تحليل الأنحدار الخطي البسيط بإستعمال أسلوب البرمجة الخطية
مع توظيف أسلوب البواقي المطلقة بطرائق تقدير أخرى

أ.م.د.نزار مصطفى الصراف
أستاذ مساعد في قسم الأحصاء
جامعة بغداد

م.د.فراس أحمد محمد المهنا
مدرس في قسم الأحصاء
جامعة بغداد

م.د.عمر عبدالمحسن علي
مدرس في قسم الأحصاء
جامعة بغداد

Dr. Nazar M. Al-Sarraf
Dep. Of Statistics
University of Baghdad
nazaralsarraf@yahoo.com

Dr. Firas A. Mohamed
Dep. Of Statistics
University of Baghdad
firasmohana@yahoo.com

Dr. Omar A. Ali
Dep. Of Statistics
University of Baghdad
dromar_72@yahoo.com

المستخلص

يتناول هذا البحث تقديم طريقة غير معتادة لتحليل الأنحدار الخطي البسيط ألا وهي بإستعمال البرمجة الخطية بأسلوب المرحلتين المتعارف عليه في فرع بحوث العمليات (Operation Research). إذ يتم إيجاد الحل الأمثل لمسألة البرمجة الخطية وذلك بإضافة متغيرات إصطناعية Artificial Variables: R_i الى معادلة المسألة ذاتها. وكذلك تم إستعمال طريقة أخرى لتحليل الإنحدار الخطي البسيط، فبدلاً عن إيجاد تقدير لمتوسط الأستجابة الشرطي وذلك بتصغير مجموع مربعات البواقي، يتم هنا إيجاد تقدير لوسيط الأستجابة الشرطي وذلك بتصغير مجموع مطلق البواقي – وهو مايطلق عليه بإنحدار الوسيط (Median Regression)–. وكذلك تم الأعتماد على مطلق البواقي كأوزان لطريقة أخرى تدخل ضمن عملية تكرارية لتقدير معلمات الأنحدار.

"Simple Regression Analysis by using Linear Programming Technique and illustration of Absolute Residuals method with another Estimation Techniques"

Abstract

This research deals with unusual approach for analyzing the Simple Linear Regression via Linear Programming by Two - phase method, which is known in Operations Research: "O.R.". The estimation here is found by solving optimization problem when adding artificial variables: R_i. Another method to analyze the Simple Linear Regression is introduced in this research, where the conditional Median of (y) was taken under consideration by minimizing the Sum of Absolute Residuals instead of finding the conditional Mean of (y) which depends on minimizing the Sum of Squared Residuals, that is called: "Median Regression". Also, an Iterative Reweighted Least Squared based on the Absolute Residuals as weights is performed here as another method to estimate the regression coefficients.

1 . المقدمة وهدف البحث

يعدّ تحليل الأنحدار الخطي البسيط (Simple Linear Regression: SLR)، من أدوات التحليل الأحصائي الشائعة من حيث وصف الظاهرة، والتنبوء لها، والسيطرة عليها. إذ تقوم أساساً بالأستناد الى مسألة إيجاد الحل الأمثل الناتج عن تصغير مجموع مربعات البواقي (Residuals: e_i)، والتي ستمثل الفرق بين القيم المشاهدة لمتغير الأستجابة Y والقيم المقدرة له والمتمثلة بالرمز ($\hat{Y}$)، وسينتج عن هذا التصغير حلاً يقدر فيه معلمتي الأنحدار β_0 و β_1 لأيجاد تقدير للمتوسط الشرطي لمتغير الأستجابة، وهو: $E[Y|X]$ ، بينما يمكن الأستعاضة عن عملية التصغير هذه بأخرى تستند الى تصغير مجموع القيم المطلقة للبواقي بمعنى إيجاد تقدير للوسيط الشرطي لمتغير الأستجابة، وهو: $Median[Y|X]$، للحصول على مقدر المعلمتين β_0 و β_1 لما للوسيط من حصانة تجاه الشوارد (Outliers) مقارنة بالمتوسط (Mean). ويمكن مقاربة هذا الوصف الى عملية تصغير دالة هدف تكون دالة للبواقي في حالة البرمجة الخطية (Linear Programming: LP) بأسلوب المرحلتين في بحوث العمليات والمقارنة به مع الأسلوبين السابقين.

