



COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

Revista Iberoamericana de Bioeconomía y Cambio Climático  
(Rev. iberoam. bioecon. cambio clim.)  
Vol. 4 num 7, 2018, pag 903-923  
ISSN electrónico 2410-7980

## Aplicación de la estadística robusta en la identificación de valores “OUTLIERS” en análisis de residuos de clorpirifos en maní. programa elaborado en Hp-Basic utilizando la calculadora gráfica HP Prime

Delgado, Gustavo <sup>a\*</sup>; Hernández, Norma <sup>b</sup>; Méndez, Leticia <sup>b</sup>; Abad Rojas<sup>b</sup>

<sup>a</sup> Consultor sobre Sistema de la Calidad en Laboratorios de Ensayos, Profesor Retirado de la Universidad Nacional Autónoma de Nicaragua, UNAN-LEÓN, Facultad de Ciencias y Tecnología, Edificio de Ciencias Básicas, León, Nicaragua.

<sup>b</sup> Laboratorio SER-AGRO S.A., Km 138 carretera Chinandega-Corinto, Tel: 505 2340 3493. (Acreditado por la ONA-MIFIC).

(\* ) Autor por correspondencia: [gusdel53@gmail.com](mailto:gusdel53@gmail.com)

DOI: <http://dx.doi.org/10.5377/ribcc.v4i8.6466>

Recibido: 15 Marzo 2018

Aceptado: 24 Agosto 2018

### Resumen

En el presente trabajo se ha aplicado la estadística robusta para la identificación de valores “outliers” en la determinación del factor recobrado en el análisis de residuos de clorpirifos en maní por cromatografía de gases con detector  $\square$ ECD. Para ello fue elaborado un programa en lenguaje HP-BASIC, basado en la estadística robusta y específicamente en el algoritmo de Huber y el cual fue editado en la calculadora gráfica de última generación HP PRIME Graphing Calculator. El programa fue ejecutado ingresando los datos obtenidos de los resultados de la evaluación del factor de recobrado para el método de análisis de clorpirifos en una muestra de referencia interna de maní, obteniéndose como resultado un estimado de los parámetros de la población y la identificación de 3 “outliers”. Este programa fue validado con uno de referencia elaborado en Excel por el Comité de Métodos Analíticos de la Royal Society of Chemistry, observándose que la desviación de los parámetros de la población ( $\mu, \sigma$ ) fueron menores que la precisión iterativa, lo que indica que no hubo diferencias significativas entre los dos programas. Por otro lado, estos resultados se compararon con los obtenidos por las pruebas clásicas de Dixon y de Grubbs, concluyendo que el método de Huber es el más sensible. El menos sensible es el test de Dixon, dado que su aplicación no detectó ningún outlier, mientras que el test de Grubbs identificó los dos valores más pequeños.

**Palabras claves:** Valores outliers; Estadística robusta; Algoritmo de Huber; Test de Dixon; Test de Grubbs; Lenguaje HP Basic; Factor recobrado; Análisis clorpirifos en maní.

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .

## **Application of robust statistics in the identification of values "OUTLIERS" in analysis of residual chlorpyrifos in peanuts. program developed in Hp-Basic using the HP Prime graphic calculator**

### **Abstract**

In the present work, robust statistics have been applied for the identification of "outliers" values in the determination of the recovered factor in the analysis of chlorpyrifos residues in peanut by gas chromatography with detector  $\square$ ECD. For this, a program was developed in the HP-BASIC language, based on robust statistics and specifically on the Huber algorithm and which was edited in the latest generation HP PRIME Graphing Calculator. The program was executed by entering the data obtained from the results of the evaluation of the recovery factor for the method of analysis of chlorpyrifos in an internal peanut reference sample, obtaining as a result an estimate of the parameters of the population and the identification of 3 "Outliers" This program was validated with a reference made in Excel by the Committee of Analytical Methods of the Royal Society of Chemistry, observing that the deviation of the parameters of the population ( $\mu, \sigma$ ) were lower than the iterative precision, which indicates that there were no significant differences between the two programs. On the other hand, these results were compared with those obtained by the classic tests of Dixon and Grubbs, concluding that the Huber method is the most sensitive. The least sensitive is the Dixon test, given that its application did not detect any outlier, whereas the Grubbs test identified the two smallest values.

**Keywords:** Outliers values; Robust statistics; Algoritmo de Huber; Dixon test; Grubbs test; HP Basic Language; Recovered factor; Chlorpyrifos analysis in peanuts.

