كشف ومعالجة القيم المتطرفة بالطريقة الحصينة ومقارنتها بطرق أخرى

أسوان محمد طيب النعيمي

كلية الإدارة والاقتصاد

جامعة الموصل

الملخص

إن وجود القيم المتطرفة يؤدي إلى إرباك كبير في تحليل البيانات في حالة استخدام الطرق التقليدية للتقدير ومن هذه الطرق طريقة المربعات الصغرى. لذلك تم اللجوء إلى تحليل البيانات باستبعاد القيم المتطرفة ومقارنتها مع تحليل البيانات كاملة. لقد تم معالجة القيم المتطرفة في البيانات بأسلوبين الأول هو زيادة حجم العينة لأعوام أخرى بحيث تصبح هذه العينة ملائمة للتقدير كما في تحليل البيانات الأصلية بعد استبعاد القيم المتطرفة. أما الأسلوب الثاني فهو استخدام الطرق الحصينة في الكشف والتقدير ومقارنة نتائج الأسلوبين مع بعضهما من حيث الدقة. ولقد تم تطبيق ما سبق على بيانات فجوات القدرة الكهربائية Mw لكافة القطاعات في محافظة نينوى للفترة (٢٠٠٤-٢٠٠٧) واستنتجنا أهمية الأسلوبين ويعتمد كل منهما حسب توفير البيانات، وإن مقياس الدقة المعتمد أكد أن حجم العينة الكبير يعطي أدق النتائج للدراسة لكون العينة تمثل المجتمع خير تمثيل.

Discovering and Treating the Outliers Values by the Inaccessible Method and Their Comparison to Other Methods

Abstract

The existence of extreme values leads to a great confusion in analyzing data in case of using the traditional methods for evaluation and one of these methods is the least square method. Thus, it was resorted to the analyzing of data by excluding the extreme values and their comparison with the complete data analysis. The extreme values were treated in data by two ways: the first is increasing the size of the sample for other years where this sample becomes suitable for evaluation as in the original data analysis after excluding the extreme values. While the second way is to use the inaccessible methods in discovering, evaluating and comparing the two ways with each other concerning the accuracy.

All this was applied on the data of electric power gaps M_w for all the sectors in Nineveh governorate for the period (2004 – 2007). We concluded the importance of the two ways and each one depends on saving data, and the accuracy meter assured that the size of the large sample gives the most accurate results for the study since the sample represents the society in the best way.

المقدمة

لقد هيمنت طريقة المربعات الصغرى الاعتيادية على تقدير معلمات الانحدار الخطي ولفترة طويلة من الزمن لما لهذه الطريقة من مزايا جيدة تميزت بها مقدراتها. ولكن الباحثون وجدوا أن هذه الطريقة تكون غير كفوءة في حالة خرق واحد أو أكثر من فروض التحليل، وذلك عندما تتوزع الأخطاء العشوائية توزيعاً غير طبيعياً نتيجة وجود القيم المتطرفة outliers والتي تعرف بكونها مشاهدات تنحرف بشكل ملحوظ عن بقية المشاهدات، وقد يعزى سبب تطرفها إلى وجود أخطاء في القراءات والحسابات والتسجيل أو حدوث ظروف غير طبيعية تؤدي إلى حدوث هذا التطرف كالكوارث والحروب وأزماتها كما هو الحال في بيانات فجوة القدرة الكهربائية Mw لكافة القطاعات في محافظة نينوى وتطرفها بسبب الحرب وهي البيانات المتعمدة في البحث. لذلك توجب على الباحثين اختيار الأسلوب الملائم للتقدير في الحصول على تصورات عن المجتمع الذي أخذت منه العينة المدروسة في حالة وجود القيم المتطرفة.

لقد تم في هذا البحث كشف ومعالجة القيم المتطرفة بأساليب عدة كاستبعاد القيم المتطرفة من البيانات وتقديرها وزيادة حجم العينة لسنوات لاحقة وتقديرها واستخدام الطرق الحصينة للتقدير أيضاً.وتم مقارنة بين نتائج الطرق المستخدمة.

مفهوم القيم المتطرفة

(Hawkins, 1980) (chatlerjee & price,1977) (الراوي,١٩٨٧)
(Rousseeuw, 1987) (Thall, 1979) (Raymond, 1986)

مصطلح يطلق على المشاهدات المتطرفة Extreme وتسمى أيضاً الخوارج اوالقيم الشاذة Outliers أو الشاردة أو الفعالة أو ملوثة Contaminated بوجود هذه المشاهدات تسمى هذه الملوثات Contaminahts.

إن المشاهدات التي تقع بعيداً عن خط الانحدار وعادة " البواقي لها كبيرة Large residuals مقارنة" ببقية المشاهدات الطبيعية الأخرى، وأنها تؤثر كثيراً على النموذج الخطي وتقديراته.

إن النقاط التي تكون لقيمة Xi في حالة الانحدار الخطي البسيط تسمى نقطة فعالة إذا كانت تقع بعيداً عن معظم القيم في المصفوفة X أي بمعنى آخر أن التطرف لا يقتصر على قيم متغير الاستجابة بل يشمل واحد أو أكثر من المتغيرات التوضيحية في تحليل الانحدار، وأن تأثيرها يكون عال جداً على مقدرات المربعات الصغرى الاعتيادية (OLS).

تأثير القيم المتطرفة على تحليل الانحدار (chatlerjee & price, 1977)
(Hawkins, 1980)
(Raymond, 1986)
(vic & Toby, 1981)

من المعلوم أن أي إحصائي تطبيقي عندما يحلل مجموعة من البيانات الحقيقية من المحتمل أن تصادفه القيم المتطرفة والتي تعرف بأنها المشاهدات التي تنحرف كثيراً عن المشاهدات الأخرى أي تكون غير منسجمة مع بقية بيانات المجموعة بظاهرة معينة وقد تكون هذه القيم كبيرة أو صغيرة أي الظاهرة تكون ملوثة Contaminants وغالباً ما تنشأ مع توزيعات ثقيلة الذيل Heavytatied Ditributions أو التوزيع المختلط Mixture Distribution.