لذا يهدف هذا البحث الى تقديم إمكانية لإستعمال أساليب تحليل بديلة للأنحدار الخطي البسيط مثل: أسلوب المرحلتين في البرمجة الخطية عبر دالة هدف وقيود يتم تمثيلها بالأستناد الى البواقي، ومقارنة ذلك بنتائج تحليل طريقة أخرى تستند الى متوسط (أو مجموع) الفروق المطلقة للبواقي. وتم بيان أنه يمكن إختزال مسألة تصغير مجموع القيم المطلقة للبواقي في الأنحدار الى مسألة تصغير دالة هدف في البرمجة الخطية، وتم كذلك تقديم طريقة أخرى تستند في الحل الى الفروق المطلقة للبواقي وتم أخذ طريقة معدّلة لها. وللوصول الى هذا الهدف تم تقسيم البحث الى أربعة أجزاء: يستهل في الجزء الأول، المقدمة وهدف البحث. أما في الجزء الثاني، فقد تم عرض مايخص الجانب النظري لحالة تحليل الأنحدار الخطي البسيط وإستعمال طريقة الفروق المطلقة الصغرى، والطريقة المعدلة عنها والتي تعتمد على إحتساب الوسيط، وثم تقديم طريقة أسلوب المرحلتين كأسلوب لحالة البرمجة الخطية، بالإضافة الى طريقة تكرارية موزونة بمطلق البواقي، وفي الجزء الثالث، تم تقديم الجانب التطبيقي المستند الى بيانات واقعية لمرضى مصابين بداء السكري Diabetes. وفي الجزء الأخير، تم تلخيص الأستنتاجات التي أفرزها البحث.

٢. الجانب النظري

لغرض تقديم الموضوع بشكل مفيد وموجز في آن واحد، فقد تم تقديم إستعمال طرائق مختلفة في تحليل الأنحدار البسيط ، وكما موضح في الفقرات أدناه.

١.٢ طريقة تصغير مجموع مربعات البواقي

لنموذج الأنحدار الخطي:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad \text{... (1)}$$

يتم إيجاد تقديرات β_0 و β_1 عن طريق تصغير مجموع مربعات البواقي المبين أدناه[3]:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \quad \text{... (2)}$$

وهي ما تدعى في بعض الأحيان بمسألة تصغير معيار L_2. وبإشتقاق المعادلة (2) أعلاه بالنسبة لـ (β_0) ولـ (β_1) وبمساواة المشتقات الجزئية المستحصلة للصفر، تتكون ماتدعى "بالمعادلات الطبيعية"ومن إشتقاقهما يتم الحصول على[3]:

$$\hat{\beta}_{1OLS} = \frac{\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right) \Big/ n}{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n} \quad \text{... (3)}$$

$$\hat{\beta}_{0OLS} = \overline{Y} - \hat{\beta}_1 \overline{X} \quad \text{... (4)}$$

أما معادلة الأنحدار التقديرية تكون كما يأتي:

$$\hat{Y}_i = \hat{\beta}_{0OLS} + \hat{\beta}_{1OLS} X_i \quad ; \quad i=1, 2, \ldots, n \quad \text{...(5)}$$

٢.٢ طريقة تصغير متوسط الفروق المطلقة للبواقي

يمكن إستقصاء الحل لأيجاد تقديرات معاملات الأنحدار بإستعمال أسلوب آخر ألا وهو تصغير مجموع القيم المطلقة للبواقي مابين قيم Y المشاهدة و قيمها $\hat{Y}$ التنبؤية، وهي ماتدعى أحياناً بطريقة تصغير معيار L_1 وسندعوها هنا ((Least Absolute Residuals: (LAR (slope)).

فلنموذج الأنحدار الخطي البسيط المذكور في المعادلة (1) أعلاه، يمكن حصر مشاهدات X و Y كأزواج مرتبة بصيغة (X_i , Y_i) إذ أن: $i=1, 2, 3, \ldots, n$ ويكون الهدف من ذلك إيجاد تقديرات β_0 و β_1 من تصغير مجموع القيم المطلقة للبواقي، وكما يأتي[1]:

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i| \quad \text{...(6)}$$

وبأخذ قيد إضافي وهو أن المعادلة التنبؤية تمر بالنقطة $(\bar{X}\ ,\bar{Y})$ لتحقيق المعادلة (5) أي أن:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

