estimación de parámetros de modelos lineales. Principalmente, la intención es comparar un método robusto de estimación, contra otro que no es robusto, pero que presenta algunas características teóricas deseables. A lo largo de este trabajo, se hace una presentación de lo que son los modelos lineales. Además se consideran tres métodos de estimación de parámetros, y se comparan dos de ellos. La simulación, nos permite hacer una comparación experimental de los métodos, sin el costo que tienen los ensayos en la realidad. Además, cuando los tratados teóricos de un tema son complejos, la simulación puede ser una buena alternativa. Uno de los métodos de estimación tratados en este trabajo, es relativamente nuevo. Los resultados de este trabajo, dan una calificación no satisfactoria a este método, sin embargo, el deseo y dedicación de explorar en un campo nuevo, debe estar en las mentes de los hombres que conforman la comunidad científica mundial, a pesar de los resultados que se puedan obtener.

1. INTRODUCCIÓN

El principal objetivo de este trabajo es realizar un estudio comparativo, por simulación, de diferentes métodos de estimación de parámetros para los modelos lineales. La idea central es verificar bajo qué condiciones debe preferirse uno u otro estimador.

Uno de los métodos más usados para estimar parámetros de modelos lineales, y uno de los más estudiados, es el de mínimos cuadrados (MC). Un estimador de mínimos cuadrados cumple con muchas características teóricas deseables que debe tener todo "buen" estimador.

Para hallar el estimador de mínimos cuadrados de los parámetros de un modelo, se procede de la siguiente manera: dada una muestra aleatoria $y_1, y_2, \cdots, y_n$, de la variable aleatoria *Y*, se trata de encontrar el valor de los parámetros del modelo, de tal forma que se minimice la suma cuadrática del error que está dada por $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, donde $\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n$ representan la estimación de la variable aleatoria *Y*, para ciertos valores de una o más variables independientes. La diferencia $e_i = y_i - \hat{y}_i, i = 1, \cdots, n$ representa la estimación del error de cada observación.

Esta suma mide en qué grado la estimación de los datos se aleja de los datos observados. La idea del método de mínimos cuadrados es intuitivamente buena, puesto que trata de minimizar una función que mide el error de estimación de los datos.

Estos estimadores presentan ciertas propiedades teóricas deseables en cualquier estimador: son insesgados, de mínima varianza, etc. Estas características teóricas son muy importantes, y se estudiarán en detalle próximamente.

A pesar que un estimador determinado con este método presenta algunos características deseables, es muy sensible a los valores aberrantes. Cuando los datos se contaminan con información no deseada, los estimadores MC pueden también perder sus propiedades, y podría resultar una estimación incorrecta de los parámetros poblacionales.

Existen otros métodos para estimar los parámetros de un modelo, como los procedimientos no paramétricos. Este procedimiento no hace suposiciones específicas sobre la distribución del error, mas bien hace supuestos generales, como que la distribución del error debe ser simétrica. Este procedimiento exige menos restricciones sobre los datos que el de los mínimos cuadrados, pero no presenta algunos características teóricas del método MC. Es usado cuando la información disponible no cumple ciertos requisitos, como que el error de estimación siga una ley de distribución normal.

Otro método de estimación de parámetros disponible es el de máxima verosimilitud. El estimador obtenido por este método para modelos lineales, coincide con el estimador de mínimos cuadrados. Este método conduce a las mismas ecuaciones que el de mínimos cuadrados.

Otro método, es el de la mínima mediana de los cuadrados (MMC). Este método presenta ciertas propiedades: es insesgado y robusto. No presenta la característica de mínima varianza, que es una propiedad deseable. Sin embargo, cuando los datos están contaminados, lo cual sucede con mucha frecuencia, este método podría ser preferible bajo ciertas condiciones.

La idea de la comparación que se va a realizar en este trabajo, es la de simular varios modelos de regresión. En cada uno de estos modelos se aplicarán los distintos métodos de estimación que se han mencionado, y se verificará como se comporta cada método en las distintas situaciones.

2. EL MÉTODO DE MÍNIMOS CUADRADOS

Uno de los métodos más utilizados en la estimación de parámetros de un modelo lineal, es el método de mínimos cuadrados.

