المربعات الصغرى المشذبة الموزونة لتقدير تأثير مياه الصرف الصحي في تلوث مياه نهر دجلة/ محافظة واسط

أ. م. د عمر عبد المحسن علي / كلية الادارة والاقتصاد ـ جامعة بغداد
الباحث/ خالد جمال كاظم

تاريخ التقديم:2018/4/26
تاريخ القبول:2018/7/31

المستخلص

في كثير من الأحيان يفشل تحليل المربعات الصغرى (LS) تماماً في حالة وجود قيم شاذة في الظواهر المدروسة، اذ ستفقد OLS خصائصها ومن ثم تفقد صفة المقدر الخطي الجيد Beast Linear Unbiased Estimator (BLUE) لِما تسببه الشواذ Outliers من تأثير سيئ على نتائج التحليل الاحصائي للبيانات اذ أن وجودها يؤدي الى إرباك كبير في تحليل البيانات في حالة إستخدام الطرائق التقليدية، ولعلاج هذه المشكلة تم تطوير أساليب إحصائية جديدة بحيث لا تتأثر بالقيم الشاذة بسهولة. هذه الطرائق تمتاز بالحصانة أو (المقاومة). ولذا كانت طريقة المربعات الصغرى المشذبة Least Trimmed Squares (LTS) كبديل جيد يحقق نتائج أكثر مقبولية وأمثليه. الاّ انه يمكن افتراض أوزان تأخذ بنظر العناية مواقع تواجد القيم الشاذة في البيانات وتحددها بشكل دقيق. ولزيادة قوة التقدير بطريقة المربعات الصغرى المشذبة الموزونة Weighted Least Trimmed Squares (WLTS) هو بأعاده الوزن لبيانات العينة حول المقدر المطلوب بصورة تكرارية وهو ما سيدعى طريقة المربعات الصغرى المشذبة المعاد وزنها Reweighted Least Trimmed Squares (RWLTS). ولتحقيق هذا البحث استدعت الحاجة الكشف والتقصّي عن تأثير التلوث مياه نهر دجلة في محافظة واسط بسبب مياه الصرف الصحي وبالذات التلوث بالمواد الصلبة غير الذائبة في الماء Total Dissolved Solids (TDS) وتأثير ثلاث ملوثات أملاح الكبريتات Sulphates (SO_4)، الكلورايدات Chlorides (Cl) والفوسفات Phosphates (PO_4) على ذلك. وتم تقديم ذلك بدراسة احصائية وتقييمها بشكل دقيق ورفعها الى الجهات المختصّة ولتحقيق هذا الهدف تم استعمال بحجم عينة (91) موقع تم سحبها وفحصها في مختبرات بلدية محافظة واسط. وتم إجراء التحليل باستعمال برنامج (MATLAB-R2015b).

المصطلحات الرئيسة للبحث/ المربعات الصغرى، القيم الشاذة، الطرائق المشذبة، تلوث مياه الانهار.

مجلة العلوم الاقتصادية والإدارية
العدد 109 المجلد 24
الصفحات 486ـ 495

*بحث مستل من رسالة ماجستير

1-1 مقدمة Introduction

تعدّ مشكلة وجود قيم شاذة (Outliers) في البيانات من المواضيع التي حازت على اهتمام كبير، وذلك لإدراك الكثير من الباحثين أن استخدام الأساليب التقليدية في تقدير معلمات نموذج الانحدار في ظل وجود هذه المشكلة ستؤدي الى مقدرات غير جيدة. لذا كان من الضروري البحث عن طرائق بديلة تكون حصينة لظهور الشواذ في البيانات. أذ فضّل بعض الباحثين عدم تغيير قيم الشواذ بل يجب ايجاد طرائق بديلة للتقدير تكون غير حسّاسة لها عند وجود الشواذ في العينة لأنه كما هو معلوم عملية تقدير المعلمات للنموذج المشخص تعتبر مؤشرا للحصول على تمثيل جيد للمجتمع الذي أخذت منه العينة قيد الدراسة. على الرغم من وجود طرائق حصينة مختلفة للتقدير الاّ أن أغلبها تشترك في نقطتين أساسيتين هما: إعطاء وزن أقل للقيمة الشاذة التي تنحرف لتقليل من تأثيرها والأخرى هي استخدام أسلوب التكرار. وغالباً ما تظهر القيم الشاذة البيانات الحقيقية بصورة كبيرة جدا وأنها غالبا ما تظهر دون أن يلاحظها أحد وذلك بسبب أن البيانات تتم معالجتها بواسطة اجهزة الكومبيوتر من دون الكشف الدقيق. والقيم الشاذة لا تقتصر على قيم متغير الاستجابة y بل يمكن أن يكون في الجزء التفسيري x أيضاً. وكلا النوعين من الشواذ في كثير من الأحيان يفشلان تحليل المربعات الصغرى (LS) تماماً وكذلك الامكان الأعظم (ML). ولعلاج هذه المشكلة تم تطوير أساليب إحصائية جديدة بحيث لا تتأثر بالقيم الشاذة بسهولة. هذه الطرائق تمتاز بالحصانة أو (المقاومة). حيث أن هذه القيم الشاذة أما أن تأتى من بيانات لتوزيعات غير متماثلة أي يكون فيها التواء عالٍ نحو اليمين أو نحو اليسار او من نوعين من التوزيعات أحدهما التوزيع الأساسي والذي يولد مشاهدات جيدة بينما الآخر يسمى التوزيع الملوث والذي يولد قيما شاذة او نتيجة لأسباب أخرى منها أخطاء يقع فيها الباحث عند تسجيل القياسات أو نتيجة وجود خلل في جهاز القياسات ولاسيما في التجارب المختبرية، أو نتيجة أخطاء في الحسابات مما يؤدي إلى ظهور القيم الشاذة.

