

Estimadores Robustos para el Vector de Medias y la Matriz de Varianzas y Covarianzas de Vectores Aleatorios Multivariados

Néstor Montaña, Msc. Gaudencio Zurita
Instituto de Ciencias Matemáticas
Escuela Superior Politécnica del Litoral
Campus Gustavo Galindo, Km. 30.5 Vía Perimetral, Apartado 09-01-5863, Guayaquil, Ecuador
nestor.montano@gmail.com

Resumen

La Estimación Robusta nace de la necesidad de estimadores que se comporten “bien” aún cuando existan variaciones en los supuestos iniciales o cuando es posible que el modelo esté “contaminado” por valores aberrantes que producen influencias en los resultados y por lo tanto conducen a estimaciones erróneas, siendo este un campo en constante desarrollo se han propuesto diversos métodos de Estimación. Este artículo presenta los resultados de un estudio tipo Monte Carlo realizado para comparar algunos Métodos de Estimación Robusta para el Vector de Medias y Matriz de Varianzas y Covarianzas de un vector aleatorio de seis variables. El propósito es evaluar el comportamiento de los estimadores bajo diversas condiciones como Contaminación total o Contaminación por Variable; además se trata de establecer una “regla empírica” donde se utilice al tamaño de la Muestra, al Sesgo y la Curtosis Muestral como elementos de decisión sobre el estimador a utilizar. Los estimadores M de Huber y Bicuadrático o Bponderado son los que mejor rendimiento presentan aunque cuando la curtosis es “alta” el Estimador MCD es el mejor.

Palabras claves: Estimación Multivariada, Robustez, Monte Carlo

Abstract

The Robust Estimation born from the need of estimators to behave "well" even when there are variations in the initial assumptions or when it is possible that the model is "contaminated" by outliers that producing influences the results and thus lead to estimates wrong, because of this is a field in constant development have been proposed various methods of estimation. This article presents the results of a Monte Carlo study realized for to compare some Robust Estimation Methods for Vector averages and Matrix of Variance and Covariance of a random vector of six variables. The purpose is to evaluate the behavior of the estimators under various conditions such as total contamination or contamination for variable, also I seek to establish a “empirical rule” that it use the size of the Sample, the Sample Bias and the Sample Kurtosis as elements of decision on the estimator one to using. The M estimator of Huber and Bisquared are those who present better performance, though when the kurtosis is "high" the Esteeming MCD is the best.

Key words: Multivariate Estimation, Robustness, Monte Carlo

1. Introducción

Para estimar los parámetros poblacionales se utiliza información obtenida a partir de los datos que nos proporciona una Muestra; en la práctica se verifica que un alto porcentaje de las mediciones que se efectúan, por diferentes razones, contienen errores u observaciones atípicas llamadas “valores aberrantes” o “extremos” pues se alejan acentuadamente del comportamiento general de las demás observaciones; bajo este escenario ¿el estimador seguirá siendo una “buena” aproximación, o se verá afectado por este particular?.

Para ilustrar lo mencionado, Considere la estatura (cm) y el peso (kg) de veinte estudiantes, datos que se muestran a continuación:

cm	kg	cm	kg
164	65	175	77
178	81	189	85
187	88	179	84
179	84	178	79
178	77	173	74
179	69	178	74
166	69	175	79
169	72	175	70
185	86	175	75
172	70	189	65

Un primer pensamiento puede llevar a verificar la existencia de valores extremos en cada variable de forma independiente, sin embargo los datos se ajustan a una Distribución Normal; sin embargo en la Figura 1 se presenta el Gráfico de Dispersión para las dos variables estudiadas en donde se observa un valor atípico, el mismo que corresponde al punto (189, 65).

Sea $\bar{\mathbf{X}}_{19}^T = (176.53, 76.73)$ el vector de medias correspondientes a los datos sin el valor atípico, y $\bar{\mathbf{X}}_{20}^T = (177.16, 76.14)$ el vector de medias correspondiente a los veinte datos, se puede ver que la diferencia entre ambos vectores es “pequeña”; sin embargo, considere las Matrices de Correlación

$$\mathbf{R}_{19} = \begin{pmatrix} 1 & \\ 0.833 & 1 \end{pmatrix} \quad \text{y} \quad \mathbf{R}_{20} = \begin{pmatrix} 1 & \\ 0.549 & 1 \end{pmatrix}$$

correspondientes a los datos sin valor atípico y a los datos completos respectivamente, se puede notar que existe “gran” influencia de este valor atípico en la correlación entre las variables.

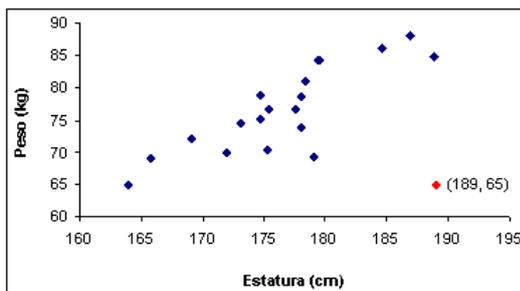


Figura 1. Gráfico de dispersión: Estatura Vs. Peso

La Matriz de Covarianzas y su correspondiente Matriz de Correlación son el punto inicial de muchos de los análisis multivariados como Componentes Principales, Análisis de Correlación Canónica, etc. Así, este ejemplo permite entender la importancia de los estimadores robustos en muestras multivariadas.

2. Estimación Robusta

Para modelar la situación en que la mayoría de las observaciones provienen de la distribución F_θ , pero una pequeña fracción ε de las observaciones son valores atípicos generados por otra distribución H , Tukey en [14] plantea se plantea la Familia de Contaminación \mathcal{F}_ε definida por:

$$\mathcal{F}_\varepsilon = \{(1 - \varepsilon)F_\theta + \varepsilon H; \theta \in \Theta\} \quad (2.1)$$

donde ε representa el porcentaje de contaminación.

Se espera que los Estimadores Robustos cumplan con dos requerimientos: Eficiencia y Estabilidad.

Se llama Eficiencia al hecho que el Estimador Robusto se comporte “bien” cuando no hay contaminación, es decir que el Estimador Robusto sea comparable con el Estimador de Máxima Verosimilitud, para el caso multivariado este requerimiento se implica lo siguiente:

i) Sea $(\hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n)$ los estimadores de localización y dispersión para una muestra de tamaño n y sean $(\hat{\boldsymbol{\mu}}_\infty, \hat{\boldsymbol{\Sigma}}_\infty)$ sus valores asintóticos. Si $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\hat{\boldsymbol{\mu}}_\infty = \boldsymbol{\mu}$ y $\hat{\boldsymbol{\Sigma}}_\infty = c\boldsymbol{\Sigma}$ donde c es una constante

ii) $(\hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n)$ deben ser asintóticamente normales

$$\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \hat{\boldsymbol{\mu}}_\infty) \xrightarrow{L} N_p(\mathbf{0}, \mathbf{V}_\mu)$$

$$\sqrt{n}(\text{vech}(\hat{\boldsymbol{\Sigma}}_n - \hat{\boldsymbol{\Sigma}}_\infty)) \xrightarrow{L} N_q(\mathbf{0}, \mathbf{V}_\Sigma)$$

donde $q = \frac{p(p+1)}{2}$ y $\text{vech}(\boldsymbol{\Sigma})$ es el vector que

contiene los q elementos de la triangular inferior de $\boldsymbol{\Sigma}$

Un estimador es estable cuando su “buen” comportamiento se preserva cuando F varía sobre \mathcal{F}_ε es decir cuando hay contaminación, para evaluar la estabilidad se han propuesto varias medidas, como Sesgo Asintótico Máximo y la Varianza asintótica Máxima que miden el “peor” comportamiento del estimador para todo $\varepsilon < \varepsilon^*$; también se tiene el Punto de Ruptura Asintótico que intuitivamente representa la mayor fracción de contaminación que el estimador puede tolerar.