وبإستعمال هذا التحويل يتم الحصول على مشاهدات البيانات بصيغة الأنحرافات عن الوسط الحسابي، وكما يأتي:

$$\begin{aligned} x_i &= X_i - \bar{X} \\ y_i &= Y_i - \bar{Y} \end{aligned} \qquad \ldots (7)$$

وعندها يمكن صياغة المسألة من عملية التصغير للمعادلة (5) الى الآتي:

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - \beta_1 x_i|$$

وهنا يتم الحصول على خطين مستقيمين وبقيمة صغرى عند نقطة $(y_i/x_i\ ,\ 0)$ وهي نقطة الحل الأمثل، وبميلين هما: $-|x_i|$ و $|x_i|$. والآن، فلتكن:

$$f(\beta) = \sum_{i=1}^{n} |y_i - \beta x_i| \qquad ;\ i=1, 2, 3, \ldots, n$$

x_{il}: يمثل قيمة x_i المقابلة للرتبة (l).

وهي دالة محدبة خطية قطعية (Convex Linear Piecewise Function)، ولها الخصائص الآتية[2]:

١. إن ميل الجزء الخطي في أقصى جهة اليسار يساوي $-\sum_{i=1}^{n} |x_i|$ ، وكذلك يكون الميل للجزء الخطي في أقصى جهة اليمين يساوي $\sum_{i=1}^{n} |x_i|$.

٢. إذا كانت النقاط من النمط (y_i/x_i) تعطي القيمة الصغرى بحيث تكون:

$$y_{i1}/x_{i1} \le y_{i2}/x_{i2} \le \ldots \le y_{in}/x_{in}$$

فإن ميل $f(\beta)$ سيزداد عند نقطة $\beta_{(K)}$ بمقدار $\sum_{i=1}^{n} |x_{iK}|$ ، إذأن:

$$\beta_{(K)} = y_{(k)}/x_{(k)} \qquad ;\ k=1, 2, \ldots, r$$

ويمكن إستحصال القيمة المثلى للـ (β) عن طريق إيجاد القيمة الصغرى $\beta_{(r)}$ بحيث يكون:

١. المقدار $-\sum_{i=1}^{n} |x_i| + 2\sum_{K=1}^{r-1} |x_{iK}|$ ذي قيمة سالبة ؛

٢. والمقدار $-\sum_{i=1}^{n} |x_i| + 2\sum_{K=1}^{r} |x_{iK}|$ ذي قيمة موجبة وعندها فقط يظهر الحل الأمثل (المقابل للرتبة r).

ويمكن إيجاز العمل الحسابي بهذه الطريقة كما يأتي:

أ. إيجاد قيم المشاهدات x_i و y_i بصيغة الأنحرافات عن الوسط الحسابي كما في المعادلة (7).

ب. تكوين النسب y_i/x_i وترتيبها تصاعدياً ومنحها رتباً (k) ، لتصبح $y_{(k)}/x_{(k)}$.

ج. إعادة النسب المكونة في الفقرة السابقة الى موقعها (تسلسلها) الأصلي حسب المشاهدات مع الأحتفاظ بالرتب الممنوحة لها.

د. الأحتكام الى المعيار:

$$S_{K-1} = -\sum_{i=1}^{n}|X_i| + 2\sum_{L=1}^{K-1}|X_{i_L}| \qquad \text{... (8)}$$

إذ أن: $i=1, 2, \ldots, n$ و $k=1, 2, \ldots, r+1$

هـ. يستقر الحل عند الحصول على الرتبة r والمقابلة للمعيار S_r ذو القيمة الموجبة.

وبهذا يكون الحل عند الخطوة $\beta_{1(r)}$ تكون تقديرات معاملات الأنحدار كالآتي:

$$\hat{\beta}_{1(r)} = \frac{y_{(r)}}{x_{(r)}} \qquad \text{...(9)}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_{1(r)} \cdot \overline{X} \qquad \text{... (10)}$$

3.2 طريقة تصغير مجموع الفروق المطلقة للبواقي (بإستعمال الوسيط)

يمكن النظر الى المعادلة (2) على إنها تبين تباين البواقي عبر المقدار $\sum_{i=1}^{n} e_i^2$ واذ يمكن التعبير عنه بالصيغة:

$$Var(e_i) = \frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2$$

إلاّ أن:

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)}{n}$$

$$\bar{e} = 0$$

ولكن يمكن التعبير عن $\sum_{i=1}^{n}(e_i - \bar{e})^2$ بالصيغة الآتية:

$$\sum_{i=1}^{n}(e_i-\bar{e})^2 \equiv \frac{\sum_{i<j}(e_i-e_j)^2}{n}$$

وبذلك يمكن إستبدال علامة "التربيع" بـ "المطلق" لإيجاد تقديرات β_0 و β_1 وعبر تحقيق معيار التصغير الآتي:

$$Minimize \sum_{i<j}|e_i-e_j|$$

وبإعادة صياغة المعادلة رقم (6) ولكن بالتعبير عنها بـ (e_i) و (e_j) يمكن الحصول على[4]:

$$\sum_{i<j}|e_i-e_j| = \sum_{i<j}|(Y_i-\beta_0-\beta_1X_i)-(Y_j-\beta_0-\beta_1X_j)|$$
$$= \sum_{i,j}|(Y_i-Y_j)-\beta_1(X_i-X_j)|$$
$$\therefore \sum_{i<j}|e_i-e_j| = \sum_{i<j}|Y_{ij}-\beta_1X_{ij}|$$

إذ أن: $Y_{ij} = Y_i - Y_j$ وإن: $X_{ij} = X_i - X_j$ ؛ لكل $i < j$.

وبهذا يكون هنالك $n(n-1)/2$ من النقاط التي يتم أخذ المعلومات منها وكما في الصيغة الآتية لتقدير معلمة مقطع الأنحدار وهي β_0 :

$$\hat{\beta}_0 = \underset{i<j}{median}\left\{\frac{1}{2}(Y_i-Y_j)\right\} \quad \text{... (11)}$$

أما معلمة الميل β_1 فكما مر في المعادلة (9) أعلاه.

فبدلاً عن إيجاد تقدير لمتوسط الأستجابة الشرطي وذلك بتصغير مجموع مربعات البواقي، يتم هنا إيجاد تقدير لوسيط الأستجابة وذلك بتصغير مجموع مطلق البواقي (وهو مايطلق عليه بإنحدار الوسيط Median Regression)، وسندعوها هنا ((LAR(median)).

4.2 <u>طريقة تكرارية بإعادة وزن البواقي</u>

يمكن إستعمال مقدرات OLS كقيم إبتدائية لعملية تقدير أخرى تستند الى تقدير موزون. وسيتم تسميتها هنا Iterative Reweighted Least Absolute Residuals: IWLAR فللنموذج المذكور في المعادلة رقم (1) وذلك بحسب الخوارزمية الآتية:

١. إيجاد تقديرات OLS كما في المعادلتين (3) و (4) أعلاه، ومنها نجد $\hat{e}$ ،

إذ أن: $\hat{e}_i = y_i - \hat{y}_i$.

2. إحتساب أوزان عددية بالأستناد الى مقلوب مطلق البواقي، وكما يأتي:

$$w_{j,n} = \frac{1}{|e_i|} \quad for \quad i = 1,2,3,\cdots,n \qquad \text{... (12)}$$

3. إيجاد المعاملات الموزونة بالطريقة التكرارية إذ ستقوم بتصغير مجموع المربعات الآتي:

$$\sum_{i=1}^{n} e_i^{*2} = \sum_{i=1}^{n} \left(\sqrt{w_i} Y_i - \beta_0 \sqrt{w_i} - \beta_1 \sqrt{w_i} X_i\right)^2 \qquad \text{... (13)}$$

$$e_i^* = \sqrt{w_i} e_i$$

وبإشتقاق المعادلة (13) أعلاه بالنسبة لـ (β_0) ولـ (β_1) وبمساواة المشتقات الجزئية المستحصلة للصفر، تتكون ماتدعى "بالمعادلات الطبيعية" وبحلها نحصل على:

$$\hat{\beta}_{1\,IWLAR} = \frac{\sum_{i=1}^{n} W_i X_i Y_i - \left(\sum_{i=1}^{n} W_i X_i\right)\left(\sum_{i=1}^{n} W_i Y_i\right) \Big/ \sum_{i=1}^{n} W_i}{\sum_{i=1}^{n} W_i X_i^2 - \left(\sum_{i=1}^{n} W_i X_i\right)^2 \Big/ \sum_{i=1}^{n} W_i} \qquad \text{... (14)}$$

$$\hat{\beta}_{0\,IWLAR} = \overline{Y}_w - \beta_{1\,IWLAR} \overline{X}_w \qquad \text{... (15)}$$

