DETECCIÓN DE OBSERVACIONES ATÍPICAS MEDIANTE TRUNCAMIENTOS: CASO UNIVARIANTE

ORTEGA DATO, Juan Fco.
Departamento de Economía y Empresa
Universidad de Castilla-La Mancha
correo-e: JuanFco.Ortega@uclm.es

RESUMEN

El estudio y protección ante la presencia de observaciones atípicas, en las muestras utilizadas para inferir parámetros de una población, es de vital interés en todo estudio estadístico. Estas observaciones producen tales efectos en los resultados, que éstos dependen finalmente de sólo una pequeña parte de la información de la que se parte, siendo ésta además la constituida por estas observaciones que no son genuinas del experimento que se está estudiando.

En este papel se propone un método de detección de observaciones atípicas, en muestras univariantes y bajo el supuesto de Normalidad, basado en unas estandarizaciones construidas usando estimadores de posición y escala definidos mediante truncamientos, los denominados α*_Truncadas* y $\alpha\beta$*_Truncadas*.

Palabras clave: Detección de outliers. Trimmed means. Estimadores de posición y escala robustos.

1 Introducción

La presencia de observaciones atípicas en las muestras utilizadas para inferir parámetros de un modelo, producen malas influencias en los resultados, proporcionando modelos erróneos. Así, un primer problema que se nos puede plantear al decidir inferir un modelo a partir de una muestra, es que ésta puede contener observaciones no deseables, observaciones atípicas, que enturbian y nos proporcionan información errónea.

Como definición de observación atípica consideraremos la dada por Barnett y Lewis (1994), donde éstas son observaciones que parecen ser sorprendentemente discordantes o inconsistentes con respecto al resto de observaciones de la muestra en la que se encuentra.

Para proteger los resultados de las posibles observaciones atípicas, se definen unos procedimientos denominados métodos de detección de "outliers". Estos métodos se basan en ciertas medidas que nos permiten averiguar si, en una muestra dada para la estimación de unos determinados parámetros de un modelo, una observación es atípica; siendo candidata a ser estudiada a fondo para su posible eliminación de la muestra.

En la mayoría de estos procedimientos se utilizan como elementos ciertos estimadores de posición y escala, de manera que las buenas propiedades o comportamientos de dichos procedimientos dependen de los comportamientos de estos estimadores. Por esta razón, a los estimadores utilizados habitualmente en los métodos de detección de "outliers" se les piden ciertas propiedades o características en presencia de observaciones atípicas, en definitiva que sean robustos.

En la literatura se pueden encontrar diferentes familias de estos estimadores robustos, tanto de posición como de escala. Una de estas familias, los L-Estimadores, contiene una subfamilia conocida por el nombre de "trimmed means" (que traduciremos por medias truncadas), definidas como media de las observaciones de la muestra tras haber eliminado un porcentaje de las mismas. Este porcentaje es determinado por el parámetro α, denominado nivel de truncamiento, por lo que a la media truncada de parámetro α se denota por α*_Truncadas*, construyéndose una familia de estimadores de posición al mover α en su conjunto de definición. Dada esta familia de estimadores de posición, es posible construir una familia de estimadores de escala sustituyendo, en la definición de la conocida varianza, medias por medias truncadas. Para ello es necesario

utilizar dos parámetros, α y β, de manera que los elementos de nueva familia serán denotados por $\alpha\beta$*_Truncadas*.

Ambas familias de estimadores, de posición y escala, son especialmente útiles cuando se aplican sobre muestras de variables con distribuciones simétricas. Esto es así, ya que, dada sus construcciones en las que se eliminan observaciones simétricamente, es bajo estas condiciones cuando sus propiedades son óptimas.

En este papel propondremos un procedimiento para detectar observaciones atípicas en modelos univariantes, bajo normalidad, utilizando como estimadores de posición las α*_Truncadas* y como estimadores de escala las $\alpha\beta$*_Truncadas*. El estudio realizado comienza, en la Sección 2, presentando las medias truncadas y la familia de estimadores de escala definida a partir de ella. En la siguiente sección, Sección 3, proponemos un método para determinar los niveles de truncamientos apropiados, para la elección de unos elementos en dichas familias. En la Sección 4 presentamos el procedimiento para detectar observaciones atípicas, así como un ejemplo que nos ilustre su uso. Por último, en la Sección 5 realizaremos unos comentarios finales a modo de resumen.

