مـــجلة جـــامعة بخت الرضا العـــلمية المحكمة، ربع السنوية، المجلد (1)، العدد (38) مارس 2025م
معامل التأثير العربي: (1.12)
ISNN: [1858-6139], Online

استخدام محاكاة مونت كارلو للمقارنة بين طريقتي انحدار المركبات الرئيسية وانحدار الحرف من خلال متوسط مربعات الخطأ (MSE)

محمد بن محمد العجره

تاريخ الاستلام: 12/01/ 2025م تاريخ النشر:30/03/2025م

مستخلص

في أنموذج الانحدار الخطي المتعدد من المفترض أن تكون متغيرات الانحدار مستقلة عن بعضها البعض عندما لا تكون متغيرات الانحدار مستقلة عن بعضها البعض ويكون هناك علاقة خطية مما يجعل الأنموذج غير مناسب وبالتالي قد تكون النتائج غير دقيقة ، لذلك يتم تطبيق طريقتين من طرق الانحدار المتحيزة وهما: طريقة انحدار الحرف (Ridge Regression) وطريقة انحدار المركبات الرئيسية Regression (Principal Components) للحصول على نتائج أكثر دقة. في هذا البحث تم استعمال محاكاة مونت كارلو لتقييم أداء كل من طريقة انحدار الحرف وطريقة انحدار المركبات الرئيسية في حالة وجود مشكلة التعدد الخطي، وتم استعمال متوسط مربعات الخطأ (MSE) كمعيار لتحديد أفضل الطرق أداءًا. توصل البحث إلى أن طريقة انحدار الحرف (RR) قدمت أداء أفضل من طريقة انحدار المركبات الرئيسية (PCR) في حالة وجود مشكلة التعدد الخطي، وأوضحت نتائج المقارنة بين طرق (RR)المختلقة أن الطريقة K_{GM} كانت الأفضل أداءً عند مستوى الارتباط (0,94، 0.25 ، 0,09.) وتوصل البحث لأن الطريقة K_{AM} قدمت أداءًا أفضل عند مستوى الارتباط (0, 16) بينما قدمت الطريقة K_{HKB} أداء أفضل عند مستوى الارتباط(0.49) وكانت الطريقة K_{DK} هي الأفضل عند مستوى الارتباط (0.81) كما توصل لأن طريقة K_{KS} قدمت أفضل أداء عند مستوى الارتباط (0.98).

كلمات مفتاحية: الانحدار الخطي المتعدد، انحدار الحرف، انحدار المركبات الرئيسية، محاكاة مونت كارلو، متوسط مربعات الخطأ.

.

Abstract

In the multiple linear regression model the regression variables are supposed to be independent with each other when the regression variables are non-independent of each other and there is a linear relationship which renders the model inappropriate and therefore the results may be inaccurate, so two regression methods Biased: the method of Ridge regression and the method of Principal Components regression to obtain more accurate results. In this research the monte Carlo simulation was used to evaluate the performance of both the method of Ridge regression and the method of Principal component in the case of the problem of linear multiplicity, the mean squares error were used as the criterion for determining the best performance methods. The research found that the method of Ridge regression a better performance than the method of Principal Components regression in the case of the problem of linear multiplicity, the results of the comparison between different RR methods showed that the K_{GM} method was the best performance at the level of correlation (0.09, 0.25, 0.94) and the research concluded that the way K_{AM} performed better at the level of correlation (0.16) while provided K_{HKB} better performance at correlation level (0.49) and K_{DK} was the best at correlation level (0.81) and found that the K_{KS} method provided the best performance at correlation level (0.98).

Keywords: Multiple linear regression, Ridge regression, Principal components regression, Monte Carlo simulation, Mean square error.

مقدمة

يعرف الانحدار بشكل عام بأنه احد الأساليب الإحصائية المهمة التي تستخدم بشكل واسع جدا لتحديد وتوضيح التأثيرات بين المتغيرات التوضيحية (التفسيرية) والمتغير التابع (متغير الاستجابة)وتستخدم أيضا للتنبؤ عن قيمة المتغير التابع بدلالة المتغيرات التوضيحية بعد إيجاد معادلة الانحدار الخطية [1]. يعتبر الانحدار الخطي المتعدد من النماذج الخطية المستخدمة بكثرة في تحليل بيانات العديد من البحوث في المجالات الاقتصادية والإدارية والاجتماعية والصحية والطبية والعلوم التطبيقية الأخرى، ويستخدم أيضا التحليل بطريقة الرسم الصندوقي(Box Plot) الذي يعتبر من الطرق المهمة في التحليل الإحصائي حيث تتبين بواسطته مدى تشتت القيم عن الوسيط