### **1. Introducción**

Actualmente, Nicaragua está ampliando el mercado de productos agroalimentarios, y entre ellos el maní es uno de los productos que más se exporta a Latinoamérica, EE. UU. y Europa Delgado G y Hernández N., (2014). Para garantizar su inocuidad se hace necesario la implementación del Sistema de la Calidad a todo lo largo de la cadena de producción, desde el cultivo hasta el procesamiento industrial. Uno de las exigencias que imponen los países compradores es la certificación del producto, para lo cual debe cumplir con los requisitos establecidos en el Codex Alimentarius, (2000), o por la legislación europea European Food Safety Authority, (2014). El laboratorio analítico juega un papel muy importante en la evaluación de la conformidad de este producto. Esto significa que se debe garantizar la calidad de los resultados de los ensayos bajo tres

---

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

principios básicos relacionados: aptos para el uso del cliente, método validado y control de calidad de acuerdo a ISO/IEC 17025(E) (2017) y Thompson M., Lowthian P.L.,(2011) . Esto implica que los resultados que se reportan en el certificado de análisis deben reflejar trazabilidad, precisión y confiabilidad.

No obstante, es muy frecuente que los datos obtenidos en las mediciones analíticas están sujetos a valores considerados discordantes con el resto de los valores de la serie, los cuales son denominados “outliers” AMC-RSC, (1989). Estos valores sospechosos deben ser investigados con el fin de evitar que el laboratorio emita certificados de análisis con resultados sesgados, ya que esto conlleva a la no conformidad del producto, provocando desconfianza en los clientes y por consiguiente pérdidas para la empresa y el país. La investigación de las causas y la aplicación de las acciones correctivas, ISO/IEC 17025(E) (2017), pertinentes son muy importantes para la garantía de la calidad de la certificación del producto. Entre las causas más frecuentes de estas no conformidades están los errores durante la transcripción o registro de los datos, los cometidos durante el proceso analítico, los debidos a la contaminación o a la transposición de la muestra.

Las dos técnicas estadísticas que comúnmente se utilizan para identificar “outliers” son el test de Dixon, Dixon W.J, (1950) y el test de Grubbs Grubbs F.E., (1969); ISO5721-2,(1994), *Tecnometrics*, 1969, 11, 1. Estas técnicas dependen de los valores extremos y del tamaño de los datos. Si se utiliza la estadística robusta, se evita este problema, ya que depende de la mediana y de la mediana de la desviación absoluta (MAD). Esta estadística se aplica en forma aproximada, evaluando el intervalo de aceptación de los datos con la mediana y el MAD multiplicada por un factor 3 según Thompson M., Lowthian P.L.,(2011). Pero para obtener una mejor estimación de los parámetros de la población y rechazar los valores discordantes con mayor certeza, es necesario aplicar el algoritmo de Huber (ISO 5721-5, (1994)), el cual consiste en calcular la media y la desviación estándar por iteraciones sucesivas, tomando como valores iniciales la mediana y el MAD. En ensayos de aptitud es muy útil para demostrar la competencia técnica de los laboratorios de ensayos. Este algoritmo ha sido programado en lenguaje FORTRAN 77 (AMC-RSC, (1989)), el cual

---

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

converge hasta obtener una media y una desviación estándar con una aproximación mínima de 0.01%. Para que este programa se ejecute es necesario disponer de una computadora dotada con este tipo de sistema de programación. Este programa se puede editar en Visual Basic integrado con Excel o con el cálculo simbólico MAPLE 2015. En cualquiera de los casos una computadora es imprescindible. Cuando se lleva a cabo un control de calidad inmediato, es decir durante el proceso de análisis de rutina, donde se necesita evaluar en forma rápida algún parámetro de calidad analítica se hace de mucha utilidad un medio de cálculo portátil, es decir una calculadora programable apropiada.

El presente estudio pretende aplicar la estadística robusta para la identificación de “outliers” en la determinación del factor de recobrado en análisis de residuos de clorpirifos en maní de exportación, aplicando un programa basado en el algoritmo de Huber, elaborado en lenguaje HP-BASIC en la calculadora gráfica de última generación HP-PRIME Graphing (HP PRIME, (2013)). Así mismo, se realiza una comparación de ésta con las pruebas estadísticas clásicas, el test de Dixon y el de Grubbs.

