Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 5 N°1 Enero-Junio de 2016 • ISSN 0121-747X / ISSN-e 2357-5749 • Artículo Investigación • Páginas 111 a 123
DOI: https://doi.org/10.15446/rev.fac.cienc.v5n1.55358

DETECCIÓN DE DATOS EXTREMOS Y DE MULTICOLINEALIDAD EN MODELOS NO LINEALES: UNA INTERFAZ GRÁFICA EN R[a]

DETECTION OF OUTLIERS AND MULTICOLLINEARITY IN NONLINEAR MODELS: A GRAPHICAL INTERFACE IN R

JUAN PALACIO[b], ELKIN CASTAÑO V.[c] [d]

Recibido 21-01-2016, aceptado 31-05-2016, versión final 08-06-2016.
Artículo Investigación

RESUMEN: El análisis de regresión es una herramienta ampliamente usada en el trabajo estadístico aplicado. En este análisis, la presencia de datos extremos o la existencia de multicolinealidad pueden introducir serias distorsiones en la estimación de parámetros y la inferencia estadística; dichos efectos han sido estudiados ampliamente en la literatura. En este artículo se presenta una herramienta construida bajo la librería shiny del paquete computacional R con el objeto de detectar este tipo de problemas en modelos de regresión no lineal, cuando se emplea estimación por mínimos cuadrados no lineales. La interfaz gráfica presentada permite especificar el modelo de regresión no lineal, realizar su estimación por mínimos cuadrados no lineales, y diagnosticar la presencia de datos extremos, o la existencia y severidad de problemas de multicolinealidad.

PALABRAS CLAVE: Datos extremos, interfaz gráfica, mínimos cuadrados no lineales, multicolinealidad, regresión no lineal.

ABSTRACT: Regression analysis is a widely used tool in the statistical work applied. In this analysis, the presence of extreme data or the existence of multicollinearity can introduce serious distortions in parameter estimation and statistical inference; these effects have been widely studied in the literature. This article describes a tool built under the shiny R library software package in order to detect such problems in nonlinear regression models, when estimation is used for nonlinear least squares is presented. The graphical interface presented allows you to specify the nonlinear regression model, make its estimate for nonlinear least squares, and diagnosing the presence of extreme data, or the existence and severity of multicollinearity problems.

KEYWORDS: Collinearity, graphical interface, nonlinear least squares, nonlinear regression, outliers.

[a]Palacio, J., & Castaño, E. (2016). Detección de datos extremos y de multicolinealidad en modelos no lineales: una interfaz gráfica en R. *Revista de la Facultad de Ciencias*, 5 (**1**), 111–123. DOI: https://doi.org/10.15446/rev.fac.cienc.v5n1.55358

[b]Estadístico, M. Sc. Ciencias-Estadística, Universidad Nacional de Colombia, Sede Medellín.
estpalac@unal.edu.co
[c]Profesor Asociado, Universidad Nacional de Colombia, Sede Medellín
[d]Profesor Titular, Universidad de Antioquia.

1. INTRODUCCIÓN

Los modelos de regresión lineal se usan frecuentemente en el análisis de datos de diversas áreas del conocimiento. En este tipo de estudios, es habitual encontrar observaciones con una influencia desproporcionada en los resultados del ajuste de dichos modelos, lo cual puede causar graves problemas en las estimaciones y la inferencia; también se pueden encontrar situaciones en las cuales exista poca variabilidad observada en las variables regresoras o relaciones de dependencia lineal entre ellas, lo cual puede conducir a problemas de multicolinealidad que frecuentemente producen una inflación artificial en la varianza de los coeficientes estimados del modelo. Belsley *et al.* (1980) proponen una serie de indicadores que permiten determinar la existencia de observaciones atípicas o problemas de multicolinealidad en un modelo lineal.

Por otra parte, uno de los supuestos asumidos en el modelo de regresión lineal es que la verdadera relación existente entre la variable dependiente y las variables independientes es de tipo lineal. Sin embargo, en muchas aplicaciones, asumir linealidad es bastante restrictivo, y la relación de dependencia podría ser mejor explicada por una relación de tipo no lineal (ver Novales (2012)). Por tanto, en algunas situaciones particulares, resulta una práctica útil y adecuada la implementación de modelos que consideren relaciones de dependencia no lineales.