2 Estimadores de posición y escala mediante truncamientos

Los estimadores de posición más utilizados en la literatura, la media y la mediana, pueden ser considerados como elementos extremos de la familia de medias truncadas. Por otro lado, dichos estimadores de posición tienen asociados dos estimadores de escala bien conocidos: el más utilizado, la desviación típica (o su cuadrado la varianza); y el denominado *mediana de desviaciones absolutas a la mediana* (*Meda*) dado en Hampel y otros (1986).

Partiendo de la estructura de la varianza y sustituyendo medias por elementos de la familia de las medias truncadas, en Ortega (2003) se propone unos elementos, denotados por $\alpha\beta$*_Truncadas*, que proporcionan una familia de estimadores de escala, donde sus elementos extremos son la 0,0*_Truncadas*, la varianza, y 50,50*_Truncadas*, concepto muy similar al cuadrado de la *Meda*. Presentemos a continuación la notación y principales propiedades de los citados estimadores de posición y escala.

2.1 Estimadores de posición mediante truncamientos

Los elementos de la familia de medias truncadas, utilizan ara su definición un número real que denotaremos por α, con éste en el intervalo [0,50) y que denominaremos nivel de truncamiento. Así, la media truncada a nivel de truncamiento α (α*_Truncada*) se construye como la media e las observaciones de la muestra eliminando el α% de las observaciones mayores y el α% de las menores. Es decir, dada una colección de n elementos $x=\{x_1,x_2,...,x_n\}$ y un valor $\alpha\in[0,50)$, se define α*_Truncada* sobre x, denotándolo por α*_Trun*(x), de la forma:

$$\alpha\_Trun(x) = \frac{1}{n-2a}\sum_{i=a+1}^{n-a} x_{[i]} \qquad (1)$$

donde $a=Int(\alpha n/100)$ y $x_{[i]}$ denota la observación en la posición i-ésima de una ordenación de menor a mayor de los elementos de x. Además, notar que si α tiende a 50, entonces α*_Truncada* tiende a la mediana, por lo que; tomando por convenio que 50*_Truncadas* es la mediana muestral, podemos definir las α*_Truncadas* para todo α en el intervalo cerrado [0,50].

Es conocido que, para cualquier α, las α*_Truncadas* son estimadores de posición *insesgados* en variables con distribuciones simétricas; que son *equivariantes afín*, con *punto de ruptura* (a+1)/n, y *punto de ruptura asintótico* de valor α/100; y que tienen *curva de sensibilidad* acotada, siempre que $a\geq1$. Por otra parte, su *eficiencia* bajo Normalidad (considerando *eficiencia relativa* con respecto a la media) puede ser aproximada mediante la forma lineal (100-0.723α)% para $\alpha\in[0,50]$, por lo que los elementos de esta familia recorren los diferentes rangos de esta propiedades; entre el 64% de la mediana y el máximo (el 100%) de la media.

2.2 Estimadores de escala mediante truncamientos

Partiendo de la estructura de la conocida varianza y sustituyendo medias por elementos de la familia de las medias truncadas, se construye (Ortega, 2003) unos elementos que proporcionan una familia de estimadores de escala con un comportamiento apropiado en presencia de observaciones atípicas.

Así, ya que la varianza se define como media de los cuadrado de las distancias de las observaciones a la media, se propone sustituir las dos medias por otras tantas medias truncadas. De esta manera, dada la colección de elementos $x=\{x_1,x_2,...,x_n\}$, los parámetros $\alpha,\beta\in[0,50]$ y conocida la familia de medias truncadas (1), se definen los elementos $\alpha\beta\_Truncadas$ sobre x, denotadas por $\alpha\beta\_Trun(x)$, como:

$$\alpha\beta\_Trun(x) = C(\alpha,\beta)\ \ \beta\_Trun(\ \{(x_i - \alpha\_Trun(x)\ )^2\}_i\ ) \qquad (2)$$

donde $C(\alpha,\beta)$, para los diferentes valores de α y β, son unos coeficientes de consistencia. Dichos coeficientes son invariantes ante cambios de posición y/o escala en la muestra que se aplican, y, bajo el supuesto de Normalidad, tiene el valor de $C(\alpha,\beta)=[\beta\_Trun(\ \chi_1^2\ )]^{-1}$.