2-1 هدف البحث Goal of the Research

يهدف البحث الى تسليط الضوء على تلوث مياه نهر دجلة في محافظة واسط بمياه الصرف الصحي وذلك لأهمية المياه في حياة جميع الكائنات الحية، وإن مياه الصرف الصحي تُعد من أخطر مصادر تلوث المياه لِما تحتويه من مواد ضارة لا يمكن التخلص منها بسهولة ،حيث تم إستعمال أسلوب التشذيب (trimming) ضد وجود القيم الشاذة، ودمجها بطرائق تقدير نموذج الانحدار المعروفة، مثل: المربعات الصغرى (OLS) Ordinary least squares لتصبح Least Trimmed Squares (LTS) والامكان الاعظم (ML) Maximum Likelihood لتصبح Maximum Trimmed Likelihood (MTL) ،حيث إن التشذيب في اللغة هو إبعاد الأجزاء التي لا يُرتجي منها فائدة ،وكذلك يهدف الى إستعمال أوزان حصينة يتم إدخالها على الطريقتين المشذبة أعلاه لتصبح Weighted Least Trimmed Squares (WLTS) و(WMTL) وWeighted Maximum Trimmed Likelihood وكذلك التعرّف في الوقت نفسه على القيم الشاذة وكشفها، ومقارنة افضلية هذا الاسلوب بإحدى معايير المقارنة مثل: Root Mean Square Error (RMSE)، (Efficiency) وذلك من خلال المقارنة بين كفاءة الطرائق المستخدمة.

3-1 القيم الشاذة Outliers

هي المشاهدات في العينة التي تكون قيمها صغيرة جدا أو كبيرة جدا لمقارنة ببقية المشاهدات في العينة حيث أن وجود هذه القيم يؤثر سلبا على كفاءة الطرائق الإحصائية (ذلك لإنها سوف تنتج مقدرات غير كفوءة أي ستفقد خصائص المقدر الجيد) ويجب الكشف عن هذه القيم للحدّ من تأثيرها [6]p.252. ويمكن تعرفها كذلك على أنها المشاهدات التي تبدو غير منطقية عندما تتم مقارنتها ببقية البيانات أو هي المشاهدات التي تقع بعيدة عن معادلة الانحدار ويكون لها خطا كبير مقارنة ببقية المشاهدات الطبيعية الأخرى. بينما ورد التعريف الدقيق للقيم الشاذة يستند على سياق الدراسة [11]p.1، وقد وضح Rousseeuw (1987) بيانياً عن كيفية تأثير القيمة الشاذة في مقدرات طريقة المربعات الصغرى وكيف أن هذه القيمة تغير من اتجاه خط الانحدار وقد عرف العديد من الباحثين القيم الشاذة. اذ عرّف Hawkins p.1 [5] المشاهدات الشاذة كالاتي: "هي المشاهدات التي تنحرف كثيرا عن المشاهدات الأخرى مما يثير الشك أنه تم توليدها بألية مختلفة ".

اما Freeman [9] p.2 فقد عرّف المشاهدات الشاذة كالاتي: "أي مشاهدة التي لم يتم توليدها بواسطة الالية التي ولدت أغلبية المشاهدات في مجموعة البيانات ". وبذا يمكن القول أن القيمة الشاذة هي المشاهدة التي تقع بعيداً عن معظم مشاهدات العينة والتي لا تنسجم أو لا تتطابق مع بقية المشاهدات الموجودة في العينة تحت الدراسة. وأن هذه القيمة الشاذة يكون لها تأثير سلبياً على المعالم المقدرة وقد وضح Huber [7]p.1 ذلك بقوله: "ان وجود مشاهد واحدة شاذة قد يهدم المزايا الجيدة لمقدرات المربعات الصغرى".