Supongamos el modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i\,p-1} + \varepsilon_i, \quad i = 1, \cdots, n \qquad \textbf{2-1}$$

donde y_i es el valor observado, ε_i tiene distribución normal con media 0 y varianza σ^2. Además $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$.

Entonces

$$E[y_i] = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i\,p-1} \qquad \textbf{2-2}$$

Ahora, el valor con el que estimaremos las observaciones y_i es su valor esperado. La estimación de y_i se denota por $\hat{y}_i$. Como no se conocen los parámetros, entonces se emplearán las estimaciones de los parámetros para obtener el valor de $\hat{y}_i$. Esto es

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_{p-1} x_{i\,p-1} \qquad \textbf{2-3}$$

La diferencia entre el valor observado y el valor estimado, es igual al error correspondiente a esa observación, es decir

$$\varepsilon_i = y_i - \hat{y}_i \qquad \textbf{2-4}$$

Ahora, se requiere de una función diferenciable, de tal manera que el error total de la estimación, sea mínimo. Si minimizamos la suma del error, obtendríamos estimaciones con un error pequeño en valor relativo, pero de magnitud grande. Si minimizamos la suma de los valores absolutos del error, entonces la función resultante no va ser diferenciable.

Se define la suma cuadrática del error

como la suma de los errores elevados al cuadrado. Se denota por *Q*, y es una medida del error total de la estimación.

El método de mínimos cuadrados consiste en establecer el valor de los β_i que minimicen la suma cuadrática del error. Esto se hace derivando parcialmente la suma cuadrática del error con respecto a cada parámetro, e igualando cada derivada a cero. Luego

$$Q = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 =$$
$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{p-1} x_{i\,p-1} \right)^2$$

2-5

$$\frac{\partial Q}{\partial \beta_0} = 0 \rightarrow$$
$$-2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{p-1} x_{i\,p-1} \right) = 0$$
$$\frac{\partial Q}{\partial \beta_j} = 0 \rightarrow$$
$$-2 \sum_{i=1}^{n} x_{ij} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{p-1} x_{i\,p-1} \right) = 0,$$
$$j = 1, \cdots, p-1$$

2-6

Este es un conjunto de p ecuaciones con p incógnitas. La resolución de este sistema de ecuaciones, daría como resultado la estimación de los parámetros β_i por el método de mínimos cuadrados.

A este conjunto de ecuaciones, se le conoce como *ecuaciones normales del modelo.*

De la ecuación correspondiente a β_0, podemos obtener dos propiedades del método de mínimos cuadrados.

$$-2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{p-1} x_{i\,p-1} \right)$$
$$= 0 \Rightarrow \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right) = 0 \Rightarrow$$
$$\begin{cases} \sum_{i=1}^{n} \varepsilon_i = 0 \\ \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \end{cases}$$

2-7

Es decir, la suma total del error de estimación de las observaciones es cero, y la suma de los valores observados es igual a la suma de los valores estimados. Esto no quiere decir que los valores observados sean iguales a los valores estimados. También implica que la media aritmética del error es cero, y que la media aritmética de los valores observados es igual a la media aritmética de los valores estimados.

El vector de error **ε** se puede conocer solamente cuando se conoce el valor del vector **β**. Como el vector obtenido por el método anterior es una estimación de los parámetros, no podemos conocer el error, pero si podemos estimarlo. Al estimador de **β** se lo denota por ***b***, mientras que al estimador del vector de error **ε**, lo conoceremos como ***e***.

La media aritmética de los e_i, representada por $\bar{e}$ no es una variable aleatoria, a pesar que es resultado de una combinación lineal de variables aleatorias, puesto que $\bar{e}$ solo puede tomar un solo valor, cero, lo cual le da el carácter de *variable determinística*, o "*variable aleatoria con varianza cero*".

Las ecuaciones normales contiene algunos términos con sumatorias, y su resolución podría ser un poco tediosa. A continuación, se muestra un método matricial equivalente al de mínimos cuadrados.

