

ESTIMADORES ROBUSTOS PARA EL VECTOR DE MEDIAS Y LA MATRIZ DE VARIANZAS Y COVARIANZAS DE VECTORES ALEATORIOS MULTIVARIADOS

¹Montaño Néstor, ²Zurita Gaudencio

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 la muestra esté "contaminada" 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 Estimator MCD is the best.

Key words: Multivariate Estimation, Robustness, Monte Carlo

RECIBIDO: Agosto, 2010
Aceptado: Septiembre, 2010

1. INTRODUCCIÓN

Para estimar parámetros poblacionales se utiliza información obtenida a partir de los datos que proporciona una Muestra; en la práctica se verifica que un alto porcentaje de las mediciones que se efectúan, por diferentes razones, contienen errores de medición 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?.

Esta situación origina la búsqueda de estimadores robustos, es decir, estimadores "poco" sensibles a errores de medición o a valores aberrantes.

2. ESTIMACIÓN ROBUSTA

Para modelar la situación en que la mayoría de las observaciones provienen de una distribución F_θ , pero una pequeña fracción ε de las observaciones son valores atípicos generados por otra distribución H , Tukey en [15] 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 la proporción de contaminación.

Se espera que los Estimadores Robustos cumplan con dos requerimientos: Eficiencia y Estabilidad. Se dice que un estimador es Eficiente si sus estimaciones son "buenas" aunque no exista contaminación, es decir que por ejemplo, el Estimador Robusto debe ser comparable con el Estimador de Máxima Verosimilitud (al que de aquí en adelante llamado Estimador Clásico). Para el caso multivariado este requerimiento implica lo siguiente:

i) Sea $(\hat{\mu}_n, \hat{\Sigma}_n)$ los estimadores de localización y dispersión para una muestra de tamaño n y sean

¹ Montaño Nestor, Ingeniero en Estadística e Informática, Escuela Superior Politécnica del Litoral (ESPOL); (e_mail: rmontano@espol.edu.ec).

² Zurita Gaudencio, M.Sc., Profesor de la Escuela Superior Politécnica del Litoral (ESPOL); Director del Centro de Estudios e Investigaciones Estadísticas ICM – ESPOL. (e_mail: gzurita@espol.edu.ec).

$(\hat{\mu}_\infty, \hat{\Sigma}_\infty)$ sus valores asintóticos. Si

$X_i \sim N_p(\mu, \Sigma)$ entonces $\hat{\mu}_\infty = \mu$ y $\hat{\Sigma}_\infty = c\Sigma$ donde c es una constante; y,

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

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

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

donde $q = \frac{p(p+1)}{2}$ y $\text{vech}(\Sigma)$ es el vector que contiene los q elementos de la triangular inferior de Σ .

Un estimador se lo considera estable si su "buen" comportamiento se preserva incluso ante la presencia de contaminación, esto es cuando F varía sobre \mathcal{F}_ε . Para evaluar la estabilidad se han propuesto varias medidas, como Sesgo Asintótico Máximo y la Varianza asintótica Máxima las cuales miden el "peor" comportamiento del estimador para todo $\varepsilon < \varepsilon^*$; también se tiene el Punto de Ruptura Asintótico donde la idea es representar 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:

Si $y = Ax + b$ entonces

$$\hat{\mu}(y) = A\hat{\mu}(x) + b$$

$$\hat{\Sigma}(y) = A\hat{\Sigma}(x)A^T$$

Se han propuesto diversos Estimadores Robustos, sin embargo para este estudio se ha escogido a considerados "más populares", a continuación definirá a cada uno de ellos.

En lo siguiente, la distancia de Mahalanobis representada por:

$$d_i = d_i(x_i, \mu, \Sigma) = (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$

2.1 ESTIMADOR M MULTIVARIADO

Maronna en [9] extiende los estimadores M propuestos por Huber en [5] a espacios p -dimensionales; así, define al estimador M como la solución de

$$\sum_{i=1}^n [W_1(d_i)](x_i - \hat{\mu}) = 0$$

$$\frac{1}{n} \sum_{i=1}^n [W_2(d_i^2)](x_i - \hat{\mu})(x_i - \hat{\mu})^T = \hat{\Sigma}$$

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

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

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

2.2 ESTIMADOR S BICUADRÁTICO MULTIVARIADO

Se define al Estimador S Bicuadrático o Bponderado multivariado como $\hat{\mu} \in R^p$ y $\hat{\Sigma} \in S_p$

que minimiza $\hat{\sigma}(d_i(x, \hat{\mu}, \hat{\Sigma}))$ con $|\hat{\Sigma}| = 1$, donde $\hat{\sigma}$ es un Estimador M univariado de Escala que satisface

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

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

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 continua se aproximan al vector de medias muestral y la matriz de varianzas y covarianzas muestral, esto implica una pérdida de robustez; Rocke en [12] 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 $\lim_{p \rightarrow \infty} \rho(d) = I(d > 1)$, donde $I(d > 1)$ es la función "indicador". El estimador T-Bicuadrático es un estimador que cumple las mismas condiciones del Estimador Bicuadrático 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 [18] indica que los estimadores MCD propuestos por Rousseeuw escogen h observaciones las cuales minimizan el determinante de la Matriz de Covarianzas Clásica, así, el estimador MCD de localización es el promedio de las h observaciones y el estimador MCD de escala es un múltiplo escalar de la Matriz de Covarianzas correspondiente a las h observaciones.