إن تعاريف القيم الشاذة تقع في خمسة فئات[12] p.1 وأن هذه الفئات تستند الى:

1. التوزيع: إن المشاهدة الشاذة تعد شاذة إذا إنحرفت عن التوزيع الأساسي وعلى سبيل المثال في حالة التوزيع الطبيعي فأن المشاهدة الشاذة تعتبر شاذة إذا كانت تبعد عن المتوسط بمقدار ثلاثة أمثال التباين.
2. المسافة: تعرف القيم الشاذة هي القيمة التي تقع على مسافة بعيدة من النسبة المئوية للبيانات.
3. التجمّع أو التكتّل: تعرف القيم الشاذة هي القيمة التي لا تلائم نمط التكتل.
4. الكثافة: يتم الكشف عن المشاهدة الشاذة من خلال المشاهدات.
5. العمق: المشاهدة الشاذة هي المشاهدة التي تقع في الطبقات الخارجية.

4-1 الكشف عن القيم الشاذة *Outlier Detection*:

يتم الكشف عن المشاهدات الشاذة في المتغيرات المستقلة في نموذج الانحدار الخطي المتعدد بواسطة المصفوفة H (the hat matrix) والتي تعرف بالصيغة الاتية [8] p.217.

$$H = x(x'x)^{-1}x'$$

فإذا كان $h_{ii} > \frac{2m}{n}$ or $h_{ii} > \frac{3m}{n}$ دل ذلك على أن المشاهدة (i) هي مشاهدة شاذة حيث أن (m) هي عدد المتغيرات المستقلة في النموذج و(n)هو حجم العينة، حيث أن h_{ii} هي عناصر القطر الرئيسي للمصفوفة (H).

2- طرائق التقدير

2-1 طريقة المربعات الصغرى في تقدير معالم نموذج الانحدار

تُعدّ طريقة المربعات الصغرى (Ordinary Least Squares: OLS) من أشهر طرق التقدير لتميّزها بخاصية أفضل تقدير خطي غير متحيز [1] p.7 (Beast Linear Unbiased Estimator: : BLUE). ولغرض تقدير معالم نموذج الانحدار:

$$Y_i = B_o + B_1X_{i1} + B_2X_{i2} + \cdots + B_kX_{ik} + U_i \quad ; \quad i = 1, 2, 3, \cdots, n$$

يتم ذلك من خلال تصغير مجموع مربعات البواقي الى أصغر ما يمكن:

$$Q = \sum_{i=1}^{n} e_i^2$$

أي يتم ذلك من خلال جعل المقدار (Q) أصغر ما يمكن ويتم ذلك عن طريق أخذ المشتقة الجزئية لكل معلمة على حدة ومساواتها الى الصفر:

$$\frac{\partial Q}{\partial B_i} = 0 \quad ; \quad i = 1, 2, \cdots, k$$

وبحلّ المعادلات نحصل على المعالم المقدرة وعليه فأنه مقدرات (OLS) يمكن الحصول عليها من الصيغة الاتية:

$$\hat{B}_{OLS} = (x'x)^{-1}x'y$$

إن نقطة الانهيار التي يمكن القول بأنها الحد الذي يصف مقدار مقاومة المقدر للبيانات الشاذة، لطريقة المربعات الصغرى تكون مساوية الى $\frac{1}{n}$ أي إنها لا تتحمل وجود الشواذ.

فإذا كان لدينا عينة بحجم (n) وكانت بيانات العينة كالاتي

$$Z = \{(x_{11}, \cdots, x_{1p}, y_1), \cdots, (x_{n1}, \cdots, x_{np}, y_n)\} \qquad \dots (2,14)$$

وكانت (T) تمثل مقدرات الانحدار أي أن $T(z) = \hat{\theta}$ فإذا تم إستبدال (m) من البيانات الاصلية حيث أن $0 \le m < n$ ببيانات أخرى عشوائية (هذه البيانات عبارة عن شواذ سيئة جداً) فأن البيانات الجديدة سيرمز لها بالرمز (Z') ، وكان مقدار التحيز $b(m;T,Z)$ الذي هو أكبر تحيز الذي يحصل بسبب التلوث

$$bias(m;T,Z) = sup_{Z'} \|T(Z') - T(Z)\| \qquad \ldots(2,15)$$

حيث أن الـsup معرفة على كل (Z') الممكنة، فإذا كان مقدار التحيز $bias(m;T,Z)$ غير محدد هذا يعني أن (m) من البيانات استطاعت أن تؤثر بشكل كبير على المقدر (T) والتي يمكن القول عنه بأن المقدر ينهار ويمكن التعبير عنها بالشكل الاتي:

$$\epsilon_n^* = min\left\{\frac{m}{n}; bias(m;T,Z)\ is\ infinite\right\}$$

2-2 طريقة المربعات الصغرى المشذبة (Least Trimmed Squares Method (LTS:

