The Administration & Economic College Journal
For Economics & Administration & Financial Studies
Vol. 12,N.2, P P. 1-28
ISSN PRINT 2312-7813
ISSN ONLINE 2313-1012
مجلة كلية الإدارة والاقتصاد للدراسات الاقتصادية
والإدارية والمالية
المجلد ١٢، العدد٤، ٢٠٢٠ ص ص ١-٢٨

Parameters Estimation of the Multiple Linear Regression Model Under Multicollinearity problem

*تقدير معلمات انموذج الانحدار الخطي المتعدد في ظل وجود مشكلة التعدد الخطي

** أ.م.د. شروق عبد الرضا سعيد السباح *** ساره ماجد حسين الكفيشي

Abstract:In order to get at a model that leads to accurate predictions, it is necessary to search for the method of selecting the most important variables to be included in the model, especially when the data under study suffer from the problem of Co-Linearity , and the presence of many phenomena in our daily life, The main objective in this paper to compare among the methods of estimate unbiase parameters of the regression model ,OLS, Unbiase Ridge Regression and Lasso Regression and we consider MSE and P-Value as measures of comparison, a random sample of 100 women's fertility was withdrawn to study the factors affecting the number of children born as dependent variable and many independent random variables is (woman age , the age at married , graduate of woman , graduate of husband , women's weight, women's use of contraception, women's smoking, Age of Husband, job of husband, Marriage Period, Number of Dead Children, Hours of Exercise per Week, hours of female sleep per day, women taking drugs, duration of breast feeding, mother's job), data were suffering from a linear multiplicity problem, It was found that the Lasso regression method has the best estimation methods for owning the MSE, followed by the method of the unbiased ridge

المستخلص : للتوصل الى إنموذج له تنبؤات دقيقة ، لابد من البحث عن الطريقة التي تقوم بمهمة اختيار المتغيرات الاكثر اهمية لتضمينها في الانموذج وخاصة عندما تكون البيانات تحت الدراسة تعاني من مشكلة التعدد الخطي ، ولوجود كثير من الظواهر في حياتنا اليومية وخاصة الاجتماعية تعاني من هذه المشكلة ، يهدف البحث للمقارنة بين طرائق تقدير معلمات انموذج الانحدار الخطي وهي (طريقة المربعات الصغرى،انحدار الحرف البيزي الغير متحيز، وانموذج انحدار لاسو)

*بحث مستل

** جامعة كربلاء – كلية الادارة والاقتصاد

***طالبة ماجستير

وقد تم اعتماد متوسط مربعات الخطأ MSE والقيمة الاحتمالية P-Value كمعايير للمقارنة، تم سحب عينة عشوائية مؤلفة من (١٠٠) مفردة تخص خصوبة المرأة لدراسة العوامل المؤثرة على عدد الاطفال المولودين (متغير الاستجابة) ، وعدة متغيرات مستقلة هي (عمر المراة ، العمر عند الزواج ، التحصيل الدراسي للمرأة ، التحصيل الدراسي للزوج ، وزن المرأة ، استخدام المراة لوسائل منع الحمل ، تدخين المرأة ، عمر الزوج ، مهنة الزوج، فترة الزواج ، عدد الاطفال المتوفين ، عدد ساعات ممارسة الرياضة في الاسبوع ، اصابة المرأة بأمراض الغدة الدرقية ، عدد ساعات نوم المرأة في اليوم ، تناول الادوية من قبل المرأة ، مدة الرضاعة الطبيعية ، مهنة الام) ، وكانت البيانات تعاني من مشكلة تعدد خطي ، وتم التوصل الى ان طريقة انحدار لاسو افضل من طريقة المربعات الصغرى وطريقة انحدار الحرف البيزية الغير متحيزة لامتلاكها اقل متوسط مربعات خطأ.

المقدمة : Introduction

يعرف الانحدار **Regression** بشكل عام بأنه تحليل للعلاقة بين متغيرين او اكثر ، وأن هذه العلاقة يمكن التعبير عنها بشكل معادلة تحتوي على متغير واحد يعرف بالمتغير التابع (متغير الاستجابة) مع واحد او أكثر من المتغيرات التوضيحية (التفسيرية).او بمعنى أخر هو مجموعة الطرائق الاحصائية التي تتعامل مع الصيغ المختلفة للنماذج الرياضية لكونه اداة احصائية يقوم ببناء أنموذج أحصائي لتقدير العلاقة بين المتغير التابع(**Dependent Variable**) (متغير الاستجابة) وعدة متغيرات توضيحية (**Explanatory Variables**) (تفسيرية) وهي المتغيرات المستقلة، ينتج عنها معادلة أحصائية توضح العلاقة بين المتغيرات ويمكن استعمال هذه المعادلة لغرض التقدير والتنبؤ بقيمة المتغير التابع(متغير الاستجابة) بأعتماد معلومات المتغيرات التوضيحية (التفسيرية) المستقلة. حين تكون العلاقة خطية في الانموذج بين متغير تابع واحد ومتغير مستقل واحد فأن هذا الانموذج هو يعد من أبسط نماذج الانحدار ويسمى أنموذج الانحدار الخطي البسيط(**Simple Linear Regression Model**) ، اما عندما تكون العلاقة خطية في الانموذج بين متغير تابع مع اثنين او اكثر من المتغيرات التوضيحية (التفسيرية) يسمى أنموذج الانحدار الخطي المتعدد(**Multiple Linear Regression Model**). وبينما يكون عدد المتغيرات المستقلة كبير جدا بحيث يمكن ان يحصل تداخل بين المتغيرات فتنشأ مشكلة التعدد الخطي بين المتغيرات التي تعجز الطرائق التقليدية في حلها وبالتالي عدم الدقة في التنبؤ، ولذلك وجدت طرائق كبديل مجدي لحل هذه المشكلة فهي تعطي سهولة ومرونة في الحسابات منها انحدار الحرف وانحدار لاسو . لغرض التوصل الى انموذج احصائي يمثل البيانات المدروسة افضل تمثيل ، فلابد من البحث عن الطرائق التي تقوم باختيار المتغيرات الداخلة في الدراسة ، لذا فأن الهدف من هذا البحث هو مقارنة تقديرات معلمات انموذجي انحدار الحرف وانحدار لاسو في ظل وجود مشكلة التعدد الخطي باستعمال بعض المعايير الاحصائية المعروفة للتوصل الى افضل الطرائق في التقدير.

هناك العديد من الدراسات والبحوث التي تناولت موضوع تقديرات معلمات انموذج الانحدار ، وخاصة طرائق انحدار الحرف، ولكن ندرة الدراسات وخصوصاً العربية التي تناولت موضوع انحدار لاسو، ففي **عام ١٩٧٢ عمما الباحثان (Lindley & Smith)** [19] النموذج الاحصائي الخطي باستعمال الطرائق البيزية وقاما بتوضيح هذه الطرائق على نموذج انحدار خطي متعدد وبينا بان الطرائق البيزية تعطي تقديرات غير متحيزة لمعاملات نموذج الانحدار الخطي المتعدد .**وفي عام ١٩٩٣ قارنا (SALEH & KIBRIA)** [24] بين مقدرات انحدار الحرف الاعتيادية ومقدرات انحدار الحرف المقيدة ومقدرات انحدار الحرف الغير مقيدة في حالة وجود مشكلة التعدد الخطي بين المتغيرات المستقلة . **وكذلك في عام ٢٠٠٥ قدر (Al-Sadoun)** [7] معاملات انموذج الانحدار الخطي المتعدد باستخدام اسلوب انحدار الحرف الاعتيادي واسلوب انحدار الحرف البيزي وقارن بين المقدرين باستخدام المحاكاة وبين بان اسلوب انحدار الحرف البيزي افضل من الحرف الاعتيادي . و**في عام ٢٠١٠ قدم (Al-Hassan)** [5] مقدر انحدار الحرف كبديل عن مقدر المربعات الصغرى في حالة وجود مشكلة التعدد الخطي ، اذ قدم طريقة جديدة لتقدير معلمة الحرف عن طريق دراسة محاكاة لاختبار اداء الطريقة المقترحة باستعمال مقايس متوسط مربعات الخطأ **MSE** وبينت نتائج المحاكاة ان المقدر بالطريقة المقترحة افضل من المقدرات الاخرى . **في عام ٢٠١٨ قام (Ali Sadig)** [6] بدراسة العوامل المؤثرة على كلفة الانتاج لمعمل سمنت المثنى باستعمال طريقة انحدارالحرف ، اذ اعتبر كلفة كمية الانتاج متغير معتمد والعوامل المؤثرة (السعر بالطن ، الطاقة الكهربائية ، الكمية المستهلكة) كمتغيرات مستقلة تعاني من مشكلة التعدد الخطي ، واستنتج بان طريقة انحدار الحرف هي الافضل في حالة البيانات المالية والاقتصادية ، وطبق الطريقة باستخدام لغة **R.**

اما التي تناولت انحدا لاسو فهي: **في عام ١٩٩٦ اقترح (Tibshirani)** [25] طريقة جديدة لتقدير النماذج الخطية اسماها **LASSO** وهي مختصرات **Least Absolute Shrinkage and Selective Operator** والتي تقلل من مجموع مربعات البواقي الخاضعة لمجموع القيمة المطلقة للمعاملات والتي تكون اقل من قيمة ثابتة ،

وبسبب طبيعة هذا القيد فانه يميل الى انتاج بعض المعاملات التي تساوي بالضبط صفر وبالتالي يعطي نماذج قابلة للتفسير، وباستخدام تجربة محاكاة بين بان **LASSO** تتمتع بخصائص افضل من طريقة الخطوات المتسلسلة وطريقة انحدار الحرف .**ايضاً في عام ١٩٩٧ اقترح (Tibshirani)** [26] طريقة جديدة لاختيار المتغيرات والتقلص في نموذج المخاطرة النسبية كوكس لتقليل الاحتمالية الجزئية اللوغاريمتية نسبة الى مجموع القيم المطلقة للمعاملات والتي تحدد بثابت انحدار لاسو ، وباستعمال المحاكاة بين بان طريقة **LASSO** اكثر دقة من طريقة الاختيار المتسلسل في التقدير. **في عام ١٩٩٨ قارن (J.Fu)** [15] بين انحدار **Bridge** وانحدار **LASSO** باستعمال تجربة محاكاة ، وبين بان انحدار لاسو افضل من انحدار **Bridge** عند قيمة مختلفة من معلمة الجزاء. **في عام ٢٠٠٩ قارن (HANS)** [14] طريقة لاسو القياسية بطريقة لاسو البيزية في ظل توزيعات اولية

مسبقة عن معاملات الانحدار مستخدماً النظرية البيزية، واوجد ان طريقة لاسو القياسية ليست بالضرورة ان تتفق مع التنبؤات النظرية البيزية . **في عام ٢٠١٧ شرح (Fonti)** [13] طريقة لاسو **LASSO** في خاصية الاختيار وهي الميزة التي نختار بها عدداً اقل من المتغيرات التوضيحية (المستقلة) لوصف متغير الاستجابة لجعل النموذج اسهل في التفسير ، وطبق طريقة **LASSO** على النماذج الخطية والنماذج الخطية المعممة عندما يكون عدد المتغيرات اكبر من عدد المشاهدات.