Los extremos de la familia $\alpha\beta\_Truncadas$ son la varianza, para el caso de $\alpha=\beta=0$, y el concepto llamado mediana de desviaciones a la mediana (*MDM*), en el caso de $\alpha=\beta=50$, definido de la forma:

$$MDM(x)\ =\ C(MDM)\ \ Med\{\{(x_i - Med(x))^2\}_i\} \qquad (3)$$

donde de nuevo $C(MDM)$ es un coeficiente consistencia y *Med* es la mediana muestral.

La raíz cuadrada de *MDM* es muy similar a la conocida *Meda*, encontrándose diferencias mínimas entre ellos causadas por la paridad del tamaño muestral. Es de destacar que, *MDM* respeta de una manera más estricta la estructura de la varianza, al considerar cuadrados de distancias y no valores absolutos, lo que en la práctica puede ser útil a la hora de manejar dicho concepto.

De las propiedades de los elementos de la familia $\alpha\beta\_Truncada$ se pueden destacar que son: *equivariantes*; con *eficiencia* (respecto a la varianza corregida) entre el 38.35% y el 100%, dependiendo de los niveles de truncamiento considerados, y con respecto a su robustez, siendo $a=Int(\alpha n/100)$ y $b=Int(\beta n/100)$, tienen *punto de ruptura* de valor $Min\{(a+1)/n,\ (b+1)/n\}$, siendo el *punto de ruptura asintótico* de la forma $Min\{\alpha/100,\ \beta/100\}$; y *curva de sensibilidad* acotada siempre que $Min\{a,b\}\geq 1$.

En definitiva, la raíces cuadradas positivas de los elementos de la familia dada en (2) pueden ser considerados como estimadores de escala, con (dependiendo de los niveles α y β) buenas propiedades ante la presencia de observaciones atípicas.

3 Elección de los niveles de truncamiento

Como hemos comprobado en la sección anterior, los parámetros α y β condicionan el comportamiento de los elementos α*_Truncada* y $\alpha\beta$*_Truncada* ante la presencia de observaciones atípicas (*robustez*) y también sus *eficiencias*. Así, debemos determinar qué valores de α y β son los posibles en cada caso, y posteriormente, cuales serían los aconsejables u óptimos en cada situación particular.

3.1 Posibles niveles de truncamiento

En Ortega (2002) se determinan los posibles valores que puede tomar α para que las α*_Truncadas* proporcionen diferentes valores entre sí. En él se denota por $Pv(x)$ a dicho conjunto sobre $x=\{x_1,x_2,...,x_n\}$, construido de la forma:

$$Pv(x) = \{0, s, 2s, \dots , ks\} \tag{4}$$

siendo $s=100/n$ y $k=Int((n-1)/2)$.

Por lo tanto, para (1) podemos considerar que $\alpha \in Pv(x)$, mientras que para el caso de las $\alpha\beta$*_Truncadas*, sobre la misma muestra x, de nuevo α y β están contenidos en este mismo conjunto, ya que la construcción de éste sólo dependen del tamaño muestral n.

En definitiva, sobre una muestra x de tamaño n, los niveles de truncamiento α y β para las familias α*_Truncada* y $\alpha\beta$*_Truncada* deben ser elegidos en el conjunto $Pv(x)$.

3.2 Niveles de truncamiento óptimos

La elección de un nivel de truncamiento para la familia α*_Truncada* es estudia en diferentes trabajos (por ejemplo; Dodge y Jurecková, 2000). Así, se pueden encontrar procedimientos que determinan este nivel dependiendo de ciertas características particulares de la muestra, mientras que otros se basan en la idea de asegurar una determinada eficiencia o robustez, dependiendo de las necesidades del estudio o de las preferencias del investigador.

En Ortega (2002) se estudia también este problema, proponiéndose un criterio que consiste en elegir el menor nivel de truncamiento de manera que, las diferencias

entre las α_*Truncadas* para valores mayores de éste sean "pequeñas"; entendiendo por esto que sean menores que una determinada cota. Éste será el método que utilizaremos en el presente trabajo.

Así, dada una muestra x de tamaño n, se define el nivel de truncamiento óptimo de cota ε al valor α_ε dado por:

$$\alpha_\varepsilon = Min \{ \alpha \in Pv(x) \ / \ | u\_Trun(x) - v\_Trun(x) | < \varepsilon, \ \forall \ u,v \geq \alpha \} \qquad (5)$$

donde el valor ε es la citada cota denominada *cota de estabilidad*, la cual pretende controlar las diferencias entre los posible valores de las α_*Truncadas*, y permitir determinar a partir de que nivel de truncamiento las diferencias de éstas pueden ser consideradas no significativas, generando así el nivel de truncamiento óptimo.