Teniendo en cuenta que al estimar un modelo de regresión no lineal mediante mínimos cuadrados no lineales, el procedimiento obtiene una aproximación lineal en la que se estiman por mínimos cuadrados ordinarios dichos parámetros, pero tomando como insumos una nueva variable respuesta y una nueva matriz de diseño, los problemas causados por datos extremos y/o multicolinealidad se presentan, en este caso, al estimar por mínimos cuadrados ordinarios esta forma lineal y por consiguiente, dichos problemas deben tenerse en cuenta y deben ser evaluados tomando como insumos esta nueva respuesta transformada y esta nueva matriz de diseño. En consecuencia, es posible extender, del modelo lineal al no lineal, las medidas de diagnóstico, tanto de datos extremos como de multicolinealidad (Palacio, 2016).

En este trabajo se presenta una interfaz gráfica, construida mediante la librería shiny de R, que calcula y muestra algunos de los indicadores propuestos por Belsley *et al.* (1980) para diagnosticar la presencia de observaciones extremas o problemas de multicolinealidad en la estimación por mínimos cuadrados ordinarios de los parámetros de un modelo no lineal. La aplicación, adicionalmente, arroja algunos gráficos de diagnóstico que ayudan a complementar los análisis. Esta interfaz recibe como insumos de parte del usuario un conjunto de datos, un modelo no lineal y un vector de valores iniciales para los parámetros a estimar. La aplicación permite seleccionar el tipo de diagnóstico que se quiere realizar sobre el modelo y sus datos, y adicionalmente, cuando se evalúa el problema de multicolinealidad, permite cuantificar el potencial impacto negativo que ésta tiene sobre la precisión

de cada uno de los parámetros estimados.

El presente trabajo incluye en su sección 2 una descripción detallada de la forma en que se obtiene la aproximación lineal del modelo no lineal, haciendo hincapié en la obtención de los insumos de dicha aproximación lineal a partir del modelo original; las secciones 3 y 4 presentan, respectivamente, las medidas utilizadas para detectar observaciones extremas o multicolinealidad en la aproximación lineal; en la sección 5 se hace una completa descripción de la interfaz gráfica construida en R; y, finalmente, en la sección 6 se presentan las conclusiones del trabajo.

2. APROXIMACIÓN LINEAL DEL MODELO NO LINEAL

Del modelo de regresión lineal bajo los supuestos tradicionales

$$y = X\beta + \epsilon \tag{1}$$

el estimador de mínimos cuadrados ordinarios para el parámetro (o vector de parámetros) β está dado por la expresión matricial

$$\beta = (X^T X)^{-1} X^T y \tag{2}$$

donde X es una matriz de dimensión $n \times (k+1) = n \times p$, que tiene en sus columnas las n observaciones de cada una de las $k+1 = p$ variables regresoras o predictoras, y y es un vector columna, formado por las n observaciones de la variable dependiente.

Considerando el caso en que la relación de dependencia es del tipo $y = f(x;\beta) + \epsilon$ donde $f(x;\beta)$ es una función no lineal en las componentes del vector β, el procedimiento de mínimos cuadrados no lineales se enfoca en resolver el problema de optimización

$$min_\beta \sum_{i=1}^{n} (y_i - f(x_i;\beta))^2 = min_\beta \sum_{i=1}^{n} \epsilon_i^2 \tag{3}$$

lo cual, al aplicar reglas de derivación, se traduce en resolver el sistema de ecuaciones

$$\left(\frac{\partial f(x_i,\beta)}{\partial \beta}\right)^T y = \left(\frac{\partial f(x_i,\beta)}{\partial \beta}\right)^T f(X;\beta) \tag{4}$$

donde la matriz gradiente $\left(\frac{\partial f(x_i,\beta)}{\partial \beta_j}\right)$ tiene n filas y k columnas (determinadas por cada una de las variables predictoras sobre las que se debe derivar parcialmente la función no lineal evaluada en la observación i), mientras que $y = [y_1, y_2, \ldots, y_n]^T$ y $f(X;\beta) = [f(x_1;\beta), f(x_2;\beta), \ldots, f(x_n;\beta)]^T$ son vectores columna de dimensión n.

En una solución por métodos numéricos, siempre es complicado saber a ciencia cierta si el tipo de solución encontrada es la adecuada (Novales, 2012). Una alternativa propuesta para afrontar esta dificultad consiste en estimar la aproximación lineal del modelo alrededor de una estimación inicial.