قدّم Peter Rousseeuw 1984عدة مقدرات حصينة من ضمنها طريقة وسيط المربعات الصغرى Least Median Squares (LMS) وطريقة المربعات المشذبة Least Trimmed Squares (LTS) . [10] p.876

أن طريقة (LTS) تشبه الى حد ما طريقة المربعات الصغرى الاعتيادية (OLS) إلا أن الفرق بينهما هو أن مربعات البواقي الكبيرة لا تدخل ضمن المجموع مما يسمح ببقاء القيم الجيدة بعيدا عن القيم الشاذة.

اذ يتم في طريقة (LTS) إيجاد $\beta \in R^p$ بحيث تجعل مجموع مربعات البواقي للمجموعة الجزئية (h) أقل ما يمكن، ويتم ذلك عن طريق إيجاد $\binom{n}{h}$ من العينات الجزئية وتُطبق على كل عينة جزئية طريقة (OLS) والعينة التي لها أقل مجموع مربعات بواقي هي التي يتم أستخدمها للحصول على المعالم المقدرة. ويمكن إيجاد المقدرات على النحو الآتي:

$$\hat{\theta}_{LTS} = min \sum_{i=1}^{h} (r^2)_{i.n}$$

حيث $(r^2)_{1.n} \le \cdots \le (r^2)_{i.n}$ هي مربعات البواقي (الأخطاء تربع ثم ترتب) :

$$h = \frac{n+p+1}{2}$$

وان (h) تقع ضمن الفترة: $\frac{n}{2} < h \le n$ ،أن (h) من البيانات لا يتم تشذيبها والباقي (n-h) تم تشذيبه.

إن نقطة الانهيار لطريقة (LTS) عندما تكون $h = [(n+p+1)/2]$ هي ، البرهان موجود في p.132 [8].

$$\epsilon_n^* = \frac{\frac{n-p}{2}+1}{n}$$

أن قيمة (h) المذكورة أعلاه تجعل نقطة الانهيار تبلغ أقصى حد p.132 [2] والذي يكون مساوي الى 50% أما إذا تجاوزت نقطة الانهيار هذا الحد فأنه لا يمكن التمييز بين الجزء الجيد عن الجزء غير الجيد (الشاذ) من البيانات. أما في حالة كون $(h = n)$ فأننا نحصل بذلك على مقدر (LS) والذي لديه نقطة انهيار تكون مساوية الى الصفر.

وبصورة عامة (h) قد تعتمد على نسبة التشذيب (α) وعلى سبيل المثال يمكن أن نعبر عنها على النحو الآتي:

$$h = n(1-\alpha) + \alpha(p+1)$$

أو

$$h = n(1-\alpha) + 1$$

حيث أن نقطة الانهيار تقريباً تعادل هذه النسبة (α).

عندما (α) تقترب من 50% سوف يتم الحصول على مقدرات (LTS) أما إذا اقتربت هذه النسبة (α) من الصفر سوف نحصل مقدرات (LS) p.134 [8].

إن مقدر (LTS) لديه مزايا عديدة على عكس مقدر (LMS) حيث أن دالة هدف (LTS) أكثر سلاسة من دالة هدف (LMS) مما يجعلها أقل حساسية للتأثيرات المحلية، وكفاءتها الإحصائية أفضل بسبب أن مقدر (LTS) هو محاذي طبيعي بينما مقدر (LMS) لديه معدل تقارب ضعيف ويساوي ($n^{-1/3}$) بينما معدل تقارب (LTS) هو ($n^{-1/2}$).

على الرغم من كل هذه المزايا الجيدة إلا أن طريقة (LTS) تطبق بصورة قليلة وذلك بسبب صعوبتها عند الحساب.

