المجلد 11 العدد 25 | مجلة جامعة الانبار للعلوم الاقتصادية والادارية | السنة/2019م

مقارنة بين طرق تقدير معلمة انحدار الحرف المعممة مع التطبيق على بيانات مرض العجز الكلوي المزمن

A Comparison among generalized ridge regression parameter estimating with application on chronic renal insufficiency

م.م.نـــدى نـــزار العبيدي
Nada Nazar Alobaidy
قسم الاحصاء والمعلوماتية/ كلية علوم الحاسوب/ جامعة الموصل

أ.م. د.زكريا يحيى نوري الجمّال
Dr. Zakariya Y. Algamal
قسم الاحصاء والمعلوماتية/ كلية علوم الحاسوب/ جامعة الموصل
zakariya.algamal@uomosul.edu.iq

تاريخ استلام البحث 2018/1/5 تاريخ قبول النشر 2018/3/30 تاريخ النشر 2019/8/19

المستخلص

يعد نموذج الانحدار احد أهم النماذج الخطية المستخدمة في العديد من المجالات العلمية. قد تتعرض المتغيرات المستقلة الداخلة في بناءه إلى ارتباط عالي بين متغيرين أو أكثر مما يؤثر سلباً على عملية تقدير معلمات الأنموذج بطريقة المربعات الصغرى الاعتيادية. يهدف هذا البحث إلى مقارنة طرق تقدير معلمات نموذج الانحدار الخطي عندما يعاني من مشكلة التعدد الخطي شبه التام عبر طريقة انحدار الحرف المعممة فضلاً عن اقتراح طريقتين لمعلمة التحيز. تم استخدام بيانات مرض العجز الكلوي المزمن كحالة تطبيقية لاثبات اهمية الطريقتين المقترحتين مقارنة بالطرق الاخرى. حيث اظهرت النتائج تفوق هاتين الطريقتين على باقي الطرق الاخرى بالاعتماد على معيار متوسط مربعات الخطأ.

Abstract

The regression model is a well-known model in several real applications. However, it is known that multicollinearity negatively affects the ordinary least squared estimator. To address this problem, a generalized ridge regression estimator has been proposed. The performance of this estimator is fully depending on the biasing parameter. In this paper, numerous selection methods of the biasing parameter are explored and investigated. Our real application results suggest that our proposed methods can bring significant improvement relative to others, in terms of mean squared error.

1.المقدمة:

يعد تحليل الانحدار الخطي من أساسيات علم الإحصاء وأسلوباً مهماً من أساليب الإحصاء التطبيقي ، ويعد واحد من الأدوات الإحصائية الأكثر استعمالاً حيث يعمل على تحليل العلاقة بين المتغيرات ويمكن وصف هذه العلاقة **على شكل أنموذج تحتوي على متغير الاستجابة y وواحد أو أكثر من المتغيرات التوضيحية (Explanatory variables)**

في تطبيقات تحليل الانحدار غالبا ما تحدث مشكلة تعدد العلاقة الخطية والتي أصبحت معروفة لدى العديد من الباحثين الإحصائيين وكذلك معرفة مشاكلها الإحصائية على معلمات أنموذج الانحدار الخطي المتعدد ، إذ تؤدي هذه المشكلة في ابسط حالتها إلى فقدان مقدرات المعلمات لأنموذج الانحدار المقدرة دقتها في التقدير ، حيث تعد هذه المشكلة من المشاكل القائمة الوجود في العديد من البيانات وان وجودها له تأثيرات على تقديرات وتباينات معاملات المربعات الصغرى الاعتيادية (Ordinary Least Square) ، لذا وجب تفادي هذه المشكلة ووضع الحلول المناسبة لها ،إذ

استعملت عدة طرق للتخلص من هذه المشكلة ومعالجتها ومن هذه الطرائق في حال كون وجود مشكلة التعدد الخطي شبه تام هي طريقة انحدار الحرف العام Generalized Ridge Regression .

استخدم الكثير من الباحثين مقدر انحدار الحرف العام حيث قام Kibria عام 2003 بدراسة حول مقدر الحرف العام (GRR) واقترح عدة طرق لتقدير معلمة الحرف K وبين انه مقدر الحرف العام (GRR) يتغلب على جميع المقدرات المتحيزة وذلك تحت معيار متوسط مربعات الخطأ (MSE) ويكون هذا المقدر اسهل في حساب قيمة معلمة الحرف K ، والتي تجعل متوسط مربعات الخطأ لمقدر الحرف العام (GRR) اقل ما يمكن .وإذا كانت القيم المثالية لمعلمة الحرف K تختلف في المعنوية من واحدة إلى أخرى فانه هذا التعميم (generalized) له إمكانية اكبر في تخفيض متوسط مربعات الخطأ (MSE) عن نظيره مقدر الحرف الاعتيادي ، وذلك لان مقدر الحرف الاعتيادي يفترض تساوي قيم الـ K .