مشكلة البحث: Problem of the Research

في الكثير من التطبيقات العملية ، يكون عدد المتغيرات المستقلة المؤثرة على المتغير المعتمد كبيرة وبعضها تكون مرتبطة مع بعضها البعض فتنشأ مشكلة التعدد الخطي بين المتغيرات وتكون التنبؤات خاطئة ، وطالما ان الطرائق الاحصائية التقليدية تفترض عدم وجود ارتباط بين المتغيرات المستقلة ، لذلك لابد من البحث عن طرائق تقوم بتقليص عدد المتغيرات المستقلة وبالتالي تقليص عدد معاملات الانحدار وبالتالي الخروج واختيار النموذج الافضل الذي تكون لديه قوة تفسيرة وتنبؤية عالية جداً.

هدف البحث: Aim of the Research

لغرض التوصل الى انموذج احصائي يمثل البيانات المدروسة افضل تمثيل ، فلابد من البحث عن الطرائق التي تقوم باختيار المتغيرات الداخلة في الدراسة ، لذا فأن الهدف من هذا البحث هو مقارنة تقديرات معلمات انموذجي انحدار الحرف وانحدار لاسو في ظل وجود مشكلة التعدد الخطي باستعمال بعض المعايير الاحصائية المعروفة للتوصل الى افضل الطرائق في التقدير.

الإنموذج الخطي : Liner Model

يدعى غالباً بإنموذج الانحدار الخطي ، وهو إنموذج يصف العلاقة بين متغير الاستجابة او المتغير المعتمد Y_i والمتغيرات المستقلة X_{ij} ، وفي حالة المتغير المستقل الواحد يدعى إنموذج الانحدار الخطي البسيط Simple Linear Regression model ، بينما في حالة متغيرين مستقلين فاكثر فيدعى بإنموذج الانحدار الخطي المتعدد Multiple Linear Regression .

وان افتراضات إنموذج الانحدار الخطي هي كالاتي :

١. متغير الاستجابة يتوزع توزيعاً طبيعياً بمتوسط μ_i وتباين σ^2 اي ان :

$$Y_i \sim N(\mu_i, \sigma^2) \quad ; i=1,2,\ldots,n$$

٢. خطية الإنموذج ، اي العلاقة بين المتغير المعتمد والمتغير او المتغيرات المستقلة علاقة خطية .

ويمكن ان نعبر عن الإنموذج الخطي بالشكل الاتي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots . + \beta_p X_{ip} + u_i \quad ; i=1,2,\ldots,n$$

اذ ان : $\beta_0, \beta_1, \beta_2 \ldots \ldots \beta_p$: معاملات الانحدار

p : عدد المتغيرات المستقلة

u_i : الخطأ العشوائي يتبع التوزيع الطبيعي بمتوسط صفر وتباين ثابت ، والاخطاء مستقلة عن بعضها ، اي نعبر عنها رياضياً :

$$u \sim N(0, \sigma^2_I)$$
$$cov(u_i, u_j) = 0$$

ويمكن ان يكتب الانموذج الخطي بصيغة المصفوفات كالاتي :

$$\underline{Y} = X\underline{\beta} + \underline{U}$$

...(١)

اذ ان :

Y : متجه nx1 يفترض ان يتوزع توزيع طبيعي بمتوسط صفر وتباين ثابت على اختلاف قيم x

X : مصفوفة المتغيرات المستقلة nxp

β : متجه المعاملات px1

U : متجه الاخطاء nx1

هدف الانحدار الخطي هو ملائمة الخط المستقيم لعدد من النقاط والتي تقلل مجموع مربعات البواقي (الأخطاء) ، اي احتواء خط مستقيم الى عدد من النقاط والتي تقلل مجموع مربعات البواقي. وتستعمل نماذج الانحدار لعدة اغراض ، منها وصف وتحليل العلاقة بين المتغيرات ، التنبؤ واختيار المتغيرات.[28] [32] [27][13]

وتقدر معلمات الأنموذج (1) بإتباع طريقة المربعات الصغرى OLS حسب الصيغة الاتية:

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

...(2)

وتكتب معادلة الانحدار التقديرية التي تعطي n من قيم متوسط الاستجابة $(\hat{y})$ كالاتي :

$$\hat{y} = X\hat{\beta}$$

...(٣)

مشكلة التعدد الخطي Multi-CoLinearity

مشكلة الارتباط الخطي المتعدد Multi-CoLinearity تحدث في نماذج الانحدار المتعدد ولاتحدث في إنموذج الانحدار البسيط ، فهي مشكلة خاصة بالمتغيرات المستقلة.

نُواجَه مشكلة التعدد الخطي عندما يرتبط بعض او كل المتغيرات المستقلة بعلاقة خطية مع بعضها البعض يصبح من الصعب فصل اثر كل متغير على المتغير المعتمد ، او عندما تكون قيمة احد المتغيرات المستقلة متساوية لكافة المشاهدات او عندما تعتمد قيمة احد المتغيرات المستقلة على قيمة واحد او اكثر من المتغيرات المستقة الاخرى في الإنموذج . او عند استخدام المتغيرات المرتدة زمنيا (متغيرات التباطؤ الزمني) كمتغيرات تفسيرية في الإنموذج. [1] [2] [7]

يمكن أن تؤدي مشكلة التعدد الخطي إلى تقديرات متحيزة وغير دقيقة لمعاملات الانحدار ، وتضخّم الأخطاء المعيارية لمعاملات الانحدار وتقلل من اختبارات t الجزئية لمعاملات الانحدار ، فضلاً عن التقليل من قابلية تنبؤ الأنموذج .

وقد يكون التعدد الخطي تاماً ، وبذلك تكون مصفوفة المعلومات $X'X$ غير تامة الرتبة . وقد يكون التعدد الخطي غير تام ، عندما تكون بعض المتغيرات المستقلة دالة في التركيبة نفسها لمتغيرات اخرى مع قيم عشوائية ، وهنا يكون محدد مصفوفة المعلومات صغيراً مما يؤدي الى تضخيم تباين المعلمات المقدرة ، كما يعجز الإنموذج عن اظهار أثر المتغيرات المستقلة منفصلة في المتغير التابع نظراً لترابطها. [3] [4]

بعض طرائق الكشف عن التعدد الخطي :

هنالك طرائق احصائية مختلفة تستخدم للكشف عن مشكلة التعدد الخطي بين المتغيرات ومن أكثر الطرائق استخداماً هي :

١- اختبار Farrar and Glauber (1967) [8] : يعتمد الاختبار على قيمة محسوبة لمربع كاي Chi square بالاعتماد على محدد مصفوفة معامل الارتباط $|R|$ وعدد المتغيرات المستقلة (p) وعدد الوحدات التجريبية (n) وبدرجات حرية $n(n-1)/2$ وهي :

$$\chi_0^2 = -\left[\left[n-1-\frac{1}{6}(2p+5)\right] ln \mid R \mid\right] \quad \ldots(4)$$

٢- عامل تضخم التباين Variance inflation factors (VIF) :

ويستعمل لقياس مدى ارتباط كل متغير مستقل مع المتغيرات الاخرى ، تم وضع هذا المقياس من قبل Marquardt (1970) ، فاذا كان :

$$\mathbf{VIF > 5}$$

فانه مؤشر على وجود مشكلة التعدد الخطي .

٣- مقياس المعيار الشرطي Conditional Index (CI) : وهو النسبة بين أكبر وأصغر جذر مميز Eigen Root الناتجة من تحليل مصفوفة المعلومات ، وقد أشار Montgomery and Peck (1982) [18] الى الشروط الواجب توفرها للحكم على وجود مشكلة التعدد الخطي وهو :

أ- اذا كان العدد الشرطي بحدود 10-5 وان اثنين او اكثر من نسب تحليل التباين اقل من ٠.٥ فانه دليل على وجود ارتباط ضعيف.

ب-اذا كان العدد الشرطي بحدود $10 \leq CI \leq ٣٠$ وان اثنين او اكثر من نسب تحليل التباين اكبر من ٠.٥ فانه دليل على وجود ارتباط معتدل.

ت-اذا كان العدد الشرطي بحدود $CI \geq ٣٠$ وان اثنين او اكثر من نسب تحليل التباين اكبر من ٠.٥ فانه دليل على وجود ارتباط خطي على درجة كبيرة.[21][4]

معالجة مشكلة التعدد الخطي

من الطرائق الشائعة للتغلب على مشكلة التعدد الخطي هي :

١- التحويل المعياري لقيم المتغيرات: ان مثل هذا النمط من التحويل يضعف التداخل الخطي ويخفض قيمة معامل الارتباط بين ازواج المتغيرات .

٢- إضافة بيانات جديدة الى البيانات الاصلية : وهذا مشابه لأسلوب زيادة حجم العينة اذ ترتفع القيم الذاتية لمحدد مصفوفة المعلومات ($|X'X|$) ، ولهذه الطريقة بعض النواحي السلبية منها تعود الى اسباب اقتصادية او تغيرات في المجتمع المدروس ، كما قد يؤدي تضخم البيانات الى تطرفها.