## 2. Revisión de la literatura (Teoría)

La técnica estadística más utilizada para investigar valores “outliers” en una serie de datos producidos en los laboratorios de ensayos es el test de Grubbs, la que es descrita en la norma ISO 5725-2 (ISO 5721-2, (1994)). Esta prueba depende del tamaño de la muestra estadística y es muy sensible para identificar dos valores “outliers” sucesivos. Para valores extremos (el valor más pequeño y el más grande) el parámetro de Grubbs ( $G_k$ ) se calcula de acuerdo a la siguiente fórmula:

$$G_k = \frac{|x_k - \bar{x}|}{s} \quad (1)$$

Donde  $k$  es **1** para el valor más pequeño y  $n$  para el más grande;  $\bar{x}$  y  $s$  son el promedio y la desviación estándar de todos los datos. El valor resultante de la aplicación de la ecuación (1) se compara con los valores de la tabla de Grubbs (Feinberg M., (1996)), al 95% o 99% de nivel de confianza. Si los valores de  $G_k$  calculados son mayores que los de las tablas de Grubbs se dice que el valor dudoso

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



es “outlier” y debe ser rechazado del resto de la serie de datos. Este rechazo debe ser fundamentado a través de un procedimiento de acciones correctivas de acuerdo a al requisito 8.7 de la norma ISO/IEC 17025:2017.

La prueba para dos valores dudosos sucesivos, ya sea los dos valores más pequeños o los dos más grandes, se aplican las ecuaciones (2) y (3) respectivamente.

$$G_L = \frac{\sum_1^{p-2} (x_i - \bar{x}_L)^2}{\sum_1^p (x_i - \bar{x})^2} \quad (2)$$

$$G_H = \frac{\sum_3^p (x_i - \bar{x}_H)^2}{\sum_1^p (x_i - \bar{x})^2} \quad (3)$$

Donde el subíndice **L** indica el valor más bajo de la serie y **H** el valor más alto,  $p$  es el tamaño de la serie de datos,  $\bar{x}_L$  es el promedio calculado de la serie para los dos valores menores excluyendo los dos mayores; y  $\bar{x}_H$  es el promedio para la serie excluyendo los dos valores más pequeños. La formulación es la siguiente:

$$\bar{x}_L = \frac{\sum_1^{p-2} x_i}{p-2} \quad (4)$$

$$\bar{x}_H = \frac{\sum_3^p x_i}{p-2} \quad (5)$$

La otra técnica estadística es la prueba de Dixon, la cual consiste en calcular el parámetro  $Q_i$  para los valores extremos. El valor más pequeño ( $x_1$ ) y el más grande por ( $x_n$ ) son verificados por las ecuaciones siguientes:

$$Q_1 = \frac{x_2 - x_1}{x_n - x_1} \quad (6)$$



$$Q_n = \frac{x_n - x_{n-1}}{x_n - x_1} \quad (7)$$

Si el valor de  $Q_i$  es mayor que el de las tablas de Dixon <sup>[11]</sup> se rechaza el valor dudoso y se dice que es “outlier”.

La diferencia principal con el test de Grubbs, es que la prueba de Dixon no puede verificar los dos valores sucesivos extremos. Por otro lado, el test de Huber tiene la ventaja sobre los dos anteriores que no depende de los valores extremos, ni del tamaño de la muestra estadística. Además, se pueden identificar simultáneamente todos los outliers y obtener una mejor estimación de los parámetros de la población. El algoritmo en que está basado esta técnica es denominado algoritmo de Huber, el cual es recomendado por la norma ISO 5725-5 (ISO 5721-5, (1994))<sup>1</sup> consiste de las siguientes etapas según Thompson M., Lowthian P.L.,(2011) e ISO 5721-5, (1994):

- Cálculo de la mediana ( $\tilde{x}_i$ ) de los  $x_i$  valores y el MAD=mediana( $|x_i - \tilde{x}_i|$ ).
- Cálculo de la desviación estándar inicial  $s_j$ , dividiendo el MAD entre 0.6745.
- Estimar los valores iniciales de los parámetros de la población ( $\mu_0 = \tilde{x}_1$  y  $\sigma_0 = s_1$ ).
- Si el valor de  $x_i$  es mayor que  $\mu_0 + 1.5 \sigma_0$ , entonces  $\tilde{x}_i = \mu_0 + 1.5 \sigma_0$ .
- Si el valor de  $x_i$  es menor que  $\mu_0 - 1.5 \sigma_0$ , entonces  $\tilde{x}_i = \mu_0 - 1.5 \sigma_0$ .
- Si el valor de  $x_i$  cae en el intervalo, es decir  $|\tilde{x}_i - \mu_0| < 1.5 \sigma_0$ , entonces  $\tilde{x}_i = x_i$ .
- Cálculo de la media mejor estimada  $\mu_j = \frac{\sum \tilde{x}_i}{n}$
- Cálculo de la desviación estándar robusta  $\sigma_j = \sqrt{\frac{\sum (\tilde{x}_i - \mu_j)^2}{0.778(n-1)}}$
- El proceso finaliza cuando  $\frac{|\mu_{j-1} - \mu_j|}{\mu_j} < 0.0001$  o  $\frac{\sigma_j}{\sigma_{j-1}} - 1 < 0.0001$
- Finalmente, los parámetros de la población estimados son:  $\mu_j$  y  $\sigma_j$ .