Este estimador es probablemente 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 x_i es una función de la “lejanía” de x_i , denotada por $t(x_i)$.

Es decir, siendo W_1 y W_2 dos funciones de ponderación, se define

$$\hat{\mu} = \frac{1}{\sum_{i=1}^n w_{i1}} \sum_{i=1}^n w_{i1} x_i$$

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

Con $w_{ij} = w_j(t(x_i))$, $j = 1, 2$

3. DETALLES DE LA SIMULACIÓN

Las medidas planteadas para evaluar 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 trabajo se generan muestras aleatorias a partir de distribuciones a las que se les manipula algunos parámetros con el objetivo de simular varias condiciones y así estudiar el comportamiento de los Estimadores Robustos bajo estas condiciones; uno de los estudios seminales en cuanto a comparar Estimadores Robustos fue el realizado por Andrews et al.[1]; en dicho estudio se dio a lugar a la noción de la tres "esquinas" para representar las posibles situaciones que se pueden 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 argumentos son el punto de partida del presente estudio, es decir, se consideraran muestras generadas a partir de la Distribución Normal Multivariada $N(0,\Sigma)$ y la Distribución Cauchy $CAU(0,\Sigma)$, considerando la diferencia entre ambas en el “peso” de sus colas.

Las muestras generadas son 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, $\varepsilon = 0.5$ 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 mayor 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 $N(0,9\Sigma)$, para seguir con la idea propuesta por Tukey en [15] y que además es la Distribución utilizada con más frecuencia 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 $U_{esf}(d)$, 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 otro lado, como se menciona en la segunda sección, el Entorno de Contaminación (2.1) indica que una proporción de vectores no siguen la distribución original, para el caso multivariado esto implica que existen dos escenarios posibles: ó todas las componentes de la observación están contaminadas ó ninguna está contaminada; 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 considera 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 . A continuación se presentan las dos Matrices de Varianzas y Convarianzas utilizadas para generar las muestras aleatorias,

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

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 $\lambda_1=1.282$, $\lambda_2=1.253$, $\lambda_3=1.120$, $\lambda_4=1.056$, $\lambda_5=0.719$ y $\lambda_6=0.697$.

La matriz Σ_1 fue utilizada por Devlin en [2], la misma tiene correlaciones que varían, es términos absolutos, entre 0 y 0.95, 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 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 [11].

3.3 CRITERIOS DE EVALUACIÓN

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

$$\bar{\mathbf{x}}^{(e)} = \frac{1}{N} \sum_{i=1}^N \bar{\mathbf{x}}^{(e)}$$

que es el Promedio de las estimaciones correspondientes a "e" para el Vector de Medias, donde $e = \{\text{Clásico, M Huber, T-Bicuadrático, Bicuadrático, MCD, DS}\}$;

$$\bar{\mathbf{x}}^{(e)}_{max} = \max \{ \bar{\mathbf{x}}^{(e)} \}$$

que representa a la estimación "más alejada" del verdadero valor del parámetro, en términos de la distancia Euclidiana, de entre los Vectores de Medias correspondientes al estimador e en las N repeticiones.

Con estos dos vectores se obtiene el Sesgo Promedio Total y Sesgo Máximo Total, haciendo

$$\text{Sesgo Promedio Total} = \text{Prom} \left(\bar{\mathbf{x}}^{(e)} \right) = \left(\bar{\bar{\mathbf{x}}}^{(e)} \right)^T \left(\bar{\mathbf{x}}^{(e)} \right)$$

$$\text{Sesgo Máximo T} = \text{Max} \left(\bar{\mathbf{x}}^{(e)} \right) = \left(\bar{\mathbf{x}}^{(e)}_{max} \right)^T \left(\bar{\mathbf{x}}^{(e)}_{max} \right)$$

es decir, calculando Distancia Euclidiana entre el verdadero valor del Vector de Medias "0" con $\bar{\bar{\mathbf{x}}}^{(e)}$ y con $\bar{\mathbf{x}}^{(e)}$ respectivamente.

Se define también

$$sd_{total} \left(\bar{\mathbf{x}}^{(e)} \right) = \sum_{i=1}^6 sd \left(\bar{\mathbf{x}}^{(e)} \right)_i$$

que es la suma de las Desviaciones Estándar de cada componente del Vector de Medias correspondiente al estimador e , de tal manera que $sd_{total} \left(\bar{\mathbf{x}}^{(e)} \right)$ representa la Desviación Total del estimador e .

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 ha elaborado un índice que es:

$$\text{Indice} = \frac{\left(\text{Prom}(\bar{\mathbf{x}}^{(e)}) + \text{Max}(\bar{\mathbf{x}}^{(e)}) + \text{sd}_{\text{total}}(\bar{\mathbf{x}}^{(e)}) \right)}{3} \quad (3.1)$$

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, nótese que se da la misma importancia a las tres medidas.