2. مشكلة تعدد العلاقة الخطية Multicollinearity Problem

يعد أنموذج الانحدار الخطي المتعدد الأكثر شيوعاً اليوم في تحليل الظواهر لاسيما مع توفر البرامج الإحصائية الجاهزة والتي جعلت إمكانية تجهيز نتائج تحليل الانحدار سهلة وسريعة، فعند دراسة أية ظاهرة يجب تحديد المتغيرات المؤثرة في تلك الظاهرة وصياغة العلاقة بين تلك المتغيرات على هيئة أنموذج ، وأن العلاقة الخطية بين عدة متغيرات أحدهما متغير الاستجابة والباقي متغيرات توضيحية يطلق عليها بـ الانحدار الخطي المتعدد، و يأخذ شكلاً رياضياً خطياً وتكون صيغته كالأتي :

$$y = X\beta + \varepsilon \qquad \dots (1)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

إذ إن :

Y : متغير الاستجابة من الدرجة $(n \times 1)$.

X : يمثل مصفوفة مشاهدات المتغيرات التوضيحية من الدرجة $(n \times p+1)$.

β : المعلمات المجهولة من الدرجة $(p+1 \times 1)$.

p : عدد المتغيرات التوضيحية .

n : عدد المشاهدات.

ε : يمثل موجه الأخطاء العشوائية من الدرجة $(n \times 1)$.

وقد بني نموذج الانحدار على مجموعة من الفروض ومن ابرز هذه الفروض أن لا توجد علاقة خطية تامة أو شبه تامة بين أي المتغيرات التوضيحية وعند تحقق هذه الفروضيه ، نحصل على أفضل تقدير لمعالم الأنموذج باستعمال طريقة المربعات الصغرى الاعتيادية (OLS) وهي من أفضل الطرق حيث تتمتع مقدراتها بخاصية أفضل تقدير خطي غير متحيز (Best Linear unbiased Estimations) (BLUE) . وفي حالة عدم تحقق أو غياب احد هذه

الفرضيه اي وجود علاقة خطية تامة أو شبه تامة بين أي المتغيرات التوضيحية فهنا تظهر ما تسمى بمشكلة التعدد

الخطي .أن مصطلح تعدد العلاقة الخطية يشير إلى وجود ارتباط بين المتغيرات التوضيحية المستعملة في نموذج الانحدار الخطي مما يجعل من الصعب فصل تأثيراتها كل على حدة على متغير الاستجابة . وعند وجود مشكلة تعدد العلاقة الخطية فأن تطبيق طريقة المربعات الصغرى تؤدي إلى مشكلة تضخم في تباينات معاملات الانحدار المقدرة ويتمثل هذا التضخم بالعناصر القطرية لمصفوفة $X'X$ لذلك نلجأ إلى استعمال الطرق المتحيزة للتخلص من هذه المشكلة (النعيمي،2005).

1.2 اساليب الكشف عن مشكلة التعدد الخطي

توجد عدة طرائق لغرض الكشف عن مشكلة التعدد الخطي في أنموذج يحتوي على متغيرين توضيحين أو أكثر منها:

1. معامل تضخم التباين(VIF) Variance Inflation Factor

يتم حساب معامل تضخم التباين لكل متغير من المتغيرات التوضيحية، إذ يستفاد منه في قياس مدى ارتباط كل متغير توضيحي مع المتغيرات الأخرى في الأنموذج، فإذا كانت قيمة معامل التضخم ($VIF > 10$) فإنه يدل على أن هناك مشكلة التعدد الخطي بين المتغيرات التوضيحية، ويكون سبباً كافياً لإهمال المتغير X_j من التحليل أو استخدام طريقة أخرى كبديل عن طريقة المربعات الصغرى، ولإيجاد قيمة معامل تضخم التباين VIFتستخدم الصيغة الآتية:

$$\mathrm{VIF_j} = \frac{1}{(1 - R_j^2)} \qquad \dots \qquad (2)$$

R_j^2: تمثل معامل التحديد لانحدار المتغير التوضيحي X_j على بقية المتغيرات التوضيحية $j = 1,2,\cdots,m$.