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .

### 3. Metodología

Para determinar el factor de recobrado se fortificaron muestras negativas de maní siguiendo las directrices del método oficial del Laboratorio SER-AGRO, S.A. SAOSA-001, (2015). Para elaborar el programa con el algoritmo de Huber fue utilizada la calculadora HP PRIME GRAPHING CALCULATOR. Para su validación se descargó de internet el programa en Excel elaborado por el Comité de Métodos Analíticos de la Royal Society of Chemistry Feinberg M., (1996).

El procedimiento para la edición del programa en la calculadora HP PRIME es el siguiente:

1. Ingresar al editor de programas y editar con el nombre: MediaRobust-00
2. Introducir serie de datos  $x_i$  en D1.
3. Cálculo de la mediana  $\tilde{x}$ : Median(D1).
4. Cálculo del MAD: Median(abs(D1(i)-Median(D1))).
5. Cálculo de la desviación estándar:  $s=MAD/0.6745$ .
6. Asignar la mediana  $\tilde{x}_1$  como el valor inicial de la media poblacional  $\mu_0$ .
7. Asignar la desviación estándar  $s$  como el valor inicial de la incertidumbre estándar de la población  $\sigma_0$ .
8. Asignar  $j=1$

Desde  $i=1$  hasta  $n$

$$\tilde{x}_i = \begin{cases} \mu_j + k\sigma_j & \text{si } x > \mu_j + k\sigma_j \\ \mu_j - k\sigma_j & \text{si } x < \mu_j - k\sigma_j \\ x & \text{si } \mu_j + k\sigma_j < x < \mu_j - k\sigma_j \end{cases}$$

El valor de  $k=1.5\sqrt{1 - 1/n}$

Calcular  $\mu_j = \text{promedio}(\tilde{x}_i)$  hasta que  $\text{abs}\left(\frac{\mu_j}{\mu_{j-1}} - 1\right) < 0.0001$

Calcular  $\sigma_j = \sqrt{\frac{\sum(\tilde{x}_i - \mu_j)^2}{0.778(n-1)}}$  hasta que  $\text{abs}\left(\frac{\sigma_j}{\sigma_{j-1}} - 1\right) < 0.0001$





COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

Asignar  $j=j+1$  hasta  $j=50$

Si  $j>50$  no hay convergencia.

9. La media y la desviación estándar robustas serán los parámetros estimados de la población:  $\mu$ ,  $\sigma$ .

10. Los valores "outliers" están fuera del intervalo  $\mu \pm 2\sigma$ .

## 4. Resultados y discusión

### 4.1 Programa de algoritmo de Huber en HP-BASIC

En base al procedimiento del algoritmo de Huber se editó el programa en la calculadora HP PRIME, nombrándolo *MediaRobust00*. El cuerpo del programa detallado se presenta a continuación.

```
EXPORT MediaRobust00()
//By G. Delgado, UNAN-León (2016)
BEGIN
  LOCAL i,n,d,x,y,s,j,SX;
  LOCAL DT,c,smad,xs,sumx,Xm,k0,xc;
  LOCAL xm0,xs0,sm,sm2,a, H15,c1;
  LOCAL l;
  LOCAL xm:=makeMat(0,50);
  D2:={};L2:={}; D3:={}; L7:={}; L8:={};
  INPUT(n,"Tamaño datos");
  EDITLIST(D1);
  median(D1)►Xm;
  MSGBOX("MED"+Xm);
  L9:=D1;
```

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .





COLEGIO DE POSTGRADUADOS

FOR i FROM 1 TO n DO



UNAN-LEÓN

$L1(i) := ABS(D1(i) - X_m);$

END;

$median(L1) \triangleright y;$

$MSGBOX("MAD=" + y);$

$smad := y / 0.6745;$

$c := 1.5;$

$xs := smad;$

$L2(1) := X_m // mediana$

$xc := smad * c;$

$c1 := c * \sqrt{1 - 1/n};$

$xm := X_m;$

$xs := smad;$

$j := 1;$

while  $j < 50$  do

$xm0 := xm;$

$xs0 := xs;$

$sm := 0;$

$sm2 := 0;$

$xc := c1 * xs;$

for  $i$  from 1 to  $n$  do

$a := \min(xm + xc, \max(xm - xc, L9(i)));$

$sm := sm + \text{evalf}(a);$

$sm2 := sm2 + (\text{evalf}(a) - xm)^2;$

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



COLEGIO DE POSTGRADUADOS  
end;



UNAN-LEÓN

```

xm:=sm/n;

xs:=(sqrt(sm2/(0.778*(n-1)))));

if ABS(xs-xs0)/xs<0.0001 then

l:=j;

MSGBOX("Iteraciones"+l);

j:=51;

end;

j:=j+1;

end;

H15:=xm;

FOR i FROM 1 TO n DO

L7(i):=H15-2*xs;

L8(i):=H15+2*xs;

END;

D2:=L7;

D3:=L8;

if l>50 then

MSGBOX("NO CONVERGE");

else

PRINT();

PRINT("*****ESTADISTICA ROBUSTA *****");

PRINT(" μHub="+H15);

PRINT(" σHub="+xs);

```

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

```

PRINT("Iteraciones "+l);

PRINT("-----");

PRINT("Presione Enter para ver OUTLIERS");

PRINT("-----");

WAIT(0);

FOR i FROM 1 TO n DO

    IF L9(i)<H15-2*xs or L9(i)>H15+2*xs THEN

        PRINT("Outlier "+"("+i+")=> "+L9(i));

    END;

END;

end;

PRINT("-----");

PRINT("Límites:");

PRINT("Li="+ (H15-2*xs));

PRINT("Ls="+ (H15+2*xs));

PRINT("");

PRINT(" G. Delgado");

END;

```

Una vez editado el programa se puede ejecutar desde el menú usuario (*Usua*). Se selecciona el nombre del programa *MediaRobust00* y se ejecuta con Enter. Los datos de entradas son los siguientes: número o tamaño de datos (*n*) e introducción de la serie de datos  $x_1...x_n$ . Los datos de salida son los siguientes: media robusta ( $\mu_{Rob}$ ), desviación o incertidumbre estándar robusta ( $\sigma_{Rob}$ ), los valores Outliers, límite mínimo y máximo de valores permitidos. Se puede obtener el gráfico de control en Aplicaciones (*Apps*). Seleccionando la aplicación *1 Var Estadística* y pulsando

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



Enter, aparece la tabla numérica de tres columnas D1, D2 y D3. Al pulsar la tecla *Symb* aparece el menú donde se deben especificar las columnas  $D_i$  y seleccionar *line*. Con la tecla *Plot* aparecerá el gráfico de control.

#### 4.2. Aplicación del programa *MediaRobust00* en recobrados para análisis de residuos de clorpirifos en maní

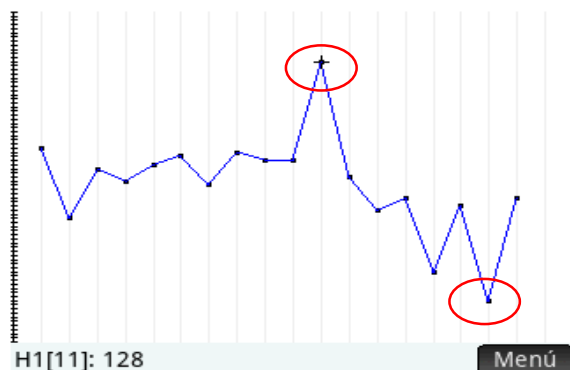
El programa *MediaRobust00* fue aplicado para identificar valores “outliers” en la estimación del factor de recobrado en análisis de residuo de clorpirifos en maní, cuyos resultados son útiles para estimar la veracidad del método de determinación de plaguicidas residuales por cromatografía de gases con detector  $\mu$ ECD. En la tabla 1 se presentan los resultados de los recobrados obtenidos en 18 días a lo largo de 3 meses y por diferentes analistas. Estos datos se introducen en el programa para obtener los valores “outliers”.

**Tabla 1.** Recobrado en porcentaje para residuos de toxafeno en maní para concentraciones 0.24  $\mu$ g/kg.

<i>Día</i>	<i>Rec%</i>	<i>Día</i>	<i>Rec%</i>
1	107	10	104
2	90	11	128
3	102	12	100
4	99	13	92
5	103	14	95
6	105	15	77
7	98	16	93
8	106	17	70
9	104	18	95



Los datos se introdujeron en la columna D1 de la aplicación *1 Var Estadística*. Para visualizar los valores dudosos se graficaron los datos pulsando la tecla *Plot*, como puede verse los puntos dentro de los círculos en la figura 1. Se observa que al menos hay 2 outliers.