Por otro lado, en la práctica no se conoce cuan contaminada está la muestra, ni que Distribución genera la contaminación, etc. solo se tiene la matriz de observaciones y el tamaño muestral, por lo que en el presente estudio también 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, el análisis detallado de los mismos es presentado en [10], en esta sección se presenta algunos de los resultados obtenidos dando un breve resumen de los más relevantes.

4.1 ANÁLISIS POR ESCENARIO SIMULADO

Las tablas de resultado se encuentran divididas en tres secciones verticales, presentando los resultados para la estimación del vector de medias, los valores propios y la matriz de varianzas y covarianzas, y en tres secciones horizontales, correspondientes a cada tamaño muestral $k=5, 10$ y 20 . En cada sección se muestra en la columna denominada “Prom” el Sesgo Promedio Total, el Sesgo Máximo Total es presentado en la columna “Max”, mientras que la Desviación Total se presenta en la columna “Desv” y en la última columna se muestra el Índice (3.1).

La Tabla I presenta los resultados obtenidos para el escenario de muestras generadas a partir de una Distribución Normal $N(0, \Sigma)$, $\Sigma = \Sigma_1$ sin contaminar;

se puede ver que los Estimadores Robustos son comparables al Estimadores de Máxima Verosimilitud, pues en todos los casos sus resultados no difieren considerablemente de los resultados del estimador de máxima verosimilitud. En la simulación realizada en [10], para muestras generadas a partir de una Distribución Normal $N(0, \Sigma)$, $\Sigma = \Sigma_2$ se notó un comportamiento parecido, 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.

La Tabla II presenta los resultados obtenidos para muestras generadas a partir de una Distribución Normal $N(0, \Sigma)$, $\Sigma = \Sigma_1$ contaminada con $\varepsilon = 0.10$ y $H = N(0, 9\Sigma_1)$, se puede ver como el Estimador Clásico estima incorrectamente a los Valores Propios y Matriz de Covarianzas, teniendo un índice más de dos veces mayor al índice alcanzado por el “peor” estimador robusto cuando $k = 10$ y 20 . En general, cuando la muestra es generada a partir de una Población Normal $N(0, \Sigma)$, $\Sigma = \Sigma_1$ contaminada con $H = N(0, 9\Sigma_1)$, o $H = U\text{esf}(d=5)$ los estimadores sobreestiman el primer y segundo valor propio para luego estimar con error “pequeño” los valores propios restantes; en 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 $\Sigma = \Sigma_2$ los Estimadores sobreestiman el primer y segundo valor propio, sin embargo, al final generalmente subestiman el menor valor propio. De acuerdo a los resultados obtenidos, el tamaño muestral influye en la estimación, pues en todos los estimadores se cumple que al aumentar el mismo, se disminuye el Sesgo Promedio Total y la Desviación Total. El estimador T-Bicuadrático 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 p “grande” y en el presente estudio se considera $p = 6$. La Tabla III muestra que a pesar de que la mayoría de los Estimadores Robustos considerados han sido construidos bajo el supuesto de contaminación simétrica, al ser sometidos a contaminación asimétrica se comportan de manera similar a cuando la misma es simétrica.

TABLA I

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados para Población Normal $N(0, \Sigma)$, $\Sigma = \Sigma_1$ sin Contaminar

		k											
		5				10				20			
		Prom	Max	Desv	Indice	Prom	Max	Desv	Indice	Prom	Max	Desv	Indice
Vector de Medias	Clásico	0.023	1.486	0.693	0.794	0.009	0.945	0.438	0.463	0.008	0.697	0.307	0.338
	MHuber	0.022	1.459	0.636	0.706	0.012	0.986	0.438	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	3.989	0.433	3.719	0.870	1.674	0.240	2.200	0.642	0.994
	MHuber	0.484	2.552	0.857	1.297	0.294	2.122	0.672	1.029	0.166	1.254	0.440	0.630
	T-Bic	1.027	4.043	1.449	2.440	0.422	2.610	0.807	1.280	0.226	1.636	0.525	0.795
	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	7.680	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	0.880	1.524	0.275	1.955	0.596	0.942
Matriz de Covarianzas	Clásico	0.666	9.776	1.955	4.136	0.373	4.539	1.228	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	1.140	0.026	1.710	0.669	0.798
	T-Bic	0.288	5.945	2.200	2.811	0.069	3.164	1.246	1.493	0.033	2.122	0.702	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

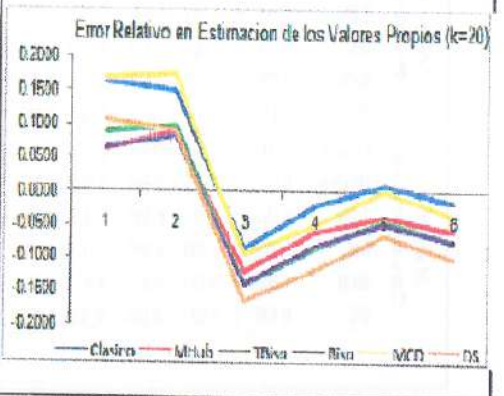
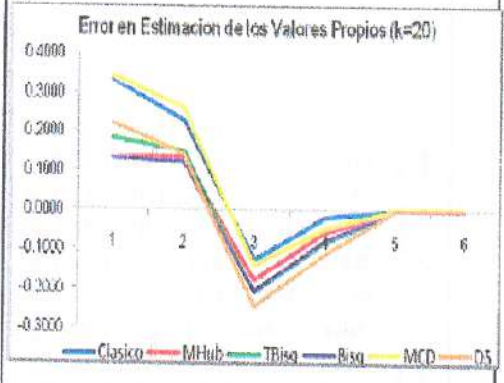


