استكشاف ومعالجة تأثير القيم الشاذة في نماذج اتخاذ القرارات الخطية غير المقيدة و نماذج الانحدار الخطي (مع التطبيق العملي)

م. رشيد بشير رحيمه
جامعة ذي قار/ كلية الإدارة والاقتصاد/ قسم الإحصاء
م. خولة عبد الحسين سوير
جامعة بغداد كلية الهندسة / قسم الميكانيك
م.م. واثق حياوي لايذ
كلية الإدارة والاقتصاد/قسم الإحصاء

الخلاصة

إن وجود القيم الشاذة ضمن مجموعة من البيانات يؤثر بشكل كبير على نتائج التحليل الإحصائي للبيانات وبالتالي على عملية اتخاذ القرار المناسب لذلك لابد من دراسة هذه القيم وطرق الكشف عنها وتقديرها .هذه المشكلة تم دراستها في بعض مسائل الانحدار الخطي ونماذج البرمجة الخطية ولكنها لم تلقى الاهتمام نفسه في نماذج البرمجة الخطية غير المقيدة (تحتوي نماذجها مغيرات حرة) والتي تعتبر من أهم المواضيع في بحوث العمليات لاستخداماتها المتعددة ولقدرتها في التعامل مع متغيرات قرار غير مقيدة بإشارة (يسمح للمتغير ان يكون سالبا أو موجبا أو صفرا) . تم في هذا البحث استكشاف وتحديد القيم الشاذة في تلك النماذج وكيف يمكن لهذه القيم أن تؤثر على الحل الأمثل وبالتالي على اتخاذ القرار المناسب . وقد استعملت طريقة المصفوفة المقدرة لتحديد القيود الشاذة ومعالجتها ليصبح النموذج المعتمد أكثر ملائمة لتحقيق الأهداف . حيث امتازت هذه الطريقة بالبساطة وسهولة الحساب فضلا عن الوضوح في تحديد المشاهدات الشاذة .

1- المقدمة :-

إن لعلم التخطيط دور بارز في تنمية البلدان نظرا لما يعكسه هذا العلم من جوانب ايجابية في عملية نجاح سياسات هذه البلدان في مختلف المجالات الاقتصادية والاجتماعية وما إلى ذلك. و تعتبر البرمجة الخطية غير المقيدة من المواضيع المهمة في بحوث العمليات للاستعمال الواسع لنماذجها في الشركات والمصانع الإنتاجية والسيطرة على الخزين وغيرها من لنماذج التي تحتوي متغيرات قرار غير مقيدة با شارة . [15],[5],[4],[3] ويعد موضوع الانحدار من المواضيع التي أصبحت ذات تطبيقات واسعة من قبل المهتمين في مختلف العلوم الاجتماعية والاقتصادية، لأنه يصف العلاقة بين المتغيرات على هيئة معادلة .[16], [5] ونظراً للتطور العلمي الحاصل في العديد من المجالات العلمية التي تستخدم جوانب النظرية الإحصائية كأدوات في تطبيقاتها العلمية وانعكاس تطور تلك النظرية عليها ، برزت أهمية جمع البيانات وحاجة الباحثين لهذه البيانات في عملية التحليل بجوانبه المختلفة بدءاً من المشكلة ثم بتقدير النموذج الملائم لها مروراً باختبار الفرضيات والمزاعم حولها وأنتهاءاً باتخاذ القرارات بناءاً على ذلك ، وللوقوف على أهمية التحليل الإحصائي في عملية التخطيط واتخاذ القرارات والتنبؤ بما يفيد الدراسة ، اهتم الباحثون بنوعية البيانات ، هذه البيانات إما أن تكون ممثلة للمشكلة بشكل صحيح ودقيق وبالتالي تعطي نتائج صحيحة أو أن تكون أقل تمثيلاً للمشكلة وذلك في حالات، من أبرزها أن تحتوي البيانات على قيم شاذة Outliers متأتية من خارج مجتمع الظاهرة تؤثر على عملية التحليل وبالتالي على اتخاذ القرار.[10], [5],[4],[2]

ذهب الباحثون بعد تشخيص مشكلة وجود القيم الشاذة ، للتعامل معها من جوانب مختلفة فمنهم من لاحظ ان هذه القيم تؤثر على النتائج وبالتالي أقترح حذف المشاهدات الشاذة وأجراء التحليل على بقية المشاهدات ، ولهذا الجانب مذاهب وطرق، ومنهم من تعامل مع الشواذ من خلال التعامل مع الطرق الحصينة Robust حيث تكون مقدرات هذه الطرق وإحصاءات الاختبار الخاصة بها قليلة الحساسية والتأثر نسبياً تجاه الشواذ عند وجودها ، وآخرين ذهبوا إلى أهمية البيانات على الرغم من وجود الشواذ لأن أي نقص في المشاهدات تعني نقص في تلك المعلومات الأمر الذي سينعكس سلبا على دقة النتائج الإحصائية عند حذف هذه القيم . [13], [12], [4],[5]

2- هدف البحث

دراسة نماذج البرمجة الخطية غير المقيدة ونماذج الانحدار الخطي وتأثير القيم الشاذة على تلك النماذج مع توضيح عملية اكتشاف تلك القيم بطريقة المصفوفة المقدرة ومعالجتها للوصول الى النموذج الأمثل .