**Figura 1.** Los puntos en los círculos son posibles “outliers”.

El programa se ejecutó desde el menú *Usua* y pulsando la tecla táctil *ok* y la tecla *Enter* se introduce el número de datos  $n: 15$  y se pulsa *ok*. Aparece la pantalla con la tabla de datos como se muestra en el cuadro de la figura 2.

Listas	
D1	
1	107
2	90
3	102
4	99
5	103
6	105
7	98
8	106
9	104
10	104
11	128

Editar   Ins   Elimi.   Tamaño   OK

**Figura 2.** Tabla de datos parciales introducidos en la columna D1.

Se revisan si los valores son correctos. Si es necesario corregirlos se utilizan las teclas direccionales de la calculadora para ubicarse en la celda y se introducen los valores corregidos. La ejecución del programa continúa pulsando la tecla *ok* y siguiendo las direcciones del cuadro de diálogo de la pantalla.

Los resultados finales del programa *MediaRobust00*, se presentan en el cuadro de la figura 3. Donde se pueden ver los parámetros de la población estimados ( $\mu, \sigma$ ), los 3 valores outliers y los límites del intervalo donde son aceptados los datos.

```
μHub=98.7648101567
σHub=7.90887973368
Iteraciones 10
-----
Presione Enter para ver OUTLIERS
-----
Outlier (11)=> 128
Outlier (15)=> 77
Outlier (17)=> 70
-----
Límites en %:
Li=82.9470506893 %
Ls=114.582569624 %
```

G. Delgado

**Figura 3.** Resultados mostrados en la pantalla HP PRIME

Los límites de control y la serie de datos incluyendo los “outliers” se pueden visualizar gráficamente en la aplicación *1 Var Estadística*. En esta aplicación se pulsa Enter y aparecen la tabla de datos con dos columnas adicionales D2 y D3. En D2 aparece el límite de control inferior y en D3 el límite superior.

Vista numérica 1Var estadística 19:25

	D1	D2	D3	D4
1	107	82.947051	114.58257	
2	90	82.947051	114.58257	
3	102	82.947051	114.58257	
4	99	82.947051	114.58257	
5	103	82.947051	114.58257	
6	105	82.947051	114.58257	
7	98	82.947051	114.58257	
8	106	82.947051	114.58257	
9	104	82.947051	114.58257	
10	104	82.947051	114.58257	

107

**Figura 4.** Pantalla de HP PRIME mostrando parcialmente la serie de datos y los límites de aceptación de los resultados de los recobrados.

Con la tecla *Plot* se obtiene la gráfica de la figura 5. Donde se observan claramente los puntos “outliers” y los límites de variación de los valores de recobrado de clorpirifos en mani.



**Figura 5.** Límites de variación del recobrado de clorpirifos en mani.

### 4.3. Validación del programa

Para validar el programa *MediaRobust00*, se utilizó el programa *Robstat* basado en el algoritmo de Huber, y elaborado en Excel por la Comisión de métodos analíticos de la “Royal Society of Chemistry”. Los datos se ingresaron en ambos programas y los resultados obtenidos se muestran en la tabla 2.

**Tabla 2.** Resultados de comparación del programa *MediaRobust00* con *Robstat* como referencia.

Parámetro	Programa de Prueba	Programa referencia	Desviación	Precisión
$\mu$	98.7657	98.7663363	-0.00064%	0.01%
$\sigma$	7.9050	7.8997	0.067 %	0.1%

Las desviaciones son menores que las precisiones o aproximaciones en el cálculo iterativo. Esto significa que no hay diferencias entre los dos programas. De esta manera se puede concluir que el programa *MediaRobust00*, elaborado en HP-BASIC es apto para ser utilizado en la identificación de “outliers” en las mediciones analíticas.

### 4.4. Comparación con los test de Dixon y Grubbs

Aplicando las pruebas de Dixon y de Grubbs al 95% de nivel de confianza para detectar “outliers”, se tienen los resultados de la tabla 3 y 4. Los cálculos se realizaron ordenando los datos de menor a mayor.