TABLA II

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados para Población Normal $N(0, \Sigma_1)$ Contaminada con $\epsilon = 0.10$ y $H = N(0.9, \Sigma_1)$

		k											
		5				10				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	0.808	0.380			
	T-Bic	0.015 ± 0.793	1.707	0.858	0.011 ± 0.455	1.078	0.515	0.012 ± 0.322	0.758	0.384			
	Bic	0.012 ± 0.651	1.424	0.680	0.008 ± 0.448	1.022	0.483	0.012 ± 0.318	0.770	0.387			
	MCD	0.016 ± 0.762	1.654	0.814	0.014 ± 0.507	1.187	0.589	0.011 ± 0.329	0.830	0.390			
	DS	0.008 ± 0.756	1.611	0.792	0.013 ± 0.505	1.119	0.548	0.014 ± 0.328	0.764	0.368			
Valores Propios	Clásico	4.043 ± 2.526	12.767	6.445	3.213 ± 1.742	13.288	6.085	2.619 ± 1.166	5.937	3.241			
	MHuber	1.057 ± 0.792	4.072	1.974	0.651 ± 0.576	2.274	1.200	0.465 ± 0.521	1.546	0.877			
	T-Bic	1.586 ± 1.446	8.285	3.759	0.885 ± 0.755	2.378	1.273	0.441 ± 0.571	1.788	0.938			
	Bic	0.819 ± 0.976	3.372	1.758	0.550 ± 0.893	2.600	1.282	0.390 ± 0.571	1.332	0.881			
	MCD	2.794 ± 1.807	9.425	4.875	1.142 ± 0.901	4.077	2.040	0.638 ± 0.624	2.249	1.170			
	DS	1.447 ± 1.442	10.816	4.568	0.727 ± 0.837	3.643	1.735	0.445 ± 0.583	1.735	0.821			
Matriz de Covarianzas	Clásico	2.666 ± 3.716	13.479	6.820	2.394 ± 2.519	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	0.887			
	T-Bic	0.630 ± 2.297	8.458	3.762	0.273 ± 1.219	3.272	1.588	0.228 ± 0.802	2.086	1.042			
	Bic	0.269 ± 1.553	4.529	2.117	0.200 ± 1.085	3.012	1.432	0.190 ± 0.763	1.900	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	1.316			
	DS	0.465 ± 2.095	9.738	4.089	0.214 ± 1.347	4.117	1.893	0.198 ± 0.862	2.205	1.086			

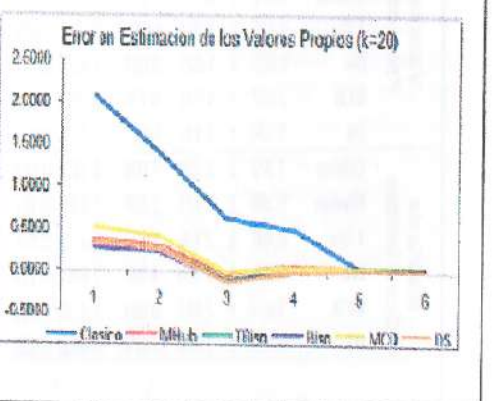
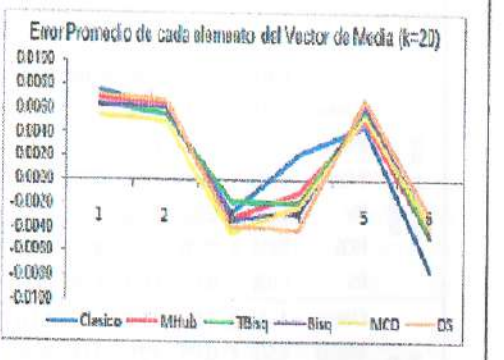


TABLA III

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados para Población Normal $N(0, \Sigma_1)$ Contaminada con $\epsilon = 0.10$ y $H=N(0.537\alpha_6, 9 \Sigma_1)$