2. إيجاد العدد الشرطي (C.N.) Condition Number

إنّ هذا المقياس يعتمد بالأساس على الجذور والمتجهات الذاتية (المميزة) ، ويمثل النسبة بين أكبر وأصغر جذر مميز ناتج عن تحليل مصفوفة المعلومات (X'X)، وان تحديد قيمة العدد الشرطي يبين لنا درجة التعدد الخطي بين المتغيرات التوضيحية، فإذا كانت قيمة العدد الشرطي واقعة بين 30<C.N.<100، دل ذلك على وجود علاقة خطية متعددة بين المتغيرات التوضيحية، وإذا كانت قيمة العدد الشرطي C.N.>100 فالتعدد على درجة مرتفعة جداً. والصيغة الرياضية لها تتمثل كالآتي:

$$\mathrm{C.N.} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \qquad \dots \qquad (3)$$

إذ إن :

λ_{max}: أكبر قيمة مميزة لمصفوفة المعلومات X'X.

λ_{min}: أصغر قيمة مميزة لمصفوفة المعلومات X'X.

(Muniz & Kibria,2009)

2.2 اساليب معالجة مشكلة تعدد العلاقة الخطية

هناك العديد من الطرائق لحل مشكله التعدد الخطي منها طريقة انحدار الحرف التي اقترحت في عام (1970) من قبل الباحثان (Hoerl & Kennard) وهي طريقة بديلة لطريقة المربعات الصغرى الاعتيادية (OLS) لتقدير معالم انموذج الانحدار الخطي المتعدد في حالة وجود مشكلة التعدد الخطي وذلك لما تعانيه طريقة المربعات الصغرى الاعتيادية من تضخم في تباينات المقدرات وتعتبر طريقة انحدار الحرف الاعتيادية (Ordinary Least Square) أكثر الطرائق شيوعاً في معالجة مشكلة التعدد الخطي الشبه التام و في عام (1995) اقترح العالمان (Akdeniz & Kaciranlar) مقدر متحيز لمعالجة مشكلة التعدد الخطي ويرمز له بالرمز (GLE) .

3.مقدر انحدار الحرف العام :

عند وجود تعدد العلاقة الخطية فأن مقدرات طريقة المربعات الصغرى تعاني من تضخم في تباينات المعلمات المقدرة وحصول عدم استقرار ، ويتمثل هذا التضخم بالعناصر القطرية للمصفوفة (X'X) ، لذلك اقترحت طريقة انحدار الحرف (RR)(Ridge Regression) إذ إن طريقة انحدار الحرف تعتبر احد بدائل التقدير عندما توجد هناك مشكلة التعدد الخطي شبه التام بين المتغيرات التوضيحية ، وهي ليست جديدة فقد قدمها لأول مرة (1970) (Horal & Kennard) واستعمالها في تقدير المعلمات لأنموذج الانحدار الخطي المتعدد حيث أوضحا أن التأثيرات غير مرغوب فيها والناتجة من التعدد الخطي يمكن تخفيضها باستعمال أسلوب (Ridge Regression) بدلاً من (OLS) و تعتبر طريقة انحدار الحرف احد طرق معالجة مشكلة التعدد الخطي شبه التام في أنموذج الانحدار الخطي العام(GLM)General Linear Regression Model .

وبعدها تم اقتراح مقدر من قبل (Horal & Kennard) عام (1970 b) يسمى مقدر انحدار الحرف العام حيث يعمل على تقليل التباين في مقدر انحدار الحرف الاعتيادي (ORR) بالإضافة إلى تقليل التحيز لمقدر ORR ، وتتمثل بإضافة قيم مختلفة إلى عناصر القطر الرئيسي للمصفوفة المعلومات (X′X) ، وتعتبر هذه المقدرات الحالة العامة لمقدرات الحرف الاعتيادية (ORR) .

