

Revista Colombiana de Estadística  
Nº 7 - 1983

## ESTIMADORES DE ESCALA: ESTUDIO COMPARATIVO.

*Jorge Martínez C.*  
Profesor Asociado  
Universidad Nacional

### 1. Introducción.

El desarrollo y estudio de las propiedades de nuevos estimadores para los parámetros de una distribución simétrica dada es un campo de la estadística al cual se le ha dedicado considerable atención en los últimos años. Gran parte de este esfuerzo se ha encaminado a la consideración del problema de localización; esto es, al diseño de nuevas medidas de "tendencia central" representadas por estimadores del parámetro de localización de una distribución o de los parámetros de una ecuación de regresión. La estimación del parámetro de escala, problema de escala, generalmente se ha considerado como una herramienta auxiliar del problema de localización dada la necesidad de estandarizar

los datos. Sin embargo, el problema de escala tiene interés por sí mismo. Por ejemplo, la eficiencia de un intervalo de confianza depende de su longitud y ésta a su vez del éxito que tengamos al estimar la escala de la distribución.

El objeto de este artículo es presentar varios estimadores de escala desarrollados o reformados recientemente y mediante su estudio comparativo, en el cual se incluyen los estimadores de uso tradicional, determinar las estadísticas que ofrecen mayores ventajas a sus usuarios. Adicionalmente, se describirán de una manera muy somera algunas de las aplicaciones en las cuales el uso de estos estimadores permite obtener procedimientos estadísticos de mayor potencia y eficiencia.

Resultados preliminares a los que se muestran en este trabajo aparecen publicados en Hoaglin et.al. (1982), para el cual el autor de este artículo colaboró con el aspecto correspondiente a la simulación. De otra parte, estos resultados confirman y amplían los obtenidos por Lax (1975) al incluir en la comparación nuevos estimadores.

## 2. Conceptos preliminares.

### 2.1. Parámetros de localización y de escala.

Consideremos una familia de distribuciones con función de densidad de probabilidad

$$f(x) = \frac{1}{\sigma} f_0\left(\frac{x-\theta}{\sigma}\right),$$

donde  $f_0$  es la "densidad estándar" de la familia,  $\theta$  un número real cualquiera y  $\sigma$  un número real positivo. En situaciones como esta se dice que  $\theta$  es el parámetro de localización y  $\sigma$  el parámetro de escala.

Si  $X$  es una variable aleatoria con parámetros de localización  $\theta$  y de escala  $\sigma$ , a  $Y = \frac{X-\theta}{\sigma}$  le corresponde la "densidad estándar", esto es, los elementos  $X$  de la familia se encuentran mediante una ampliación o reducción (cambio de escala) y una translación (cambio de localización) de la variable estándar de la familia. Vale la pena anotar que si  $\sigma$  es un parámetro de escala, dado cualquier número positivo  $k$ ,  $k\sigma$  es también un parámetro de escala.

A manera de ilustración se presenta la familia de distribuciones uniforme con densidad

$$f(x) = \frac{1}{2a}, \quad -a < x < a, \quad a \neq 0.$$

Si se define  $f_0(x) = \frac{1}{2}$ ,  $-1 < x < 1$ , como la densidad estándar de la familia, se obtiene que

$$f(x) = \frac{1}{a} f_0\left(\frac{x}{a}\right), \quad -a < x < a$$

o sea que  $a$  es un parámetro de escala.

## 2.2. Estimadores de escala.

Un estimador de escala es una estadística no negativa  $W$  tal que

$$W(a+bx_1, a+bx_2, \dots, a+bx_n) = |b| W(x_1, x_2, \dots, x_n)$$

Como puede observarse, un estimador de escala no responde a cambios en la localización de los datos.

Existen algunas propiedades que hacen un estimador de escala más atractivo que otro. Se mencionan a continuación algunas de las propiedades deseables para  $W$ :

- Proporcionalidad a la dispersión de los valores  $x_1, x_2, \dots, x_n$  obtenidos en la muestra.
- Resistencia a un número moderado de observaciones extremas.
- Alta eficiencia cuando se le compara con otros estimadores de escala en distribuciones que contemplen el modelo tradicional (normal) y modelos que posiblemente puedan representar de una manera satisfactoria los datos objeto del análisis estadístico.

