المجلة العراقية للعلوم الاحصائية (20) 2011
عدد خاص بوقائع المؤتمر العلمي الرابع كلية علوم الحاسوب والرياضيات
ص ص [360– 370]

طريقة مقترحة لاختيار أفضل معادلة انحدار بالاعتماد على معامل الارتباط الجزئي

غزوان هاني محمود*

الملخص

يتناول هذا البحث طرح طريقة مقترحة لاختيار أفضل معادلة انحدار وذلك بافتراض وجود مجموعة من المتغيرات التوضيحية والمطلوب اختيار بعض تلك المتغيرات للحصول على المعلومات التي نحصل عليها لو استخدمنا جميع هذه المتغيرات. تم في هذا البحث عرض أهم الطرائق المستخدمة لاختيار أفضل معادلة انحدار، وكذلك عرض الطريقة المقترحة التي تعتمد على قيمة معامل الارتباط الجزئي. وتم تطبيق الطريقة المقترحة على بيانات حقيقية ومقارنة نتائجها مع أفضل طريقة لاختيار أحسن معادلة انحدار والمتمثلة بطريقة الاختيار التدريجي. تم الاعتماد على معيار متوسط مربعات الخطأ (mse)، وقد تميزت الطريقة المقترحة بسهولة التطبيق ودقة النتائج.

A New Suggested Method for the Selection of Best Regression Equation Depending on the Partial Correlation coefficient

Abstract

This research suggested a new heuristic method to select the best regression equation. Assuming that we have a set of predictor variables. The choice of some of those variables is required to obtain the information that we get it if we used all those variables. The research presents the used methods to select the best regression equation also presents a new method that depends on the partial correlation coefficient. This paper compares the new method with stepwise regression procedure by using real data depending on criteria mean square error (mse).The new method showed ease of use and high efficiency.

* مدرس/كلية علوم الحاسوب والرياضيات/جامعة الموصل.
تاريخ الاستلام:1 /2011/12   تاريخ القبول:2011/12/25   تاريخ النشر:2011/12/31

1- المقدمة:

تعد عملية تحديد معادلة الانحدار في حال وجود عدد كبير من المتغيرات التوضيحية وحجم عينة كبير عملية معقدة، لذا فقد قدم الباحثون بعض الطرائق التي تمكننا من إيجاد معادلة انحدار وبأقل عدد من المتغيرات التوضيحية، ونظرا إلى أن إدخال أعدادا كبيرة من المتغيرات التوضيحية في المعادلة يكلف جهدا ووقتا وكلفة كبيرة، فانه من الأفضل إن نختار معادلة تحتوي على أقل عدد ممكن من المتغيرات التوضيحية وليكن(k) للحصول على المعلومات كما لو أننا استخدمنا جميع المتغيرات قيد الدراسة (m) حيث أن k≤m.

إن اعتماد جميع المتغيرات في أية دراسة لتحليل الانحدار يقود إلى نتائج أدق واشمل وقريبة من الواقع، من المعلوم أيضا إن إدخال عدد كبير من المتغيرات في معادلة الانحدار لأية ظاهرة أو دراسة يكلف وقتا وجهدا وثمنا. كما أن بعض المتغيرات غير أساسية في تأثيرها في المتغير المعتمد أو يصاحبها أخطاء كبيرة في القياس أو يكون تأثيرها مماثل لتأثير متغيرات أخرى وابعد من ذلك فان العديد من هذه المتغيرات يكون لها ارتباط داخلي عال مع بعضها البعض مما يؤدي إلى جعل تأثيرها غير معنوي مما يدعو إلى استبعاد مثل هذه المتغيرات. وهنا تكمن صعوبة مسالة تحليل الانحدار في اختيار مجموعة المتغيرات التوضيحية المتضمنة في معادلة الانحدار، فالباحث يحتاج إلى توفيق بين تقليل المتغيرات التوضيحية تجنبا لزيادة تكاليف الحصول على المعلومات وبين زيادة المتغيرات التوضيحية للحصول على نتائج تنبؤية أفضل.

