جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية - كلية الادارة والاقتصاد
قسم الاحصاء

تقدير اقل معدل تباين صفري حصين

مع التطبيـــــــق

رسالة مقدمة الى

مجلس كلية الادارة والاقتصاد في جامعة القادسية وهي جزء من متطلبات نيل درجة ماجستير في علوم الإحصاء

قدمتها الطالبة

سناء جبار طعمه

بإشراف

أ. د . علي جواد كاظم

1444هـ

2023م

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

﴿وَإِذَا سَأَلَكَ عِبَادِي عَنِّي فَإِنِّي قَرِيبٌ أُجِيبُ دَعْوَةَ الدَّاعِ إِذَا دَعَانِ
فَلْيَسْتَجِيبُوا لِي وَلْيُؤْمِنُوا بِي لَعَلَّهُمْ يَرْشُدُونَ ﴾

صدق الله العلي العظيم

سورة البقرة:186

الإهـــــــــداء

إلى... من كان أملي رضاه وغايتي حبه ورجائي غفرانه الله رب العالمين.

إلى.. ذي الخلق العظيم نبينا محمد وآله الطيبين الطاهرين وصحبه ﷺ.

إلى... من كللهما الله بالهيبة والوقار إلى من علماني العطاء دون انتظار إلى من احمل اسميهما بكل فخر وأعتزاز.

الى الرجل الفاضل الذي تعب من أجل وصولي الى ما أنا عليه الان، أبي الغالي

(أطال الله في عمره)

الى أعظم امرأة كان دعاؤها سر نجاحي وحنانها بلسم جراحي الى امي الغالية

(أطال الله في عمرها)

الى من ترعرعت معهم، ونما غصني بينهم ...أخواني واخواتي .

إلى... أخي الشهيد(حيدر جبار طعمة)وأخي (جعفر جبار طعمة) لروحهم المغفره والسلوان.

إلى من شجعني على مواصلة مسيرتي العلمية رفيق دربي زوجي الغالي واهله ...

الى زينة الحياة ابنائي.....بنين وحنين واحمد ومحمد .

الى الاستاذ الفاضل الذي ساندني وبذل جهده معي أ.د. علي جواد كاظم

إلى... جميع من كانت له يد العون والمساندة وكل من تمنى لي النجاح

أهدي ثمرة جهدي المتواضع ولهم مني جزيل الشكر والأمتنان.

الباحث

الشكر والأمتنــان

الحمد والشكر لله ملء السموات والارض وما بينهما، الذي أعانني بفضله ونعمته على إتمام هذا البحث المتواضع، وأفضل الصلاة والسلام على سيد المرسلين نبينا وحبيبنا محمد وآله الطيبين الطاهرين الذين ما خاب من تمسك بهم وأمن من لجأ اليهم .
أما بعد حمد الله وشكره على نعمائه، فبالشكر تدوم النعم أجد نفسي مدينة لكثير من الخيرين في إتمام هذه الرسالة، فبعضهم من أبدى لي الدعم بقلبه ودعائه، وبعضهم بلسانه بالتشجيع والتأييد، فجزاهم الله عني خير الجزاء.

يدفعني واجب الوفاء والاحترام والتقدير الى تقديم جزيل الشكر والامتنان العميق الى الأستاذ الفاضل الدكتور (علي جواد كاظم) لتفضلة بالإشراف على هذه الرسالة ولما بذله من جهود قيمة وتوجيهات بناءة وآراء صائبة وملاحظات مفيدة طيلة فترة إعداد الرسالة بغية إخراجها بالشكل الذي عليه الآن، لقد منحني وقته الثمين رغم انشغالاته الكثيرة... فجزاه الله خير الجزاء.

كما اتقدم بالشكر وفائق الاحترام الى اساتذتي الافاضل في قسم الاحصاء (أ.د.محمد حبيب الشاروط، أ.د طاهرريسان دخيل، أ.د.أحمد نعيم، أ.د.مهند السعدون ، أ.م.د. حسن سامي عريبي، د. فاضل حميد الحسيني، أ.م.د. طه حسين علي، أ.م.د. بحر كاظم ، أ.م. سيف حسام، م.د. اسعد ناصر، م.عفراء عباس، م.م.حميدة نعيم) لما قدموه من جهود علمية متميزة خلال فترة الدراسة فجزاهم الله خير الجزاء..

كما أتوجه بالشكر والأمتنان الى السيد رئيس وأعضاء لجنة المناقشة الموقرة التي تكرمت علينا بموافقتها لمناقشة هذة الرسالة وما تحملوه من عناء المراجعة والتقويم على الرغم من مشاغلهم، وكل ملاحظاتهم تعد هدايا تزين رسالتي نحو الاجمل.

واقدم شكري واحترامي الى رفقاء مرحلة الماجستير كافة داعية الله لهم بالتوفيق والنجاح والمزيد من التقدم.

واخيرا اتقدم بشكري وتقديري واعتذاري لكل الذين فاتني ذكرهم وأسال الله لهم دوام الموفقية والنجاح..

المستخلص Abstract

ان زيادة عدد المتغيرات يؤدي الى زيادة تعقيد النموذج وقد يقود ذلك الى مشكلة تعدد الابعاد (Curse dimensionality)، هذه المشكلة قادت الباحثين للعمل على تقليص هذه الابعاد العالية للبيانات. ان بعض المتغيرات التوضيحية ليس لها تأثير معنوي على المتغير المعتمد وكذلك بعض هذه المتغيرات لها ارتباط داخلي فيما بينها وهذا يتطلب استبعاد مثل هذه المتغيرات من اجل زيادة دقة النموذج، وهنالك طريقتان لتقليص الابعاد هما طريقة اختيار المتغيرات(Variables Selection)(V.S)وطريقة استخلاص المتغيرات(Variables extractions). تحت افتراضات نظرية SDR(Sufficient dimension reduction)عمل الباحثون على اقتراح طرائق لتقليل الابعاد ومنها دمج طرائق SDR مع طرائق التنظيم ((Regularization method))وطرائق التنظيم تعني إضافة حد جزاء للتحكم في تعقيد النموذج اذ يقلل بشكل كبير من تباين النموذج، ومن هذه الطرائق (SMAVE-AdEN) (AlkenaniوRahman عام (2020)وهي طريقة لاختيار متغير تحت افتراضات نظرية SDR.طريقة SMAVE-AdENعبارة عن مزيج من الشبكة المرنة المتكيفة(Adaptive elastic net)(AdEN) مع طريقة تقليل الأبعاد الفعالة (Minimum average variance estimator)(MAVE) لتقدير متوسط التباين الأدنى.

تكون هذه الطريقة فعالة عندما تكون الارتباطات عالية بين المتغيرات. لكن طريقة SMAVE-AdEN ليست حصينة وهي طريقة حساسة تتأثرعند وجود قيم شاذة في البيانات، لأنها تستخدم معيار المربعات الصغرى .

اقترحنا هنا طريقة اختيار متغير حصينة تحت أفتراضات SDRتدعى (RSMAVE-AdEN). لا تتأثر بالقيم الشاذة الموجودة في كل من المتغيرات التفسيرية والاستجابة.

تم التحقق من كفاءة الطريقة المقترحة من خلال المحاكاة واستخدام البيانات الحقيقية.

قائمة المحتويات

قائمة الجداول Tables

قائمة **الاشكال Shapes**

قائمة المختصرات

الرمز المختصر	المصطلح العلمي باللغة الانكليزية	المصطلح العلمي باللغة العربية
AdEN	Adaptive elastic net	الشبكة المرنة التكيفية (المكيفة)
Ad LASSO	Adaptive least absolute shrinkage and selection operator	اقل تقليص مطلق ومعامل الاختيار المكيف
AIC	Akaike information criterion	معيارمعلومات اكايكي
ALMAVE	Sparse MAVE with adaptive Lasso penalty	مقدر أقل معدل تباين مع حد جزاء لاسو التكيفي
Ave. 0's	Average number of zero coefficients	معدل عدد المعاملات المصفرة
BIC	Bayesian information criterion	معيارمعلومات بيز
CD	Curse of dimensionality	مشكلة لعنه الابعاد
CMF	Conditional mean function	دالة المتوسط الشرطي
CMS	Central mean subspace	الفضاء الجزئي للمتوسط المركزي
DR	Dimension reduction	تقليل(اختزال) الابعاد
DRS	Dimension reduction subspace	فضاء تقليل الابعاد
EN	Elastic net	الشبكة المرنة
GR	Graphical regression	الانحدار الرسومي
GCV	Generalized Cross validation	معيار العبور(التقاطع) الشرعي العام
GSR	Graduate student rates	معدلات تخرج الطلاب
HD	High dimension	الابعاد العالية
IHT	Iterative Hessian Transformation	تحويل هيسين التكراري
i.i.d	independent identically distributed	مستقلة بشكل مماثل
Lasso	Least absolute shrinkage and selection operator	اقل تقليص مطلق ومعامل اختيار(لاسو)
LC's	Linear combinations	تركيبات خطية
LS	Least squared	المربع الاصغر
MAVE	Minimum average variance estimator	مقدر اقل معدل تباين
MCP	Minimax concave penalty	دالة جزاء أقل اعضم المقعرة
MCP-SMAVE	Spars MAVE with MCP penalty	مقدر اقل معدل تباين الصفري مع دالة جزاء أقل اعضم المقعرة
ME	Model selection	اختيار النموذج
MSE	Mean squared error	متوسط مربعات الخطأ
MLE	Maximum Likelihood estimator	مقدر الامكان الاعظم
OLS	Ordinary least squared	المربعات الصغرى الاعتيادية
OP's	Oracle properties	خصائص اوراكل
OPG	Outer product of gradients	الناتج الخارجي للتدرجات
PACS	Pairwise absolute clustering and sparsity	المجموعات الزوجية المطلقة

		والتناثر
PHd	principal Hessian directions	اتجاهات هيسين الرئيسية
PE	Prediction error	خطأ التنبؤ
P-MAVE	Penalized MAVE	مقدر اقل معدل تباين الجزائي
RSIR	Regularized Sliced inverse regression	الانحدار المقطعي المعكوس المنتظم
R SMAVE	Robust Sparse MAVE	مقدر أقل معدل تباين متناثر حصين
R SMAVE-EN	Robust Sparse MAVE with EN penalty	مقدر اقل معدل تباين متناثر حصين مع الشبكة المرنة
SAVE	Sliced average variance estimator	تقدير معدل التباين المقطعي
SCAD-MAVE	Sparse MAVE with SCAD penalty	مقدر اقل معدل تباين متناثر مع جزاء سكادا
SD	Standard deviation	الانحراف المعياري
SDR	Sufficient dimension reduction	تقليل الابعاد الكافي
SIR	Sliced inverse regression	انحدار عكسي شرائحي
SMAVE	Sparse MAVE	مقدر اقل معدل تباين متناثر
SMAVE-EN	Sparse MAVE with EN penalty	مقدر اقل معدل تباين متناثر مع دالة جزاء الشبكة المرنة
SMAVE-AdEN	Sparse MAVE with AdEN	مقدر اقل معدل تباين متناثر مع الشبكة المرنة المتكيفة او التكيفية
SSIQR	Sparse sliced inverse quantile regression	الانحدار القسيمي المعكوس الشرائحي الصفري
SSIR	Sparse sliced inverse regression	الانحدار العكسي الشرائحي الصفري
SSIR-AL	Sparse sliced inverse regression with adaptive Lasso	الانحدار العكسي الشرائحي الصفري مع دالة جزاء لاسو التكيفي
SSIR-EN	Sparse Sliced inverse regression with Elastic net penalty	انحدار عكسي شرائح صفري مع دالة جزاء الشبكة المرنه
SSIR-PACS	Sparse sliced inverse regression with PACS	انحدار عكسي شرائحي صفري معPACS
V.S	Variable selection	اختيار متغير

الفصل الاول

Chapter one

المقدمة

مشكلة الرسالة

هدف الرسالة

الأستعراض المرجعي

الفصل الاول

(المقدمة –مشكلة الرسالة- هدف الرسالة – الأستعراض المرجعي)

المقدمـــــــــــــــة (Introduction)

ان علم الاحصاء اداة مهمة لخدمة الانسان وتطويره في مختلف المجالات، إذ اصبح علماً بارزاً في مختلف مجالات العلوم التطبيقية. وان خطوات هذا العلم قسمت على مراحل عدة وان اهم المراحل التي تستند اليها دقة النتائج المتوخاة من النموذج الاحصائي المعتمد بالدراسة هي مرحلة تقدير معالم النموذج الاحصائي، إذ ان مقدرات النموذج هي الاساس التي تعطي صفة الظاهرة المدروسة ،والنموذج الاحصائي يختلف من حيث الهدف والشكل فربما يكون نموذج سلسلة زمنية أو نموذج تصميم او نموذج توزيع احتمالي او نموذج انحدار وغيرها.

تم اقتراح العديد من الطرائق لتقدير معاملات نموذج الانحدار في حالة استيفاء الفرضيات الأساسية .وكان اشهر تلك الاساليب طريقة المربعات الصغرى (OLS)(Ordinary Least Square)التي تستخدم في الانحدار الخطي البسيط والانحدار الخطي المتعدد لتقدير معالم النموذج لان مقدرات هذه الطريقة تتصف بأنها تمتلك أفضل تقدير خطي غير متحيز (Best Linear Unbiased Estimator). تعد دراسة الانحدار عندما يكون هناك عدد كبير من المتغيرات وحجم عينة كبير عملية صعبة ومعقدة ، لأنها تزيد من تعقيد نموذج الانحدار، مما دفع الباحثين إلى استخدام عملية اختيار المتغير(Variable selection)لان بعض المتغيرات التوضيحية(explanatory variable) تكون غير اساسية في تأثيرها على المتغير المعتمد(dependent variable) او يكون تأثيرها مماثلاً لتأثيرات متغيرات اخرى وان العديد من هذه المتغيرات يكون لها ارتباط داخلي مع بعضها البعض مما

يؤدي الى ظهور مشكلة التعدد الخطي (Multicollinearity) وبهذا يكون تأثيرها غير معنوي، مما يدعو الى استبعاد المتغيرات غير المعنوية واختيار المتغيرات المعنوية لزيادة دقة تنبؤ النموذج.

في معظم مجالات الدراسة نستخدم طرائق تقدير المعلمات لتحليل العلاقة الموجودة بين مجموعة من المتغيرات التوضيحية والمتغير التابع، بافتراض ان دالة التوقع الشرطي هي معلومات عن العديد من المعلمات ،ولكن اذا كانت أفتراضات النموذج غير متحققة ،فقد تكون التقديرات الناتجة مضللة

للغاية(Horowitz وآخرون،2002). تتمثل الطريقة المألوفة المستخدمة للتعامل مع البيانات عالية الأبعاد (HD)(High Dimension) في نموذج الانحدار في إزالة بعض المتغيرات دون فقدان أي معلومات ودون الحاجة إلى نموذج معلمي معين مسبقًا. ،غالباً ما تؤدي الابعاد العالية الى مشكلة تعدد الابعاد(Curse dimensionality) والتي يجب معالجتها بطريقة صحيحة. هذه المشكلة قادت الباحثين للعمل على تقليل الأبعاد العالية للبيانات ،من خلال توفر أساليب فعالة للتعامل مع المشكلة. فقد اقترح Cookفي عام(1998) نظرية (SDR)(Sufficient dimension reduction)، وهذا الاسلوب ذو اهمية عالية بوصفه أحد الادوات الفعالة لمعالجة مسألة تحليل البيانات ذات الابعاد العالية. تم تقديم العديد من طرائق تقليل الأبعادSDR،على سبيل المثال طريقة MAVE (**Minimum average variance estimator**)(Xiaوآخرون,2002) . ومع ذلك ، فإن النتائج تكون مجموعات خطية من جميع المتغيرات التوضيحية الاصلية. لذلك هذه الطرائق تعاني من صعوبة تفسير التقديرات الناتجة. أن اختيار أهم مجموعة فرعية صغيرة من المتغيرات التوضيحية تجعل تفسير النتائج أمرًا سهلاً ،ونحصل على نماذج اقل تكلفة وتعطي فهماً جيداً لمجموعة البيانات. وعلاوة على ذلك، يمكن أن يؤدي اختيار المتغيرات المهمة إلى تحسين دقة التنبؤ بالنموذج. تم اقتراح العديد من الطرائق للجمع بين طرائق SDR وطرائق التنظيم (Regularization method) وهذه الطرائق قادرة على التعامل مع البيانات عالية الابعاد وذلك بأضافه قيد معين على المعلمات و تقليص بعض المعاملات وتجعل الاخرى مساوية للصفر، وتعطي نموذجاً متبعثراً(Sparse Model) يتضمن اقل عدد ممكن من المتغيرات وقابلاً للتفسير، و *تستخدم* هذه الطرائق *من* قبل العديد من الباحثين لحل مشكلة عدم الاستقرار التي تعاني منها الطرائق التقليدية(Classical methods) وتعمل على تطوير قدرة تفسير النموذج فضلاً عن دقه التنبؤ من خلال الاختيار التلقائي إذ يتم اختيار المتغير وتقدير المعلمات في وقت واحد.