		k											
		5				10				20			
		Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	0.036 ± 0.757	1.033	0.365	0.033 ± 0.534	1.243	0.613	0.054 ± 0.377	0.919	0.450			
	MHuber	0.018 ± 0.614	1.384	0.372	0.022 ± 0.432	0.964	0.473	0.018 ± 0.313	0.690	0.340			
	T-Bic	0.027 ± 0.760	1.714	0.394	0.016 ± 0.453	1.111	0.527	0.009 ± 0.317	0.594	0.337			
	Bic	0.012 ± 0.641	1.449	0.701	0.017 ± 0.433	1.044	0.499	0.011 ± 0.313	0.590	0.335			
	MCD	0.016 ± 0.740	1.525	0.794	0.019 ± 0.475	1.227	0.574	0.010 ± 0.329	0.599	0.343			
	DS	0.019 ± 0.766	1.689	0.311	0.019 ± 0.514	1.175	0.539	0.011 ± 0.307	0.605	0.345			
Valores Propios	Clásico	3.924 ± 2.694	13.366	6.361	3.097 ± 1.705	8.199	4.333	2.665 ± 1.184	6.454	3.438			
	MHuber	0.596 ± 0.610	2.477	1.294	0.332 ± 0.609	1.507	0.853	0.254 ± 0.492	1.167	0.630			
	T-Bic	1.341 ± 1.332	6.705	2.793	0.875 ± 0.739	2.808	1.341	0.456 ± 0.574	1.916	0.982			
	Bic	0.774 ± 1.003	3.152	1.943	0.652 ± 0.693	2.241	1.162	0.409 ± 0.559	1.716	0.895			
	MCD	2.426 ± 1.629	7.117	3.724	1.097 ± 0.919	3.486	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	1.339	0.452 ± 0.603	1.819	0.958			
Matriz de Covarianzas	Clásico	2.600 ± 3.690	13.959	6.716	2.322 ± 2.547	9.000	4.850	2.180 ± 1.834	6.691	3.568			
	MHuber	0.044 ± 1.275	3.384	1.588	0.032 ± 0.900	2.257	1.093	0.021 ± 0.666	1.550	0.739			
	T-Bic	0.352 ± 2.177	6.547	3.026	0.256 ± 1.198	3.113	1.522	0.244 ± 0.845	2.104	1.084			
	Bic	0.136 ± 1.539	4.379	2.018	0.190 ± 1.096	2.845	1.374	0.210 ± 0.790	2.051	1.020			
	MCD	1.460 ± 2.742	9.052	4.421	0.653 ± 1.473	3.940	2.022	0.444 ± 0.943	3.005	1.464			
	DS	0.138 ± 1.927	10.695	4.309	0.150 ± 1.280	3.300	1.597	0.202 ± 0.884	2.197	1.094			

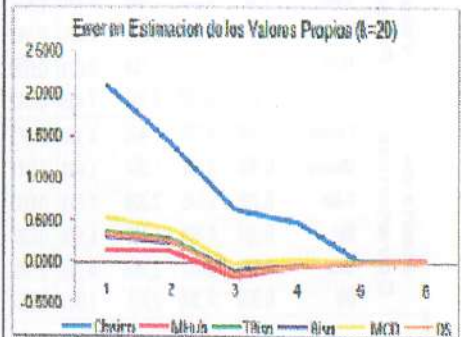
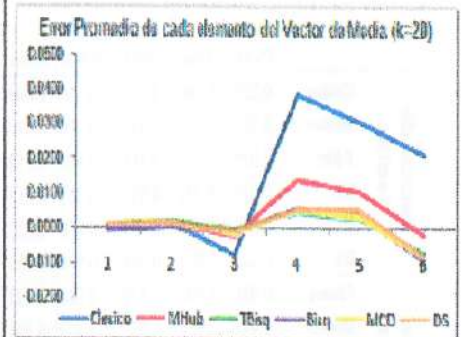


TABLA IV

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados para Población Normal $N(0, \Sigma_1)$ Contaminada por variable con $\epsilon = 0.10$ y $H=U(-5,5)$

		k											
		5				10				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	0.025 ± 0.606	1.320	0.651	0.014 ± 0.432	1.006	0.465	0.008 ± 0.302	0.576	0.329			
	T-Bic	0.020 ± 0.748	1.521	0.782	0.016 ± 0.464	1.095	0.525	0.009 ± 0.316	0.706	0.344			
	Bic	0.026 ± 0.629	1.370	0.875	0.014 ± 0.443	0.987	0.461	0.008 ± 0.308	0.885	0.394			
	MCD	0.019 ± 0.733	1.633	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 ± 0.498	1.159	0.557	0.013 ± 0.325	0.734	0.357			
Valores Propios	Clásico	1.742 ± 1.165	4.739	2.568	1.153 ± 0.820	3.408	1.794	0.705 ± 0.502	2.053	1.117			
	MHuber	0.575 ± 0.779	2.221	1.182	0.394 ± 0.538	1.736	0.869	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	0.818			
	Bic	0.800 ± 1.023	3.889	1.927	0.504 ± 0.674	2.731	1.303	0.294 ± 0.529	1.595	0.799			
	MCD	2.607 ± 1.694	8.126	4.109	1.097 ± 0.905	3.623	1.872	0.454 ± 0.930	1.932	0.975			
	DS	1.135 ± 1.135	6.041	2.771	0.665 ± 0.783	2.598	1.348	0.351 ± 0.529	1.803	0.811			
Matriz de Covarianzas	Clásico	1.372 ± 2.289	7.008	3.580	0.973 ± 1.579	4.792	2.448	0.612 ± 0.891	2.753	1.452			
	MHuber	0.269 ± 1.173	2.668	1.437	0.191 ± 0.868	2.099	1.049	0.101 ± 0.624	1.546	0.757			
	T-Bic	0.481 ± 2.106	7.273	3.313	0.194 ± 1.167	3.352	1.571	0.092 ± 0.774	2.000	0.952			
	Bic	0.275 ± 1.593	4.397	2.080	0.133 ± 1.038	3.018	1.398	0.051 ± 0.727	1.925	0.911			
	MCD	1.629 ± 2.817	8.813	4.418	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	3.476	1.818	0.054 ± 0.860	2.311	1.075			