3- نماذج الانحدار الخطية Linear Regression Models [1], [4],[5],[6],[20]

تلك النماذج التي تظهر فيها المعلمات المجهولة بصورة خطية وتكتب بصورة عامة كما يلي:

$$Y = \sum_{i=0}^{n} \beta_i X_i + \in$$

اذ ان $X_0 = 1$ ، وتشمل :-

3.1-نموذج الانحدار الخطي البسيط Simple Linear Regression Model

هو النموذج الذي يوضح العلاقة الخطية بين متغير الاستجابة (Y_i) Response Variable ومتغير توضيحي واحد (X_i) Explantory Variable ويمكن التعبير عن العلاقة بالشكل :

$$Y_i = \beta_0 + \beta_1 X_i + \in_i \qquad , \quad i = 1,2,\ldots..,n$$

اذ ان $\in_i$ يمثل الخطأ العشوائي

3.2 – نموذج الانحدار الخطي المتعدد Multiple Linear Regression Model

هو النموذج الذي يوضح العلاقة الخطية بين متغير الاستجابة (Y_i) و K من المتغيرات التوضيحية (X_j) ويمكن التعبير عن العلاقة بصورة عامة كالأتي:

$$\underline{Y} = X\underline{\beta} + \underline{U} \qquad , \; i = 1,2,\ldots.,n; \; j = 1,2,\ldots...,k$$

اذ ان

$\underline{Y}$: متجه لمشاهدات متغير الاستجابة وذات بعد (nx1)

X : مصفوفة لمشاهدات المتغيرات التوضيحية وذات سعة ((nx(k+1)) .

$\underline{\beta}$: متجه لمعلمات النموذج وذات بعد ((k+1)x1)).

$\underline{U}$: متجه للأخطاء العشوائية وذات بعد (nx1)

3.3- الصيغة الرياضية العامة للبرمجة الخطية غير المقيدة[14] ,[6],[3]

يمكن وضع صيغة ثابتة للبرنامج الخطي غير المقيد بالإشارة حيث يتألف البرنامج الخطي من نموذج رياضي يتضمن دالة الهدف **Z** والتي قد تكون بحالة تكبير (Max) أو تصغير

(Min) والقيود التي من الممكن إن تأخذ العلاقة الآتية($\geq$ ، = ، $\leq$) وإن بعض المتغيرات او جميع المتغيرات x_j المطلوب اتخاذ القرار بشأنها تكون غير مقيدة بإشارة (يسمح للمتغيرات ان تكون سالبة او موجبة او اصفار) ويمكن تمثيل هذه الصيغة كالأتي:

Max or (Min) Z $= c_1x_1 + c_2x_2 + \dots + c_nx_n$ دالة الهدف

Sub. to

$$\left.\begin{array}{l}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \ (\leq \ = \geq) \ b_1 \\
a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \ (\leq \ = \geq) \ b_2 \\
. \qquad\qquad\qquad\qquad\qquad . \\
. \qquad\qquad\qquad\qquad\qquad . \\
. \qquad\qquad\qquad\qquad\qquad . \\
a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \ (\leq \ = \geq) \ b_m \\
x_1, x_2, \dots, x_n \ \text{unrestricted in signs}
\end{array}\right\}$$

القيود

حيث ان a_{ij}, b_i, c_j ثوابت تحدد سياق المسألة

$j = 1, 2, \dots, n, \ i = 1, 2, \dots, m$

وان x_j متغيرات المسألة المطلوب اتخاذ القرار بشأنها

وبالإمكان وضع الصيغة العامة للبرمجة الخطية المعرفة أعلاه بصيغة المصفوفات بالشكل الأتي:-

$$\begin{array}{l}
\text{Max or (min) } \mathbf{Z} = CX \\
\qquad AX \ (\leq \ = \geq) \ B \\
\qquad X \quad \text{unrestricted in signs}
\end{array}$$

حيث إن

$X = (x_1, x_2, \dots, x_n)'$

$B = (b_1, b_2, \cdots, b_m)'$

$C = (c_1, c_2, \dots, c_n)$

وان $A=[a_{ij}]$ مصفوفة بأبعاد $m \times n$ $(j = 1, \dots, n; \ i = 1, \dots, m)$

X: تمثل متجه المتغيرات الذي يمثل حل المسألة.

A : مصفوفة معاملات القيود.

X: متجه يمثل معاملات دالة الهدف.

B: متجه يمثل قيم متغيرات الحل الأساس للمسألة.

أما التعبير CX فهو يمثل دالة الهدف.

ويمكن ايجاد الحل الأمثل لنموذج البرمجة الخطية غير المقيدة باتباع طريقة السمبلكس الاعتيادية أو المعدلة .

ان قيود نموذج البرمجة الخطية الممثل بالمقدار

A X = b

يمثل بشكل عام معادلة الانحدار (والتي تكتب عادة بشكل XB = Y) لذلك يمكن إيجاد الحل لهذه المعادلة كما نجده لأي معادلة انحدار أو نجده بطريقة السمبلكس أو السمبلكس المعدلة على وفق نموذج البرمجة الخطية كما سنلاحظ ذلك لاحقاً .

وعند الحل بطريقة نموذج الانحدار سيتم تحديد معاملات الانحدار B لتحدد العلاقة بين متغيرات النماذج المستقلة والمتغير التابع لها . ان تحديد تلك العلاقة يسمى تحليل الانحدار .

وتحليل الانحدار عبارة عن وسيلة إحصائية تتبع لتحديد العلاقة بين متغير مستقل واحد او أكثر والمتغير التابع .

ان B هي الحل لمعادلة الانحدار (XB = Y) وهي تقابل الحل X في صيغة البرمجة الخطية (AX =b) المستخرج بطريقة السمبلكس او السمبلكس المعدلة لكن هذا لا يعني تطابق الحلين دائماً إذ ان حل الطريقتين يتطابق فقط عندما تحمل القيود علامة المساواة وهذا يعني ان قيم المتغيرات المهملة S_i ستكون اصفاراً في جدول الحل الأمثل وقد لا تحمل القيود علامة المساواة ويتطابق الحل وهذا سيعني ان المتغيرات المهملة هي أصفار ويمكن التحقق من ذلك لو تم الحل بطريقة السمبلكس او السمبلكس المعدلة اذ ستكون المتغيرات المهملة اصفاراً في جدول الحل الأمثل ولتوضيح ذلك سنأخذ المثال الأتي :

مثال (1)

$$Min\, Z = 75X_1 + 120X_2$$

$$S.t$$
$$50X_1 + 100X_2 \geq 4200$$
$$100X_1 + 150X_2 \geq 6000$$
$$X_1 ، X_2 \text{ unrestricted in signs}$$

تم حل النموذج أعلاه باستخدام برنامج (win QSB) فكانت النتائج كما يلي :-

$X_1 = -12$ $X_2 = 48$ $S_1 = 0$ $S_2 = 0$

أما عند تطبيق معادلة الانحدار XB=Y المقابلة لمعادلة قيود البرمجة الخطية AX=b .

