regression and Bayesian ridge regression models

مقارنة تقدير معلمات أأنموذجي انحدار الحرف المتحيزة وانحدار الحرف البيزي

أ.م.د. شروق عبد الرضا سعيد السباح
Ass. Prof. Shrook A.S.AL-Sabbah
shrookalsabah1968@gmail.com

سارة ماجد حسين الكفيشي
Sara Majid Hussein Al-Kufaishi
Saramajid106@gmail.com

الملخص :

ان مهمة اختيار المتغيرات تعد من الامور المهمة لبناء أأنموذج يؤدي تنبؤات دقيقة , لاسيما عندما تكون البيانات تحت الدراسة تعاني من مشكلة التعدد الخطي , ولوجود كثير من الظواهر في حياتنا اليومية لاسيما الاجتماعية تعاني من هذه المشكلة , لذا جاء البحث بهدف المقارنة بين طرائق تقدير معلمات أأنموذج انحدار الحرف المتحيزة(الاعتيادية) وأأنموذج انحدار الحرف البيزي , وتمت المقارنة بين الطريقتين عن طريق معيار متوسط مربعات الخطأ MSE والقيمة الاحتمالية P-Value وطبقت الطريقتين عملياً بسحب عينة عشوائية مؤلفة من (100) 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, مدة الزواج , عدد الاطفال المتوفين , عدد ساعات مارسة الرياضة في الاسبوع , اصابة بامراض الغدة الدرقية , عدد ساعات نوم المراة في اليوم , تناول الادوية لدى المرأة , مدة الرضاعة الطبيعية , مهنة الام) , وقد اتضح من الاختبا ان البيانات تعاني من مشكلة تعدد خطي .

وتم التوصل الى ان طريقة انحدار الحرف البيزي افضل من طريقة انحدار الحرف المتحيزة(الاعتيادية) لامتلاكها اقل متوسط مربعات خطأ.

Abstract

The task of selecting variables is important to build a model that leads to accurate predictions, especially when the data under study suffer from the problem of linear multiplicity, and the presence of many phenomena in our daily life, especially the social suffer from this problem, In order to compare the methods of estimating the parameters of the regression model of the biased ridge regression and the Bayesian ridge regression. As a comparison of the two methods through the mean squares error MSE and the probability value P-Value. In order to apply the methods in practice, a random sample of 100 women's fertility was withdrawn to study the factors affecting the number of children born as dependent variable and many independent random variables is (woman age , the age at married ,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 Bayesian regression method has the best estimation methods for owning the MSE .

الكلمات المفتاحية :أأنموذج خطي متعدد , انحدار الحرف المتحيزة , انحدار الحرف البيزية

المقدمة : Introduction

الانحدار **Regression** بشكل عام هو 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 أأنموذج أحصائي لتقدير العلاقة بين المتغير التابع(**Variable Dependent**) (متغير الاستجابة) وعدة متغيرات توضيحية (**Explanatory Variables**) (تفسيرية) وهي المتغيرات المستقلة، ينتج عنها أأنموذج أحصائي يوضح 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 الأأنموذج بين متغير تابع واحد ومتغير مستقل واحد فأن هذا الأأنموذج هو يعد من أبسط نماذج الانحدار ويسمى أأنموذج الانحدار الخطي البسيط(**Simple Linear Regression Model**) ، اما عندما تكون العلاقة خطية في الأأنموذج بين متغير تابع مع اثنين او اكثر من المتغيرات التوضيحية (التفسيرية) يسمى أأنموذج الانحدار الخطي المتعدد(**Multiple Linear Regression Model**). وبينما يكون عدد المتغيرات المستقلة كبير جدا باذ 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 يمكن عدها كبديل مجدي لحل هذه المشكلة فهي تعطي سهولة ومرونة في الحسابات منها انحدار الحرف وانحدار لاسو . لغرض التوصل الى أأنموذج 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 أأنموذجي انحدار الحرف المتحيزة(الاعتيادية) والانحدار البيزي في ظل وجود مشكلة التعدد الخطي باستعمال بعض المعايير الاحصائية المعروفة للتوصل الى افضل الطرائق في التقدير.

يوجد العديد من الدراسات والبحوث التي اهتمت بموضوع تقديرات معلمات أأنموذج الانحدار , لاسيما طرائق انحدار الحرف, لذلك سيتم عرض بعض الدراسات التي تخص موضوع البحث , ف **في عام 1970 اقترح Hoerl & (Kennard)** [11] طريقة لتقدير معاملات أنموذج الانحدار المتعدد مستنداً الى اضافة كمية موجبة صغيرة لقطر مصفوفة المعلومات **X'X** اطلق عليه اسم اثر الحرف , اذ ان اضافة هذه الكمية الموجبة يزيد من امكانية جعل البيانات متعامدة ومن ثم الحصول على تقديرات افضل لمعاملات إأنموذج الانحدار . **عام 1972 عما الباحثان (Lindley & Smith)** [16] الأنموذج الاحصائي الخطي باستعمال الطرائق البيزية وقاما بتوضيح هذه الطرائق على أنموذج انحدار خطي متعدد وبينا بان الطرائق البيزية تعطي تقديرات غير متحيزة لمعاملات أنموذج الانحدار الخطي المتعدد . **في عام 1975 قدم (Arthur) وآخرون** [13] خوارزمية لاختيار معلمة الحرف **k** في انحدار الحرف وبينت الخوارزمية عن طريق المحاكاة بان متوسط مربعات الخطأ ومربعات الخطأ في طريقة انحدار الحرف اقل من متوسط مربعات الخطأ ومربعات الخطأ في طريقة المربعات الصغرى الاعتيادية .. **عام 1981 وصفا (Bruce & Hann)** [6] طريقة انحدار الحرف في حالة وجود مشكلة التعدد الخطي على أنموذج الانحدار الخطي المتعدد واستخدما ثلاثة معايير في اختيار افضل مقدر هي اثر الحرف , عامل تظخم التباين ومحدد مصفوفة الارتباط .**في عام 2003 استعمل (KIBRIA)** [14] طرائق مختلفة لتقدير معلمة الحرف في انحدار الحرف , وكذلك قدم مقدرات جديدة بالاعتماد على طريقة انحدار الحرف المعممة وقارن بين تلك الطرائق طريقة المربعات الصغرى الاعتيادية عن طريق المحاكاة باستعمال معيار متوسط مربعات الخطأ **MSE**. **عام 2007 اقترحا (ALKhamisi & Shukur)** [4] اسلوب جديد للحصول على معلمة الحرف **K** , وتم اختبار الاسلوب المقترح عن طريقة تجارب محاكاة مونت–كارولو مختلفة لاعداد مشاهدات مختلفة , واستخدما معيار قوة الارتباط