على سبيل المثال اقترح الباحثان Alkenaniو Rahman عام(2020) طريقة SMAVE-EN ، حيث جمع الباحثان بين طريقة MAVE (Minimum Average Variance Estimator)المقترحة من قبل الباحث Xiaوآخرين عام(2002)مع الشبكة المرنة EN المقترحة من قبل Zouو Hastieعام(2005)وتمتاز هذه الطريقة بأن لديها القدرة على التعامل مع المتغيرات التي تكون بشكل مجاميع(groups) وذات ارتباط عالي تحت أفتراضاتSDR.

واقترح الباحثان Alkenaniو Rahman عام(2020) طريقة SMAVE-AdEN، فقد جمع الباحثان بين طريقة MAVE (Minimum Average Variance Estimator) مع الشبكة المرنة التكيفيةAdEN(Adaptive Elastic Net) التي تم اقتراحها من قبل الباحثين Zouو Zhang عام(2009)لإنتاج طريقةSMAVE-AdEN وهذه الطريقة تتميز بانها تعطي تقديرات دقيقة عندما تكون الارتباطات عالية بين المتغيرات علاوة على ذلك ، يتم اختيار المتغير وتقدير المعلمات في آن واحد. وبرغم من هذه المزايا الجيدة لهذه الطريقة الا انها تفقد كفاءتها في حالة وجود قيم شاذة في بياناتها. أن حساسية الطريقة المذكورة للقيم الشاذة والتأثر بها دعانا الى أن نقترح في هذه الرسالة طريقة حصينة وهي (RSMAVE-AdEN) ، والتي يمكنها تقدير المعلمات واختيار المتغيرات في وقت واحد ، ولا تتأثر بوجود القيم الشاذة والطريقة المقترحة تعمل ضمن افتراضات نظرية اختزال البعد الكافي. وتم التحقق من فعالية الطريقة المقترحة من خلال دراسة المحاكاة وتحليل البيانات الحقيقية.

وقد تضمنت هذه الرسالة أربعة فصول وعلى النحو الآتي:

الفصل الاول يتضمن المقدمة، مشكلة الرسالة، هدف الرسالة ،والاستعراض المرجعي لأهم الدراسات السابقة ذات الصلة بموضوع الرسالة.

الفصل الثاني ويتضمن الجانب النظري ويضم مراجعة بعض طرائق اختيار المتغير V.S ضمن افتراضاتOLS.واستعراض موجزلـ SDR،MAVE،SMAVE-EN،SMAVE-AdEN والطريقة الحصينة المقترحة RSMAVE-AdEN.

الفصل الثالث ويتضمن مبحثين ، المبحث الاول دراسة جانب المحاكاة للطريقة المقترحة بينما تضمن المبحث الثاني دراسة البيانات الحقيقية لعينة الدراسة وتطبيق الطريقة المقترحة، ومن خلال دراسة المحاكاة وتحليل البيانات الحقيقية تم التحقق من فعالية الطريقة المقترحة.

الفصل الرابع ويتضمن أهم الاستنتاجات والتوصيات التي انبثقت عن الرسالة، فضلاً عن بعض المقترحات لدراسات مستقبلية.

2.1 مشكلة الرسالة (Problem of the thesis)

تكمن مشكلة الرسالة ان طريقة SMAVE-AdEN غير حصينة وحساسة للقيم الشاذة لذا أقترحنا طريقة حصينة RSMAVE-AdEN لا تتأثر بالقيم الشاذة وتعطي نتائج دقيقة.

3.1هدف الرسالة (Aim of the thesis)

الهدف من دراستنا هو تطوير طريقة SMAVE-AdEN (Alkenani و Rahman ,2020)التي تتأثر بشكل كبير بوجود القيم الشاذة في البيانات. لذلك تم اقتراح طريقة حصينة RSMAVE-AdEN للتعامل مع القيم الشاذة في المتغيرات التوضيحية ومتغيرات الاستجابة تحت أفتراضات نظرية اختزال البعد الكافي.

4.1 الأستعراض المرجعي (Literature Review)

من أجل معرفة أهم الدراسات والبحوث التي تناولت البيانات ذات الابعاد العاليةHD(High Dimensional data)والمشاكل المترتبة عليها، بمعنى عندما تكون الابعاد كبيرة تظهر مشكلة تعدد الابعاد(Curse of dimensionality)، يهدف الباحثون الى تقليل ابعاد p متجه المتغيرات التوضيحيةx دون فقدان معلومات الانحدار ويتم ذلك من خلال اختيار مجموعة فرعية للمتغيرات المؤثرة في النموذج حيث تتولد مجموعة من الفضائات الجزئية وكل فضاء جزئي هو يمثل تركيبة خطية لمجموعة من المتغيرات الاصلية ويشار اليها بالرمز$S_{y/x}$ إذ تضم جميع معلومات الانحدار لـ(y/x)، وهنالك عدد من الطرائق المقترحة نذكر منها طريقةSIR(Sliced inverse regression) التي اقترحها Liفي عام(1991)تم توظيف هذه الطريقة في مجالات مختلفة مثل المعلوماتية الحيوية والتسويق والاقتصاد ،وهي طريقة فعالة لتقليل الأبعاد وفعالة في التعامل مع البيانات عالية الابعاد(HD). وفي هذه الطريقة يتم استبدال المتغيرات الأصلية بمجموعات خطية منخفضة الأبعاد من المتغيرات التوضيحية دون أي فقدان لمعلومات الانحدار ودون الحاجة إلى التحديد المسبق لنموذج أو توزيع خطأ. ومع ذلك ، فإنها تعاني من حقيقة أن كل مكون SIR هو مزيج خطي من جميع المتغيرات التوضيحية الأصلية ؛ وبالتالي ، غالبًا ما يكون من الصعب تفسير النتائج المستخرجة.

في حين اذا كانت دالة المتوسط قيد الاهتمام(Cookو2002,Li)أقترحوا فكرة تقدير $S_{E(y/x)}$ بدلاًمن $S_{y/x}$ العديد من الطرائق تم اقتراحها لغرض تقدير $S_{E(y/x)}$ ،ومنها طريقة MAVE المقترحة من قبل Xiaوآخرين في عام(2002)وهذة الطريقة تعتمد على النماذج شبة المعلمية لاسيما للبيانات عالية الابعاد و تتصف بالمرونة اي قابلة للتطبيق على مجموعة واسعة من النماذج، وتم اثبات كفاءتها نظرياً وكذلك من خلال دراسة المحاكاة والبيانات الحقيقية.

تم اقتراح العديد من الطرائق للجمع بين طرائق تقليل البعد الكافي(SDR) مع طرائق التنظيم من قبل عدد من الباحثين .

ففي عام(2005) قدم (Niوآخرون) طريقة SSIR(**Sparse sliced inverse regression**) وهذه الطريقة ناتجة من الجمع بين طريقة Lasso(Tibshirani ،1996) وطريقة SIR(Li،1991) وهي قادرة على التعامل مع المتغيرات عندما تكون الارتباطات عالية بين المتغيرات التوضيحية وتم اثبات فعاليتها نظرياً ومن خلال دراسة المحاكاة وتحليل البيانات الحقيقية.

وفي عام(2007) قدم Li طريقةSSIR والتي تجمع بين فكرة Lasso مع مجموعة من طرق SDR لإنتاج حلول دقيقة ومتفرقة، وتم التحقق من فعالية الطريقة المقترحة نظرياً وعن طريق المحاكاة و البيانات الحقيقية.

وفي عام(2008) قام الباحثان (Wang وYin) بدمج طريقة (Lasso) مع (MAVE) لإنتاج(SMAVE) وهذة الطريقة بإمكانها تقدير الابعاد وهي طريقة غير حصينة تتأثر بالقيم الشاذة وانها اعتمدت على Lassoالذي يعطي تقديرات متحيزة للمعاملات الكبيرة ولا تمتلك خصائص اوراكل (**Oracle properties**).

وفي عام(2013)قام الباحثان (Alkenani وYu) بدمج طريقة MAVE مع الدوال الجزائية(SCAD)و(ALasso)و(MCP)لاقتراح (SCAD-MAVE و ALMAVE وMCP-MAVE)على التوالي تعطي هذة الطرق حلول متفرقة ودقيقة، ولا تحتاج اي توزيع معين ، وتمت معرفة كفاءة هذه الطرائق من خلال دراسة المحاكاة وتحليل البيانات الحقيقية.

وفي عام (2013)اقترح (Wangوآخرون)، طريقة P-MAVE طريقة مقدر اقل معدل تباين جزائي Penalized-MAVE إذ تكون الطريقة المقترحة قادرة على اختزال الابعاد واختيار المتغير.الا ان

هذه الطريقة تقل دقة التقدير فيها عندما يكون عدد المتغيرات كبيراً ولم تحل هذه المسألة من قبل الباحثين واعدت مقترحاً لدراسة مستقبلية .

وفي عام (2019) اقترح الباحثان (Alkenani وMalik) طريقة LQMAVE الذي يجمع بين

طريقة QMAVE وطريقة Lasso ، إذ عرف عن طريقةQMAVE طريقة جيدة لاختزال الابعاد تحت افتراضات الانحدار القسيمي لكن النتائج تكون عبارة عن تركيبة خطية لجميع المتغيرات التوضيحية الاصلية وهذا يجعل تفسير النتائج المقدرة صعباً ولهذا تم الجمع بين الطريقتين للتخلص من المتغيرات التوضيحية غير المهمة الموجودة ضمن التراكيب الخطية وقد اثبت الباحثان فعالية الطريقة من خلال دراسة المحاكاة و تحليل البيانات الحقيقية.

وفي عام(2020) اقترح الباحثان (Alkenani وRahman)، طريقة SMAVE-ENالتي تجمع بين MAVE و EN ، وتمتاز هذة الطريقة بأن لديها القدرة على التعامل مع المتغيرات التي تكون بشكل مجاميع ذات ارتباط عالي ، وتحت أفتراضات نظرية أختزال البعد الكافي .

واقترح الباحثان أيضًا طريقة SMAVE-AdEN التي تجمع بين طريقة MAVE وطريقة الشبكة المرنة التكيفية AdEN، لإنتاج تقديرات متفرقة ودقيقة عندما تكون الارتباطات عالية بين المتغيرات التوضيحية. وقد تم اثبات كفاءة الطريقة من خلال دراسة المحاكاة وتحليل البيانات الحقيقية الخاصة بأهم المتغيرات المهمة المؤثرة في المستوى العلمي لطلبة الدراسات العليا في جامعة القادسية .

وفي عام(2020) اقترح الباحثان(Alkenani وAbdulkadhim)طريقة SSIR-EN عندما جمع الباحثان طريقة الشبكة المرنة EN مع طريقة الانحدار المعكوس الشرائحي SIRهذه الطريقة توفر دقة تنبؤ جيدة وتفسيرات سهلة للنموذج من خلال دراسة المحاكاة وتحليل البيانات الحقيقية تحت أفتراضات نظرية اختزال البعد الكافي.

وفي عام(2021) اقترح (Alkenaniو Salman) طريقة SSIR-AL ، وقام الباحثان بدمج طريقة adaptive Lasso مع طريقة الانحدار المعكوس الشرائحي SIRويمكن لهذه الطريقة تحقيق خصائص اوراكل (**Oracle properties**) ، وتمتد الى الانحدارات غير الخطية وكذلك متعددة الابعاد دون الحاجة الى اي نموذج محدد وهي طريقة فعالة من الناحية الحسابية وكذلك من ناحية التقدير والاختيار في كل من المحاكاة والبيانات الحقيقية .

ومع ذلك ، وبرغم الميزات الجيدة لكل طريقة الا ان هذة الطرائق ليست حصينة وحساسة لوجود القيم الشاذة في المتغيرات. لذلك قدمت العديد من الدراسات(الطرق) الحصينة .

ففي عام (2006)اقترح الباحثان (Cizekو Hardle) الطريقة الحصينة RMAVE(**Robust** Minimum Average Variance Estimator) **Sparse** تعمل هذه الطريقة بنفس الجودة مثل الطرق الاصلية بالنسبة للبيانات العادية وحصينة بالنسبة للبيانات الشاذة وتكون سهلة التنفيذ وتم اثبات فعاليتها نظرياً ومن خلال دراسة المحاكاة والبيانات الحقيقية.

وفي عام (2013)اقترح الباحثان(Wangو Yao) الطريقة الحصينة(RSMAVE) وهي طريقة اختيار متغير حصينة خالية من النماذج تجمع بين الطريقة الجزائيةLasso و طريقة تقليل الابعادMAVE مع خوارزمية تقدير فعالة لتعزيز قابليتها للتطبيق العملي، ومن خلال دراسة المحاكاة وتحليل البيانات الحقيقية ومقارنة هذة الطريقة مع طريقةMAVEوSMAVEوعندما يكون النموذج متناثراً والقيم الشاذة موجودة في متغيرات الاستجابة ،يكون لطريقة RSMAVE اداء افضل على باقي الطرائق .

وفي عام(2017) قدم Alkenaniمع Dikheel طريقةRPACS(Robust **Pairwise absolute clustering and sparsity**) هذه الطريقة حصينة تجاة القيم الشاذة اي لا تتاثر بها وذلك عن طريق استخدام طريقة التقديرMMالحصينة، يتم فيها تقدير معلمات الانحدار واختيار المتغيرات المهمة في وقت واحد ، وتم استخدام دراسة المحاكاة واثنين من تطبيقات البيانات الحقيقية لتقييم فعالية هذه الطريقة.

وفي عام (2020)أقترحAlkenani طريقة RSSIR وهو طريقة اختيار متغير(VS)،وهي طريقة حصينة لا تتأثر بوجود القيم الشاذة اي مقاومة للقيم الشاذة ،حيث يتم استبدال دالة الخسارة التربيعية بمعيار توكي وايضا يتم استبدال التقديرات الكلاسيكية لمصفوفة المتوسط والتغاير بالمتوسط والتغاير الكروي وهما مقياسان حصينان للموقع والتشتت، تتمتع هذه الطريقة بدقة تنبؤ عالية ولها اداء افضل لتقدير المعلمات واختيار المتغيرات وقد بينت هذه الطريقة فعاليتها من خلال نتائج دراسة المحاكاة وتحليل البيانات الحقيقة.

وفي عام (2021) اقترح Alkenani الطريقة الحصينة (RSSIR-PACS). يظهر من خلال دراسة المحاكاة وتحليل البيانات الحقيقية ان هذه الطريقة لديها دقة تنبؤية عالية لجميع النسب المئوية التي استخدمت للتلوث وقدرة عالية على تحديد المجموعات المرتبطة ذات الصلة.

وايضاً في عام(2022)اقترح الباحثان Alkenaniو Aljobori الطريقة الحصينة (RSMAVE-EN). ان هذه الطريقة تتمتع بسلوك جيد في اختيار المتغير ودقة تقدير حتى مع وجود القيم الشاذة في yوx فضلاً عن ان هذه الطريقة لها نتائج جيدة ومتسقة مع جميع حالات التلوث من خلال المقارنة مع نتائج طرائق اخرى، وتم التحقق من فعالية هذه الطريقة من خلال نتائج الدراسات التي ظهرت لكل من عملية المحاكاة وتحليل البيانات الحقيقية. يمكن أن تقدر الاتجاهات في دالة متوسط الانحدار و تحديد المتغيرات المشتركة في وقت واحد ، في حين أنها حصينة لوجود القيم الشاذة المحتملة في كل من المتغيرات التابعة والمستقلة.

المساهمة في هذه الرسالة هي اقتراح طريقة حصينة يطلق عليها(RSMAVE-AdEN) يتم فيها استبدال دالة خسارة المربعات الصغرى في SMAVE-AdEN بمعيار توكي biweight Tukey. إنها طريقة فعالة عندما تكون الارتباطات عالية بين المتغيرات التوضيحية تحت أفتراضات SDR،فضلاً عن انها تتعامل مع القيم الشاذة الموجودة في كل من yوx.