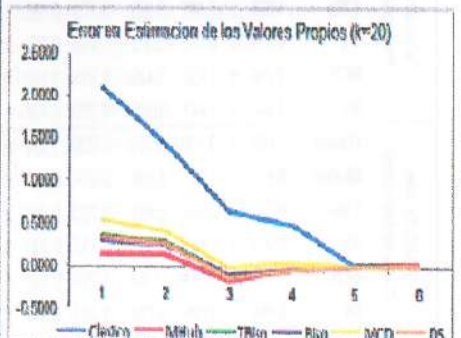
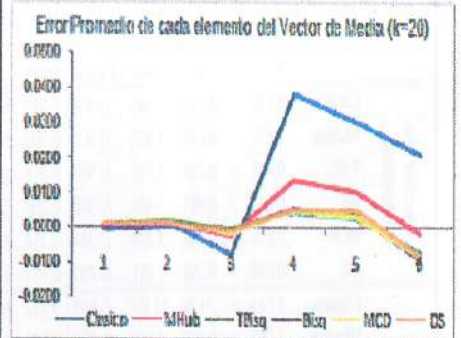
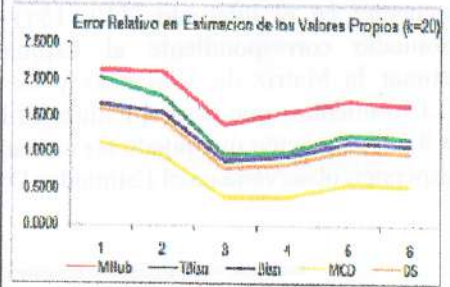
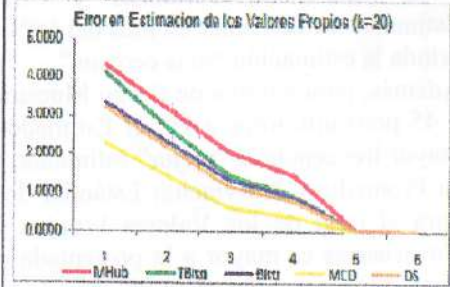


TABLA V

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados para Población Cauchy $CAU(0, \Sigma)$, $\Sigma = \Sigma_1$ sin Contaminar

	5				10				20			
	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice	Prom	Desv	Max	Indice
Vector de Medias	Clásico	77.378 ± 4.978	--	--	1.913 ± 4.855	716.581	241.053		0.984 ± 4.843	391.010	132.279	
	MHuber	0.047 ± 0.868	2.361	1.132	0.023 ± 0.620	1.428	0.030		0.011 ± 0.429	0.967	0.469	
	T.Bic	0.041 ± 0.952	3.014	1.338	0.022 ± 0.663	1.613	0.768		0.026 ± 0.479	1.048	0.518	
	Bic	0.045 ± 0.840	2.218	1.035	0.018 ± 0.590	1.367	0.859		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	1.000	0.512	
	DS	0.034 ± 0.835	1.940	0.937	0.016 ± 0.597	1.514	0.709		0.016 ± 0.426	0.960	0.484	
Valores Propios	Clásico	--	--	--	--	--	--		--	--	--	
	MHuber	8.641 ± 6.787	51.030	22.229	6.753 ± 4.074	26.143	12.323		5.906 ± 2.740	19.202	9.203	
	T.Bic	10.176 ± 8.344	93.205	37.409	6.729 ± 5.045	25.051	12.275		5.225 ± 2.866	23.001	10.371	
	Bic	6.918 ± 6.212	40.263	17.793	5.017 ± 3.549	18.338	9.001		4.390 ± 2.411	14.457	7.088	
	MCD	9.065 ± 7.366	42.272	18.575	4.008 ± 3.068	13.959	7.012		2.746 ± 1.789	9.368	4.631	
	DS	11.871 ± 10.976	190.074	70.940	4.785 ± 3.969	36.038	15.057		4.115 ± 2.392	12.980	6.492	
Matriz de Covarianzas	Clásico	--	--	--	--	--	--		--	--	--	
	MHuber	6.998 ± 6.972	46.980	20.317	5.665 ± 4.496	23.372	11.178		6.121 ± 3.079	16.931	8.377	
	T.Bic	6.705 ± 8.220	79.492	31.473	5.107 ± 5.390	26.590	13.029		4.298 ± 3.555	19.172	9.009	
	Bic	4.797 ± 6.068	42.679	17.815	3.970 ± 1.992	19.123	9.029		3.698 ± 2.841	13.856	6.798	
	MCD	6.130 ± 7.542	45.319	18.883	2.868 ± 3.637	15.785	7.420		2.144 ± 2.295	9.708	4.710	
	DS	8.329 ± 9.669	170.344	62.744	3.630 ± 4.011	30.214	12.619		3.413 ± 2.799	13.110	6.441	



Así también, los estimadores han sido construidos siguiendo el entorno de contaminación (2.1), sin embargo, al contaminar por variable, según lo explicado en la sección 3.1, los Estimadores brindan "buenas" estimaciones, véase la Tabla IV; no obstante, en el presente estudio no se analiza las consecuencias que puede tener la 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 V presenta el caso cuando se generan muestras a partir de una Cauchy sin contaminar. Además, en estos casos, el algoritmo utilizado para el Estimador M de Huber no siempre converge a una solución. Los valores propios son siempre sobreestimados; en la parte derecha de la Tabla V se muestra los errores absolutos y relativos de la estimación de los valores propios para $k=20$.