ويكون اختيار أفضل معادلة انحدار بطرائق عدة تتفاوت في أهميتها وفي دقتها وهي خاضعة إلى حد كبير للقرار الشخصي للباحث. وهناك مقاييس للمفاضلة بين المعادلات المنتخبة ومنها:

أ- قيمة معامل التحديد المتعدد (R_p^2)

The Coefficient of Multiple Determination

وتحسب كالآتي:

$$R_p^2 = \frac{SSR(X_1X_2 \ldots X_p)}{SST} = 1 - \frac{SSe(X_1X_2 \ldots X_p)}{SST} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

ب- قيمة متوسط مربعات الخطأ (mse_p)

Residual Mean Square

وتحسب كالآتي :

$$mse_p = \frac{SSe(X_1X_2 \ldots X_p)}{n-p} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

ج- قيمة إحصائية مالو (C_p)

وتحسب كالآتي :

$$C_p = \frac{SSe(X_1X_2 \ldots X_p)}{msR(X_1X_2 \ldots X_m)} - (n-2p) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

2- طرائق اختيار أحسن معادلة انحدار :[9, 8, 6,5]

هناك عدة طرائق لاختيار أحسن معادلة وهي:

1- طريقة كل الانحدارات الممكنة: (All Possible Regression Procedure)

تعتبر طريقة كل الانحدارات الممكنة من الطرائق المطولة لاختيار أفضل معادلة انحدار لأنها تعتمد على إيجاد كل المعادلات الانحدارية الممكنة باستخدام كل المجاميع الممكنة من المتغيرات التوضيحية. فإذا كان لدينا m من المتغيرات التوضيحية فان مجموع عدد المعادلات الكلية سيكون 2^m . وبعد الحصول على جميع المعادلات يتم اختيار أفضل معادلة باستخدام احد معايير المفاضلة بين المعادلات.

طريقة الاختيار العكسي(الحذف العكسي):(The Backward Elimination Procedure)

يمكن تلخيص عمل هذه الطريقة بصياغة معادلة انحدار تحتوي على جميع المتغيرات التوضيحية، ثم نحذف المتغيرات التوضيحية غير المعنوية من المعادلة واحدا تلو الأخر بالاعتماد على قيمة F المحسوبة. فإذا كانت قيمة اقل F جزئية محسوبة أقل من قيمةF الجدولية فهذا دليل على عدم معنوية ذلك المتغير، وعليه يمكن حذفه من معادلة الانحدار. وبعد حذف المتغير الأول تصاغ معادلة الانحدار للمتغيرات المتبقية ثم نحسب قيم F الجزئية لبقية المتغيرات ونجري المقارنة وهكذا. أما إذا كانت قيمة اقل F جزئية أكبر من قيمة F الجدولية فهذا دليل على معنوية جميع المتغيرات التوضيحية.

3- طريقة الاختيار المباشر (الأمامي): (The Forward Elimination Procedure)

في هذه الطريقة تبدأ معادلة الانحدار بدون وجود أي متغير توضيحي، ثم نضيف المتغيرات الواحد تلو الآخر. وأول المتغيرات التي تضاف إلى المعادلة هو المتغير الذي يمتلك أعلى قيمة F محسوبة والتي تكون أكبر من قيمة F الجدولية، والمتغير الثاني الذي يدخل على معادلة الانحدار هو المتغير الذي يمتلك F جزئية بوجود المتغير الأول أكبر من بقية المتغيرات وأكبر من قيمة F الجدولية. وهكذا نستمر بإضافة المتغير الذي له أعلى F جزئية وأكبر من F الجدولية. ونتوقف عن الإضافة عندما تكون F الجزئية أقل من قيمة F الجدولية.