$$\therefore X = \begin{bmatrix} 50 & 100 \\ 100 & 150 \end{bmatrix} \quad , \quad Y = \begin{bmatrix} 4200 \\ 6000 \end{bmatrix}$$

$$\because \hat{B} = (\grave{X}X)^{-1}\grave{X}Y \quad . \quad \hat{B} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\because \grave{X}X = \begin{bmatrix} 50 & 100 \\ 100 & 150 \end{bmatrix}\begin{bmatrix} 50 & 100 \\ 100 & 150 \end{bmatrix} = \begin{bmatrix} 12500 & 20000 \\ 20000 & 32500 \end{bmatrix}$$

$$(\grave{X}X)^{-1} = \begin{bmatrix} 0.0052 & -0.0032 \\ -0.0032 & 0.002 \end{bmatrix}$$

$$\therefore \hat{B} = \begin{bmatrix} 0.0052 & -0.0032 \\ -0.0032 & 0.002 \end{bmatrix}\begin{bmatrix} 50 & 100 \\ 100 & 150 \end{bmatrix}\begin{bmatrix} 4200 \\ 6000 \end{bmatrix}$$

$$= \begin{bmatrix} -12 \\ 48 \end{bmatrix}$$

$$X_1 = -12 \qquad X_2 = 48$$

وكما نرى فأن الحل متطابق لكلا الطريقتين لان قيم المتغيرات المهملة (S_1 , S_2) في جدول الحل الأمثل بطريقة السمبلكس كانت اصفاراً .

4- بعض الدراسات الخاصة بموضوع البحث

ان موضوع القيم الشاذة من المواضيع المهمة والذي تناوله الباحثين في مختلف المجالات ومنها الإحصاء وبحوث العمليات وسنذكر هنا بعض الدراسات والبحوث التي تتعلق بموضوع البحث .

قام(نبيل ناسي)عام 2001 بتقييم كفاءة طرق تقدير القيم الشاذة لنماذج الانحدار [7]

في عام 2001 درس (Milan M. , Militký J.) النقاط الشاذة في نماذج الانحدار التربيعي الاعتيادية .[18]

في عام 2005 اقترح كل من(Baddeley و Turner و Moller و Hazelton)طرق إحصائية جديدة في تحليل البواقي باستخدام الرسم البياني، إضافة إلى تشخيص المشاهدات المؤثرة. [11]

في عام 2005 درس (انكين) القيم الشاذة في بعض نماذج الانحدار ألا خطية والمقارنة بينهما ضمن نطاق معين من التطبيقات .[5]

وفي عام 2005 قدم كل من (Wei و Kosorok) طريقة لنوع من أنواع الدوال الحساسة التي لها أمكانية تبيان المشاهدة التي تساهم بقدر في تأثير الإخفاء أثناء عملية تشخيص القيم الشاذة داخل نماذج الانحدار. [23]

في عام 2006 اهتم (رحيم جبار ظاهر) بدراسة الكشف عن القيم الشاذة في بيانات تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة وكيف يمكن لهذه القيم ن تؤثر على التحليل الإحصائي للتجربة . [8]

في عام 2007 اقترحت (رشا جلال متلف) بعض الطرق لتحديد القيود الشاذة في مسائل البرمجة الخطية ونماذج الانحدار الخطي .[6]

في عام 2007 تناول (هاني عبد الله حسن) مسالة احتساب تأثيرات الإغراق والإخفاء الناجمة عن وجود مشاهدات شاذة .[4]

قارن كل من(Arezoo B., Habshah M., Mojtaba G. and Samaneh E.)عام 2010 العديد من الطرق للكشف عن النقاط الشاذة في نماذج الانحدار الخطية . [9]

5- إيجاد النقاط أو القيود الشاذة بطريقة المصفوفة المقدرة (Hat Matrix)

عند تقدير النماذج الخطية بطريقة المربعات الصغرى يمكن دائماً معرفة كم هو ابتعاد النقاط Y عن قيمتها المقدرة $\hat{Y}$ وكذلك يمكن ملاحظة النقاط المتطرفة جداً أو الشاذة عن بقية النقاط الأخرى .

إن المصفوفة المقدرة H العائدة إلى مبتكر هذه التقنية John W .Tukey سنة 1972 بإمكانها أن تحدد لنا النقاط المتطرفة أو الشاذة لذلك سنورد في البند اللاحق اشتقاق المصفوفة H من معادلة الانحدار .

5.1- إيجاد المصفوفة المقدرة H [21], **[1],[6]**

إن معادلة الانحدار بشكل عام ممثلة بالمعادلة

$$Y = XB + e \qquad \ldots(2.1)$$

وتقدير Y سيكون

$$\hat{Y} = X\hat{B} \qquad \ldots(2.2)$$

حيث إن e تمثل الخطأ العشوائي موزعاً توزيعاً طبيعياً بمتوسط مقداره صفر وتباين مقداره σ^2

إن تقدير B هو

$$\hat{B} = (X'X)^{-1}X'Y$$

ولو عوضنا $\hat{B}$ بالمعادلة (2.2) سيكون لدينا

$$\hat{Y} = X(X'X)^{-1}X'Y \qquad \ldots(2.3)$$

إذن

$$\hat{Y} = HY \qquad \ldots(2.4)$$

حيث

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

أي إن H قد قدرت Y في المعادلة (2.4) لذا فأن H تسمى المصفوفة المقدرة (بكسر الدال) إلى Y .

فإذا كان X_{nxp} حيث n تمثل عدد الصفوف و p تمثل عدد المتغيرات .