إذ أن:

$$\overline{Y}_w = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i} \quad ; \quad \overline{X}_w = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

$$\hat{Y}_{(IWLAR)_i} = \hat{\beta}_{0(IWLAR)} + \hat{\beta}_{1(IWLAR)} X_i \qquad \text{... (16)}$$

٥.٢ البرمجة الخطية بأسلوب المرحلتين في تحليل الأنحدار الخطي

تعنى طريقة البرمجة الخطية بالأساس على صياغة مشكلة في أحد ميادين العمل (أو لظاهرة معينة) تتضمن وجود قيود ومواد متاحة، ويتم إيجاد أنسب (أمثل) توزيع لهذه المواد على العوامل (أو المتغيرات) الداخلة في دالة الهدف. ويمكن إجراء مقارنة بين متطلبات أسلوبي تحليل الأنحدار في الأحصاء مع البرمجة الخطية في بحوث العمليات كما في الجدول الآتي:

جدول (1) يبين مقارنة بين أسلوبي الأنحدار الخطي والبرمجة الخطية

الصفة	طريقة تحليل الأنحدار الخطي البسيط	البرمجة الخطية بأسلوب المرحلتين
أ. التجميعية (Additivity)	يفضل توفرها، إذ أن يجب أن تكون مكونات نموذج الأنحدار من النوع التجميعي. وهو مايلاحظ من وجود حد المقطع β_0 مضافاً الى حد الميل $\beta_1 x$.	يشترط توفرها، ومعناها أن لا يوجد تفاعل (Interactions) بين الفعاليات أو الأنشطة وهي مايمثلها عوامل (أو متغيرات) دالة الهدف.
ب. اللاسلبية (Non-Negative)	لايشترط توفرها، إذ معظم سواء لقيم المشاهدات أو لقيم معملات الأنحدار قد تكون موجبة أو سالبة.	يشترط توفرها، يجب أن تكون قيم عوامل (أو متغيرات) القرار موجبة (أو غير سالبة). فمن غير المعقول الحصول على كميات مادية ملموسة بقيم سالبة.
ج. قابلية القسمة (أوالكسرية) Divisibility (Fractionality)	يشترط توفرها، إذ أنه غالباً مايتم التعامل في الأنحدار مع متغيرات كمية مستمرة (Continuous) قابلة للكسور.	يشترط توفرها،إذ تكون الأعداد الداخلة في حل مسألة البرمجة الخطية قيماً صحيحة، بل يعني إمكانية الحصول على قيم كسرية (مستمرة) للعوامل (أو المتغيرات). أما إذا كان من الصعب إنتاج كسور من المنتج فعند ذاك يتم اللجوء الى البرمجة الصحيحة (Integer Programming).
د. التأكد (Certainty)	لايشترط توفرها، فالعكس صحيح. إذ أن التحليل يقوم على مبدأ عدم التأكد (Uncertainty).	يفضل توفرها، أي أن هناك معرفة تامة بمدى مساهمة العوامل الداخلة في دالة الهدف ، وإحتياجاتها، والمصادر من المواد المتاحة.
هـ. التناسبية (Proportionality)	يشترط توفرها، وهو مايلاحظ من إعطاء وزن لمساهمة المتغير التوضيحي x في تفسير معادلة الأنحدار التقديرية $\hat{y}$.	يشترط توفرها، أي أن معيار الأنجاز في حل المسألة الخطية يكون حاصل جمع المساهمات المختلفة في دالة الهدف، بحيث تتناسب الكميات المستعملة في مختلف المواد مع إحتياجات العوامل المختلفة لكل من هذه المواد.

ومن أهم أنواع البرمجة الخطية مايدعى بأسلوب المرحلتين. إذ يتم إيجاد الحل الأمثل لمسألة البرمجة الخطية وذلك بإضافة متغيرات إصطناعية Artificial variables الى معادلة المسألة ذاتها. ومن الجدير بالذكر أن هذه الطريقة تستعمل في حالة وجود قيود بحالة (مساواة =) أو (أكبر أو يساوي ≥).