وأن لتلك المشاهدات المتطرفة في الانحدار الخطي البسيط والمعبر عنها هذا النموذج بالصيغة

$$Y_i = B_0 + B_1 X_i + U_i \ldots (1)$$

$$i = 1,2,3,......n$$

التي تكون القيمة X_i تسمى نقطة فعالة إذ كانت تقع بعيداً عن معظم القيم في المصفوفة X، أي بمعنى آخر أن التطرف لا يقتصر على متغير الاستجابة بل يشمل واحد أو أكثر من المتغيرات التوضيحية في تحليل الانحدار، وكذلك الحال لتأثير المشاهدات المتطرفة على الانحدار الخطي المتعدد والمعبر عن هذا النموذج بالصيغة

$$Y_i = B_0 + B_1 X_{1i} + B_{2i} X_{2i} + ...B_p X_{pi} + U_i ...(2)$$

$$i = 1,2,........,n$$

أن تشخيص القيم المتطرفة في متعدد المتغيرات ليس بالأمر السهل كما في حالة المتغير الأحادي إذ يتم اتخاذ المشاهدة المتطرفة جانباً عن مشاهدات العينة. ولمعالجة المشاهدات المتطرفة هناك أسلوبان يسيران باتجاهين متعاكسين الأسلوب الأول الكشف عن المشاهدات المتطرفة حيث يتم استبعادها عن المشاهدات ومن ثم استخدام التقديرات التقليدية، أما الأسلوب الثاني يتم الكشف عن القيم المتطرفة باستخدام طرق الاستكشاف من ثم استخدام الطرق الحصينة للتقدير.

ومن الصيغتان (١) ، (٢) لا يمكن أن ترتبط بعلاقة خطية دقيقة مع قيمة لمشاهدة X_{pi} في كل محاولة من المحاولات المتكررة مما يؤدي إلى إضافة متغير جديد يسمى بالحد العشوائي أو الخطأ العشوائي Random Variable.

كما أوضح Rousseeuw (1987) تأثير القيم المتطرفة على مقدرات المربعات الصغرى وأوضح بالرسم كيف أن المشاهدة المتطرفة الواحدة تغير اتجاه خط المربعات الصغرى.

إن وجود القيم المتطرفة في البيانات يؤدي إلى إرباك كبير في تحليل البيانات في حالة استخدام الطرق الكلاسيكية في التقدير ومن هذه الطرائق طريقة المربعات الصغرى.

لقد أوضح Huber (1973) تأثير القيم المتطرفة على مقدرات المربعات الصغرى من خلال مقولته المشهورة أن وجود قيمة متطرفة واحدة يهدم المزايا الجيدة لمقدرات المربعات الصغرى كما أنها تسحب إليها التوافق للمربعات الصغرى.

كشف ومعالجة القيم المتطرفة (الدباغ,1999)(الراوي,١٩٨٧) (Huber, 1973) (Rand, 1997) (Raymnd, 1986)(Vic & Toby,1981)

لمعالجة المشاهدات المتطرفة هناك أسلوبان.

الأسلوب الأول هو الكشف عن المشاهدات المتطرفة ثم إبعادها عن المشاهدات ومن ثم استخدام التقديرات التقليدية ويمكن اكتشاف القيم المتطرفة بيانياً عندما نرسم الرسم البياني eis(الخطأ القياسي للمشاهدات) ضد Y_i فالنقاط التي تقع خارج 2 ± تعد من القيم المتطرفة كما أن هناك اختبارات إحصائية للكشف عن هذه القيم المتطرفة حيث أن:-

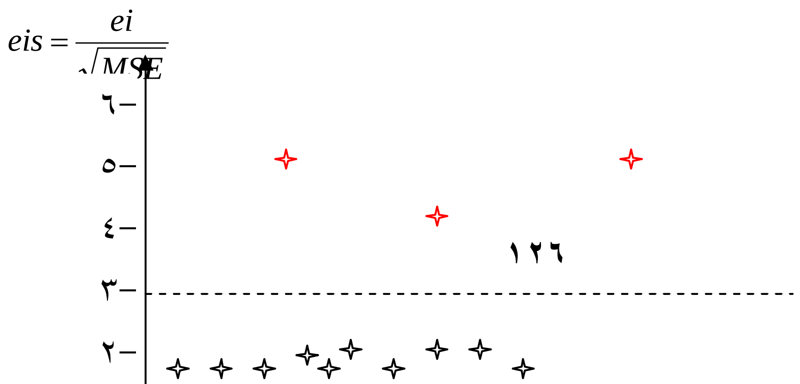
$$eis = \frac{ei}{\sqrt{MSE}} ...(3)$$

$$i = 1,2,3,........, n$$

حيث أن

ei   تمثل خطأ المشاهدة

MSE   متوسط مربع الخطأ



Y_i

✧ القيم المتطرفة   **الشكل (١)**

رسم بياني يبين القيم المتطرفة

إن تأثير القيم المتطرفة يمكن معرفته في تحليل البيانات كاملة ومن ثم تحليل البيانات باستثناء القيم المتطرفة بالمقارنة بمتوسط الخطأ النسبي المطلق MAPE في كلتا الحالتين

ولمعالجة هذه الحالة يمكن حذف القيم المتطرفة إذا كانت هذه القيم متأتية من خطأ تسجيل المشاهدات أو من وضع الأجهزة أي تصحح بالحذف وإلا فيجب جمع بيانات أكثر لتكون القرارات أقرب إلى الواقع.

أما الأسلوب الثاني:-   (صالح,٢٠٠١)(عزان,٢٠٠١)

(Huber, 1973) (Rousseuw, 1987)

إن الكشف عن القيم المتطرفة في البيانات أخذت مساحة واسعة في البحث الإحصائي بسبب كون الاستدلال الإحصائي المبني على أساس التوزيع الطبيعي يكون حساساً تجاه القيم المتطرفة.