٣- حذف المتغيرات المسببة للتعدد الخطي او استبدالها : يعاب على هذه الطريقة بحصول مشكلة في توصيف البيانات وعدم استقرارية المعلمات المقدرة وظهور اخطاء معيارية فيها.

٤- اتباع احدى طرائق الاحصاء البديلة مثل : المربعات الصغرى المقيدة او طريقة الدمج في السلاسل الزمنية او طريقة المركبات الرئيسة او طريقة انحدار الحرف ، انحدار لاسو. [10] وسيتناول البحث مقارنة طريقة انحدار الحرف مع انحدار لاسو كونه من الطرائق الحديثة.

انحدار لاسو: Lasso Regression

قدمت هذه الطريقة لأول مرة في الادب الجيوفيزياوي في عام (١٩٨٢) [4] ، ثم اعيد اكتشافها بشكل مستقل في عام (١٩٩٦) من قبل الباحث روبرت تيبشراني Ropert Tibshirani حيث قام بصياغة هذه الطريقة وقدم الكثير من الافكارحول ادائها [32][27]

لاسو اقل تقلص مطلق واختيار العامل، هي مختصرات للكلمات الانكليزية (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator) ، وهي دالة جزاء لإنموذج الانحدار الخطي وهي طريقة لتقدير معلمات إنموذج الانحدار وكذلك لاختيار وتنظيم المتغيرات الداخلة في الإنموذج لزيادة الدقة التفسيرية لنماذج الانحدار المستخدمة في تحليل الظاهرة محل الدراسة من خلال عمليات ملائمة الإنموذج لاختيار مجموعة فرعية من المتغيرات المشتركة في الإنموذج النهائي بدلاً من استخدامها كلها ، ففي طريقة لاسو يتم تصغير مجموع مربعات الاخطاء العشوائية لاعلى حد من مجموع القيم المطلقة لمعاملات إنموذج الانحدار [13][12][27] [29].

صممت LASSO اصلاً لنماذج المربعات الصغرى Least squares models ، حيث تكشف LASSO عن كمية كبيرة من سلوك المقدر عن طريق معامل لاسو او ما يسمى بالعتبة الناعمة Soft Thresholding بما في ذلك علاقة مقدر لاسو مع مقدر بانحدار الحرف Ridge Regression ومقدر افضل اختيار مجموعة جزئية من المتغيرات Best subset selection والتي تماثل طريقة الخطوات المتسلسلة Stepwise selection ، ويكشف ايضاً (كما في الانحدار الخطي) انه لايجب ان تكون تقديرات معامل لاسو فريدة (وحيدة) اذا كانت المتغيرات المستقلة تعاني

من مشكلة التعدد الخطي . وان طريقة لاسو لها القدرة على اختيار مجموعة جزئية تعتمد على صيغة القيد ، وعلى الرغم من انه تم تعريف لاسو للمربعات الصغرى الا انه يمكن بسهولة استعمال طريقة لاسو في مجموعة واسعة في كثير من النماذج الاحصائية منها النماذج الخطية المعممة ومعاملات التقدير المعممة ونماذج المخاطر النسبية ومقدرات M، ويمكن ان تستعمل لاسو في كثير من المجالات مثل الهندسة والاحصاءات البيزية والتحليل المحدب[18] .

قبل طريقة انحدار لاسو كانت الطريقة الاكثر استخداماً لاختيار المتغيرات المستقلة التي يتم تضمينها ضمن الإنموذج هي طريقة الاختيار المتسلسل Stepwise Selection والتي تعمل على تحسين دقة الإنموذج في حالات معينة خاصة عندما يكون لبعض المتغيرات المستقلة علاقة قوية بمتغير الاستجابة، والذي يجعل التنبؤ غير دقيق ، فضلاً طريقة انحدار الحرف Ridge Regression الاكثر شعبية التي تستعمل لتحسين دقة التنبؤ للإنموذج الانحدار . فهو يعمل على تحسين خطأ التنبؤ عن طريق تقليص معاملات الانحدار الكبيرة من اجل تقليل التكرار ولكنه لايقوم بالاختيار المشترك وبالتالي لايساعد على جعل الإنموذج اكثر قابلية على التفسير. بينما لاسو يستطيع تحقيق كلا هذين الهدفين عن طريق جعل مجموعة القيم المطلقة لمعاملات الانحدار لها مقادير اقل من قيمة ثابتة ، مما يُجبِر بعض المعاملات لتكون مساوية للصفر، مع اختيار إنموذج ابسط لايتضمن تلك المعاملات [11].

مبدأ انحدار لاسو: LASSO Regression Principle

ان مبدأ طريقة انحدار لاسو هو تصغير مجموع مربعات البواقي وفقاً الى قيد يمثل المجموع المطلق للمعاملات والتي تكون اصغر من ثابت معين ، اذ يضع لاسو قيداً على مجموع القيم المطلقة لمعاملات الإنموذج Coefficients ، بحيث يجب ان يكون المجموع اقل من قيمة ثابتة (الحد الاعلى) من اجل القيام بذلك تطبق لاسو عملية التقلص (تنظيم) اذ انها تعاقب معاملات الانحدار وتقلص بعضها الى الصفر ، واثناء عملية اختيار المتغيرات سيتم تحديد المتغيرات التي تحتوي على رمز غير صفري بعد عملية التقلص وستكون جزءاً من الإنموذج والهدف من هذه العملية هو تقليل خطأ التنبؤ .

ان في طريقة لاسو توجد معلمة ضبط (تسوية) هي التي تتحكم في قوة معاقبة(جزاء) معاملات الانحدار وتحتل اهمية كبيرة في ذلك فعندما تكون معلمة الضبط كبيرة بشكل كافي تضطر المعاملات الى ان تكون مساوية للصفر ، وهذه مقيدة في تقليل المتغيرات في الإنموذج ، اي بمعنى كلما كانت قيمة معلمة الضبط كبيرة معناه عدد اكبر من المعاملات المساوية للصفر. واذا كانت معلمة الضبط مساوية للصفر ، سنحصل على انحدار المربعات الصغرى الاعتيادية OLS Regression. [30][9][27][5].

مميزات انحدار لاسو : LASSO Regression Advantages

هنالك العديد من المميزات في استخدام طريقة لاسو كالاتي:

١. يمكن ان يوفر لاسو دقة تنبؤية جيدة للغاية لان تقليص وازالة المتغيرات يمكن ان يخفض التباين دون زيادة كبيرة في التحيز ، وهذا مفيد بشكل خاص عندما يكون لدينا عدد قليل من المشاهدات وعدد كبير من المتغيرات.

٢. يساعد لاسو على زيادة امكانية تفسير الإنموذج من خلال القضاء على المتغيرات غير ذات صلة التي لاترتبط بمتغير الاستجابة.

فبذلك تعتبر طريقة لاسو طريقة لاختيار وتنظيم المتغيرات الداخلة في الإنموذج الانحداري.[29]

صيغة انحدار لاسو: LASSO Regression Formula

تم تقدير معلمات انحدار لاسو طبقاً لمبدأ المربعات الصغرى من الصيغة الاساسية كالاتي :

ليكن لدينا عينة مؤلفة من N حالات كل حالة مؤلفة من P من المتغيرات المستقلة ومتغير متعمد واحد y_i ، وليكن $\underline{x}$ يمثل متجه المتغيرات المستقلة للحالة j^{th} ، فيكون هدف انحدار لاسو هو حل المعادلة الاتية :

$$\min\{\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2\} \text{ Subject to } \sum_{j=1}^{p}|\beta_j| \leq t \quad [12]$$
$$\dots (5)$$

اذا ان:

t : تمثل معلمة حرة تحدد مسبقاً والتي تعيين مقدار التسوية(التقليص).

X: مصفوفة المتغيرات المستقلة.

وان $X_{ij} = (x_i)_j$

وان x_i^T هي الصف j^{th} من المصفوفة X

فيمكن كتابة صيغة لاسو بالشكل الاتي :

$$\min_{\beta_0,\beta}\{\frac{1}{N}||y - \beta_0 I_N - X\beta||_2^2\} \quad \text{Subject to} \quad ||\beta|| \leq t$$
$$\dots(6)$$

اذا ان:

$||\beta||_p = (\sum_{i=1}^{N}|\beta_i|^p)^{1/p}$ هو الطول القياسي ℓ^p وان I_N متجه الواحدات (Nx1) .

يرمز $\bar{x}$ للمتوسط القياسي لنقاط البيانات x_i و $\bar{y}$ متوسط المتغير المعتمد(متغير الاستجابة y_i) والتقدير $\beta_0 = \bar{y} - \bar{x}_i^T\beta$ اذا ان :

$$y_i - \beta_0 - x_i^T\beta = y_i - (\bar{y} - x_i^T\beta) - x_i^T\beta = (y_i - \bar{y}) - (x_i - \bar{x})^T\beta$$

وبالتالي فانه من الطبيعي العمل مع المتغيرات التي تم جعلها مركزية (جعل متوسطها يساوي صفر) اضافة الى المتغيرات المستقلة تكون معيارية مثالية Typically standardizes $\sum_{i=1}^{N} x_i{}^2 = 1$.