A las estimaciones brindadas por el estimador DS en varias ocasiones les 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 implica 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

En esta parte del estudio se trata de establecer una "regla empírica" en la cual se utilice el tamaño de la Muestra, el Sesgo y la Curtosis Muestral como elementos de decisión sobre que estimador utilizar en ese caso.

A cada muestra generada se le ha calculado Sesgo y Curtosis, además se determinan la distancia Euclidiana entre el verdadero valor del Vector de Medias, la Matriz de Covarianzas y los Valores Propios asociados a la misma con la estimación obtenida por cada Método; a partir de esto se establece el estimador "más cercano" en cada muestra.

En las Tablas VI y VII se presenta un resumen de los resultados obtenidos para muestra tamaño 30 y 60 respectivamente; en las mencionadas tablas se puede observar los tres estimadores que con mayor frecuencia presentan la menor distancia entre el valor del parámetro y la estimación; para cada estimador se presenta la distancia Promedio y su Desviación Estándar, además de la Frecuencia Relativa de ser el estimador "más cercano" al valor real.

Así, por ejemplo, con tamaño de muestra igual a 30, cuando el Sesgo es menor a 45 y la Curtosis se encuentra en el intervalo [20, 65] el 25,3% de las

ocasiones 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 superiores a 45 pero inferiores 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, 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 causada por la alta dispersión observada en el Estimador DS.

De acuerdo a la Tabla VII, 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 alcanza la proporción 0.688 de ser el estimador "más cercano" a los Valores Propios; es 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 sesgos mayores a 90, el Estimador MCD.

De acuerdo a los resultados obtenidos, para $n=120$, cuando la Curtosis es menor a 230 el Estimador M de Huber o el Estimador Bicuadrático son los que con mayor frecuencia se constituyen en los "mejores" estimadores, mientras que para Curtosis mayor a 230 el mejor estimador es con mayor frecuencia el MCD.

TABLA VI
Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados en función del Sesgo y la Curtosis para tamaño muestral 30

Sesgo Muestral			Curtosis Muestral					
			[20, 65]		[65, 110]		[110, 155]	
			Distancia	Frecuencia	Distancia	Frecuencia	Distancia	Frecuencia
[0, 45]	Localización	1 Clas	0.503 ± 0.174	0.258	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 ± 0.161	0.210	DS	0.537 ± 0.202	0.174	
	Valores Propios	1 MHub	1.154 ± 1.155	0.535	MHub	3.598 ± 4.828	0.369	
		2 Bis	1.212 ± 1.181	0.328	Bis	3.510 ± 4.879	0.308	
		3 DS	1.919 ± 2.271	0.068	DS	5.185 ± 10.074	0.164	
	Matriz de Covarianzas	1 MHub	1.229 ± 1.020	0.648	MHub	3.414 ± 4.317	0.419	
		2 Bis	1.288 ± 1.030	0.338	Bis	3.287 ± 4.354	0.299	
		3 DS	1.871 ± 1.866	0.058	DS	4.788 ± 8.726	0.168	
[45, 90]	Localización	1 DS	0.598 ± 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	0.948 ± 0.254	0.178	
	Valores Propios	1 DS	10.648 ± 13.286	0.321	DS	12.381 ± 16.178	0.366	
		2 Bis	7.640 ± 8.283	0.278	Bis	9.384 ± 8.217	0.229	
		3 MHub	8.372 ± 8.229	0.191	MCD	9.807 ± 5.930	0.195	
	Matriz de Covarianzas	1 DS	8.538 ± 11.628	0.322	DS	11.075 ± 14.074	0.376	
		2 Bis	8.899 ± 5.442	0.278	Bis	8.987 ± 5.385	0.220	
		3 MHub	7.691 ± 5.541	0.184	MCD	8.540 ± 5.075	0.204	
[90, 135]	Localización	1 DS	0.618 ± 0.238	0.334	DS	0.827 ± 0.218	0.288	
		2 MHub	0.688 ± 0.265	0.251	MHub	0.881 ± 0.248	0.259	
		3 Bis	0.654 ± 0.247	0.181	Bis	0.648 ± 0.240	0.222	
	Valores Propios	1 DS	11.190 ± 12.920	0.422	DS	13.489 ± 10.911	0.333	
		2 Bis	8.857 ± 8.306	0.238	MCD	10.478 ± 7.010	0.241	
		3 MCD	8.368 ± 5.647	0.187	MHub	11.112 ± 8.280	0.204	
	Matriz de Covarianzas	1 DS	10.044 ± 11.236	0.428	DS	12.128 ± 8.652	0.352	
		2 Bis	8.102 ± 5.480	0.234	MCD	9.352 ± 6.914	0.259	
		3 MCD	8.367 ± 5.090	0.180	Bis	9.249 ± 5.689	0.204	
[135, 180]	Localización	1 DS	0.618 ± 0.262	0.371				
		2 Bis	0.638 ± 0.257	0.257				
		3 MHub	0.666 ± 0.256	0.149				
	Valores Propios	1 Bis	10.580 ± 7.907	0.314				
		2 DS	13.448 ± 13.000	0.257				
		3 MCD	10.188 ± 8.006	0.228				
	Matriz de Covarianzas	1 Bis	8.532 ± 8.750	0.371				
		2 DS	12.125 ± 11.515	0.267				
		3 MCD	8.142 ± 5.181	0.228				