وبالرغم من وجود طرق حصينة مختلفة إلا أن أغلبها تشترك بنقطتين أساسيتين أحدهما إعطاء وزن أقل للمشاهدات المتطرفة إذ وجدت وذلك للتقليل من تأثيرها، والأخرى هي استخدام أسلوب التكرار.

يتم الكشف عن القيم المتطرفة باستخدام طرق الاستكشاف الحصينة ومن ثم استخدام الطرق الحصينة للتقدير.

وقد اعتمدنا **طريقة الكشف الأولى** التي استخدمها Rousseuw, Leroy لتشخيص القيم المتطرفة في الانحدار الخطي المتعدد وكما يلي:-

١- نوجد قيم التقديرية لمتغير الاستجابة.

$$\hat{Y} = \hat{B}_0 + \hat{B}_1 x_{1i} + \hat{B}_2 x_{2i} + \hat{B}_3 x_{3i} + \dots\dots\dots + \hat{B}_p x_{p_i} \dots (4)$$

$i = 1,2,3,.......n$

٢- نحسب البواقي المعيارية للطريقة الحصينة كما يأتي:

$$\frac{r_i}{\hat{\sigma}} ...(5)$$

$i = 1,2,3,.........n$

$$\hat{\sigma} = K \sqrt{med \quad r_i^{\,2}} ...(\ 6 \ )$$

$i = 1,2,3,.........n$

حيث أن

K قيمة ثابتة موجبة

med ri الوسيط للبواقي

٣-يستخدم الرسم البياني لاكتشاف القيم المتطرفة حيث تكون قيم البواقي على المحور العمودي وعلى المحور الأفقي اما أن تكون تسلسل البيانات Index of the observation أو القيم التنبؤية لمتغير الاستجابة Estimated response تبين من الشكل (١) أن نقطة القطع البواقي 2.5 ± حيث أن المشاهدة التي تقع خارج هذه المنطقة تعد مشاهدة متطرفة.

طريقة الكشف الثانية دالة Huber

لقد اقترح الباحثون عدد من المشتقات $\Psi(.)$ من الدوال $\rho(.)$بحيث تجعل نتائج التقدير جيدة ولا تتأثر بوجود القيم المتطرفة وفيما يلي دالة مهمة لهذا النوع من المقدرات والمعرفة بدلالة الدالة $\Psi_{(u)}$

$$\Psi_{(u)} = \begin{cases} u & |u| < h \\ h\,\mathrm{sgn}(u) & |u| \geq h \end{cases} ...(7)$$

حيث أن h ثابت القطع Tunin constant ويأخذ القيم 1.5 و ١.٧٠ و ٢.٠٨.

وهناك عدة طرائق حصينة تستخدم لتقدير معلمات النموذج الخطي ومنها مقدرات (M) تقديرات نوع الأماكن الأعظم ومقدرات (R) تقديرات الراتب ومقدرات (L) تقديرات الاحصاءات المرتبة ومقدرات (M المكيفة) تقديرات نوع الأماكن الأعظم المكيف.

وسنتناول في هذا البحث طريقة دالة Andrews إذ تسمى أحياناً بدالة الجيب (Sin function) كتقدير بعد الكشف.

$$\Psi_{(u)} = \begin{cases} \sin(u/a) & |u| < \mathrm{II} \\ 0 & |u| \geq \mathrm{II} \end{cases} ...(8)$$

دقة مقدرات معلمات الانحدار

(Arthanari & et.al , 1981)

(Makridakis & et.al, 1998)

إن من مقاييس الدقة الشائعة الاستخدام هي متوسط مربع الخطأ Mean squares Error (MSE) والخطأ النسبي Percentage Error (PE) ومتوسط الخطأ النسبي المطلق Mean Absolute percentage Error (MAPE) ومتوسط مجموع الأخطاء المطلقة Mean sum of Absolute Error (MSAE). وسنستخدم في هذا البحث متوسط الخطأ النسبي المطلق(MAPE) وهو متوسط الأخطاء النسبية المطلقة لمجموعة من البيانات التي أخذت دون إشارة، هو المقياس الوحيد للدقة المستخدم بصورة عامة في الطرق الكمية للتقدير، ويأخذ هذا المعيار الشكل الآتي:

$$MAPE = \frac{\sum_{i=1}^{n} \left| PE_i \right|}{n} \ldots(9)$$

$$i = 1,2,3,\ldots\ldots\ldots\ldots n$$

حيث أن PE_i الخطأ النسبي للمشاهدة.

فـروض الخطـأ العشـوائي (Arthanari &et.al,1981) (كاظم،١٩٩٩)

الفروض التي يجب توفرها في الأخطاء العشوائية:-

١. إن متوسط قيم المتغير العشوائي (U) يساوي صفر

$$E\left(\underline{U}\right) = 0 \ldots(10)$$

٢- إن تباين قيم المتغير العشوائي (U) يكون ثابتاً وفي كل فترة زمنية

$$E\left(\underline{U}\ \underline{U}^T\right) = \sigma^2 In \ldots(11)$$

٣- الأخطاء (Ui) تتوزع طبيعياً بمتوسط صفر وتباين (σ^2)

$$Ui \sim N(0,\sigma^2) \ldots(12)$$

٤- القيم المختلفة للمتغير العشوائي U_i تكون مستقلة عن بعضها البعض أي أن قيمة العنصر العشوائي U_i في أية فترة لا تعتمد على قيمة في فترة أخرى.

$$E(UiU_j) = Cov\left(UiU_j\right) = 0 \ldots(13)$$

$\forall\ i \neq j.$

٥- الاستقلالية بين المتغيرات التوضيحية والأخطاء العشوائية

$$E(U_i X_i) = 0 \ldots (14)$$

٦- إن U_i متغير عشوائي حقيقي، إذ أن أية قيمة من قيم (U) وفي أية فترة زمنية تعتمد على الصدفة وقد تكون هذه القيم سالبة أو موجبة أو مساوية للصفر.