طريقة الاختيار التدريجي: (The Stepwise Selection Procedure)

إن اختيار المتغيرات لصياغة معادلة انحدار باستخدام طريقة الاختيار المتدرج هو عبارة عن توفيق أو دمج بين طريقة الاختيار المباشر وطريقة الاختيار العكسي، ففي الخطوة الأولى يتم اختيار المتغير التوضيحي الذي يمتلك أعلى قيمة F محسوبة من بين جميع المتغيرات التوضيحية وإضافته إلى معادلة الانحدار، فإذا كانت F المحسوبة أكبر من F الجدولية نثبت

المتغير في معادلة الانحدار أما إذا كانت F المحسوبة أقل من F الجدولية لا ندخل المتغير التوضيحي ونتوقف عن الاختيار. والخطوة الثانية يتم حساب قيمة F الجزئية لبقية المتغيرات التوضيحية بوجود المتغير التوضيحي الذي تم اختياره في الخطوة الأولى واختيار المتغير الذي يمتلك أعلى قيمة F جزئية ليدخل معادلة الانحدار مع المتغير الأول، أما الخطوة الثالثة فتجري عملية الاختيار العكسي لاختبار أهمية وجود المتغيرات المختارة في الخطوتين السابقتين بالاعتماد على قيمة F الجزئية.

3-الطريقة المقترحة:[1]

من اجل التعرف على هذه الطريقة لابد لنا من توضيح الأساس الذي تم الاعتماد عليه لاقتراح هذه الطريقة، تعتمد هذه الطريقة على استخدام معامل الارتباط الجزئي بين المتغير المعتمد والمتغيرات التوضيحية، ويمكن توضيح هذه الطريقة باتجاهين:

الأول: الأسلوب العكسي ويتم فيه:

1- إيجاد مصفوفة الارتباط R بين المتغير المعتمد والمتغيرات التوضيحية.

$$R=\begin{bmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_m} \\ . & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_m} \\ . & . & r_{x_2x_2} & \cdots & r_{x_2x_m} \\ . & . & . & . & . \\ . & . & . & . & r_{x_mx_m} \end{bmatrix}$$

2- إيجاد معكوس مصفوفة الارتباط R^{-1}.

$$R^{-1}=\begin{bmatrix} c_{yy} & c_{yx_1} & c_{yx_2} & \cdots & c_{yx_m} \\ . & c_{x_1x_1} & c_{x_1x_2} & \cdots & c_{x_1x_m} \\ . & . & c_{x_2x_2} & \cdots & c_{x_2x_m} \\ . & . & . & . & . \\ . & . & . & . & c_{x_mx_m} \end{bmatrix}$$

3- نوجد معامل الارتباط الجزئي للمتغير المعتمد مع كل متغير من المتغيرات التوضيحية بثبوت المتغيرات الأخرى. وبصورة عامة فان معامل الارتباط الجزئي بين المتغيرين i وj بعد جعل تأثير جميع المتغيرات الأخرى ثابتاً هو:

$$r_{ij}.(all\,over\,\mathrm{var}\,iable)=\frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}}$$

حيث إن c_{ij} و c_{ii}و c_{jj} هي عناصر معكوس مصفوفة الارتباط.

4- ترتيب القيم المطلقة لمعامل الارتباط الجزئي المحسوبة في الخطوة (3) ترتيباً تنازلياً.

5- إيجاد معادلة الانحدار باستخدام جميع المتغيرات وحساب قيمة متوسط مربعات الخطأ (mse).

6- حذف المتغير الذي يمتلك أقل قيمة معامل ارتباط جزئي مطلقة ومن ثم نحسب (mse) للمعادلة التي تضم المتغيرات التوضيحية المتبقية.

7- نقارن قيمة (mse) للمعادلة في الخطوة (5) مع قيمة (mse) للمعادلة في الخطوة (6) . فإذا كانت قيمة متوسط مربعات الخطأ لجميع المتغيرات أقل من قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الذي يمتلك أقل معامل ارتباط جزئي، فهذا دليل على معنوية جميع المتغيرات ونتوقف عن الحذف. أما إذا كانت أكبر فهذا دليل على عدم معنوية (أهمية) المتغير المحذوف. ثم نحذف المتغير التوضيحي الذي يمتلك ثاني اقل معامل ارتباط جزئي ونقارن قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الثاني مع قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الذي يمتلك أقل معامل ارتباط جزئي وهكذا.......

ثانيا: الأسلوب الأمامي وفيه يتم:

إتباع نفس خطوات الخوارزمية السابقة عدا الخطوات الآتية:

5- يتم إيجاد معادلة الانحدار لكل من المتغير المعتمد والمتغير التوضيحي الذي يمتلك اكبر معامل ارتباط جزئي مع حساب قيمة (mse).