خصائص تساوي التغاير (Equivariance Properties):

توجد هناك ثلاثة أنواع من تساوي التغاير مرتبة من الأعلى الى الأسفل حسب الأولوية وهذه الخصائص هي:

1.انحدار متساوي التغاير (Regression Equivariant): يقال للمقدر بأنه انحدار متساوي التغاير إذا كان:

$$T(\{(x_i, y_i + x_i v)\ ; i = 1, \cdots, n\}) = T(\{(x_i\ , y_i)\ ; i = 1, \cdots, n\})$$

حيث أن v هو متجه عمودي، إن إنحدار متساوي التغاير هو ترجمة متساوي التغاير لمقدر الموقع في متعدد المتغيرات ولكن بصياغة مختلفة.

2.قياس متساوي التغاير (Scale Equivariant): يقال للمقدر T بأنه قياس متساوي التغاير إذا كان

$$T(\{(x_i, cy_i)\ ; i = 1, \cdots, n\}) = cT(\{(x_i, y_i)\ ; i = 1, \cdots, n\})$$

لأي ثابت c ، هذا يؤدي الى أن الملائمة بشكل أساسي مستقلة عن اختيار وحدة القياس لمتغير الاستجابة y.

3.انحدار قياس متساوي التغاير (Affine Equivariant): يقال للمقدر T بأنه انحدار قياس متساوي التغاير إذا كان:

$$T(\{(x_i\ , A, y_i); i = 1, \dots, n\} = A^{-1}T\{(x_i\ , A, y_i); i = 1, \dots, n\})$$

لأي مصفوفة A غير مفردة (Non- singular)، حيث يمكننا القول بأن انحدار قياس متساوي يعني أن التحويل الخطي لـ x_i سيحول المقدر T تبعا لذلك بسبب أن

$$\hat{y}_i = x_i T = (x_i A)(A^{-1} x_i)$$

هذا يسمح لنا باستخدام نظام إحداثي أخر للمتغيرات التوضيحية من دون أن تؤثر على القيمة التقديرية لـ y_i.

إن مقدر (LTS) يحقق الثلاث أنواع المذكورة أعلاه أي أنه انحدار متساوي التغاير وقياس متساوي التغاير وانحدار قياس متساوي التغاير على النحو التالي: انحدار متساوي التغاير إذا كان:

$$\sum_{i=1}^{h} ((y_i + x_i v - x_i\{v + \theta\})^2)_{i:n} = \sum_{i=1}^{h} ((y_i - x_i\theta)^2)_{i:n}$$

لأي متجه عمودي v ، قياس متساوي التغاير وانحدار قياس متساوي التغاير تكون مماثلة.

إن مقدر (LTS) ينتمي الى عائلة مقدرات انحدار قياس متساوي التغاير(affine equivariant) [4].

3-2 المربعات الصغرى المشذبة الموزونة (Weighted Least Trimmed Squares) :p.8 [8]p.17[3]

لغرض زيادة كفاءة مقدرات (LTS) فأنه من الممكن حساب مقدرات موزونة لهذه الطريقة، وينبغي أن يجمع هذا الاجراء بين حصانة المقدرات الابتدائية (الحصينة) والدقة العالية في التقدير.

افترض أن $\hat{B}_o$ و $\hat{\sigma}_o$ هي المقدرات الابتدائية (الحصينة) التي تم الحصول عليها من خلال طريقه (LTS) لمعلمات نموذج الانحدار والانحراف المعياري للبواقي على التوالي، إن الهدف هو بناء أوزان صعبة الرفض تحدد أي من الحالات يجب أن يتم وزنها أو تشذيبها (أي إن الحالة i التي سوف تشذب يكون وزنها مساوى الى الصفر $w_i = 0$). ويتم تحديد الحالة الشاذة عن طريق البواقي المعيارية باستخدام الطريقة الحصينة

$$\left|\frac{r_i}{\hat{\sigma}}\right|$$

اذ أن

$$r_i = y_i - \hat{y}_i$$

$$\hat{\sigma} = \frac{1}{c_1}\sqrt{\frac{1}{n}\sum_{i=1}^{h}(r_i)^2}$$

حيث أن $\hat{\sigma}$ هو الانحراف المعياري للبواقي بالنسبة الى طريقة (LTS) و(c_1) هو معامل تصحيح، أن الحالة (i) ستعتبر حالة شاذة إذا وفقط إذا كانت البواقي المعيارية لها كبيرة (أن هذه النسبة لا تعتمد على وحدات القياس الاصلية).

وبعد أن تم تحديد الحالات الشاذة يمكن أن نطبق طريقه المربعات الصغرى الموزونة (WLS) ويكون الوزن على النحو الاتي:

$$w_i = \begin{cases} 1 & if \ \left|\frac{r_i}{\hat{\sigma}}\right| \le 2 \cdot 5 \\ 0 & if \ \left|\frac{r_i}{\hat{\sigma}}\right| > 2 \cdot 5 \end{cases}$$

إن الطرف الايسر من المتباينة والمساوي الى (2.5) هو اعتباطي ولكنه معقول اذ أنه في حالة التوزيع الطبيعي فأن هناك عدد قليل من البواقي يكون أكبر من $\hat{\sigma}\ (2 \cdot 5)$. إن المقدرات الناتجة لا تزال تمتلك نقطة الانهيار العالية، ولكنها ذات كفاءة عالية.