الفصـــــــــــل الثـــــــــــاني

Chapter two

الجـــــــــانب النظـــــــــري

المبحث الاول

المبحث الثاني

الفصل الثاني

الجانب النظري/ المبحث الاول

1.1.2 اختيــار المتغيــر (Variable selection)

تعد نماذج تحليل الانحدار من بين الادوات الاحصائية الاكثر شيوعاً في الدراسات التطبيقية ولاسيما الدراسات التي تحتوي على بيانات ومتغيرات كبيرة منها الرياضيات التطبيقية والجينات الوراثية والهندسة الالكترونية وغيرها، لذلك كان من الضروري للباحثين ايجاد وتطوير عدة طرائق لتقدير معلمات هذه النماذج وكذلك اختيار العوامل المهمة و المؤثرة في الدراسات التطبيقية ، لذلك هنالك حاجة متزايدة لتقليل عدد هذه المتغيرات. إذ يعتقد ان جزءاً فقط من هذه المتغيرات مهم والبعض الاخر غير مهم .من الصعب او حتى من غير المجدي صياغة نموذج حدودي مع عدد كبير من المتغيرات المشتركة لذا يلعب اختيار المتغير دوراً مهماً في تحليل هذة البيانات عالية الابعاد، ليس فقط من اجل تفسير أفضل للنموذج ولكن من أجل دقة أعلى حيث يجعل تفسير النموذج سهلاً وايضاً يوفر نموذجاً منخفض التكلفة(Guyonو Elisseeff، 2003) .يمكن أن يتضمن نموذج الانحدار عددًا كبيرًا من المتغيرات التوضيحية و لا نعرف اياً من المتغيرات التوضيحية يمكن أن تؤثر على المتغير التابع، لذلك، فإن المهمة الرئيسة هي إنشاء نموذج انحدار يضم عدداً من المتغيرات التوضيحية المهمة (Hesterbergوآخرون،2008).

من ناحية أخرى، ان النموذج بعدد كبير من المتغيرات التوضيحية يكون مكلفاً، وان نموذج الانحدار بعدد محدود من المتغيرات التوضيحية يكون اسهل في التحليل وفي الفهم وكذلك وجود عدد من المتغيرات التوضيحية ذات الارتباطات الداخلية العالية يمكن ان تضيف قوة تنبؤية قليلة للنموذج لذلك يلعب اختيار المتغيرV.S(Variable Selection) دوراً مهماً في تحليل البيانات عاليه الابعاد لانه يسعى الى اختزال المتغيرات غير المهمة والتقليل من التحيز. تم اقتراح العديد من الطرائق من قبل الباحثين لاختيار المتغير V.S لتحقيق الأهداف المذكورة. وتنقسم طرائق اختيار المتغير على نوعين هما: الطرائق التقليدية وطرائق التنظيم.

2.1.2 الطرائق التقليدية لاختيار المتغير (Classical V.S methods)

نقدم هنا لمحه موجزة عن الطرائق التقليدية إذ ان الباحث غالبا مايضطر للبحث عن المتغيرات التوضيحية التي يعتقد انها تؤثر بشكل معنوي على المتغير المعتمد تحت الدراسة ،والاسباب التي تدفع الباحث للبحث عن المتغيرات التوضيحية المهمة هي ان بعض المتغيرات تكون غير اساسية في العلاقة، ويمكن ان تكون فيها اخطاء كبيرة في القياس ،وتأثيرها يمكن ان يكون مماثلاً لتأثير متغيرات اخرى، لذلك يرغب الباحث في تقليص عدد المتغيرات التوضيحية التي تستخدم بالنموذج النهائي. واخيرا لابد من الاشارة الى ان هنالك عدداً كبيراً من الطرائق التقليدية المقترحة لاختيار المتغيرات التي تتفاوت في اهميتها وفي دقتها(الراوي ،1987) من اجل الوصول الى النموذج النهائي الذي يضم المتغيرات التوضيحية المهمة ذات التأثير المعنوي على الانحدار تم اقتراح عدد من الأساليب التقليدية لأختار المتغير V.S في الأدبيات فعلى سبيل المثال لا الحصر:-

- الاختيار الأمامى أو المباشر(Forward selection Procedure(FSP)
- طريقة الاختيار التدريجيSWS ((Stepwise selection) (Efroymson،1960)
- طريقة الحذف المعاكس(BEP) (The Backward elimination procedure)
- طريقة معيارمعلومات أكايكي AIC (Akaike Information) (Akaike، 1973)
- معيارمعلومات بيزBIC (Bayesian Information Criteria)(Schwars، 1978)

هذه الأساليب لها عدة عيوب مثل عدم الاستقرار والتباين العالي واستهلاك الوقت لان الإجراءات منفصلة لاختيار المتغير وتقدير المعالم اي عملية اختيار المتغير وتقدير المعالم لا تتم في آن واحد. وبالتالي فإن الانموذج الناتج لديه دقة تنبؤ ضعيفة (Breiman، 1996).

1.2.1.2الاختيار الأمامى أو المباشر(FSP) (Forward Selection Procedure)

يبدأ هذا الإجراء بدون متغيرات مستقلة في النموذج ثم يضيف متغيرات مستقلة يتم تحديدها للمعادلة واحدة تلو الأخرى. يتم إضافة المتغير الأكثر أهمية أولاً بناءً على المقارنة (Fالمحسوبه) لكل متغير بقيمة (Fالجدولية). يتم حساب قيمة (Fالمحسوبه) لكل خطوة وبعد التحقق من أن القيمة أكبر من

(Fالجدولية) يتم إدخال المتغير المعني في المعادلة. تستمر هذه العملية في إظهار المتغيرات واحدًا تلو الآخر حتى الوصول إلى القيمة (F_l المحسوبة) أقل من القيمة (Fالجدولية) وفقًا للمعادلة (1-2)الاتية:

$$F^* = \frac{SSR(x1)}{\frac{SSE(x1)}{n-1}}$$

إذ SSR: تمثل الانحرافات الموضحة

SSE: تمثل الانحرافات غير الواضحة ، n :تمثل حجم العينة.

2.2.1.2 طريقة الاختيار التدريجي (Stepwise selection method)

اقترحEfroymson عام (1960)خوارزمية تسمى تقنية الاختيار التدريجي ،والتي تعد واحدة من أكثرطرائق أختيار المجموعات الفرعية شهرة واستخداماً على نطاق واسع. يتم تعريفها على أنها عملية تلقائية لاختيار النماذج في الحالات التي يوجد فيها عدد كبير من المتغيرات التوضيحية المحتملة ،وثم يتم تنفيذ الطريقة بشكل رئيس في تحليل الانحدار.

هذه الطريقة اقترحت لتحسين كفاءة طريقة الاختيار الامامي**Forward selection** و نقطة التمييز بين كلا الطريقتين هي أن جميع المتغيرات المستقلة في نهاية كل خطوة يتم فحصها بناءً على اختيار (Fالمحسوبة)،و إعادة التقييم مرة أخرى لأن هناك علاقات قوية بين المتغيرات المستقلة التي تم تقديمها في الخطوات السابقة. وبالتالي ، فقد عد هذا الإجراء منهجًا جيدًا لاختيار أفضل معادلة انحدار.

3.2.1.2 طريقة الحذف المعاكس(BEP) (The Backward elimination procedure)

تعد واحدة من أبسط طرق اختيار المتغير V.S ، بدءاً من نموذج كامل يأخذ في الاعتبار جميع المتغيرات التوضيحية التي سيتم تضمينها في معادلة الانحدار . ثم يتم حذف المتغيرات التوضيحية من المعادلة واحدة تلو الأخرى بالاعتماد على قيمة(F)الجدولية فالخطوة الاولى تتم من خلال اضافة كل المتغيرات التوضيحية الى المعادلة وبعدها نقوم بحساب قيم (F) المحسوبة داخل المعادلة لكل متغير ،حتى تبقى المتغيرات المستقلة المهمة فقط. وتعرف هذه الطريقة بالمعادلة الاتية:

$$F_{i\ partial} = \frac{SSR\left[\frac{xi}{all\ other\ explanatory\ variables}\right]}{\frac{SSE(x1,\ldots,xk)}{n-k-1}} \quad , \qquad (2\text{-}2)$$

SSR: مجموع الانحرافات الموضحة،

SSE: مجموع الانحرافات غير الموضحة(مجموع مربعات الاخطاء).

وبعد مقارنةFالجزائية مع F الجدولية لكل متغير على حدا ، إذ يتم حذف المتغير من المعادلة عندما تكون الجدولية(F)> (F)المحسوبه. وبعدها تبدأ الخطوة الثانية ونقوم بحساب F المحسوبة لباقي المتغيرات التوضيحية من الخطوة الاولى ويتم مقارنتها مع Fالجدولية بدرجة حرية (1) للبسط و(n-k-1) للمقام. ويحذف المتغير الذي يمتلك اقلF المحسوبه مقارنة معF الجدولية .وتستمرالعملية حتى الحصول على أقل قيمة الجدولية(Fالجدولية) <(F المحسوبه) تتوقف العملية.

4.2.1.2 معيارمعلومات أكايكيAIC (Akaike Information Criteria)

تم اقتراح AIC بواسطة (Akaike، 1973) وهو مقياس لمقارنة جودة النماذج وتحديد أي منها هو أنسب نموذج للبيانات. وبالتالي ، فإنه يوفر أداة لاختيار النموذج. وأفضل نموذج وفقًا لـ AIC هو النموذج الذي يحتوي على أدنى قيمة AIC(لان اقل قيمة تعطي اقل خطأ)،و لأنه يعمل على مبدأ تصغير الاخطاء اي يقيس مقدار المعلومات المفقودة وكلما كان اقل كان افضل ، ويتم استخدام المعادلة الآتية:

$$AIC(K)=-2\,Ln(L)+2K \quad , \qquad (2\text{-}3)$$

حيث L :هي قيمة MLE للنموذج وk :هي عدد معلمات النموذج.

5.2.1.2 معيار معلومات بيز(BIC) (Bayesian Information Criteria)

تم اقتراح طريقة BIC بواسطة (Schwarz، 1978) وهي معيار لاختيار نموذج من مجموعة معينة من النماذج. في الواقع BIC مشابهه لمعيار AIC ، لكن الاختلاف بينهما هو أن BIC يتضمن حجم العينة (n)الذي يمنح BIC ميزة على AIC (Carlos و Sergioc ، 2012). أفضل نموذج وفقًا لـ BIC هو النموذج الذي يحتوي على أقل قيمة BIC ويتميز ببساطة الحسابية واداءه الفعال ويتم توضيحه بالمعادلة الآتية:

$$BIC=-2Ln(L)+KLn(n) \quad , \qquad (2\text{-}4)$$

K: يمثل عدد المتغيرات التوضيحية ، n :يمثل حجم العينة.

3.1.2 طرائق التنظيم (Regularization methods)

لقد تم بذل قدر كبير من الجهد لتطوير طرائق التنظيم لاختيار متزامن للمتغير وتقدير المعالم ،طرائق التنظيم تخفف من تحيزات النمذجة وتحقيق دقة تنبؤ أعلى في النماذج عالية الابعادHDعن طريق تقليص المعاملات وتقديم تقديرات ذات مغزى حتى اذا كان النموذج يتضمن عدداً كبيراً من المتغيرات او مترابطين للغاية بالعلاقات الخطية ،تجعل عمليات تنظيم النماذج المعقدة أقل تعقيداً وبالتالي قدم Donohoو Johnstonعام (1994) أول استخدام لطرائق التنظيم لـ V.S .والتنظيم يعني أضافة دالة جزاء للتحكم في تعقيد النموذج الذي يستخدم مصطلح الجزاء (Penalty)أنه يقلل بشكل كبيرمن تباين النموذج مع عدم وجود زيادة كبيرة في التحيز، ذكرالباحثان (Fanو Li)في عام(2001) ان دالة الجزاء(Penalty function) يجب ان تتمتع بخصائص ثلاث وهي عدم التحيز(unbiasedness) والتناثراوالتبعثر (Sparsity)والاستمرارية(Continuity). بهذه الطرائق يتم تنفيذ V.S من خلال تقديرالمعلمات(Wangو Yin، 2008)، حيث توفر طرائق التنظيم أداة يمكننا من خلالها تطوير قدرة تفسير النماذج ودقة التنبؤ. وهذه الطرائق يتم فيها الاختيار التلقائي للمتغيرات وتقدير المعلمات في آن واحد. وفيما ياتي بعض الطرائق الجزائية :-

1- طريقةLasso (Least Abolute Shrinkage and Selection Oprator) اقترحت من قبل Tibshirani عام(1996)

2- طريقة(EN)(The Elastic Net method)للباحثين(Zouو Hastie عام(2005) .

3- طريقةSCAD(The smoothly clipped absolute deviation method) للباحثين (Fanو Li)عام(2001).

4- طريقةALasso(The adaptive Lasso method) للباحث (Zou)عام(2006) .

5- طريقةAdEN (The adaptive elastic net) للباحثين (Zou وZhang)عام(2009).

6- طريقة MCP (Minimax Concave Penalty) للباحثين(Zhuو Yu)عام(2010).

وغيرها من الطرائق. وتمتاز هذه الطرائق بأن لديها استقراراً أعلى مقارنة بالطرائق التقليدية، علاوة على ذلك، يتم تنفيذ عملية اختيار المتغيرV.S وتقديرالمعلمات في آن واحد(Alkenaniو Yu، 2013).

ومن المصطلحات المستخدمة مع طرق التنظيم(الجزائية)

1- معلمة الضبط او معلمة الجزاء (Tuning parameter)أو(Tuning Penalty)

ويرمز لها بالرمز(λ) وهي المعلمة التي تتحكم في مقدار اختزال المعلمات واختيار مجموعات فرعية من المتغيرات المدرجة في النموذج النهائي(Desboulets واخرون ،2018)،فإذا كانت قيمة هذه المعلمة تساوي صفراً($\lambda =0$)او تميل للصفر هذا يعني عدم فرض جزاء على المعاملات، اي نحصل على مقدرات طريقة (OLS) المربعات الصغرى الاعتيادية ،وتعطينا أختزالاً اكبر عندما تكون قيمة(λ) كبيرة، وهنالك عدة طرق لاختيار معلمة الجزاء (λ) وهي طريقة التقاطع الشرعي(GCV)(Generalized Cross Validation) (2006,Zou).ومن المعايير الاخرى لاختيار معلمة الضبط هو معيار اكاكي للمعلومات (AIC)(1973,Akaike) ،ومعيار بيزي للمعلومات(BIC)(Schwars,1978)، وتمتاز (λ) بانها:-

- λ هي معلمة الاختزال.
- التحكم والسيطرة بحجم المعاملات .
- التحكم والسيطرة بكمية الجزاء.
- كلما تقترب قيمتها من zero نحصل على مقدرات (OLS).
- وكلما تقترب قيمتها من ($\lambda \to \infty$)نحصل على نموذج معلمة المقطع الصادي β_0فقط.

2- خاصية اوراكل (Oracle property)

عندما يمتلك المقدّر بعض الخصائص ، يطلق عليه خاصية أوراكل. الخصائص هي: -

1- خاصية الاتساق(consistency) .