ويمكن اعادة كتابة الصيغة (٦) بالشكل الاتي :

$$\min_{\beta_0,\beta}\{\frac{1}{N}||y - X\beta||_2^2\} \quad \text{Subject to} \quad ||\beta||_1 \leq t \quad \ldots(٧)$$

ويكون بصيغة مضاعف لاكرانج كالاتي:

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_1\} \quad \ldots(٨)$$

λ هي المعلمة التي تتحكم في قوة الجزاء (العقوبة) على معلمات الانحدار . [30] [2] [22]

خصائص مقدر انحدار لاسو Lasso Estimator properties

هنالك بعض الخصائص لمقدر لاسو يمكن ان ندرجها كالاتي : [25] [32] [20]

1 . التعامد الطبيعي للمتغيرات المستقلة Orthonormal Covariates

لنفرض ان المتغيرات المشتركة متعامدة طبيعياً بحيث ان $(x_i|x_j) = \delta_{ij}$ حيث ان :

$(.|.)$ حاصل الضرب الداخلي Inner product و δ_{ij} كرونجر دلتا Kroncher delta تعبر عن :

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

وباستعمال طريقة Subgradient التكرارية والتي هي احدى طرائق التحليل التعددي لحل مسائل الاقل تحدب نحصل على[2] :

$$\hat{\beta}_j = S_{N\lambda}\left(\hat{\beta}_j{}^{OLS}\right) = \hat{\beta}_j{}^{OLS}\text{Max}(0,1 - \frac{N\lambda}{|\hat{\beta}_j{}^{OLS}|})$$

اذ ان :

$$\hat{\beta}_j{}^{OLS} = (X^T X)^{-1}X^T Y$$

وان $S_{N\lambda}$ يشير الى العتبة الناعمة Smooth Threshold

ونظراً لانه يحول القيم الى صفر (جعلها صفراً تماماً اذا كانت صغير بما فيه الكفاية) بدلاً من وضع قيم اصغر الى الصفر وترك القيم الاكبر دون المساس كعامل عتبة صلب يرمز له $H_{N\lambda}$ وهذا يمكن ان يكون مقارنة مع انحدار الحرف فيكون الهدف هو تصغير المقدار الاتي:

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_2^2\} \quad \ldots(٩)$$

ويؤدي الى :

$$\hat{\beta}_j = (1 + N\lambda)^{-1}\hat{\beta}_j^{OLS}$$

وطالما ان انحدار الحرف تقلص كل المعاملات بالعامل المتغير لـ $(1 + N\lambda)^{-1}$ ولا يضع اي من المعاملات صفراً.

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_0\}$$

...(١٠)

حيث $||\cdot||_0$ هوالطول ℓ^0 الذي يعرف بانه $||z|| = m$ اذا كانت m مركبات بالضبط لـ z تكون غير صفرية ويكن ان تكتب كالاتي :

$$\hat{\beta}_j = H_{\sqrt{N\lambda}}\left(\hat{\beta}_j^{OLS}\right) = \left(\left|\hat{\beta}_j^{OLS}\right| > \sqrt{N\lambda}\right)$$

حيث ان $H_{\sqrt{N\lambda}}$ حد العتبة الصلب وان I هي الدالة المؤشرة Indicator function .

ونلاحظ من ذلك بان تقديرات لاسو تجمع بين مميزات انحدار الحرف وانحدار افضل اختيار جزئي فهي تحول جميع المعاملات الى الصفر بقيمة ثابتة وتضبطها الى الصفر اذا وصلت اليها.[25][32]

2. المتغيرات المستقلة مترابطة Correlated Covariates

بالعودة الى الصيغة العامة للاسو التي قد لاتكون فيها المتغيرات المشتركة المختلفة مستقلة ، ففي الحالة التي يكون فيها اثنين من المتغيرات المستقلة i,j متطابقين لكل حالة بحيث أن: $x_i = x_j$ فأن قيم المعلمات B_i و B_j التي تصغر دالة هدف لاسو لاتكون محددة بصورة فريدة، بل هنالك $\hat{B}_i, \hat{B}_j \geq 0$ فانه اذا كان $s \in [0,1]$ فباستبدال $\hat{B}_i$ بـ $s(\hat{B}_i + \hat{B}_j)$ و بـ $(1 - s)(\hat{B}_i + \hat{B}_j)$ مع الابقاء على بقية $\hat{B}_i$ وسوف نحصل على حل جديد وتستمر دالة لاسو باجراء تقليص للمعاملات .[25][32]

انحدار الحرف البيزية (الغير متحيزة) **Unbiased Ridge Regression**

في طريقة انحدار الحرف المتحيزة يتم تقدير معلمات الانموذج المتعدد وفق الصيغة الاتية :

$$\hat{\beta}_R = \left(X'X + kI_p\right)^{-1}X'Y$$

حيث اثبتت هذه الطريقة فعاليتها في معالجة مشكلة التعدد الخطي الذي يؤدي الى كبير حجم تباين المقدرات ويُحدث ارباك في العلاقة السببية بين المتغيرات المستقلة والمتغير المعتمد في انموذج الانحدار. ولكن في طريقة انحدار الحرف الغير متحيزة سيتم ادخال المعلومات السابقة Prior في إنموذج انحدار الحرف للتخلص من مشكلة التحيز في تقدير المعلمات. [14] [2] [4]

اقترح Swindel (1976) [16] أدخال المعلومات السابقة Prior Information في صيغة انحدار الحرف ، بذالك سوف تكون صيغة انحدار الحرف البيزية كما يلي:

$$\hat{\beta}_R = (X'X + KI_P)^{-1}(X'Y + KJ)$$

...(11)

حيث نلاحظ من الصيغة (11) انه تم ادخال الثابت K مع المتجه J إلى المتجه X'Y والذي يتمثل بالمعلومات الاولية (Prior information).

ونلاحظ ايضاً إنه في حالة $K = 0$ سوف نحصل على تقدير المربعات الصغرى الاعتيادية OLS.

ومن خلال مفهوم نظرية بيز المتمثل بالآتي:

Posterior dist. $\propto$ Prior Prob.×Likelihood

نلاحظ من الصيغة أعلاه انه للحصول على التوزيع النهائي يجب توفير الاحتمال الأولي ودالة الترجيح أي توفير معلومات أولية حول المقدر من اجل الحصول على الاحتمال النهائي.

فاذا كان لدينا المتجه J والذي يمثل متجه المعلومات الأولية المعلوماتي فأنه قيمة K تعطي مقدرات تمتلك MSE اقل من MSE للمربعات الصغرى. وإذا كان J أي متجه عشوائي(غير معلوماتي) فان K تعطي ايضاً مقدرات تمتلك MSE اقل من MSE للمربعات الصغرى.

وكما معروف من خلال فرضيات الإنموذج الخطي، ، ان الخطأ u_i يتوزع توزيع طبيعي بمعدل ٠ ، وتباين σ^2_{In}, أي أن $u_i \sim N(0, \sigma^2_{In})$ ، وفان تقديرات المربعات الصغرى تتوزع توزيع طبيعي بمعدل B وتباين $\sigma^2(X'X)^{-1}$.

فلنفرض انه المتجه J المتضمن المعلومات السابقة له توزيع طبيعي بمعدل B، ومصفوفة التباين – التباين مشترك V أي $J \sim N(B, V)$.

وعلى فرض أن V هي مصفوفة تامة الرتبة (Full rank) وتمثل مصفوفة التباين – التباين المشترك، بذلك فان التقدير المحدب convex estimator :

$$B(C, J) = C\widehat{B}_{Ols} + (I - C)J \quad \dots (12)$$

اذ أن:

$B(C, J)$ هو التقدير المحدب وهو دالة ذات قيمة حقيقية محدبة ومن خواصها أنه اذا كان لاي نقطتين في مجال هذه الدالة مثل x و y ، فاذا كان الخط المستقيم الذي يصل بين اي نقطتين على رسمها البياني يقع فوق الرسم البياني للدالة ، فأذا كان:

$$B(C, J) = cf(x) + (1 - c)f(y)$$

فأن : $cf(x) = f(x)$

I هي مصفوفة أحادية ذات رتبة P×P

C هي مصفوفة ذات بعد P×P .

نلاحظ من الصيغة (12) أن المصفوفة C مصفوفة مجهولة ، فلايجاد هذه المصفوفة فأننا سوف نجد متوسط مجموع مربعات الخطأ للتقدير المحدب والذي يكون بالصيغة الاتية:[17] [21] [23] [31]

$$\widehat{B}_{Bays} = (X'X + \sigma^2 V_0^{-1})^{-1} (X'Y + \sigma^2 V_0^{-1} B_0) \quad \dots\dots\dots(١٣)$$

الجانب التطبيقي : Applied Side

وصف البيانات Data Description

تضمنت بيانات البحث المعلومات المتوفرة في سجلات زيارات النساء الى المراكز الصحية التابعة لدائرة صحة بابل ، حيث تم مراجعة المراكز الصحية الموجودة في مركز محافظة بابل وهي مركز صحي مرجان للرعاية الصحية الاولية ومركز صحي الخالصة للرعاية الصحية الاولية ومركز صحي شهداء نادر النموذجي للرعاية الصحية الاولية ومركز صحي الاساتذة النموذجي للرعاية الصحية الاولية ومركز صحي بابل التدريبي للرعاية الصحية الاولية ، وكذلك من الاستمارات الاحصائية الخاصة برعاية الام والطفل الموجودة في قسم الصحة العامة للرعاية الصحية الاولية حيث سحبت عينة عشوائية بسيطة مؤلفة من ١٠٠ امراة لدراسة العوامل المؤثرة على عدد الاطفال المولودين (متغير الاستجابة) Y ، بينما كانت المتغيرات المستقلة تتمل بالاتي :

X_1 : عمر المراة

X_2 : العمر عند الزواج

X_3 : التحصيل الدراسي للمرأة

X_4 : التحصيل الدراسي للزوج

X_5 : وزن المرأة

X_6 : استخدام المراة لوسائل منع الحمل

X_7 : تدخين المرأة

X_8 : عمر الزوج

X_9 : مهنة الزوج

X_{10} : فترة الزواج

X_{11} : عدد الاطفال المتوفين

X_{12} : عدد ساعات ممارسة الرياضة في الاسبوع

X_{13} : اصابة بامراض الغدة الدرقية

X_{14} : عدد ساعات نوم المراة في اليوم

X_{15} : تناول الادوية من قبل المرأة

X_{16} : مدة الرضاعة الطبيعية

X_{17} : مهنة الام

اختبار مشكلة التعدد الخطي :

تم اختبار البيانات من حيث كونها تعاني من مشكلة التعدد الخطي ام لا سيتم استعمال الاختبارات الاتية :

١- **اختبار فارار-كلوبر** Farrar and Glauber **لاختبار الفرضية الاتية :**

H_0: لاتوجد مشكلة التعدد الخطي

H_1: توجد مشكلة التعدد الخطي

ومن خلال تطبيق الصيغة رقم **(4)** وكالاتي :

مصفوفة الارتباط بين المتغيرات المستقلة هي :