فأن المصفوفة المقدرة ستكون H_{nxn} وان عناصرها تمثل مصفوفة تغاير $\hat{Y}$ (covariance) وان مقدار الخطأ e لكل المشاهدات هو

$$e = Y - \hat{Y}$$

$$Var(\hat{Y}) = \sigma^2 H \qquad , \quad Var(e) = \sigma^2 (1-H)$$

وان العناصر القطرية من H (أي h_{ii}) تعطينا المؤشر لكل عنصر من عناصر مشاهدات Y بحيث :

$$\sum_{i=1}^{n} \mathbf{hii} = P$$

5.2–اختبار معنوية القيود والنقاط الشاذة [1], [2],[6] ,[22]

هنالك عدة اختبارات بإمكانها أن تحدد لنا شذوذ أو عدم شذوذ النقطة أو القيد بعضها يعطي مؤشراً أولياً على الشذوذ مثل الاختبار الأول (اختبار Thumb) أما إذا أردنا التأكد أكثر فبالإمكان تطبيق الاختبار الثاني (اختبار F) حيث سيؤكد لنا هذا الاختبار نتيجة الاختبار الأول أو ينفيه أما الاختبار الثالث (Cook's Di) فسيكون تأكيداً لما توصل إليه الاختبار الأول والثاني او نفيهما . $\frac{2p}{n}$،والاختبارات الثلاثة هي كالأتي :–

الاختبار الأول:

إن h_{ii} (أو للسهولة سنقول h_i) التي تتجاوز $\frac{2p}{n}$ ستكون هي القيمة ذات التأثير العالي في القطر والتي تقابل السطر (القيد) الشاذ والمقابل للمتغير التابع Y . يعد هذا الاختبار مؤشراً أولياً على شذوذ هذا القيد أو هذه النقطة فإذا كانت h_i اكبر بكثير عن $\frac{2p}{n}$ فان احتمال الشذوذ كبير وكما يقول المثل فالعنصر الشاذ يشار له بالبنان أو بالأصبع لذلك يسمى هذا الاختبار rule of the thumb أو قانون الإبهام .

وإذا كان قليلاً فان احتمال الشذوذ سيكون خفيفاً ويمكن التأكد من هذا الشذوذ بتطبيق الاختبار الثاني أما إذا لم يكن اكبر من $\frac{2p}{n}$ فهذا يعني إن h_i ليست شاذة .

الاختبار الثاني: يمكن تطبيق اختبار F للتأكد من قيم hi المعنوية (الشاذة) اذ سيتم إيجاد

$$F_{Ci} = \frac{\left(h_i - \frac{1}{n}\right)/(P-1)}{(1-h_i)/(n-P)} \quad \sim \quad F_t\,\alpha\ \ (p-1\ ,\ n-p)$$

يسمى الطرف الأيمن F_c القيمة المحسوبة إلى F

ويسمى الطرف الأيسر F_t قيمة F الجدولية

حيث: $h_i \neq \frac{1}{n}\ ,\ h_{ii} \neq 1\ ,\ \ n > p$

وسنأخذ $\alpha = 0.05$ أو $\alpha = 0.01$ من جدول F

فإذا كانت $F_c > F_t$ فهذا سيعني إن هذه النقطة أو هذا القيد الذي قلنا انه قد يكون شاذاً هو فعلاً شاذ عدا ذلك فالنقطة أو القيد غير شاذين.

الاختبار الثالث :

يمكن تطبيق اختبار Cook's Di للتأكد من قيم h_i (الشاذة) اذ إن اختبار Cook's Di مرتبط ارتباطاً كبيراً مع قطر H . فأن اختبار Cook's Di يعتمد على عناصر قطر H في اختيار الشاذ . اذ يمثل الشاذ في هذا الاختبار اكبر عنصر في عناصر Cook's Di وقانون اختبار Cook's Di هو :

$$Di = \{\ e_i / s\,(1 - h_{ii})^{1/2}\ \}^2\ \ h_{ii} / p\,(1 - h_{ii})$$

حيث إن:

e : يمثل الخطأ العشوائي الذي ينتج من طرح المشاهدات الحقيقية من المشاهدات المقدرة .

s : يمثل الجذر التربيعي إلى s^2 .

h_{ii} : يمثل العناصر القطرية في H .

P : يمثل عدد المتغيرات في المسألة .

إن الاختبارات الثلاثة هذه قد لا تتفق جميعها في تشخيص شذوذ أو عدم شذوذ نقطة أو قيد فقد يشير احد الاختبارات إلى شذوذ في حين لا يشير إلى ذلك الاختبار الأخر لكن اتفاق الجميع على الشذوذ أو عدمه يؤكد قوة ذلك الاعتقاد في حين إذا أشارت إحدى الاختبارات إلى الشذوذ

ونفته أخرى فمعنى هذا إن الشذوذ لم يكن بالشدة او القوة بحيث تؤكده بقية الاختبارات لذلك فهو شذوذ ضعيف .

مثال(2)/ معمل نجارة عمار يستطيع صناعة ثلاثة أنواع من الأرائك ربح كل نوع منها (10 $ ، 15 $ ،$ 20) على التوالي ، أن المسلك التكنولوجي لها يمر بثلاثة ورش، الورشة الأولى هي ورشة النجارة وفيها يتم تقطيع الأجزاء الخشبية بالقياسات المطلوبة ويحتاج كل نوع منها إلى ساعتين والورشة الثانية هي ورشة التجميع وفيها يتم تثبيت الهيكل الخشبي للأريكة ويحتاج كل نوع منها إلى (1,1, 2) ساعة على التوالي ، والورشة الثالثة هي ورشة التغليف ويتم فيها تغليف الأريكة بالأسفنج والقماش ويحتاج كل نوع منها إلى (1 , 2 , 2) ساعة على التوالي ، والزمن المتاح للورش الثلاث في الأسبوع هو (48 ,48 , 50) ساعة وأن الطلب على المنتج الثاني لا يقل عن (5) قطع وعلى المنتج الثالث لا يقل عن (25) قطعة في الأسبوع ، فما هو المزيج السلعي الأمثل الذي يحقق أعلى الأرباح لصاحب المعمل .