Por otro lado, en varios de los métodos de análisis multivariados se trabaja con transformaciones lineales de las variables, entonces todos los estimadores tratados cumplen la propiedad de equivarianza, esto es: Sea $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ entonces

$$\hat{\boldsymbol{\mu}}(Y) = \mathbf{A}\hat{\boldsymbol{\mu}}(X) + \mathbf{b}$$

$$\hat{\boldsymbol{\Sigma}}(Y) = \mathbf{A}\hat{\boldsymbol{\Sigma}}(X)\mathbf{A}^T$$

Existe una gran variedad de Estimadores Robustos propuestos, sin embargo se ha escogido a los considerados más populares, a continuación se presentan los estimadores considerados en el estudio.

Sea la distancia de Mahalanobis $d_i = d_i(\mathbf{x}_i, \mu, \Sigma) = (\mathbf{x}_i - \mu)^T \Sigma^{-1} (\mathbf{x}_i - \mu)$

2.1 Estimador M Multivariado

Maronna en [9] generaliza los estimadores M univariados propuestos por Huber en [5], así define a los estimadores M como soluciones de

$$\sum_{i=1}^n W_1(d_i)(\mathbf{x}_i - \hat{\mu}) = 0$$

$$\frac{1}{n} \sum_{i=1}^n W_2(d_i)(\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T = \hat{\Sigma}$$

donde W_1 y W_2 no son necesariamente iguales. Nótese que el Estimador M se puede interpretar como una Matriz de Covarianza ponderada y un Vector de Medias ponderado, donde las ponderaciones dependen de la Distancia de Mahalanobis.

Se utiliza funciones W_1 y W_2 de tipo Huber, esto es:

$$W_1(d_i) = \begin{cases} 1 & d_i \leq k \\ \frac{k}{d_i} & d_i > k \end{cases} \quad W_2(d_i) = \frac{\{W_1(d_i)\}^2}{\beta}$$

2.2 Estimador S Bicuadrático Multivariado

Se define al Estimador S Bicuadrático o Biponderado multivariado como un estimador con escala dada por

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{d_i}{\hat{\sigma}}\right) = \delta \quad \text{donde:}$$

i) $\hat{\sigma}$ es un Estimador univariado de Escala Robusto talque $\hat{\sigma}(d_i(X, \hat{\mu}, \hat{\Sigma})) = \min$ y $|\hat{\Sigma}| = 1$.

ii) $\rho(t) = \min\{1, 1 - (1 - t)^3\}$.

Se puede demostrar que, bajo ciertas condiciones, los estimadores S son una particularización de los Estimadores M cuya función de ponderación tiende a cero para distancias "grandes".

2.3 Estimador S T-Bicuadrático

Mientras mayor sea el número de variables, los estimadores S con función de ponderaciones continúa se aproximan al vector de medias muestral y la matriz de varianzas y covarianzas muestral, esto implica una pérdida de robustez; Rocke en [11] considera este problema y propone un tipo de estimador con función dependiente del número de variables, en particular propuso una familia de estimadores cuya función ρ cumple que $\lim_{p \rightarrow \infty} \rho(d) = I(d > 1)$. El estimador T-

Bicuadrático es un estimador que cumple las mismas condiciones del Estimador Bicuadrático pero con

$$\rho(t) = \begin{cases} 0 & \text{para } 0 \leq t \leq 1 - \gamma \\ \left(\frac{t-1}{4\gamma}\right) \left[3 - \left(\frac{t-1}{\gamma}\right)^2\right] + \frac{1}{2}; & \text{para } 1 - \gamma < t < 1 + \gamma \\ 1 & \text{para } t \geq 1 + \gamma \end{cases}$$

2.4 Estimador Covarianza de Mínimo Determinante MCD

Zuo en [17] indica que los estimadores MCD propuestos por Rousseeuw escogen h observaciones las cuales minimizan el determinante de la Matriz de Covarianzas Clásica, así, los estimador MCD de localización es el promedio de las h observaciones y el estimador MCD de escala es un múltiplo escalar de su Matriz de Covarianzas correspondiente a las h observaciones.

Este estimador es talvez el más popular debido (en parte) al "rápido" cálculo de los estimadores por parte de varios algoritmos desarrollados.

2.5 Estimador Stahel-Donoho (DS)

Maronna en [8] indica que la idea del estimador Stahel-Donoho, propuesto por Stahel (1981) y Donoho (1982) es que un "valor aberrante" multivariado debe serlo también en alguna proyección univariada. Entonces el estimador DS, es un vector de Media y Matriz de Covarianzas ambos ponderados de tal manera que la ponderación de \mathbf{x}_i es una función de la "lejanía" de \mathbf{x}_i , denotada por $t(\mathbf{x}_i)$.

Es decir, sean W_1 y W_2 dos funciones de ponderaciones, se define

$$\hat{\mu} = \frac{1}{\sum_{i=1}^n w_{i1}} \sum_{i=1}^n w_{i1} \mathbf{x}_i$$

$$\hat{\Sigma} = \frac{1}{\sum_{i=1}^n w_{i2}} \sum_{i=1}^n w_{i2} (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T$$

con $w_{ij} = W_j(t(\mathbf{x}_i))$, $j = 1, 2$

3. Detalles de la Simulación

Las medidas planteadas para medir la Robustez son definidas Asintóticamente, quedando sin explicar el comportamiento de dichos Estimadores en Muestras Finitas, esto es lo que se explora en el presente estudio utilizando Simulación Matemática.

3.1 Diseño del Estudio

En este estudio se generan muestras aleatorias a partir de distribuciones en las que se manipula una serie de parámetros con el objetivo de simular varias condiciones y así estudiar el comportamiento de los Estimadores Robustos bajo estas condiciones; tal vez el primer estudio comparativo de Estimadores

Robustos fue el realizado por Andrews et al. en [1], en dicho estudio se dio a lugar a la noción de las tres "esquinas" para representar las posibles situaciones que se puede encontrar en la práctica, se propuso probar los Estimadores Robustos bajo un enfoque muy optimista (Generando una Distribución Normal), muy pesimista (a través de una Cauchy) y la última prueba consistía en 20 datos, 19 de los cuales eran generados a partir de una Normal Estándar y el último generado a partir de una $N(0, 100)$. Los dos primeros casos son el punto de partida del presente estudio, es decir, se consideraran muestras generadas a partir de la Distribución Normal Multivariada $N(\mathbf{0}, \Sigma)$ y la Distribución Cauchy $CAU(\mathbf{0}, \Sigma)$ ambas difieren principalmente en el "peso" de sus colas.