بين المتغيرات المستقلة ومعيار متوسط مربعات الخطأ لمقارنة أداء المقدرات المقترحة مقارنة بمقدرات معروفة تحت شرط المعايير المستخدمة وتبين من نتائج المحاكاة ان واحداً على الاقل من المقدرات المقترحة له اقل متوسط مربعات خطأ من مقدرات المربعات الصغرى المقترحة لدى **Hoerl HK** و **Kenerd** . **في عام** 2012 **اثبتا** (**Wichern & Churchill)** [20] ان طريقة انحدار الحرف تقلل من التقديرات المتطرفة لطريقة المربعات الصغرى الاعتيادية اذ قارنا بين الطريقتين باستخدام تجارب محاكاة مونت–كارلو. **العام 2015 اقترحا (Duzan & Shariff)** [8] طريقة جديدة لتقدر معلمة الحرف في ظل شرط وجود اثنين او اكثر من العلاقات بين المتغيرات التي تؤدي الى مشكلة التعدد الخطي اذ هدفت الدراسة الى تحديد معلمة الحرف **K** باذ يكون حد تباين معلمات الانحدار اكبر من الزيادة في مربع التحيز, بالاضافة الى ان قيم معلمات الحرف الغير صفرية لاي من متوسط مربعات الخطأ لمعلمات الانحدار اقل من تباين طريقة المربعات الصغرى لنفس المعلمات.

وفي هذا البحث سيتم مقارنة طريقة انحدار الحرف الاعتيادية وطريقة الانحدار البيزية التي لم يتم التطرق

اليها سابقا لدى الباحثين.

مشكلة البحث: Problem of the research

الكثير من الدراسات التطبيقية , ومنها الدراسات الاجتماعية يكون عدد المتغيرات المستقلة المؤثرة في المتغير المعتمد كبيرة وبعضها غير معروف ومتداخلة فيما بينها فتنشأ مشكلة التعدد الخطي بين المتغيرات ويؤدي ذلك الى تنبؤات خاطئة , وكثيراً من الطرائق استعملت لتلافي مشكلة التعدد الخطي بين المتغيرات المستقلة ولكنها لا تحل هذه المشكلة بصورة نهائية , لذلك لابد من البحث عن طرائق مهمتها حل مشكلة التعدد الخطي واختيار عدد المتغيرات المستقلة التي لها تأثير كبير والخروج بالأنموذج الافضل الذي تكون لديه قوة تفسيرة وتنبؤية عالية جداً. فجاء هذا البحث للمقارنة بين بعض طرائق تنظيم واختيار المتغيرات لتحديد الافضل منها.

هدف البحث: Aim of the research

الهدف من هذا البحث هو مقارنة تقديرات معلمات أأنموذجي انحدار الحرف المتحيزة (الاعتيادية) وانحدار الحرف البيزي في ظل وجود مشكلة التعدد الخطي باستعمال بعض المعايير الاحصائية المعروفة للتوصل الى افضل طريقة في التقدير.

الإأنموذج الخطي Liner Model

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

وان افتراضات إأنموذج الانحدار الخطي هي كالاتي :

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

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

2. خطية الإأنموذج , اي العلاقة بين المتغير المعتمد والمتغير او المتغيرات المستقلة علاقة خطية .

ويمكن ان نعبر عن الإأنموذج 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 وسوف نفترض له متوسطاً صفراً وتبايناً ثابتاً , والاخطاء مستقلة عن بعضها , اي نعبر عنها رياضياً :

$$u \sim N(0, \sigma^2_I)$$

$$cov(u_i, u_j) = 0$$

ويمكن ان يكتب بصيغة المصفوفات كالاتي :

$$\underline{Y} = X\underline{\beta} + \underline{U} \quad \ldots(1)$$

اذ ان :

Y : متجه nx1 يفترض ان يتوزع توزيعا طبيعيا بمتوسط B_0+B_1Xi وتباين ثابت σ^2_u

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.[10] [19]

وتقدر معلمات الأأنموذج (1) باتباع طريقة المربعات الصغرى OLS حسب الصيغة الاتية:

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad \ldots(2)$$

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

$$\hat{y} = X\hat{\beta} \quad \ldots(3)$$

مشكلة التعدد الخطي Multi-CoLinearity

مشكلة الارتباط الخطي المتعدد Multi-CoLinearity تحدث في نماذج الانحدار المتعدد ولاتحدث في إأنموذج الانحدار البسيط. فهي مشكلة خاصة بالمتغيرات المستقلة.[1]

نُواجَه مشكلة التعدد الخطي عندما يرتبط بعض او كل المتغيرات المستقلة بعلاقة خطية يصبح من الصعب فصل اثر كل متغير على المتغير المعتمد , او عندما تكون قيمة احد المتغيرات المستقلة متساوية المشاهدات كافة او عندما تعتمد قيمة احد المتغيرات المستقلة على قيمة واحد او اكثر من المتغيرات المستقة الاخرى في الإأنموذج . او عند استخدام المتغيرات المرتدة زمنيا (متغيرات التباطؤ الزمني) كمتغيرات تفسيرية في الإأنموذج.

تؤدي مشكلة التعدد الخطي إلى تقديرات متحيزة وغير دقيقة لمعاملات الانحدار ، وتضخّم أخطاء معيارية لمعاملات الانحدار وتقلل من اختبارات t الجزئية لمعاملات الانحدار ، فضلاً عن التقليل من قابلية تنبؤ الأأنموذج .[2]

قد 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 ما يؤدي الى تضخيم تباين المعلمات المقدرة ، كما يعجز الإأنموذج عن اظهار أثر المتغيرات المستقلة منفصلة في المتغير التابع نظراً لترابطها.

بعض طرائق الكشف عن التعدد الخطي :

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

1- اختبار Farrar and Glauber (1967) : يعتمد الاختبار على قيمة محسوبة لمربع كاي Chi square بالاعتماد على محدد مصفوفة معامل الارتباط |R| وعدد المتغيرات المستقلة (p) وعدد الوحدات التجريبية (n) وبدرجة حرية $\frac{n(n-1)}{2}$ وهي :

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

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

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

VIF > 5

فانه مؤشر على وجود مشكلة التعدد الخطي .

3- 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) 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 0.5 فانه دليل على وجود ارتباط ضعيف.

ب- اذا كان العدد الشرطي بحدود $10 \leq CI \leq 30$ وان اثنين او اكثر من نسب تحليل التباين اكبر من 0.5 فانه دليل على وجود ارتباط معتدل.