6- نضيف المتغير التوضيحي والذي يمتلك ثاني اكبر معامل ارتباط جزئي مع حساب قيمة (mse).

7- نقارن قيمة (mse) للمعادلة في الخطوة (5) مع قيمة (mse) للمعادلة في الخطوة (6) . فإذا كانت قيمة متوسط مربعات الخطأ في الخطوة (5) أقل من قيمة متوسط مربعات الخطأ في الخطوة (6) ، فهذا دليل على معنوية المتغير الأول وعدم معنوية المتغير المضاف ومن ثم يمكننا الجزم بعدم معنوية باقي المتغيرات ونتوقف عن إضافة المتغيرات. أما إذا كانت أكبر فهذا دليل على معنوية المتغير المضاف. ثم نضيف المتغير التوضيحي الذي يمتلك ثالث اكبر معامل ارتباط جزئي ونقارن قيمة متوسط مربعات الخطأ بعد إضافة المتغير التوضيحي الذي يمتلك ثاني اكبر معامل ارتباط جزئي مع قيمة متوسط مربعات الخطأ بعد إضافة المتغير التوضيحي الذي يمتلك ثالث اكبر معامل ارتباط جزئي وهكذا.......

4- الجانب التطبيقي:[7, 4, 3, 2]

تناول الجانب التطبيقي من البحث المقارنة بين الطريقة الاعتيادية (طريقة الانحدار المتدرج) باعتبارها أفضل طريقة لاختيار أحسن معادلة انحدار. وتم الاعتماد على أربعة أمثلة حقيقية يتضمن كل منها أعداداً مختلفة من المتغيرات التوضيحية فضلا عن اختلاف أحجام العينات وكما يأتي:

المثال (1):

تناول الباحث [النعيمي، 2005] مجموعة البيانات المتعلقة بمرض الثلاسيميا كدالة لعشرة متغيرات توضيحية وهي:

Y : عمر العظم (شهر).

X1 : العمر الحقيقي (شهر). X2 : عمر المريض عند المرض (شهر).

X3 : تضخم الكبد (سنتمتر). X4 : هيموكلوبين الدم.

X5 : مكداس الدم (خلايا الدم المضغوط). X6 : الخلايا الشبكية.

X7 : أرومة حمراء. X8 : الهيموكلوبين الجيني.

X9 : عدد وحدات الدم. X10 : بداية نقل الدم حسب العمر (شهر).

ولاختيار أفضل معادلة نتبع الخطوات الآتية:

1- إيجاد مصفوفة الارتباط:

$$R = \begin{bmatrix}
1.00000 & 0.85386 & 0.03502 & 0.31885 & 0.11462 & 0.08069 & -0.03668 & 0.44844 & -0.18403 & 0.79195 & 0.05971 \\
 & 1.00000 & 0.00603 & 0.26222 & 0.27219 & 0.22059 & -0.07598 & 0.38606 & -0.26084 & 0.96404 & -0.00700 \\
 & & 1.00000 & 0.06287 & 0.09279 & 0.07443 & 0.18025 & 0.17829 & 0.16153 & -0.08654 & 0.78830 \\
 & & & 1.00000 & -0.17493 & -0.19371 & -0.11521 & 0.18532 & -0.10822 & 0.27757 & 0.13688 \\
 & & & & 1.00000 & 0.94988 & 0.05738 & 0.19721 & 0.00613 & 0.08991 & 0.06566 \\
 & & & & & 1.00000 & 0.11324 & 0.17153 & 0.04427 & 0.06164 & 0.05173 \\
 & & & & & & 1.00000 & 0.07934 & 0.27254 & -0.10244 & 0.18436 \\
 & & & & & & & 1.00000 & 0.04555 & 0.30869 & 0.12319 \\
 & & & & & & & & 1.00000 & -0.30601 & 0.20045 \\
 & & & & & & & & & 1.00000 & -0.10857 \\
 & & & & & & & & & & 1.00000
\end{bmatrix}$$