Usando la expansión de Taylor de primer orden para $f(x_i;\beta)$ alrededor de una estimación inicial β^*, se tiene

$$y_i \approx f(x_i;\beta^*) + \left(\frac{\partial f(x_i;\beta}{\partial \beta}\right)_{\beta=\beta^*} (\beta - \beta^*) + \epsilon_i \tag{5}$$

Haciendo el cambio de variable $y_i^* = y_i - f(x_i;\beta^*) + \left(\frac{\partial f(x_i;\beta}{\partial \beta}\right)_{\beta=\beta^*} \beta^*$, y generando simultáneamente datos para las k variables definidas en el gradiente $\left(\frac{\partial f(x_i;\beta}{\partial \beta}\right)_{\beta=\beta^*}$, se puede estimar el modelo lineal dado por

$$y_i^* \approx \left(\frac{\partial f(x_i;\beta}{\partial \beta}\right)_{\beta=\beta^*} \beta + \epsilon_i \tag{6}$$

por mínimos cuadrados ordinarios.

De esta forma, dado un valor inicial β^* para el estimador, se puede construir la variable y_i^* , así como las k variables que componen el valor del gradiente de la función $f(x_i;\beta)$ en el punto $\beta = \beta^*$. Las realizaciones de estas nuevas variables están en función de las observaciones muestrales de y_i y x_i y del vector de valores iniciales β^*. A continuación, se estima a través de mínimos cuadrados ordinarios el modelo lineal que tiene como variable respuesta a y_i^* y como variables explicativas las componentes de la matriz gradiente, con lo cual se obtiene una nueva estimación para β (Seber & Wild, 2003).

El nuevo estimador obtenido por mínimos cuadrados para β, denotado como $\widetilde{\beta}$ sería:

$$\widetilde{\beta} = \beta^* + \left[\left(\frac{\partial f(x_i;\beta)}{\partial \beta}\right)^T \left(\frac{\partial f(x_i;\beta)}{\partial \beta}\right)\right]^{-1}_{\beta=\beta^*} \left(\frac{\partial f(x_i;\beta)}{\partial \beta}\right)^T \xi_i^* \tag{7}$$

donde $\xi_i^* = y_i^* - f(x_i;\beta^*)$ son los nuevos residuales obtenidos con la estimación inicial β^*. Esta expresión proporciona la nueva estimación $\widetilde{\beta}$ a partir de β^*. Con esta nueva estimación $\widetilde{\beta}$ como valor inicial, se repite el proceso de manera idéntica. Finalmente, tras repetir este proceso varias veces, se obtiene la estimación definitiva para β, denotada $\widehat{\beta}$.

A partir de $\widehat{\beta}$ y de las observaciones muestrales para y_i y x_i, es posible construir el modelo lineal que tiene como variable respuesta a y_i^* y como variables explicativas a las componentes de la matriz gradiente $\left(\frac{\partial f(x_i;\beta)}{\partial \beta}\right)_{\beta=\widehat{\beta}}$. Este modelo lineal tiene una importancia trascendental puesto que es sobre éste sobre quien se realizarán todos los diagnósticos sobre presencia de valores extremos y

problemas de multicolinealidad (Palacio, 2016).

Con el fin de simplificar la notación, el anterior modelo se puede expresar como

$$y_i^* \approx W_i\beta + \epsilon_i; i = 1, 2, 3, \ldots, n \tag{8}$$

donde

$$W_i = \left(\frac{\partial f(x_i;\beta)}{\partial \beta}\right)_{\beta=\widehat{\beta}} \tag{9}$$

3. DETECCIÓN DE DATOS EXTREMOS EN EL MODELO NO LINEAL

Tomando como insumo la aproximación lineal $y_i^* \approx W_i\beta + \epsilon_i$ encontrada para el modelo no lineal $y_i = f(x_i;\beta) + \epsilon_i$, los datos extremos con incidencia en la estimación de un modelo no lineal pueden ser detectados usando las medidas propuestas por Belsley *et al.* (1980) para el modelo lineal, entre las que se encuentran la diagonal de la matriz de proyección, los residuales estandarizados y estudentizados, los DFBETAs, los DFFITs y el COVRATIO.
Tales medidas son calculadas tomando como insumos la matriz de pseudo datos W y la nueva variable respuesta y_i^*.

3.1. Diagonal Matriz de Proyección

Denotados como h_{ii} los elementos de la diagonal de la matriz de proyección de mínimos cuadrados, $H = W(W^TW)^{-1}W^T$, tienen una importancia sustancial en la determinación de los valores predichos, puesto que $\widehat{y}^* = W\widehat{\beta} = Hy^*$.

Los h_{ii} calculados como $h_{ii} = w_i(W^TW)^{-1}w_i^T$, dan indicios sobre una observación extrema cuando su valor calculado es mayor a $2p/n$ (Belsley *et al.*, 1980).