ت- اذا كان العدد الشرطي بحدود $CI \geq 30$ وان اثنين او اكثر من نسب تحليل التباين اكبر من 0.5 فانه دليل على وجود ارتباط خطي على درجة كبيرة.[3][17]

معالجة مشكلة التعدد الخطي

من الطرائق الشائعة للتغلب على مشكلة التعدد الخطي هي :

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

2- 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.

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

4- 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. [16]

انحدار الحرف الاعتيادية (المتحيزة) Ordinary Ridge Regression (Biased)

تتميز طريقة انحدار الحرف بإيجاد قيمة ثابتة k تدعى بمعلمة التحيز Biasing Parameter , وهي كمية موجبة صغيرة تضاف الى عناصر قطر مصفوفة المعلومات $X'X$ ، وفائدة ذلك هو تقليل قيم عناصر قطر معكوس مصفوفة المعلومات الذي يؤدي الى خفض قيم تباينات المعلمات المقدرة . عند ابتعاد المتغيرات المستقلة عن الاستقلالية اي عند ارتفاع قوة الارتباط بين ازواج المتغيرات المستقلة، فإن اضافة الثابت k بقيم صغيرة تعمل على تغيير سريع في قيم المعاملات المقدرة ، ومع زيادة قيمة k تبدأ تلك القيم بالأستقرار تدريجياً الى ان تصل الى حد يكون التغير فيها طفيفاً وثابت الاشارة ، وكلما كان استقرار المعلمات سريعاً دل على ان المتغيرات المستقلة قريبة من الاستقلالية .

ان مقدرات الحرف الاعتيادية تعطي تقديرات للمعلمات تقلل مجموع مربعات الخطأ الى جانب تحقيق القيد $\sum_{j=1}^{p} {\beta_j}^2 \leq s$ كالاتي:

$$Min \quad \sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \quad \text{Subject to:} \quad \sum_{j=1}^{p} {\beta_j}^2 \leq s \quad ...(5)$$

وباستعمال مضاعف لاكرانج *Lagrange Multiple* يمكن ان نكتب المعادلة السابقة كالاتي :

$$Min \sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + k(\sum_{j=1}^{p} {\beta_j}^2 - s)$$

اذ ان s ثابت ليس له اي تاثير على الحل , فيمكن كتابة المعادلة بالشكل الاتي :

$$Min \sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + k\sum_{j=1}^{p}\beta_j{}^2$$
...(6)

اذا ان المعادلة أنفاً مكونة من جزأين هما :

$\sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2$ الجزء الخاص بمربع الخطأ *Sum Square Error term(SSR)*

$k\sum_{j=1}^{p}\beta_j{}^2$ الجزء الخاص بالجزاء(العقوبة) *Penalty Term*

اي ان :

$$Min \sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \mathrm{k}\sum_{j=1}^{p}\beta_j{}^2$$
$$= RSS + k\sum_{j=1}^{p}\beta_j{}^2$$
…(7)

ويتم الحصول على القيم المقدرة لمعلمات الأنموذج بطريقة انحدار الحرف عن طريق تصغير مجموع مربعات الاخطاء الجزائية الاتية :

$$(Y - XB)'(Y - XB) \qquad +k \qquad \| \beta \|_2^2$$
…(8)

وذلك باشتقاق المعادلة (8) بالنسبة الى β ونساوي المشتقة بالصفر ينتج :

$$-2X'(Y - XB) + 2k\beta = 0$$
$$-2X'Y - 2X'X\beta + 2k\beta = 0$$
$$-2X'Y - 2\beta(X'X + kI_P) = 0$$
$$X'Y = \beta(X'X + kI_P)$$
$$\hat{\beta}_{\mathrm{R}} = \left(\mathrm{X'X} + k\mathrm{I_p}\right)^{-1}\mathrm{X'Y}$$
…(9)

إذ ان:

$\hat{\beta}_{\mathrm{R}}$: متجه المعلمات المقدرة بأسلوب انحدار الحرف

K : معلمة التحيز (معلمة الحرف)

وتتمثل العلاقة بين مقدرات الحرف والمربعات الصغرى الاعتيادية بالاتي:

$$\hat{\beta}_{\mathrm{R}} = \left[\mathrm{I_p} + \mathrm{kI_p}(\mathrm{X'X})^{-1}\right]^{-1}\hat{\beta}_{\mathrm{ols}}$$
$$= (\mathrm{Z_R})\hat{\beta}_{\mathrm{ols}} \qquad \ldots(10)$$

إذ أن :

$$Z_R = \left[I_p + kI_p(X'X)^{-1}\right]^{-1}$$

من معادلة (10) يتبين ان مقدرات انحدار الحرف هي تحويل خطي لمقدرات المربعات الصغرى ، وعند اخذ التوقع لمقدرات انحدار الحرف يتضح بأنها مقدرات متحيزة لـ β :

$$E(\hat{\beta}_R) = E(Z_R\hat{\beta}_{ols}) = Z_R\hat{\beta}_{ols} \quad \ldots(11)$$

ان قيمة متوسط مربعات الخطأ لمقدرات انحدار الحرف هي :

$$MSE_R = \text{variance}\,(\hat{\beta}_R) + (\text{bias in}\,\hat{\beta}_R)^2 \quad \ldots(12)$$

عند زيادة قيمة k فأن مقدار التحيز يزداد والتباين يقل ، ولذلك يجب اختيار قيمة K باذ يكون الانخفاض في قيمة التباين اكثر من الارتفاع في مقدار مربع التحيز ، عند ذلك يكون متوسط مربعات الخطأ لانحدار الحرف أقل من التباين لمقدرات المربعات الصغرى الاعتيادية . ان زيادة قيمة k تؤدي الى انخفاض قيمة معامل التحديد R^2 ، من هنا يتضح بأن مقدرات انحدار الحرف ليس من الضروري ان تعطي افضل مطابقة للبيانات ، اذ اننا نبحث عن افضل معادلة ذات مقدرات ثابتة . [7]

طرائق اختيار قيمة معلمة التحيز Choosing Biasing parameter هنالك عدة طرائق لاختيار معلمة التحيز هي :

1- طريقة Hoerl et al (1975) [3] :

$$k = (m\hat{\sigma}^2)/(\hat{\beta}'_{ols}\,\hat{\beta}_{ols}) \quad \ldots(13)$$

وقد اشار المشهداني (1994) الى تحوير في الطريقة (13) لتصبح كالاتي :

$$k = [(m-2)\hat{\sigma}^2]/(\hat{\beta}'_{ols}\,\hat{\beta}_{ols}) \quad \ldots(14)$$

m : عدد المتغيرات التجريبية .