El resultado del trabajo propone como cota de estabilidad una dependiente del tamaño de la muestra y de un estimador de escala robusto, que la proteja de las posibles contaminaciones. Así, bajo el supuesto de Normalidad y mediante un estudio de simulación sin contaminaciones, se llega a la conclusión de que la cota de estabilidad para una muestra x de tamaño n será:

$$\varepsilon_n = 1.7350 \ n^{-0.4746} \ S_R \qquad (6)$$

donde S_R es un estimador de escala robusto sobre los datos de x, aconsejándose, en el trabajo citado, el uso de la *Meda* como este estimador.

En definitiva, dada una muestra x de tamaño n, tomaremos como α óptimo el valor denotado por α_0 donde:

$$\alpha_0 = Min \{ \alpha \in Pv(x) \ / \ | u\_Trun(x) - v\_Trun(x) | < \varepsilon_n, \ \forall \ u,v \geq \alpha \} \qquad (7)$$

Para la elección de los niveles de truncamiento óptimos en $\alpha\beta$_*Truncada* utilizaremos $\alpha=\alpha_0$, y para β seguiremos el procedimiento antes expuesto, realizando algunas variaciones. Así, proponemos construir β_0 como:

$$\beta_0 = Min \{ \alpha \in Pv(x) \ / \ | \alpha_0 u\_Trun(x) - \alpha_0 v\_Trun(x) | < \varepsilon'_n, \forall \ u,v \geq \beta \} \qquad (8)$$

donde ε'_n es la cota de estabilidad pero ahora para las $\alpha\beta$_*Truncadas.*

Ya que las $\alpha\beta$_*Truncadas* son *equivariantes afín* y bajo el supuesto de Normalidad, decimos que $\varepsilon'_n = \hat{\varepsilon}'_n \ S_R^2$, de forma similar a la construcción de ε_n, donde

S_R es un estimador de escala robusto sobre x y donde $\hat{\varepsilon}'_n$ es la cota para $\alpha\beta\_Truncadas$ sobre muestras de variables normales tipificadas. Dicho valor se calcula mediante simulación de la forma:

$$\hat{\varepsilon}'_n = Percentil_{99} \{ Max_{v>u} \{ | u\_Trun(x') - v\_Trun(x') | \} \}$$

para x' muestras de variables que se comportan como χ^2_1. Realizando estas simulaciones se obtiene en definitiva que:

$$\varepsilon'_n = 2.5332 \; n^{-0.2464} \; S_R^2 \tag{9}$$

S_R es un estimador de escala robusto sobre los datos de x, aconsejándose, como en el trabajo citado, el uso de la *Meda* como este estimador.

4 Detección de observaciones atípicas

Una de las vías de estudio de la presencia de observaciones atípicas es el uso de los métodos de detección de "outliers", los cuales se basan en ciertas medidas que nos permiten averiguar si, en una muestra dada para la estimación de unos determinados parámetros de un modelo, una observación es atípica, o si por el contrario su presencia no produce ninguna sorpresa, siendo admisible en la muestra del fenómeno que queremos analizar.

Un procedimiento sencillo, utilizado habitualmente en muestras univariantes, es el considerado en Rousseeuw y Leroy (1987) en su rutina PROGRESS. Éste utiliza una medida asociado a cada observación de la muestra que deja al descubierto la similitud entre dicha observación y la mayoría de observaciones de la muestra mediante una estandarización de las observaciones iniciales. Esta estandarización la define mediante la diferencia entre las observaciones y un determinado representante o "centroide" de ellas, estimador de posición, corregida por un elemento de dispersión de la muestra, estimador de escala.

Siguiendo esta idea, en nuestro caso definiremos una medida mediante una estandarización de las observaciones, utilizando como estimadores de posición las $\alpha\_Truncadas$ y como estimadores de escala $+\sqrt{\alpha\beta\_Truncadas}$ para ciertos niveles de truncamiento. Notar que en el caso extremo, donde α=β=50, se obtendrá la

estandarización utilizada en PROGRESS, salvando las diferencias entre $+\sqrt{MDM}$ y *Meda*, cuyo procedimiento tiene las mejores características en *robustez* pero con la mínima *eficiencia* heredada de la mediana.