En este estudio, las muestras generadas serán contaminadas, para ello se considera la Familia de Contaminación (2.1) donde $\varepsilon = 0; 0.05; 0.1$ y 0.3 de tal manera que se observa el comportamiento de los estimadores analizados bajo condiciones de:

- No Contaminación, $\varepsilon = 0$, caso que permitirá compararlos con los Estimadores de Máxima Verosimilitud,
- Contaminación Moderada, $\varepsilon = 0.05$ y $\varepsilon = 0.1$, que parece ser lo más cercano a la realidad y
- Contaminación Extrema, $\varepsilon = 0.3$, cuyos resultados se pueden utilizar para verificar si el Punto de Ruptura de los Estimadores es menor a 0.3 .

Además, para la Distribución H que genera la contaminación en el modelo (2.1) se ha escogido tres Distribuciones:

- Distribución Normal Multivariada $N(\mathbf{0}, 9\Sigma)$, que sigue la idea propuesta por Tukey en [14] y que es la Distribución frecuentemente utilizada en este tipo de estudios.
- Distribución Normal Multivariada $N(0.537\alpha_6, 9\Sigma)$, es decir una Contaminación Asimétrica donde α_6 representa el vector propio correspondiente al menor valor propio asociado a la Matriz de Varianzas y Covarianzas; este tipo de contaminación es utilizada por Devlin en [2]; y
- Distribución Uniforme Esférica $Uesf(d)$, donde d representa la distancia hacia el origen; esta Distribución se utiliza bajo el supuesto que una observación errónea puede producirse en cualquier punto con igual probabilidad, por lo que se escoge esta Distribución para modelar lo planteado.

Por otro lado, como se menciona en el capítulo dos del presente trabajo, el Entorno de Contaminación (2.1) indica que una proporción de vectores no siguen la distribución original, para el caso multivariado esto implica que todas las componentes de la observación están contaminadas o No están contaminadas; esto sin embargo no es necesariamente lo que ocurre en la realidad, pues se puede pensar que los errores se dan en una o varias componentes de la observación, difícilmente en todas; en base a esto se considera otro tipo de contaminación, la misma que contamina cada

variable independientemente, es decir, mientras en el modelo (2.1) una observación tiene probabilidad ε de estar contaminada, el segundo modelo considerado indica que cada componente de la observación tiene probabilidad ε de estar contaminada. Las distribuciones utilizadas para generar la contaminación por variable son: Distribución Normal $N(0, 9\sigma_i)$ y la Distribución Uniforme $U(-5, 5)$.

3.2 Parámetros de la Simulación

Además de lo explicado en la sección 3.1, para el presente estudio se trabaja con $p=6$, donde p representa el número de variables y el tamaño muestral será $n=kp$ donde $k=5; 10$ y 20 .

Para Σ se utilizan dos Matrices de Varianzas y Covarianzas, la primera igual a:

$$\Sigma_1 = \begin{pmatrix} 1 & & & & & \\ .950 & 1 & & & & \\ .300 & .100 & 1 & & & \\ 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & -.499 & 1 & \\ 0 & 0 & 0 & -.499 & -.499 & 1 \end{pmatrix}$$

a la misma que le corresponde los valores propios $\lambda_1 = 2.029, \lambda_2 = \lambda_3 = 1.499, \lambda_4 = 0.943, \lambda_5 = 0.028$ y $\lambda_6 = 0.002$ La segunda matriz considerada es

$$\Sigma_2 = \begin{pmatrix} 1 & & & & & \\ .08 & 1 & & & & \\ .10 & .12 & 1 & & & \\ .12 & -.10 & -.08 & 1 & & \\ -.10 & -.08 & .08 & .10 & 1 & \\ -.08 & .10 & -.01 & .08 & .12 & 1 \end{pmatrix}$$

cuyos valores propios son $\lambda_1 = 1.282, \lambda_2 = 1.253, \lambda_3 = 1.120, \lambda_4 = 1.056, \lambda_5 = 0.719, \lambda_6 = 0.697$.

La matriz Σ_1 fue utilizada por Devlin en [2], la misma tiene correlaciones que varían entre .10 y .95, de tal manera que se prueba la habilidad del estimador para detectar esta variedad de correlaciones, además, λ_6 es muy cercano a cero lo que indica que la matriz es "casi" singular, esto permite medir el comportamiento de los estimadores en condiciones de "casi" singularidad, finalmente, al igual que las correlaciones, los valores propios asociados a Σ_1 también varían ampliamente.

La matriz Σ_2 en cambio, presenta correlaciones y valores propios con poca diferencia, así se mide la capacidad del estimador ante matrices con estas poca variabilidad en sus elementos y los valores propios asociados a la misma.

Por otro lado, para la simulación de Monte Carlo cada escenario se repite $N = 500$ veces y el software utilizado para el efecto es R versión 2.8.0 [10].

3.3 Criterios de Evaluación

Con el propósito de evaluar la Eficiencia y Estabilidad de los Estimadores estudiados; se define

$$\bar{\bar{\mathbf{x}}}^{(e)} = \left\{ \frac{1}{N} \sum_1^N \bar{\mathbf{x}}^{(e)} \right\} \text{ el Vector Promedio de las}$$

estimaciones para el Vectores de Medias correspondiente al estimador e en las N repeticiones, $\bar{\mathbf{x}}_{\max}^{(e)} = \max \{ \bar{\mathbf{x}}^{(e)} \}$ el vector que representa el Vector “más alejado”, en términos de la distancia Euclidiana, de los Vectores de Medias correspondientes al estimador e en las N repeticiones, entonces se obtiene el Sesgo Promedio Total y Sesgo Máximo Total calculando la Distancia Euclidiana entre el Vector de Medias Real con $\bar{\bar{\mathbf{x}}}^{(e)}$ y $\bar{\mathbf{x}}_{\max}^{(e)}$ respectivamente; se define también $sd_{Total}(\bar{\mathbf{x}}^{(e)}) = \sum sd[\bar{\mathbf{x}}^{(e)}]_i$ la Suma de las Desviaciones Estándar de cada componente del Vector de Medias correspondiente al estimador e , tal que $sd_{Total}(\bar{\mathbf{x}}^{(e)})$ representa la Desviación Total.

Por último, bajo el supuesto que el mejor estimador es el que presente un mejor balance entre: su comportamiento promedio, su “peor” comportamiento y la dispersión de sus estimaciones, se establece un índice que es la media aritmética entre el Sesgo Promedio Total, Sesgo Máximo Total y la Desviación Total, de tal manera que el estimador que presente menor índice será el más Eficiente y Estable en cada escenario estudiado, estableciendo igual importancia a las tres componentes.