$\hat{\sigma}^2$: تباين المجتمع المقدر بطريقة المربعات الصغرى من البيانات الأصلية .

$\hat{\beta}_{ols}$: المعلمات المقدرة بطريقة المربعات الصغرى من البيانات الاصلية .

2- وضع Hoerl and Kennard (1976) [12] طريقة تكرارية بالأعتماد على المعادلة (8) التي ستعطى اول قيمة لـ k (لنرمز لها بـ k_0) والتي بواسطتها يتم ايجاد مقدرات لمعلمات انحدار الحرف بتطبيق المعادلة (9) ، ومن ثم ايجاد قيمة جديدة لـk_{p+1} (اذ أن : $p = 0,1,2,\ldots$) وذلك بتطبيق المعادلة (12) أدناه ، ثم نعود مرة اخرى لتطبيق المعادلة (9) ، وهكذا نستمر بالعملية التكرارية حتى تتحقق المقارنة .

$$k_{p+1} = (m\hat{\sigma}^2)/[\hat{\beta}'_{R(k)}\,\hat{\beta}_{R(k)}] \quad \ldots(15)$$

$$(k_{p+1} - k_p)/k_p \le 20(T^{-1.3}) \quad \ldots(16)$$

اذ ان :

$$T = \text{tr}(X'X)^{-1}/m$$

3- **أثر الحرف Ridge trace** : هو مخطط يحتوي على m من المنحنيات تمثل مسار المعلمات المقدرة عند كل قيمة من قيم k_p ، المحور العمودي يمثل قيم معلمة التحيز k_p ، والمحور الأفقي يمثل قيم المعلمات المقدرة β_{jp} ، يتم اختيار قيمة k التي عندها تبدأ المنحنيات بالاستقرار .[7]

ولوحظ تأثير سلبي في وضوح مسار منحنيات أثر الحرف تسببه زيادة في عاملين ، هما:

1- عدد المتغيرات .
2- الاختلاف في قيم المعلمات المقدرة المقابلة لقيم معلمة التحيز.

لذا اقترح وضع المتغيرات ذات قيم مطلقة متقاربة للمعلمات المقدرة في مجاميع مختلفة ، كل مجموعة تعطي مخططاً لأثر الحرف، ومن كل مخطط تعين قيمة لمعلمة التحيز ، وبذلك نحصل على مدى لقيم معلمة التحيز .

4- **تدوير محور أثر الحرف** : يعطي أثر الحرف عدد m من المنحنيات بمسارات مستقلة عن بعضها . لنفترض أن المنحنيات وضعت في اسطوانة ، ووزعت المنحنيات بزوايا متساوية على القاعدة ، ويرتفع كل منحنى بمسار تحدده قيم معلماته المقدرة المقابلة لقيم معلمة التحيز k_p . ويحدد ارتفاع الاسطوانة بين أكبر وأصغر قيمة لـk_p الواقعة على المركز.

عند النظر الى المنحنيات من أعلى الاسطوانة نلحظ اختلاف في مسارها ، وقد يتجاوز بعضها مركز الاسطوانة الى الجانب الأخر . وعند النظر من الجانب ، نلحظ تغير في مسار بعض المنحنيات في مرحلة ما بعد القيم الصغيرة لمعلمة التحيز محدثة بذلك تقاطع المنحنيات في مواقع عدة ، وحتى تعين حدود التقاطعات وجب النظر الى الاسطوانة من الجانب مع تدويرها ، ان القيمة المثالية لمعلمة التحيز تكون قريبة من مركز مدى التقاطعات.

ان الطريقتين (3) و (4) تعطيان مدى لقيمة معلمة التحيز , لذا نطبق الاسلوب الاتي لغرض تعيين قيمة معلمة التحيز:

عند كل قيمة من قيم مدى معلمة التحيز يطبق تحليل الانحدار , وتحدد اصغر قيمة k عندما يتم التغلب على مشكلة التعدد الخطي , وان قيمة k المثالية هي تلك القيمة التي تبعد مقاييس التعرف على مشكلة التعدد الخطي عن حدودها العليا دون احداث تأثير يذكر في مقاييس المفاضلة بين نتائج تحليل الانحدار مثل الدلالة الاحصائية لمعلمات الأنموذج و قيم R^2 و MSE .[17]

اختيار المتغيرات في معادلة انحدار الحرف التقديرية

هنالك طريقتان لاختيار أفضل المتغيرات لتكون في معادلة انحدار الحرف التقديرية ، هما :

1. أشار (Hoerl and Kennard (1970b الى استبعاد المتغير اذا اتصفت قيم معلماته المقدره المقابلة لقيم k_p بإحدى الصفات الاتية :

1- عدم استقرارها
2- انحدارها الى الصفر
3- ذات قيم قياسية صغيرة مقارنة بالمعلمات المقدرة للمتغيرات الاخرى.

ان هذا الاسلوب يعتمد على ملاحظة قيم المعلمات المقدرة لكل متغير بصورة منفردة ، وهذا بدوره يلغي العملية التعويضية لمتغير اخر او مجموعة من المتغيرات في اظهار تأثيرها بديلة عن المتغير المدروس .

2: أشار اليها Montgomery and Peck (1982) وهي الاستعانة بمخطط أثر الحرف ، وهنا تظهر سيطرة متغيرات ذات قيم عالية المعلمات المقدرة على المخطط ، ما يؤدي الى عدم اعطاء صورة واضحة عن مسار المنحنيات الأخرى .

للتغلب على القصور في الطريقتين المشار اليهما ، نقوم بإنشاء مجاميع أثر الحرف ، ومنها قد يتم اختيار بعض المتغيرات من كل مجموعة حسب الملاحظات التي أشار اليها (Hoerl and Kennard (1970b لاستبعاد المتغير.