الحل :- نفرض أن عدد وحدات النوع الأول من الأرائك $=X_1$

عدد وحدات النوع الثاني من الأرائك $=X_2$

عدد وحدات النوع الثالث من الأرائك $=X_3$

فيكون النموذج الرياضي لنموذج البرمجة الخطية للمشكلة كالأتي :-

$$Max\ Z = 10X_1 + 15X_2 + 20X_3$$
$$S.t$$
$$2X_1 + 2X_2 + 2X_3 \leq 48$$
$$X_1 + X_2 + 2X_3 \leq 48$$
$$X_1 + 2X_2 + 2X_3 \leq 50$$
$$X_2 \geq 5$$
$$X_3 \geq 25$$
$$\mathrm{X_1 ، X_2 , X_3\ unrestricted\ \ in\ signs}$$

تم حل النموذج أعلاه باستخدام برنامج (win QSB) فكانت النتائج كما يلي :

$$\mathrm{X_1} = -10 \quad , \quad X_2 = 5 \quad ,X_3 = 25$$
$$\because H = X(\grave{X}X)^{-1}\grave{X}$$

$$\grave{X}X = \begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 2 & 1 & 0 \\ 2 & 2 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 7 & 8 \\ 7 & 10 & 10 \\ 8 & 10 & 13 \end{bmatrix}$$

$$\therefore (\grave{X}X)^{-1} = \frac{1}{23} \begin{bmatrix} 30 & -11 & -10 \\ -11 & 14 & -4 \\ -10 & -4 & 11 \end{bmatrix}$$

$$\therefore H = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \frac{1}{23} \begin{bmatrix} 30 & -11 & -10 \\ -11 & 14 & -4 \\ -10 & -4 & 11 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 2 & 1 & 0 \\ 2 & 2 & 2 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.869 & 0.174 & 0.087 & -0.087 & -0.260 \\ 0.174 & 0.434 & 0.217 & -0.217 & -0.348 \\ 0.087 & 0.217 & 0.608 & 0.391 & 0.174 \\ 0.087 & 0.217 & 0.391 & 0.608 & -0.174 \\ -0.260 & -0.348 & -0.174 & -0.174 & 0.478 \end{bmatrix}$$

يمكن كتابة العناصر القطرية للمصفوفة H كما يأتي :-

ت	1	2	3	4	5
h_{ij}	0.869	0.434	0.608	0.608	0.478

ولو طبقنا الاختبار الأول (Thumb test):-

$$\frac{2p}{n} = \frac{2 \times 3}{5} = 1.2$$

وأن المقدار $\frac{2p}{n}$ لا يشير إلى وجود قيم شاذة .

ولو طبقنا الاختبار الثاني :-

$$F_{Ci} = \frac{(h_i - \frac{1}{n})/(P-1)}{(1-h_i)/(n-P)} \quad \therefore F_{C1} = 5.106\ ,\ F_{C2} = 0.413\ , F_{C3} = 1.041\ , F_{C4} = 1.041\ , F_{C5} = ($$

$$\because F_{t\ 0.05} = 19 \qquad F_{t\ 0.01} = 99$$

لا توجد قيم شاذة وفق لهذا الاختبار.وعند تطبيق الاختبار الثالث نحصل على :-

$$\hat{B} = (\grave{X}X)^{-1} \grave{X}\, Y$$

$$= \frac{1}{23}\begin{bmatrix} 30 & -11 & -10 \\ -11 & 14 & -4 \\ -10 & -4 & 11 \end{bmatrix}\begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 2 & 1 & 0 \\ 2 & 2 & 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 48 \\ 48 \\ 50 \\ 5 \\ 25 \end{bmatrix}$$

$$= \begin{bmatrix} -3.67 \\ 3.65 \\ 23.95 \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

فإذا افترضنا نموذج الانحدار لهذا المثال سيكون كالأتي :-

$$\widehat{Y}_1 = B_1X_1 + B_2X_2 + B_3X_3$$

حيث قيم (X_1, X_2, X_3) تؤخذ من المصفوفة X .

$$\therefore \widehat{Y}_1 = -3.67 \times 2 + 3.65 \times 2 + 23.95 \times 2 = 47.44$$
$$e_1 = Y_1 - \widehat{Y}_1 = 48 - 47.44 = 0.56$$
$$\widehat{Y}_2 = -3.67 \times 1 + 3.65 \times 1 + 23.95 \times 2 = 47.68$$
$$e_2 = Y_2 - \widehat{Y}_2 = 48 - 47.68 = 0.32$$
$$\widehat{Y}_3 = -3.67 \times 1 + 3.65 \times 2 + 23.95 \times 2 = 51.33$$
$$e_3 = Y_3 - \widehat{Y}_3 = 50 - 51.33 = -1.33$$
$$\widehat{Y}_4 = -3.67 \times 0 + 3.65 \times 1 + 23.95 \times 0 = 3.65$$
$$e_4 = Y_4 - \widehat{Y}_4 = 5 - 3.65 = 1.35$$
$$\widehat{Y}_5 = -3.67 \times 0 + 3.65 \times 0 + 23.95 \times 1 = 23.95$$
$$e_5 = Y_5 - \widehat{Y}_5 = 25 - 23.95 = 1.05$$

يمكن تلخيص النتائج بالجدول الأتي :-

ت	Y_i	$\widehat{Y}_i$	e_i	e_i^2
1	48	47.44	0.56	0.3136
2	48	47.68	0.32	0.1024
3	50	51.33	−1.33	1.7689
4	5	3.65	1.35	1.8225
5	25	23.95	1.05	1.1025
				$\sum e_i^2 = 5.1099$

SSE=5.1099

$$S^2 = \frac{SSE}{n-2} = 1.7033$$

$$S = \sqrt{S^2} = 1.305$$

$$D_i = \left[e_i / S(1-h_i)^{1/2}\right]^2 [h_i/P(1-h_i)]$$

$$\therefore D_1 = 3.108\,, D_2 = 0.027\,, D_3 = 1.3699\,, D_4 = 1.4114, D_5 = 0.3785$$

∴القيد الأول يعتبر شاذ حسب هذا الاختبار .