Sin embargo, en situaciones reales, no se conoce cuan contaminada esta la muestra, ni que Distribución genera la contaminación, etc. solo se tiene la matriz de observaciones y el tamaño muestral, por lo que se explora el comportamiento de los estimadores en función del Sesgo y Curtosis Muestral, con el propósito de que dichas medidas junto con el tamaño de la muestra sirvan como criterios para decidir que estimador utilizar.

4. Resumen de Resultados

Se simula en total 184 escenarios, en esta sección se presenta un resumen de los resultados obtenidos.

4.1 Análisis por escenario simulado

En la Tabla 1 se muestra el Sesgo Promedio Total, Sesgo Máximo Total, Desviación Total e Índice planteado para los Estimadores estudiados en el escenario de muestras generadas a partir de una Distribución Normal sin contaminar, se puede ver entonces que los Estimadores Robustos son comparables a los Estimadores de Máxima Verosimilitud para este caso, comportamiento parecido se da cuando $\Sigma = \Sigma_2$, es decir, se verifica que los estimadores Robustos Estudiados son Eficientes.

Considerando el Sesgo Promedio Total como medida de la “bondad” del Estimador, en los escenarios simulados el Estimador Clásico presenta su “peor” rendimiento al tratar de estimar la Matriz de Covarianzas y los Valores Propios asociados a la misma.

En general, cuando $\Sigma = \Sigma_1$ y la muestra es generada a partir de una Población Normal, los estimadores en general sobreestiman el primer y segundo valor propio para luego estimar con error “pequeño” los valores propios restantes, para este caso, los Estimadores Clásico y MCD son los que sobreestiman con mayor error el primer valor propio, sin embargo, el Estimador MCD reduce considerablemente su error mientras aumenta el tamaño muestral. Mientras que, cuando $\Sigma = \Sigma_1$ los Estimadores sobreestiman el primer y segundo valor propio, sin embargo, al final generalmente subestiman el menor valor propio, para esta matriz se puede observar que para $k=10$ y 20 que el estimador Clásico se aleja de los Estimadores Robustos presentando mayor error en si estimación

El tamaño muestral influye en la estimación, pues en todos los estimadores se cumple que al aumentar el tamaño muestral se disminuye el Sesgo Promedio

Tabla 1. Resultados para Población Normal $N(\mathbf{0}, \Sigma)$, $\Sigma = \Sigma_1$ sin Contaminar

		5				k				20			
		Prom	Max	Desv	Indice	Prom	Max	Desv	Indice	Prom	Max	Desv	Indice
Vector de Medias	Clásico	0.023	1.496	0.693	0.794	0.009	0.945	0.433	0.463	0.008	0.697	0.307	0.338
	Mhuber	0.022	1.459	0.636	0.706	0.012	0.966	0.433	0.470	0.010	0.761	0.322	0.364
	T-Bic	0.025	1.747	0.811	0.861	0.010	1.098	0.479	0.529	0.011	0.805	0.337	0.384
	Bic	0.023	1.495	0.650	0.722	0.012	1.019	0.443	0.491	0.012	0.815	0.328	0.385
	MCD	0.023	1.801	0.779	0.868	0.010	1.223	0.507	0.580	0.011	0.828	0.356	0.398
	DS	0.028	1.961	0.848	0.946	0.015	1.219	0.523	0.586	0.013	0.867	0.366	0.415
Valores Propios	Clásico	0.804	7.920	1.425	3.383	0.433	3.719	0.870	1.674	0.240	2.200	0.542	0.994
	Mhuber	0.484	2.552	0.857	1.297	0.294	2.122	0.672	1.029	0.186	1.264	0.440	0.630
	T-Bic	1.027	4.843	1.449	2.440	0.422	2.610	0.807	1.280	0.226	1.636	0.525	0.796
	Bic	0.613	3.922	0.896	1.810	0.348	2.639	0.725	1.237	0.211	1.416	0.505	0.711
	MCD	1.457	7.660	1.744	3.621	0.589	4.041	0.942	1.857	0.268	2.164	0.546	0.993
	DS	1.278	18.588	2.392	7.419	0.542	3.171	0.860	1.524	0.275	1.955	0.596	0.942
Matriz de Covarianzas	Clásico	0.666	9.778	1.965	4.136	0.373	4.539	1.229	2.047	0.197	2.778	0.779	1.251
	Mhuber	0.060	3.566	1.357	1.661	0.041	2.422	0.958	1.141	0.026	1.710	0.659	0.798
	T-Bic	0.288	5.945	2.200	2.811	0.069	3.164	1.246	1.493	0.033	2.122	0.782	0.979
	Bic	0.065	4.243	1.445	1.918	0.027	2.872	1.080	1.327	0.025	1.850	0.725	0.867
	MCD	1.593	9.632	2.940	4.721	0.548	4.731	1.532	2.270	0.193	2.491	0.884	1.189
	DS	0.952	17.308	2.761	7.007	0.062	3.921	1.455	1.813	0.027	2.305	0.884	1.072

Tabla 2. Resultados para Población Normal $N(\mathbf{0}, \Sigma)$, $\Sigma = \Sigma_1$ Contaminada $\varepsilon = 0.10$ y $H = N(0.537\mathbf{a}_6, 9\Sigma_1)$

	5				k				20			
	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	0.032 ± 0.792	1.648	0.824	0.013 ± 0.572	1.241	0.609	0.014 ± 0.387	0.926	0.442		
	Mhuber	0.016 ± 0.649	1.455	0.707	0.009 ± 0.457	1.061	0.509	0.012 ± 0.321	0.809	0.380		
	T-Bic	0.015 ± 0.793	1.707	0.838	0.011 ± 0.455	1.078	0.515	0.012 ± 0.322	0.758	0.364		
	Bic	0.012 ± 0.651	1.424	0.696	0.008 ± 0.448	1.022	0.493	0.012 ± 0.318	0.770	0.367		
	MCD	0.016 ± 0.762	1.664	0.814	0.014 ± 0.507	1.187	0.569	0.011 ± 0.329	0.830	0.390		
	DS	0.008 ± 0.756	1.611	0.792	0.013 ± 0.505	1.119	0.546	0.014 ± 0.328	0.764	0.368		
Valores Propios	Clásico	4.043 ± 2.526	12.767	6.445	3.213 ± 1.742	13.298	6.085	2.619 ± 1.166	5.937	3.241		
	Mhuber	1.057 ± 0.792	4.072	1.974	0.651 ± 0.676	2.274	1.200	0.465 ± 0.521	1.646	0.877		
	T-Bic	1.566 ± 1.446	8.265	3.759	0.685 ± 0.755	2.378	1.273	0.441 ± 0.571	1.798	0.936		
	Bic	0.919 ± 0.976	3.372	1.758	0.550 ± 0.698	2.600	1.282	0.390 ± 0.571	1.682	0.881		
	MCD	2.794 ± 1.807	9.425	4.675	1.142 ± 0.901	4.077	2.040	0.638 ± 0.624	2.249	1.170		
	DS	1.447 ± 1.442	10.816	4.588	0.727 ± 0.837	3.643	1.736	0.445 ± 0.583	1.735	0.921		
Matriz de Covarianzas	Clásico	2.666 ± 3.716	13.479	6.620	2.394 ± 2.619	11.817	5.610	2.129 ± 1.748	6.455	3.444		
	Mhuber	0.511 ± 1.217	5.035	2.254	0.361 ± 1.042	2.799	1.401	0.298 ± 0.734	1.870	0.967		
	T-Bic	0.530 ± 2.297	8.458	3.762	0.273 ± 1.219	3.272	1.588	0.228 ± 0.802	2.096	1.042		
	Bic	0.269 ± 1.553	4.529	2.117	0.200 ± 1.085	3.012	1.432	0.190 ± 0.763	1.950	0.968		
	MCD	1.725 ± 2.884	11.164	5.258	0.692 ± 1.514	4.683	2.296	0.422 ± 0.892	2.630	1.315		
	DS	0.465 ± 2.095	9.738	4.099	0.214 ± 1.347	4.117	1.893	0.198 ± 0.852	2.205	1.085		