فلنموذج الأنحدار الخطي البسيط المذكور في المعادلة (1) وللمعادلات الطبيعية المذكورة بالصيغة (2) أعلاه، ولأزواج المشاهدات(X_i , Y_i) إذ أن: i=1, 2, 3, …, n ويكون الهدف من ذلك إيجاد تقديرات β_0 و β_1. وبذا يمكن إستعمال أسلوب البرمجة الخطية الآتية لحل مسألة التصغير:

Minimize $f(\beta) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$

subject to :

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

$$\beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

إذ يتم تطبيق المرحلة الأولى (Phase I) من أسلوب المرحلتين (two – phase) كأحد أساليب البرمجة الخطي لحل المعادلة أعلاه، وكما يأتي:

إنطلاقاً من حقيقة أن: $\max\{f(x)\} = \min\{-f(x)\}$ **يمكن عرض المسألة كالآتي:**

Maximize $Z_0 = -R_1 - R_2$

subject to :

$$n\beta_0 + c_1\beta_1 + R_1 = c_2$$

$$c_1\beta_0 + d_1\beta_1 + R_2 = d_2$$

R_1 & R_2 : متغيرات إصطناعية (Artificial variables)

$R_j \geq 0$; j= 1, 2

$\beta_i \geq 0$; i=0, 1

ويمكن كتابة المسألة أعلاه بالصيغة:

Maximize $Z_0 = (n + c_1)\beta_0 + (d_1 + c_1)\beta_1 - (d_2 + c_2)$

subject to :

$$n\beta_0 + c_1\beta_1 + R_1 = c_2$$

$$c_1\beta_0 + d_1\beta_1 + R_2 = d_2$$

$R_j \geq 0$; j= 1, 2

$\beta_i \geq 0$; i=0, 1

إذ أن:

$c_1 = \Sigma x_i$, $c_2 = \Sigma y_i$, $d_1 = \Sigma x_i^2$, $d_2 = \Sigma x_i y_i$

أما الجدول الأبتدائي لحل مسألة البرمجة الخطية أعلاه، فهو:

جدول (2) يبين الجدول الأساسي لبدأ الحل بأسلوب المرحليتن بأستعمال معلومات الأنحدار

B.V.	β_0	β_1	R_1	R_2	R.H.S.
R_1	n	c_1	1	0	c_2
R_2	c_1	d_1	0	1	d_2
Z	-n-c1	$-d_1-c_1$	0	0	$-d_2-c_2$

أما عملية تحسين الحل فهي كما في طريقة المرحلتين العادية تماماً (للأستزادة راجع المصدر5).

٥.٢ <u>معايير الكفاءة المستعملة</u>

تم إستعمال ثلاث معايير لمقارنة كفاءة النموذج، وجودة المعيار تعتمد على صغر قيمته فالمعيار ذو القيمة الأصغر يكون أفضل، وهي:

١. متوسط مربعات الخطأ Mean Square Error: MSE

وصيغته

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}$$

٢. متوسط الفروق المطلقة للخطأ Mean Absolute Error: MAE

وصيغته

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n}$$

٣. متوسط الفروق المطلقة النسبية للخطأ Mean Absolute Percentage of Error: MAPE

وصيغته

$$MAPE = \sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

٣. الجانب العملي

١.٣ البيانات

إستند البحث الى بيانات حقيقية عن مرض السكري ولعينة بحجم (n=12) مريض، وكان متغير الأستجابة Y عبارة عن نسبة السكر داخل كرية الدم الحمراء، في حين يمثل المتغير التوضيحي X أعمار المرضى (بالسنوات).

٢.٣ النتائج

تم إستعمال لغة MATLAB version 2008a البرمجية في تحليل بيانات ووضع برنامج خاص لهذا الغرض للطرائق الخمس الآنفة الذكر في الجانب النظري.

فعند تطبيق الصيغ (7) و (8) و (9) و (10)، كتصغير لمتوسط مطلق البواقي LAR(Slope) تم الحصول على نتائج الجدول (3) الآتية:

جدول (3) يبين نتائج تحليل الأنحدار بأسلوب الرتب

i	Y_i	X_i	y_i	x_i	y_i/x_i	Rank
1	7.00	41	-0.60	4.83	-0.124	1
2	6.00	27	-1.60	-9.17	0.175	10
3	7.60	46	0.00	9.83	0.000	4
4	7.70	52	0.10	15.83	0.006	5
5	7.80	68	0.20	31.83	0.006	6
6	7.70	45	0.10	8.83	0.011	7
7	8.40	24	0.80	-12.17	-0.066	3
8	7.30	17	-0.30	-19.17	0.016	8
9	7.50	36	-0.10	-0.17	0.600	12
10	6.60	19	-1.00	-17.17	0.058	9
11	9.30	17	1.70	-19.17	-0.089	2
12	8.90	42	1.30	5.83	0.223	11
Σ	91.80	434	---	---	---	---