عند نجاح انحدار الحرف في التغلب على تعدد العلاقة الخطية معناه ان التحويل الخطي لطريقة المربعات الصغرى الاعتيادية قد أفلح في تقريب البيانات الى التعامد ، وبذلك استوفيت الشروط للسماح بالتعامل بالمثل مع طريقة المربعات الصغرى.[1]

انحدار الحرف البيزية (الغير متحيزة) Expand Ridge Regression (unbiased)

استعملت طريقة انحدار الحرف المتحيزة التي تم شرحها في الكثير من الدراسات , اذ اثبتت هذه الطريقة فعاليتها في معالجة مشكلة التعدد الخطي الذي يؤدي الى كبر حجم تباين المقدرات ويُحدث ارباك في العلاقة السببية بين المتغيرات المستقلة والمتغير المعتمد في أنموذج الانحدار. ولكن في طريقة انحدار الحرف الغير متحيزة سيتم ادخال المعلومات السابقة Prior في إأنموذج انحدار الحرف للتخلص من مشكلة التحيز في تقدير المعلمات. [2][9]

عند تطبيق طريقة المربعات الصغرى OLS لبيانات غير متعامدة فأننا سوف نحصل على مقدرات غير مستقرة وتمتلك مجموع مربعات خطأ MSE كبير. فللتعامل مع مشكلة التعدد الخطي, تستعمل طريقة انحدار الحرف الاعتيادية (المتحيزة) والتي طورت لدى ((Horel and Kennard(1970 a,b) . اذ تتلخص هذه الطريقة بإضافة الثابت K إلى مصفوفة $(X'X)$ قبل اخذ المعكوس لها وكما يأتي:

$$\widehat{B}(KI, J) = \hat{\beta}_R = (X'X + KI_P)^{-1}(X'Y) \quad \dots (17)$$

اقترح Swindel (1976) [9] أدخال المعلومات السابقة Prior Information في صيغة انحدار الحرف , بذالك سوف تكون صيغة انحدار الحرف البيزية كما يأتي:

$$\hat{\beta}_R = (X'X + KI_P)^{-1}(X'Y + KJ) \quad \dots (18)$$

اذ نلحظ من الصيغة (18)انه تم ادخال الثابت K مع المتجه J إلى المتجه X'Y والذي يتمثل بالمعلومات الاولية (Prior information).

ونلحظ من الصيغة (18) إنه في حالة $K = 0$ سوف نحصل على تقدير المربعات الصغرى الاعتيادية OLS.

وعن طريق مفهوم نظرية بيز المتمثل بالآتي:

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

نلحظ من الصيغة انفا 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(ليس معلوماتي) فان K تعطي ايضاً مقدرات تمتلك MSE اقل من MSE للمربعات الصغرى.

وكما معروف عن طريق فرضيات الإأنموذج الخطي, , ان الخطأ u_i يتوزع توزيع طبيعي بمعدل 0 , وتباين σ^2_{In}, أي أن $u_i \sim N(0, \sigma^2_{In})$, وفان تقديرات المربعات الصغرى تتوزع توزيعاً ً طبيعياً ً 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 :

$$B(C,J) = C\widehat{B}_{Ols} + (I - C)J$$
… (19)

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 .

نلحظ من الصيغة (19) أن المصفوفة C مصفوفة مجهولة , فلإيجاد هذه المصفوفة فأننا سوف نجد متوسط مجموع مربعات الخطأ للتقدير المحدب والذي يكون بالصيغة الاتية[5] [15][18]:

$$\widehat{B}_{Bays} = (X'X + \sigma^2 V_0^{-1})^{-1} (X'Y + \sigma^2 V_0^{-1} B_0) \quad \ldots.(20)$$

الجانب التطبيقي Applied Side

وصف البيانات Data Description

تضمنت بيانات البحث المعلومات المتوفرة في سجلات زيارات النساء الى المراكز الصحية التابعة لدائرة صحة بابل , اذ تم مراجعة المراكز الصحية الموجودة في مركز محافظة بابل وهي:

1- مركز صحة مرجان للرعاية الصحية الاولية

2 - مركز صحة الخالص للرعاية الاولية

3- مركز صحة شهداء نادر الأنموذجي للرعاية الصحية الاولية

4- مركز صحة الاساتذة الأنموذجي للرعاية الصحية الاولية

5- مركز صحة بابل التدريبي للرعاية الصحية الاولية , وكذلك من الاستمارات الاحصائية الخاصة برعاية الام والطفل الموجودة في قسم الصحة العامة للرعاية الصحية الاولية اذ سحبت عينة عشوائية بسيطة مؤلفة من 100 امراة لدراسة العوامل المؤثرة على عدد الاطفال المولودين وهومتغير كمي (الاحياء +المولودين ثم توفوا) (متغير الاستجابة) Y , بينما كانت المتغيرات المستقلة تتمل بالاتي :

X_1 : عمر المراة متغير كمي بالسنوات

X_2 : العمر عند الزواج متغير كمي بالسنوات

X_3 : التحصيل الدراسي للمرأة متغير وصفي من النوع الرتبي(1- امية 2- تقرأ وتكتب 3- ابتدائية 4- متوسطة-5- اعدادية 6- دبلوم 7-بكالوريوس 8- بدلوم عالي فما فوق)

X_4 : التحصيل الدراسي للزوج متغير وصفي من النوع الرتبي (1- امي 2- يقرأ ويكتب 3- ابتدائية4- متوسطة-5- اعدادية 6- دبلوم 7-بكالوريوس 8- بدلوم عالي فما فوق)

X_5 : وزن المرأة متغير كمي بالكغم

X_6 : استخدام المراة لوسائل منع الحمل متغير وصفي من النوع الاسمي (1- تستعمل 2- لاتستعمل)

X_7 : تدخين المرأة متغير وصفي من النوع الاسمي (1- تدخن 2- لاتدخن)

X_8 : عمر الزوج متغير كمي بالسنوات

X_9 : مهنة الزوج متغير وصفي من النوع الاسمي (1- كاسب 2- موظف 3- عسكري 4- استاذ جامعي)

X_{10} : مدة الزواج متغير كمي بالسنوات

X_{11} : عدد الاطفال المتوفين متغير كمي

X_{12} : عدد ساعات مارسة الرياضة في الاسبوع متغير كمي

X_{13} : اصابةالمرأة بأمراض الغدة الدرقية متغير وصفي من النوع الاسمي (1- مصابة 2- غير مصابة)

X_{14} : عدد ساعات نوم المراة في اليوم متغير كمي

X_{15} : تناول الادوية لدى المرأة متغير وصفي من النوع الاسمي (1- تناولت- 2- لم تتناول)

X_{16} : مدة الرضاعة الطبيعية متغير كمي بالاشهر

X_{17} : مهنة الام متغير وصفي من النوع الاسمي (1- ربة بيت 2- موظفة)

اختبار مشكلة التعدد الخطي :

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

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

H_0: لاتوجد مشكلة التعدد الخطي

H_1: توجد مشكلة التعدد الخطي

وعن طريق تطبيق الصيغة **(4)** وكالاتي:

ان 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	-0.78	-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 0.05 **n=100,** و **P=17** كانت **(164.216)** وهي اقل من القيمة المحسوبة لذلك نرفض فرضية العدم اي انه توجد مشكلة التعدد الخطي .