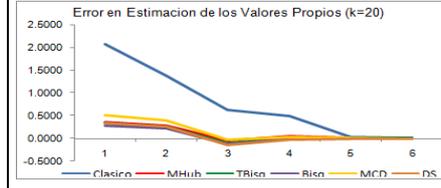
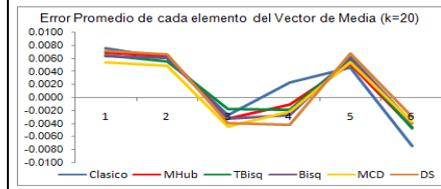


Tabla 3. Resultados para Población Normal $N(\mathbf{0}, \Sigma)$, $\Sigma = \Sigma_1$ Contaminada $\varepsilon = 0.10$ y $H = N(0.537\mathbf{a}_6, 9\Sigma_1)$

	5				k				20			
	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	0.066 ± 0.757	1.833	0.885	0.063 ± 0.534	1.243	0.613	0.054 ± 0.377	0.919	0.450		
	Mhuber	0.018 ± 0.614	1.384	0.672	0.022 ± 0.432	0.964	0.473	0.018 ± 0.313	0.690	0.340		
	T-Bic	0.027 ± 0.760	1.714	0.834	0.018 ± 0.453	1.111	0.527	0.009 ± 0.317	0.684	0.337		
	Bic	0.012 ± 0.641	1.449	0.701	0.017 ± 0.433	1.044	0.498	0.011 ± 0.313	0.680	0.335		
	MCD	0.016 ± 0.740	1.626	0.794	0.019 ± 0.475	1.227	0.574	0.010 ± 0.329	0.688	0.343		
	DS	0.019 ± 0.765	1.649	0.811	0.019 ± 0.514	1.175	0.569	0.011 ± 0.337	0.685	0.345		
Valores Propios	Clásico	3.924 ± 2.694	13.365	6.661	3.097 ± 1.703	8.199	4.333	2.665 ± 1.184	6.464	3.438		
	Mhuber	0.596 ± 0.810	2.477	1.294	0.392 ± 0.609	1.587	0.863	0.254 ± 0.492	1.167	0.638		
	T-Bic	1.341 ± 1.332	5.705	2.793	0.675 ± 0.739	2.608	1.341	0.456 ± 0.574	1.918	0.982		
	Bic	0.774 ± 1.003	3.152	1.643	0.552 ± 0.693	2.241	1.162	0.409 ± 0.559	1.718	0.895		
	MCD	2.426 ± 1.629	7.117	3.724	1.097 ± 0.919	3.485	1.834	0.669 ± 0.628	2.909	1.402		
	DS	1.132 ± 1.291	12.580	5.001	0.676 ± 0.814	2.528	1.338	0.452 ± 0.603	1.819	0.958		
Matriz de Covarianzas	Clásico	2.600 ± 3.590	13.959	6.716	2.322 ± 2.547	9.080	4.650	2.180 ± 1.834	6.691	3.568		
	Mhuber	0.044 ± 1.275	3.384	1.588	0.032 ± 0.900	2.257	1.063	0.021 ± 0.665	1.530	0.739		
	T-Bic	0.352 ± 2.177	6.547	3.025	0.256 ± 1.198	3.113	1.522	0.244 ± 0.845	2.104	1.064		
	Bic	0.136 ± 1.539	4.379	2.018	0.190 ± 1.086	2.845	1.374	0.210 ± 0.790	2.061	1.020		
	MCD	1.460 ± 2.742	9.062	4.421	0.653 ± 1.473	3.940	2.022	0.444 ± 0.943	3.006	1.464		
	DS	0.136 ± 1.927	10.865	4.309	0.150 ± 1.290	3.350	1.597	0.202 ± 0.884	2.197	1.094		

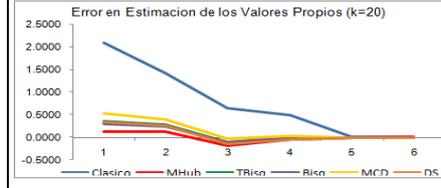
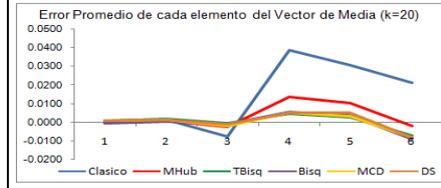


Tabla 4. Resultados para Población Normal $N(\mathbf{0}, \Sigma)$, $\Sigma = \Sigma_1$ Contaminada por variable $\varepsilon = 0.10$ y $H = U(-5, 5)$

	5				k				20			
	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	0.017 ± 0.695	1.464	0.725	0.016 ± 0.470	1.034	0.507	0.009 ± 0.321	0.714	0.348		
	Mhuber	0.026 ± 0.606	1.320	0.651	0.014 ± 0.432	1.008	0.485	0.008 ± 0.302	0.678	0.329		
	T-Bic	0.020 ± 0.748	1.521	0.783	0.016 ± 0.464	1.095	0.525	0.009 ± 0.316	0.708	0.344		
	Bic	0.026 ± 0.629	1.370	0.675	0.014 ± 0.443	0.987	0.481	0.008 ± 0.308	0.685	0.334		
	MCD	0.019 ± 0.733	1.533	0.761	0.015 ± 0.494	1.130	0.547	0.011 ± 0.324	0.687	0.341		
	DS	0.024 ± 0.725	1.663	0.804	0.013 ± 0.496	1.159	0.557	0.013 ± 0.335	0.724	0.357		
Valores Propios	Clásico	1.742 ± 1.166	4.799	2.589	1.153 ± 0.820	3.408	1.794	0.706 ± 0.562	2.623	1.117		
	Mhuber	0.575 ± 0.779	2.221	1.192	0.394 ± 0.538	1.736	0.889	0.248 ± 0.440	1.186	0.625		
	T-Bic	1.433 ± 1.374	6.355	3.054	0.618 ± 0.733	2.773	1.375	0.326 ± 0.537	1.594	0.819		
	Bic	0.888 ± 1.023	3.869	1.927	0.504 ± 0.674	2.731	1.303	0.284 ± 0.529	1.595	0.799		
	MCD	2.607 ± 1.594	8.126	4.109	1.087 ± 0.905	3.623	1.872	0.464 ± 0.568	1.892	0.975		
	DS	1.136 ± 1.136	6.041	2.771	0.665 ± 0.783	2.598	1.348	0.351 ± 0.579	1.803	0.911		
Matriz de Covarianzas	Clásico	1.372 ± 2.299	7.008	3.560	0.973 ± 1.579	4.792	2.448	0.612 ± 0.991	2.753	1.452		
	Mhuber	0.269 ± 1.173	2.868	1.437	0.191 ± 0.858	2.099	1.049	0.101 ± 0.624	1.546	0.757		
	T-Bic	0.481 ± 2.186	7.273	3.313	0.194 ± 1.167	3.352	1.571	0.062 ± 0.774	2.000	0.952		
	Bic	0.275 ± 1.568	4.397	2.080	0.133 ± 1.038	3.018	1.396	0.051 ± 0.727	1.955	0.911		
	MCD	1.629 ± 2.817	8.813	4.419	0.622 ± 1.461	4.111	2.064	0.248 ± 0.847	2.446	1.180		
	DS	0.190 ± 1.886	7.079	3.052	0.122 ± 1.266	3.475	1.618	0.054 ± 0.860	2.311	1.075		