2- خاصية التناثر او التبعثر(Sparsity)

2- خاصية تقارب ألامثلية (Optimal Asymptoticacly)

ويقال للطريقة بأنها تمتلك خاصية أوراكل(Oracle) عندما تكون لديها القدرة على أختيار النموذج الحقيقي باحتمال مقدارة 1.(Fanو Li,2001)

1.3. 1.2 طريقة لاسو (Lasso method)

ان Lasso هي اختصار لـ(Least Abolute Shrinkage and Selection Oprator) تم اقتراح Lasso من قبل Tibshiraniعام(1996) إنها طريقة فعالة وقوية وهي جزء من عائلة المربعات الصغرى الجزائية حيث تعمل في آن واحد(نفس الوقت) على اختيار المتغيرات وتقدير المعالم ، و يقوم Lasso بتعيين جزاء للمعاملات في النموذج الخطي ، بمعنى ان طريقة Lasso تمتاز بخاصية رائعة هو انه يمكن تقليص بعض المعاملات الى الصفر تماماً وبالتالي يمكن ان تحقق تلقائياً اختيار المتغير. اوضح االباحثان(Fanو Li،2001) ان Lasso ينتج تقديرات متحيزة للمعاملات الكبيرة وبالتالي لايتمتع بخاصية اوراكل(Oracle property)، ومن عيوب Lassoأنه اذا كان لدينا pمن المتغيرات التوضيحية وn من المشاهدات وكانتp>n فأنه يختارالعددn من المتغيرات التوضيحية على الاكثر. وطريقة Lasso تعد قوية وفعالة لمعالجة البيانات ذات الابعاد العالية(HD) (High Dimension) ، يتم الحصول على هذا المقدّر عن طريق إضافة دالة جزاء إلى دالة خسارة المربعات الصغرى كما في المعادلة الآتية:

$$\hat{\theta}(Lasso) = arg\,min \sum_{i=1}^{n}(y_i - \theta^T x_i)^2 + \lambda \sum_{k=1}^{p} |\theta_k| , \quad (2\text{-}5)$$

الجزء الاول من المعادلة اعلاه يمثل دالة خسارة المربعات الصغرى، والجزء الثاني يمثل دالة جزاء Lasso. و(λ) تمثل معلمة الضبط (tuning parameter)او معلمة الجزاء(penalty parameter) ،وقيمتها أكبر من الصفر، وان(λ) تلعب دوراً اساسياً في عملية أختيار المتغير المعنوي أذ انها تسيطر على درجة الانكماش(التقلص) للمقدر لذا وجب تحديدها بشكل دقيق(Wangوآخرون ،2013) وتعمل على التحكم بشدة الجزاء بمعنى عندما تكون قيمةλ كبيرة فأنها تعطي اختزالاً اكبر في مدى تعقيد النموذج وتوفر معايير لاختيار المتغير ومقارنة النموذج من خلال فرض بعض القيود على المعلمات وكلما زادت قيمةλ فان معاملات الانحدار تتقلص الى الصفر ، λ> 0 ،اي تعطي نماذج قابلة للتفسير(Alkenaniو Yu، 2013) ، ويتم تحديد قيمةλ من خلال GCV(Generalized Cross Validation) كما في المعادلة الآتية:

$$\text{GCV}= \frac{RSS}{n\{1-\text{p}(\lambda)/nx\}^2} \quad , \qquad (2\text{-}6)$$

إذ

$$RSS=\sum_{i=1}^{n}(y_{i-\theta^T}\, x_i)^2, \qquad (2\text{-}7)$$

p(λ): يمثل العدد الفعال للمعلمات، تؤدي القيمة الأكبر لـ P(λ) إلى مزيد من التضخم (penalization).

p :يمثل عدد المتغيرات، i = 1, …,p ،

n :يمثل حجم العينة.

2. 3.1.2 طريقة لاسو التكيفي ALasso (Adaptive Lasso Method)

تم اقتراح لاسو التكيفي (ALasso)من قبل الباحث Zou عام (2006)، أمتداداً الىLasso لانه ومن المعروف ان تقديرات Lasso منحازة للمعاملات الكبيرة(اي يفرض نفس القيود على كل المعاملات وهذا ينتج تقديرات غير متسقة) ولا يمتلك خاصية اوراكل(Oracle property)(Li و Fan، 2001) .يتحكم لاسو التكيفي في تحيز تقديرات Lasso حيث يعمل على تعيين أوزان تكيفية مختلفة للمعاملات الجزائية في دالة الجزاء وهذا يؤدي الى زيادة الجزاء للمعاملات التي تقترب من الصفر وبعدها اختزال التحيز في تقدير الدالة وتحسين دقة اختيار المتغير. وبين Zou أن طريقة (adaptive Lasso) لها مزايا علىLasso بأنها تمتلك خاصية اوراكل لذلك، يمكن تقليل تحيز التقديرات عندما نكون قادرين على اختيار الأوزان بحيث يكون للمتغيرات التوضيحية ذات المعاملات الكبيرة أوزان أصغر، يتم تعريف تقديرات لاسو التكيفية كما في المعادلة الآتية:

$$\hat{\theta}(\text{adaptive Lasso})=\arg\min\sum_{i=1}^{n}(y_{i-\theta^T x_i})^2 + \lambda\sum_{k=1}^{p}\widetilde{w}_{k|\theta_k|}, \qquad (2\text{-}8)$$

$\widetilde{w}_k$ وتمثل دالة الاوزن التكيفية ،ويعبر عنه بالمعادلة $\widetilde{w}_k=1/|\hat{\theta}|^{\vartheta}$ ، $|\hat{\theta}|$ يمثل القيمة المطلقة للمعلمة المقدرة و ϑ قيمته اكبر من الصفر .

3.3.1.2طريقة الشبكة المرنة(EN) (Elastic Net Method)

إن الشبكة المرنة هي طريقة انحدار تم اقتراحها من قبل الباحثين (Zouو Hastie)في عام (2005) كنسخه محسنة الى Lasso، اذ بين الباحثان ان طريقة Lasso غير مستقرة عندما يكون بين

المتغيرات التوضيحية ارتباط عالٍ، ولتعزيز القدرة التنبؤية لـLasso تم اقتراح طريقة EN، وهي مزيج من دالة جزاء Lasso ودالة جزاء Ridge. وأشار الباحثون إلى بعض العيوب في عمل Lasso في بعض الحالات ، مثل:

1. إذا كان عدد المتغيرات(p)أكبر من حجم العينة(n) ، أي(p> n) ، فإن Lasso يختارعلى الاكثر n متغيرات التوضيحية.
2. . في حالة وجود مجموعة من المتغيرات تتميز بالارتباط العالي فيما بينها ، يختار Lasso متغيراً واحداً فقط بشكل اعتباطي من المجموعة.

يقلل Ridge معاملات المتنبئين المرتبطين تجاه بعضهم بعضاً (يتعامل مع مشكلة التعدد الخطي) ، بينما يختار Lasso واحدًا من بين المتنبئين المرتبطين. وبالتالي فقد أظهر الباحثون أن طريقة EN تتفوق على طريقة Lasso. ويتم تعريف تقديرات EN كما في المعادلة الآتية:

$$\hat{\theta}\ (EN)= \text{Arg min}\sum_{i=1}^{n}(y-\theta^T\ x_i)^2\ \ +\lambda_1\sum_{k=1}^{p}\theta_k^2+\lambda_2\ \ \sum_{k=1}^{p}|\theta_k| \quad , (2\text{-}9)$$

حيث الجزء الاول من المعادلة اعلاه يمثل دالة الخسارة لطريقة المربعات الصغرى(OLS)، والجزء الثاني يمثل دالة Ridge ، والجزء الثالث يمثل دالة جزاء Lasso.

و $0 \le \lambda_1 و \lambda_2$ هما معلمات الضبط(tuning parameters)

P:هو عدد المتغيرات ،k=1,2,…p

4.3.1.2 طريقة سكاد (SCAD) Smoothly clipped absolute deviation method

تم اقتراح هذة الطريقة في عام(2001) من قبل(LiوFan) وأشار الباحثان إلى أن مقدر(SCAD) يتميز بخاصية أوراكل(**Oracle properties**) وانها طريقة ذات أهمية عالية بسبب استخدام الخصائص الاحصائية والحسابية الممتازة ويمكن التعبير عنها بالمعادلة الآتية:

$$P_{SCAD\lambda,C}(\theta)=\begin{cases}\lambda\theta & if\ \theta\le\lambda\\ \frac{c\lambda\theta-0.5\theta^2}{c-1} & if\lambda<\theta\le c\lambda\\ \frac{\lambda^2(c+1)}{2} & otherwise,\end{cases} \quad , \qquad (2\text{-}10)$$

ومشتقتها الاولى بالنسبة لـ (θ) يعبر عنها بالمعادلة الآتية:

$$P'_{SCAD\lambda}(\theta) = \begin{cases} \lambda & if\ \theta \leq \lambda \\ \dfrac{c\lambda - \theta}{c-1} & if\ \lambda < \theta \leq c\lambda \\ 0 & otherwise, \end{cases}$$

(2-11)

يقلل SCAD الانحدارالجزائي (penalized regrssion)كما في المعادلة الآتية:

$$\hat{\theta}(\text{SCAD})=\arg\min\sum_{i=1}^{n}(y_i - x_i^T\theta)^2 + n\sum_{k=1}^{p} P_{SCAD\lambda,C}(|\theta_k|), \qquad (2\text{-}12)$$

حيث2< c وهي معلمة ضبط ثابتة،و$\lambda \geq 0$ وهما معلمات الضبط(tuning parameters) .

P:يمثل عدد المتغيرات،k=1,..,p. ، n :يمثل حجم العينة ،i=1,2,…,n

5. 3.1.2 طريقة الشبكة المرنة التكيفية(AdEN) (Adaptive Elastic Net Method)

اقترح الباحثان (ZhangوZou)في عام (2009)طريقة AdENحيث تتعامل مع العلاقات الخطية المتداخلة بشكل افضل من الطرق الاخرى ،وايضاً لمعالجة القيود المفروضة على الشبكة المرنة(EN)التي لا تملك خاصية اوراكل من خلال الجمع بين مزايا طريقة (Adaptive Lasso) ومزايا (Ridge)،Adaptive Lasso يمتلك خاصية اوراكل، بينما يتعامل (Ridge) مع العلاقات الخطية المتداخلة. إذ يتم بناء اوزان تكيفية بواسطة المعادلة الآتية:-

$$W_k^* = (|\hat{\theta}_k\,(\text{EN})| + 1/n)^{-\gamma} \quad , \qquad (2\text{-}13)$$

$\hat{\theta}_k\,(\text{EN})$ ويمثل مقدر الشبكة المرنة التكيفية،حيث$\gamma > \frac{2V}{1-V}$ (ثابت موجب)،$0 \leq V < 1$،

يتم تعريف مقدر الشبكة المرنة التكيفية بالمعادلة الآتية:-

$\hat{\theta}$ (AdEN)=arg min$\sum_{i=1}^{n}(y_i - \theta^T xi)^2 + \lambda_1 \sum_{k=_1}^{p} \theta_k^2 + \lambda_2 \sum_{k=1}^{p} w_k^* |\theta_k|$,(2-14)

الجزء الاول من المعادلة اعلاه تمثل دالة الخسارة لطريقة المربعات الصغرى(OLS)،والجزء الثاني يمثل دالة جزاءRidge ،والجزء الثالث يمثل دالة جزاء Adaptive Lasso.

حيثp: يمثل عدد المتغيرات ، k=1,……,p

n: يمثل حجم العينة، i=1,…..,n

λ_1،λ_2: معلمات التنظيم (tuning parameters)وهي مسؤولة بشكل مباشر عن تباين التقديرات.

w_k^* معادلة بناء دالة الاوزان التكيفية وتمت الاشارة اليها بالمعادلة(13-2).

6.3.1.2 طريقة MCP (Minimax Concave Penalty Method)

تم اقتراح الطريقة الجزائيةMCPمن قبل الباحث Zhang(2010) طريقة سريعة ودقيقة وغير متحيزة لاختيار متغيرفي انحدارات(HD)الخطية.يبدأMCP نسبة الجزاء(ROP)(Rate of Penalization)الذي يقابل ذلك في Lasso إذ يقلل باستمرار حدLassoحتى يكون صفراً، وقيمتها بين$[0, \infty)$ ويعبر عنها بالمعادلة الآتية:

$$P_{MCP\lambda,C}\theta = \begin{cases} \lambda\theta - \frac{\theta^2}{2C} & \text{if } \theta \leq c\lambda \\ \frac{1}{2}c\lambda^2 otherwise, \end{cases}$$

(2-15)

والمشتقة الاولى بالنسبة لـ(θ) يعبر عنها بالمعادلة الاتية:

$$(p'_{MCP\lambda,C}(\theta)=\begin{cases} \lambda - \frac{\theta}{c} sing(\theta) if\ \theta \leq c\lambda \\ 0 \quad otherwise \end{cases} \quad , \quad (2\text{-}16)$$

حيث1< c و0 $\geq \lambda$ وهما معلمات الضبط(tuning parameters).

يقلل MCP الانحدار الجزائي (penalized regrssion)كما في المعادلة الآتية:

$\hat{\theta}$(MCP)=arg min$\sum_{i=1}^{n}(y_i - x_i^T\,\theta)^2$ +n$\sum_{k=1}^{p} p_{MCP\lambda,C}(|\theta_k|)$, (2-17)

حيث1< c و$0 \geq \lambda$ وهما معلمات الضبط(tuning parameters)

P:يمثل عدد المتغيرات،.k=1,..,p وn:يمثل حجم العينة ،

i=1,2,…,n

4.1.2 استخلاص المتغير (Variable extraction)

يهدف هذا الاسلوب الى تحويل المتغيرات الاصلية البيانات من الفضاء الأصلي إلى مساحة ذات خصائص مشتركة بين المتغيرات ، أو بعبارة أخرى إنها عملية تؤدي إلى تقليل الأبعاد حيث إنها تقلل البيانات ذات الأبعاد العالية (H.D)إلى تلك ذات الأبعاد الأقل. هنالك العديد من طرائق (Sufficient dimension reduction (SDR لاستخراج المتغيرات وتقليل الأبعاد دون فقدان الكثير من المعلومات. ونذكر هنا طرائق استخراج المتغيرعندما يكون الفضاء الجزئي المركزي قيد الاهتمام (Central Subspace) (CS) $S_{y/x}$(Yuو,Zhu 2013)ونذكر منها:-

- الانحدار الشرائحي المعكوس(SIR) (Li، 1991).
- مقدر معدل التباين المقطعي(SAVE)The sliced average variance estimation ، (Cook وWeisberg ، 1991) .
- الاتجاهات الرئيسية (PHd) Principal Hessian directions (Li، 1991) وغيرها.

وطرائق استخراج المتغير عندما يكون الفضاء الجزئي للمتوسط المركزيCMS(Central mean subspace) $S_{Ey/x}$ قيد الاهتمام ونذكر منها طريقة مقدر اقل معدل تباين MAVE (Xiaوآخرون، 2002) .

5.1.2 تقليل البعد الكافي (SDR) Sufficient dimension reduction

لوحظ مؤخرًا أن ظاهرة البيانات عالية الأبعاد ظاهرة يومية وسائدة في عدد من العلوم، مثل العلوم البيولوجية وعلوم الارض والعلوم الطبية والهندسية والزراعة حيث يكون عدد المتغيرات كبيراً جداً، و من الصعوبة استخدام الأساليب الإحصائية التقليدية في عملية التقدير والتنبؤ، وبالتالي يجب اختزال عدد المتغيرات التوضيحية وتقليصها عن طريق اختيار مجموعة فرعية من المتغيرات التي فعلاً تكون مؤثرة بالنموذج، أي الهدف هو إيجاد أبعاد كافية (SDR) للمتغيرات التوضيحية المؤثرة فعلاً في النموذج، وبالتالي هي طريقة فعالة لتحديد المتغيرات المهمة ولكن عيب هذه الطرائق هو انها تنتج مزيجاً خطياً من جميع المتغيرات التوضيحية، هذة الحقيقة تجعل تفسير التقديرات الناتجة غير سهلة ، لذلك فهي تعاني من صعوبة في تفسير النتائج .

منذ العمل الرائد لـCook عام(1998)كانت (SDR) ذات اهمية كبيرة بوصفها أحد الادوات الفعالة لمعالجة قضية تحليل البيانات عالية الابعاد. كانت الفكرة الرئيسة لـ(SDR)في مشكلة الانحدار هوتحويل البيانات من فضاء عالي الابعاد الى فضاء منخفض الابعاد بحيث يحتفظ التمثيل منخفض الابعاد بالخصائص ذات المعنى للبيانات الاصلية، أي استبدال المتغيرات التوضيحية ببضعة تركيبات خطية مع الحفاظ على جميع معلومات الانحدارمن اجل الحصول على متوسط الفضاء الجزائي المركزCMS(Central mean subspace)، وحل مشكلة تعدد الابعاد (Curse of dimensionality)و كانت هنالك العديد من الاساليب المقترحة ضمن أفتراضات نظرية(SDR) لغرض ايجاد الفضاء الجزئي المتوسط المركزي(CMS)(Central mean subspace) $S_{E(Y/X)}$ فضاء فرعي مركزي (Sy/x)، وبالتالي ،كانت هنالك العديد من الاساليب المقترحة لـ(SDR) ومنها طريقة SAVE تقدير معدل التباين الشرائحي (Cookو Weisberg،1991) وهي واحدة من طرائق تقليل الابعاد لتقليل ابعاد المتغيرات التوضيحية، طريق(SIR) المقترحة من قبل Liفي عام(1991) وهي طريقة فعالة لتقليل الأبعاد وفعالة في التعامل مع البيانات عالية الابعاد(HD). يستبدل المتغيرات الأصلية بمجموعات خطية منخفضة الأبعاد من المتغيرات التوضيحية دون أي فقدان لمعلومات الانحدار ودون الحاجة إلى التحديد المسبق لنموذج أو توزيع خطأ. ومع ذلك ، فإنها تعاني من حقيقة أن كل مكون SIR هو مزيج خطي من جميع المتغيرات الأصلية ؛ وبالتالي ، غالبًا ما يكون من الصعب تفسير النتائج المستخرجة.