**Tabla 3.** Test de Dixon para detección de “outliers” al 95% de nivel de confianza.

xi	Qi	Q <sub>0.95</sub>	Valor outlier si Qi>Q <sub>0.95</sub>
x <sub>1</sub> =70 x <sub>18</sub> =128	$Q_1 = \frac{77-70}{128-70}=0.1207$ $Q_{18} = \frac{128-107}{128-70}=0.3621$	0.567	Ninguno

**Tabla 4.** Test de Grubbs(\*) para “outliers” al 95% de nivel de confianza.

xi	Gi	G <sub>0.95</sub>	Valor outlier si G>G <sub>0.95</sub>
x <sub>1</sub> =70 y x <sub>2</sub> =77 x <sub>18</sub> =128	$G_{1,2} = \frac{17427.29}{2435.83}=0.5860^*$ $G_{18} = \frac{128-98.22}{112.31716}=2.488$	0.4025 2.651	70 Y 77

\*) Prueba realizada para dos valores sucesivos.

Como puede observarse claramente, el test de Dixon es el menos sensible, el Test de Grubbs identificó los dos valores más pequeños como “outliers”. Mientras que con el método de Huber se encontraron tres outliers: el valor máximo de la serie (128) y los dos valores más pequeños (70 y 77). Lo que indica que este último método es mucho más sensible. El test de Dixon es muy útil para tamaño de serie de datos pequeñas (n≤5), el test de Grubbs se aplica para n≤10 y el test de Huber funciona para n≥6.

Con el fin de confirmar la validez del programa en HP BASIC se utilizó un programa elaborado en cálculo simbólico MAPLE 2015 denominado *MediaHuber00* (Delgado, G, 2015). Los resultados muestran que no hay diferencias entre los parámetros de la población estimados con los dos programas (desviación de 0.05% para la incertidumbre estándar y 0.0009% para la media poblacional). Si se toma la media y la desviación estándar clásicas y se elabora el gráfico de control

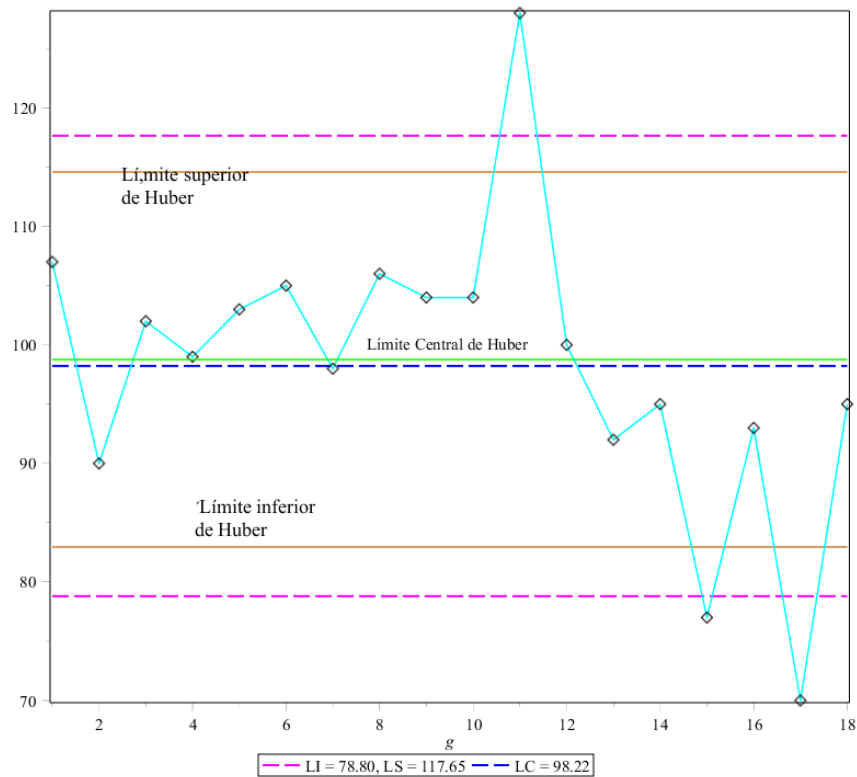
Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



con rango móvil al 95% de nivel de confianza (ver anexo, según ASTM D 6299-02, 2002 y WADSWORTH H, STEPHENS K, GOFFREY A. B., 2005), el intervalo de los límites será mayor que el del método de Huber, tal a como lo muestran las líneas punteadas de la figura 6. En la misma gráfica se observa que este intervalo se reduce aplicando la estadística robusta. Esto verifica que el método de Huber es más sensible para la identificación de “outliers”. Es importante destacar que con este gráfico también se identificaron los tres puntos “outliers” al 95% de nivel de confianza. Son los puntos que están fuera de los límites de control.