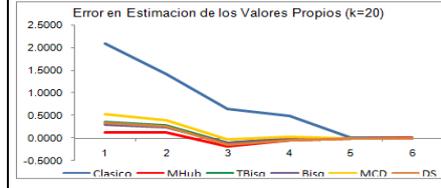
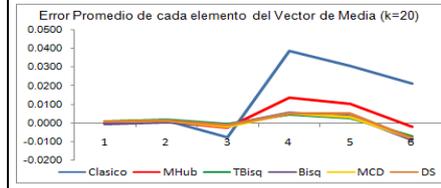


Tabla 5. Resultados para Población Normal CAU($\mathbf{0}, \Sigma$), $\Sigma = \Sigma_1$ sin Contaminar

	5				k				20			
	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	77.378 ± 4.978	---	---	1.913 ± 4.655	716.591	241.053	0.984 ± 4.843	391.010	132.279		
	Mhuber	0.047 ± 0.898	2.361	1.102	0.023 ± 0.620	1.428	0.690	0.011 ± 0.429	0.967	0.469		
	T-Bic	0.041 ± 0.962	3.014	1.339	0.022 ± 0.663	1.613	0.766	0.026 ± 0.479	1.048	0.518		
	Bic	0.045 ± 0.840	2.218	1.035	0.018 ± 0.590	1.367	0.658	0.022 ± 0.407	0.935	0.454		
	MCD	0.054 ± 0.939	2.241	1.078	0.013 ± 0.629	1.621	0.754	0.015 ± 0.439	0.960	0.512		
	DS	0.034 ± 0.835	1.940	0.937	0.016 ± 0.597	1.514	0.709	0.016 ± 0.425	0.950	0.464		
Valores Propios	Clásico	---	---	---	---	---	---	---	---	---		
	Mhuber	8.841 ± 6.787	51.060	22.229	6.753 ± 4.074	26.143	12.323	5.906 ± 2.740	19.202	9.283		
	T-Bic	10.176 ± 8.344	93.705	37.409	6.729 ± 5.045	25.051	12.275	5.225 ± 2.886	23.001	10.371		
	Bic	6.918 ± 6.212	40.263	17.798	5.017 ± 3.649	18.338	9.001	4.390 ± 2.411	14.457	7.086		
	MCD	9.065 ± 7.386	42.272	19.575	4.009 ± 3.068	13.959	7.012	2.746 ± 1.789	9.358	4.631		
	DS	11.871 ± 10.876	190.074	70.940	4.765 ± 3.569	36.838	15.057	4.115 ± 2.382	12.980	6.492		
Matriz de Covarianzas	Clásico	---	---	---	---	---	---	---	---	---		
	Mhuber	6.996 ± 6.972	46.960	20.317	5.665 ± 4.496	23.372	11.178	5.121 ± 3.079	16.931	8.377		
	T-Bic	6.705 ± 8.220	79.492	31.473	5.107 ± 5.390	28.590	13.029	4.298 ± 3.556	19.172	9.009		
	Bic	4.797 ± 6.068	42.879	17.915	3.970 ± 3.992	19.123	9.029	3.698 ± 2.841	13.856	6.798		
	MCD	6.130 ± 7.542	45.319	19.863	2							

Total y la Desviación Total.

El estimador T-Bicuadrático o T-Bponderado es casi siempre superado por los demás Estimadores Robustos, situación que era previsible pues este estimador fue construido para mejorar el comportamiento de los estimadores S para “grandes” número de variables.

La Tabla 3 muestra que a pesar de que la mayoría de los Estimadores Robustos considerados en el presente estudio han sido construidos bajo el supuesto de contaminación simétrica, al ser sometidos a contaminación asimétrica se comportan de manera similar a cuando la contaminación es simétrica.

Así también, los estimadores han sido construidos siguiendo el entorno de contaminación (2.1), sin embargo al contaminar por variable, según lo explicado, los Estimadores brindan “buenas” estimaciones, veáse la Tabla 4; sin embargo, en el presente estudio no se analiza las consecuencias que puede tener una contaminación por variable en las Técnicas de Análisis Multivariado.

Cuando las Muestras son generadas a partir de la Distribución Cauchy, el Estimador Clásico en todos los casos brinda las estimaciones más distantes de cada parámetro poblacional, la Tabla 5 presenta el caso cuando se genera muestras a partir de una Cauchy sin contaminar. Además, en estos casos, el algoritmo del Estimador M de Huber puede no converger a una solución. Los valores propios son siempre sobreestimados, en la parte derecha de la Tabla 5 se muestra los errores absolutos y relativos de la estimación de los valores propios para cuando k=20.

A las estimaciones brindadas por el estimador DS en varias ocasiones le corresponde un Sesgo Promedio Total menor al Sesgo presentado por los demás estimadores, sin embargo su “peor” estimación puede incluso encontrarse más alejada que la “peor” estimación utilizando el método Clásico; esto genera que no presente un Balance adecuado y no sea considerado el mejor estimador (de acuerdo al índice planteado) en los escenarios simulados.

4.2 Análisis por Tamaño Muestral, Sesgo y Curtosis

El objetivo en esta parte del estudio es tratar de establecer una “regla empírica” donde se utilice al tamaño de la Muestra, al Sesgo y la Curtosis Muestral como elementos de decisión sobre el estimador a utilizar.

En cada muestra generada se ha calculado el Sesgo y Curtosis Muestral, además la distancia Euclidiana entre los valores reales del Vector de Medias, Matriz de Covarianzas y Valores Propios asociados a la misma versus la estimación brindada por cada Método; a partir de ello se establece el estimador que menor distancia presenta en cada caso. Los resultados se resumen a continuación.