La suma cuadrática del error, no es más que el producto interno del vector del error consigo mismo, y este se puede obtener multiplicando su transpuesta por el mismo. Luego se tiene que

$$Q = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} =$$
$$\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} =$$
$$\varepsilon_1^{\,2} + \varepsilon_2^{\,2} + \cdots + \varepsilon_n^{\,2} =$$
$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

2-8

Donde, de la representación matricial del modelo **Y** = **Xβ** + **ε**, se tiene que el **ε**=**Y** - **Xβ**. Es decir el vector $\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta}$ representa el vector de los valores estimados.

La aplicación del método de mínimos cuadrados en su forma matricial sería

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \mathbf{0} \rightarrow$$
$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(\mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \right) = \mathbf{0}$$

2-9

Aquí, el símbolo de derivada indica derivación parcial de Q con respecto a cada uno de los elementos de **β**. El tercer término es la transpuesta del segundo, y además ambas matrices son de dimensión 1 x 1. Esto quiere decir que ambos términos son iguales. Con estas consideraciones procederemos a resolver la derivada y a igualar a cero

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left(\mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}\right)$$
$$= -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

2-10

Luego, las "ecuaciones normales" del modelo serían

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y} \qquad \textbf{2-11}$$

Luego, si **b** es la solución de este sistema de ecuaciones, entonces **b** sería la estimación de mínimos cuadrados del vector de parámetro $\boldsymbol{\beta}$.

Estas ecuaciones son las mismas que se habían obtenido anteriormente, pero en notación simplificada.

3. El método MMC

En el método de mínimos cuadrados, se intenta hallar el estimador que minimice la suma cuadrática total. Esto es

$$\min_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\varepsilon_i^{\,2}(\boldsymbol{\beta})\right\} \qquad \textbf{3-1}$$

Es decir, cada $\boldsymbol{\beta}$ posible, tiene su propio vector de error ($\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$). Luego, el estimador de mínimos cuadrados **b** es tal que, la función $\boldsymbol{\varepsilon}^T(\boldsymbol{\beta})\ \boldsymbol{\varepsilon}(\boldsymbol{\beta})$ tiene un mínimo absoluto en $\boldsymbol{\beta} = \mathbf{b}$.

En el método de la mínima mediana de los cuadrados (MMC), lo que cambia es la función objetivo. En lugar de ser la suma cuadrática del error, tenemos como función objetivo la mediana de los elementos del vector de error correspondiente a $\boldsymbol{\beta}$, es decir

$$\min_{\boldsymbol{\beta}}\left\{mediana\left(\varepsilon_i^{\,2}(\boldsymbol{\beta})\right)\right\} \qquad \textbf{3-2}$$

Esta función es continua, pero no es diferenciable, es menos sensible a los valores aberrantes, y por tanto tiene mayor ruptura, y es más robusta.

Estadísticamente, la mediana muestral es

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}; & n \text{ impar} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right); & n \text{ par} \end{cases}$$

Para lograr la más alta ruptura posible en la estimación, en lugar de minimizar la mediana, minimizaremos el estadístico de orden *h*, tal que *h* podría ser

$$h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor \qquad \textbf{3-3}$$

donde p es el número de parámetros del modelo lineal, y n es el número de observaciones.

En 1990, Cook y Hawkins recomendaron que se debe investigar la estabilidad de un estimador MMC sobre un imagen de valores de h.

Nótese, que por la definición de h, el estadístico de orden h, es muy cercano a la mediana. De hecho, si n es impar, y solo tengo dos parámetros, este estadístico coincidiría con la mediana.

Entonces, el estimador MMC se reduce a

$$\min_{\boldsymbol{\beta}} \left\{ \varepsilon_h{}^2 (\boldsymbol{\beta}) \right\} \qquad \textbf{3-4}$$

Este estimador, no es muy fácil de calcular. A continuación describiremos como se calcula el estimador MMC, según el algoritmo de Stromberg.