## 3. Algunos estimadores de escala.

Los estimadores de escala que se presentan en

este artículo se clasifican básicamente en tres categorías:

- Estimadores basados en estadísticas de orden.
- Estimadores correspondientes a la varianza de un estimador  $M$  de localización.
- Estimadores  $M$  de escala.

En la parte restante de la presente sección se describe de una manera general cada uno de los grupos de estimadores mencionados anteriormente y se discuten las propiedades particulares de los estimadores de mayor interés para este estudio.

### 3.1. Estimadores basados en las estadísticas de orden.

De acuerdo a la forma en que estos estimadores utilizan las estadísticas de orden, Lax (1975) propuso considerarlos en dos grupos. En el primero se colocan las estadísticas que usan la distancia de cada dato a un estimador de localización dado. Como ejemplos se pueden mencionar en primera instancia la muy conocida desviación estándar muestral  $S$ , obtenida a partir de

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 ;$$

la desviación absoluta ( $DA$ ) definida por

$$DA = \frac{1}{n} \sum_{i=1}^n |X_i - \tilde{X}| ,$$

donde  $\tilde{X}$  es la mediana de la muestra; y, la desviación absoluta mediana

$$DAM = \underset{i}{Mediana}\{|X_i - \tilde{X}|\}$$

En el segundo grupo se encuentran los estimadores que utilizan las diferencias de estadísticas de orden simétricas. Tal es el caso de la estadística  $S^*$  discutida por Hettsmansperger y McKean (1978) definida de la siguiente manera:

$$S^* = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} (n-2i+1) (X_{(n-i+1)} - X_{(i)})$$

donde  $\lfloor \frac{n}{2} \rfloor$  es el mayor entero menor o igual que  $\frac{n}{2}$  y  $X_{(i)}$  es la  $i$ -ésima estadística de orden. Otro estimador correspondiente a este grupo y de gran uso en el análisis exploratorio de datos es la llamada distancia entre cuartos (DC). Es una versión simplificada de la distancia intercuartílica. Se define como

$$DC = X_{(n-f+1)} - X_{(f)}$$

con  $f = \frac{\lfloor \frac{n+1}{2} \rfloor + 1}{2}$ . El valor de  $f$  puede ser  $k$  ó  $k + \frac{1}{2}$ , siendo  $k$  un entero. En caso de que  $f$  sea igual a  $k + \frac{1}{2}$ , el promedio de  $k$  y  $(k+1)$ ,  $X_{(f)}$  corresponde al

promedio de  $X_{(k)}$  y  $X_{(k+1)}$ .

Como se verá más adelante, el estimador  $S$  presenta algunas dificultades frente a datos no distribuidos normalmente. Por este motivo se ha sugerido modificar  $S^2$  teniendo en cuenta el promedio de las desviaciones cuadráticas del  $100(1-2\beta)\%$  de las estadísticas de orden centrales, respecto de la media de las observaciones que quedan al descartar  $100\alpha\%$  a cada lado de la muestra. Este estimador es denotado por  $S_2^2$  y se le conoce como la varianza muestral recortada en  $100\beta\%$  respecto a la media recortada en  $100\alpha\%$ . En este artículo se considerará el caso en el cual  $\alpha = .20$  y  $\beta = .10$ .

### 3.2. Estimadores de escala basados en la varianza de estimadores $M$ de localización.

Dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  tomada de una distribución dada con parámetro de localización  $\theta$ , se dice que  $\hat{\theta}_n$  es un estimador  $M$  de localización si éste es la solución de la ecuación

$$\sum_{i=1}^n \psi(X_i - \hat{\theta}_n) = 0$$

para alguna función  $\psi$  dada. Su nombre se debe a que los estimadores de máxima verosimilitud son un caso especial de los estimadores  $M$ ; la función  $\psi(x)$  es la derivada del logaritmo natural de la función de densidad, esto es,  $\psi(x) = f'(x)/f(x)$ .