وأن الفكرة الأساسية للتحليل الصندوقي هي عرض البيانات بالرسم [14] ، أنموذج الانحدار الخطي المتعدد كغيره من النماذج الإحصائية الخطية يقوم على توافر مجموعة من الفروض الإحصائية وعند تخلف واحداً منها أو أكثر يتعرض الأنموذج إلى عدة مشكلات أهمها ظهور ما يسمى بمشكلة عدم ثبات الأخطاء أو مشكلة عدم تجانس التباين أو تخلف فرض استقلال الأخطاء الذي ينتج عنه ما يعرف بمشكلة الارتباط التسلسلي أو الارتباط الذاتي للأخطاء ومنها أيضا مشكلة تخلف الفرض الخاص بعدم وجود ارتباط خطي بين المتغيرات التوضيحية أو ما يعرف بالتعدد الخطي (Multicollinearity) ويعبر عنها أيضا بعدم التعامد (Non-Orthogonal) بين أعمدة مصفوفة المتغيرات التوضيحية [2] $(X'X)$. تعتبر مشكلة التعدد الخطي(Multicollinearity) بين المتغيرات التوضيحية واحدة من أهم وأكثر المشكلات التي تقف عقبة أمام الباحثين عند استخدام تحليل الانحدار الخطي المتعدد وهي تنشأ عندما يتضمن أنموذج الانحدار أكثر من متغير توضيحي وتكون هناك علاقة ارتباط تامة أو قوية جداً بين اثنين أو أكثر من هذه المتغيرات أو بين جميع المتغيرات [2] . في هذا البحث تم استعمال طريقتي انحدار الحرف وانحدار المركبات الرئيسية لمعالجة هذه المشكلة وللحصول على مقدرات دقيقة في أنموذج الانحدار الخطي المتعدد.

الدراسات السابقة

في عام(2006م) قدم Norliza Adnan وآخرون[7] دراسة بعنوان: " دراسة مقارنة على بعض طرق معالجة مشكلة التعدد الخطي" هدفت الدراسة لتقييم أداء ثلاث طرق لتخطي مشكلة التعدد الخطي هي طريقة انحدار الحرف(RR)، انحدار المركبات الرئيسية(PCR) ، وانحدار المربعات الصغرى الجزئية(PLSR)، وتم استخدام محاكاة مونت كارلو ولغرض المقارنة بين أداء هذه الطرق تم الاعتماد على معيار المقارنة متوسط مربعات الخطأ(MSE)، تم توليد عدد(P= 2 , 4 , 6 , 50)من المتغيرات التوضيحية وإحجام عينات(n= 20 , 30, 40 001، 08، 06،)حيث تم التوصل إلى أن طريقة انحدار الحرف قدمت أداء أفضل من انحدار المركبات الرئيسية وانحدار المربعات الصغرى الجزئية في حالة إدخال عدد محدود أو كبير من المتغيرات إلى النموذج(p=2 or 50) كما تم التوصل الى أن طريقتي انحدار الحرف وانحدار المربعات الصغرى الجزئية ذات أداء أفضل من طريقة انحدار المركبات الرئيسية في حالة إدخال عدد متوسط من المتغيرات للنموذج(p=4 or 6).

في عام (2008م) قدم Yazid M. Al-Hassan[10] دراسة بعنوان: " تقييم لبعض مقدرات الحرف باستخدام طريقة مونت كارلو " هدفت الدراسة إلى تقييم بعض مقدرات الحرف وفقاً لاختيار معلمة الحرف ، حيث تم استخدام سبعة طرق لتقدير معلمة الحرف هي) K_{HK} , K_{HSL} , K_{AM} , K_{GM} , K_{KS}(، تم توليد عدد(P= 50 , 10 , 20) من المتغيرات التوضيحية عند مستويات مختلفة من الارتباط (0.99، 0.9، 0.8 ، 0.7) وأحجام عينات(n= 15 , 30 ، 25) وتم استخدام محاكاة مونت كارلو ولتحديد أفضل الطرق أداءً تم استخدام متوسط مربعات الخطأ) MSE(حيث تم التوصل الى أن المقدر (K_{GM}) كان الأفضل أداءً عند القيم المتوسطة لمعامل ارتباط المتغيرات التوضيحية وفي حالة القيم العليا لمعامل الارتباط كان المقدر (K_{HKB}) الأفضل أداءً وفي حالة القيم العالية جداً لمعامل الارتباط (0.99) جميع المقدرات باستثناء المقدر (K_{AM}) كانت لها أداءً أما أفضل أو مثل أداء المقدر (K_{GM})

في عام (2016م) قام الباحث Eledum[12] باستعمال طريقة انحدار الحرف(RR) وطريقة انحدار المركبات الرئيسية(PCR) لمعالجة مشكلة التعدد الخطي في نماذج الانحدار، إذ تم إجراء المقارنة للطريقتين من خلال تجارب المحاكاة باستعمال معيار المقارنة متوسط مربعات الخطأ(MSE) وتم التوصل إلى أن طريقة انحدار الحرف(RR) أفضل من طريقة انحدار المركبات الرئيسية(PCR) لأنها تمتلك اقل (MSE).

مشكلة البحث

في حالة كون متغيرات الانحدار غير متعامدة ومرتبطة ارتباطاً قوياً فأنه لا يمكن اعتماد أنموذج الانحدار الخطي المتعدد بصيغته العادية وذلك بسبب ظهور مشاكل في المتغيرات التوضيحية ، ومن هذه المشاكل هي مشكلة التعدد الخطي(Multicollinearity)

التي تجعل الأنموذج غير مناسب وبالتالي قد تكون النتائج غير دقيقة لذلك يتم استعمال طرائق لمعالجتها مثل طريقة انحدار الحرف وانحدار المركبات الرئيسية.