Primero, el estimador MMC, minimiza el cuadrado del *h-ésimo* residuo mayor, o el cuadrado del residuo de orden h, de un conjunto dado de datos. Entonces, el estimador MMC debe minimizar el máximo de los cuadrados de los residuos, para algún subconjunto de tamaño h. Esto quiere decir, que el estimador MMC vendría a ser una estimación por el método de Chebyshev” (ó minimax), pues intenta hallar un subconjunto de los datos originales que minimice el máximo de los cuadrados residuales.

Aparentemente, habría que examinar todos los conjuntos de tamaño h, de los n datos. El número de subconjuntos a examinar sería $\binom{n}{h}$, lo cual es bastante elevado, porque h es un valor cercano a la mitad de n.

Sin embargo, Cheney demostró en 1962, que el estimador de Chebyshev de todos los datos, es el estimador de Chebyshev para un subconjunto de tamaño p+1. Esto quiere decir, solamente hay que revisar $\binom{n}{p+1}$, lo cual es mucho menor que $\binom{n}{h}$.

Luego para cada subconjunto de tamaño p+1, tendremos una matriz de diseño y una matriz de observaciones. Por cada $\boldsymbol{\beta}$, tendremos un residuo e_i ($\boldsymbol{\beta}$).

Supongamos ahora que para un subconjunto E, la matriz de diseño es $\mathbf{X}_E$ y la matriz de observaciones es $\mathbf{Y}_E$. Sea $\hat{\boldsymbol{\beta}}_{EMC}$, el estimador de mínimos cuadrados correspondiente a los puntos del conjunto E, esto es

$$\hat{\boldsymbol{\beta}}_{EMC} = \left(\mathbf{X}_E{}^T\mathbf{X}_E\right)^{-1}\mathbf{X}_E{}^T\mathbf{Y}_E \quad \text{3-5}$$

A continuación, defínase δ como

$$\delta = \frac{\sum_{i=1}^{p+1} e_i{}^2\left(\hat{\boldsymbol{\beta}}_{EMC}\right)}{\sum_{i=1}^{p+1}\left|e_i\left(\hat{\boldsymbol{\beta}}_{EMC}\right)\right|} \qquad \text{3-6}$$

Ahora, sea $\mathbf{s}$ un vector columna de dimensión p+1, tal que $s_i = \operatorname{sgn}\left(e_i\left(\hat{\boldsymbol{\beta}}_{EMC}\right)\right)$. Luego, la estimación de Chebyshev para el conjunto de datos E sería la solución del siguiente sistema lineal:

$$\left(\mathbf{X}_E{}^T\mathbf{X}_E\right)\hat{\boldsymbol{\beta}}_{EC} = \mathbf{X}_E{}^T\left(\mathbf{Y}_E - \delta\mathbf{s}\right) \quad \text{3-7}$$

El algoritmo e Stromberg consiste en examinar todos los subconjuntos posibles de tamaño p+1, y para cada subconjunto calcular $\hat{\boldsymbol{\beta}}_{EC}$. Luego el estimador LMS exacto, $\hat{\boldsymbol{\beta}}_{MMC}$ sería aquel $\hat{\boldsymbol{\beta}}_{EC}$, que aplicado con los datos originales, produzca el menor error cuadrático de orden h.

4. Comportamiento de los estimadores en condiciones normales

La tabla que se muestra a continuación, está clasificada en las filas, por el tamaño de la muestra, y en las columnas por el número de corridas. En cada elemento de la tabla, se encuentran el promedio de los estimadores de los parámetros correspondiente a un número de corridas y tamaño muestral. Los intervalos de confianza definidos son de 1 desviación estándar. Estas estimaciones son hechas con el método de mínimos cuadrados.

de corridas

Tamaño de la muestra

	100	200	500
	2.936±0.181	3.097±0.147	3.018±0.0157

16	-3.992±0.037	-4.017±0.034	-4.000±0.036
	2.014±0.054	1.988±0.047	1.994±0.049
30	3.031±0.103	2.980±0.103	3.012±0.106
	-4.010±0.016	-3.990±0.016	-4.000±0.016
	2.000±0.031	1.994±0.030	1.996±0.029
49	3.011±0.077	3.031±0.075	3.011±0.084
	-3.998±0.011	-4.003±0.011	-4.002±0.011
	1.993±0.016	1.994±0.014	1.998±0.016

Tabla 4-1

Obsérvese que los valores de la primera columna de cada elemento de la tabla, son cercanos a 3, -4 y 2, es decir, a los parámetros. La segunda columna de cada elemento, tiene solo valores positivos. Esto es porque representa la estimación de la varianza correspondiente a cada estimador de parámetros.