Huber (1981) muestra que bajo ciertas condiciones, la distribución asintótica de

$$\sqrt{n}(\hat{\theta}_n - \theta)$$

es normal con media cero y varianza

$$A(\theta, F) = \frac{\int \psi^2(x-\theta) dF}{\left[ \int \psi'(x-\theta) dF \right]^2} \quad (1)$$

siendo  $F$  la función de distribución considerada en el modelo estadístico.

La versión muestral  $A_n$  de esta varianza se encuentra a través de la distribución empírica  $F_n$ . Como es bien conocido,  $F_n$  es una función escalonada con saltos de magnitud  $k/n$  en los valores  $x_i$  los cuales corresponden a datos que se repitan  $k$  veces. Por lo tanto  $dF$  es  $k/n$  en los valores  $x_i$  y cero en cualquier otro punto, de manera que las integrales de (1) se convierten en sumatorias. De esta manera se obtiene que

$$A_n = \frac{n \sum_{i=1}^n \psi^2(x_i - \hat{\theta})}{\left[ \sum_{i=1}^n \psi'(x_i - \hat{\theta}) \right]^2} \quad (2)$$

Dado que los valores  $x_i$  deben estandarizarse con algún estimador de escala preliminar  $w$ , resulta que

$$A_n = \frac{n(kw)^2 \sum_{i=1}^n \psi^2(Z_i)}{\left[ \sum_{i=1}^n \psi'(Z_i) \right]^2} \quad (3)$$



siendo  $Z_i = \frac{X_i - \hat{\theta}}{k \cdot W}$  y  $k$  una constante apropiada para cada caso. El término  $(kW)^2$  del numerador de la expresión (3) aparece como consecuencia del reemplazo de  $(x - \theta)$  por su valor estandarizado en (1).

La forma particular de cada estimador depende de la función  $\psi$  que se utilice. En este artículo se presentan solo dos de estas funciones. Una descripción de las diferentes alternativas propuestas puede encontrarse en Andrews et. al (1972) y Huber (1981).

En primera instancia se considera la función bicuadrada sugerida por Beaton y Tukey (1964).

$$\psi(x) = \begin{cases} x^2(1-x^2) & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

Calculando esta función y su derivada en  $Z_i = \frac{X_i - \bar{X}}{9(DAM)}$ , y haciendo los reemplazos correspondientes en (3) se obtiene el estimador "S bicuadrado" el cual corresponde a la raíz cuadrada de

$$S_b^2 = \frac{n \sum_{|Z_i| < 1} (X_i - \bar{X})^2 (1 - Z_i^2)^4}{\left[ \sum_{|Z_i| < 1} (1 - Z_i^2) (1 - 5Z_i^2) \right]^2}$$

El valor de  $k = 9$  es explicable debido a que  $E(DAM) \approx \frac{2}{3}\sigma$  en una distribución normal con desviación estándar  $\sigma$ , esto hace que  $E(9DAM) \approx 6\sigma$ .

La otra función tenida en cuenta en este estudio, es la propuesta por Andrews (1972)

$$\psi(x) = \begin{cases} \text{sen}(x/\pi) & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

Con esta función y un valor de  $k$  igual a 2.1 se origina el estimador  $S_w$  cuya fórmula puede deducirse fácilmente de la expresión (3).