ويمكن أن نلخص نتائج الاختبارات الثلاثة للمثال كما في الجدول أدناه :-

ت	h_{ij}	الاختبار الأول	الاختبار الثاني	الاختبار الثالث
1	0.869	**1.2** غير شاذ	**5.106** غير شاذ	**3.108** شاذ
2	0.434	**1.2** غير شاذ	**0.413** غير شاذ	**0.027** غير شاذ
3	0.608	**1.2** غير شاذ	**1.041** غير شاذ	**1.3699** غير شاذ
4	0.608	**1.2** غير شاذ	**1.041** غير شاذ	**1.4114** غير شاذ
5	0.478	**1.2** غير شاذ	**0.532** غير شاذ	**0.3785** غير شاذ

مثال(3) معمل التضامن لصناعة الشبابيك الحديدية ، فكان أنتاج شباك من قياس (1.5×2) م بثلاثة أنواع فالنوع الأول من الحديد المربع والنوع الثاني من الحديد المربع وحديد Tوالنوع الثالث من حديد T فقط وأن ربح المنتجات الثلاث (25, 30, 40) دولار على التوالي وأن أنتاج كل شباك يمر بأربعة مراحل هي مرحلة تقطيع الحديد بالقياسات المطلوبة وكل منتج يحتاج إلى (3 , 2, 2.5) ساعة على التوالي ومرحلة لحم الأجزاء وكل منتج يحتاج إلى (1 , 1.5 , 3) ساعة على التوالي ومرحلة صب وعمل النقشات داخل الشبابيك وكل منتج يحتاج إلى (2 , 3 , 4) ساعة على التوالي ومرحلة التشطيب والصباغة وكل منتج يحتاج إلى (2 , 4 , 6) ساعة على التوالي والزمن الكلي المتاح للمراحل الأربعة في الأسبوع هو (42 , 32 , 40 , 36) ساعة وأن أنتاج المنتج الثاني لا يقل عن (5) قطعة و أنتاج المنتج الثالث لا يزيد عن (20)قطعة في الأسبوع ، فما هو المزيج السلعي الذي يحقق أعلى الأرباح لصاحب المعمل .

الحل : - نفرض أن عدد وحدات النوع الأول $=X_1$

عدد وحدات النوع الثاني $=X_2$

عدد وحدات النوع الثالث $=X_3$

فيصبح النموذج كالأتي :-

$$Max\ Z = 25X_1 + 30X_2 + 40X_3$$

$$S.t$$

$$2X_1 + 2.5X_2 + 3X_3 \le 42$$

$$X_1 + 1.5X_2 + 3X_3 \le 32$$

$$2X_1 + 3X_2 + 4X_3 \le 40$$

$$2X_1 + 4X_2 + 6X_3 \le 36$$

$$X_2 \ge 5$$

$$X_3 \le 20$$

X_1 ،X_2 ,X_3 unrestricted in signs

تم حل النموذج أعلاه باستخدام برنامج (win QSB) فكانت النتائج كما يلي :

$$X_1 = 21.5 \qquad ,X_2 = 5 \quad , \qquad X_3 = -4.5$$

$$H = X(\grave{X}X)^{-1}\grave{X}$$

$$\grave{X}X = \begin{bmatrix} 2 & 1 & 2 & 2 & 0 & 0 \\ 2.5 & 1.5 & 3 & 4 & 1 & 0 \\ 3 & 3 & 4 & 6 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2.5 & 3 \\ 1 & 1.5 & 3 \\ 2 & 3 & 4 \\ 2 & 4 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 13 & 20.5 & 29 \\ 20.5 & 34.5 & 48 \\ 29 & 48 & 71 \end{bmatrix}$$

$$\therefore (\grave{X}X)^{-1} = \frac{1}{111.25} \begin{bmatrix} 145.5 & -63.5 & -16.5 \\ -63.5 & 82 & -29.5 \\ -16.5 & -29.5 & 28.25 \end{bmatrix}$$

$$\therefore H$$

$$= \begin{bmatrix} 2 & 2.5 & 3 \\ 1 & 1.5 & 3 \\ 2 & 3 & 4 \\ 2 & 4 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \frac{1}{111.25} \begin{bmatrix} 145.5 & -63.5 & -16.5 \\ -63.5 & 82 & -29.5 \\ -16.5 & -29.5 & 28.25 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 & 2 & 0 & 0 \\ 2.5 & 1.5 & 3 & 4 & 1 & 0 \\ 3 & 3 & 4 & 6 & 0 & 1 \end{bmatrix}$$

يمكن كتابة العناصر القطرية للمصفوفة H كما يأتي:-

ت	1	2	3	4	5	6
h_{ij}	0.658	0.263	0.341	0.746	0.737	0.254

ويمكن أن تلخيص نتائج الاختبارات الثلاثة للمثال كما في الجدول أدناه :-

ت	h_{ij}	الاختبار الأول	الاختبار الثاني	الاختبار الثالث
1	0.658	**1**غير شاذ	2.15غير شاذ	12.06غير شاذ
2	0.263	**1**غير شاذ	0.195غير شاذ	14.905غير شاذ
3	0.341	**1**غير شاذ	0.401غير شاذ	0.0001غير شاذ
4	0.746	**1**غير شاذ	3.452غير شاذ	165.693غير شاذ
5	0.737	**1**غير شاذ	3.275غير شاذ	472.703غير شاذ
6	0.254	**1**غير شاذ	0.177غير شاذ	532.001 شاذ

مثال (4) مصنع أحمد لصناعة المضخات المائية الكبيرة ينتج نوعين من المضخات ربح كل منهما (60, 75) دولار على التوالي ، يمر كل نوع منهما بسبعة ورش هي ورشة السباكة ويتم فيها صب أجزاء المضخة الداخلية والخارجية ويحتاج كل نوع منها إلى (90 , 105) دقيقة على التوالي ورشة التثقيب ويتم فيها تثقيب أجزاء المضخة بالثقوب وبالأقطار المطلوبة ويحتاج كل نوع منها إلى (15 , 45) دقيقة على التوالي ، وورشة التفريز فتح المسننات الداخلية وبالمواصفات والقياسات المطلوبة ويحتاج كل نوع منها إلى (30 , 75) دقيقة على التوالي ، وورشة الصباغة ويتم فيها صبغ أجزاء المضخة ويحتاج كل نوع منها إلى (45 , 60) دقيقة على التوالي ، وورشة اللف ويتم فيها لف ملف المضخة الرئيسي والثانوي ويحتاج كل نوع منها إلى (105 , 30) دقيقة على التوالي ، وورشة التجميع والفحص ويتم فيها تجميع أجزاء المضخة وبعد ذلك فحص المضخة ومطابقتها للمواصفات المطلوبة ويحتاج كل نوع منها إلى (135 , 105) دقيقة على التوالي ، وورشة التغليف ويتم فيها تغليف المضخة لتصبح جاهزة للتسويق ويحتاج كل نوع منها إلى (90 60 دقيقة) على التوالي ،وأن الوقت المتاح للورش (10500,21000,12000,13500,9000,15000,18000) دقيقة على التوالي في الشهر ، فما المزيج السلعي الذي يحقق أعلى الأرباح .