De esta manera, dada una muestra $x=\{x_1,x_2,...,x_n\}$ consideramos la familia de estandarizaciones, moviendo los niveles de truncamientos $\alpha,\beta\in[0,50]$, siguientes:

$$\mathbf{DT}_{\alpha\beta} = \frac{x_i - \alpha\_Trun(x)}{+\sqrt{\alpha\beta\_Trun(x)}} \tag{10}$$

donde, si $\alpha\beta\_Trun(x)=0$ tomaremos en el cociente la varianza muestral, y si ésta es también 0 entonces, como ya hemos dicho, todas las observaciones de la muestra serán iguales y por lo tanto no existen observaciones atípicas.

Dichas estandarizaciones son independientes de cambios de posición y/o escala en la muestra x, y dependen de los niveles de truncamiento utilizados, de manera que estos determinarán sus características en *robustez* y *eficiencia*. De las posibles elecciones de los niveles de truncamiento, y por lo tanto de las medidas en (10), tomaremos en cada caso los proporcionados por el procedimiento de la sección anterior, denotados por α_0 y β_0.

Bajo el supuesto de Normalidad y de que los estimadores de posición y escala mediante truncamientos son buenas aproximaciones de los parámetros del modelo, *DT* se comporta como una $N(0,1)$, considerándose valores anómalos para éstas a aquellos mayores o iguales en valor absoluto a una cierta cota, determinada mediante un nivel de significación γ. Así, observaciones x_i con valor absoluto de *DT* mayores o iguales a esta cota serán consideradas observaciones atípicas en x.

Notar que $DT^2_{\alpha\beta}(x_i)$, para diferentes niveles α y β, es la distancia euclídea entre x_i y $\alpha\_Trun(x)$ corregida por $\alpha\beta\_Trun(x)$, por lo que serán denominadas distancias por truncamientos. Dichas distancias, bajo los mismos supuestos que antes, se comportan como una χ^2_1, por lo que el criterio antes expuesto para las estandarizaciones *DT* es posible sustituirlo por uno para dichas distancias.

En la práctica, tomaremos las distancias por truncamientos como elementos para la detección de observaciones atípicas, considerando como nivel de significación γ=0.01, de manera que valores de DT^2 mayores o iguales a 6.66 serán considerados

como anómalos, y las observaciones con dichas distancias observaciones atípicas. Veamos a continuación un ejemplo, utilizados en la literatura, que proporcione una visión práctica del procedimiento propuesto.

Ejemplo: En Davies and Gather (1993) se estudia un conjunto de observaciones, que denotaremos por x, que representan un muestra de una variable que se supone Normal. Dichas observaciones son:

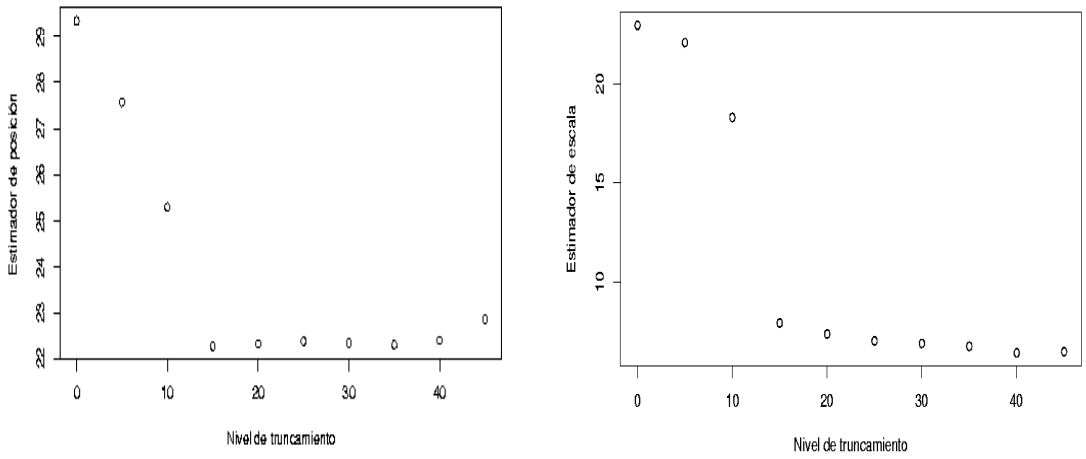
22.6 28.8 26.8 81.5 19.1 15.2 24.1 23.6 9.1 79.5

18.6 78.8 23.1 11.9 20.1 20.3 17.3 25.8 14.1 26.5

donde denotaremos por x_i a la observación i-ésima en x (tomada primero de izquierda a derecha y después de arriba abajo) en el conjunto anterior.