2- عامل تضخم التباين **VIF** اذ كانت نتائج هذا المقياس كالاتي:

جدول(1) 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

اذ ظهرت زيادة قيمة هذا المقياس عن 5 لكل من المتغيرات الاتية : (X1: عمر المرأة , X4: التحصيل الدراسي للزوج , X8:عمر الزوج , X10 :مدة الزواج)

3- مقياس المعيار الشرطي Condition Index : اذ اظهرت نتائج التحليل الآتي:

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 (مدة الزواج) قيمها تتراوح بين$10 \leq CI \leq 30$, بينما كانت المتغيرات x11 (عدد الاطفال المتوفين) و x12 (عدد ساعات مارسة الرياضة في الاسبوع) و x13 (اصابة المرأة بأمراض الغدة الدرقية), بينما كانت المتغيرات X14 (عدد ساعات نوم المرأة في اليوم) و X15 (تناول الادوية لدى المرأة) و X16 (مدة الرضاعة الطبيعية) و X17 (مهنة الام) اكبر من 30 .

وعن طريق المؤشرات السابقة تبين لنا وجود مشكلة التعدد الخطي بين المتغيرات المستقلة المشار اليها في جدول (1).

التقدير بطريقة المربعات الصغرى الاعتيادية OLS:

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

$$\hat{y}=b_0+b_1 x_1+b_2 x_2+b_3 x_3+b_4 x_4+b_5 x_5+b_6 x_6+b_7 x_7+b_8 x_8+b_9 x_9+b_{10} x_{10}+b_{11} x_{11}+b_{12} x_{12}+b_{13} x_{13}+b_{14} x_{14}+b_{15} x_{15}+b_{16} x_{16}+b_{17} x_{17} \quad \ldots(29)$$

وباستخدام برنامج ماتلاب لايجاد معادلة الانحدار التقديرية , كانت النتائج كما مبينة في الجدول (2)

جدول (2) تقديرات المربعات الصغرى الاعتيادية وقيمة t المحسوبة والقيمة الاحتمالية لطريقة المربعات الصغرى الاعتيادية OLS

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 (2) نجد ان المتغيرات (X_5 وزن المرأة , X_6 استخدام وسائل منع الحمل , X_7 تدخين المرأة , X_{10} مدة الزواج, X_{11} عدد الاطفال المتوفين , X_{13} اصابة المرأة بأمراض الغدة الدرقية , X_{15} تناول الادوية لدى المرأة و X_{16} مدة الرضاعة الطبيعية و X_{17} مهنة الام) كانت معنوية لانها تمتلك قيمة P-Value اقل من مستوى المعنوية 0.05 , في حين ان باقي المتغيرات (X_1 عمر المرأة , X_2 العمر 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 مارسة الرياضة وX_{14} عدد ساعات نوم المرأة) كانت غير معنوية بسبب ان قيمة P-Value اكبر من مستوى المعنوية 0.05 .

ويمكن كتابة معادلة الانحدار المقدرة بطريقة المربعات الصغرى الاعتيادية كالاتي :

$$\hat{y} = \mathbf{-0.61700\ x_1 + 0.0538\ x_2 - 0.11139\ x_3 - 0.0197\ x_4 - 0.0616\ x_5 + 2.5289\ x_6 + 1.0886\ x_7 - 0.0757\ x_8 - 0.6970\ x_9 + 0.3519\ x_{10} + 1.3074\ x_{11} - 0.0849\ x_{12} - 2.4201 x_{13} - 0.0814 x_{14} + 1.1597\ x_{15} + 0.1647\ x_{16} + 1.5553\ x_{17}} \ldots.(30)$$

ونلحظ من المعادلة (30) x_1 (عمر المرأة) غير معنوي لان القيمة الاحتمالية له التي تساوي (**0.4173**) اكبر من مستوى المعنوية 0.05 وهذه النتيجة لا توافق للمنطق فلا يمكن ان يكون العمر غير مؤثر على درجة خصوبة المرأة , وكذلك نلحظ 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 معامل انحدار المتغير **x10** البالغ (**0.3519**) اشارته الجبرية موجبة وهو (مدة الزواج) , وايضاً المتغير $\mathbf{x_{15}}$ (تناول الأدوية) البالغ (**1.1597**) اشارته موجبه فلايمكن 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 .

علماً ن جدول تحليل التباين وفق هذه الطريقة يكون كالاتي :

جدول(3) جدول تحليل التباين للأنموذج 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**	**R^2**	**0.83**		

من الجدول (3) نلحظ معنوية الأنموذج القيمة الاحتمالية .Sig (0.000) لاختبار F اقل من مستوى المعنوية 0.05.

التقدير بطريقة انحدار الحرف الاعتيادية(المتحيزة): Biased Ridge Regression

باستخدام برنامج ماتلاب لايجاد معادلة الانحدار التقديرية , كانت النتائج كما مبينة في الجدول

جدول(4) تقديرات معاملات الانحدار والخطأ المعياري وقيمة t المحسوبة والقيمة الاحتمالية لطريقة انحدار الحرف الاعتيادية(المتحيزة) عند معلمة الحرف K=0.09

Variable	Estimates	S.E.	T	P-Value
x1	**-0.1978**	**0.0192**	**-10.302083**	**0.00022**
x2	**-0.084**	**0.0273**	**-3.0769231**	**0.00198**
x3	**-0.105**	**0.0332**	**-3.1626506**	**0.00628**
x4	**-0.033**	**0.1427**	**-0.2312544**	**0.88696**
x5	**-0.061**	**0.0194**	**-3.1443299**	**0.00764**
x6	**-2.575**	**0.619**	**-4.1599354**	**0.00058**
x7	**1.1114**	**0.431**	**2.57865429**	**0.01021**
x8	**-0.081**	**0.0366**	**-2.2131148**	**0.00391**
x9	**-0.727**	**0.4199**	**-1.7313646**	**0.06731**
x10	**0.2065**	**0.0581**	**3.55421687**	**0.0067**
x11	**1.2415**	**0.3064**	**4.05189295**	**0.0008**
x12	**-0.079**	**0.0523**	**-1.5105163**	**0. 9846**
x13	**-2.574**	**0.4734**	**-5.4372624**	**0.00046**
x14	**-0.179**	**0.0786**	**-2.2773537**	**0.00295**
x15	**1.1746**	**0.4953**	**2.37149203**	**0.00296**
x16	**0.1687**	**0.0647**	**2.60741886**	**0.00246**
x17	**1.6545**	**0.4486**	**3.68814088**	**0.00156**