En la Tabla 6 y 7 se muestra la tabla de resumen de los resultados obtenidos para estimadores estudiados

en función de los valores de Sesgo y Curtosis para muestra tamaño 30 y 60 respectivamente, en las Tablas se puede observar los tres estimadores que con mayor frecuencia presentan la menor distancia entre el valor del parámetro y la estimación brindada por los mismos; para cada estimador se presenta la distancia Promedio y su Desviación Estándar, además de la Frecuencia Relativa de ser el estimador con “más cercano” al valor real.

Así, por ejemplo cuando el Sesgo Muestral es menor a 45 y la Curtosis Muestral se encuentra en el intervalo [20, 65) el 25,3% de las veces el estimador Clásico es el que presenta la menor distancia para la estimación del Vector de Medias, siendo su distancia Promedio igual a 0.503±0.174. Mientras que para los Valores Propios y la Matriz de Varianzas y Covarianzas el Estimador M de Huber en más del 50% de los casos brinda la estimación “más cercana”. Además, para valores de Sesgo Muestral mayores a 45 y menores a 90, el Estimador DS es con mayor frecuencia el “mejor” estimador, sin embargo su Promedio y Desviación Estándar de la distancia para el caso de los Valores Propios y Matriz de Covarianzas es mayor a la presentada por los otros dos estimadores mostrados en cada caso, por ejemplo, cuando la Curtosis se encuentra en el intervalo [110, 155) la distancia promedio correspondiente al Estimador DS al estimar la Matriz de Varianzas y Covarianzas es 11.075 mientras que para el estimador Bicuadrático es 8.397, situación que puede ser causado por la alta dispersión observada en el Estimador DS.

De acuerdo a la Tabla 7, cuando el Sesgo Muestral es mayor a 90 y menor a 180 y la Curtosis Muestral se encuentra en el intervalo [130, 220) el estimador MCD tiene una frecuencia igual a 0.688 de ser el estimador

Tabla 6. Resultados en función del Sesgo y la Curtosis para tamaño muestral 30

		Curtosis Muestral						
		[20, 65)			[65, 110)			
		Distancia	Frecuencia	Distancia	Frecuencia			
Sesgo Muestral	[0, 45)	Localizacion	1 Clas	0.503 ± 0.174	0.253	MHub	0.502 ± 0.205	0.239
			2 MHub	0.450 ± 0.147	0.212	Bis	0.502 ± 0.211	0.205
			3 Bis	0.459 ± 0.151	0.210	DS	0.537 ± 0.202	0.174
		Valores Propios	1 MHub	1.154 ± 1.155	0.535	MHub	3.598 ± 4.828	0.389
			2 Bis	1.212 ± 1.191	0.323	Bis	3.510 ± 4.979	0.306
			3 DS	1.919 ± 2.271	0.068	DS	5.185 ± 10.074	0.164
		Matriz de Covarianzas	1 MHub	1.229 ± 1.020	0.548	MHub	3.414 ± 4.317	0.419
			2 Bis	1.266 ± 1.030	0.338	Bis	3.287 ± 4.334	0.289
			3 DS	1.871 ± 1.865	0.058	DS	4.739 ± 8.726	0.156
Sesgo Muestral	[45, 90)	Localizacion	1 DS	0.596 ± 0.231	0.304	DS	0.815 ± 0.238	0.339
			2 MHub	0.619 ± 0.247	0.264	MHub	0.855 ± 0.254	0.262
			3 Bis	0.620 ± 0.252	0.185	Bis	0.849 ± 0.254	0.178
		Valores Propios	1 DS	10.648 ± 13.266	0.321	DS	12.391 ± 16.178	0.366
			2 Bis	7.840 ± 6.283	0.279	Bis	9.334 ± 6.217	0.228
			3 MHub	8.372 ± 6.229	0.191	MCD	9.607 ± 5.930	0.195
		Matriz de Covarianzas	1 DS	9.538 ± 11.528	0.322	DS	11.075 ± 14.074	0.376
			2 Bis	6.899 ± 5.442	0.278	Bis	8.397 ± 5.365	0.220
			3 MHub	7.891 ± 5.541	0.184	MCD	8.540 ± 5.075	0.204
Sesgo Muestral	[90, 135)	Localizacion	1 DS	0.619 ± 0.236	0.334	DS	0.627 ± 0.216	0.286
			2 MHub	0.666 ± 0.255	0.251	MHub	0.661 ± 0.248	0.259
			3 Bis	0.654 ± 0.247	0.191	Bis	0.849 ± 0.240	0.222
		Valores Propios	1 DS	11.190 ± 12.920	0.422	DS	13.493 ± 10.911	0.333
			2 Bis	8.957 ± 6.306	0.238	MCD	10.479 ± 7.010	0.241
			3 MCD	9.368 ± 5.947	0.167	MHub	11.112 ± 6.280	0.204
		Matriz de Covarianzas	1 DS	10.044 ± 11.238	0.429	DS	12.128 ± 9.662	0.352
			2 Bis	8.102 ± 5.460	0.224	MCD	9.362 ± 5.914	0.259
			3 MCD	8.387 ± 5.080	0.180	Bis	9.249 ± 5.668	0.204
Sesgo Muestral	[135, 180)	Localizacion	1 DS	0.619 ± 0.262	0.371			
			2 Bis	0.638 ± 0.257	0.257			
			3 MHub	0.888 ± 0.256	0.143			
		Valores Propios	1 Bis	10.560 ± 7.907	0.314			
			2 DS	13.448 ± 13.000	0.257			
			3 MCD	10.199 ± 6.006	0.229			
		Matriz de Covarianzas	1 Bis	9.532 ± 6.750	0.371			
			2 DS	12.125 ± 11.515	0.257			
			3 MCD	8.142 ± 5.161	0.229			