إن تحقيق الفروض السابقة من الناحية العلمية تعتبر حالة مثالية (وهي نادراً ما تتحقق) فإذا لم تتحقق فرضية واحدة أو أكثر فإن هذا سيؤدي إلى أن المقدرات B تكون غير دقيقة، لقد جرت العادة باستخدام طريقة المربعات الصغرى الاعتيادية إن هذه الطريقة لا تعتمد على معرفة التوزيع الاحتمالي للأخطاء العشوائية وهي كثيرة الاستعمال ويطلق عليها المربعات الصغرى الاعتيادية ordinary Least squares (OLS) إذا تحققت الفروض الخاصة بها.

جمع البيانات (جاسم، ٢٠٠٨)

أخذت البيانات من مديرية مبيعات الطاقة/ توزيع كهرباء نينوى، والمعتمدة في رسالة ماجستير .والبيانات تمثل فجوات القدرة الكهربائية M_w لكافة القطاعات في محافظة نينوى للفترة (٢٠٠٤-٢٠٠٧) ولستة متغيرات وهي:-

X_1 فجوة القطاع المنزلي

X_2 فجوة القطاع التجاري

X_3 فجوة القطاع الحكومي

X_4 فجوة القطاع الصناعي

X_5 فجوة القطاع الزراعي

Y أشهر السنة

التحليل الإحصائي

الأسلوب الأول:-

أولاً: الكشف عن القيم المتطرفة للبيانات الأصلية لعام ٢٠٠٤ بالطريقة التقليدية.

١– نجد القيم التقديرية لمتغير الاستجابة من خلال المعادلة التقديرية وهي:–

$$\hat{Y} = -6.8200 - 0.0199X_1 + 0.730X_2 + +0.163X_3 + 0.0377X_4 + 0.313Xs$$

٢– ثم نجد البواقي المعيارية وفق الصيغة (٣)

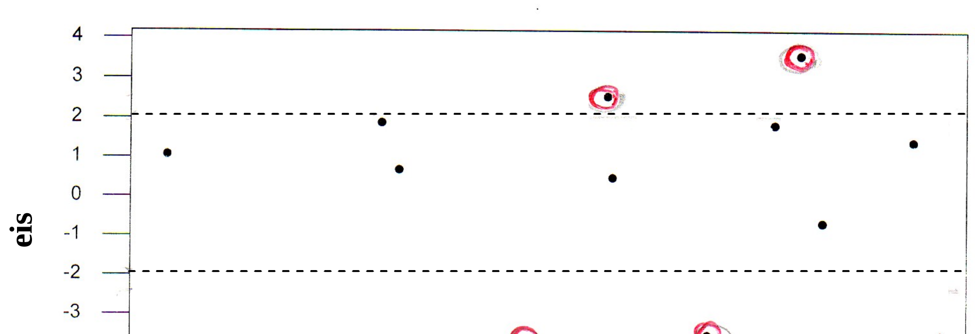
الجدول (١)

يمثل البواقي المعيارية والقيم المتطرفة فيها مرتبة حسب الشهر لعام ٢٠٠٤.

Index	Month	eis
1	كانون الثاني	-5.7812 *
2	شباط	-3.59507 *
3	آذار	1.06927
4	نيسان	-3.50233 *
5	أيار	0.68745
6	حزيران	1.88669
7	تموز	0.49116
8	آب	-0.64972
9	أيلول	2.55314 *
10	تشرين الأول	1.83849
11	تشرين الثاني	1.42153
12	كانون الأول	3.5060 *

*

تمثل القيم المتطرفة للبيانات الأصلية لعام ٢٠٠٤



Index(month)

الشكل (2)

القيم المتطرفة لبيانات فجوات القدرة الكهربائية Mw الأصلية لعام ٢٠٠٤

إذ يلاحظ أن القيم المتطرفة في الجدول (١) للبواقي تكون أكبر من (±٢) لشهر كانون الثاني، شباط، نيسان، أيلول، كانون الأول، نسبَةً لبقية الأشهر من السنة ٢٠٠٤. والشكل(٢) يوضح أيضاً القيم المتطرفة لبيانات فجوات القدرة الكهربائية Mw لنفس السنة أعلاه.

ثانياً- معالجة القيم المتطرفة بالطريقة التقليدية

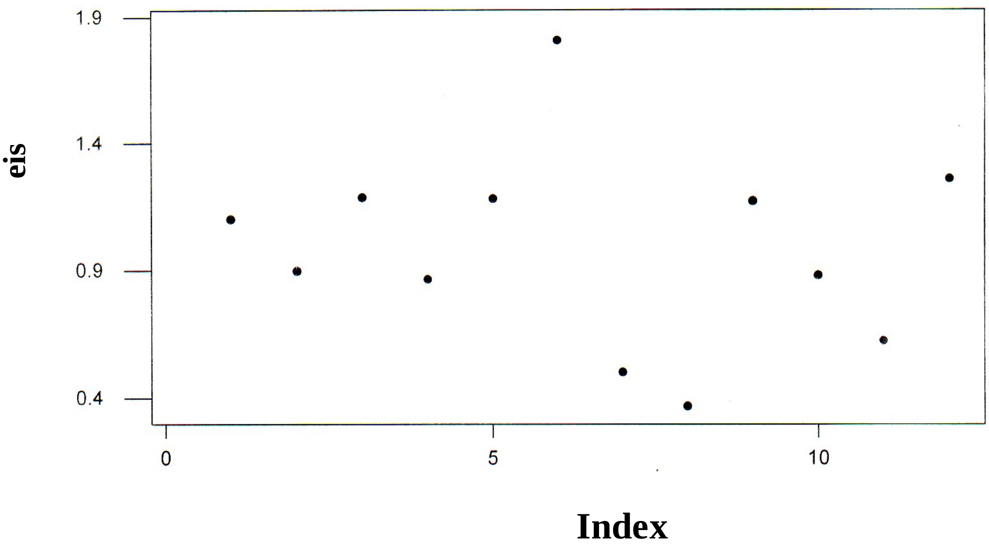
وذلك باستبعاد القيم المتطرفة من البيانات الأصلية أي استبعاد بيانات شهر كانون الثاني، شباط، نيسان، ايلول، كانون الأول من بيانات عام ٢٠٠٤ التي تعتبر القيم المتطرفة من البيانات الأصلية إذ يلاحظ المعادلة التقديرية لمتغير الاستجابة بعد استبعاد القيم المتطرفة من البيانات الأصلية لعام ٢٠٠٤ هي:

$$\hat{Y} = -7.03 - 0.00838X_1 + 0.473X_2 + 0.128X_3 + 0.0231X_4 + 0.242X_5$$

الجدول (٢)يمثل البواقي المعيارية بالطريقة الاعتيادية بعد استبعاد القيم المتطرفة لبيانات فجوة القدرة الكهربائية Mw لعام ٢٠٠٤

Index	Month	eis

1	آذار	-0.10935
2	أيار	0.244188
3	حزيران	-0.098108
4	تموز	-0.173344
5	آب	-0.132458
6	تشرين الأول	0.523455
7	تشرين الثاني	-0.254381



الشكل (3)يمثل استبعاد البواقي المتطرفة لبيانات فجوة القدرة الكهربائية Mw لمحافظة نينوى ولعام ٢٠٠٤

والشكل رقم (3) يوضح حذف القيم المتطرفة للبواقي.

ومن المعلوم ليس دائماً ممكن استبعاد القيم المتطرفة وإجراء التحليل الإحصائي لأنه قد تكون البيانات مهمة ولا يمكن إهمالها وتؤثر على النتائج وعلى إعطاء صورة حقيقية لواقع الدراسة، بغض النظر عن عدم وجود قيم متطرفة وإجراء التحليل بالطرق الاعتيادية.

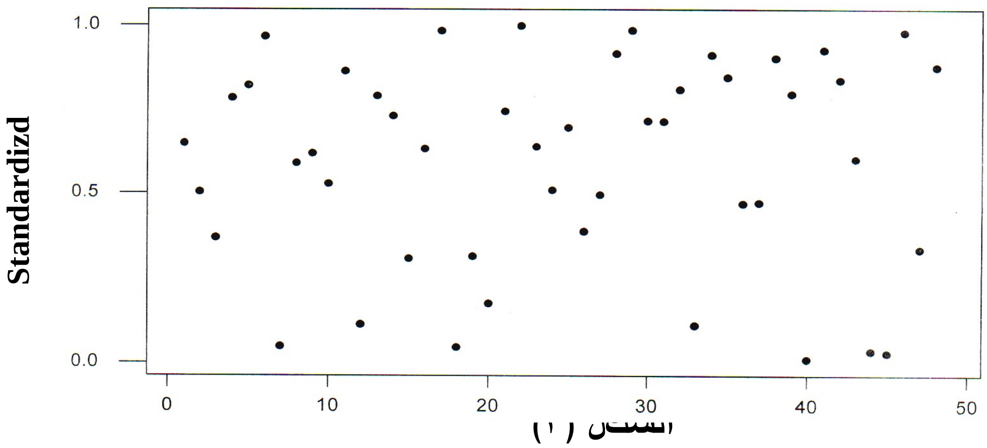
ثانياً: زيادة حجم العينة لعلاج القيم المتطرفة:لذلك اعتمدنا زيادة حجم العينة لسنوات لاحقة أخرى وهي ٢٠٠٥، ٢٠٠٦، ٢٠٠٧ لتلافي استبعاد القيم المتطرفة من جهة والحصول على نتائج واقعية تخدم الدراسة من جهة أخرى، وتعتبر هذه طريقة معالجة للقيم المتطرفة.

وبعد زيادة حجم العينة للأعوام ٢٠٠٥، ٢٠٠٦، ٢٠٠٧ لكافة القطاعات، تم الكشف عن البواقي وفق الصيغة (٣) وملاحظة عدم وجود قيم متطرفة تزيد عن (٢±). والجدول رقم(٣) يوضح ذلك.

أما المعادلة التقديرية لمتغير الاستجابة بعد زيادة حجم العينة للأعوام ٢٠٠٥، ٢٠٠٦، ٢٠٠٧.

$$\hat{y} = -3.39 + 0.007x_1 + 1.91x_2 + 0.079x_3 - 0.0313x_4 + 0.0513x_5$$

والشكل (٤) يوضح عدم وجود قيم متطرفة للبواقي المعيارية بعد زيادة حجم العينة بالأعوام أعلاه.



Estimed response

الشكل (٤)

عدم وجود قيم متطرفة للبواقي المعيارية تتجاوز (٢±) بعد زيادة حجم العينة للأعوام ٢٠٠٤، ٢٠٠٥، ٢٠٠٦، ٢٠٠٧.

الأسلوب الثاني

١- الكشف عن القيم المتطرفة بالطريقة الحصينة التي استخدمها Leroy، Rousseeu

إذ تعتمد الخطوات التالية:- نجد المعادلة التقديرية لمتغير الاستجابة وللأسلوب الثاني وهي

$$\hat{Y} = -7.90 - 0.0017x_1 + 0.13x_2 + 0.267x_3 - 0.0080x_4 + 0.316x_5$$

ثم نحسب البواقي المعيارية للطريقة الحصينة **Leroy، Rousseeu** كما في المعادلة (٦)

الجدول (٤)