R=

1.00									
-0.04	**1.00**								
-0.43	**0.33**	**1.00**							
-0.24	**0.25**	**0.71**	**1.00**						
0.47	**0.02**	**-0.11**	**0.00**	**1.00**					
0.12	**0.16**	**-0.03**	**0.16**	**-0.21**	**1.00**				
-0.06	**-0.21**	**-0.03**	**-0.18**	**-0.17**	**0.34**	**1.00**			
0.78	**0.06**	**-0.23**	**-0.18**	**0.67**	**0.06**	**0.02**	**1.00**		
-0.01	**0.41**	**0.37**	**0.54**	**0.15**	**0.03**	**-0.10**	**0.16**	**1.00**	
0.84	**-0.32**	**-0.٧٨**	**-0.35**	**0.60**	**0.00**	**0.00**	**0.85**	**-0.10**	**1.00**

0.46	**0.08**	**-0.04**	**0.04**	**0.35**	**0.00**	**-0.13**	**0.78**	**-0.11**	**0.40**	**1.00**					
-0.25	**0.13**	**0.00**	**0.17**	**-0.01**	**0.04**	**0.05**	**-0.20**	**0.17**	**-0.23**	**-0.36**	**1.00**				
0.20	**-0.16**	**0.01**	**0.18**	**-0.21**	**0.50**	**0.43**	**0.06**	**0.00**	**0.09**	**-0.07**	**-0.04**	**1.00**			
0.22	**0.13**	**-0.27**	**-0.16**	**-0.13**	**0.34**	**0.04**	**0.07**	**-0.03**	**0.03**	**-0.23**	**0.16**	**0.13**	**1.00**		
-0.27	**-0.07**	**0.09**	**-0.05**	**-0.50**	**0.50**	**0.57**	**-0.29**	**-0.18**	**-0.27**	**-0.33**	**0.10**	**0.33**	**0.14**	**1.00**	
-0.09	**0.09**	**0.04**	**0.01**	**-0.11**	**0.32**	**-0.03**	**-0.08**	**0.11**	**-0.08**	**-0.12**	**0.21**	**0.06**	**0.20**	**-0.04**	**1.00**
-0.17	**0.05**	**0.51**	**0.43**	**0.16**	**-0.21**	**-0.03**	**0.17**	**0.46**	**-0.08**	**-0.21**	**0.09**	**0.08**	**-0.05**	**-0.15**	**0.00**	**1.00**

وان محدد مصفوفة الارتباطات يساوي **(1.07E -05)** لذا فان قيمة كاي سكوير ستكون كالآتي:

$$\chi_0^2 = -\left[\left[100 - 1 - \frac{1}{6}(2(17) + 5)\right]\ln(1.07\mathrm{E} - 05)\right]$$
$$= 1058.600546$$

ومن جداول مربع كاي فان القيمة الجدولية لمربع كاي بدرجة حرية **P(P-1)/2=136** وعند مستوى معنوية ٠.٠٥ **n=100,** و **P=17** كانت **(164.216)** وهي اقل من القيمة المحسوبة لذلك نرفض فرضية العدم اي انه توجد مشكلة التعدد الخطي .

٢- عامل تضخم التباين VIF **حيث كانت نتائج هذا المقياس كالاتي:**

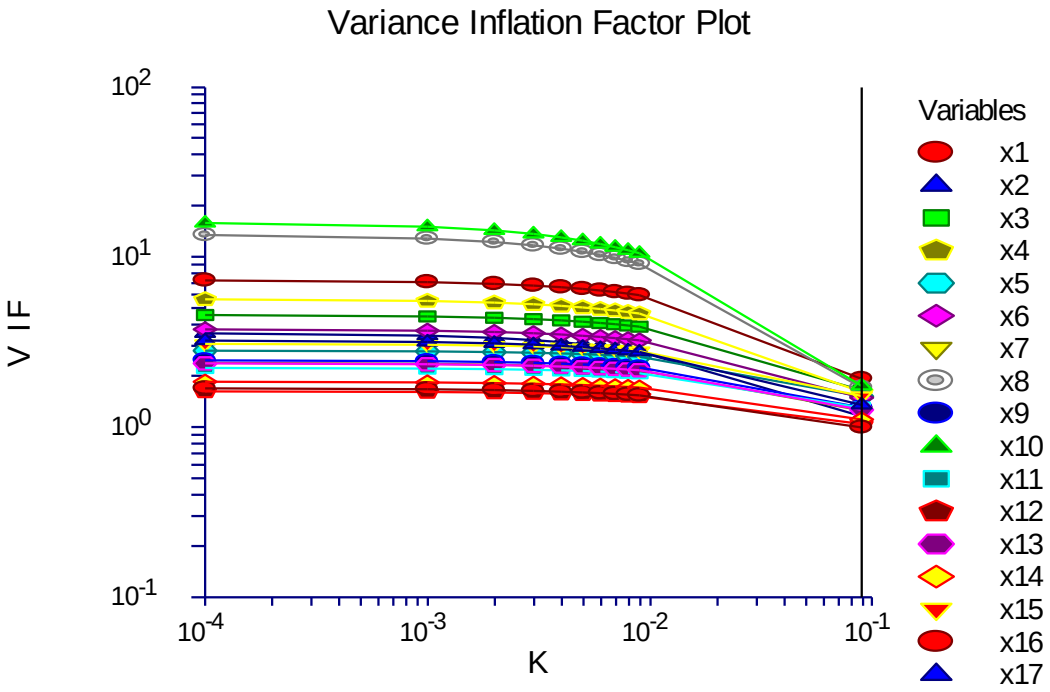
جدول(١) عامل تضخم التباين للمتغيرات

Variable	VIF	Remark
x1	**7.322**	**Co-Linearity**
x2	**3.575**	**Non-Co-Linearity**
x3	**4.582**	**Non-Co-Linearity**
x4	**5.672**	**Co-Linearity**
x5	**2.826**	**Non-Co-Linearity**
x6	**3.773**	**Co-Linearity**
x7	**2.395**	**Co-Linearity**

x8	**13.550**	**Co-Linearity**
x9	**2.479**	**Non -Co-Linearity**
x10	**15.936**	**Co-Linearity**
x11	**2.236**	**Non-Co-Linearity**
x12	**1.628**	**Non-Co-Linearity**
x13	**2.379**	**Non-Co-Linearity**
x14	**1.855**	**Non-Co-Linearity**
x15	**3.099**	**Non-Co-Linearity**
x16	**1.695**	**Non-Co-Linearity**
x17	**3.241**	**Non-Co-Linearity**

حيث ظهرت زيادة قيمة هذا المقياس عن ٥ لكل من المتغيرات الاتية :

(X1:عمر المرأة ، X4:التحصيل الدراسي للزوج ، X8: عمر الزوج ،X10: فترة الزواج)



شكل رقم (١) عامل تضخم التباين للمتغيرات المستقلة

٣- مقياس المعيار الشرطي Condition Index : : حيث اظهرت نتائج التحليل الآتي:

جدول (٢) قيم المعيار الشرطي للمتغيرات المستقلة

Variable	CI
x1	**3.981**
x2	**6.201**
x3	**7.494**
x4	**8.857**

x5	**13.624**
x6	**16.473**
x7	**17.189**
x8	**19.851**
x9	**25.475**
x10	**30.281**
x11	**31.389**
x12	**35.474**
x13	**45.245**
x14	**58.864**
x15	**68.543**
x16	**77.924**
x17	**103.187**

ومن الجدول اعلاه نلاحظ ان المتغيرات الاتية : x5 (وزن المرأة) و x6 (استخدام المرأة لوسائل الحمل) و x7 (تدخين المرأة) وx8 (عمر الزوج) و x9 (مهنة الزوج) و x10 (فترة الزواج) قيمها تتراوح بين$30 \leq CI \leq ١٠$ ، بينما كانت المتغيرات x11 (عدد الاطفال المتوفين) و x12 (عدد ساعات ممارسة الرياضة في الاسبوع) و x13 (اصابة المرأة بأمراض الغدة الدرقية)، بينما كانت المتغيرات X14 (عدد ساعات نوم المرأة في اليوم) و X15 (تناول الادوية من قبل المرأة) و X16 (مدة الرضاعة الطبيعية) و X17 (مهنة الام) اكبر من ٣٠

ومن خلال المؤشرات السابقة تبين لنا وجود مشكلة التعدد الخطي بين المتغيرات المستقلة .

التقدير بطريقة المربعات الصغرى الاعتيادية OLS:

حسب المتغيرات الداخلة في الدراسة ستكون معادلة الانحدار الخطي البسيط كالاتي :

$$\hat{y}=b_0 +b_1 x_1 + b_2 x_2+ b_3 x_3+ b_4 x_4+ b_5 x_5+ b_6 x_6+ b_7 x_7+ b_8 x_8 + b_9 x_9+ b_{10} x_{10}+ b_{11} x_{11}+ b_{12} x_{12}+ b_{13} x_{13}+ b_{14} x_{14}+ b_{15} x_{15}+ b_{16} x_{16}+ b_{17} x_{17} \quad \ldots(1)$$

وباستخدام برنامج ماتلاب لايجاد معادلة الانحدار التقديرية ، كانت النتائج كما مبينة في الجدول :

جدول (3) تقديرات المربعات الصغرى OLS الاعتيادية وقيمة t المحسوبة والقيمة الاحتمالية

Variable	Estimate	S.E.	T	P-Value
x1	**-0.6170**	**0.7569**	**-0.8152**	**0.4173**
x2	**0.0538**	**0.0813**	**0.6618**	**0.5100**
x3	**-0.1139**	**0.1434**	**-0.7946**	**0.4291**

x4	**-0.0197**	**0.1937**	**-0.1016**	**0.9193**
x5	**-0.0616**	**0.0260**	**-2.3721**	**0.0200**
x6	**2.5289**	**0.8210**	**3.0802**	**0.0028**
x7	**1.0886**	**0.4380**	**2.4857**	**0.0150**
x8	**-0.0757**	**0.0613**	**-1.2347**	**0.2205**
x9	**-0.6970**	**0.4174**	**-1.6699**	**0.0988**
x10	**0.3519**	**0.0624**	**5.6412**	**0.0000**
x11	**1.3074**	**0.3153**	**4.1469**	**0.0001**
x12	**-0.0849**	**0.0605**	**-1.4035**	**0.1643**
x13	**-2.4201**	**0.6300**	**-3.8416**	**0.0002**
x14	**-0.0814**	**0.2542**	**-0.3201**	**0.7497**
x15	**1.1597**	**0.5277**	**2.1977**	**0.0308**
x16	**0.1643**	**0.0652**	**2.5200**	**0.0137**
x17	**1.5553**	**0.5811**	**2.6765**	**0.0090**

من ملاحظة الجدول (3) نجد ان المتغيرات (x5 وزن المرأة ، X_6 استخدام وسائل منع الحمل ، X_7تدخين المرأة ، X_{10} فترة الزواج، X_{11} عدد الاطفال المتوفين ، X_{13} اصابة المرأة بأمراض الغدة الدرقية ، X_{15} تناول الادوية من قبل المرأة و X_{16} مدة الرضاعة الطبيعية و X_{17} مهنة الام) كانت معنوية لأنها تمتلك قيمة P-Value اقل من مستوى المعنوية 0.05 ، في حين ان باقي المتغيرات (X_1 عمر المرأة ، X_2 عمر المرأة عند الزواج،X_3 التحصيل الدراسي للمرأة ، X_4 التحصيل الدراسي للزوج ، X_8 عمر الزوج بالسنوات ،X_9 مهنة الزوج ، X_{12} عدد ساعات ممارسة الرياضة وX_{14} عدد ساعات نوم المرأة) كانت غير معنوية بسبب ان قيمة P-Value اكبر من مستوى المعنوية ٠.٠٥ .