3.2. Residuales

Los residuales se utilizan para detectar aquellos datos sospechosos que afectan indebidamente los resultados de la regresión. Se consideran los residuales estandarizados y estudentizados, los cuales se calculan, respectivamente, como

$$\epsilon_{si} = \frac{\epsilon_i}{s\sqrt{1-h_{ii}}} \tag{10}$$

$$\epsilon_i^* = \frac{\epsilon_i}{s(i)\sqrt{1-h_{ii}}} \tag{11}$$

donde s y $s(i)$ representan a la desviación estándar en los casos donde se tienen en cuenta todos los datos y donde se omite la fila i, respectivamente.

Aquellas observaciones, cuyo residual asociado tenga una magnitud mayor a 2 deben recibir una atención especial (Belsley *et al.*, 1980)

3.3. DFBETAs

Esta medida analiza el cambio producido en cada coeficiente al suprimir la i-ésima fila de los datos. Teniendo en cuenta que β representa el vector de parámetros estimados calculado con todos los datos y $\beta(i)$ representa al mismo vector, omitiendo en la estimación la fila i; el DFBETAs está dado por

$$DFBETAs_{ij} = \frac{c_{ji}}{\sqrt{\sum_{k=1}^{n} c_{jk}^2}} \frac{\epsilon_i}{s(i)(1-h_{ii})} \tag{12}$$

donde

$$C = (W^T W)^{-1} W^T \tag{13}$$

Belsley *et al.* (1980), sugieren prestar especial atención a aquellas observaciones con DFBETAS asociados cuyo valor absoluto sea mayor a $\frac{2}{\sqrt{n}}$

3.4. DFFITs

Una medida que permite comprender mejor los efectos en la predicción cuando se elimina una observación es el DFFITs, calculado como

$$DFFITs_i = \sqrt{\left[\frac{h_{ii}}{1-h_{ii}}\right]} \frac{\epsilon_i}{s(i)\sqrt{1-h_{ii}}} \tag{14}$$

Belsley *et al.* (1980), sugieren prestar especial atención a aquellas observaciones con DFFITs asociado cuyo valor absoluto sea mayor a $2\sqrt{p/n}$.

3.5. COVRATIO

Esta medida compara, a través del cociente entre sus determinantes, la matriz de covarianza calculada con todos los datos y la matriz de covarianza que resulta al eliminar la i-ésima fila. Valores del cociente de determinantes cercanos a la unidad se pueden interpretar como señal de que las dos matrices de covarianza están cerca, o que la matriz de covarianza original es insensible a la eliminación de la fila i (Belsley *et al.*, 1980). El hecho de que el estimador s^2 de σ^2 también cambia con la eliminación de la i-ésima observación se incorpora comparando las matrices $s^2(W^T W)^{-1}$ y

Tabla 1: Índices de condición y proporciones de varianza

Valor Singular Asociado	$\text{var}(\widehat{\beta_1})$	$\text{var}(\widehat{\beta_2})$	...	$\text{var}(\widehat{\beta_p})$
μ_1	π_{11}	π_{12}	...	π_{1p}
μ_2	π_{21}	π_{22}	...	π_{2p}
...	...	...	...	...
μ_p	π_{p1}	π_{p2}	...	π_{pp}

$s(i)^2(W(i)^TW(i))^{-1}$, mediante la razón de sus determinantes. El COVRATIO se puede expresar como

$$COVRATIO = \frac{1}{\left[\frac{n-p-1}{n-p} + \frac{\epsilon_i^{*2}}{n-p}\right]^p (1-h_{ii})} \tag{15}$$

Belsley *et al.* (1980), proponen investigar aquellos puntos con $|COVRATIO-1|$ cercanos o mayores a $3p/n$.

4. DETECCIÓN DE PROBLEMAS DE MULTICOLINEALIDAD EN EL MODELO NO LINEAL

Tomando como insumo la aproximación lineal $y_i^* \approx W\beta + \epsilon_i$ encontrada para el modelo no lineal $y_i = f(x_i;\beta) + \epsilon_i$, el análisis sobre la existencia de problemas de multicolinealidad se enfoca en la matriz W, calculando los índices de condición de ésta y la proporción de varianza de cada coeficiente estimado que es debida a cada uno de estos índices.