2- إيجاد معكوس مصفوفة الارتباط

$$R^{-1}=\begin{bmatrix}
6.1776 & -15.1621 & 0.46982 & -0.25477 & 4.2667 & -1.8063 & -0.10019 & -0.51406 & 0.14987 & 9.7705 & 0.11534 \\
 & 78.5668 & -1.27600 & 0.16037 & -23.4929 & 10.5005 & 0.30333 & -0.69363 & -1.04511 & -62.8402 & -3.14502 \\
 & & 2.76322 & 0.12620 & 0.1384 & 0.0399 & -0.08670 & -0.28756 & 0.05722 & 0.9162 & -2.10425 \\
 & & & 1.24399 & 0.1264 & 0.1367 & 0.12742 & -0.14012 & 0.05517 & -0.2680 & -0.31517 \\
 & & & & 17.9156 & -13.4316 & 0.47385 & -0.09449 & 0.60184 & 18.8140 & 0.81915 \\
 & & & & & 12.2354 & -0.56587 & 0.00550 & -0.42066 & -8.5003 & -0.35470 \\
 & & & & & & 1.15959 & -0.07876 & -0.24104 & -0.2075 & -0.12111 \\
 & & & & & & & 1.37840 & -0.13077 & 0.6482 & 0.21892 \\
 & & & & & & & & 1.24342 & 1.2311 & -0.14172 \\
 & & & & & & & & & 53.3095 & 2.99508 \\
 & & & & & & & & & & 2.98652
\end{bmatrix}$$

3− إيجاد معامل الارتباط الجزئي للمتغير المعتمد مع كل متغير من المتغيرات التوضيحية.

$$r_{ij}.(\text{all other varible})=\frac{-c_{ij}}{\sqrt{c_{ii}\,c_{jj}}}$$

$$r_{yx_1.(\text{all other varible})}=\frac{15.1621}{\sqrt{(6.1776)\ (78.5668)}}=0.688 \quad , \quad r_{yx_2.(\text{all other varible})}=\frac{-0.46982}{\sqrt{(6.1776)\ (2.76322)}}=-0.114$$

$$r_{yx_3.(\text{all other varible})}=\frac{0.25477}{\sqrt{(6.1776)\ (1.24399)}}=0.092 \qquad r_{yx_4.(\text{all other varible})}=\frac{4.2667}{\sqrt{(6.1776)\ (17.9156)}}=-0.406$$

$$r_{yx_5.(\text{all other varible})}=\frac{1.8063}{\sqrt{(6.1776)\ (12.2354)}}=0.208 \qquad r_{yx_6.(\text{all other varible})}=\frac{0.10019}{\sqrt{(6.1776)\ (1.15959)}}=0.037$$

$$r_{yx_7.(\text{all other varible})}=\frac{0.51406}{\sqrt{(6.1776)\ (1.37840)}}=0.176 \qquad r_{yx_8.(\text{all other varible})}=\frac{-0.14987}{\sqrt{(6.1776)\ (1.24342)}}=-0.054$$

$$r_{yx_9.(\text{all other varible})}=\frac{-9.7705}{\sqrt{(6.1776)\ (53.3095)}}=-0.538 \qquad r_{yx_{10}.(\text{all other varible})}=\frac{-0.11534}{\sqrt{(6.1776)\ (2.98652)}}=-0.027$$

4− نرتب قيم معامل الارتباط الجزئي ترتيبا تنازليا بعد اخذ القيمة المطلقة فيكون لدينا التسلسل الآتي:

$$X_1=0.688, X_9=0.538, X_4=0.406, X_5=0.208, X_7=0.176, X_2=0.114, X_3=0.092, X_8=0.054, X_6=0.037, X_{10}=0.027$$

5− نوجد معادلة الانحدار للمتغير المعتمد وجميع المتغيرات التوضيحية ثم نحسب قيمة mse (mse= 257)

6− نحذف المتغير الذي يمتلك أقل معامل ارتباط جزئي وهو المتغير العاشر ونوجد معادلة الانحدار لبقية المتغيرات ونحسب قيمة mse (mse= 255)