أن مقدرات انحدار الحرف الاعتيادية (Ordinary Ridge Regression) تتمثل بإضافة قيمة ثابتة K إلى عناصر القطر الرئيسي للمصفوفة X′X ، وقد افترض في هذا المقدر وجود مصفوفة zمصفوفة ذات درجة $(p \times p)$ وأعمدتها تمثل المتجهات المميزة $(v_1, v_2, v_3, \dots, v_p)$ لمصفوفة (X'X) ، وأن $z'z = I$ ومن هنا فأن :

$$z'(X'X)z = \Lambda = diag(\lambda_1, \lambda_2, \dots \lambda_z)$$

إذ بإمكاننا إعادة كتابة الأنموذج الخطي العام (1) بالشكل :

$$y = X^*\alpha + \varepsilon \qquad \dots (4)$$

إذ أن :

$$x^* = XP \quad , \quad \alpha = z'\beta$$

z : تمثل مصفوفة متعامدة أعمدتها تمثل المتجهات المميزة المقابلة للجذور المميزة للمصفوفة المعلومات (X'X)

حيث ان :

$$Z'Z = ZZ' = I$$

وأن تقدير المربعات الصغرى الاعتيادية لـ (α) يعطي كآلاتي :

$$\hat{\alpha}_{OLS} = (X^{*'}X^*)^{-1} \chi^{*'} y$$

$$= \Lambda^{-1} \chi^{*\prime} y \qquad \dots \quad (5)$$

علما إن معاملات الانحدار الأصلية لـ (OLS) هي :

$$\underline{\hat{\beta}}_{OLS} = Z\underline{\hat{\alpha}}_{OLS}$$

وأن مقدرات الحرف الاعتيادية (ORR) لـ α تعطى بالشكل :

$$\hat{\alpha}_{ORR} = \left(\chi^{*\prime}\chi^{*} + kI\right)^{-1} \chi^{*\prime} y$$

$$= (\Lambda + kI)^{-1} \Lambda\, \hat{\alpha}_{OLS}$$

$$\underline{\hat{\beta}}_{ORR} = P\underline{\hat{\alpha}}_{ORR}$$

و انه في حالة كون k= 0 فانه مقدرات الحرف الاعتيادي(ORR) تتحول إلى مقدرات المربعات الصغرى الاعتيادية.

إما مقدر انحدار الحرف العام (GRR) (Generalized Ridge Regression Estimator) لـ (α) فيعطى بالصيغة الآتية :

$$\hat{\alpha}_{GRR} = \left(\chi^{*\prime}\chi^{*} + \underline{K}\right)^{-1} \chi^{*\prime} y$$

$$= \left(\Lambda + \underline{K}\right)^{-1} \chi^{*\prime} y \qquad \dots \quad (6)$$

$$= \left(I + \underline{K}\,\Lambda^{-1}\right)^{-1} \hat{\alpha}_{OLS}$$

وعليه فأن :

$$\hat{\beta}_{GRR} = Z\hat{\alpha}_{GRR}$$

إذ إن :

K : تمثل مصفوفة قطرية بالمدخلات $\left(k_1, k_2, k_3, \dots, k_p\right)$، وهذه قيم الـ K تعمل على تقليل MSE لمقدر انحدار الحرف العام (Generalized Ridge Regression).وفي حالة كون K متساوية فأن مقدر الحرف العام(GRR) تتحول إلى مقدر(ORR).

و أن التباين لمقدرات انحدار الحرف العام (GRR) بالشكل الأتي :

$$Var\left(\hat{\alpha}_{GRR}\right) = C_K\, Var\left(\hat{\alpha}_{OLS}\right) C_K'$$

$$= \hat{\sigma}^2 \left(C_K\, \Lambda^{-1}\, C'_K\right) \qquad \dots \quad (7)$$

حيث أن :

$$C_K = \left(1 + K\Lambda^{-1}\right)^{-1}$$

$\hat{\sigma}^2$: مقدر المربعات الصغرى الاعتيادية لـ σ^2 .

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'_{OLS} X^{*\prime} Y}{n - m - 1}$$

وكما نلاحظ من المعادلة (6) أن مقدر الحرف العام هو متحيز للمعلمة الأصلية ، وأن مقدار التحيز هو:

$$Bias\,(\hat{\alpha}_{GRR}) = E\,(\hat{\alpha}_{GRR}) - \alpha$$

$$= C_K\,\alpha - \alpha \qquad = (C_K - I)\,\alpha$$

و مصفوفة متوسط مربعات الخطأ لمقدرات انحدار الحرف العام (GRR) بالشكل الأتي :

$$MSE\,(\hat{\alpha}_{GRR}) = Var(\hat{\alpha}_{GRR}) + \left(Bias(\hat{\alpha}_{GRR})\right)^2$$

$$MSE\,(\hat{\alpha}_{GRR}) = \hat{\sigma}^2\,\left(C_K\,\Lambda^{-1}\,C'_K\right) + (C_K - I)\alpha\alpha'(C_K - I)'$$

$$= \hat{\sigma}^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + K_i)^2} + \sum_{i=1}^{p} \frac{K_i^2\,\alpha_i^2}{(\lambda_i + K_i)^2} \qquad \ldots (8)$$