Esta información es convenientemente presentada con la estructura sugerida en la Tabla 1.
Los índices de condición considerados grandes (mayores a 30) identifican el número de dependencias aproximadas existentes entre las columnas de la matriz de datos W. Por otra parte, la determinación de proporciones grandes en la descomposición de varianza (mayores a 0.5) asociadas con un alto índice de condición identifican aquellas variables que están involucradas en las dependencias correspondientes, y la magnitud de estas proporciones en conjunto con un alto índice de condición proporciona una medida del grado en que la correspondiente estimación de regresión ha sido afectada por la presencia de multicolinealidad (Belsley *et al.*, 1980).

5. INTERFAZ GRÁFICA EN R

Los datos DNase, almacenados en la librería datasets de R, consisten en 15 observaciones de la densidad óptica (density) obtenidas mediante una prueba ELISA, utilizando diferentes niveles de

concentración (conc) de la proteína DNase. Se trata de explicar la densidad óptica observada a partir de la concentración de la proteína, utilizando un modelo no lineal dado por:

$$density = \frac{Asym}{1 + exp\left(\frac{xmid - log(conc)}{scal}\right)} \tag{16}$$

Donde $Asym$, $xmid$ y $scal$ son parámetros desconocidos.

Los procedimientos descritos para diagnosticar presencia de datos extremos y/o problemas de multicolinealidad en este modelo de regresión no lineal son implementados en una interfaz gráfica creada con la librería *shiny* de R.

Shiny funciona haciendo un llamado al directorio de trabajo del R, en el cual deben estar almacenados, dentro de una carpeta titulada con el nombre asignado a la aplicación, ciertos archivos que incluyen, entre otros, las tablas de datos, los programas con que se calculan los resultados y se construyen los gráficos mostrados en las salidas de la aplicación, además de otros archivos que permiten personalizar el diseño gráfico de la aplicación y las tablas e imágenes mostradas en ella.
La aplicación debe ser llamada desde la consola del R a través del comando
runApp("NombreAplicación").
Una vez ejecutada esta línea, se abre una página de navegación en la cual se visualiza la aplicación en su estado inicial, tal como se muestra en la Figura 1.
Esta aplicación consta de dos pestañas llamadas "Ingresar Modelo Nuev" y "Modelos Implementados", que cumplen funciones diferentes. En la pestaña "Ingresar Modelo Nuevo", que está seleccionada por defecto, la aplicación permite ingresar una nueva tabla de datos y especificarle un modelo no lineal junto con valores iniciales para cada uno de los parámetros a estimar.
La tabla de datos se lee con el botón "Seleccionar archivo", el cual está en la parte superior de la pantalla gris que aparece en el costado izquierdo de la ventana. Al leer la tabla de datos, en este caso llamada "DatosPrueba.txt", en la pantalla aparecerá un encabezado con las primeras seis filas de la tabla, algunas medidas descriptivas para cada variable y algunos gráficos que resumen, individualmente y por pares, el comportamiento de las variables. La forma como se presentan estos resultados se muestra en la Figura 2.
En el mismo recuadro donde se selecciona la tabla de datos, aparecen dos espacios en blanco donde se deben especificar, respectivamente, el modelo no lineal y una lista con valores iniciales para los parámetros que se deben estimar. Esto debe hacerse tal como se muestra en la Figura 2.

Una vez especificada adecuadamente la tabla de datos y el modelo, se oprime el botón "Calcular" que aparece en la parte inferior del recuadro.