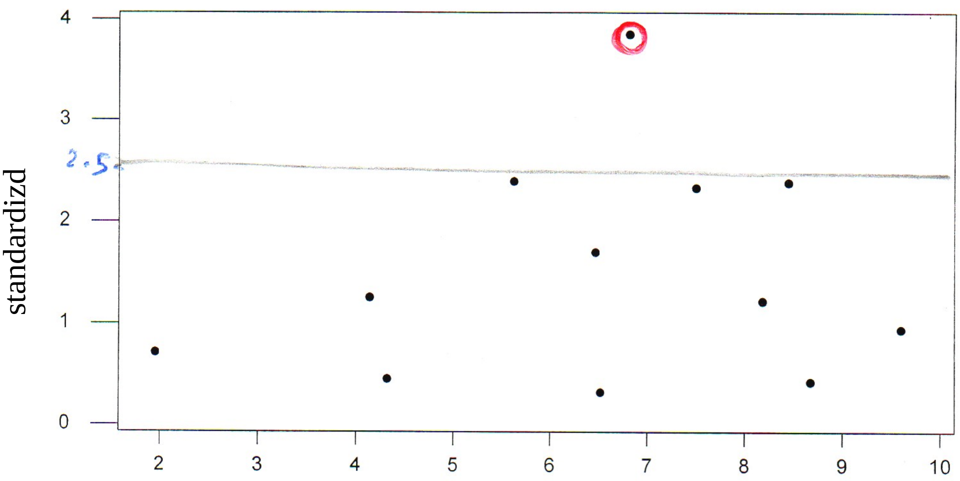
يمثل كشف البواقي المعيارية بالطرق الحصينة Rousseeuw, Leroy مرتبة حسب الشهر لبيانات عام ٢٠٠٤.

Index	Month	eis
1	كانون الثاني	3.85414 *
2	شباط	2.39671
3	آذار	0.71284
4	نيسان	0.33489
5	أيار	0.45830
6	حزيران	0.25779
7	تموز	0.32744
8	آب	0.43314
9	أيلول	1.70209
10	تشرين الأول	1.22566
11	تشرين الثاني	0.94769
12	كانون الأول	2.38706

* تمثل القيمة المتطرفة

الجدول (٣) يمثل عدم وجود قيم متطرفة للبواقي المعيارية تتجاوز (٢±) بعد زيادة حجم العينة للأعوام ٢٠٠٤، ٢٠٠٥، ٢٠٠٦، ٢٠٠٧ـ لبيانات الفجوة القدرة الكهربائية في نينوى

Years	Index	Month	eis
2004	1	كانون الثاني	0.648657
	2	شباط	0.503855
	3	آذار	0.367320
	4	نيسان	0.783518
	5	أيار	0.821737
	6	حزيران	0.966635
	7	تموز	0.047863
	8	آب	0.589138
	9	أيلول	0.618356
	10	تشرين الأول	0.528201
	11	تشرين الثاني	0.863611
	12	كانون الأول	0.111816
2005	13	كانون الثاني	0.789982
	14	شباط	0.729842
	15	آذار	0.306048
	16	نيسان	0.631364
	17	أيار	0.983525
	18	حزيران	0.045599
	19	تموز	0.312569
	20	آب	0.173549
	21	أيلول	0.743944
	22	تشرين الأول	0.998733
	23	تشرين الثاني	0.6380430
	24	كانون الأول	0.508967
2006	25	كانون الثاني	0.694908
	26	شباط	0.385897
	27	آذار	0.494509
	28	نيسان	0.916459
	29	أيار	0.985581
	30	حزيران	0.715038
	31	تموز	0.713226
	32	آب	0.808945
	33	أيلول	0.109677
	34	تشرين	0.912388
	35	تشرين الثاني	0.485259
	36	كانون الأول	0.468161
2007	37	كانون الثاني	0.470590
	38	شباط	0.902896
	39	آذار	0.795184
	40	نيسان	0.010271
	41	أيار	0.927176
	42	حزيران	0.837504
	43	تموز	0.600636
	44	آب	0.034363
	45	أيلول	0.028499
	46	تشرين الأول	0.979438
	47	تشرين الثاني	0.332532
	48	كانون الأول	0.876418



Estimed response

الشكل رقم (5)الكشف بالطريقة الحصينة

Rousseeuw, Leroy وجود القيمة المتطرفة لبيانات فجوة القدرة الكهربائية Mw

محافظة نينوى لعام ٢٠٠٤

إذ يلاحظ أن القيمة الأولى في جدول (٤) للبواقي أكبر من ٢.٥+ وتمثل القيمة المتطرفة لبيانات فجوة القدرة الكهربائية Mw لكافة القطاعات لـ ١٢ شهر في محافظة نينوى لعام ٢٠٠٤.

ومن الشكل (5) يوضح القيمة المتطرفة لبيانات فجوات القدرة الكهربائية Mw في شهر كانون الثاني ولكافة القطاعات نسبية لبقية الأشهر في السنة ٢٠٠٤.

٢- معالجة القيم المتطرفة بالطريقة الحصينة Anderws :من المعلوم أن طريقة المربعات الصغرى غير كفوءة لتقدير معلمات النموذج الخطي المتعدد إذا كان هناك خرق الفروض الخاصة لطريقة المربعات الصغرى، وذلك بسب وجود القيم المتطرفة في بيانات العينة. لذلك سيتم اللجوء إلى اسلوب التقدير الحصين بتطبيق طريقة دالة Anderws لمقدرات M وبعد استخدام المعادلة (٨) لبيانات فجوات القدرة الكهربائية لكافة القطاعات في محافظة نينوى لعام ٢٠٠٤ مصنفة حسب الشهر.