TABLA VII

Estimadores robustos para el vector de medias y la matriz de varianzas y covarianzas de vectores aleatorios multivariados
Resultados en función del Sesgo y la Curtosis para tamaño muestral 60

				Curtosis Muestral						
				[40, 130]			[130, 220]			
				Distancia	Frecuencia		Distancia	Frecuencia		
[0, 90]	Localización	1 Clas	0.383 ± 0.243	0.220		MHub	0.402 ± 0.158	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	0.894 ± 1.569	0.484		MCD	4.004 ± 2.707	0.579		
		2 Bis	0.885 ± 1.407	0.246		DS	5.346 ± 4.914	0.155		
		3 DS	1.144 ± 1.631	0.107		MHub	6.078 ± 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	0.158		
		3 DS	1.201 ± 1.432	0.083		DS	4.827 ± 4.347	0.152		
	[90, 180]	Localización	1 MHub	0.435 ± 0.160	0.288		MHub	0.423 ± 0.148	0.278	
			2 DS	0.420 ± 0.151	0.258		DS	0.414 ± 0.148	0.259	
			3 Bis	0.423 ± 0.155	0.198		Bis	0.418 ± 0.153	0.212	
Valores Propios		1 MCD	4.593 ± 2.542	0.688		MCD	4.661 ± 2.488	0.753		
		2 DS	6.056 ± 4.611	0.181		DS	8.185 ± 4.328	0.158		
		3 Bis	6.183 ± 3.609	0.090		Bis	6.357 ± 3.430	0.069		
Matriz de Covarianzas		1 MCD	4.225 ± 2.212	0.708		MCD	4.301 ± 2.193	0.779		
		2 DS	5.567 ± 4.061	0.179		DS	5.707 ± 3.872	0.149		
		3 Bis	5.705 ± 3.201	0.078		Bis	5.889 ± 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	0.428 ± 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	0.780		
		2 DS	6.514 ± 4.988	0.183		DS	6.840 ± 4.262	0.113		
		3 Bis	6.647 ± 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 ± 3.746	0.097		
		3 Bis	6.184 ± 3.431	0.084		Bis	6.657 ± 3.458	0.018		
	[270, 360]	Localización	1 DS	0.436 ± 0.150	0.318					
			2 MHub	0.456 ± 0.152	0.278					
			3 Bis	0.447 ± 0.160	0.250					
Valores Propios		1 MCD	5.011 ± 2.528	0.722						
		2 DS	6.345 ± 3.730	0.208						
		3 Bis	6.539 ± 3.277	0.056						
Matriz de Covarianzas		1 MCD	4.707 ± 2.329	0.778						
		2 DS	5.994 ± 3.430	0.181						
		3 Bis	6.201 ± 3.061	0.028						

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 $k=20$ el Estimador MCD es el que "mejor" estima la Matriz de Covarianzas y los Valores propios, todo ello en base al índice planteado en 3.2.

Para $\varepsilon=0.30$ los estimadores Robustos estudiados, se comportan de forma parecida a cuando $\varepsilon=0.05$ y $\varepsilon=0.10$, se concluye entonces que el Punto de Ruptura, para $p=6$ variables, es mayor a 0.3

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

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

- [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 Willey & Sons, New York-USA.
- [9]. MARONNA, R. (1976). "Robust M-Estimators of Multivariate Location and Scatter", *The Annals of Statistics*; 4, 51-67.
- [10]. MONTAÑO, N. (2009) "Estimadores Robustos para el Vector de Medias y la Matriz de Varianzas y Covarianzas de Vectores Aleatorios Multivariados", Tesis de Grado.
- [11]. R DEVELOPMENT CORE TEAM (2008). "R: A language and environment for statistical computing. R Foundation for Statistical Computing", Vienna, Austria.
- [12]. ROCKE D. (1996) "Robustness properties of multivariate location and shape in high dimension", *The Annals of Statistics*; 24, 1327-1345.
- [13]. RENCHER, A. (1998). "Multivariate Statistical Inference and Applications", John Willey & Sons, New York-USA.
- [14]. SEBER, G. (1984). "Multivariate observation", John Willey & Sons, New York-USA.
- [15]. 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.
- [16]. YOHAI, V. (2004). "Notas de probabilidad y estadística", [en línea] Universidad de Buenos Aires, Buenos Aires - Argentina.
- [17]. ZAMAR, R. (1994) "Estimación Robusta", *Estadística Española*; 36, 327-387.
- [18]. 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.