وقد أثبت كلاً من (Viond & Ullah) عام 1981م بأن

$$MES\left((\hat{\alpha}_{GRR})\right) \le MES\left((\hat{\alpha}_{OLS})\right)$$

4.اساليب تقدير معلمة الحرف K

اقترح العديد من الباحثين عده طرائق جديدة لتقدير المعلمة K منها ما يأتي :

1.4الطريقة المقترحة من قبل (Hoerl and kennard,1970):

$$k_i(HK) = \frac{\hat{\sigma}^2}{\hat{\gamma}_i^2} \qquad ,i = 1,2,,\ldots,p \qquad \ldots \quad (9)$$

2.4 الطريقة المقترحة من قبل (Nomura,2011)

$$k_i(N) = \frac{\hat{\sigma}^2}{\hat{\gamma}_i^2}\,\{1 + [1 + \lambda_i(\hat{\gamma}_i^2/\hat{\sigma}^2)]\} \quad i = 1,2,,\ldots,p \qquad \ldots \quad (10)$$

3.4 الطريقة المقترحة من قبل (Troskie and Chalton ,1996) :

$$k_i(TC) = \frac{\lambda_i\hat{\sigma}^2}{\lambda_i\hat{\gamma}_i^2 + \hat{\sigma}^2} \qquad i = 1,2,,\ldots,p \qquad \ldots(11)$$

4.4 الطريقة المقترحة من قبل (Firinguetti,1999) :

$$k_i(F) = \frac{\lambda_i\hat{\sigma}^2}{\lambda_i\hat{\gamma}_i^2 + (n-p)\,\hat{\sigma}^2} \qquad i = 1,2,,\ldots,p \qquad \ldots(12)$$

5.4 الطريقة المقترحة من قبل (Hockin et ai.,1999) :

$$k_i(HSL) = \hat{\sigma}^2\,\frac{\sum_{I=1}^{p}\ (\lambda_i\hat{\gamma}_i)^2}{\sum_{I=1}^{p}\ (\lambda_i\hat{\gamma}_i^2)^{\,2}} \qquad i = 1,2,,\ldots,p \qquad \ldots(13)$$

6.4 الطريقة المقترحة من قبل (Al-Hassan,2010) :

$$k_i(AH) = \hat{\sigma}^2 \frac{\sum_{I=1}^{p} (\lambda_i \hat{\gamma}_i)^2}{\sum_{I=1}^{p} (\lambda_i \hat{\gamma}_i^2)^2} + \frac{1}{\lambda_{max}} \quad i = 1,2,,\dots,p \qquad \dots(14)$$

7.4الطريقة المقترحة من قبل (Dorugade,2014) :

$$\dots(15) \qquad k_i(D) = \frac{2\hat{\sigma}^2}{\lambda_{max}\hat{\gamma}_i^2} \quad i = 1,2,,\dots,p$$

5. الطرق المقترحة:

من اجل اتمام وتحقيق هدف البحث، تم اقتراح استخدام الطريقة من قبل الباحثين (Bhat and Vidya,2016) في حالة مشكلة التعدد الخطي شبه التام عبر طريقة انحدار الحرف المعممةوالمعرفة بالمعادلتين الاتيتين :

$$k_{new1} \dots(16) \qquad =\frac{p\hat{\sigma}^2}{\hat{\gamma}'\hat{\gamma}} + \frac{1}{\lambda_{max}\hat{\gamma}'\hat{\gamma}} = k_{HKB} + \frac{1}{\lambda_{max}\hat{\gamma}'\hat{\gamma}}$$

$$k_{new2} \dots(17) \qquad =\frac{p\hat{\sigma}^2}{\hat{\gamma}'\hat{\gamma}} + \frac{1}{2(\sqrt{\lambda_{max}/\lambda_{min}}} = k_{HKB} + \frac{1}{2CN}$$

6.الجانب التطبيقي:

لغرض اتمام الفائدة المرجوة من البحث ، تم التطبيق على بيانات تتبع التوزيع الطبيعي والتي أخذت من بيانات استخدمت من قبل (عبدالله واخرون،2011) حول مرض الفشل الكلوي المزمن وعلاقته بإنزيم ميتالوإندوببتايديز. حيث تم جمع (73) نموذج دم لأشخاص مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وتم سحب نماذج الدم لمجموعة المرضى من قبل اجراء عملية الغسيل الكلوي التي تستغرق (4−3) ساعات، وقد شخصت حالة المرضى من قبل اطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي – وحدة الكلية الاصطناعية، تراوحت اعمارهم بين (80−20) سنة ، وتتضمن (38) نموذجا للذكور و (35) نموذجا للإناث، ودونت المعلومات الخاصة بالمرضى على وفق استمارة استبيان خاصة لكل مريض اعدت لهذا الغرض لسنة 2013، حيث سجلت الدراسة ثمانية متغيرات توضيحية والتي يعتقد بان لها تأثير في متغير الاستجابة الذي يمثل نسبة إنزيم ميتالوإندوببتايديز. يوضح الجدول (1) وصف المتغيرات التوضيحية المستخدمة في الدراسة.

جدول (1) : وصف المتغيرات المستقلة المستخدمة في الدراسة

رمز المتغير التوضيحي	وصف المتغير التوضيحي	وحدة القياس
X1	الجنس	(ذكر =1، انثى =2)
X2	العمر	سنوات
X3	مدة المرض	الايام
X4	الوراثة	(نعم=1، كلا=2)
X5	القلب	(نعم=1، كلا=2)

X6	السكري	(نعم=1، كلا=2)
X7	التدخين	(نعم=1، كلا=2)
X8	الضغط	(نعم=1، كلا=2)
X9	امراض اخرى	(نعم=1، كلا=2)
X10	نسبة اليوريا	(ملي مول/لتر)
X11	نسبة البروتين الكلي	غرام/100 ميليلتر
X12	نسبة الالبومين	غرام/100 ميليلتر
X13	نسبة الكلوبيولين	غرام/100 ميليلتر
X14	نسبة الكالسيوم	غرام/100 ميليلتر
X15	نسبة البوتاسيوم	غرام/100 ميليلتر
X16	نسبة الصوديوم	غرام/100 ميليلتر
X17	نسبة الخارصين	غرام/100 ميليلتر
X18	نسبة النحاس	غرام/100 ميليلتر
X19	نسبة المغنسيوم	غرام/100 ميليلتر

ولغرض اختبار التوزيع الاحتمالي المناسب لمتغير الاستجابة فقد تم استخدام اختبار مربع كاي لحسن المطابقة. إذ اظهرت نتيجة هذا الاختبار بان متغير الاستجابة يتبع التوزيع طبيعي بمتوسط مقداره 17.634 وتباين مساوي الى 7.033، حيث بلغت قيمة الاختبار 9.853 وبقيمة p-value مساوية الى 0.1309 عند مستوى معنوية 0.05.

وعليه فان نموذج الانحدارالخطي المقدر الملائم لوصف متغير الاستجابة مع المتغيرات التوضيحية يعرف بالشكل الاتي

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1X_1 + \hat{\beta}_2X_2 + \hat{\beta}_3X_3 + \hat{\beta}_4X_4 + +\hat{\beta}_5X_5 + \hat{\beta}_6X_6 + \hat{\beta}_7X_7 + \hat{\beta}_8X_8 + \hat{\beta}_9X_9 + \hat{\beta}_{10}X_{10} + \hat{\beta}_{11}X_{11} + \hat{\beta}_{12}X_{12} + \hat{\beta}_{13}X_{13} + \hat{\beta}_{14}X_{14} + \hat{\beta}_{15}X_{15} + \hat{\beta}_{16}X_{16} + \hat{\beta}_{17}X_{17} + \hat{\beta}_{18}X_{18} + \hat{\beta}_{19}X_{19}$$

لغرض التحري على وجود مشكلة التعدد الخطي، تم استخدام معيار العدد الشرطي condition number (CN) ، فاذا كانت قيمة هذا المعيار اكبر من 100 فهذا يدل على ان مشكلة التعدد الخطي موجودة. بالاعتماد على الجذور المميزة ، فانه يمكن حساب العدد الشرطي بالشكل الرياضي التالي:

$$C.N = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

حيث $\lambda_{\max}$ تمثل الجذر المميز الأكبر للمصفوفة (X^TX) وان $\lambda_{\min}$ تمثل الجذر المميز الأصغر للمصفوفة (X^TX). الجدول (2) يوضح الجذور المميزة للمتغيرات التوضيحية المدروسة. من خلال الجدول (2) نلاحظ بان قيمة معيار العدد الشرطي مساوية ال 14028.662 وهي اكبر بكثير من 100 مما يدل على ان مشكلة تعدد العلاقة الخطية تكن موجودة في البيانات الحقيقية المدروسة.