هدف البحث

يهدف هذا البحث إلى معالجة مشكلة التعدد الخطي التي من الممكن أن تظهر بين المتغيرات التوضيحية من خلال استعمال طريقة انحدار الحرف(Ridge Regression: RR) وطريقة انحدار المركبات الرئيسية(Principal Components Regression: PCR) في أنموذج الانحدار الخطي المتعدد للحصول على مقدرات اكثر دقة من طريقة المربعات الصغرى الاعتيادية في أنموذج الانحدار الخطي المتعدد ، وباستعمال أسلوب المحاكاة يتم المقارنة بين الطريقتين من خلال معيار متوسط مربعات الخطأ (MSE).

فرضيات البحث

1/ لا تقدم طريقة انحدار الحرف (RR) أداء أفضل من طريقة انحدار المركبات الرئيسية (PRC)عند وقوع الأنموذج تحت تأثير مشكلة التعدد الخطي .

2/ لا تظهر طريقة تقدير معلمة الحرف أداء أفضل مقارنة مع طرائق تقدير معلمة الحرف الأخرى.

الجانب النظري للبحث

أنموذج الانحدار الخطي المتعدد Multiple Linear Regression Model :

بفرض أن المتغير Y يعبر عن المتغير التابع ، والمتغيرات($X_1 , X_2 , \dots , X_k$) تعبر عن K من المتغيرات التوضيحية ، وأن عدد المشاهدات هي n فأن المشاهدة التابعةY_i , $i = 1 , 2 , \dots , n$ يمكن التعبير عنها كدالة خطية في مجموعة المشاهدات المفسرة ($X_{i1} , X_{i2} , \dots , X_{ip}$) كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad \dots (1)$$

حيث أن($\beta_0 , \beta_1 , \beta_2 , \dots , \beta_k$) تعبر عن معاملات الانحدار ، ε_i يعبر عن الخطأ العشوائي للمشاهدة رقم i ، $i = 1 , 2 , \dots ,$ n وحيث أن عدد المشاهدات هي n ، ويكون لدينا n من المعادلات يمكن صياغتها في صورة مصفوفات كما يلي [16] :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \cdots & x_{k1} \\ 1 & x_{22} & x_{32} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & x_{3n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$Y = X\beta + E \qquad \dots\dots..(2)$$

حيث أن:

Y: يعبر عن متجه المشاهدات التابعة ، وهو من درجة(nx1) والعنصر رقم i في هذا المتجه هو Y_i .

X: تمثل مصفوفة المشاهدات التوضيحية (المفسرة) ، وهي من درجة((k+1) x n)، والصف رقم i في هذه المصفوفة) ... X_{i2} X_{i1} 1 X_{ik} : هو)

β: يعبر عن متجه معاملات الانحدار ، وهو من الدرجة((k+1)x1)

E: يعبر عن متجه الأخطاء العشوائية ، وهو من درجة(nx1()، والعنصر رقم i هو الخطأ العشوائيε_i .

طريقة انحدار الحرف Ridge Regression Method

تعتبر طريقة انحدار الحرف احد طرق معالجة مشكلة التعدد الخطي للأنموذج العام (General Linear Model)(GLM) ، وتتلخص هذه الطريقة بإضافة كمية صغيرة موجبة تقع قيمتها بين الصفر والواحد1(0>K <)إلى العناصر القطرية لمصفوفة المعلومات(X'X()) للحصول على مقدرات أكثر دقة ، حيث تعمل هذه الطريقة على فك الارتباطات بين المتغيرات التوضيحية [5] ، وتستخدم الصيغة الآتية

في إيجاد تقديرات قيم() βباستخدام طريقة انحدار الحرف على أن يتم تحويل المتغير المعتمد والمتغيرات التوضيحية الى صيغتها المعيارية[9] :

$$\beta(K) = (X'X + KI_p)^{-1}X'y \qquad \dots (3)$$

حيث أن (I) مصفوفة الوحدةIdentity Matrix.

عندما تكون قيمة $K = 0$ فأن تقديرات معلمات طريقة انحدار الحرف تساوي تقديرات معلمات طريقة المربعات الصغرى الاعتيادية ، وعندما تكون($K < 0$) فأن مقدرات انحدار الحرف تميل الى الاستقرار عند قيمة معينة كنسبة للتغيرات في البيانات ولكنها تكون متحيزة ، كما أن متوسط مربعات الخطأ تكون اقل من متوسط مربعات الخطأ لطريقة المربعات الصغرى الاعتيادية لهذا نقبل بمقدار معين من التحيز مقابل تقليل التباين للمقدرات.