ويمكن كتابة معادلة الانحدار المقدرة بطريقة المربعات الصغرى الاعتيادية كالاتي :

$$\hat{y} = -0.61700\,x_1 + 0.0538\,x_2 - 0.11139\,x_3 - 0.0197\,x_4 - 0.0616\,x_5 + 2.5289\,x_6 + 1.0886\,x_7 - 0.0757\,x_8 - 0.6970\,x_9 + 0.3519\,x_{10} + 1.3074\,x_{11} - 0.0849\,x_{12} - 2.4201x_{13} - 0.0814x_{14} + 1.1597\,x_{15} + 0.1647\,x_{16} + 1.5553\,x_{17} \quad \ldots(2)$$

ونلاحظ من المعادلة (3-2) x_1 (عمر المرأة) غير معنوي لان القيمة الاحتمالية له التي تساوي **(0.4173)** اكبر من مستوى المعنوية ٠.٠٥ وهذه النتيجة لا توافق للمنطق فلا يمكن ان يكون العمر غير مؤثر على درجة خصوبة المرأة ، وكذلك نلاحظ ان معامل المتغير $\mathbf{x_6}$ (استعمال وسائل منع الحمل) البالغ **(2.5289)** اشارته الجبرية موجبة (علاقة طردية) اي كلما استخدمت وسائل منع الحمل زادت خصوبتها وهذا غير منطقي ، وان اشارة متغير (فترة الزواج) موجبة هو صحيح فكلما زادت فترة الزواج اثر في امكانية زيادة عدد الاطفال ، وايضاً المتغير $\mathbf{x_{15}}$ (تناول الأدوية) البالغ **(1.1597)** اشارته موجبة فلا يمكن ان تتناول المرأة الادوية ويزيد اخصابها .

وكما هو ملاحظ من النتائج ان طريقة المربعات الصغرى اعطت تقديرات غير دقيقة لمعاملات الانحدار فهي لم تمثل واقع الظاهرة المدروسة حيث اظهرت نتائج الطريقة ان بعض المتغيرات غير مهمة ولكن في الواقع هي مهمة ولكن بناء الانموذج يعجز عن إظهار اثر كل منها بشكل منفصل نظراً لارتباط هذه المتغيرات بعضها ببعض.

علماً ان جدول تحليل التباين وفق هذه الطريقة هو كالاتي :

جدول(٤) جدول تحليل التباين للنموذج المقدر بطريقة المربعات الصغرى

S.O.V	D.f.	SSE	MSE	F	Sig.
Model	**17**	**727.793**	**42.811**	**22.6896**	**0.0000**
Error	**82**	**154.717**	**1.8868**		
Total	**99**	**882.510**			
MSE	**1.8868**	$\mathbf{R^2}$	**0.83**		

من الجدول (٤) نلاحظ معنوية الانموذج القيمة الاحتمالية .Sig (٠.٠٠٠) لاختبار F اقل من مستوى المعنوية ٠.٠٠٥.

التقدير بطريقة انحدار الحرف البيزية(غير المتحيزة) Unbiased Ridge Regression

لايجاد تقديرات معاملات الانحدار بطريقة انحدار الحرف البيزية سنعتبر بان عناصر منتجه المعلومات الاولية J تمثل معدل تقديرات معلمات المربعات الصغرى الاعتيادية كالاتي Swindle(1976)[8]:

J'=[0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376, 0.2376]

وكانت النتائج كما مبينة في الجدول (٥) :

جدول (٥) تقديرات معاملات الانحدار والخطأ المعياري وقيمة t المحسوبة والقيمة الاحتمالية لطريقة انحدار الحرف البيزية(الغير المتحيزة)عند قيمة معلمة الحرف K=0.09

Variable	Estimates	S.E.	T	P-Value
x1	-0.2456	0.0181	-13.5691	0.0000
x2	-0.0854	0.0189	-4.5185	0.0008
x3	-0.135	0.0311	-4.3408	0.0004
x4	0.015	0.1422	0.1055	0.8964
x5	-0.0751	0.0156	-4.8141	0.0006
x6	-2.719	0.6048	-4.4957	0.0002
x7	0.0034	0.3433	0.0099	0.8302
x8	-0.1134	0.0265	-4.2792	0.0002
x9	-0.0559	0.3234	-0.1729	0.6453
x10	-0.1823	0.045	-4.0511	0.0005
x11	0.9935	0.1064	9.3374	0.0005
x12	0.0336	0.023	1.4609	0. 8960
x13	-2.089	0.3228	-6.4715	0.0003
x14	-0.194	0.1055	-1.8389	0.0174
x15	-1.2245	0.4043	-3.0287	0.0005
x16	0.0008	0.0612	0.0131	0.9957
x17	0.0001	0.3985	0.0003	0.9935

ومن ملاحظة الجدول (٥) نجد ان المتغيرات (X_1 عمر المرأة ، X_2 عمرالمرأة عند الزواج ، X_3 التحصيل الدراسي للمرأة ، X_5 وزن المرأة ، X_6 استخدام وسائل منع الحمل ، X_8 عمر الزوج بالسنوات، X_{10} فترة الزواج، X_{11} عدد الاطفال المتوفين ،X_{13} اصابة المرأة بأمراض الغدة الدرقية ، X_{14} عدد ساعات نوم المرأة و X_{15} تناول الادوية من قبل المرأة) كانت معنوية لان قيمة P-Value اقل من مستوى المعنوية 0.05 ، في حين ان باقي المتغيرات (X_4 التحصيل الدراسي للزوج ، X_7 تدخين المرأة، X_9 مهنة الزوج ،X_{12} عدد ساعات ممارسة الرياضة ،X_{16} مدة الرضاعة الطبيعية و X_{17} مهنة الام) كانت غير معنوية بسبب ان قيمة P-Value اكبر من مستوى المعنوية ٠.٠٥ .

وبالتالي فان معادلة الانحدار المقدرة بطريقة انحدار الحرف البيزية كالاتي :

$$\hat{y}= -0.2456x_1 -0.0854x_2 -0.135x_3 +0.015x_4 -0.0751x_5 -2.719x_6 +0.0034x_7 - 0.1134x_8 -0.0559x_9-0.1823x_{10} +0.9935x_{11} + 0.0336x_{12} -2.089x_{13} - 0.194x_{14}-1.2245x_{15} +0.0008x_{16} +0.0001x_{17} \quad \ldots.(٣)$$

وان جدول تحليل التباين لطريقة انحدار الحرف الغير متحيزة كان كالاتي:

جدول(٦) جدول تحليل التباين للنموذج المقدر بطريقة انحدار الحرف الغير متحيزة

S.O.V	D.f.	SSE	MSE	F	Sig.
Model	**17**	**739.621**	**43.5071**	**24.9695**	**0.0000**
Error	**82**	**142.889**	**1.7424**		
Total	**99**	**882.51**			
MSE	**1.7424**	**R^2**	**0.838**		

من الجدول (٦) نلاحظ معنوية الانموذج القيمة الاحتمالية .Sig (٠.٠٠٠) لاختبار F اقل من مستوى المعنوية ٠.٠٥ ، ونلاحظ ايضاً ان قيمة متوسط مربعات الخطأ البالغ (١.٧٤٢٤) قد انخفضت عن قيمة متوسط مربعات الخطأ لطريقة المربعات الصغرى البالغة (١.٨٨٦٨). ونلاحظ ايضاً ان هذه الطريقة اعطت نتائج منطقية مناسبة للواقع الظاهرة المدروسة . ولكنها اعطت عدم معنوية معاملات الانحدار للمتغيرات X4 التحصيل الدراسي للزوج و X7 تدخين المرأة وX9 مهنة الزوج و X12 عدد ساعات ممارسة الرياضة في الاسبوع و X16 مدة الرضاعة الطبيعية و X17 مهنة الام حيث نلاحظ ان تأثيراتها قليلة جداً كما هو مبين من قيمها **(0.015، 0.0034، 0.0559- ، 0.0336، 0.0008، 0.0001)** على التوالي مقارنة ببقية معاملات الانحدار

.

التقدير بطريقة انحدار لاسو : Lasso Regression

باستخدام برنامج ماتلاب لايجاد معادلة الانحدار التقديرية ، كانت النتائج كما مبينة في الجدول (٧) .