### 3.3. Estimadores $M$ de escala.

Una estadística  $w$  es un estimador  $M$  de escala si se cumple que

$$\sum_{i=1}^n \chi\left(\frac{y_i}{w}\right) = 0 \quad (4)$$

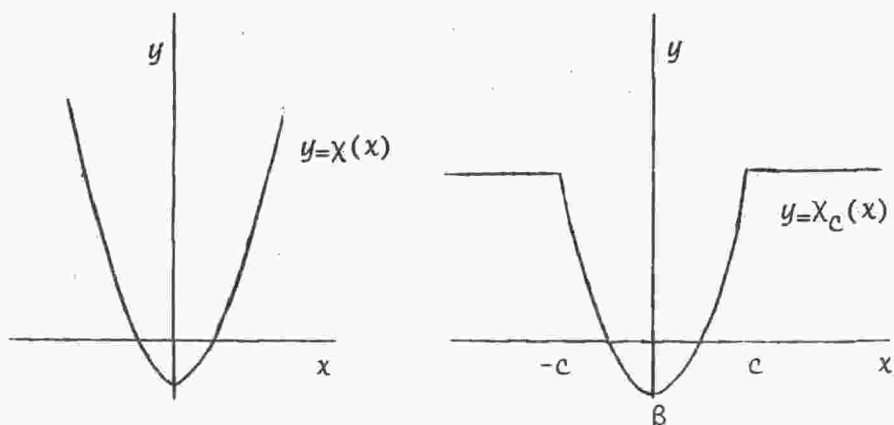
para alguna función  $\chi$ . Los valores  $y_i$  son las desviaciones de los datos  $x_i$  respecto de un estimativo de localización, o sea que el estimativo de localización de los  $y_i$  es cero.

Es fácil ver que si  $\chi(x) = x^2 - 1$  y  $y_i = x_i - \bar{X}$ , entonces  $w^2$  es la varianza muestral  $S^2$ .

Con el ánimo de limitar el efecto de observaciones que se desvían considerablemente del resto de los datos, H u b e r (1981) propuso la función:

$$\chi_c(x) = \begin{cases} x^2 - \beta & |x| \leq c \\ c^2 - \beta & |x| > c \end{cases}$$

con  $c$  una constante dada y  $\beta$  determinada de tal manera que el estimador resultante sea eficiente en algún modelo específico, generalmente el normal. En la Gráfica 1 se puede analizar la influencia que cada una de las funciones mencionadas anteriormente le permite a una observación que se aleja considerablemente de las demás (mayor que  $c$ ). Vale la pena anotar que si  $\beta$  toma el valor de 1 y  $c$  es suficientemente grande el estimador de escala obtenido es  $S$  el cual, como es bien conocido, es eficiente bajo la hipótesis de normalidad. Al acotar los valores de  $x$  es necesario reajustar el valor de  $\beta$  para mantener la eficiencia. En este trabajo se considerará el caso en el cual  $\beta = 1$  y  $c = 1.4$  para ilustrar la pérdida de eficiencia en el modelo normal simultáneamente con el buen desempeño en otros modelos. El estimador obtenido usando estas constantes es denotado por  $S_1$ .



Gráfica 1

Ejemplos de funciones  $X$  utilizadas en el desarrollo de estimadores  $M$  de escalas extremas.

#### 4. Comparación de estimadores.

El desempeño de cada uno de los estimadores presentados en la sección anterior es analizado en términos de las eficiencias obtenidas bajo algunas distribuciones y de la resistencia a observaciones externas.

Con el fin de estudiar la eficiencia de los estimadores bajo diferentes formas distribucionales se han escogido las distribuciones usadas en el extenso estudio comparativo de estimadores de localización presentado por Andrews et al. (1972). Estas tres distribuciones contemplan el modelo tra-

dicional de normalidad y algunas de sus desviaciones más comunes. Específicamente las distribuciones consideradas son:

- $N$ : la distribución normal estándar.
- $S$ : el cociente de una variable con distribución normal estándar y otra distribuida uniformemente en el intervalo  $(0,1)$  e independiente de la primera. Esta distribución tiene colas considerablemente más largas que las de la distribución normal, las cuales dan margen a observaciones alejadas del centro de los datos sin llegar a ser consideradas como observaciones extremas.
- $M$ : Es una distribución generada con  $(n-m)$  observaciones de una distribución normal estándar y  $m$  de una distribución normal con media 0 y varianza  $\sigma^2$ . En el presente trabajo se consideran solo muestras de tamaño 20 de las cuales 19 observaciones provienen de la normal estándar y 1 de una normal con  $\sigma^2 = 100$ .