لنفترض أن هناك مصفوفة متعامدة (D) و $D'CD = \Delta$ ، حيث$C = X'X$ $\Delta = \mathrm{diag}(\theta_1, \theta_2, \dots, \theta_p)$ تحتوي على القيم المميزة للمصفوفة C ، حيث يكون أنموذج الانحدار الخطي العام [9] :

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

حيث أن $X^* = XD$ و $\alpha = D'\beta$ عليه فأن مقدرات المربعات الصغرى الاعتيادية تكتب بالشكل الاتي:

$$\hat{\alpha} = \Delta^{-1} X^{*\prime} y \qquad \dots\dots\dots\dots(5)$$

وعليه يمكننا القول أن مقدرات انحدار الحرف تعطى وفق الصيغة الاتية:

$$\hat{\alpha}(K) = (X^{*\prime}X^* + K)^{-1} X^{*\prime} y \qquad \dots\dots(6)$$

حيث أن $K = diag(k_1, k_2, \dots, k_p)$, $k_i > 0$

وأن متوسط مربعات الخطأ لطريقة انحدار الحرف يكتب بالشكل الاتي:

$$MSE(\hat{\alpha}(K)) = \sigma^2 \sum_{i=1}^{p} \frac{\theta_i}{(\theta_i + K_i)^2} + \sum_{i=1}^{p} \frac{K_i^2 \alpha_i^2}{(\theta_i + K_i)^2} \qquad \dots\dots(7)$$

$$K_i = \frac{\sigma^2}{\alpha_i^2}$$

حيث σ^2 يمثل تباين الخطأ و α_i يمثل ith عنصر من α .

يهدف انحدار الحرف إلى ايجاد قيمة K التي من شأنها خفض التباين بشكل كبير مقابل الزيادة في قيمة التحيز، نقدم ادناه بعض الطرق لتقدير معلمة الحرف K .

طريقة Hoerl و Kennard:

اقترح كل من Hoerl و Kennard في ورقتهما التي نشرت في عام(1970) إيجاد قيمة K وفقاً للمعادلة (8) التي تؤدي إلى خفض MSE كأول معادلة لحساب قيمة معلمة الحرف [9].

$$\hat{K}_{HK} = \frac{\hat{\sigma}^2}{\max(\hat{\alpha}^2)} \qquad \dots\dots\dots(8)$$

طريقةHoerl و Kennard و Baldwin :

في عام(1975) قدم كل من Hoerl و Kennard و Baldwin صيغة جديدة لتحديد معلمة الحرف المثلى كتقدير ل K وتمثلها المعادلة(9)[9] .

$$K_{HKB} = \frac{P\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{\alpha}_i^2} \qquad \dots\dots(9)$$

حيث أن:

$\hat{\alpha}$ و $\hat{\sigma}^2$ يتم الحصول عليهما من حل طريقة المربعات الصغرى الاعتيادية.

طريقة Lawless و Wang :

قدم كل من Lawless و Wang طريقتهما لايجاد معلمة الحرف في ورقتهما التي نشرت في عام(1976) ويلاحظ أن طريقتهما تعتمد نوعا ما على K_{HKB} حيث تمثلها المعادلة(10)[9] .

$$\widehat{K}_{LW} = \frac{P\hat{\sigma}^2}{\sum_{j=1}^{p} \theta_j \, \hat{\alpha}_j^{\,2}} \quad .....(10)$$

طريقة Hocking وSpeed وLynn :

اقترح كل من Hocking و Speed و Lynn في عام(1976)طريقة جديدة لتحديد معلمة الحرف كما في المعادلة) 11([9] :

$$K_{HSL} = \hat{\sigma}^2 \frac{\sum_{i=1}^{p}(\theta_i \hat{\alpha}_i)^2}{(\sum_{i=1}^{p} \theta_i \hat{\alpha}_i^{\,2})^2} \quad .....(11)$$

طريقةKhalaf وShukur :

اقترح كل من Khalaf و Shukur في عام(2005) طريقة جديدة كتعديل ل K_{HK} لتحديد معلمة الحرف وفقاً للمعادلة (12) : [9]

$$K_{ks} = \frac{\max(\theta_i)\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \max(\theta_i).\max(\hat{\alpha}_i)^2} \quad .....(12)$$

طريقةDeragid و Kashier's :

في عام(2010) اقترح كل من Deragid و Kashier's طريقة جديدة لايجادK تعتبر تعديلا ل K_{HKB} حيث قاما بطرح عامل التضخم VIF بعد ضربه في حجم العينة والمعادلة هي: [11]

$$K_D = max\left(0, \frac{P\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} - \frac{1}{n(\mathrm{VIF_j})_{\max}}\right) \quad .....(13)$$

حيث إذا كان $(\mathrm{VIF_j})_{max}$ كبيراً جداً فأن K_D تقريباً تكون مساوية ل K_{HKB} ، وإذا كان $(\mathrm{VIV_j})_{max}$ قريباً من الواحد فأن K_D يساوي الصفر أو $K_{HKB} - \frac{1}{n}$ ، عليه فأن $0 \geq K_D \leq K_{HKB}$

طريقةKibria's:

في عام2003() اقترح Kibria's استخدام الوسط الحسابي والوسط الهندسي والوسيط لايجاد معلمة الحرف: [51]

$$\widehat{K}_{AM} = \frac{1}{p}\sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}} \quad .....(14)$$

$$\widehat{K}_{GM} = \frac{\hat{\sigma}^2}{(\prod_{i=1}^{p} \hat{\alpha}^2{}_i)^{\frac{1}{p}}} \quad .....(15)$$

$$K_{MED} = \mathrm{Medion}\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}}\right\} \quad , \quad i=1, 2, \dots, p \quad .....(16)$$

Principal Components Regression Method طريقة انحدار المركبات الرئيسية

أن تحليل المركبات الرئيسية هو من أساليب متعدد المتغيرات ، والذي عرفه الباحث(Karl Pearson) سنة (1901)التي تتضمن هذه الطريقة للتخلص من مشكلة التعدد الخطي ، والذي تم تحسينه من قبل(Harold Hoteling) عام (1933)[8] ، تقوم طريقة المركبات الرئيسية على تحويل المتغيرات التوضيحية الأصلية المرتبطة دون حذف أي منها إلى متغيرات جديدة متعامدة تسمى بالمركبات الرئيسية(Principal Components) ، وكل مركبة رئيسية عبارة عن تركيب خطي في المتغيرات المستقلة الأصلية[4]، يتم ترتيب المركبات الرئيسية وفقاً لحجم التباين الذي تستطيع كل مركبة تفسيرية بواسطة المتغيرات التي تتضمنها ، فالمركبة الأولى هي المركبة ذات التباين الأكبر ويليه الثانية وهكذا .

لغرض تحويل المتغيرات التوضيحية(X) المرتبطة إلى متغيرات جديدة) P(غير مرتبطة(Orthogonal) توجد لدينا مصفوفة متعامدة ولتكن(V) تحقق الشروط الآتية: [5]

$1 - V'V = VV' = I$
$2 - V'(X'X)V = \Delta$

حيث أن:

Δ: مصفوفة قطرية ذات مرتبة(kxk) للجذور المميزة(θ_j) لمصفوفة المعلومات($X'X$) ، وان $\theta_1 \geq \theta \leq$ (θ_2 $\theta_3 \ldots \geq \theta_K$

V: مصفوفة ذات مرتبة(kxk) متعامدة(Orthogonal) أعمدتها تمثل المتجهات المميزة للمصفوفة($X'X$).

واعتمادا على المصفوفة V يتم الحصول على مجموعة جديدة من المتغيرات التوضيحية على هيئة تراكيب خطية تدعى بالمركبات الرئيسية)(P)(Principal Components وتكتب بالشكل التالي:

$$P_j = \sum_{j=1}^{k} V_j X_k \qquad \ldots\ldots\ldots(17)$$

ويمكن كتابته بصيغة المصفوفات كالاتي:

$$\mathrm{P} = \mathrm{XV} \qquad \ldots\ldots\ldots(18)$$

حيث أن:

P: تمثل مصفوفة المركبات الرئيسية($P_1 , P_2 ,\ldots , P_k$).

X: تمثل مصفوفة المتغيرات التوضيحية($X_1 , X_2 ,\ldots , X_k$).

V: تمثل مصفوفة المتجهات المميزة للمصفوفة($X'X$).

وتكونهذه التراكيب الخطية على هيئة دوال خطية بدلالة المتغيرات التوضيحية المرتبطة يتم من خلالها الحصول على متغيرات توضيحية جديدة غبر مرتبطة تدعى بالمركبات الرئيسية(Principal Components).

لنفرض أن:

$$P = XV$$

□′□ = □

عليه بالامكان كتابة النموذج بدلالة المركبات الرئيسية(P) أي بدلالة المتغيرات التوضيحية الجديدة المتعامدة كالاتي:

$$\mathrm{Y} = \mathrm{P}\alpha + \mathrm{M} \qquad \ldots\ldots\ldots(19)$$

حيث أن:

P: مصفوفة من مرتبة(nxk) للمركبات الرئيسية (المتغيرات المتعامدة)وبناءًا على ذلك فأن المركب الرئيسي (P) يكتب بالشكل التالي:

$$P_j = XV_j \qquad \ldots\ldots(20)$$

وبتعويض قيمة(P) تصبح المعادلة(20) بالشكل التالي:

$$\hat{\alpha} = \Delta_r^{-1} \mathrm{V_r X Y} \qquad \ldots\ldots\ldots(21)$$

وبالتعويض عن المصفوفتين(Δ ,)V نحصل على مايأتي **[5]**

$$\begin{bmatrix} \hat{\alpha}_r \\ \hat{\alpha}_{k-r} \end{bmatrix} = \begin{bmatrix} \Delta_r & 0 \\ 0 & \Delta_{k-r} \end{bmatrix}^{-1} \begin{bmatrix} V_r \\ V_{k-r} \end{bmatrix} \mathrm{X'Y} \qquad \ldots\ldots(22)$$