Para calcular las distancias por truncamientos sobre x, el procedimiento propuesto anteriormente sugiere que: primeramente se calcule el estimador de posición mediante truncamientos, para lo cual es necesario determinar el nivel de truncamiento óptimo para dicho estimador, que denotábamos por α_0, y para lo cual necesitamos la cota de estabilidad (ε_n) en este caso; seguidamente, debemos determinar el estimador de escala mediante truncamientos sobre x, para lo cual es necesario elegir el segundo nivel de truncamiento óptimo, denotado por β_0, por lo que ahora necesitamos determinar su correspondiente cota de estabilidad (ε'_n). Así, conocidos los niveles α_0 y β_0, podemos construir las distancias $DT^2_{\alpha_0\beta_0}$ sobre x que nos proporcionan información sobre la presencia de posibles observaciones atípicas en dicha muestra.

Para calcular los niveles de truncamiento α_0 y β_0, antes debemos determinar qué valores pueden tomar. Así, dado que el tamaño muestral de los datos del ejemplo es n=20 y utilizando (4) tenemos que, s=5 y k=9 por lo que el conjunto de posibles niveles de truncamiento sobre x es $Pv(x)$={0,5,10,15,...,45}, de manera que α_0 y β_0 están contenidos en este conjunto.



(a) (b)

Figura 1: Elección de α_0 y β_0 para la construcción del estimador de posición (a) y el estimador de escala (b) mediante truncamientos, para los datos de Davies y Gather (1993).

Si $\alpha_0 \in Pv(x)$, aplicando el procedimiento propuesto por Ortega (2002) se obtiene, en primer lugar, un gráfico de los estimadores de posición $\alpha\_Truncada$ sobre la muestra x, recogido en la Figura 1a. Siguiendo (6), la cota de estabilidad en x es de valor ε_n=2.55, para n=20 y donde el estimador de escala considerado es la $Meda(x)$=6.08. Utilizando la ecuación (7) se obtiene como nivel de truncamiento óptimo el valor α_0=15.

En definitiva, el procedimiento proporciona una estimación de parámetro de posición $15\_Trun(x)$=22.28, valor más idóneo que la media muestra, de valor 29.34.

Para construir el estimador de escala, debemos elegir el nivel de truncamiento $\beta_0 \in Pv(x)$, para ello, en la Figura 1b se presenta un gráfico de las $\alpha_0\beta\_Truncadas$ sobre x. Siguiendo el procedimiento propuesto, la cota de estabilidad es de valor ε'_n = 44.79, y utilizando la ecuación (8) se obtiene como nivel de truncamiento óptimo el valor β_0=15.

En definitiva, el procedimiento proporciona una estimación para el parámetro de escala de valor $+\sqrt{15{,}15\_Trun(x)} = 7.91$, valor más idóneo que la desviación típica de los datos que es 22.4.

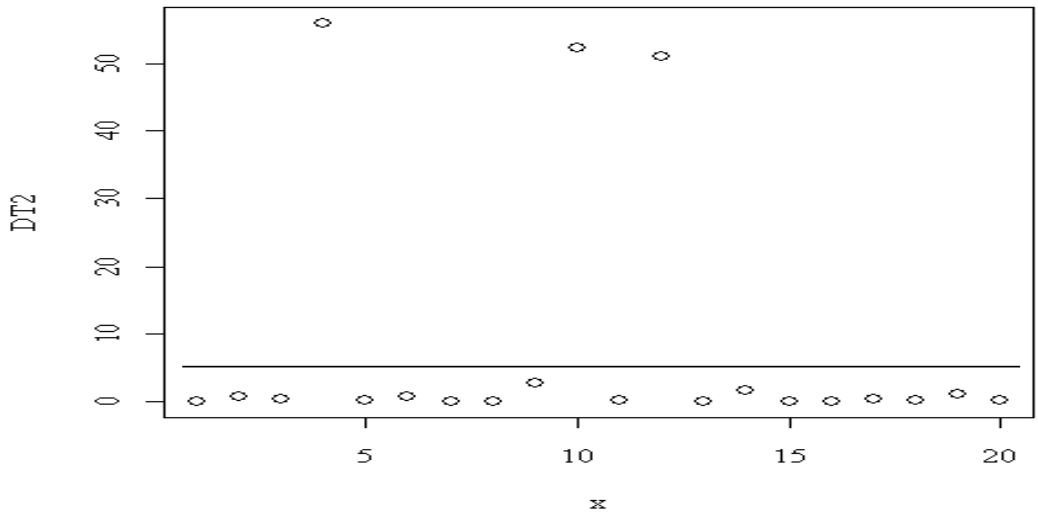


Figura 2: *Distancias por truncamientos* para los datos de Davies y Gather (1993).