جدول (2): الجذور المميزة للمتغيرات التوضيحية قيد الدراسة

القيمة	الجذر المميز
3.40×10^{8}	$\lambda1$
1.12×10^{6}	$\lambda2$
3.04×10^{4}	$\lambda3$
5.78×10^{3}	$\lambda4$
3.91×10^{3}	$\lambda5$
0.080	$\lambda6$
0.0169	$\lambda7$
0.296	$\lambda8$
0.274	$\lambda9$
0.111	$\lambda10$
9.05	$\lambda11$
6.42	$\lambda12$
4.97	$\lambda13$
3.85	$\lambda14$
2.59	$\lambda15$
2.40	$\lambda16$
2.23	$\lambda17$
2.09	$\lambda18$
1.73	$\lambda19$

لغرض استخدام نموذج الانحدار الخطي بوجود مشكلة التعدد الخطي، فانه تم تقدير معلمات هذا النموذج باستخدام مقدر الانحدار المعمم وبالاعتماد على طرق تقدير المصفوفة K التي سبق وان تم استعراضها في الجانب النظري وكذلك مقارنة هذه الطرق مع الطريقة المقترحة. الجدول (3) والجدول (4) يوضح نتائج قيم المعلمات المقدرة و متوسط مربعات الخطأ للنموذج المقدر بالطرق المستخدمة.

الجدول رقم (3):المعلمات المقدرة بالطرق المستخدمة

	(OLS)	(AH)	(D)	(HK)	(HSL)	(TC)	(F)	(N)	New1	New2
$\hat{\beta}1$	0.416	0.101	0.416	0.0615	0.101	0.0185	-0.01754041	0.842	-0.0175404	-1.01×10^{9}
$\hat{\beta}2$	0.03.66	0.0334	0.0366	0.0338	0.0334	0.0365	0.038456712	3.29×10^{8}	0.0384567	-1.22×10^{4}
$\hat{\beta}3$	-1.69	-0.534	-1.69	-0.133	-0.534	-0.139	-0.16583079	1.02×10^{7}	-0.1658308	1.43×10^{4}
$\hat{\beta}4$	5.53×10^{5}	6.25×10^{5}	5.53×10^{5}	5.13×10^{5}	6.25×10^{5}	5.21×10^{5}	6.94281×10^{5}	1.46×10^{4}	6.943E-05	4.70×10^{5}
$\hat{\beta}5$	0.377	0.165	0.377	0.342	0.165	0.382	0.442038728	7.85×10^{4}	0.4420387	-4.62×10^{4}
$\hat{\beta}6$	0.371	0.154	0.371	0.2.92	0.154	0.297	0.297866422	4.16×10^{3}	0.2978664	-9.95×10^{4}
$\hat{\beta}7$	-0.201	-0.0383	-0.201	-0.891	-0.0383	-0.0628	-0.11942822	1.69×10^{3}	-0.1194282	3.03×10^{3}
$\hat{\beta}8$	0.305	0.236	0.305	0.021	0.236	0.106	0.281712836	1.84×10^{4}	0.2817128	2.64E-02
$\hat{\beta}9$	0.758	0.185	0.758	0.0436	0.185	0.149	0.592324943	4.55×10^{4}	0.5923249	-7.47E-02
$\hat{\beta}10$	-1.22×10^{3}	1.85×10^{3}	-1.22×10^{3}	1.78×10^{3}	1.85×10^{3}	-6.33×10^{4}	-0.00408066	0.0395	-0.0040807	-3.10×10^{3}
$\hat{\beta}11$	0.896	1.03	0.896	0.454	1.03	0.823	0.918766999	0.0.04	0.918767	1.45
$\hat{\beta}12$	0.629	0.235	0.629	0.689	0.235	0.499	0.553117062	0.377	0.5531171	0.656
$\hat{\beta}13$	1.84	1.16	1.84	1.86	1.16	1.64	1.649050165	1.05	1.6490502	1.59
$\hat{\beta}14$	-8.22×10^{3}	2.18×10^{3}	- 8.22×10^{3}	1.78×10^{3}	2.18×10^{3}	- 6.61×10^{3}	-0.01484635	7.02×10^{4}	-0.014846	5.54×10^{3}
$\hat{\beta}15$	0.295	0.192	0.295	0.145	0.192	0.203	0.236886106	0.187	0.2368861	0.139
$\hat{\beta}16$	5.13×10^{3}	5.69×10^{3}	5.13×10^{3}	5.02×10^{3}	5.69×10^{3}	4.86v	0.004819732	7.48×10^{3}	0.0048197	5.75×10^{3}
$\hat{\beta}17$	0.168	0.164	0.168	0.175	0.164	0.154	0.130631554	0.220	0.1306316	0.0470
$\hat{\beta}18$	0.100	0.110	0.100	0.0952	0.110	0.0898	0.086145279	0.1.83	0.0861453	0.0722
$\hat{\beta}19$	0.0262	0.170	0.0262	0.219	0.170	0.277	-0.01754041	3.29×10^{8}	0.2865077	0.0767