وعلى افتراض ان) Δ_{k-r}^{-1}(مساوية للصفر ، عليه فأن مقدرات طريقة المربعات الصغرى) OLS(الى) α_r(تكتب بالشكل الاتي:

$$\widehat{\alpha_r} = \Delta_r^{-1} \mathrm{V_r X'Y} \qquad \ldots\ldots(23)$$

وبما ان لدينا في الجانب التقديري:

$$\widehat{\alpha_r} = \mathrm{V'_r\, b}$$

وعليه يمكننا القول ان مقدرات المركبات الرئيسية(PrincipalComponents) تعطى وفق الصيغة الاتية:

$$\mathrm{B_{p.c}} = \mathrm{V_r}\widehat{\alpha_r} \qquad \ldots\ldots(24)$$

وبتعويض المعادلة)(23)بالمعادلة)24(نحصل على الاتي:

$$b_{p.c} = V_r \Delta_r^{-1} V_r X'Y \qquad \ldots\ldots\ldots(25)$$

والمعادلة(24)تمثل مقدرات متجه المعالم بأستخدام طريقة المركبات الرئيسية.

وان متوسط مربعات الخطأ()MSE)لمقدرات المركبات الرئيسية يكتب بالشكل الاتي:

$$MSE(b_{p.c}) = \sigma_u{}^2 \sum_{i=1}^{r} \theta_i{}^{-1} + \sum_{i=r+1}^{k} (V'_i B)^2 \qquad .....(26)$$

7الجانب التجريبي :

7-1المقدمة :Introduction

يمكن تعريف المحاكاة بصورة عامة على أنها عبارة عن الحلول للمشكلات الرياضية من خلال بناء أنموذج مشابه للمشكلة الأصلية، وقد تم استعمال هذا الأسلوب كثيرا في مجالات الاحصاء المختلفة لدراسة وتطوير الطرائق الاحصائية المختلفة فيما بينها ، يتم استخدام المحاكاة كأسلوب للتحليل عند عدم التمكن من استخدام أساليب التحليل

[3] الأخرى كما وانه يوفر على الباحثين الكثير من الجهد

والكلفة والوقت .

لقد تعددت استعمالات أساليب المحاكاة في المجالات المختلفة ، ولاسيما بعد التطور السريع الذي حصل في استعمال الحاسبة الالكترونية ، اذ أن المحاكاة تسهل الكثير من العمليات الرياضية في الجوانب التطبيقية التي تحتاج الى جهد نظري لغرض اشتقاقها كما وتسهم في ايجاد الحلول للمعادلات التفاضلية المعقدة وبعض التكاملات المعقدة، كما وان استعمال اسلوب المحاكاة يؤدي الى تطوير أنموذج النظام وذلك من خلال ملاحظة التغيرات التي تطرأ على صياغة المشكلة في حالة تنفيذها عمليا ، كما وتمتاز عملية المحاكاة بالمرونة إذ تعطي القدرة على التجريب والأختبار من خلال تكرار العملية لمرات كثيرة بتفسير المدخلات الخاصة بعملية التقدير في كل مره[17] .

وتوجد طرائق مختلفة للمحاكاة ومن هذه الطرائق هي الطريقة التناظرية(Analog Method) والطريقة المختلطة (Mixed Method ، وطريقة مونت كارلو(Monte Carlo Method) وتعد طريقة مونت كارلو من أهم هذه

الطرائق وأكثرها شيوعا، حيث تستخدم في حل المسائل التي تتخللها عمليات عشوائية حيث يصعب عمل تجارب طبيعية وكذلك يصعب وضع صيغة رياضية معينة لها ، لذا يستعمل اسلوب المحاكاة بواسطة العينة و يتم اخذ عينة عشوائية من المجتمع لتمثيل الظاهرة بدلاً من المجتمع الحقيقي **[5]** .

7-2خطوات اجراء المحاكاة:

1. توليد المتغيرات التوضيحية من خلال استعمال اسلوب مونت–كارلوMont-Carlo() في المحاكاة اذ تم الاستعانة ببرنامج)MATLAB)حيث يعتبر من البرامج ذا القدرة الكبيرة في المجال البرمجي والرياضي حيث تم توليد (20) من المتغيرات التوضيحية وفق التوزيع الطبيعي وحدوث مشكلة التعدد الخطي حسب الصيغة الاتية[9] :

$$X_{ij} = (1 - r^2)^{\frac{1}{2}} Z_{ij} + rZ_{ip+1} \quad , \quad i = 1, \dots, n \quad , \quad j = 1, \dots, p \qquad .....(26)$$

حيث أن:

Z_{ij}: الاعداد العشوائية المولدة والتي تتبع التوزيع الطبيعي القياسي.

P: يمثل قيمة الارتباط بين المتغيرات التوضيحية.

2. يتم الحصول على n للمتغير التابع y على النحو التالي[9]

$$Y_i = \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \square_{\square}\square_{\square\square} + \square_{\square} \quad , \quad \square = 1, \dots, n \qquad .....(27)$$

3. يتم توليد الأخطاء العشوائية على وفق التوزيع الطبيعيσ^2(, 0.)
4. من أهم العوامل الآخرة التي يتم اختيارها هي كالآتي:
   أ. اختيار حجم للعينة المفترضة وهي(50).