3- الجانب التطبيقي

3-1 وصف البيانات والعينة

تم الحصول على عينة بحجم 91 من مختبر دائرة مجاري واسط حيث هذا العدد يمثل المحطات التي توجد في مدينة الكوت حصرا وكانت هذه العينة تحتوي على متغير معتمد (TDS) وثلاث ملوثات مستقلة وهي (PO4,CL,SO4), حيث تم سحب بيانات هذه العينة من قبل المختصين وتم سحبها من مواقع مختلفة لمدينة الكوت لعام 2015.

المواد الذائبة الكلية Total Dissolved Solids (TDS) [13]:

المواد الصلبة الذائبة هي كافة المواد التي تمر من خلال أوراق الترشيح المختبري مشتملة على المواد الغروية وتشكل المواد الصلبة الذائبة في المياه حوالي 90% والباقي على شكل غروي. تشكل المواد العضوية حوالي 40% من أجمالي المواد الصلبة الذائبة والباقي غير عضوي والجزء الغروي يكون في الغالب عضوي التركيب. تعد المواد الصلبة الذائبة من أهم معايير صلاحية المياه والري أن زيادة تركيز هذه المواد يؤثر على نمو النباتات اذ تحد من كميات المياه التي تمتصها الجذور لتقوم بعملية النتح وتقلل ومن ثم كمية المياه المنتقلة الى الأوراق والمياه المالحة تؤدي الى تراكم الاملاح في التربة مما يستوجب غسل التربة بعد كل موسم.

الكلوريدات Chloride(CL) [13]:

تعد الكلوريدات من الايونات السالبة المهمة الموجودة في المياه الطبيعية وعند تواجدها بتراكيز عالية يكتسب الماء تأثير تأكلي قد تظهر على الانابيب والمنشأة المعدنية، ويتم فحص الكلوريدات عند استخدام الماء للري في المياه المعالجة وذلك لتأثيره على أوراق المحاصيل الحساسة ويكون على شكل املاح الصوديوم والبوتاسيوم والكالسيوم ويعطي الطعم المالح إذا ارتبط مع أيون الصوديوم وشكل ملح كلوريد الصوديوم (ملح الطعام) NaCl ، ولا يعطي الطعم المالح إذا ارتبط مع الايونات الأخرى مثل الكالسيوم وأيون المغنيسيوم، وتحتوى فضلات الانسان مثل العرق والادرار على (6) غرام من الكلوريدات للشخص الواحد في اليوم بسبب ملح الصوديوم الذى لا يتغير عبر مروره في الجهاز الهضمي لذلك ينطرح مع الادرار، وإن مصدر الكلوريدات تأتى من الترسبات والصخور (أي من تحلل الصخور الموجودة في الأنهر وتحلل الرمال ومن ادرار الانسان) ولا يمكن تقدير نتيجة الكلوريدات من اللون إلا من الحجم.

الكبريتات Sulphate SO4 [13]:

أن وجود الكبريتات في المياه نتيجة المخلفات الصناعية منها السيارات، الأسمدة، الكيمياويات الغير عضوية، البلاستك، تكرير البترول، الصلب، الصناعات الورقية. إن تأثير الكبريتات في المياه يسبب العسرة الدائميه وتكون على شكل كبريتات الكالسيوم والمغنيسيوم. وتحديد صلاحية المياه للري والبناء (ضروري لنمو النبات، ويؤثر على نوعية الكونكريت ويعمل على تفتته) ويعطي رائحة غير مقبولة للماء ويسبب تأكل الانابيب.

الفوسفات Phosphate(PO4) [13]:

إن المصدر الرئيسي للفوسفيت في المياه العادمة هو الأسمدة ومساحيق الغسيل وللكشف عنها يتم الاعتماد على اللون حيث أن شدة اللون الأصفر تتناسب مع تركيز الفوسفيت الموجود في المياه العادمة ويستخدم جهاز سبكتروفوتوميتر لقياس الامتصاصية أو النفاذية ومن خلالها نستطيع أن نعرف تركيز الفوسفيت باستخدام الرسوم البيانية المعدة مسبقا. حيث أن الفوسفات يفيد في تقييم الطاقة البيولوجية للمياه السطحية وتشغيل محطات التنقية ودراسة تلوث الجداول.