A continuación mostramos una tabla similar a la anterior, pero las estimaciones son realizadas por el método de mínima mediana de los cuadrados.

de corridas

Tamaño de la muestra	100	200	500
16	2.641±0.387	3.150±0.385	2.934±0.354
	-3.926±0.084	-4.003±0.100	-3.989±0.090
	2.083±0.105	1.944±0.097	2.016±0.095
30	2.884±0.235	3.017±0.275	3.020±0.259
	-4.005±0.048	-3.998±0.047	-4.008±0.046
	2.037±0.064	1.993±0.068	2.010±0.066
49	3.073±0.202	2.974±0.208	
	-4.018±0.034	-3.989±0.036	
	2.001±0.033	1.994±0.034	

Tabla 4-2

Nótese que el promedio de las estimaciones es cercano al valor de los parámetros. Además, las desviaciones de esta tabla son significativamente mayores que las desviaciones de la tabla 4-1. Esto muestra que la varianza de los estimadores obtenidos por el método de mínimos cuadrados es menor que la varianza de los estimadores obtenidos por el método de la mínima mediana de los cuadrados.

Luego, bajo condiciones normales, la estimación obtenida por el método de mínimos cuadrados es más precisa que la estimación obtenida por el método de la mínima mediana de los cuadrados.

5. COMPORTAMIENTO DE ESTIMADORES CON POBLACIONES CONTAMINADAS

La generación de errores sin contaminación, se realizaba en la sección anterior, empleando una distribución normal estándar. La contaminación se realiza tomando en promedio el 100α% de las veces, muestras de una población normal, con media cero y varianza 16.

En la tabla que se muestra a continuación, se encuentran los valores de las estimaciones de los parámetros, tanto para el método de mínimos cuadrados (MC), como para el método de la mínima mediana de los cuadrados (MMC). Además, se muestran los porcentajes de cobertura de intervalos de confianza para cada caso. Estos intervalos son de 68% de confianza. Cada fila, corresponde a un nivel de contaminación diferente.

	Mínimos Cuadrados		Mínima Mediana de los Cuadrados	
0	2.980±0.103	68.00%	3.017±0.275	69.00%
	-3.990±0.016	65.50%	-3.998±0.047	68.50%
	1.994±0.030	66.00%	1.993±0.068	66.00%
0.05	3.014±0.139	68.50%	2.930±0.258	68.50%
	-4.007±0.021	73.00%	-4.002±0.047	70.00%
	2.004±0.040	71.50%	2.028±0.061	68.00%
0.10	2.957±0.178	75.00%	2.930±0.265	72.00%
	-4.000±0.032	73.50%	-4.002±0.045	70.00%
	2.011±0.049	71.00%	2.028±0.069	71.50%
0.15	2.937±0.202	70.50%	3.011±0.290	70.00%
	-3.994±0.036	69.00%	-3.998±0.048	70.50%
	2.020±0.054	68.50%	2.001±0.071	70.00%
0.20	3.116±0.254	71.00%	2.939±0.257	70.50%
	-3.999±0.038	68.50%	-3.997±0.049	71.50%
	1.975±0.060	69.50%	2.020±0.069	69.00%

Tabla 5-1

Como puede apreciarse, las coberturas son similares, tanto para mínimos cuadrados, como para mínima mediana de los cuadrados. Además, nótese que lo peor que le puede pasar a la varianza del estimador de mínimos cuadrados, es llegar a ser la misma que la varianza del estimador de mínima mediana de los cuadrados.

A continuación, se muestra un tabla donde, el nivel de contaminación es 0.10, pero los datos del cual se toma contaminar la muestra, tienen una distribución normal con medias diferentes de cero, pero varianza uno. Las filas representan las medias de la contaminación.