أما عند تطبيق محتويات الجدول (2) فيما يخص أسلوب البرمجة الخطية، فتم الحصول على النتائج الآتية:

جدول (4) يبين مراحل تطبيق لأسلوب البرمجة الخطية المكافيء لتحليل الأنحدار الخطي

B.C.	B.V.	β_0	β_1	R_1	R_2	R.H.S.	Ratio

-1	R_1	12	434	1	0	91.8	0.212
-1	R_2	434	18454	0	1	3328.9	0.180
	Z_j-C_j	-446	-18888	0	0	-3420.7	

B.C.	B.V.	β_0	β_1	R_1	R_2	R.H.S.	Ratio
-1	R_1	1.79	0	1	-0.0235	13.511	7.53
0	β_1	0.0235	1	0	0.00006	0.1804	7.67
	Z_j-C_j	-1.79	0	-1	0.023517	-13.511	

B.C.	B.V.	β_0	β_1	R_1	R_2	R.H.S.	Ratio
0	β_0	1	0	0.55765	-0.0131	7.5346	
0	β_1	0	1	-0.0131	3.6262	0.0032	
	Z_j-C_j	0	0	0	0	0	

ويمكن تقديم نتائج تقديرات معلمات الأنحدار بالطرائق الخمس المستعملة في هذا البحث بإيجاز بالجدول الآتي:

جدول (5) يبين تقديرات معلمتي الأنحدار بالطرائق الخمس

الطريقة / المعلمة	OLS	IRWLS	LAR(slope)	LAR(median)	أسلوب المرحلتين (Two-phase)
$\beta 0$	7.5346	7.3747	7.4000	7.6500	7.5346
β_1	0.0032	0.0063	0.0060	0.0130	0.0032

أما نتائج معايير الكفاءة للمقارنة بين طرائق التقدير، فقد تم عرضها بالجدول الآتي:

جدول (6) يبين قيم معايير كفاءة النموذج الخطي البسيط

الطريقة / المعيار	OLS	WLS	LAR(slope)	LAR(median)	أسلوب المرحلتين (Two-phase)
MSE	0.6336	0.6177	0.6223	0.8987	0.6336
MAE	0.7702	0.7747	0.7731	1.0133	0.7702
MAPE	1.0113	0.9765	0.9862	1.4747	1.0113

٤. الأستنتاجات والتوصيات

١. أفرز التحليل بالبرمجة الخطية Two-Phase نتائجاً مطابقة لطريقة OLSمن حيث تقدير

معلمتي الأنحدار أو من حيث معايير الكفاءة للنموذج.

٢. قد تفوقت طريقة IWLAR بمعياري MSE و MAPE على ما سواها من الطرائق الأخرى، وتأتي بالمرتبة الثانية بعدها طريقة LAR(slope) ولمعياري MSE و MAPE بشكل جيد.

٣. تفوقت طريقة OLS على ماسواها من الطرائق بمعيار MAE.

٤. يوصي الباحثون بدراسة تعميم أستعمال أسلوب المرحلتين في الأنحدار الخطي المتعدد (Multiple Regression).

٥. يوصي الباحثون بتقصّي حالة أن تكون معلمات الأنحدار غير مقيدة بأشارة (Unrestricted).

المصادر

1. **Armstrong, Ronald D., and Kung, Mabel Tam; (1978); "Least Absolute Estimates for a Simple Regression Problem"; Applied Statistics; Vol. 27; No. 3; pp. 363-366.**

2. **Barrodale, I., and Roberts, F.D.K.; (1973); "An Improved Algorithm for Discrete L_1 Linear Approximation"; SIAM J. Numer. Anal. 10, 839.**

3. **Draper, N.R., and Smith, H.; (1998); "Applied Regression analysis"; 3rd edition John Wiley & Sons. ; New – York.**

4. **Gutenbrunner, C., and Jureckova, J.; (1992); " Regression rank scores and regression quanriles"; Annals Statistics; 43, 410-414.**

5. **Taha, Hamdy A.; (2007); "Operations Research: An Introduction"; 8th edition , Pearson Education Inc., New-Jersey.**