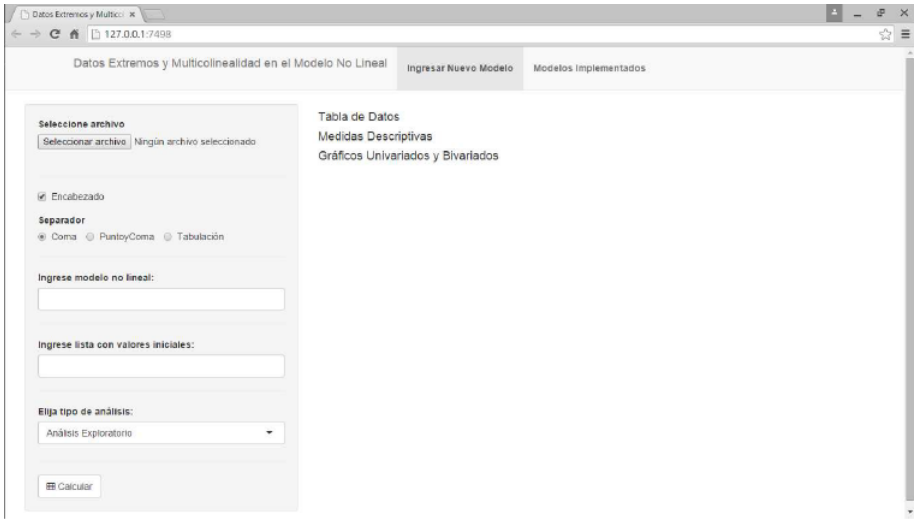

Figura 1: Aplicación en Blanco

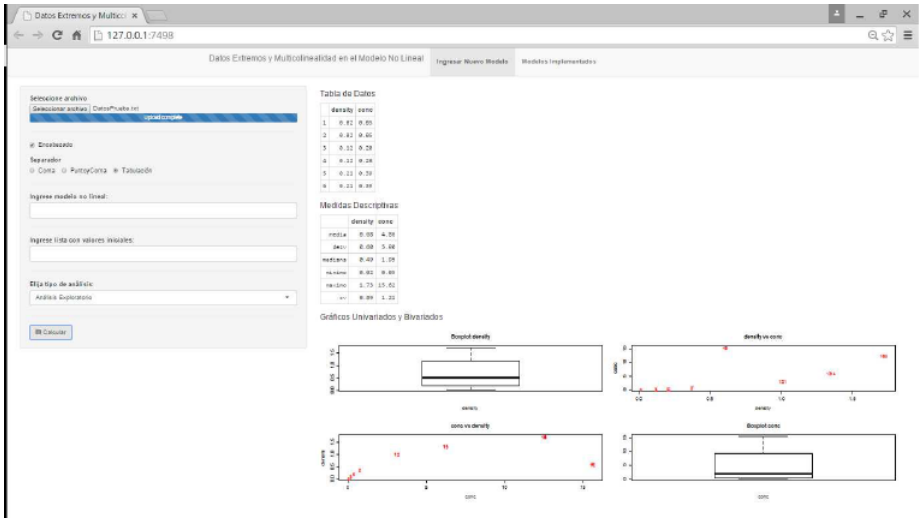

Figura 2: Resultados Descriptivos

Al oprimir este botón, se calculan todas las medidas necesarias para diagnosticar presencia de datos extremos o problemas de multicolinealidad.

Ahora bien, en el menú desplegable que tiene como título "Elija tipo de análisis", se puede seleccionar alguna de las siguientes cuatro opciones: Análisis Exploratorio, Estim. No Lineal, Detección de Datos Extremos o Detección de Multicolinealidad.

Figura 3: Especificación de Parámetros

La opción "Análisis Exploratorio", que viene seleccionada por defecto, muestra el encabezado de la tabla de datos, un breve análisis descriptivo univariado y los diagramas de dispersión para cada par de variables estudiadas. La forma como aparecen estos resultados en la aplicación para la tabla de datos "DatosPrueba.txt", se muestra en la Figura 2.

El tipo de análisis "Estim. No Lineal" presenta un resumen detallado de las estimaciones por mínimos cuadrados no lineales obtenidas para el modelo no lineal especificado. También presenta el encabezado de una tabla que tiene como columnas la variable respuesta transformada y las variables pseudo regresoras obtenidas a través del proceso de linealización del modelo no lineal. Finalmente muestra el resumen del modelo lineal estimado con esta tabla, que debe coincidir plenamente con los resultados del modelo no lineal. La forma como se muestran estos resultados para la tabla de datos "DatosPrueba.txt" y el modelo no lineal $density \sim Asym/(1+exp(xmid-log(conc))/scal)$ se presenta en la Figura 4.

Al seleccionar como tipo de análisis la opción "Detección de Datos Extremos", se presenta una tabla con aquellas observaciones cuyo DFBETAs, DFFITs, Razón de Covarianza, Distancia de Cook (definida en (Cook, 1977)) o h_{ii} supere el límite establecido por Belsley *et al.* (1980) para ser considerada extrema. Adicionalmente se presentan algunos gráficos de diagnóstico entre los que se incluyen los gráficos de regresión parcial. Las Figuras 5 y 6 muestran la forma en que se presentan estos resultados en la aplicación para el ejemplo estudiado.