	Mínimos Cuadrados		Mínima Mediana de los Cuadrados	
0	2.980±0.103	68.00%	3.017±0.275	69.00%
	-3.990±0.016	65.50%	-3.998±0.047	68.50%
	1.994±0.030	66.00%	1.993±0.068	66.00%
1	3.081±0.113	63.00	3.037±0.306	66.00
	-3.998±0.019	68.00	-3.981±0.052	65.00
	2.004±0.030	64.00	2.000±0.069	68.00
2	3.169±0.131	73.50%	2.994±0.307	84.50%
	-3.996±0.025	83.00%	-3.995±0.054	85.00%
	2.006±0.035	81.50%	2.009±0.071	87.00%

Tabla 5-2

6. CONCLUSIONES

1. Al comparar las tablas 4-1 y 4-2, podemos apreciar que bajo condiciones en la que se cumplan los supuestos iniciales, el estimador de mínimos cuadrados resulta preferible que el estimador de la mínima mediana de los cuadrados. Para este caso, los resultados teóricos, resultaron concordantes

con los resultados experimentales (simulaciones).

2. De la tabla 5-1, podemos concluir que el estimador de la mínima mediana de los cuadrados es mucho más robusto que el de mínimos cuadrados, puesto que la varianza del estimador MMC no varía significativamente, a pesar de la contaminación. En cambio, la varianza del estimador MC, aumenta conforme se aumenta el nivel de contaminación. Esto, también concuerda con la teoría.
3. En cuanto al sesgo, ambos estimadores presentan sesgos numéricamente iguales. Además, el sesgo en cada uno es nulo. Por tanto, podemos concluir que los dos estimadores son insesgados.
4. A pesar que la varianza del estimador MC aumenta, y la del estimador MMC se mantiene, la varianza del estimador MC siempre es estadísticamente menor que la del estimador MMC. El único caso en el que las varianzas de ambos estimadores son numéricamente iguales, es cuando el nivel de contaminación es alto (0.20). Esto quiere decir, que a pesar de la contaminación, el estimador MC es tan bueno como el estimador MMC.
5. En la tabla 5-2, es la que contenía contaminación con desplazamiento. La varianza del estimador MMC tuvo un aumento mayor que la varianza del estimador MC. Esto quiere decir, que para este tipo de contaminación, ambos métodos son similares en robustez. Otra vez, resulta mejor el estimador de mínimos cuadrados.
6. Según la teoría expuesta en este trabajo, el estimador MMC es más robusto que el de MC, y supuestamente el método MMC debería ser mejor en condiciones de contaminación que MC. La robustez se cumple como se predijo, pero la varianza del estimador MMC es tan grande, que a pesar de no aumentar con la contaminación, de todos modos siempre es mayor que la varianza del estimador MC, la cual aumenta, conforme aumenta la contaminación. Se Esperaba que el estimador MMC sea mucho mejor

que el estimador MC, en circunstancias de contaminación de datos. Sin embargo, aquí se muestra que, no siempre se obtiene lo que uno espera hallar.

7. BIBLIOGRAFÍA

1. Rao C. Radhakrishna, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Estados Unidos, 1973.

2. Graybill Franklin A., *Theory And Application of the Linear Model*, Duxbury Press, California, 1976.

3. Draper N. R., Smith H., *Applied Regression Analysis*, John Wiley & Sons, Estados Unidos, 1980.

4. Seber G., *Linear Regression Analysis*, John Wiley & Sons, Estados Unidos, 1980.

5. Mitra Amitava, *Fundamentals of Quality Control and Improvement*, Macmillan, NewYork, ____.

6. Hollander Myles, Wolfe Douglas A., *Nonparametric Statistical Methods*, John Wiley & Sons, Estados Unidos, 1973.

7. Hawkins Douglas M., Simonoff Jeffrey S., Stromberg Arnold J., *Distributing a computationally intensive estimator: the case of exact LMS Regression*, Computational Statistics, Tennessee, 1994.

8. Fishman George S., *Principles of Discrete Event Simulation*, John Wiley & Sons, Estados Unidos, 1978.

Autor: Francisco Vera

Coautor: Gaudencio Zurita