7- بما إن قيمة متوسط مربعات الخطأ بعد حذف المتغير العاشر أقل من متوسط مربعات الخطأ لجميع المتغيرات ، إذن المتغير العاشر غير معنوي ويمكن حذفه.
نحذف المتغير الذي يمتلك ثاني أقل معامل ارتباط جزئي وهو المتغير السادس ونوجد معادلة الانحدار لبقية المتغيرات ونحسب قيمة mse (mse= 254)
بما إن قيمة متوسط مربعات الخطأ بعد حذف المتغير السادس أقل من متوسط مربعات الخطأ للمتغيرات المتبقية ، إذن المتغير السادس غير معنوي ويمكن حذفه.
نحذف المتغير الذي يمتلك أقل معامل ارتباط جزئي ثالث وهو المتغير الثامن ونوجد معادلة الانحدار ونحسب قيمة mse (mse= 252)
بما إن قيمة متوسط مربعات الخطأ بعد حذف المتغير الثامن أقل من متوسط مربعات الخطأ للمتغيرات المتبقية ، إذن المتغير الثامن غير معنوي ويمكن حذفه.
نحذف المتغير الذي يمتلك أقل معامل ارتباط جزئي رابع وهو المتغير الثالث ونوجد معادلة الانحدار ونحسب قيمة mse (mse= 253)

بما إن قيمة متوسط مربعات الخطأ بعد حذف المتغير الثالث أكبر من متوسط مربعات الخطأ للمتغيرات المتبقية ، إذن المتغير الثالث معنوي ولا يمكن حذفه. ونتوقف عن الحذف أي أن بقية المتغيرات وهي ($X_1, X_9, X_4, X_5, X_7, X_2, X_3$) هي متغيرات معنوية ومهمة في النموذج. والجدول (1) يوضح نتائج الحل باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج.

الطريقة المقترحة		طريقة الانحدار المتدرج	
Mse	المتغيرات	mse	المتغيرات
257	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$	404	X_1
255	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$	381	X_1, X_7
254	$X_1, X_2, X_3, X_4, X_5, X_7, X_8, X_9$	356	X_1, X_7, X_4
252	$X_1, X_2, X_3, X_4, X_5, X_7, X_9$	269	X_1, X_7, X_4, X_9
		261	X_1, X_7, X_4, X_9, X_5
		253	$X_1, X_7, X_4, X_9, X_5, X_2$

الجدول (1) نتائج اختيار أفضل معادلة انحدار باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج

المثال (2) :

تناول الباحث [شاكر، 1998] مجموعة البيانات المتعلقة بإنتاج السكر الأبيض كدالة لتسعة متغيرات توضيحية وهي:

Y = كمية السكر الأبيض المنتج (طن).

X1 = عدد أيام العطل خلال الموسم (يوم).   X2 = مجموع زمن التوقف في الإنتاج (يوم).

X3 =عدد أيام العمل الفعلية في الموسم (يوم). X4 = إجمالي كمية البنجر المورد إلى المعمل (طن).

X5 = إجمالي كمية البنجر المصنع (طن).   X6 = إجمالي كمية الأتربة والأحراش (طن).

X7 = معدل إنتاجية الدونم (طن/دونم).   X8 = توقفات بسبب عدم وجود بنجر (ساعة).

X9 = عطلات ميكانيكية (ساعة).

تم في هذا المثال الحل باستخدام الأسلوب الأمامي والجدول (2) يوضح نتائج الحل باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج:

الطريقة المقترحة		طريقة الانحدار المتدرج	
Mse	المتغيرات	Mse	المتغيرات
282153	X_3	282153	X_3

الجدول (2) نتائج اختيار أفضل معادلة انحدار باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج

المثال (3):

تناول الباحث [سعيد، 1996] مجموعة البيانات المتعلقة باستهلاك الوقود في توليد الطاقة الكهربائية في المحطات الحرارية كدالة لكميات ستة عناصر وهي:

Y : الطاقة الكهربائية المنتجة في المحطات الحرارية (1000 مك واط/ ساعة).

X1 : النفط الخام (100 ألف غالون).   X2 : زيت البنكر–س (100 ألف غالون).

X3 : زيت الوقود (100 ألف غالون) .   X4 : زيت الديزل (100 ألف غالون).

X5 : زيت الغاز (100 ألف غالون).   X6 : الغاز الطبيعي (100 إلف متر مكعب).