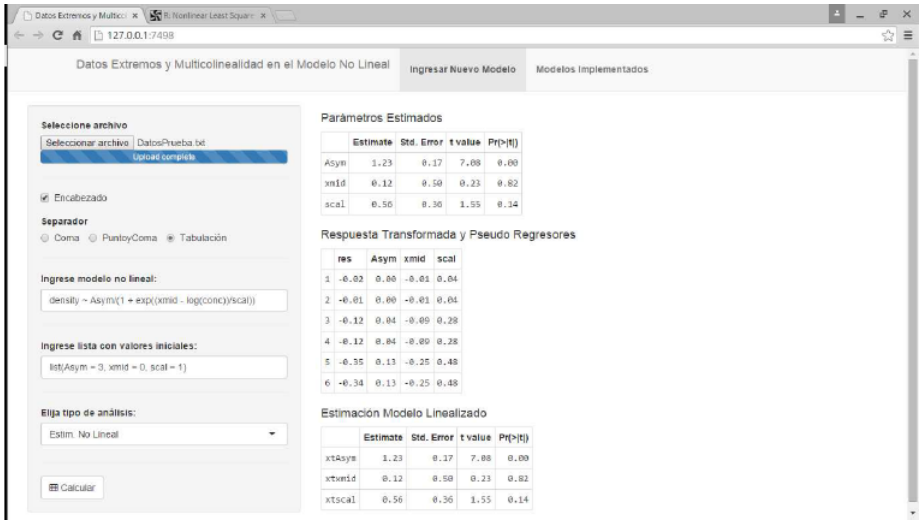

Figura 4: Modelo Estimado

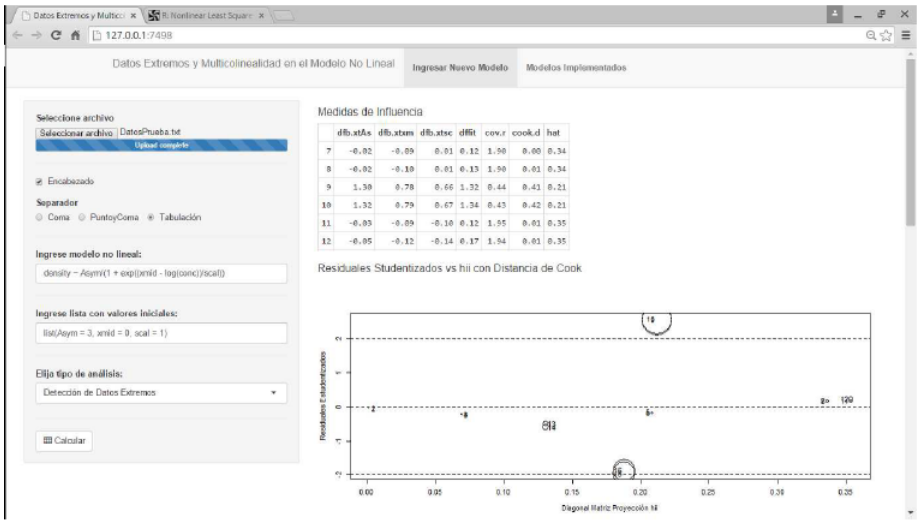

Figura 5: Medidas de Influencia

Según las medidas de referencia, las observaciones $7, 8, 9, 10, 11$ y 12 presentan un comportamiento diferente con respecto al presentado por la población general.

Finalmente, al seleccionar como tipo de análisis la opción "Detección de Multicolinealidad", se presenta la matriz de correlación entre las variables pseudo regresoras y una tabla donde se muestran los índices de condición y la proporción de varianza de cada coeficiente asociada a cada uno de los índices. La forma como se presenta esto se muestra en la Figura 7.

No se observa ningún índice de condición llamativamente grande, lo cual no permite evidenciar la presencia de problemas de multicolinealidad en los datos.