Una vez definidas las distribuciones en las cuales se hace la comparación de los estimadores en términos de su eficiencia, se debe decidir de qué manera se realiza esta comparación. Como es bien conocido, la eficiencia relativa de dos estimadores de localización se calcula con el cociente de las varianzas de los dos estimadores. En el caso de estimadores de escala las varianzas deben

ajustarse de alguna manera con el valor esperado del estimador, ya que dos medidas de escala no necesariamente estiman el mismo parámetro. Por ejemplo  $E(S) \approx .98\sigma$  y  $E(DC) \approx 1.345\sigma$  en una distribución normal. Para evitar el problema de este ajuste Lax D.A.(1975) ha propuesto usar la varianza del logaritmo natural del estimador  $W$ . Una comparación basada en esta varianza elimina el efecto del promedio del estimador ya que

$$V[\ln(KW)] = V[\ln K + \ln W] = V[\ln W].$$

Tabla 1

$V(\ln W)$  para muestras de tamaño 20

Estimador	D i s t r i b u c i ó n		
	N	R	M
S	.0277	1.2756	.2560
DA	.0312	.7340	.0813
DAM	.0741	.1081	.0734
DC	.0650	.1121	.0644
S*	.0283	.8453	.1098
S <sub>b</sub>	.0323	.1066	.0373
S <sub>w</sub>	.0356	.1062	.0392
S <sub>1</sub>	.0560	.0962	.0556
S <sub>2</sub>	.0340	.3277	.0337
V <sub>F</sub> *	.0277	.0962	.0337

La Tabla 1 presenta los valores de  $V[\ln W]$  para los

estimadores y las distribuciones descritas anteriormente. Cada valor de la tabla ha sido obtenido en base a 5000 muestras simuladas de tamaño 20. En esta tabla también se presentan los valores  $V_F^*$  correspondientes al estimador cuya varianza es mínima con la distribución  $F$ .

La eficiencia  $E$  de cada estimador  $W$  en una distribución  $F$  se define como

$$E_F(W) = \frac{V_F^*}{V[\ln W]} \times 100.$$

El desempeño de cada estimador  $W$  se evalúa con el mínimo de las tres eficiencias  $E_N(W)$ ,  $E_S(W)$  y  $E_M(W)$ . Este valor se denomina la trieficiencia y representa la eficiencia garantizada que se tiene al utilizar este estimador en caso de que los datos sean generados por alguna de las distribuciones consideradas. Las eficiencias y trieficiencia de cada estimador aparecen en la Tabla 2.

Se considera ahora la comparación de los estimadores en términos de su resistencia a observaciones extremas. Con este fin se calcula el valor de  $W_x = W(x_1, x_2, \dots, x_{19}, x)$  con  $x_1, x_2, \dots, x_{19}$  igual a los valores esperados de las estadísticas de orden de una muestra de tamaño 19 en una distribución normal y  $x$  una observación variable.