ب .اختيار القيم الافتراضية لمعاملات الارتباط وهي(,0.0916، 0.25، 0.081.94، 0.49 ،0.98).

5.تقدير معلمات أنموذج الانحدار الخطي المتعدد وفق طرائق التقدير التي تم عرضها في الجانب النظري والمقارنة بين هذه الطرائق باستخدام معيار المقارنة متوسط مربعات الخطأ(MSE).

وبتطبيق الخطوات أعلاه تم الحصول على النتائج في الجداول(1 و 2 و 3 و 4) وكما يلي:

جدول (1)

P=0.16			P=0.09		
MSE	K	Method	MSE	K	Method
0.38261	0.00176000	AM	0.359291	0.00210000	GM
0.382618	0.00106300	MED	0.359301	0.00140000	ME
0.382638	0.00134000	GM	0.359300	0.00140000	HKB
0.38265	0.000110000	HKB	0.359331	0.0003806	DK
0.382674	0.00068350	DK	0.359310	0.00038660	HK
0.382684	0.00049513	HK	0.359336	0.00038064	KS
0.382681	0.00049513	KS	0.359352	0.00010313	HSL
0.382702	0.00001989	LW	0.359350	0.00001858	LW
0.387702	0.00860000	HSL	0.359784	0.01055000	AM
0.49 Based on the 16 compone PC			0.481254 Based on the 18 compone PC		

جدول (2)

P=0.49			P=0.25		
K	MSE	Method	K	M	Method
0.451950	0.0002450	HK	0.441420	0.0084500	GM
0.451951	0.0004607	GM	0.441450	0.0008490	MED
0.4519530	0.0004358	HKB	0.441457	0.0003658	DK
0.4519540	0.0002456	KS	0.441472	0.0005185	AM
0.4519510	0.0002456	DK	0.441500	0.0003700	HK
0.451956	0.0004798	MED	0.441510	0.0003700	KS
0.451959	0.0004887	AM	0.441550	0.0003700	HKB
0.451964	0.0001220	HSL	0.441668	0.00011550	HSL
0.451967	0.00006013	LW	0.445230	0.0000123	LW
0.48 Based on the 16 compone PC			0.471 Based on the 17 compone PC		

جدول (3)

P=0.94			P=0.81		
MSE		Method	MSE	K	Method
0.361297		GM	0.333500	0.0000041	LW
0.0001770		MED	0.333700	0.0001980	KS
0.361383		HKB	0.33371	0.0001971	HK
0.0001784		HK	0.33270	0.0001971	DK
0.361381		DK	0.33361	0.0001525	HSL
0.0001766		KS	0.333714	0.0002302	HKB
0.361380		AM	0.405	0.01070	GM
0.00017709		LW	0.46363	0.0109603	MED
0.361363		HSL	0.44653	0.0108074	AM
0.00017709					
0.361447	0.0000016				
0.361441	0.0001775				
.0361459	0.000003616				
0.3661948	0.00				
0.4821 Based on the 7 compone PC			0.472 Based on the 14 compone PC		

ISNN: [1858-6139], Online

جدول (4)

P=0.98		
MSE	K	Method
0.41006	0.0001548	KS
0.410096	0.0001632	MED
0.4101126	0.0001648	GM
0.410132	0.4101779	AM
0.00015483	0.0001633	HKB HSL
0.4104009	0.00016108	DK
0.412405	0.00016763	HK
0.412415	0.00016763	LW
0.41699	0.000047	
0.4876	Based on the 14 compone	PC

النتائج الواردة بالجداول(1) و(2) و(3) و(4) توضح قيم MSE للمقارنة بين طريقة انحدار المركبات الرئيسية (PCR) وطريقة انحدار الحرف(RR) وفقاً لطرق تقدير معلمة الحرف K ، عند مستويات مختلفة للارتباط بين المتغيرات التوضيحية(0.09 ، 0.16 ، 0.25 ، 0.94 ، 0.81 ، 0.49 ، 0.98) وحجم عينة n=50 ، حيث نجد أن قيم MSE عند مستويات الارتباط المختلفة(.0.09 ، 0.16 ، 0.25 ، 0.94،0.81 ، 0.49 ، 0.98)تبين أن جميع طرق

انحدار الحرف قدمت أداءًا أفضل من أداء طريقة انحدار المركبات الرئيسية.

وعند مقارنة طرق RR عند حجم العينة(n=50) نجد أن طريقة K_{GM} هي الأفضل عند مستوى الارتباط(، 94.090.0 ، 25.0)وعند مستوى الارتباط0(.16)كانت الطريقة K_{AM} هي الأفضل وعند مستوى الارتباط(49.0)كانت

أفضل الطرق أداءًا هي K_{HKB} بينما عند مستوى الارتباط(81.0)(كانت طريقة K_{DK} هي الافضل وعندما يكون مستوى الارتباط(98.0)تعتبر طريقة K_{KS} هي الأفضل.

النتائج

1/ قدمت طريقة انحدار الحرف(RR) أداء أفضل من أداء طريقة انحدار المركبات الرئيسية (PCR) عند وقوع الأنموذج تحت تأثير مشكلة التعدد الخطي.