، وطريقة(PHd) The principal Hessian directions)المقترحة من قبلLi في عام (1992)كانت هذة الطريقة سهلة للغاية وسريعة الحساب ومع ذلك فهي تفرض بعض الافتراضات الاحتمالية على المتغيرات التوضيحية وتتطلب حجم عينة كبيرنسبياً. قدمت الدراسة التي أجراها Li وCookعام(2002)نهجاً لـ(CMS) متوسط الفضاء الجزئي المركزي ($S_{E(y/x)}$) ولغرض تقدير $S_{E(y/x)}$ولدراسة نموذج الانحدار مع متغير الاستجابة،$Y \in R^1$و$X_{px1} \in R^P$ متجة المتغيرات التوضيحية ،ولنفرض النموذج الآتي:

$$y=f(\theta^T X)+ \varepsilon \qquad (2\text{-}18)$$

$$y=f(x_1,x_2,.,.,.,x_{p)} + \varepsilon \qquad (2\text{-}19)$$

حيثf :رمز دالة الربط ((link function غير معلومة .

y:متغير الاستجابة

ε : يمثل حد الخطأ يمتلك متوسطاً مقداره 0 وتبايناً مقداره σ^2 ، بدون خسارة اي معلومات من دالة المتوسط المركزي (Centeral mean function)(CMF)(Liو Cook, 2002) والذي يرمز له بالرمز($S_{E\ (y/x)}$) اي ان

$$f(x_1,x_2,.,.,.,x_p)=E(y/x) \qquad (2\text{-}20)$$

و $E(\varepsilon /x)=0$

تهدف (SDR) إلى ايجاد المجموعة الجزئية او المجموعة الفرعية S (subset)من مساحة المتغيرات التوضيحية، إذ

$$Y \perp\!\!\!\perp E(y/x)|P_S(x) , \qquad (2\text{-}21)$$

إذ ⫫تشير الى الاستقلالية الاحصائية(statistical independence)،

و p(.) تشير الى معامل الاسقاط(projection operator)،

وتسمى المساحات الفرعية بمعادلة (2-21)بمتوسط البعد الكافي (cookو Li ,2002).

إذا كانت d = dim (S)

و $\theta=(\theta_1,\theta_2,..,\theta_d)$ هو اساسS(وتمثل متجه المعلمات غير الصفرية في النموذج)، فيمكن استبدال X بالمجموعة الخطية

$$\theta_1^T X, \theta_2^T X, \dots., \theta_d^T X,$$

d≤p ، دون فقدان المعلومات عنE(y/x).

تم اقتراح العديد من الطرائق لتقدير $S_{E(y|x)}$ واحدى هذة الطرائق هي (MAVE) المقترحة من قبل Xiaوآخرون عام(2002) .

2-1-5-1طريقة (MAVE) (Minimum average variance estimation)

تم اقتراح طريقMAVE من قبلXiaوآخرين عام(2002) وهي طريقة التقدير بأدنى معدل تباين وتستخدم هذه الطريقة على مدى واسع من نماذج الانحدار وهي احدى طرائقSDR لتقليل الابعاد، وهذة الطريقة تتمتع بميزات اهمها المرونة وقابليتها على الاتحاد مع طرائق اخرى، و قدرتها على اختيار المتغيرات وتقدير المعالم انياً وكذلك توفر خوارزمياتها وسهولة تنفيذها .

بحيث تكون θهي حل:

$$\min_{\theta} \{E[y- E(y|\theta^T x)]^2 \}, \qquad (2\text{-}22)$$

حيث$\theta^T\theta = I_d$ وهو شرط الاختزال

والتباين الشرطي لـ θ^T X هو

$$\sigma_\theta^2(\theta^T x)=E[\{y-E(y/\theta^T x)\}^2|\theta^T x] \quad , \quad (2\text{-}23)$$

$$\min E[y-E(y/\theta^T x)]^2 = \min E\{\sigma_\theta^2(\theta^T x) \quad , \quad (2\text{-}24)$$

بالنسبة لـ x_0 ، يمكن التقريب الخطي الموضعي لـ($\theta^T x_0$)σ_θ^2على النحو الاتي:

$$\sigma_\theta^2(\theta^T x_0) \approx \sum_{i=1}^n \{y_i - E(y_i/\theta^T x_i)\}^2 w_{i0}\}$$

$$\approx \sum_{i=1}^n [y_i - (a_{0+} b_0^T \theta^T (X_i - X_0))]^2 w_{i0} \quad , \quad (\mathbf{2-25})$$

وانW_{i0}:هي دالة تعيين المسافة بين X_iوX_0 وانها تلعب دوراً حيوياً في البحث عن اتجاهات (SDRالفعال حيث ان $a_0+b_0^T \theta^T(xi-x0)$ هو تركيب خطي موضعي (Local linear expansion) لـ$E(y_i/\theta^T x_i)$ عند x_0.

و$0 \geq w_{i0}$ هي اوزان كيرنل المتمركزة عند $\theta^T x_0$

حيث$\sum_{i=1}^n w_{i0} = 1$،

$$w_{ij} = k_h\{\hat{\theta}^T(X_i - X_j)\} / \sum_{i=1}^n k_h\{\hat{\theta}^T(X_i - X_j)\} \quad , \quad (2\text{-}26)$$

حيث$K_h(.)= h^d k(./h)$

و d هي ابعاد (.) k،

K(.) هي دالة كيرنل الدالة الكاوسية Gaussian متعددة الأبعاد المكررة (Brillinger ،1983) ، وكالاتي:

$K(x)= \frac{1}{\sqrt{2\pi}} \quad e^{-x^2/2}$

$h_{opt}=A(d)n^{-1/(4+d)}$ هي عرض النطاق الترددي

$A(d)=\{\frac{4}{(d+2)}\}^{1/(4+d)}$

و h: هي معلمة التمهيد(smoothing parameter)،يسمى عرض النطاق الترددي (bandwidth)الذي يتحكم في نعومة التقدير وتحيزه (Xiaوآخرون، 2002) كما في المعادلة الاتية:

$$Min(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i-\{a_j+b_j^T\theta^T(x_i-x_j)\}]]^2 w_{ij}), \qquad (2\text{-}27)$$

2-5-1-2 طريقة Sparse MAVE (SMAVE)

على الرغم من أن طريقة MAVE هي طريقة فعالة لتقليل الأبعاد ، إلا أن مخرجاتها لا تزال عبارة عن مجموعات خطية من جميع المتغيرات، لذلك فإنها تعاني من الصعوبة في التفسير كما تفعل طرائق (DR) الأخرى. لهذا اقترحت عدة طرائق للجمع بين طرق V.S وطرائق(SDR) في خطوة واحدة ،ففي عام(2008) اقترح الباحثانWang و Yin طريقة(SMAVE) تجمع بين طريقة Lassoوطريقة MAVE.لقد اضافوا حد جزاء الى دالة الخسارة MAVEفي معادلة(2-27) للحصول على تقدير متفرق او متناثر ،يتميز SMAVE بمزايا على Lasso لأنه يمتد إلى الإعدادات متعددة الأبعاد وغير الخطية دون افتراض أي نموذج معين. وتعرف SMAVE بالمعادلة الاتية:

$$\min(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i-\{a_j+b_j^T\theta^T(x_i-x_j)\}]^2 W_{ij}+\lambda\sum_{k=1}^{p}|\theta_{m,k}|), \qquad (2\text{-}28)$$

و m=1,...,d

وافترضوا أن d معروفة ثم اقترحوا أنه يمكنهم تقدير d وفقاً لنسخة جديدة من معيارBI C،و(λ) هومعامل تنظيم غيرسالبة يتحكم في مقدار التقلص او الانكماش (shrinkage).

3-5-1-2طريقة SMAVE-EN Sparse MAVE with EN penalty

في عام(2020) اقترح الباحثان(Alkenani و Rahman) طريقة(SMAVE-EN). إذ قام الباحثان بدمج طريقة MAVE(Xiaو آخرين، 2002) مع طريقة EN (ZouوHastie ، 2005) لإنتاج تقديرات متفرقة ودقيقة عندما تكون المتنبئات شديدة الارتباط وتتعامل مع العلاقات الخطية المتداخلة يتم تعريف طريقة SMAVE-ENبالمعادلة الاتية:

$$(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i-\{a_j+b_j^T\theta^T(x_i-x_j)\}]^{2}w_{ij})+\lambda_1\|\theta_m\|_2^2+\lambda_2\|\theta_m\|_1$$

(2-29)

حيث $\|.\|_2^2$ هيL_2 هي معيار متعلق بـ(Ridge)

و $\|.\|_1$ هوL_1 معيار المتعلق بـ(Lasso).

λ_1 و λ_2 هما معلمات ضبط الشبكة المرنة(tuning parameters)،التي تتحكم في مقدار التقلص او الانكماش(shrinkage).

4-5-1-2 طريقةSMAVE-AdEN

في عام(2020) اقترح الباحثان(Alkenani و Rahman) طريقة جديدة(SMAVE-AdEN)،(Sparse minimum average variance estimation via the adaptive elastic net) إذ تم دمج طريقة تقليل البعد الكافي (SDR) وهي MAVE (Xia واخرون عام 2002)مع طريقة Adaptive Elastic Net (Zou وZhang،2009) التي هي عبارة عن دمج طريقة انحدار الحرف (Ridge Regression) مع طريقة(Lasso penalty function Adaptive) للحصول على تقدير متناثر ودقيق ويتم تعريف طريقة SMAVE-AdENبالمعادلة الاتية:

$$(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i-\{a_j+b_j^T\theta^T(x_i-x_j)\}]^{2}w_{ij})+\lambda_1\|\theta_m\|_2^2+w_k^*\lambda_2\|\theta_m\|_{1,}$$

(2-30)

حيثp: يمثل عدد المتغيرات ، k=1,……,p

n: يمثل حجم العينة، i=1,…..,n

λ_1،λ_2: هما معلمات التنظيم (tuning parameters)وهي مسؤولة بشكل مباشر عن تباين التقديرات، اي التي تتحكم في مقدار التقلص او الانكماش(shrinkage).

w_k^* دالة الاوزان التكيفية كما في المعادلة(13-2).

ولكن هذه الطرائق تعمل ضمن افتراضات ومنها التوزيع الطبيعي لحد الخطأ و عند وجود قيم شاذة في البيانات فهي تعطي مقدرات غير كفوءة ، لذلك زاد أهتمام الاحصائيين في السنوات الاخيرة على معالجة حالة القيم الشاذة في البيانات من خلال طرائق التقدير الحصينة او بالطرائق الحصينة للتقدير(Robust Estimation Methods)وسنوضح ذلك في المبحث الثاني من هذا الفصل .

الفصل الثاني

الجانب النظري/ المبحث الثاني

اولاً: الطرائق الحصينة ضمن افتراضاتSDR

نذكر هنا لمحة موجزة عن بعض الطرائق الحصينة ضمن افتراضات SDR كالطريقة الحصينةR SMAVEو الطريقة الحصينة R SMAVE-EN .

ثانياً: الطريقة المقترحة The proposed method

الطريقة المقترحة هي R SMAVE-AdEN(Robust S MAVE with AdEN)

2.2.2مقدمة عن الطرائق الحصينة

ان اكثر الطرائق المتبعة لتقدير معالم النموذج الاحصائي هي الامكان الاعظم (Maximum (ML liklihooh)،وطريقة المربعات الصغرى(OLS)(Ordinary Least Sequared)وطريقة العزوم (M.OM)(Method of moments) وغيرها. ولكن هذه الطرائق تعمل ضمن افتراضات ومنها التوزيع الطبيعي لحد الخطأ ولذلك فهي تعطي مقدرات غير كفوءة عند وجود قيم شاذة في البيانات، لذلك زاد أهتمام الاحصائيين في السنوات الاخيرة على معالجة حالة القيم الشاذة في البيانات .أو بمعنى اخر عند وجود قيم شاذة في البيانات كيف يتم التعامل معها ؟ والجواب يتم التعامل معها من خلال طرق التقدير الحصينة او بالطرائق الحصينة للتقدير(Robust Estimation Methods) حيث يتم الحصول على مقدرات حصينة ذات كفاءة عالية مقارنه بالطرائق الاعتيادية في حالة وجود قيم شاذة في البيانات كما يفترض ان تكون مقدرات الطريقة الحصينة قريبة جداً من مقدرات الطريقة الاعتيادية عند عدم وجود قيم شاذة .

3-2-2 طريقة Robust SMAVEالحصينة

على الرغم من أن طريقةSMAVE لها مزايا مقارنة بالطرائق الموجودة الا انها غير حصينه تجاه القيم الشاذة بسبب استخدام معيار المربعات الصغرى، قدم الباحثان Cizekو Hardle عام(2006) بدراسة حساسية طريقة MAVEللقيم الشاذة واقترحا تعزيزاً حصيناً لـ MAVEعن

طريق استبدال المربعات الصغرى المحلية بتقدير -L او-M.يمكن تعريف طريقة RMAVE الحصينة بالمعادلة الاتية:

$$\sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T\,\theta^T(X_i - X_j)\}]w_{ij}\ , \qquad (2\text{-}31)$$

حيث(.) p دالة خسارة حصينة. واقترح الباحثان Wangو Yaoعام(2013)طريقة RSMAVE واضافوا حد جزاء للمعادلة(2.31) لتصبح المعادلة كالاتي:

$$\sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T\,\theta^T(X_i - X_j)\}]w_{ij} + \sum_{K=1}^{d}\lambda_k\,|\theta_k|\ , \quad (2\text{-}32)$$

حيث(.) p : دالة خسارة حصينة.$|.|_1$ إيرمز الى λ_k L$_1$-norm

λ_k : معلمات التنظيم ،k=1,2,….,d

اقترح Alkenaniعام(2020) طريقة(RSSIR)اختيار متغير حصين في طريقة SIRباستخدام Tukey’s Biweight معيار توكي والتغاير الكرويball covariance.

4-2-2 طريقة Robust SMAVE-EN

قدم الباحثان Alkenaniو Aljoborعام(2021) دراسة عن حساسية طريقة SMAVE-EN للقيم الشاذة واقترحا تعزيزاً حصيناً لـSMAVE -EN- التي يمكن أن تقدر الاتجاهات في دالة متوسط الانحدار و تحديد المتغيرات المشتركة في وقت واحد ، في حين أنها حصينة لوجود القيم الشاذة المحتملة في كل من المتغيرات التابعة والمستقلة. يمكن تعريف طريقةRSMAVE-EN الحصينة بالمعادلة الاتية:

$$\sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T\,\theta^T(x_i - x_j)\}]\,w_{ij}) + \lambda_1\,\|\theta_m\|_2^2 + \lambda_2\|\theta_m\|_{1,}$$

(2-33)

حيث (.):p يمثل دالة خسارة حصينة

حيثλ_1, λ_2:هما معلمات الضبط(tuning parameters)

حيث $p(.)$ يمثل Tuke'y biweight للحصول على تقدير حصين في كلا المتغيرات المستقلة ومتغير الاستجابة عندما تتصف مشتقة دالة الخسارة بانها (aredescending derivative) ، فإن دالة الخسارة تكون حصينة ومقاومة للقيم الشاذة في xوy (Rousseeuwو Yohai، 1984). ومعيار Tukey biweight له هذه الخاصية(Tukey،1960).لذلك فأن طريقةRSMAVE-EN ليست حساسة للقيم الشاذة في xوy.كما في المعادلة (2-33) وهي نسخة حصينة مقارنه بمعادلة رقم(2-29)

وذلك من خلال استبدال دالة خسارة المربعات الصغرى في(2-29)بمعيار Tukey biweight.

5-2-2 طريقة Robust SMAVE-AdEN

على الرغم من أن طريقةSMAVE-AdEN المقترحة من قبل الباحثين(Alkenani و Rahman) في عام(2020) لها مزايا جيدة لاختيار المتغيرات وتقدير المعلمات وسهلة التنفيذ وتمتلك دقة تنبؤ جيدة مقارنة بالطرائق الموجودة الا انها غير حصينة أتجاه القيم الشاذة واقترحنا تعزيزاً حصيناً لـSMAVE-AdEN عن طريق استبدال المربعات الصغرى المحلية بمعيار Tukey biweight .يمكن تعريف طريقة RSMAVE- AdEN المقترحة الحصينة بالمعادلة الاتية:

$$\sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T \theta^T (x_i - x_j)\}]\, w_{ij}) + \lambda_1 \|\theta_m\|_2^2 + w_k^* \lambda_2 \|\theta_m\|_{1,}$$

(2-34)

حيث(.) p تمثل دالة خسارة حصينة، λ_1وλ_1 هما معلمات الضبط، نختار دالة(.) p بوصفها معيار Tukey biweight للحصول على تقدير حصين في كل من المتغيرات المستقلة ومتغير الاستجابة، و عندما تتصف مشتقة دالة الخسارة بانها (A redescending derivative) فأن دالة الخسارة حصينة ومقاومه للقيم الشاذة فيxوy(Rousseeuwو yohai، 1984).