والجدول (3) يوضح نتائج الحل باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج:

الطريقة المقترحة		طريقة الانحدار المتدرج	
Mse	المتغيرات	Mse	المتغيرات
79347	$X_1, X_2, X_3, X_4, X_5, X_6$	473194	X_6
76087	X_6, X_1, X_3, X_5, X_4	135986	X_6, X_1
		108109	X_6, X_1, X_3
		90168	X_6, X_1, X_3, X_4
		76087	X_6, X_1, X_3, X_4, X_5

الجدول (3) نتائج اختيار أفضل معادلة انحدار باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج

المثال (4):

تناول الباحث[Hald ,1952] مجموعة البيانات المتعلقة بالحرارة المنبعثة كدالة لكميات أربعة عناصر في مزيج مادة الأسمنت وهي:

Y : الحرارة المنبعثة،

X1 : كمية الومينات الكالسيوم الثلاثي. X2 : كمية سيليكات الكالسيوم الثلاثي.

X3 : كمية الومينات الكالسيوم الرباعي الحديديكي. X4 : كمية سيليكات الكالسيوم الثنائي.

والجدول(4) يوضح نتائج الحل باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج:

الطريقة المقترحة		طريقة الانحدار المتدرج	
Mse	المتغيرات	mse	المتغيرات
5.98	X_1, X_2, X_3, X_4	80.4	X_4
5.33	X_1, X_2, X_4	7.5	X_4, X_1
		5.33	X_4, X_1, X_2

الجدول (4) نتائج اختيار أفضل معادلة انحدار باستخدام الطريقة المقترحة وطريقة الانحدار المتدرج

5-الاستنتاجات:

1- تم اقتراح طريقة جديدة لاختيار أفضل معادلة انحدار ذات كفاءة عالية وسهولة بالاستخدام.

2- إمكانية استخدام الطريقة المقترحة لاختيار أفضل معادلة انحدار بغض النظر عن طبيعة البيانات.

3- نوصي بإجراء دراسات مماثلة لاختيار أفضل معادلة انحدار.

4- نوصي باستخدام الطرائق الجديدة لاختيار أفضل معادلة انحدار، ومن ثم مقارنتها مع الطرائق السابقة لمعرفة الطريقة الأكفأ لاختيار أفضل معادلة انحدار.

6-المصادر:

1- الراوي، خاشع محمود، (1987)، "المدخل إلى تحليل الانحدار"، مديرية دار الكتب للطباعة والنشر، جامعة الموصل، العراق.

2- النعيمي، أسوان محمد طيب رشيد، (2005)، "اختيار المتغيرات في انحدار الحرف"، رسالة ماجستير(غير منشورة)، كلية علوم الحاسوب والرياضيات، جامعة الموصل، العراق.

3- سعيد، هيفاء عبد الجواد، (1996)، "طرق التعرف على تعدد العلاقة الخطية وكيفية معالجتها بطرائق التقدير المتحيزة"، رسالة ماجستير(غير منشورة)، كلية الإدارة والاقتصاد، جامعة الموصل، العراق.

4- شاكر، صالح مؤيد، (1998)، "العلاقة بين انحدار الحرف والتحليل الذاتي لمصفوفة الارتباط المتضخمة"، رسالة ماجستير(غير منشورة)، كلية الإدارة الاقتصاد، جامعة الموصل، العراق.

5- كاظم، أموري هادي والدليمي، محمد مناجد عيفان، (1988)، "مقدمة في تحليل الانحدار المتعدد"، مطبعة جامعة بغداد، العراق.

6- Derksen, S. & Keselman, H. J., (1992), "Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables", British Journal of Mathematical & Statistical Psychology, Vol. 45, 265-282.

7- Hald, A., (1952), "Statistical Theory with Engineering Applications", Wiley.

8- Hurvich, C.M. & Tsai, C.L., (1990), "The impact of model selection on inference in linear regression", American Statistician, Vol.44, 214-217.

9- Whittingham, M. J., Philip A. S., Richard B. B., and Robert P. F., (2006), “Why do we still use stepwise modelling in ecology and behaviour?” Journal of Animal Ecology, Vol. 75, 1182–1189.