ومن ملاحظة الجدول (4) نجد ان المتغيرات (X_1 عمر المرأة ,X_2 العمر عند الزواج ,X_3 التحصيل الدراسي للمرأة , X_5 وزن المرأة ,X_6 استخدام وسائل منع الحمل , X_7 تدخين المرأة, X_8 عمر الزوج بالسنوات, X_9 مهنة الزوج , X_{10}مدة الزواج , X_{11} عدد الاطفال المتوفين ,X_{13} اصابة المراة بامراض الغدة الدرقية , X_{14} عدد ساعات نوم المرأة , X_{15} تناول الادوية لدى المراة , X_{16} مدة الرضاعة الطبيعية و X_{17} مهنة الام) كانت معنوية لان قيمة P-Value لها اقل من مستوى المعنوية 0.05 , في حين ان باقي المتغيرات (X_4 التحصيل الدراسي للزوج , X_{12} عدد ساعات مارسة الرياضة) كانت غير معنوية بسبب ان قيمة P-Value اكبر من مستوى المعنوية 0.05 . ونلحظ ايضا ان بعض معاملات الانحدار ظهرت باشارة جبرية غير صحيحة كالمتغير x_{15} (تناول الادوية لدى المرأة) ما يدل على ان طريقة انحدار الحرف الاعتيادية لم تعالج مشكلة التعدد الخطي بشكل كامل .

وبالتالي فان معادلة الانحدار المقدرة بطريقة انحدار الحرف المتحيزة كالاتي :

$$\hat{y}= -0.1978\ x_1 -0.084x_2 -0.105x_3 -0.033x_4 -0.061x_5 -2.575x_6 + 1.1114x_7 -0.081x_8 -0.727x_9+ 0.2065x_{10} + 1.2415x_{11} -0.079x_{12} -2.574x_{13} -0.179x_{14}+1.1746x_{15}+0.1687x_{16}+1.6545x_{17}\ldots..(31)$$

وان جدول تحليل التباين لطريقة انحدار الحرف المتحيزة كان كالاتي:

جدول(5) جدول تحليل التباين للأنموذج المقدر بطريقة انحدار الحرف المتحيزة

S.O.V	D.f.	SSE	MSE	F	Sig.
Model	17	734.189	43.1876	23.875	0.0000
Error	82	148.321	1.8088		
Total	99	882.51			
MSE	1.8088	R^2	0.82		

من الجدول (5) نلحظ معنوية الأنموذج القيمة الاحتمالية .Sig (0.000) لاختبار F اقل من مستوى المعنوية 0.05 , ونلحظ ايضاً ان قيمة متوسط مربعات الخطأ البالغة (1.8088) قد انخفضت عن قيمة مربعات الخطأ لطريقة المربعات الصغرى البالغة (1.8868) .

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

لايجاد تقديرات معاملات الانحدار بطريقة انحدار الحرف البيزية سنعد بان عناصر منتجه المعلومات الاولية J تمثل معدل تقديرات معلمات المربعات الصغرى الاعتيادية كالاتي(1976)Swindle [9]:

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 (6) :

جدول (6) 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 (6) نجد ان المتغيرات (X_1 عمر المرأة , X_2 العمر عند الزواج , X_3 التحصيل الدراسي للمرأة , X_5 وزن المرأة , X_6 استخدام وسائل منع الحمل , X_8 عمر الزوج بالسنوات, X_{10} مدة الزواج, X_{11} عدد الاطفال المتوفين X_{13}, اصابة المراة بامراض الغدة الدرقية , X_{14} عدد ساعات نوم المرأة و X_{15} تناول الادوية لدى 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_9 مهنة الزوج , X_{12} عدد ساعات مارسة الرياضة, X_{16} مدة الرضاعة الطبيعية و X_{17} مهنة الام) كانت غير معنوية بسبب ان قيمة P-Value اكبر من مستوى المعنوية 0.05 .

وبالتالي فان معادلة الانحدار المقدرة بطريقة انحدار الحرف البيزية كالاتي :

$$\hat{y} = -0.2456x_1 -0.0854x_2 -0.135x_3 -0.015x_4 -0.0751x_5 -2.719x_6 -0.0034x_7 -0.1134x_8 -0.0559x_9 - 0.1823x_{10} -0.9935x_{11} -0.0336x_{12} -2.089x_{13} -0.194x_{14} -1.2245x_{15} -0.0008x_{16} -0.0001x_{17} \quad \ldots.(32)$$

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

جدول (7) جدول تحليل التباين للأنموذج 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 (7) نلحظ معنوية الأنموذج القيمة الاحتمالية .Sig (0.000) لاختبار F اقل من مستوى المعنوية 0.05 , ونلحظ ايضاً ان قيمة متوسط مربعات الخطأ البالغ (1.7424) قد انخفضت عن قيمة متوسط مربعات الخطأ لطريقة المربعات الصغرى البالغة (1.8868) وطريقة انحدار الحرف المتحيزة البالغ (1.8088) , اي ان طريقة انحدار الحرف البيزية عالجت مشكلة التعدد الخطي مقارنة بطريقتي المربعات الصغرى الاعتيادية وطريقة انحدار الحرف الاعتيادية . ونلحظ 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 وX_9 مهنة الزوج و X12 عدد ساعات مارسة الرياضة في الاسبوع و X16 مدة الرضاعة الطبيعية و X17 مهنة الام اذ نلحظ ان تاثيراتها قليلة جداً كما هو مبين من قيمها (**-0.015, -0.0034, -0.0559 , -0.0336, -0.0008, -0.0001**) على الترتيب مقارنة ببقية معاملات الانحدار.

مقارنة بين انحدار الحرف المتحيزة وانحدار الحرف غير المتحيزة

الجدول الاتي يبين مقارنة بين معاملات الانحدارالمقدرة بطريقة انحدار الحرف المتحيزة وانحدار الحرف غير المتحيزة :

جدول (8) مقارنة بين معاملات الانحدار المقدرة بطريقة المربعات الصغرى وانحدار الحرف الاعتيادية وانحدار الحرف البيزية

Method	OLS	Ridge Biased	Ridge Unbiased
x1	-0.6170	-0.1978	-0.2456
x2	0.0538	-0.084	-0.0854
x3	-0.1139	-0.105	-0.135
x4	-0.0197	-0.033	-0.015
x5	-0.0616	-0.061	-0.0751
x6	2.5289	-2.575	-2.719
x7	1.0886	1.1114	-0.0034
x8	-0.0757	-0.081	-0.1134
x9	-0.6970	-0.727	-0.0559
x10	0.3519	0.2065	-0.1823
x11	1.3074	1.2415	-0.9935
x12	-0.0849	-0.079	-0.0336
x13	-2.4201	-2.574	-2.089
x14	-0.0814	-0.179	-0.194
x15	1.1597	1.1746	-1.2245
x16	0.1643	0.1687	-0.0008
x17	1.5553	1.6545	-0.0001