الحل :-

نفرض أن عدد وحدات النوع الأول من المضخات $=X_1$

عدد وحدات النوع الثاني من المضخات $=X_2$

يصبح النموذج كالأتي :-

$$\text{Max } Z = 60X_1 + 75X_2$$
$$90X_1 + 105X_2 \leq 18000$$
$$15X_1 + 45X_2 \leq 15000$$
$$30X_1 + 75X_2 \leq 9000$$
$$45X_1 + 60X_2 \leq 13500$$
$$105X_1 + 30X_2 \leq 12000$$
$$135X_1 + 105X_2 \leq 21000$$
$$60X_1 + 90X_2 \leq 10500$$
$$X_1 ، X_2 \text{ unrestricted in signs}$$

تم حل النموذج أعلاه باستخدام برنامج (win QSB) فكانت النتائج كما يلي :

$$X_1 = 100 \qquad , X_2 = 50$$
$$H = X(\grave{X}X)^{-1}\grave{X}$$
$$\because (\grave{X}X)^{-1} = \frac{1}{436590000}\begin{bmatrix} 42300 & -37800 \\ -37800 & 44100 \end{bmatrix}$$
$$\therefore H = \begin{bmatrix} 90 & 105 \\ 15 & 45 \\ 30 & 75 \\ 45 & 60 \\ 105 & 30 \\ 135 & 105 \\ 60 & 90 \end{bmatrix} \frac{1}{436590000}\begin{bmatrix} 42300 & -37800 \\ -37800 & 44100 \end{bmatrix}\begin{bmatrix} 90 & 15 & 30 & 45 & 105 & 135 & 60 \\ 105 & 45 & 75 & 60 & 30 & 105 & 90 \end{bmatrix}$$

يمكن كتابة العناصر القطرية للمصفوفة H كما يأتي:-

ت	1	2	3	4	5	6	7
h_{ij}	0.262	0.109	0.265	0.092	0.613	0.424	0.231

ويمكن أن تلخيص نتائج الاختبارات الثلاثة للمثال كما في الجدول أدناه :-

ت	h_{ij}	الاختبار الأول	الاختبار الثاني	الاختبار الثالث
1	0.262	**0.57** غير شاذ	0.807 غير شاذ	0.002 غير شاذ
2	0.109	**0.57** غير شاذ	−0.189 غير شاذ	0.233 شاذ
3	0.265	**0.57** غير شاذ	0.830 غير شاذ	0.041 غير شاذ

4	0.092	**0.57** غير شاذ	−0.290 غير شاذ	0.027 غير شاذ
5	0.613	**0.57** شاذ	6.074 غير شاذ	0.181 غير شاذ
6	0.424	**0.57** غير شاذ	2.440 غير شاذ	0.008 غيرشاذ
7	0.231	**0.57** غير شاذ	0.573 غير شاذ	0.571 غيرشاذ

6-معالجة القيود الشاذة :-

بعد كشف القيد أو القيود الشاذة في نموذج البرمجة الخطية سيتخذ متخذ القرار قرارا يخص هذا القيد أو القيود وهنالك عدة بدائل يمكن أتخاذها من قبل متخذ القرار وهذه البدائل هي :-

1-قرار بإبقاء النموذج كما هو.

2- قرار رفع القيد الشاذ من النموذج.

3- قرار معالجة القيد الشاذ.

ففي المثال الرابع فأن القيد الخامس (ورشة اللف) هو قيد شاذ وعند ملاحظة زمن لف ملف المضخة الأولى (105) دقيقة وهو زمن مبالغ فيه أو يمكن تقليله إلى (80) دقيقة مثلا فنحصل على نموذج تكون نتائجه هي

$$X_1 = 134.6154, X_2 = 26.9231 , Z = 10096.15$$

وهذا يعني زيادة في ربح صاحب المصنع ،و أما عند زيادة الزمن المتاح للورشة من (12000) دقيقة إلى (15000) دقيقة مع بقاء معاملات المتغيرات كما هي فنحصل على نفس النتائج التي حصلنا عليها سابقا ، أما عندما نقوم بحذف هذا القيد باعتبار أنه قيد غير مهم فنحصل أيضا على نفس النتائج التي حصلنا عليها سابقا .

7-الاستنتاجات

من خلال استخدام طريقة المصفوفة المقدرة لاستكشاف وتحديد القيم الشاذة تم التوصل الى :

1- إن طريقة المصفوفة المقدرة التي كانت تستخدم لإيجاد القيم الشاذة في نماذج البرمجة الخطية استخدمت هنا في تحديد القيم الشاذة في نماذج البرمجة الخطية غير المقيدة .

2- تعتبر الطريقة من أسهل الطرق في تحديد القيم الشاذة .

3- تعتمد الطريقة على اسلوب المصفوفات الذي ساعد كثيرا في اختصار العمليات الحسابية .

4- ساعدت الطريقة متخذ القرار في تحديد القيم الشاذة ومعالجتها بما يخدم تحقيق الأهداف المطلوبة .

5- ان تحديد القيد الشاذ جعلنا نعدل بالنموذج أكثر من مرة إلى ان أصبح لدينا أفضل نموذج يؤدي الى أعلى الأرباح أو اقل التكاليف.