2/ أظهرت طريقة تقدير معلمة الحرف K_{GM} أفضل أداء عند مستوى الارتباط94 (.0 ، 0.25 ، 0.09) مقارنة مع طرائق تقدير معلمة الحرف الاخرى بينما عند مستوى الارتباط(0.16) أظهرت طريقة تقدير معلمة الحرف K_{AM} أفضل أداء مقارنة مع طرائق تقدير معلمة الحرف الاخرى.

3/ أظهرت طريقة تقدير معلمة الحرف K_{HKB} أفضل أداء عند مستوى الارتباط (0.49)وأظهرت طريقة تقدير معلمة الحرف K_{DK} أفضل أداء عند مستوى الارتباط0.81 ( ) بينما عند مستوى الارتباط()0.98 أظهرت طريقة تقدير معلمة الحرف K_{KS} أفضل أداء مقارنة مع طرائق تقدير معلمة الحرف الاخرى.

التوصيات

1/ ضرورة أستخدام طريقة انحدار الحرف)RR(في حال وقوع الأنموذج تحت تأثير مشكلة التعدد الخطي لانها أعطت أفضل)MSE(.

2/ ضرورة إجراء بحوث اخرى لدراسة PCR و RR، والعلاقة بين مستويات أخرى لشدة التعدد الخطي وقيمMSE.

المراجع

البياتي، محمود مهدي حسن(. 2012) ، " تطبيق عملي لتحليل البيانات الاحصائية بأستعمال البرنامج (SPSS) " ، الجزيرة للطبع والنشر / جامعة بغداد.

جبريل ، محمد سليمان محمد(، 2014م) . " التعدد الخطي أسبابه تأثيراته والمعالجة بأنحدار الحافة وانحدار المركبات الرئيسية مع التطبيق على بيانات افتراضية " ، أطروحة دكتوراه كلية الدراسات العليا ، جامعة السودان للعلوم والتكنولوجيا.

خطار ، جبران عبد الأمير ومحمد ، كوركيس شهيد واسماعيل ، محمد سالم(. 2016) ، " تقدير معلمات الانحدار الخطي المتعدد بأستعمال الطرائق الحصينه(دراسة مقارنة) " ، مجلة القادسية لعلوم الحاسوب والرياضيات ، المجلد 8 ، العدد 1.

شاكر ، هديل حميد(. 2019) ، " استعمال المركبات الرئيسية وطرائق اخرى لتقدير الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي مع تطبيق عملي " ، رسالة ماجستير في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

كاظم ، اموري هادي ومسلم ، باسم شليبه(. 2002م) ، " القياس الاقتصادي المتقدم النظرية والتطبيق ، " مطبعة الطيف ، بغداد.

يحيى ، مزاحم محمد وعبدالله ، محمود حمدون(. 2007)، " تشخيص التعدد الخطي واستخدام انحدار الحرف في أختيار متغيرات دالة الاستثمار الزراعي في العراق للفترة1980–2000 " ، مجلة تكريت للعلوم الادارية والاقتصادية ، المجلد 3 ، العدد 8.

Adnan, N., M.H. Ahmad, and R. Adnan, (2006). " A comparative study on some methods for handling multicollinearity problems " , Matematika, 22(2):PP.109-119. Aguilera, A.M., and Escabias, M and M.J. Valderrama,(2006). " Using Principal Components for estimating Logistic regression with high dimensional Multicollinearity data " , Computational statistics data 50 , 1905-.4291

Al-Hassan, Y.M., Al-Kassab, M. M.,A., (2009). " Monte Carlo Comparison between Ridge and Principal Components Regression Methods " , Applied Mathematical Sciences , 3(42): PP.2085-2098.

Al-Hassan, Y.M., (2008). " A Monte Carlo Evaluation of Some Ridge Estimators " , Jordan Jornal for Applied Sciences, 10(2):PP.101-.011

Dorugade, A.V. and D.N. Kashid, (2010). " Alternative Method for Choosing Ridge Parameter for Regression " , (2016). Applied Mathematical Sciences, 4(9): PP.447-456.

Eledum, H.y.,(2016). "Acomparison Study of Ridge Regression and Principal Components Regression with Application " , University of Tabuk 3(8):283.

Hussin, M. M. (1989). " Some studies of a graphical method in statistical data analysis; subjective judgments in the interpretation of Boxplots-Unpuplished Ph.D. Thesis " , Keele university , U.K.

Hussin, M. M. (2008). " An Experiments on the Boxplot " , Statistics Dept. College of Admin. & Econ. Baghdad University.

Kibria, B.G., (2003). " Performance of some new ridge regression estimators " , Communications in Statistics-Simulation and Computation, 32(2):PP.419-435.

Ozdemir, Y.A., Esin,A.A., (2007). " Parameter Estimation in Multiple Linear Regression Models Using Ranked Setsampling " , Commun.Fac.Sci.Univ.Ank.Series. Thompson, J.,R., (2000) . " Simulation a Modeler's Approach " , John Wiley and Sons , Ins.