Una vez que tenemos el estimador de posición y el de escala, es posible construir las distancias por truncamientos $DT^2_{\alpha_0\beta_0}$ para los datos de *x*, utilizando (10). En la Figura 2 se representan los valores de dichas distancias para las diferentes observaciones de *x*, junto con la cota para estas distancias, de valor 6.66. Como resultado, el procedimiento propuesto detecta como observaciones atípicas las de índices 4, 10 y 12.

Notar que, en el caso de haber utilizado las medidas de posición y escala no robustas, la media y la desviación típica, ninguna de las observaciones hubiera sido detectada como anómala. Por otro lado, mientras que la media de la muestra original es de valor 29.34 y la desviación típica es de valor 21.83, eliminando dichas observaciones atípicas se obtiene que la media muestral es 20.41 y la desviación típica 6.4, valores más similares a los obtenidos con el método robusto sobre las observaciones originales.

Para mostrar un punto de vista diferente del procedimiento propuesto, oculto en los datos *x* del ejemplo, supongamos que x_{12}=-78.8. En este caso, se obtiene que α_0=0, según el procedimiento de elección del nivel de truncamiento para el estimador de posición, mientras que β_0 sigue siendo 15. Notar que, la elección de α_0 se ajusta con objeto de que el estimador de posición no sea influenciado por la posibles observaciones atípicas, por lo que el hecho de que este nivel sea 0 no significa que no hayan observaciones anómalas, sino que, en el caso de que las hubiera, éstas no influyen en su determinación.

5 Conclusiones

El problema por la presencia de observaciones atípicas, en muestras utilizadas para inferir parámetros de un modelo, es de especial interés. Así, si utilizamos estimadores clásicos en muestras contaminadas, los resultados está muy influenciado por las contaminaciones, de manera que los parámetros determinados serán erróneos.

En este trabajo se propone un procedimiento, sencillo y de fácil implementar en cualquier paquete estadístico, para la detección de observaciones atípicas, bajo Normalidad y en muestras univariantes. El resultado del procedimiento propuesto es una medida, denominada distancia por truncamientos, basada en las α*_Truncada*s y las $\alpha\beta$*_Truncada*s, definidas a su vez mediante las conocidas medias truncadas. Esta distancia, para cada observación de la muestra, nos informa de lo apropiado que es suponer que dicha observación es una realización del experimento que se está estudiando.

Bibliografía

1. Andrews, D.F. y otros (1972): *Robust Estimates of Location*. Princeton University Press.
2. Barnett, V. y Lewis, T. (1994): *Outliers in Statistical Data*. 3st. Ed. Wiley and Sons.
3. Davies, L. y Gather, U. (1993): "The identification of multiple outliers". J. of the American Statistical Association, Vol. 88, No. 423.
4. Hampel, F.R. y otros (1986): *Robust Statistics: The approach based on influence functions*. Ed. Wiley and Sons. N.Y.
5. Hoaglin, D.C. y otros (2000): *Understanding Robust and Exploratory Data Analysis*. Ed. Wiley and Sons.
6. Ortega, J.Fco. (2002): "Elección de un nivel de truncamiento óptimo para la familia de medias truncadas". Documento de Trabajo de la Facultad de CC. Económicas y Empresariales de Albacete. Universidad de Castilla-La Mancha.
7. Ortega, J.Fco. (2003): "Familia de estimadores de escala mediante truncamientos". Documento de Trabajo de la Facultad de CC. Económicas y Empresariales de Albacete. Universidad de Castilla-La Mancha.
8. Rousseeuw, P.J. y Croux C. (1993): "Alternatives to the Median Absolute Deviation". J. of the American Statistical Association, No. 424.
9. Rousseeuw, P.J. y Leroy, A.M. (1987): *Robust Regression and Outlier Detection*. Ed. Wiley and Sons.