Al hacer la gráfica de  $x$  vs  $W_x$  se puede ana-

Tabla 2

Eficiencias y trieficiencias de cada estimador en muestras de tamaño 20

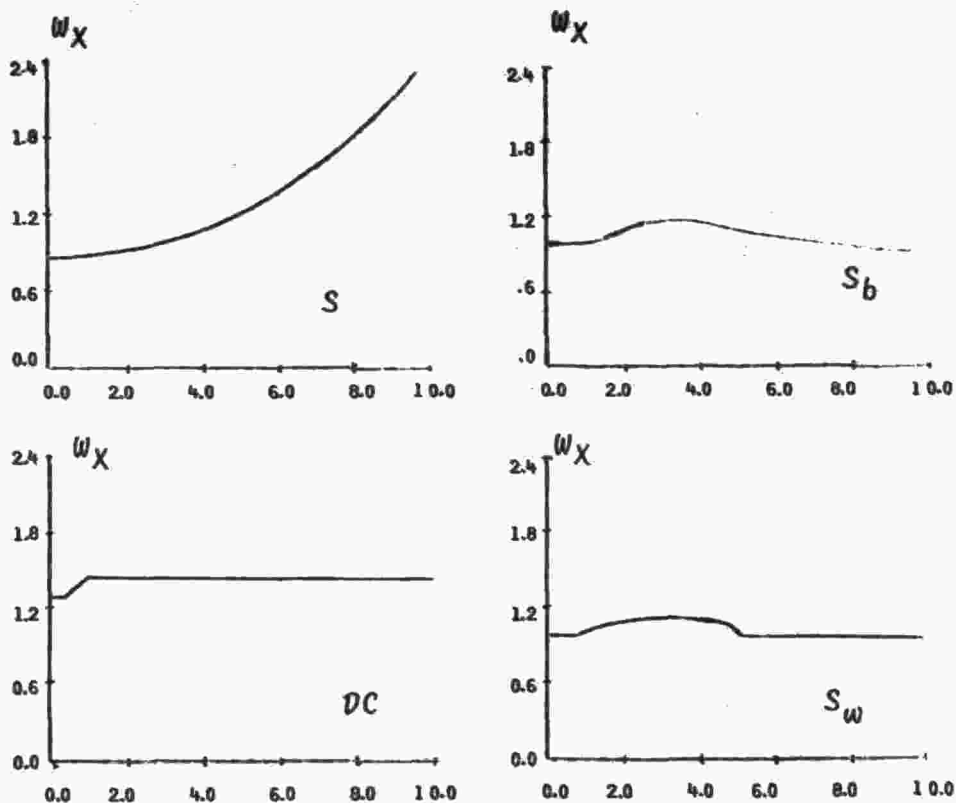
<i>Estimador</i>	D i s t r i b u c i ó n			
	<i>N</i>	<i>R</i>	<i>M</i>	<i>Trieficiencia</i>
<i>S</i>	100.0	7.6	13.2	7.6
<i>DA</i>	88.8	13.1	41.4	13.1
<i>DAM</i>	37.4	89.0	46.0	37.4
<i>DC</i>	42.6	85.8	52.3	42.6
<i>S*</i>	97.9	11.4	30.7	11.4
<i>S<sub>b</sub></i>	85.8	90.2	90.3	85.8
<i>S<sub>w</sub></i>	77.8	90.6	86.0	77.8
<i>S<sub>1</sub></i>	49.5	100.0	60.6	49.5
<i>S<sub>2</sub></i>	81.4	29.4	100.0	29.4

lizar la sensibilidad de cada estimador  $w$  con respecto a la observación variable  $x$ . Las curvas de sensibilidad para algunos de los estimadores se encuentran en la Gráfica 2 p.33.

## 5. Resultados.

La Tabla 2 muestra varios resultados de interés. En primera instancia se puede observar un desempeño "óptimo" de la desviación estándar muestral  $S$  con datos distribuidos normalmente frente a las ba-





Gráfica 2

Curvas de sensibilidad para algunos estimadores.

jas eficiencias obtenidas cuando la distribución que genera los datos da cabida a observaciones que bajo el modelo normal serían rechazadas por ser valores "muy extremos". Este hecho da una voz de alerta al uso generalizado de  $S$  con cualquier conjunto de datos aparentemente simétricos (e inclusive no simétricos!). Las modificaciones que se le hacen a  $S$  para obtener los estimado-

res  $S_1$  y  $S_2$  dan buenos resultados con distribuciones de colas alargadas.  $S_1$  resulta ser eficiente cuando la distribución acepta varias observaciones "extremas", mientras que  $S_2$  lo es cuando en la muestra se encuentra solo una observación que no concuerda con las demás.

Un estimador que sale relativamente favorecido de la comparación planteada en la Tabla 2, especialmente si además se tiene en cuenta la simplicidad de su cálculo, es la distancia entre cuartos  $DC$ . Precisamente esta característica y la resistencia a observaciones extremas mostrada en la Gráfica 2 son las razones por las cuales el uso de  $DC$  es básico en el análisis exploratorio de datos. Detalles sobre esta medida de escala y sus aplicaciones pueden encontrarse en los libros de Tukey (1977) y Hoaglin, Mosteller y Tukey (1982).