إذ يلاحظ المعادلة التقديرية لمتغير الاستجابة بعد معالجة القيم المتطرفة بالطريقة الحصينة هي

$$\hat{Y} = -14.1 + 0.01872\ x_1 - 0.30\ x_2 + 0.315\ x_3 - 0.0143\ x_4 + 0.526\ x_5$$

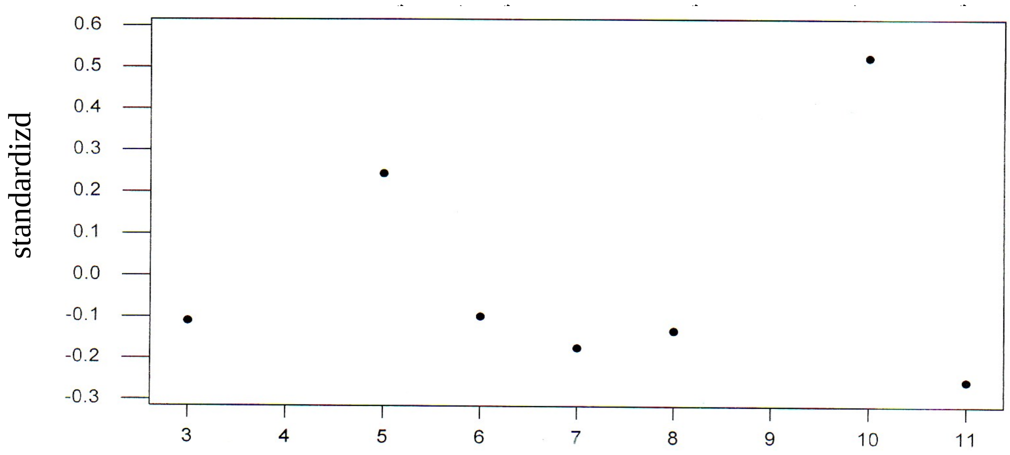
٣- الكشف عن القيم المتطرفة بعد المعالجة بالطريقة الحصينة لـ Huber

بعد معالجة القيم المتطرفة للبيانات بالطريقة الحصينة تم الكشف بطريقة أخرى بطريقة دالة Hubers بالاعتماد على الصيغة (٧) للبواقي المعيارية فكانت النتائج كما يلي:

الجدول (٥) يمثل كشف البواقي المعيارية بعد المعالجة بالطريقة الحصينة Huber مرتبة حسب الشهر لبيانات عام ٢٠٠٤.

Index	Month	Sie
1	كانون الثاني	1.10073
2	شباط	0.89927
3	آذار	1.18646
4	نيسان	0.86581
5	آيار	1.181350
6	حزيران	1.80888
7	تموز	0.50277
8	آب	0.37012
9	أيلول	1.17009
10	تشرين الأول	0.88179
11	تشرين الثاني	0.62593
12	كانون الأول	1.25694

إذ يلاحظ في الجدول (٥) لا يوجد أي قيمة من قيم البواقي تتجاوز (٢.٥ +، ٢.٥−) لنفس البيانات السابقة الذكر بعد معالجتها بأسلوب التقدير الحصين بطريقة دالة Huber لمقدرات (M).



الشكل (6)

يوضح عدم تجاوز قيم البواقي عن (+٢.٥ ، -٢.٥) للبيانات بعد معالجتها باسلوب التقدير الحصين بطريقة دالة Anderus لمقدرات (M).

بعد الكشف لا وجود للقيم المتطرفة بعد معالجتها بالمقدر الحصين بدالة Huber لبيانات فجوة القدرة الكهربائية في محافظة نينوى لعام ٢٠٠٤.

مقارنة دقة الأساليب المعتمدة بمقياس (MAPE).

في حالة التقدير التطبيقية تعامل الدقة كمعيار لاختبار طريقة التقدير، وهناك عدة أسباب تؤدي إلى عدم الدقة، مثل البيانات الغير كافية أو استخدام اسلوب لا يطابق نوع البيانات.

فالدقة هي المعيار الأكثر استخداماً لتقويم الانجاز لطرق التقدير، وهي تعكس صحة التنبؤ.

والجدول رقم (٦) يبين القيم المقدرة لمعلمات الطرق المختلفة للأسلوب الأول والثاني. كما يلاحظ مقياس الدقة متوسط الخطأ النسبي (MAPE) للأسلوبين أعلاه، حيث يلاحظ أنه كلما قل MAPE يعني أفضل ويمكن اعتماد الأسلوب الذي يكون فيه قيمة المقياس أقل، أما النسبة للبيانات المعتمدة في البحث يلاحظ جميع الطرق ذات دقة جيدة والمميزة بين الطرق هي الأسلوب الأول المعالجة بزيادة حجم العينة لعام ٢٠٠٤، ٢٠٠٥، ٢٠٠٦، ٢٠٠٧ حيث أن MAPE تساوي (٠.٠٢٠٥١٩٩) ويليها الأسلوب الثاني بالطرق الحصينة إذ بلغت على التوالي (٠.٠٨٣٣٣٣) و (٠.٠٨٣٣٣٤٦). ويليها الأسلوب الأول بالطريقة التقليدية إذ بلغت أيضاً على التوالي (0.1898225) و (٠.١٤٢٨٥٧١) وهذا يؤكد صحة وجهة نظرنا في اعتماد الطريقة التقديرية بزيادة حجم العينة للبيانات لأنها تمثل المجتمع خير تمثيل.

ويليها استخدام طرق التقدير الحصينة لكونها تعالج القيم المتطرفة.

ويمكن استخدام الطرق التقليدية بالأسلوب الأول في حالات معينة كأن يمكن استبعاد القيم المتطرفة في حالة تكون سبب القيم المتطرفة خطأ في تسجيل المشاهدات أو من وضع الأجهزة.