من خلال النتائج المستحصل عليها وبالاعتماد على معيار متوسط مربعات الخطأ، نلاحظ تفوق الطريقتين المقترحتين على باقي الطرق المستخدمة كونها تعطي اقل قيمة لمتوسط مربعات الخطأ. اذا ابرزت الطريقة المقترحة الثانية والموضحة بالمعادلة (17) تفوقا عن الطريقة المقترحة الاولى والموضحة بالمعادلة (16). اضافة الى ذلك، نلاحظ ورود الطريقة المقترحة من قبل (Firinguetti,1999) والموضحة بالمعادلة (12) بالمرتبة الثانية من حيث قيمة متوسط مربعات الخطأ وكذلك اعتبار طريقة المربعات الصغرى (OLS) كأسوأ طريقة كونها اعطت اعلى قيمة لمتوسط مربعات الخطأ.

جدول رقم (4):قيم متوسطات مربعات الخطأ

MSE	القيمة
OLS	317.0201
HK	185.1684
KN	188.7354
TC	178.1288
F	174.4322
HSL	177.4329
AH	177.4329
Fi	174.6588
$New1$	164.0574
$New2$	163.9138

7.الاستنتاجات والتوصيات :

1- تمثل الطريقتين المقترحتين لتقدير قيمة المصفوفة K الحل الأمثل في مقدر الحرف المعمم عندما تكون مشكلة التعدد الخطي حاضرة بين المتغيرات المستقلة.

2- كانت طريقة مقدرات المربعات الصغرى الاعتيادية الأضعف أداء من بين طرق التقدير السابقة والمقترحة نتيجةَ لكبر قيم متوسط مربعات الخطأ الخاصة بها مقارنةَ بالطرق الأخرى.

3- من خلال الاستنتاجات التي توصل اليها البحث، **نوصـــي** باعتماد الطرق المقترحة في تقدير معلمات مقدر انحدار الحرف المعمم عندما تعاني الدراسات الطبية من التعدد الخطي.

المصادر:

1.الصالحي، حنين مراد يوسف ،2014، "مقارنة بين طرائق تقدير انحدار الحرف العامة في معالجة مشكلة التعدد الخطي شبه التام مع تطبيق عملي"، رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة بغداد.

2.النعيمي ، أسوان محمد طيب 2005 ،"اختيار المتغيرات في انحدار الحرف"، رسالة ماجستير مقدمة إلى كلية علوم الحاسبات والرياضيات (غير منشورة) ، جامعة الموصل .

3.عبد الله ، لقاء سعيد و ذكرى علي علوش واسراء عبد الحق الجراح ،2011 ،" دراسة إنزيم ميتالواندوببتايديز وعلاقته بمرض العجز الكلوي المزمن"، مجلة علوم الرافدين، المجلد 22 ، العدد4 ، ص 71-87-2011.

4.Alheety, M. I. and B. M. G. Kibria (2009)." On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors". Surveys in Mathematics and its Applications, 4, 155-167.

5.Arthur , E. Hoerl , Robert W. Kennard (1970-a) ,"Ridge Regression Biased Estimation for Nonorthogonal Problems" ; Technometrics Vol.12 , No.1 , pp.55-60.

6. Akdeniz, F., Kacıranlar, S. (1995)." On the almunbiased generalized Liu estimator and unbiased estimation of the bias and MSE". *Common. Statist. Theor. Meth.* 24:1789–1797

7.Arthur, E. Hoerl , Robert W. Kennard (1970-b) , "Ridge Regression Applications to Nonorthogonal Problem" ; Technometrics , Vol.12 , No.1 , pp.69-82 .

8.Satish Bhat and Vidya, R. (2016) , " A Comparative Study on the Performance of New Ridge Estimators ", Pak.j.stat.oper.res. Vol.XII No.2 2016 pp317-325.

9. Satish Bhat & Vidya Raju (2017) ," A class of generalized ridge estimators", Department of Statistics, Yuvaraja's College, University of Mysore, Mysore, Karnataka, India.