معيارTukey biweightلها هذه الخاصية (Tukey،1960).لذلك فأن الطريقة المقترحة RSMAVE-AdENليست حساسة للقيم الشاذة فيxوy كما في المعادلة(2-34)وهي طريقة حصينة وتم ذلك عن طريق استبدال دالة خسارة المربعات الصغرى في معادلة(2-30)بمعيار توكي Tukey biweight ، ويمكن التعبير عن Tukey biweight بالمعادلة الاتية:

$$p_c(U)=\begin{cases} \frac{c^2}{\sigma}\{1-\left[1-\left(\frac{u}{c}\right)^2\right]^3\} & \text{if } |u|\le c \\ \frac{c^2}{\sigma} & \text{if } |u|>c \end{cases} \quad , \quad (2\text{-}35)$$

حيث c:هي ضبط مستوى الحصانة للحصول على 95%كفاءة مقاربة عند التوزيع الطبيعي القياسي، يفترض ان قيمة cهي4.685.

6-2-2 خوارزمية طريقة RSMAVE-AdEN الحصينة

يمكن الحصول على تقديراتRSMAVE-AdEN وفقاً للخوارزمية الآتية:

1- نفرض m=1 و $\theta = \theta_0$ اي متجه اعتباطي(Px1).

2- θ متجه معلوم، نجد متجه الحل(a_j,b_j) حيث انj=1,2,…,n

$$\sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T\,\theta^T(X_i - X_j)\}]w_{ij} \ ,$$

3- بعد الحصول على($\hat{a}_j, \hat{b}_j$)،j=1,2,..,n، نجد حل $\theta_{mRSMAVE-\ AdEN}$من خلال المعادلة الآتية(2-36):

$$\min_{\theta:\theta^T\theta=I} \sum_{j=1}^{n}\sum_{i=1}^{n} p[y_i - \{a_j + b_j^T(\hat{\theta}_1, \hat{\theta}_2, .. \hat{\theta}_{M-1}, \hat{\theta}_M)^T(x_i - x_j)\}]\ w_{ij})$$

$$+\lambda_1\ \|\theta_m\|_2^2 + w_k^*\lambda_2\|\theta_m\|_{1,}$$

4- الان نضع $\hat{\theta}_M$ في العمود mth منθ وتستمر الخطوات 2و3 الى ان يتحقق التقارب.

5- نحدث θ بـ$(\hat{\theta}_{1RSMAVE-\ AdEN}, \hat{\theta}_{2RSMAVE-\ AdEN}, .. \hat{\theta}_{MRSMAVE-\ AdEN}, \hat{\theta}_0)$ ولتكن m تساوي m+1.

6- اذاكانت m<d نعيد الخطوة رقم 2 الى الخطوة رقم 5 الى أن نصل الى التقارب اي m=d.

7-2 اختيار معلمة الضبط Tuning parameter selection

هنالك بعض معاييرالمعلومات على سبيل المثال معيار معلومات اكايكي Akaike's (AIC) information criterion (1973,Akaike) ،معيار معلومات بيز Bayesian information criterion (BIC) (1978,Schwars) ومعيارالمعلومات المتبقية residual information criterion (RIC) (Shiو Tsai،2002) غالباً ماتستخدم لاختيار معلمة الضبط وفقاً للصيغ الآتية:

$$AIC=n\log(RSS/n)+2P(\lambda) \quad (2\text{-}37)$$

$$BIC=n\log(RSS/n)\log(n)p(\lambda) \quad (2\text{-}38)$$

$$RIC=\{n-p(\lambda)\}\log(RSS/n-p(\lambda))+p(\lambda)\{\log(n)-1\}+4/(n-p(\lambda))\}, \quad (2\text{-}39)$$

إذ p (λ) تشير الى عدد المعلمات غير الصفرية .

يتم تعريفRSSبالمعادلة(2-40) الآتية:

$$RSS=\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i-\{\hat{a}_j+\hat{b}_j^T(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\dots,\hat{\theta}_{m-1},\hat{\theta}_m(x_i-x_j)\}]^2\ w_{ij},$$

بين الباحثان(Shiو Tsai) في عام (2002)أن استخدام RIC لاختيار λ يعطي اداء افضل وهو معيار ثابت. في هذة الدراسة استخدمنا نسخة حصينة منRICوالتي اقترحها الباحث Alkenaniفي عام(2020) كما في المعادلة(2-41) الآتية:

$$RRIC=\{n-p(\lambda)\log(RRSS/n-p(\lambda))+p(\lambda)\{\log(n)-1+\{4/(n-p(\lambda))\},$$

ونحصل علىRRSSمن خلال المعادلة (2-42) الآتية:

$$RRSS=\sum_{j=1}^{n}\sum_{i=1}^{n}p[y_j-\{a_j+b_j^T\theta^T(x_i-x_j)\}]w_{ij}$$

إذ أن $p(.)$ هو معيار Tukey biweight.

الفصل الثالث

Chapter three

دراسة المحاكاة والبيانات الحقيقية

Simulation study and Real data

الفصل الثالث / المبحث الاول (الجانب التجريبي)

3-1-1 مقدمة دراسة المحاكاة

تناول هذا الفصل الجانب التجريبي والجانب التطبيقي، تم استخدام الأمثلة العددية في المحاكاة و البيانات الحقيقية لتقييم أداء الطريقة المقترحة بالمقارنة مع الطرائق الأخرى والغرض من ذلك هو تقييم أداء الطريقة RSMAVE-AdEN المقترحة. قارنا الطريقة المقترحة RSMAVE-AdENمع الطرائق الموضحة في الفصل الثاني من هذه الرسالة، وهي طريقة SMAVE-AdEN (Alkenani وRahman،2020)وطريقةRSMAVE-EN (Alkenaniو Aljobori عام(2021)). تم إجراء المقارنة لإظهار سلوك الطريقة المقترحة RSMAVE-AdEN من حيث دقة التنبؤ وأختيار المتغير(V. S).للتحقق من كفاءة الطريقة المقترحة وامكانياتها في أختيار المتغيرات سوف نعتمد على معيار متوسط مربعات الخطأ(MSE) ومعدل عدد المعاملات الصفرية (Ave0's) لكل مثال.استخدمنا طريقة RIC الحصينة (RRIC). حيث تم اقتراح هذه النسخة الحصينة من RIC بواسطة Alkenaniعام (2020)والتي تم شرحها في الفصل الثاني تم حسابSMAVE-EN بأستخدام كودRالذي تمت كتابته بواسطة Alkenani and Rahmanعام(2020) ،تم حساب RSMAVE باستخدام كود R الذي تمت كتابته بواسطة Wang وYao (2013).بينما ، تم حساب -RSMAVE EN باستخدام كود R الذي تمت كتابته بواسطة Alkenani وAljobori عام(2021) .

قامت الباحثة بكتابة كود بلغة Rلحساب الطريقة المقترحةRSMAVE-AdEN تستند نتائج التقدير إلى (200) تكرار للبيانات. فضلاً عن ذلك ، كان توزيع المتغيرات التوضيحية وحد الخطأ كما في التوزيعات الآتية:-

1- التوزيع الطبيعي المعياري N(0,1)

2- توزيع t مع درجة حرية $t_3/\sqrt{3}$

3- $(1-\alpha)\ N(0,1)+\alpha\ N(0,10^2)$.

4- $(1-\alpha)\ N(0,1)+\alpha\ U(-50,50)$

فيما يخص التوزيعات في الحالة4و3 فأن$(1-\alpha)$% من البيانات تأتي من التوزيع الطبيعي القياسي وα %من توزيعات أخرى. (Wangو Yao, 2013)

المثال الاول: للنموذج الآتي:-

$y = 1+2(\theta^T x +3)\times \log(3|\theta^T x |+ 1) +\varepsilon ,$

إذ ،d=1, n=50,100, 200, p=40,

P:يمثل عدد المتغيرات

n:تمثل حجم العينة

d:تمثل عدد الابعاد

$$\theta = (\underbrace{0,..,0}_{10}, \underbrace{2,..,2}_{10}, \underbrace{0,..,0}_{10}, \underbrace{2,..,2}_{10})^T$$

حيث الارتباط لـ(X_i,X_j) يساوي 0.5 ، (Alkenani and Rahman،2021)

المثال الثاني: نعتمد نفس النموذج في المثال الاول

$y = 1+2(\theta^T x +3)\times \log(3|\theta^T x |+ 1) +\varepsilon ,$

إذ d=1, n=50,100, 200, p=40,

$$\theta= (\underbrace{3,..,3}_{15}, \underbrace{0,..,0}_{25})^T$$

$x_i= z_1+\varepsilon_i, \quad i=1,....,5$

$x_i=z_2+\varepsilon_i, \quad i=6,...,10,$

$x_i=z_3+\varepsilon_i \quad i=11,...,15,$

Xi تتوزع توزيع طبيعي N (0,1) ،حيث ان Xi تمتاز بأنها مستقلة بشكل مماثل(independent (iid identically distributed، حيث z يتبع نفس التوزيع مثل x وε ،والارتباط لـ (X_i،X_j) =0.8.

عندما15,…,1=i ،لدينا ثلاث مجموعات في هذا النموذج ، يوجد داخل كل مجموعة خمسة متغيرات (Alkenani و Rahman، 2021).

جدول رقم (3-1)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination) 5%للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x)\rvert$
1	RSMAVE-AdEN	14	0.157068	0.933459
	RSMAVE-EN	14	0.158234	0.936187
	SMAVE-AdEN	13	0.154574	0.946613
2	RSMAVE-AdEN	18	0.175321	0.915193
	RSMAVE-EN	16	0.469284	0.888995
	SMAVE-AdEN	11	0.936427	0.847767
3	RSMAVE-AdEN	18	11.20884	0.773604
	RSMAVE-EN	13	12.08385	0.463382
	SMAVE-AdEN	11	15.22138	0. 753192
4	RSMAVE-AdEN	18	0.204964	0.876533
	RSMAVE-EN	13	0.623045	0.747963
	SMAVE-AdEN	11	4.838745	0.708216

جدول رقم (3-2)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination) 10%للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x)\rvert$
1	RSMAVE-AdEN	14	0.157068	0.933459
	RSMAVE-EN	14	0.155096	0.926187
	SMAVE-AdEN	13	0.152438	0.936613
2	RSMAVE-AdEN	15	0.264667	0.903206
	RSMAVE-EN	13	1.726713	0.883613
	SMAVE-AdEN	11	2.003863	0.768843
3	RSMAVE-AdEN	14	12.14738	0.363678
	RSMAVE-EN	13	13.14674	0.309315
	SMAVE-AdEN	11	16.20868	0.260222
4	RSMAVE-AdEN	14	1.204639	0.814252
	RSMAVE-EN	12	2.996614	0.721139
	SMAVE-AdEN	11	4.875485	0.556117

جدول رقم (3-3)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination) 15%للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x) \rvert$
1	RSMAVE-AdEN	15	0.157068	0.933459
	RSMAVE-EN	12	0.1596096	0.926187
	SMAVE-AdEN	11	0.148138	0.956613
2	RSMAVE-AdEN	15	0.475902	0.922226
	RSMAVE-EN	13	2.514682	0.883024
	SMAVE-AdEN	11	2.387057	0.881686
3	RSMAVE-AdEN	14	13.32657	0.467028
	RSMAVE-EN	12	13.98386	0.366709
	SMAVE-AdEN	11	16.69545	0. 359005
4	RSMAVE-AdEN	14	1.202639	0.776301
	RSMAVE-EN	12	1.977723	0.500568
	SMAVE-AdEN	11	2.487216	0.478585

جدول رقم (3-4)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination) 20%للتوزيعين 3و4.

dist.	Method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x) \rvert$
1	RSMAVE-AdEN	14	0.197349	0.925614
	RSMAVE-EN	14	0.197546	0.913779
	SMAVE-AdEN	13	0.194436	0.937507
2	RSMAVE-AdEN	16	0.576296	0.884486
	RSMAVE-EN	15	2.917019	0.892740
	SMAVE-AdEN	11	3.346055	0.878714
3	RSMAVE-AdEN	16	14.20022	0.346731
	RSMAVE-EN	15	14.68767	0.266762
	SMAVE-AdEN	11	18.55533	0.325716
4	RSMAVE-AdEN	16	2.294025	0.789229
	RSMAVE-EN	14	2.885186	0.763167
	SMAVE-AdEN	11	4.402461	0.732669

جدول رقم (3-5)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباط وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث(Contamination) 5% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	14	0.073717	0.957768
	RSMAVE-EN	14	0.074843	0.956479
	SMAVE-AdEN	13	0.061121	0.969947
2	RSMAVE-AdEN	17	0.098880	0.975045
	RSMAVE-EN	14	0.231806	0.952287
	SMAVE-AdEN	11	0.382312	0.867518
3	RSMAVE-AdEN	16	12.54264	0.783236
	RSMAVE-EN	14	14.55765	0.786991
	SMAVE-AdEN	11	18.75063	0.683299
4	RSMAVE-AdEN	16	0.156871	0.972751
	RSMAVE-EN	14	0.605715	0.960329
	SMAVE-AdEN	11	1.376104	0.829236

جدول رقم (3-6)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination) 10% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	16	0.103626	0.957768
	RSMAVE-EN	15	0.103893	0.956479
	SMAVE-AdEN	12	0.101244	0.966947
2	RSMAVE-AdEN	17	0.188978	0.925891
	RSMAVE-EN	15	0.287284	0.894435
	SMAVE-AdEN	11	0.474869	0.817289
3	RSMAVE-AdEN	16	12.26655	0.674032
	RSMAVE-EN	15	13.27088	0.542085
	SMAVE-AdEN	11	18.85417	0.400728
4	RSMAVE-AdEN	17	0.227871	0.961538
	RSMAVE-EN	15	1.437581	0.940663
	SMAVE-AdEN	12	3.993977	0.850105

جدول رقم (3-7)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث(Contamination) 15% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	16	0.193284	0.957768
	RSMAVE-EN	14	0.194843	0.956479
	SMAVE-AdEN	11	0.192211	0.959694
2	RSMAVE-AdEN	17	0.220935	0.945471
	RSMAVE-EN	15	0.390582	0.931226
	SMAVE-AdEN	11	0.558628	0.929079
3	RSMAVE-AdEN	17	12.68152	0.664402
	RSMAVE-EN	15	13.24196	0.589092
	SMAVE-AdEN	11	19.64228	0.537557
4	RSMAVE-AdEN	17	0.66985	0.932853
	RSMAVE-EN	15	3.958482	0.842503
	SMAVE-AdEN	11	9.682328	0.723096

جدول رقم (3-8)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination)20 % للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	17	0.267159	0.969435
	RSMAVE-EN	15	0.294910	0.962514
	SMAVE-AdEN	11	0.269501	0.977096
2	RSMAVE-AdEN	16	0.246898	0.967210
	RSMAVE-EN	14	0.433529	0.965530
	SMAVE-AdEN	11	0.966594	0.945146
3	RSMAVE-AdEN	17	14.78032	0.393099
	RSMAVE-EN	15	15.31797	0.385594
	SMAVE-AdEN	10	19.97485	0.303590
4	RSMAVE-AdEN	16	0.998669	0.891183
	RSMAVE-EN	15	4.892254	0.362656
	SMAVE-AdEN	11	9.872421	0.268450

جدول رقم (3-9)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)5% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x)\rvert$
1	RSMAVE-AdEN	17	0.030207	0.989215
	RSMAVE-EN	15	0.091847	0.977498
	SMAVE-AdEN	11	0.026728	0.994402
2	RSMAVE-AdEN	16	0.042231	0.983752
	RSMAVE-EN	14	0.156892	0.975321
	SMAVE-AdEN	11	0.228871	0.927181
3	RSMAVE-AdEN	17	0.307871	0.949726
	RSMAVE-EN	15	0.664467	0.925868
	SMAVE-AdEN	11	1.353373	0.923392
4	RSMAVE-AdEN	17	0.048539	0.973941
	RSMAVE-EN	15	0.193108	0.963446
	SMAVE-AdEN	11	0.500422	0.952597

جدول رقم(3-10)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)10% للتوزيعين 3و4..