Tabla 6. Resultados en función del Sesgo y la Curtosis para tamaño muestral 60

		Curtosis Muestral						
		[40, 130]			[130, 220]			
		Distancia	Frecuencia	Distancia	Frecuencia			
Sesgo Muestral	[0, 90]	Localización	1 Clas	0.383 ± 0.243	0.229	MHub	0.402 ± 0.156	0.265
			2 MHub	0.315 ± 0.110	0.208	DS	0.395 ± 0.147	0.252
			3 TBis	0.335 ± 0.120	0.176	Bis	0.397 ± 0.152	0.160
	Valores Propios	1 MHub	0.994 ± 1.569	0.494	MCD	4.004 ± 2.707	0.579	
		2 DS	0.985 ± 1.407	0.246	DS	5.346 ± 4.914	0.155	
		3 DS	1.144 ± 1.631	0.107	MHub	6.076 ± 4.204	0.131	
	Matriz de Covarianzas	1 MHub	1.053 ± 1.415	0.537	MCD	3.700 ± 2.366	0.152	
		2 Bis	1.042 ± 1.248	0.257	MHub	5.657 ± 3.801	0.156	
		3 DS	1.201 ± 1.432	0.093	DS	4.927 ± 4.347	0.152	
[90, 180]	Localización	1 MHub	0.435 ± 0.160	0.289	MHub	0.423 ± 0.148	0.276	
		2 DS	0.420 ± 0.151	0.258	DS	0.414 ± 0.148	0.258	
		3 Bis	0.423 ± 0.155	0.198	Bis	0.416 ± 0.153	0.212	
	Valores Propios	1 MCD	4.593 ± 2.542	0.888	MCD	4.661 ± 2.468	0.753	
		2 DS	6.056 ± 4.611	0.181	DS	6.195 ± 4.328	0.158	
		3 Bis	6.183 ± 3.609	0.090	Bis	6.357 ± 3.430	0.068	
	Matriz de Covarianzas	1 MCD	4.225 ± 2.212	0.706	MCD	4.301 ± 2.193	0.779	
		2 DS	5.567 ± 4.081	0.178	DS	5.707 ± 3.872	0.148	
		3 Bis	5.705 ± 3.201	0.078	Bis	5.899 ± 3.075	0.005	
[180, 270]	Localización	1 MHub	0.441 ± 0.159	0.288	DS	0.417 ± 0.151	0.280	
		2 DS	0.421 ± 0.153	0.258	MHub	0.429 ± 0.133	0.274	
		3 Bis	0.425 ± 0.153	0.200	Bis	0.422 ± 0.138	0.194	
	Valores Propios	1 MCD	4.890 ± 2.705	0.738	MCD	5.256 ± 2.946	0.790	
		2 DS	6.514 ± 4.988	0.163	DS	6.840 ± 4.262	0.113	
		3 Bis	6.847 ± 3.855	0.073	Bis	7.143 ± 3.938	0.081	
	Matriz de Covarianzas	1 MCD	4.507 ± 2.371	0.759	MCD	4.858 ± 2.555	0.855	
		2 DS	6.011 ± 4.426	0.157	DS	6.336 ± 3.745	0.097	
		3 Bis	6.184 ± 3.431	0.064	Bis	6.857 ± 3.458	0.016	
[270, 360]	Localización	1 DS	0.436 ± 0.150	0.319	DS	0.436 ± 0.150	0.319	
		2 MHub	0.456 ± 0.152	0.278	MHub	0.447 ± 0.160	0.250	
		3 Bis	0.447 ± 0.160	0.250	Bis	0.447 ± 0.160	0.250	
	Valores Propios	1 MCD	5.011 ± 2.528	0.722	MCD	5.011 ± 2.528	0.722	
		2 DS	6.345 ± 3.730	0.208	DS	6.345 ± 3.730	0.208	
		3 Bis	6.539 ± 3.277	0.056	Bis	6.539 ± 3.277	0.056	
	Matriz de Covarianzas	1 MCD	4.707 ± 2.329	0.778	MCD	4.707 ± 2.329	0.778	
		2 DS	5.994 ± 3.430	0.181	DS	5.994 ± 3.430	0.181	
		3 Bis	6.201 ± 3.061	0.028	Bis	6.201 ± 3.061	0.028	

“más cercano” a los Valores Propios, seguido por el estimador DS y Bicuadrático.

Nótese que el estimador que con mayor frecuencia es el mejor estimador de la Matriz de Covarianzas y los Valores Propios asociados a la misma es, para sesgo mayores a 90, el Estimador MCD

Para tamaño muestral igual a 120, cuando la Curtosis Muestral sea menor a 230 el estimador M de Huber o Bicuadrático son los que con ayor frecuencia son los “mejores” estimadores, mientras que para Curtosis Muestral sea mayor a 230 el mejor estimador es con mayor frecuencia el MCD

5. Conclusiones

Se confirma la sensibilidad de los Estimadores de Máxima Verosimilitud para el Vector de Medias y Matriz de Varianzas y Covarianzas ante desviaciones de la Distribución Normal Multivariada.

Cuando la muestra es generada a partir de la Distribución Normal Multivariada, para cualquier tamaño muestral los Estimadores M de Huber y Bicuadrático son generalmente los que alcanzan el menor índice al estimar el Vector de Medias, Matriz de Covarianzas y Valores Propios.

Cuando se considera muestras generadas a partir de una Distribución Cauchy, el Estimador Bicuadrático presenta el mejor comportamiento para $k=5$, mientras que para $k=10$ y 20 el Estimador MCD es el que “mejor” estima la Matriz de Covarianzas y los Valores propios, todo ello en base al índice planteado.

Para $\varepsilon = 30$ los estimadores Robustos estudiados, se comportan de forma parecida a cuando $\varepsilon = 0.05, 0.10$, se concluye entonces que el Punto de Ruptura, para 6 variables, es mayor a 0.3.

Finalmente, se recomienda la realización de próximos estudios que complementen el presente trabajo.

6. Referencias

- [1] ANDREWS, D. et al. (1972). Robust Estimates of Location: Survey and Advances. Princeton Univ. Press
- [2] DEVLIN, S., GNANADESIKAN, R., Y KETTENRING J. (1981). “Robust estimation of dispersion matrices and Principal Components”, Journal of the American Statistical Association; 76, 354-362.
- [3] HAMPEL, F. (1971). “A general qualitative definition of robustness”, The Annals of Mathematical Statistics; 42, 1887–1896
- [4] HAMPEL, F. (1974). “The influence curve and its role in robust estimation”, Journal of the American Statistical Association; 69, 383–393
- [5] HUBER, P. (1964). “Robust estimation of a location parameter”, The Annals of Mathematical Statistics; 35, 73–101
- [6] HUBER, P. (1972). “The 1972 wald lecture. Robust Statistics: A Review”, The Annals of Mathematical Statistics; 43, 1041–1067
- [7] HUBER, P. (1979). “Robust Statistical Procedures”, Society for industrial and Applied Mathematics
- [8] MARONNA, R., MARTIN, D. Y YOHAI, V. (2006). “Robust Statistics: Theory and Methods”, John Wiley & Sons, New York–USA
- [9] MARONNA, R. (1976). “Robust M-Estimators of Multivariate Location and Scatter”, The Annals of Statistics; 4, 51–67
- [10] R DEVELOPMENT CORE TEAM (2008). “R: A language and environment for statistical computing. R Foundation for Statistical Computing”, Vienna, Austria.
- [11] ROCKE D. (1996) “Robustness properties of multivariate location and shape in high dimension”, The Annals of Statistics; 24, 1327–1345
- [12] RENCHER, A. (1998) “Multivariate Statistical Inference and Applications”, John Wiley & Sons, New York–USA
- [13] SEBER, G. (1984) “Multivariate observation”, John Wiley & Sons, New York–USA
- [14] TUKEY, J. (1960) “A survey of sampling from contaminated distributions”, Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling”, Olkin Ediciones Stanford University Press, Stanford. CA.
- [15] YOHAI, V.. (2004) “Notas de probabilidad y estadística”, [en línea] Universidad de Buenos Aires, Buenos Aires – Argentina
- [16] ZAMAR, R. (1994) “Estimación Robusta” , Estadística Española; 36, 327–387
- [17] ZUO, Y. (2006) “Robust location and scatter estimators in multivariate analysis” (invited book chapter to honor Peter Bickel on his 65th Birthday), The Frontiers in Statistics, Imperial College Press