ان الهدف من هذا الاجراء هو معرفة تأثير المواد الصلبة الذائبة (الاملاح) في مياه الصرف الصحي والتي تشمل الكبريتات والكلوريدات والفوسفات والتي تصبها محطات الصرف الصحي في مياه النهر مما يؤثر على الانسان والنبات والتربة.

2-3 النتائج

بعد أن تم الحصول على البيانات من الدائرة المذكورة بصورة مسبقة، تم تحليلها والحصول على النتائج الاتية.

معالم النموذج ونقطة الانهيار عند كل طريقة

Param. / Method		b_o	b_1	b_2	b_3	B.P.
OLS		-4.4179	1.0845	27.6734	2.5566	0.0110
		(1.5490) *	(0.1031) *	(8.8522) *	(0.1613) *	
LTS		-1.9541	1.0371	20.7881	2.2445	0.4890
		(0.8956) *	(0.2723) *	(3.3526) *	(0.1128) *	
WLTS	C=1.5	-1.1879	1.2527	18.2203	2.0501	0.4890
		(0.7445) *	(0.2245) *	(2.4883) *	(0.1260) *	
	C=2.1	-0.3088	1.2197	16.8118	1.9162	
		(0.7991) *	(0.2391) *	(2.8910) *	(0.1409) *	
	C=2.3	-0.1402	1.2204	15.2296	1.9145	
		(0.8776) *	(0.2626) *	(3.1514) *	(0.1540) *	
	C=2.5	-1.1829	0.8966	18.2565	2.2593	
		(0.8614) *	(0.2557) *	(3.2015) *	(0.1042) *	

***: تمثل الانحراف المعياري لكل معلمة عند كل طريقة**

- **في الجدول أعلاه يلاحظ الانحراف المعياري لكل معلمة بإستخدام طريقة LTSهو أصغر من الانحراف المعياري لكل معلمة بإستخدام طريقة OLS وهذا يدل على أن طريقة LTS تخلصت من القيم الشاذة وكانت أكثر مقبولية.**
- **عند النظر الى نتائج طريقة WLTS وبالتحديد عند نقطة القطع (C=1.5) نلاحظ ان الانحراف المعياري لكل معلمة هو اقل من الانحراف المعياري عند طريقة LTS وذلك لأن عند نقطة القطع (C=1.5) تم التخلص من المشاهدات التي كانت لها النسبة $\left|r_i/\hat{\sigma}\right|$ أكبر من 1.5 أي قللت تشتت البيانات وأصبحت البيانات أكثر تجانس.**

تباين الخطأ والنموذج ومعيار الكفاءة وRMSE

Method \ Param.		$v(\hat{y})$	$v(e)$	*RMSE*	(eff_{model})*
OLS		658.7724	30.4359	5.4865	100
LTS		26.6867	1.7318	1.3022	4.0510
WLTS	C=1.5	13.5262	0.7238	0.8378	2.0532
	C=2.1	13.6372	1.0434	1.0092	2.0701
	C=2.3	12.9844	1.2702	1.1144	1.9710
	C=2.5	27.7671	1.4754	1.2014	4.2150

*: تمثل كفاءة النموذج عند القيم التقديرية للنموذج $(\hat{y})$

- يلاحظ من خلال الجدول أعلاه ان تباين الخطاء عند طريقة WLTS عند نقطة القطع (C=1.5) هو اقل ما يمكن ومن ثم فان أفضل نتائج هي نتائج طريقة WLTS عند نقطة القطع (C=1.5).

باستخدام المصفوفة ***h*** تم الكشف عن القيم الشاذة في المتغيرات المستقلة اذ تم الكشف عن كل متغير بصورة منفردة لمعرفة القيم الشاذة في كل متغير وأيضا معرفه قيمتها، وتم وضع هذه القيم الشاذة في الجدول التالي.

القيم الشاذة في كل متغير بواسطة المصفوفة ***h***

cl	So4
8295	5102
4109	2839
3446	

1-4 الاستنتاجات

1. إن النتائج التي تم الحصول عليها بطريقه LTS أفضل من نتائج طريقة OLS حيث يمكن ملاحظة ذلك من خلال معيار RMSE، ويمكن ملاحظة ذلك أيضا من خلال كفاءة النموذج.
2. إن استخدام الاوزان أدت الى تحسين التقدير ويمكن ملاحظة ذلك من خلال معيار RMSE وأفضل نتيجة هي عندما كانت نقطة القطع مساوية الى C=1.5.