جدول (٧) تقديرات معاملات الانحدار والخطأ المعياري وقيمة t المحسوبة والقيمة الاحتمالية لطريقة انحدار لاسو عند λ=0.0900

Variable	Estimates	S.E.	t	P-Value
x1	**-0.247**	**0.015**	**-16.4667**	**0.0000**
x2	**-0.088**	**0.0149**	**-5.90604**	**0.0002**
x3	**-0.139**	**0.028**	**-4.96429**	**0.0043**
x4	**0**	**-**	**-**	**-**
x5	**-0.077**	**0.015**	**-5.13333**	**0.0000**
x6	**-2.729**	**0.604**	**-4.51821**	**0.0002**
x7	**0**	**-**	**-**	**-**

x8	**0.121**	**0.0254**	**4.76772**	**0.0002**
x9	**0**	**-**	**-**	**-**
x10	**-0.187**	**0.032**	**-5.84375**	**0.0005**
x11	**0.939**	**0.1052**	**8.92586**	**0.0002**
x12	**0**	**-**	**-**	**-**
x13	**-2.083**	**0.3132**	**-6.6507**	**0.0003**
x14	**-0.189**	**0.105**	**-1.8000**	**0.0189**
x15	**-1.222**	**0.4012**	**-3.04561**	**0.0005**
x16	**0**	**-**	**-**	**-**
x17	**0**	**-**	**-**	**-**

ومن ملاحظة الجدول (٧) نجد ان المتغيرات (X_1 عمر المرأة ،X_2 عمرالمرأة عند الزواج ، X_3 التحصيل الدراسي للمرأة ، X_5 وزن المرأة ، X_6 استخدام وسائل منع الحمل ،X_8 عمر الزوج بالسنوات ، X_{10} فترة الزواج ، X_{11} عدد الاطفال المتوفين ، X_{13} اصابة المراة بامراض الغدة الدرقية، X_{14} عدد ساعات نوم المرأة و X_{15} تناول الادوية من قبل المراة) كانت معنوية لان قيمة P-Value اقل من مستوى المعنوية 0.05 ، في حين ان باقي المتغيرات (X_4 التحصيل الدراسي للزوج، X_7 تدخين المرأة، X_9مهنة الزوج ، X_{12} عدد ساعات ممارسة الرياضة، X_{16} مدة الرضاعة الطبيعية وX_{17} مهنة الام) قد قلصت من الانموذج (اصبحت معاملات انحدارها تساوي صفر) .

وستكون معادلة الانحدار المقدرة بطريقة انحدار لاسو كالاتي :

$$\hat{y} = -0.247x_1 - 0.0881x_2 - 0.139x_3 - 0.077x_5 - 2.729x_6 + 0.121x_8 - 0.187x_{10} + 0.939x_{11} - 2.083x_{13} - 0.189x_{14} - 1.222x_{15} \quad \ldots(٤)$$

وان جدول تحليل التباين لطريقة انحدار لاسو كان كالاتي:

جدول(٨) جدول تحليل التباين للنموذج المقدر بطريقة انحدار لاسو

S.O.V	D.f.	SSE	MSE	F	Sig.
Model	**17**	**751.931**	**44.2311**	**27.777**	**0.0000**
Error	**82**	**130.579**	**1.5923**		
Total	**99**	**882.51**			
MSE	**1.5923**	**R^2**	**0.852**		

من الجدول (8) نلاحظ معنوية الانموذج القيمة الاحتمالية .Sig (٠.٠٠٠) لاختبار F اقل من مستوى المعنوية ٠.٠٥ ، ونلاحظ ايضاً ان قيمة متوسط مربعات الخطأ البالغ (1.5923) قد

انخفضت عن قيمة متوسط مربعات الخطأ لطريقة المربعات الصغرى البالغة (١.٨٨٦٨) وطريقة انحدار الحرف الغير متحيزة البالغه (١.٧٤٢٤) .

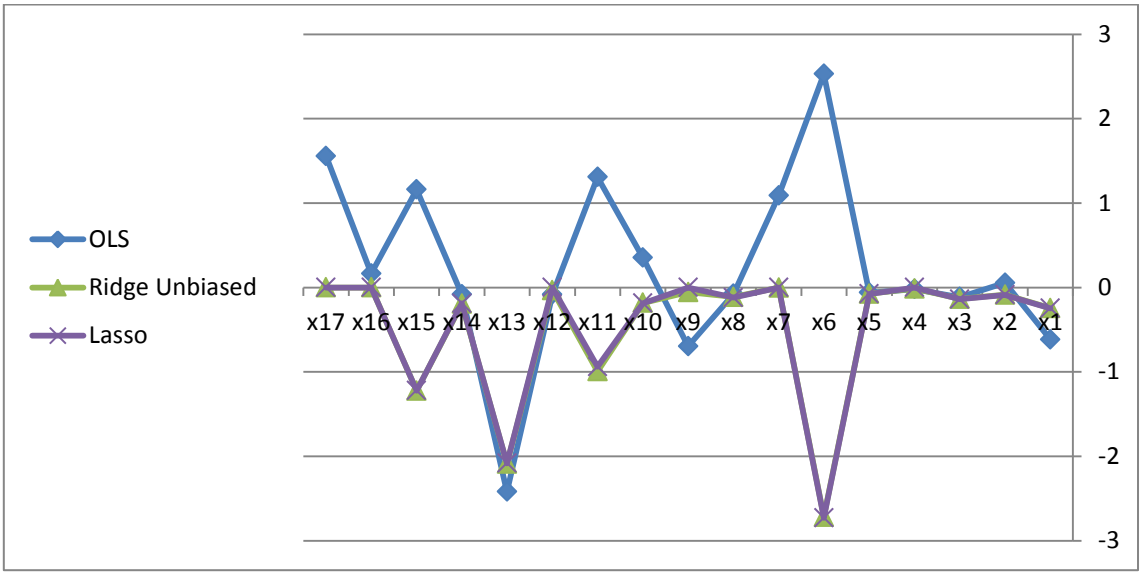
ونلاحظ ان طريقة انحدار لاسو قد قلصت المتغيرات الداخلة في الانموذج الى ١١ متغيرات فقط واستبعدت الـ (٦) متغيرات الباقية ، حيث اجبرت معاملاتها ان تكون مساوية للصفر وهي X_4 التحصيل الدراسي للزوج و **X7** تدخين المرأة و X_9 مهنة الزوج و **X12** عدد ساعات ممارسة الرياضة في الاسبوع و **X16** مدة الرضاعة الطبيعية و **X17** مهنة الأم ، وكما لاحظنا ان هذه المعاملات كانت قد ظهرت غير معنوية وبتأثيرات قليلة في طريقة انحدار الحرف البيزية الغير متحيزة .

مقارنة بين طريقة المربعات الصغرى الاعتيادية وطريقة انحدار الحرف البيزية وانحدار لاسو :

الجدول التالي يبين مقارنة بين معاملات الانحدار المقدرة بالطرائق الثلاثة :

جدول (9) مقارنة بين معاملات الانحدار المقدرة بطريقة المربعات الصغرى وانحدار الحرف البيزية وانحدار لاسو

Method	OLS	Ridge Unbiased	lasso
x1	-0.6170	-0.2456	-0.247
x2	0.0538	-0.0854	-0.088
x3	-0.1139	-0.135	-0.139
x4	-0.0197	0.015	0
x5	-0.0616	-0.0751	-0.077
x6	2.5289	-2.719	-2.729
x7	1.0886	0.0034	0
x8	-0.0757	-0.1134	0.121
x9	-0.6970	-0.0559	0
x10	0.3519	-0.1823	-0.187
x11	1.3074	0.9935	0.939
x12	-0.0849	0.0336	0
x13	-2.4201	-2.089	-2.083
x14	-0.0814	-0.194	-0.189
x15	1.1597	-1.2245	-1.222
x16	0.1643	0.0008	0
x17	1.5553	0.0001	0



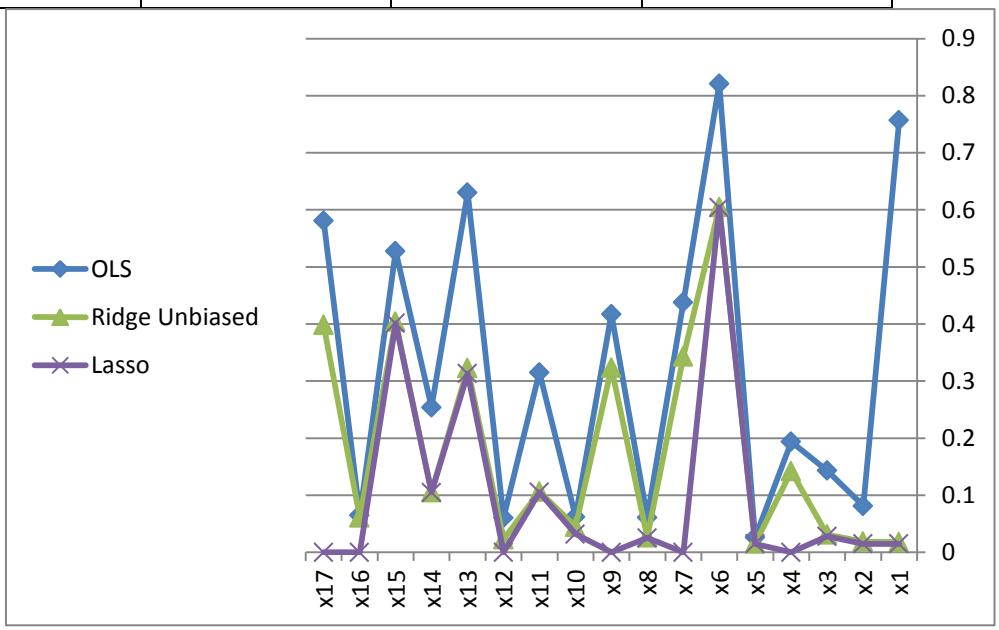
الشكل (٢) مقدرات معاملات الانحدار المقدرة بطريقة المربعات الصغرى وانحدار الحرف البيزية وانحدار لاسو

ونلاحظ من الشكل (٢) تقارب طريقتي انحدار الحرف البيزية الغير متحيزة وطريقة انحدار لاسو في حين ابتعاد طريقة المربعات الصغرى الاعتيادية بشكل كبير .