جدول (٦)يبين القيم المقدرة للمعلمات بالطرق المختلفة وMAPE لبيانات فجوة القدرة الكهربائية Mw لمحافظة نينوى

ت	المتغيرات والثابت	الأسلوب الأول الكشف بالطريقة التقليدية لعام ٢٠٠٤	الأسلوب الأول المعالجة باستثناء القيم المتطرفة لعام ٢٠٠٤	الأسلوب الأول المعالجة بزيادة حجم العينة للأعوم ٢٠٠٤، ٢٠٠٥، ٢٠٠٦، ٢٠٠٧	الأسلوب الثاني الكشف بالطريقة الحصينة Reusseuw للقيم المتطرفة لعام ٢٠٠٤	الأسلوب الثاني معالجة بالطريقة الحصينةAnderws والكشف بطريقة Huber للقيم المتطرفة لعام ٢٠٠٤
١	الثابت	-٦.٨٢٠٠	-٧.٠٣٠٠	-٣.٣٩٠٠	-٧.٩٠٠٠	-١٤.١٠٠
٢	فجوة القطاع المنزلي	-٠.٠١٩٩	-٠.٠٠٨٣٨	٠.٠٠٧٠	-٠.٠٠١٧	٠.٠١٨٧٢
٣	فجوة القطاع التجاري	٠.٧٣٠٠	٠.٤٧٣٠٠	١.٩١٠٠	٠.١٣٠٠	-٠.٣٠٠٠
٤	فجوة القطاع الحكومي	٠.١٦٣٠	٠.١٢٨٠٠	٠.٠٧٩٠	٠.٢٦٧٠	٠.٣١٥٠
٥	فجوة القطاع الصناعي	٠.٠٣٧٧	٠.٠٢٣١	-٠.٠٣١٣	-٠.٠٠٨٠	-٠.٠١٤٣
٦	فجوة القطاع الزراعي	٠.٣١٣٠	٠.٢٤٢٠	٠.٣١٣	٠.٣١٦٠	٠.٥٢٦٠
	MAPE	٠.١٨٩٨٢٢٥	٠.١٤٢٨٥٧١	٠.٠٢٠٥١٩٩	٠.٠٨٣٣٣٤٦	٠.٠٨٣٣٣٣٣

لاستنتاجات

١- إن تحقيق فروض التحليل من الناحية العلمية تعتبر حالة مثالية وهي نادراً ما تتحقق، حيث أن القيم المتطرفة الفعلية تؤثر على صلاحية الاختبارات التقليدية، لكن تكون غير حساسة للانحرافات الطفيفة عن الحالة الطبيعية.

٢- إن طرق التحليل باستبعاد القيم المتطرفة من البيانات الكاملة ممكن اعتمادها في حالة أن القيم المتطرفة متأتية من خطأ تسجيل المشاهدات أو من وضع الأجهزة.

٣- إن طرق التحليل باعتماد البيانات كاملة وزيادة البيانات بزيادة حجم العينة يساعد في اعتماد طرق التقدير التقليدية في التحليل، وتكون هذه الطريقة جيدة لتمثيلها المجتمع خير تمثيل لكون حجم العينة كبير.

٤- إن طرق الكشف والتقدير بالطرق الحصينة ممكنه وتحتاج غالباً إلى تكرار عديد، ويمكن استخدامها في حالة البيانات الكاملة بوجود القيم المتطرفة، ولكن معيار الدقة متوسط مربع الخطأ (Mean squares error (MSE في الطرق الحصينة يولد مشاكل فهو لا يسهل المقارنة عبر السلاسل الزمنية، وكذلك فإنه غير ملائم تطبيقياً أن نجري مقارنات بين طرق التقدير المختلفة لأنه يعطي وزناً كبيراً للأخطاء الكبيرة مقارنةً بالأخطاء الصغيرة لأن الأخطاء تربع قبل أخذ مجموعها.

لذا تم اعتماد مقياس الدقة متوسط الخطأ النسبي MAPE في جميع طرق التقدير المعتمدة في البحث بأسلوبنا الأول والثاني الذي أكد وجهة نظر الباحثة أنه حجم العينة الكبير يعطي أدق النتائج للدراسة لكون العينة تمثل المجتمع خير تمثيل.

المصادر

أ- المصادر العربية:

– جاسم، يسرى حازم (٢٠٠٨). "تقدير فجوة الطلب على الطاقة الكهربائية في محافظة نينوى والتنبؤ حتى عام (٢٠١٠). " رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة الموصل.

– الدباغ، ظافر عاصم مصطفى، (١٩٩٩). "تحليل تباين حصين للنماذج الخطية"، رسالة دكتوراه (غير منشورة)، مقدمة إلى كلية الإدارة والاقتصاد-جامعة بغداد.

– الراوي، خاشع محمود، (١٩٨٧). "المدخل إلى تحليل الانحدار" مديرية دار الكتب للطباعة والنشر- جامعة الموصل.

– صالح، ذكاء يوسف عزيز (٢٠٠١). "مقارنة بعض الطرائق الحصينة لتقدير معلمات نموذج الانحدار الخطي المتعدد"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

– عزان، علي سالم موسى، (٢٠٠١). "مناقشة نظرية وتجريبية في تقويم المقدرات الحصينة"، أطروحة دكتوراه (منشورة) مقدمة إلى كلية الإدارة والاقتصاد_الجامعة المستنصرية.

– كاظم، أموري هادي ومحمود، عصام خضير، (١٩٩٩). "طبيعة البيانات الإحصائية وبناء النماذج القياسية" دار وائل للنشر، الأردن.

ب- المصادر الأجنبية

- Arthanari, T.S.and Dodae. Y.(1981), "Mathe matical Programming in statistics". John wil and sons Inc.

- Chatlerjee. S. and price. (1977), "Regression Analysis By Example", John Wiley and Sons., New York.

- Hawkins, D.M. (1980), "Identification of outliers", chapman and Hall, London, New York.

- Huber, P.J.(1973), "Robust Regression: Asymptotics conjectures and Monte cario", Ann of statist, Vol.1, No.5, PP. 799-821.

- Makridakis, S; Steren, C.W and Rob, J.H. (1998), "Forecasting Methods and Application" Third Ed,. John Wiley and, sons Ic.

- Rand, R.W. (1997), "Introduction to Robust Estimation and Hypothesis Testing", San Diego London Boston.

- Raymond, H.M. (1986), "Classical and Modern Regression with Applications", Pws Publishers.

- Rousseeuw. Peter . (1987), “Robust Regression and Outlier Detection", John Wiley And Sons, New York.

- Thall, P.F. (1979), "Robust M-Estimator of parameter, with Application to Exponential Distribution", JASA, Vol. 74, pp. 147-152.

- Vic, B and Toby, L. (1981), "Outliers in statistical Data", John wiley and sons. New York.