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x)\rvert$
1	RSMAVE-AdEN	15	0.040297	0.972215
	RSMAVE-EN	15	0.051849	0.977498
	SMAVE-AdEN	14	0.036728	0.978402
2	RSMAVE-AdEN	17	0.049932	0.983752
	RSMAVE-EN	16	0.159899	0.975321
	SMAVE-AdEN	11	0.524359	0.927181
3	RSMAVE-AdEN	17	0.407872	0.949726
	RSMAVE-EN	16	0.699467	0.925868
	SMAVE-AdEN	11	2.354373	0.923392
4	RSMAVE-AdEN	18	0.098588	0.973941
	RSMAVE-EN	16	0.199984	0.963446
	SMAVE-AdEN	12	0.587731	0.952597

جدول رقم (3-11)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)15% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	18	0.060207	0.969215
	RSMAVE-EN	16	0.092857	0.967498
	SMAVE-AdEN	11	0.056828	0.974402
2	RSMAVE-AdEN	17	0.100786	0.979623
	RSMAVE-EN	16	0.176433	0.973905
	SMAVE-AdEN	10	0.377609	0.947102
3	RSMAVE-AdEN	18	1.346873	0.952862
	RSMAVE-EN	16	1.920887	0.720138
	SMAVE-AdEN	17	4.938828	0.628651
4	RSMAVE-AdEN	18	0.649869	0.980888
	RSMAVE-EN	16	1.823538	0.941332
	SMAVE-AdEN	11	3.883485	0.782316

جدول رقم (3-12)النتائج للمثال الاول، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث(Contamination) 20% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	16	0.079686	0.877217
	RSMAVE-EN	15	0.087566	0.870353
	SMAVE-AdEN	14	0.028916	0.892206
2	RSMAVE-AdEN	16	0.192842	0.930622
	RSMAVE-EN	14	0.263451	0.923899
	SMAVE-AdEN	11	0.523425	0.910156
3	RSMAVE-AdEN	17	1.9598191	0.331425
	RSMAVE-EN	17	2.6746414	0.312821
	SMAVE-AdEN	11	5.1421151	0.310381
4	RSMAVE-AdEN	18	1.558712	0.849755
	RSMAVE-EN	17	2.023293	0.799965
	SMAVE-AdEN	11	5.133745	0.723814

بالمقارنة من خلال نتائج الجداول بين الطرائق الثلاثة المستخدمة اعلاه للمثال الاول نلاحظ ما يأتي:-

1-الطريقة المقترحة RSMAVE-AdEN هي الافضل ولجميع حالات التلويث (5%، 10% ،15%،20%) ولحجوم العينات المختلفة (200،100،50)حيث اعطت اقل قيمة لـAMSE ،وتأتي بعدها الطريقةRSMAVE-EN بينما كانت الطريقة SMAVE-AdEN الاكثر تأثرا بالقيم الشاذة فقد اعطت اعلى قيمة لـ MSE.

2- وايضاً أعطت طريقتنا المقترحة أعلى ارتباط بين المؤشر المقدر والمؤشر الحقيقية مقارنة بطريقة(RSMAVE-EN) و طريقة (SMAVE-AdEN).

جدول رقم (3-13)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination)5% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\lvert Corr\ (\theta^T x, \hat{\theta}^T x)\rvert$
1	RSMAVE-AdEN	22	0.268495	0.942695
	RSMAVE-EN	21	0.280064	0.940263
	SMAVE-AdEN	15	0.222348	0.950288
2	RSMAVE-AdEN	22	0.078353	0.971359
	RSMAVE-EN	20	0.356162	0.961048
	SMAVE-AdEN	14	1.280471	0.886569
3	RSMAVE-AdEN	21	1.928861	0.628161
	RSMAVE-EN	21	2.928952	0.377362
	SMAVE-AdEN	14	4.595167	0.624616
4	RSMAVE-AdEN	20	2.510605	0.895776
	RSMAVE-EN	20	3.558851	0.615418
	SMAVE-AdEN	15	6.354191	0.685529

جدول رقم (3-14)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين ($\theta^T x, \hat{\theta}^T x$)عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination)10% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr ($\theta^T x, \hat{\theta}^T x$) |
1	RSMAVE-AdEN	21	0.152904	0.9530087
	RSMAVE-EN	20	0.165082	0.9436942
	SMAVE-AdEN	13	0.150423	0.9834779
2	RSMAVE-AdEN	21	0.085041	0.9622362
	RSMAVE-EN	20	0.207196	0.9583329
	SMAVE-AdEN	14	3.551601	0.5124867
3	RSMAVE-AdEN	20	2.030684	0.9140596
	RSMAVE-EN	19	2.561672	0.6870368
	SMAVE-AdEN	14	6.818137	0.6310854
4	RSMAVE-AdEN	20	2.903068	0.8706734
	RSMAVE-EN	19	3.505554	0.8030579
	SMAVE-AdEN	15	6.411132	0.7854364

جدول رقم (3-15)النتائج للمثال الثاني، استناداً الى معدل المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين ($\theta^T x, \hat{\theta}^T x$)عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination)15%بالنسبة للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr ($\theta^T x, \hat{\theta}^T x$) |
1	RSMAVE-AdEN	20	0.043324	0.9445082
	RSMAVE-EN	20	0.042791	0.9431955
	SMAVE-AdEN	18	0.039354	0.9653098
2	RSMAVE-AdEN	20	0.066173	0.9379881
	RSMAVE-EN	21	0.167092	0.9255567
	SMAVE-AdEN	21	5.049433	0.8359557
3	RSMAVE-AdEN	19	2.455416	0.7902962
	RSMAVE-EN	16	2.787979	0.6502261
	SMAVE-AdEN	15	6.920646	0.5927749
4	RSMAVE-AdEN	22	3.094979	0.8488805
	RSMAVE-EN	22	3.857083	0.7731059
	SMAVE-AdEN	19	8.657814	0.7803584

جدول رقم(3-16)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباط عندما يكون حجم العينة n=50، p=40 ونسبة التلويث (Contamination)20% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	23	0.083674	0.788038
	RSMAVE-EN	22	0.091334	0.799522
	SMAVE-AdEN	17	0.086436	0.890959
2	RSMAVE-AdEN	23	0.094028	0.973395
	RSMAVE-EN	22	0.085978	0.958437
	SMAVE-AdEN	16	0.081624	0.864496
3	RSMAVE-AdEN	23	3.142763	0.969693
	RSMAVE-EN	22	3.672405	0.935889
	SMAVE-AdEN	20	8.632434	0.790441
4	RSMAVE-AdEN	23	3.104059	0.979781
	RSMAVE-EN	22	3.925261	0.929539
	SMAVE-AdEN	20	8.971697	0.734279

جدول رقم (3-17)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination)5% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	22	0.204056	0.944004
	RSMAVE-EN	21	0.210365	0.937518
	SMAVE-AdEN	18	0.201173	0.950409
2	RSMAVE-AdEN	22	0.145295	0.956032
	RSMAVE-EN	22	0.207786	0.955059
	SMAVE-AdEN	18	0.165431	0.887389
3	RSMAVE-AdEN	22	3.424541	0.793849
	RSMAVE-EN	21	3.820391	0.735499
	SMAVE-AdEN	19	8.482613	0.703441
4	RSMAVE-AdEN	23	3.620744	0.849991
	RSMAVE-EN	22	4.276505	0.791489
	SMAVE-AdEN	15	8.983258	0.763346

جدول رقم (3-18)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination)10% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\|Corr\ (\theta^T x, \hat{\theta}^T x)\|$
1	RSMAVE-AdEN	23	0.248207	0.944208
	RSMAVE-EN	22	0.275993	0,949595
	SMAVE-AdEN	21	0.241677	0.950898
2	RSMAVE-AdEN	22	0.117998	0.935793
	RSMAVE-EN	21	0.109298	0.899450
	SMAVE-AdEN	19	0.112874	0.785773
3	RSMAVE-AdEN	25	3.577161	0.669866
	RSMAVE-EN	23	3.850184	0.460525
	SMAVE-AdEN	18	9.197707	0.218881
4	RSMAVE-AdEN	24	0.959013	0.894399
	RSMAVE-EN	22	1.266276	0.889945
	SMAVE-AdEN	18	1.443551	0.872766

جدول رقم (3-19)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination)15% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	$\|Corr\ (\theta^T x, \hat{\theta}^T x)\|$
1	RSMAVE-AdEN	23	1.218732	0.948252
	RSMAVE-EN	22	1.234733	0.924981
	SMAVE-AdEN	17	1.222457	0.967045
2	RSMAVE-AdEN	23	0. 974028	0.973395
	RSMAVE-EN	22	0.985978	0.958437
	SMAVE-AdEN	18	0.971624	0.864496
3	RSMAVE-AdEN	23	3.841062	0.769693
	RSMAVE-EN	22	3.962851	0.535889
	SMAVE-AdEN	19	10.42763	0.290441
4	RSMAVE-AdEN	23	1.500726	0.8084853
	RSMAVE-EN	22	2.801536	0.7560389
	SMAVE-AdEN	19	4.518206	0.7882598

جدول رقم (3-20)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=100، p=40 ونسبة التلويث (Contamination)20% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	22	0.216098	0.950181
	RSMAVE-EN	20	0.215056	0.944314
	SMAVE-AdEN	18	0.204848	0.957774
2	RSMAVE-AdEN	22	0.137837	0.964781
	RSMAVE-EN	22	0.263157	0.956237
	SMAVE-AdEN	18	0.795216	0.945542
3	RSMAVE-AdEN	23	3.916592	0.709429
	RSMAVE-EN	22	4.177384	0.616185
	SMAVE-AdEN	15	10.74476	0.621263
4	RSMAVE-AdEN	23	2.048529	0.795993
	RSMAVE-EN	20	3.089233	0.781643
	SMAVE-AdEN	17	5.640814	0.781243

جدول رقم (3-21)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSEوقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)5% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	23	0.325259	0.915231
	RSMAVE-EN	22	0.337096	0.911084
	SMAVE-AdEN	16	0.327409	0.927016
2	RSMAVE-AdEN	23	0.092877	0.935677
	RSMAVE-EN	22	0.211598	0.925903
	SMAVE-AdEN	14	0.730230	0.847927
3	RSMAVE-AdEN	23	3.958434	0.456506
	RSMAVE-EN	22	4.530445	0.554282
	SMAVE-AdEN	19	10.75319	0.583353
4	RSMAVE-AdEN	23	2.577491	0.814019
	RSMAVE-EN	22	3.928073	0.769431
	SMAVE-AdEN	20	7.175188	0.764019

جدول رقم (3-22)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's) و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)10% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	23	0.071234	0.945763
	RSMAVE-EN	22	0.072623	0.946581
	SMAVE-AdEN	18	0.030669	0.958693
2	RSMAVE-AdEN	23	0.116038	0.941422
	RSMAVE-EN	23	0.224998	0.920752
	SMAVE-AdEN	18	0.484876	0.902862
3	RSMAVE-AdEN	23	4.653389	0.375417
	RSMAVE-EN	22	5.916264	0.275915
	SMAVE-AdEN	16	11.68551	0.272556
4	RSMAVE-AdEN	23	3.329034	0.912834
	RSMAVE-EN	22	4.630682	0.912362
	SMAVE-AdEN	17	7.512814	0.908459

جدول رقم (3-23)النتائج للمثال الثاني، استناداً الى معدل عدد المعاملات الصفرية (Ave. 0's)و AMSE وقيمة الارتباطCorrelation بين $(\theta^T x, \hat{\theta}^T x)$عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)15% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr $(\theta^T x, \hat{\theta}^T x)$ |
1	RSMAVE-AdEN	23	0.111717	0.954849
	RSMAVE-EN	22	0.120227	0.953033
	SMAVE-AdEN	18	0.106576	0.964729
2	RSMAVE-AdEN	23	0.113793	0.953986
	RSMAVE-EN	22	0.124278	0.951537
	SMAVE-AdEN	16	0.174209	0.947592
3	RSMAVE-AdEN	23	4.864345	0.544538
	RSMAVE-EN	22	6.356027	0.426109
	SMAVE-AdEN	17	11.95086	0.489896
4	RSMAVE-AdEN	23	3.633082	0.696182
	RSMAVE-EN	22	4.972578	0.698649
	SMAVE-AdEN	14	8.003712	0.720524

جدول رقم (3-24)النتائج للمثال الثاني، استناداً الى معدل المعاملات الصفرية (Ave. 0's)وAMSE وقيمة الارتباطCorrelation بين ($\theta^T x, \hat{\theta}^T x$) عندما يكون حجم العينة n=200، p=40 ونسبة التلويث (Contamination)20% للتوزيعين 3و4.

dist.	method	Ave.0's	AMSE	|Corr ($\theta^T x, \hat{\theta}^T x$) |
1	RSMAVE-AdEN	22	0.069115	0.792521
	RSMAVE-EN	21	0.067746	0.797423
	SMAVE-AdEN	19	0.054486	0.895586
2	RSMAVE-AdEN	22	0.106779	0.945405
	RSMAVE-EN	22	0.160059	0.944359
	SMAVE-AdEN	17	0.312197	0.927191
3	RSMAVE-AdEN	22	4.929224	0.553322
	RSMAVE-EN	22	5.183032	0.460796
	SMAVE-AdEN	16	12.26506	0.487122
4	RSMAVE-AdEN	23	3.831312	0.691816
	RSMAVE-EN	22	5.174935	0.653381
	SMAVE-AdEN	15	8.959484	0.661416

بالمقارنة من خلال نتائج الجداول بين الطرائق الثلاثة المستخدمة اعلاه نلاحظ ما يأتي:-

1-الطريقة المقترحة RSMAVE-AdEN هي الافضل ولجميع حالات التلويث (5%، 10% ،15%،20%) ولحجوم العينات المختلفة (50،100،200) إذ اعطت اقل قيمة لـMSE ،وتاتي بعدها الطريقةRSMAVE-EN بينما كانت الطريقة SMAVE-AdEN الاكثر تأثرا بالقيم الشاذة أذ اعطت اعلى قيمة لـ MSE.

2- وايضاً أعطت طريقتنا المقترحة أعلى ارتباط بين المؤشر المقدر والمؤشر الحقيقية مقارنة بطريقة(RSMAVE-EN)و طريقة (SMAVE-AdEN).

3- نلاحظ من نتائج الجداول ان قيمة MSE تزداد بزيادة قيمة التلويث في معظم تجارب المحاكاة.

الفصل الثالث/ المبحث الثاني (البيانات الحقيقية)

3-2-1المقدمة

في هذا المبحث تم جمع بيانات حقيقية عن مرض السكري متمثلاً بالمتغير المعتمد Y و المتغيرات المستقلة (العمر ،الوزن، ..الخ) بهدف التعرف على أداء الطريقة المقترحة RSMAVE-AdEN والطرق المدروسةRSMAVE-EN ،SMAVE-AdEN باستخدام البيانات الحقيقية، ونذكر هنا نبذة مختصرة عن مرض السكري وهو من الامراض التي تعاني منها البشرية بشكل عام .وقد لوحظ مؤخراً تزايد اعداد المصابين بهذا المرض في العراق. وعلى الرغم من أن مرض السكري غير معدٍ الا انه مزمن ويمكن أن يصيب معظم الاشخاص من مختلف الاعمار، ترجع أسبابه الى عدة عوامل ومنها العوامل العضوية نتيجة خلل خلقي في الشخص المصاب ،والعوامل الوراثية تسهم بشكل كبير في امكانية الاصابة وعوامل اخرى. يسبب مرض السكري نقصاً دائماً او نسبياً في الانسولين في الدم ،والذي يتم اطلاقة بشكل طبيعي من البنكرياس، يجعل الخلايا تقاوم الانسولين أو عندما يكون الجسم غير قادر على استخدام الانسولين المنتج بشكل صحيح(الانسولين هو هرمون ينقل الجلوكوز لأجراء عمليات التمثيل الغذائي التي تعد حيوية للطاقة التي تستخدمها أجهزة الجسم لتعمل)والانسولين هو الهرمون الذي ينظم مستوى الجلوكوز في الدم يؤدي انخفاضه إلى زيادة مستوى السكر في الدم وهذا يؤدي الى خلل في أداء بعض اجهزة الجسم وحدوث مضاعفات خطيرة مع طول مدة الاصابة التي تختلف من شخص لأخر، مثل امراض القلب والاوعية الدموية وارتفاع ضغط الدم ، ارتفاع نسبة الدهون في الدم بالاضافة الى التأخر في التئام الجروح وكذلك كثرة التبول والتعب والارهاق ، العطش والجوع الشديد ،جفاف الفم واضطراب الرؤيا(الربيعان،(2022).