2-4 التوصيات

1. تكثيف الاهتمام حول وجود القيم الشاذة في بيانات الظاهرة المدروسة وذلك بسبب صعوبة توفر الظروف المثالية للظاهرة المدروسة نتيجة التغيرات التي تحدث حول هذه الظاهرة.
2. إستخدام طرائق التقدير الحصينة الأخرى على سبيل المثال طريقة (MML) Maximum Median Likelihood، وكذلك معرفه خصائص مقدرات هذه الطرائق وكذلك حساسيتها وتأثرها تجاه وجود الشواذ.
3. التقليل من إستخدام المواد المسببة للتلوث وعدم إستخدامها بكثرة للحفاظ على البيئة.
4. زيادة كفاءة محطات معالجة مياه الصرف الصحي في معالجة التلوث والاهتمام بها من ناحية الأجهزة المختبرية والمعدات اللازمة والمتطورة لعمل هذه المحطات على أتم وجه، ومحاوله التقليل من وجود هذه الملوثات وعدم رميها في مياه الأنهر بصورة مباشرة الا بعد التخلص من هذه الملوثات الضارة.
5. زيادة الوعي البيئي ويتم ذلك من قبل مختصين في مجال البيئة والصحة حيث يتم ذلك عن طريق عمل دورات تثقيفية في مجال المحافظة على البيئة أو عن طريق عمل بعض المنشورات أو الإعلانات الدعائية والتي تهدف الى زيادة الوعي البيئي لدى كافة الناس.

المصادر

[1] كاظم ، أموري هادي "القياس الاقتصادي , "عمان .دار زهران ،2009.

[2] A. Giloni, M. Padberg, "Least Trimmed Squares Regression, Least Median Squares Regression, and Mathematical Programming," Math. Comput. Model., vol. 35, no. 2, pp. 1043–1060, 2002.

[3] C. Pavel," Reweighted Least Trimmed Squares: An Alternative to One-Step Estimator", (Center Discussion Paper; Vol. 91). Tilburg University: Econometrics, 2010.

[4] E. B. Balogun, X. Huang, Y. Lin, M. Liao, and M. F. Adaramola, "Regression Estimation Modelling Techniques on Static Solar Photovoltaic Module," vol. 5, no. 4, pp. 451–461, 2015.

[5] M. Hawkins, "Identification of Outliers", Chapman & Hall in 1980.

[6] M. L. Tiku and A. D. Akkaya, "Robust Estimation and Hypothesis Testing." 2004.

[7] P. J. Huber. "Robust Regression: Asymptotics Conjectures and Monte carlo", Jstor,1973.

[8] P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection." p. 347, 1987.

[9] P. R. Freeman, "On the Number of Outliers in Data from a Linear Model," Trab. Estad. Y Investig. Oper., vol. 31, no. 1, pp. 349–365, 1980.

[10] P. Rousseeuw, "Least Median of Squares Regression," J. Am. Stat. Assoc., vol. 79, no. 388, pp. 871–880, 1984.

[11] V. Hautam , K. Ismo,F Pasi "Outlier Detection Using k-Nearest Neighbour Graph".

[12] W. J. A. K. H. T. J. Han, "Mining Top-n Local Outliers in Large Databases," Proc. seventh ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '01, pp. 293–298, 2001.

[13] American Public Health Association, & American Water Works Association. (1989). *Standard methods for the examination of water and wastewater*. American public health association.

Weighted Least Squares Estimation of the Effect of Wastewater Pollution of Tigris River / Wasit Governorate

Abstract

The analysis of Least Squares: LS is often unsuccessful in the case of outliers in the studied phenomena. OLS will lose their properties and then lose the property of Beast Linear Unbiased Estimator (BLUE), because of the Outliers have a bad effect on the phenomenon. To address this problem, new statistical methods have been developed so that they are not easily affected by outliers. These methods are characterized by robustness or (resistance). The Least Trimmed Squares: LTS method was therefore a good alternative to achieving more feasible results and optimization. However, it is possible to assume weights that take into consideration the location of the outliers in the data and determine them accurately. In order to increase the Weighted Least Trimmed Squares: WLTS, the weight of the sample data on the estimation is repeated. In order to perform this research, the need for detection and investigation of the impact of pollution of the Tigris River in Wasit Governorate has been called for by wastewater, particularly Total Dissolved Solids: TDS as dependent variable, and the impact of three covariates Sulfates: SO_4, Chloride: Cl and Phosphate: PO_4 pollutants. The evaluation was done in in a precise manner and submitted to the competent authorities. In order to achieve this objective, a sample of (91) positions were drawn and checked in the laboratories of Wasit Governorate.

Keywords: Least squares, Outliers, Trimmed Methods, River's Pollution.