La ventaja de los estimadores  $S_w$  y  $S_b$  es bien clara. Además de ser estimadores altamente resistentes, como lo muestra la Gráfica 2, garantizan eficiencias considerables. Por ejemplo  $S_b$  tiene una eficiencia mínima del 85.8% en caso de que los datos provengan de alguna de las distribuciones consideradas o de combinaciones de ellas. Las propiedades descritas de  $S_b$  han motivado su aplicación en el desarrollo reciente de nuevos procedimientos estadísticos. Gross (1976) la utiliza en la construcción de intervalos de confianza, para el parámetro

de localización, con longitud relativamente estable ante la presencia de datos generados por diferentes distribuciones. Shoemaker (1982) encuentra aplicaciones en el análisis de componentes de varianzas. Martínez e Iglewicz (1981), mediante la comparación de  $S_b^2$  con una modificación de  $S^2$ , desarrollan un método estadístico para la verificación de la hipótesis de normalidad en un conjunto de datos. El reemplazo de  $S$  por  $S_b$  permite a los mismos autores incrementar la potencia de varios de los métodos usados en la detección de observaciones extremas.

## 6. Conclusiones.

El presente artículo muestra las fallas de la desviación estándar muestral  $S$  cuando se le usa con distribuciones que se alejan de la acostumbrada hipótesis de normalidad. Este estimador de escala resulta ser poco resistente a observaciones extremas y pierde de una manera drástica la eficiencia absoluta obtenida en el modelo normal, cuando se le utiliza con datos que aunque siguen siendo simétricos no se ajustan a esta distribución. Una excelente alternativa es el uso de  $S_b$ , la desviación estándar bicuadrada. Este estimador de escala además de ser bastante resistente a observaciones discordantes con el resto de los datos, garantiza altas eficiencias a medida que los datos

muestran alejamiento de hipótesis de normalidad. Las propiedades mencionadas junto con las buenas experiencias obtenidas en sus aplicaciones permiten recomendar a  $S_b$  como un buen estimador de escala. Otro estimador de escala que merece especial atención es la distancia entre los cuartos  $DC$ . Aunque es mucho menos eficiente que  $S_b$ , tiene gran resistencia a observaciones extremas y es muy simple de calcular. Estimadores de escala como los mencionados han permitido desarrollar procedimientos para el análisis de datos que tratan de cumplir a cabalidad uno de los objetivos fundamentales de la estadística: detectar de una manera eficiente cambios reales en los datos a pesar de la presencia de la inevitable variabilidad experimental.

\* \*

#### BIBLIOGRAFIA

- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. y Tukey, J.W., *Robust estimates of location: survey and advances*. Princeton University Press, Princeton, (1972).
- Beaton, A.E. y Tukey, J.W., "The fitting of power series and polynomials, illustrated on band spectroscopic data". *Technome*

- trics* 16, 147-185 (1974).
- Gross, A.M., "Confidence interval robustness with longtailed symmetric distributions", *J. Amer. Statist. Assoc.* 71, 409-416 (1976).
- Hettmansperger, J.P. y Mc Kean, J.W., "Statistical influence based on ranks", *Psychometrika* 43, 69-79 (1978).
- Hoaglen, D.C., Mosteller, F., y Tukey, J.W., *Underestanding Robust and Exploratory Data Analysis*, John Wiley e hijos, Nueva York, (1982).
- Huber, P.J., *Robust Statistics*. John Wiley e hijos, Nueva York, (1981).
- Iglewicz, B. y Martínez, J., "Outlier detection using robust measures of scale", *J. Statist. Comput. Simul.* 15, 285-293 (1982).
- Lax, D.A. "An interim report of a Monte Carlo study of robust estimators of width", Reporte Técnico 93, Serie 2, Departamento de Estadística, Universidad de Princeton. (1975).
- Martínez, J. e Iglewicz, B. "A test por departure from normality based on a biweight estimator of scale", *Biometrika*, 68, 331-333 (1981).

Shoemaker, L.H. y Hettmansperger, T.P., "Robust estimates and test for the one and two sample scale problems", *Biometrika*, 69, 47-53, (1982).

\* \* \*