جدول (9) الخطأ المعياري لمعاملات الانحدار المقدرة بطريقة المربعات الصغرى وانحدار الحرف الاعتيادية وانحدار الحرف البيزية

Method	OLS	Ridge Biased	Ridge Unbiased
x1	0.7569	0.0192	0.0181
x2	0.0813	0.0273	0.0189
x3	0.1434	0.0332	0.0311
x4	0.1937	0.1427	0.1422
x5	0.0260	0.0194	0.0156
x6	0.8210	0.619	0.6048
x7	0.4380	0.431	0.3433
x8	0.0613	0.0366	0.0265
x9	0.4174	0.4199	0.3234
x10	0.0624	0.0581	0.045
x11	0.3153	0.3064	0.1064
x12	0.0605	0.0523	0.023
x13	0.6300	0.4734	0.3228
x14	0.2542	0.0786	0.1055
x15	0.5277	0.4953	0.4043
x16	0.0652	0.0647	0.0612
x17	0.5811	0.4486	0.3985

ونلحظ من الجدول (9) ان الاخطاء المعيارية لطريقة انحدار الحرف البيزية كانت اقل من الاخطاء المعيارية للطرائق الباقية.

الجدول (10) متوسط مربعات الخطأ والقيمة الاحتمالية لطريقة المربعات الصغرى الاعتيادية وانحدار الحرف الاعتيادية وانحدار الحرف البيزية في التقدير

Method	MSE	P-Value
O.L.S	1.8868	0.000
Biased Ridge	1.8088	0.000
Unbiased Ridge	1.7424	0.000

الاستنتاجات : Conclusions

عن طريق ماتم التوصل اليه من نتائج في الجانب التطبيقي من مقارنة بين النماذج الثلاثة توصل الباحث الى الاستنتاجات الاتية :

1- ظهور عدم معنوية بعض المتغيرات المستقلة في معادلة الانحدار المقدرة رغم اهميتها بسبب وجود مشكلة التعدد الخطي بين المتغيرات المستقلة.
2- لم تعالج طريقة انحدار الحرف المتحيزة مشكلة التعدد الخطي بشكل كامل بينما طريقة انحدار الحرف البيزية عالجتها بشكل كامل .
3- افضلية طريقة انحدار الحرف البيزي على طريقة انحدار الحرف المتحيزة.

التوصيات : Recommendations

عن طريق ما تم عرضه من استنتاجات نوصي بالاتي :

1- التوسع في دراسة طريقة انحدار الحرف البيزي ومقارنته مع الطرائق الحديثة في تقدير النماذج الاحصائية مثل انحدار لاسو والتي تلعب دور فعال في تحديد المتغيرات المستقلة الاكثر اهمية للأنموذج.
2- التوسع في اجراء دراسات وبحوث لمقارنة طرائق الحرف البيزية مع طرائق اخرى مثل المركبات الرئيسة .

المصادر References

1. النعيمي , اسوان محمد طيب ; دبدوب , مروان عبد العزيز, (2006) , " طرائق مقترحة في انحدار الحرف " , المجلة العراقية للعلوم الاحصائية , (10), 85-106.
2. عبودي , عماد حازم , علي , حميد يوسف , (2017) , " مقارنة مقدري Hurber Lasso و Hurber Elastic Net باستخدام المحاكاة " , مجلة الكوت للعلوم الاقتصادية والادارية , العدد (28) , الجزء الاول , بحث مستل من اطروحة دكتوراه .
3. مزاحم, محمد يحيى , (2005) , " استخدام المكونات الرئيسية وانحدار الحرف في تقدير معادلة السعر العالمي للقمح للمدة من 1961-2002) , مجلة تكريت للعلوم الادارية والاقتصادية , الجلد -1 , العدد-1. , 146-156.

4. ALKHAMISI M. A. , SHUKUR G., (2007) , " A Monte Carlo Study of Recent Ridge Parameters " , Communications in Statistics—Simulation and Computation, Taylor & Francis, 36: 535–547.
5. B.F.,swindle , (1976).Good Ridge Estimators Based on prior Information. Communications in Statistis,11,1065-1075.
6. Bare, B. Bruce; Hann David W., (1981). "(Applications of Ridge Regression in Forestry ", Forest $ci. Vol. 27, No. 2, 1981, pp. 339-348.
7. C. Montgomery, Douglas; A. Peck, Elizabeth; Vinining G. Geoffery , (2012). "Introduction to linear regression analysis" , Fifth Edition, A JOHN WILEY & SONS, INC., PUBLICATION1 .
8. Duzan , Hanan; Shariff, Nurul Sima Mohammed, (2015) , " Ridge Regression for solving the Multicollinearity Problem: review of Methods and Models" , Journal of Applied Statistics, 15(3) : 392-404.
9. F. Benee,Swindle , (2013) , " Good ridge estimators based on prior information", Moskow State Univ Bibliote, Taylor & Francis.
10. Fonti Valeria, (2017) , " Feature Selection using LASSO" , Research Paper in Business Analytics, VU Amsterdam .
11. Hoerl, Arthur .E. and Kennard, Robert W., (1970a)."Ridge regression: Biased estimation for non-orthogonal Problems", T Econometrics Journal , Vol.12, No.1,55-67.
12. Hoerl, Arthur .E. and Kennard, Robert W., (1970b)."Ridge regression: Applications to Non-Orthogonal Problems, ", T Econometrics Journal, Vol.12, No.1,69-82.
13. Kannard , Arthur E. Boart W.; F. Baldwin, Kent, (1975), " Ridge Regression , Some solutions " , Communications in statistics , 4(2), 105-123.
14. Kibria, B. M. Golam , (2003) , " (Performance of Some New Ridge Regression Estimators" , Communications in , Simulation and Computation. Vol. 32, No. 2, pp. 419–435.
15. L.J, Pliskin , (1987) "A Ridge type Estimator and good prior Means. Communication in Statistics, 16,3427-3429.
16. 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.
17. Mason, R.L.F. Gunst; J.T. Wabster, (1975), ' Regression analysis and problems of multicoleanirutiy " , comM. sTATST,4,3,277.
18. Robert H. Crouse & Chun Jin.(1995)."Unbiased RIDGE Estimation With prior Information And Ridge Trace"Commun.Statist.-Theory Meth.27(9),2341-2354.
19. Tibshirani , Ryan , (2013) , " Modern regression " , Optional reading: ISL 6.2.1, ESL 3.4.1, Data Mining: 36-462/36-662.
20. W. Wichern , Dean ; A. Churchill ,Gilbert , (2012) ."A Comparison of Ridge Estimators", T Econometrics Journal, 301-311.