جدول (10) الخطأ المعياري لمعاملات الانحدار المقدرة بطريقة المربعات وانحدار الحرف البيزية وانحدار لاسو

Method	OLS	Ridge Unbiased	Lasso
x1	0.7569	0.0181	0.015
x2	0.0813	0.0189	0.0149
x3	0.1434	0.0311	0.028
x4	0.1937	0.1422	-
x5	0.0260	0.0156	0.015
x6	0.8210	0.6048	0.604
x7	0.4380	0.3433	-
x8	0.0613	0.0265	0.0254
x9	0.4174	0.3234	-
x10	0.0624	0.045	0.032
x11	0.3153	0.1064	0.1052
x12	0.0605	0.023	-
x13	0.6300	0.3228	0.3132
x14	0.2542	0.1055	0.105
x15	0.5277	0.4043	0.4012
x16	0.0652	0.0612	-

x17	**0.5811**	**0.3985**	-



الشكل (٣) الخطأ المعياري لمقدرات معاملات الانحدار المقدرة بطريقة المربعات الصغرى وانحدار الحرف البيزية وانحدار لاسو

ونلاحظ من الجدول (10) والشكل (٣) ان الاخطاء المعيارية لطريقة لاسو كانت اقل من الاخطاء المعيارية للطرائق الباقية.

والجدول (11) يمثل مقارنة بين الطرائق الثلاثة بالاعتماد على معيار متوسط مربعات الخطأ ومعيار معامل التحديد والقيمة الاحتمالية P-Value كالاتي:

الجدول (11) متوسط مربعات الخطأ والقيمة الاحتمالية للطرائق الثلاثة في التقدير

Method	MSE	R^2	P-Value
O.L.S	**1.8868**	**0.83**	**0.000**
Unbiased Ridge	**1.7424**	**0.838**	**0.000**
Lasso	**1.5923**	**0.852**	**0.000**

ونلاحظ من الجدول (١١) ان افضل طريقة في تقدير معلمات انموذج الانحدار هي طريقة لاسو بمتوسط مربعات خطأ **1.5923** ثم طريقة انحدار الحرف البيزية بمتوسط مربعات خطأ **1.7424** واخيرا طريقة المربعات الصغرى بمتوسط مربعات خطأ 1.8868 ونلاحظ ايضاً معنوية الطرائق الثلاثة في التقدير لان قيمة P-Value اقل من مستوى المعنوية ٠.٠٥ .

الاستنتاجات : Conclusions

من خلال ما تم التوصل اليه من نتائج في الجانب التطبيقي توصلنا الى الاستنتاجات الاتية :

١- ظهور عدم معنوية بعض المتغيرات المستقلة على الانموذج رغم اهميتها بسبب وجود مشكلة التعدد الخطي بين المتغيرات المستقلة.

٢- افضلية طريقة انحدار لاسو على طريقة انحدار الحرف البيزية الغير متحيزة.

٣- تقارب طريقتي انحدار الحرف البيزية وطريقة انحدار لاسو

٤- في طريقة لاسو ، تم تقليص المتغيرات المستقلة الاقل اهمية في الانموذج ($\mathbf{X_4}$ التحصيل الدراسي للزوج و **X7** تدخين المرأة و X_9 مهنة الزوج و **X12** عدد ساعات ممارسة الرياضة في الاسبوع و **X16** مدة الرضاعة الطبيعية و **X17** مهنة الأم) وقلصت طريقة انحدار لاسو معاملات الانحدار التي ظهرت ذات تأثير قليل في طريقة انحدار الحرف البيزية .

التوصيات : Recommendations

من خلال ما تم عرضه من استنتاجات نوصي بالاتي :

١- التوسع في دراسة طريقة انحدار لاسو كونها من الطرائق الحديثة في تقدير النماذج الاحصائية والتي تلعب دور فعال في تحديد المتغيرات المستقلة الاكثر اهمية للنموذج.

٢- التوسع في اجراء دراسات وبحوث لمقارنة طرائق لاسو الاعتيادية مع لاسو البيزية .

٣- تطبيق طريقة انحدار لاسو على بيانات اقتصادية لبيان فعاليتها في معالجة مشكلة التعدد الخطي.

المصادر : References

١. النعيمي ، اسوان محمد طيب ; دبدوب ، مروان عبد العزيز، (٢٠٠٦) ، " طرائق مقترحة في انحدار الحرف " ، المجلة العراقية للعلوم الاحصائية ، (١٠)، ٨٥-١٠٦.

٢. عبودي ، عماد حازم ، علي ، حميد يوسف ، (٢٠١٧) ، " مقارنة مقدري Hurber Lasso و Hurber Elastic Net باستخدام المحاكاة " ، مجلة الكوت للعلوم الاقتصادية والادارية ، العدد (٢٨) ، الجزء الاول ، بحث مستل من اطروحة دكتوراه .

٣. مزاحم، محمد يحيى; محمود ، حمود عبد الله ، (٢٠٠٧) " تشخيص التعدد الخطي واستخدام انحدار الحرف في اختيار متغيرات دالة الاستثمار الزراعي في العراق للفترة ١٩٨٠ – ٢٠٠٠ ، مجلة تكريت للعلوم الادارية والاقتصادية ، الجلد -٣ ، العدد-٨. ، ١٧١-١٨٧.

٤. مزاحم، محمد يحيى ، (٢٠٠٥) ، " استخدام المكونات الرئيسية وانحدار الحرف في تقدير معادلة السعر العالمي للقمح للفترة من ١٩٦١-٢٠٠٢) ، مجلة تكريت للعلوم الادارية والاقتصادية ، الجلد -١ ، العدد-١. ، ١٤٦-١٥٦.

5. Al-Hassan ,Yazid M., (2010) , " Performance of a new ridge regression estimator" , Journal of the Association of Arab Universities for Basic and Applied Sciences , 9, 23–26 .

6. Ali , Sadig Mohommed BAGER, (2018) , " Ridge Regression for Addressing of the Multicollinearity Problem with Application in Cost of Production" , 3rd Central & Eastern European LUMEN International Conference New Approaches in Social and Humanistic Sciences | NASHS 2017| Chisinau, Republic of Moldova | June 8-10, 2017, 57-63.
7. Al-Sadoun , Muhannad Faiz , (2005) . " Empirical Bayes and Bayes for Ridge regression" , Quarterly Specialized Refereed Journal, vol. (7) no. (1).
8. E. Farrar Donald , R. Glauber Robert, (1967) , " Multicollinerity in regression analysis: The problem revisited " , The Review of Economics and Statistics, Vol. 49, No. 1 (Feb., 1967), pp. 92-107 .
9. Buhlmann, Peter ; van de Geer, Sara,(2011), " Statistics for High-Dimensional Data , Methods, Theory and Applications, Springer , Heidelberg Dordrecht London New York.
10. C. Montgomery, Douglas; A. Peck, Elizabeth; Vinining G. Geoffery , (1982). "Introduction to linear regression analysis" , Fifth Edition, A JOHN WILEY & SONS, INC., PUBLICATION .
11. Dyar, M.D.; Carmosino ,M.L.; Breves, E.A., Ozanne, M.V.; Clegg S.M.; Wiens, R.C., (2012)" Comparison of partial least squares and lasso Regression techniques as applied to laser-induced breakdown spectroscopy of geological samples" , Spectrochimica Acta Part B 70 51–67.
12. F. Benee,Swindle , (2013) , " Good ridge estimators based on prior information", Moskow State Univ Bibliote, Taylor & Francis.
13. Fonti Valeria, (2017) , " Feature Selection using LASSO" , Research Paper in Business Analytics, VU Amsterdam .
14. Hans ,Christ , (2009) , " Bayesian lasso regression" , Biometrika ,96, 4, Biometrika Trustpp. 835–845 .
15. J. Fu Wenjiang , (1998) , " Penalized Regressions: The Bridge Versus the Lasso " , Journal of Computational and Graphical Statistics, Volume 7, Number 3, Pages 397-416.
16. L.J, Pliskin , (1987) "A Ridge type Estimator and good prior Means. Communication in Statistics, 16,3427-3429.
17. Leamer ,E.E.(1973). Multicollinearity: A Bayesian interpretation",Rev.Econ.Statist.,55,371-380.
18. Li Fan; Yang Yiming ; P. Xing ,Eric, (2006) , " From Lasso regression to Feature vector machine ",Pittsburgh, PA USA 15213, fhustlf,yiming,epxingg@cs.cmu.edu.
19. Lindley, D.V.; Smith, A.F.M, (1972), " Bayes Estimation for the linear Model" , Journal of Royal Statistical Society , Series B (Methodological) , Volume 34, Issue 1 , 1-41.
20. Malo ,Pekka, (2018), " Modern Regression + Analysis: Feature + selection and +regularization techniques " , Aalto University , school of

businesses , Aalto BIZ / Department of Information and Service Management.
21. Mason, R.L.F. Gunst; J.T. Wabster, (1975), ' Regression analysis and problems of multicoleanirutiy " , comM. sTATST,4,3,277.
22. Ranstam ,J.; Cook , J. A., (2018) , " Lasso Regression " , BJS Statistical Editors .
23. Robert H. Crouse & Chun Jin.(1995)."Unbiased RIDGE Estimation With prior Information And Ridge Trace"Commun.Statist.-Theory Meth.27(9),2341-2354.
24. Saleh ,A.K.Md. Ehsanes; Kibria ,B.M. Golam, (1993), " Performance of some new Preliminary tests Ridge Regression estimators and other properites" , COMMUN. STATIST.-THEORY METH., 22(10), 2747-2764 .
25. Tibshirani , Robert, (1996) ," Regression Shrinkage and Selection via the Lasso" , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.
26. Tibshirani , Robert, (1997) ," The LASSO method for variable selection in Cox Model " , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.
27. Tibshirani , Ryan , (2013) , " Modern regression " , Optional reading: ISL 6.2.1, ESL 3.4.1, Data Mining: 36-462/36-662.
28. W. Wichern , Dean ; A. Churchill ,Gilbert , (2012) ."A Comparison of Ridge Estimators", T Econometrics Journal, 301-311.
29. Wang, Hansheng; Li ,Guodong ; Tsai ,Chih-Ling , (2007) , " Regression coefficient and autoregressive order shrinkage and selection via the lasso" , J. R. Statist. Soc. B ,69, Part 1, pp. 63–78.
30. Yi Congrui ; Huang Jian, (2016), " Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quintile Regression " , arXiv:1509.02957v2 [stat.CO] 20.
31. Zellner ,A.(1971).An introduction to Bayesian Inference in Econometrics,Wiley,New york.
32. Zou Hui ; Hastie Trevor, (2005) " Regularization and variable selection via the elastic net", J. R. Statist. Soc. B ,67, Part 2, pp. 301–320 .