6 -من الضروري الكشف عن وجود النقاط او القيود الشاذة للبيانات ومعالجتها وذلك لأن بقاءها يزيد من قيمة مربعات الخطأ ويؤدي الى عدم جودة النموذج والى عدم دقة وكفاءة الحل

7- إن نماذج البرمجة الخطية غير المقيدة بالإشارة تعتبر أكثر مرونة من نماذج البرمجة الخطية المقيدة بالإشارة لأنها تتعامل مع متغيرات قرار حرة إي يسمح لها ان تكون سالبة او موجبة او صفر وبالتالي يمكن خفض متغير وزيادة متغير أخر والحصول على أعلى الأرباح او اقل التكاليف .

8- ان نموذج البرمجة الخطية غير المقيدة يصبح مساويا لنموذج الانحدار عندما تكون قيم المتغيرات المهملة في جدول الحل الأمثل تساوي صفر .

المصادر

1- ابو صالح ، محمد صبحي " الطرق الإحصائية " دار اليازوري للنشر ،عمان الأردن ،2009 .

2- النعيمي دار وائل للنشر مقدمة في الإحصاء 2010

3- الشمرتي ، حامد سعد نور ،" بحوث العمليات مفهوما وتطبيقا " دار وائل للنشر عمان الأردن ،2009 .

4- حسن ،هاني عبد الله " تأثيرات الإغراق و الإخفاء على مؤشر الاختبارات الناجمة عن تلوث عينة تخضع لتوزيع طبيعي" أطروحة دكتوراه في الإحصاء ، كلية الإدارة والاقتصاد ، الجامعة المستنصرية. 2007

5-ياغوبيان ،انكين انترانيك هايك " استكشاف وتقدير القيم الشاذة في بعض النماذج اللاخطية" ، أطروحة دكتوراه في الإحصاء ، كلية الإدارة والاقتصاد ، الجامعة المستنصرية. 2005 .

6- متلف ، رشا جلال " البحث عن القيود الشاذة في نماذج الانحدار الخطي و نماذج البرمجة الخطية "، رسالة ماجستير رياضيات تطبيقية ، قسم العلوم التطبيقية الجامعة التكنولوجية 2007

7 - ناسي ، نبيل " تقييم كفاءة طرق تقدير القيم الشاذة لنماذج الانحدار " اطروحة دكتوراه- كلية الإدارة والاقتصاد- جامعة بغداد. (2001)

8- ظاهر ،رحيم جبار " دراسة تأثير القيم الشاذة في بيانات تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة"، مجلة القادسية للعلوم الإدارية والاقتصادية المجلد 8 العدد 4 لسنة 2006 .

9-Arezoo B., Habshah M., Mojtaba G.and Samaneh E.," A Comparison of Various Influential Points Diagnostic Methods and Robust Regression Approaches: Reanalysis of Interstitial Lung Disease Data", Applied Mathematical Sciences, Vol. 4, no. 28,pp. 1367 – 1386,(2010)

10- Andersen R., Modern Methods for Robust Regression, Sara Miller McCune, SAGE publications, The United States of America , (2008.)

11- Baddeley, A.,Turner,R.,Møller, J.& Hazelton, M. "Residual Analysis for Special Point Processes" J.R.Statist.Soc.B,67,pp.1-35. http://www,rss.org.uk/pdf. (2005)

12- Christmann A. and Steinwart I.," Consistency and robustness of kernel based regression", Bernoulli, vol 13:pp. 799–819, (2007).

13- Debruyne, M., Hubert, M. and Van Horebeek, J.," Detecting Influential Observations in Kernel PCA" Dept. of mathematics and computer science, University of Antwerpen,2008

14- Hamdy A.T.," operations research an introduction " seventh ed. Canada ,by Maxwell publishing company ,(2004) .

15- Maronna R.A.," Principal components and orthogonal regression based on robust scales", Tech- nometrics,vol 47: pp.264–273,(2005).

16- Meloun M, Militky´ J, Hill M, Brereton RG. Crucial problems in regression modeling and their solutions. The Analyst vol .127,pp.433–450, 2002.

17- Milan M. , Martin H., ´ Militky J., Jana , V., and Jan S.," New methodology of influential point detection in regression model building for the prediction of metabolic clearance rate of glucose", Clin Chem Lab Med; Vol . 42 no.(3),pp.311–322 ,(2004).

18- Milan M., Militký J.," Detection of single influential points in OLS regression model building" , Analytica Chimica Acta vol. 439 ,pp.(169–191), (2001).

19- Pena D., A new statistics for influence in linear regression. Technometrics, vol. 47 no.(1), pp.1-12. (2005)

20- Pison G. and Van S.," Diagnostic plots for robust multivariate methods. Journal of Computational and Graphical Statistics,vol 13:pp.310–329,(2004).

21- Rousseuw P. J., and Leroy A. M., Robust Regression and Outlier Detection, New York, John Willy,(2003).
22- Wilcox R. R.," Introduction to Robust Estimation and Hypothesis Testing", 2nd edition, Elsevier academic press, USA, (2005).
23- Wei , W.H and Kosorok , M.R " Interaction influence function : masking unmasked " Ann . of Stat. , vol . 135 ,no.4 pp. 1215-1223, (2005)

Detecting and treatment the Effect of Outlier Values in Linear Decision Making Models with Unrestricted Variables and Linear Regression Models (Case study)

Rasheed Basheer Reheima / University of Thi Qar /College of Administration and Economic /statistical Department
Watheq Hayawi Laith / University of Thi Qar /College of Administration and Economic /statistical Department

abstrac

The appearance of outlier values in the set of data effect the result of the statistical analysis of data .Then the correct of decision making there for study and estimate the outlier values the detection methods. This problem was studied in some of linear programming and linear regression but did not studied in linear programming with unrestricted variables which is important subject in many fields of operations research and its ability in treatment with unrestricted variables decision. In this paper we detect and estimate the outlier values in those models and how these values effect the solution of the outlier values .The hat matrix was used to determine and process the outlier constraint so that the new model will be more suitable to match the requirements .The result show that the simplicity and accurate calculations also the clearly in determine the outlier values .