**Figura 6.** Puntos representando los datos de recobrados de clorpirifos en maní y gráficos de control con parámetros de la población estimados clásicamente (líneas





COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

punteadas) y por el método de Huber. Gráfica obtenida en MAPLE con el programa *MediaHuber00*<sup>[15]</sup>.

## 5. Conclusiones

El diseño y elaboración del programa basado en el algoritmo de Huber en lenguaje HP-BASIC (mezcla de QBASIC y Pascal) utilizando la calculadora de última generación HP-PRIME y validado con el programa de referencia del Comité de Métodos Analíticos de la Royal Society of Chemistry, fue aplicado a la identificación de valores “outliers” en la determinación de residuos de clorpirifos en maní por cromatografía de gases con detector  $\mu$ ECD. El algoritmo se fundamenta en la estadística robusta que no depende ni de valores extremos ni del tamaño de la muestra. Este método es mucho más sensible que los métodos clásicos, test de Grubbs y test de Dixon. El programa editado en la calculadora programable HP PRIME es útil para la inmediatez del control de calidad en análisis de rutina, ya que en la mayoría de los casos se necesita de medios de cálculos rápidos para evaluar parámetros de calidad de los procesos analíticos.

Por otro lado, desde el punto de vista académico este trabajo es muy útil por la facilidad de contar con una herramienta de cálculo portátil que puede ser aplicado en los cursos de quimiometría o de estadística aplicada. Lo importante a señalar es que el emulador o calculadora virtual HP PRIME se puede descargar de internet Wadsworth H, Stephens K, Goffrey A. B., (2005) .

## Reconocimiento

Vaya el agradecimiento de los autores al Laboratorio SER AGRO, S.A. por permitirnos la difusión del presente estudio, dentro del contexto de la mejora continua del Sistema de Gestión de la Calidad ISO/IEC 17025.

## Referencias

AMC-RSC, (1989), "Robust Statistics", *Analyst*, Vol. 14, 1693-1697.

---

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .



COLEGIO DE POSTGRADUADOS



UNAN-LEÓN

AMC Technical Brief (2004), Royal Soc. of Chem. No. 6.

ASTM D 6299-02, (2002), "Applying Statistical Quality Techniques" Norma Internacional, EE.UU.

Codex Alimentarius, (2000), "Residuos de plaguicidas presentes en alimentos y piensos", 32ª Reunión, Países Bajos.

Delgado G y Hernández N., (2014) "Evaluación de la homogeneidad de la muestra y estudio intralaboratorio de la precisión intermedia en la determinación de aflatoxinas en maní", Universitas UNAN-León, Vol 5, 1, 11-25.

Delgado, G. (2015), Programa elaborado en MAPLE 2015, Universidad Nacional Autónoma de Nicaragua, León (UNAN-León).

Dixon W.J, (1950), Ann. Math. Statistics. Vol. 21, 488.  
<https://doi.org/10.1214/aoms/1177729747>

European Food Safety Authority, (2014), EFSA Journal, Vol 12, 12, 3942.

Feinberg M., (1996), "La Validation des Méthodes d'Analyse", Masson, París, pag. 383.

Grubbs F.E., (1969), Tecnometrics, 1969, 11, 1.  
<https://doi.org/10.1080/00401706.1969.10490657>

a) ISO 5721-2, (1994) "Accuracy of measurements methods and results", Norma Internacional, Ginebra, Suiza;

b) ISO 5721-5, (1994) "Accuracy of measurements methods and results", Norma Internacional, Ginebra, Suiza.

## Anexo

### Etapas para elaborar una carta de control de rango móvil

Los límites de control se calculan en base a  $2\sigma$ . El algoritmo consta de las siguientes etapas:

1. Cálculo del promedio de la serie de valores  $x_1, x_2, \dots, x_n$ .

$$\bar{x} = \frac{\sum x_i}{n}$$

2. Calculo del promedio de rango móvil:

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .

$$\overline{MR} = \frac{\sum_1^{n-1} |x_{i+1} - x_i|}{n - 1}$$

3. Cálculo de los límites de control:

$$Linf = \bar{x} - \frac{2 \overline{MR}}{1.128}$$

$$Lsup = \bar{x} + \frac{2 \overline{MR}}{1.128}$$

---

Copyright (c) 2018 Revista Iberoamericana de Bioeconomía y Cambio Climático.



Este trabajo de la Revista Iberoamericana de Bioeconomía y Cambio Climático del Centro de Investigación en ciencias agrarias y economía aplicada de la UNAN-León/ COLPOS México está licenciado bajo una Licencia Internacional Creative Commons 4.0 Atribución-NoComercial-CompartirIgual .