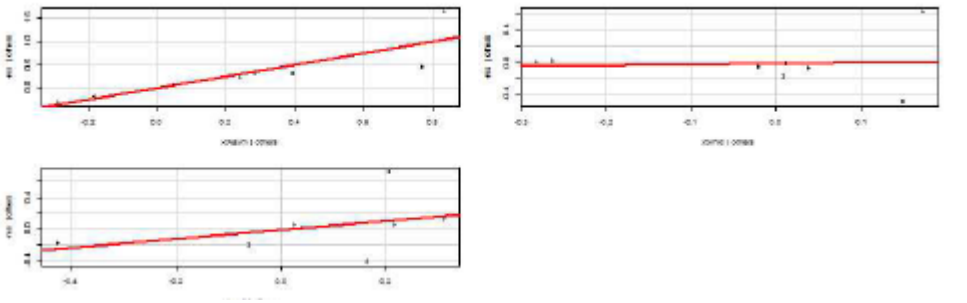

Figura 6: Gráfico de Regresión Parcial

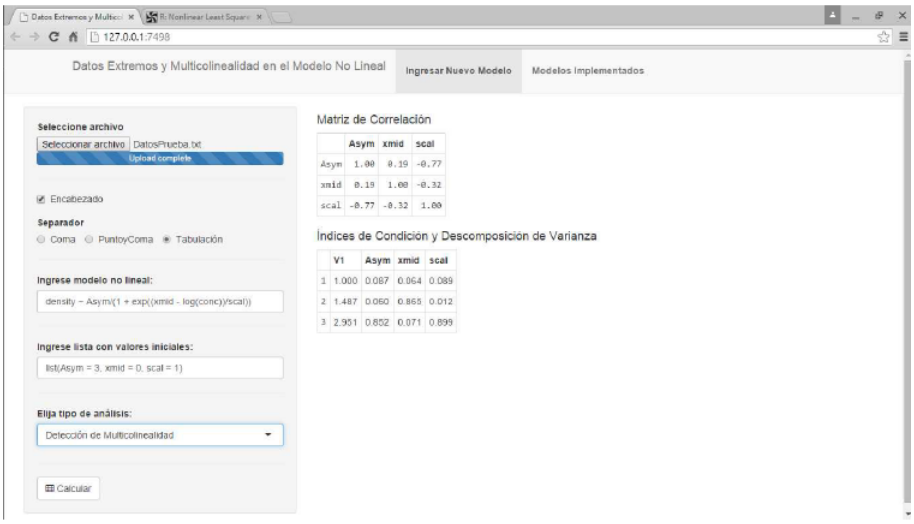

Figura 7: Diagnóstico Multicolinealidad

6. CONCLUSIONES

Los indicadores utilizados por Belsley *et al.* (1980) para diagnosticar presencia de datos influyentes o atípicos en un modelo lineal están basados tanto en la variable respuesta como en el conjunto de variables regresoras. Al utilizar estos mismos indicadores sobre la variable respuesta transformada y^* y las variables pseudo regresoras utilizadas en el modelo linealizado, estos indicadores pasan a depender, adicionalmente, de la forma no lineal del modelo, al ser W, la matriz de pseudo diseño, función de las derivadas parciales de ésta. Adicional a esto, al ser la fila i de la matriz W función de la fila i de X, una observación (o fila) extrema en el conjunto de datos originales, podría también serlo en el conjunto de datos transformados.

La metodología descrita por Belsley *et al.* (1980) para detectar problemas de multicolinealidad en la estimación de un modelo lineal, se basa únicamente en las propiedades numéricas de la matriz de diseño X. Análogamente, al aplicar esta metodología en el caso del modelo no lineal, luego de

realizar la respectiva aproximación lineal, el análisis debe centrarse en las propiedades de la matriz pseudo regresora W . De esta forma, al estar W estrechamente ligada con la forma funcional del modelo no lineal a través de sus derivadas parciales, la metodología termina basándose también en la forma como se relacionan las variables regresoras originales con la variable respuesta de interés.

En los modelos de regresión no lineal, frecuentemente, el tipo de diagnósticos realizados está enfocado principalmente en pruebas de normalidad, homogeneidad de varianza, significancia de parámetros estimados, etc; haciendo que los problemas ocasionados por la presencia de datos extremos o multicolinealidad reciba, en algunos casos, un trato menos exhaustivo. La implementación de una interfaz gráfica que, a partir de un modelo no lineal y una tabla de datos, calcula automáticamente indicadores numéricos y gráficos para realizar un completo diagnóstico sobre presencia de datos extremos y problemas de multicolinealidad (se evalúo adicionalmente el impacto negativo de este problema sobre la precisión de los coeficientes estimados), es un gran aporte para los analistas de regresión, quienes mediante esta herramienta pueden hacer visibles algunos problemas que pueden estar condicionando la precisión de los resultados finales y, por consiguiente, las conclusiones que de ellos se hacen.

El procedimiento utilizado en este artículo para detectar datos atípicos o problemas de multicolinealidad está basado únicamente en estimación mediante mínimos cuadrados no lineales; otros algoritmos de estimación como máxima verosimilitud o métodos de estimación robusta pueden ser implementados y comparados en trabajos futuros.

Referencias

Belsley, Kuh and Welsch (1980). Regression Diagnostics. *Wiley Inter- Science*.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

Novales (2012). Econometría. *McGraw-Hill*.

Palacio, J. E. (2016). Detección de Datos Influenciales y Multicolinealidad en el Modelo No Lineal *(Tesis de maestría)*. Universidad Nacional de Colombia, Medellín.

Seber, G. A. F. and Wild, C. J. (2003). Nonlinear Regression. *Wiley Inter- Science*.