2-2-3 عينة الدراسة و وصف بيانات الدراسة

تم جمع البيانات لعينة تتكون من(216) شخصاً مصاباً بمرض السكري ، وسبب اختيار نسبة السكر في الدم بوصفه متغيراً معتمداً(y) لان جميع هذه العوامل لها تأثير على نسبة السكر في الدم. استندت آليه جمع البيانات إلى استمارة ولقاءات مباشرة مع المرضى الموجدين في مستشفى الديوانية التعليمي/ مركز السكري والعيون ،إذ تم تنظيم استمارة لجمع البيانات ومن ثم تم جدولة بيانات هذه الاستمارة والجدول رقم(3-25) يوضح اسماء المتغيرات التوضيحية والمتغير المعتمد .

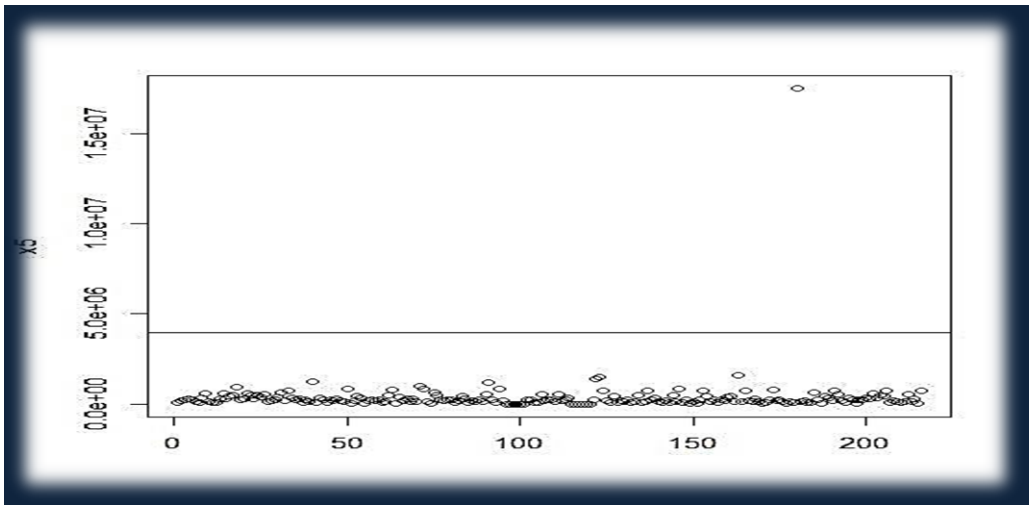
جدول رقم(3-25) يوضح المتغير المعتمد مع المتغيرات التوضيحية

أسم المتغير	الرمز
نسبة السكر في الدم	Y
العمر	X_1
الوزن	X_2
الطول	X_3
الدخل الشهري	X_5
الجنس	X_6
عدد افراد الاسرة	X_7
نسبة الكلسترول بالدم	X_8
يعاني من مشكلة نفسية	X_9
الحالة الزوجية	X_{10}
التعرض لصدمة	X_{11}
كاسب ام موظف	X_{12}
طبيعة الغذاء(نباتي، حيواني)	X_{13}
هل يعاني من اعاقة	X_{14}
فصيلة الدم	X_{15}
الاصابة بأمراض اخرى	X_{16}
الوراثة(هل مرض السكري وراثة بالعائلة)	X_{17}
التدخين	X_{18}
المستوى الدراسي	X_{19}
اليوريا	X_{20}
الكرياتين	X_{21}

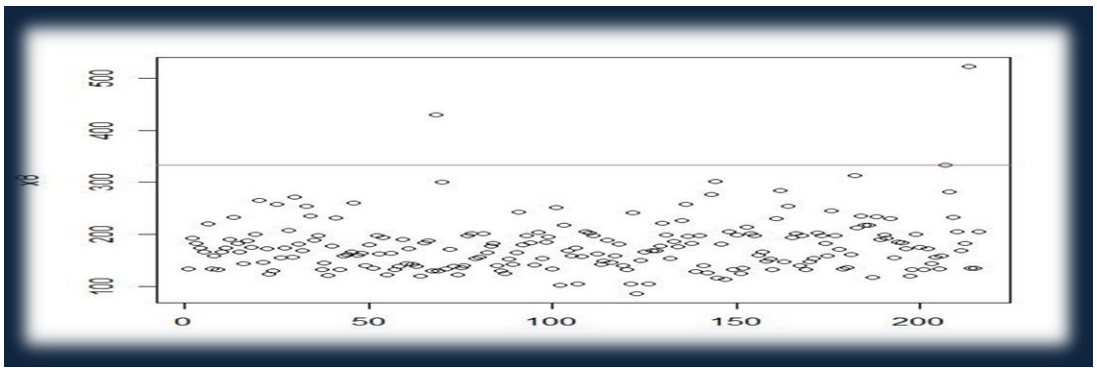
3-2-3 اختبار وجود القيم الشاذة للبيانات الحقيقية

لأجراء هذا الاختبار نستخدم معيار$(\mu \pm 3\sigma)$ لتحديد القيم الشاذة لكل متغيرفي البيانات الحقيقية. (2013,Lehmann).

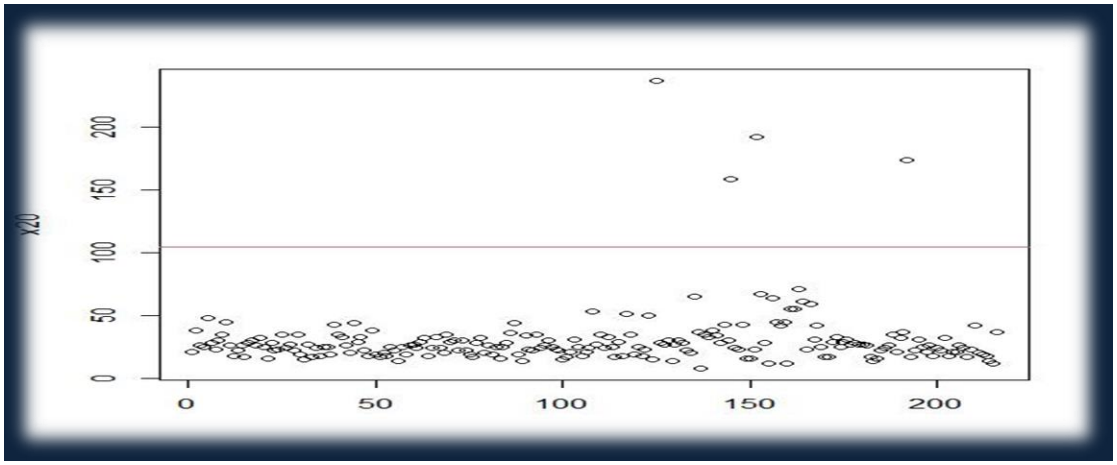
وندرج هنا رسم لشكل المتغيرات التي تضم قيماً شاذة وكالاتي:-



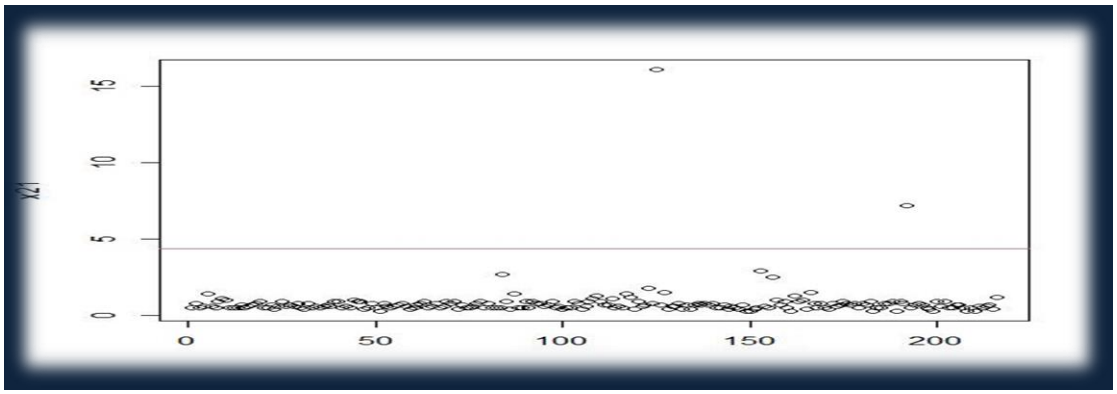
الشكل (1) يوضح القيم الشاذة في المتغيرX_5



الشكل (2) يوضح القيم الشاذة في المتغيرX_8



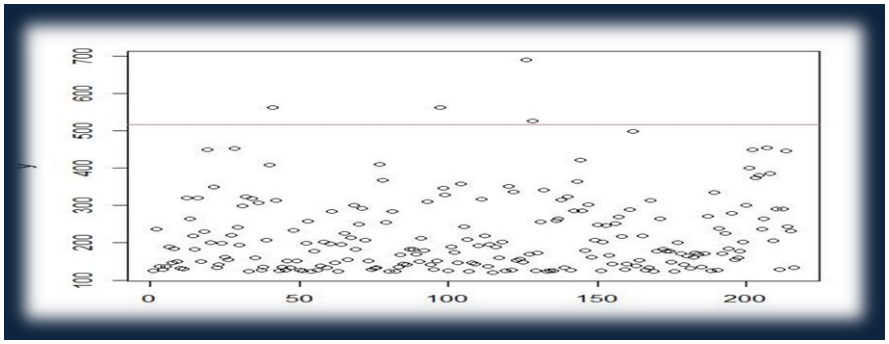
الشكل (3) يوضح القيم الشاذة في المتغير X_{20}



الشكل (4) يوضح القيم الشاذة في المتغير X_{21}

فقد اوضحت الاشكال اعلاه ان بعض المتغيرات كانت تضم قيماً شاذة مثل**X_5 وX_8وX_{20}و X_{21}** لهذا لا نحتاج الى القيام بتلويث البيانات.

وكذلك المتغير yيضم قيم شاذة وكما موضح بالشكل الاتي:-



الشكل (5) يوضح القيم الشاذة في المتغيرy

3-2-4 نتائج البيانات الحقيقية

من خلال تحليل البيانات الحقيقية لمرضى السكري ظهرت النتائج الموضحة بالجدول(3-26) قيمة معلمات نموذج الانحدار(Beta) .

الجدول(3-26) قيمة معلمات نموذج الانحدار

المتغيرات	SMAVE-AdEN	RSMAVE-EN	RSMAVE-AdEN
X_1	0.125907	0.120561	0.026029
X_2	-0.298978	0.091451	-0.016476
X_3	0.107531	0	0
X_5	0	0.207853	-0.977794
X_6	0.056054	-0.118131	0
X_7	0.181579	0.2541433	0
X_8	0.181579	0.2644151	-0.066873
X_9	0.203986	-0.136066	0
X_{10}	-0.170008	0.176567	-0.026008
X_{11}	-0.143819	0.359058	-0.036109
X_{12}	0.080801	-0.439647	0.013790
X_{13}	0.306188	-0.269757	0.013790
X_{14}	-0.043024	0.324821	-0.174384
X_{15}	-0.308703	0.134192	-0.028113
X_{16}	0.024399	0.041981	-0.013252
X_{17}	-0.071182	0.205518	.0252340
X_{18}	0.347748	-0.280087	0.032861
X_{19}	0.243710	-0.025385	0.033309
X_{20}	0.603145	-0.208113	0.035343
X_{21}	0	-0.213411	-0.016350

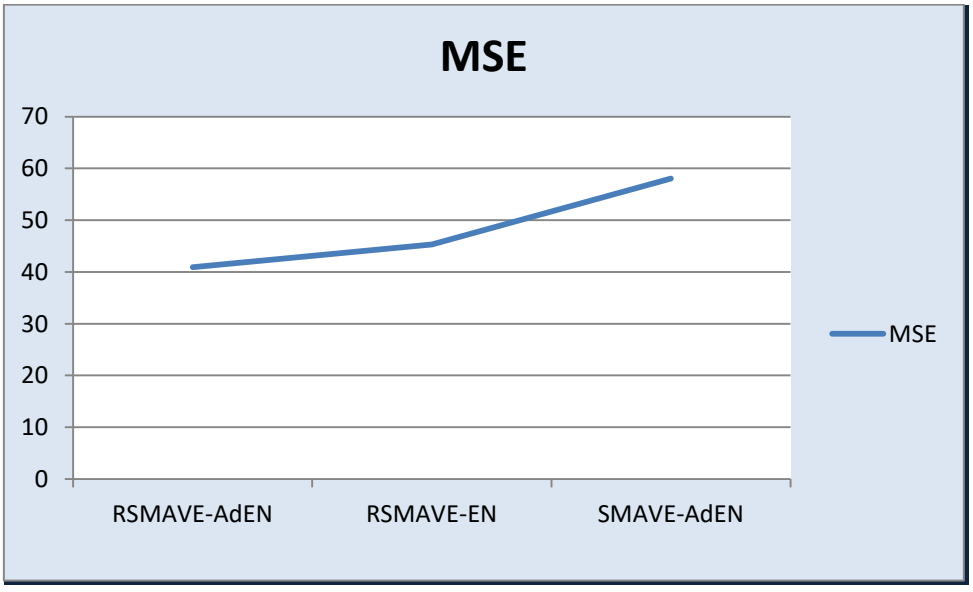
يوضح الجدول(3-26) ان عدد المتغيرات الغير مهمة(4) للطريقة المقترحة-RSMAVE AdENو(1)للطريقةRSMAVE-EN و(2)للطريقةSMAVE-AdEN .

الجدول رقم (3-27)يبين عدد الاصفار وMSE لتحليل البيانات الحقيقية لمرض السكري

method	عدد الاصفار	MSE
RSMAVE-AdEN	4	40.92771
RSMAVE-EN	1	45.30264
SMAVE-AdEN	2	58.02755

يوضح الجدول رقم (3-27) عدد الاصفار ومتوسط مربعات الخطأ MSE لتحليل البيانات الحقيقية لمرض السكري و للطرائق الثلاث إذ نلاحظ ان الطريقة المقترحةRSMAVE-AdEN لديها اداء افضل وتتفوق على الطرائق الاخرىRSMAVE-EN وSMAVE-AdEN فقد اعطت الطريقة المقترحة أقل متوسط مربعات الخطأ MSE وهذا سلوك جيد.

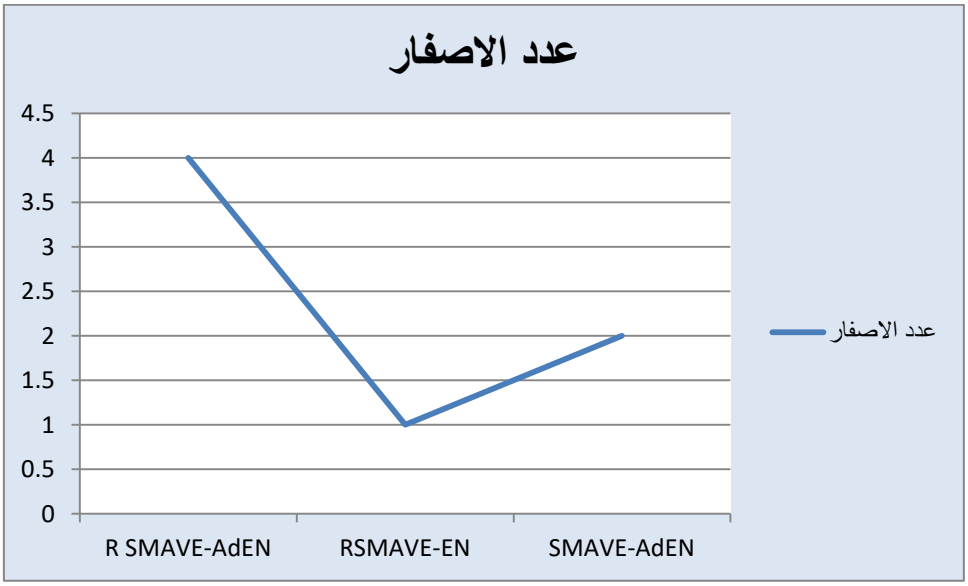
والشكل (6) يوضح قيمة MSE للطريقة المقترحةRSMAVE-AdEN و-RSMAVE EN وSMAVE-AdEN .



الشكل(6)

يبين قيمة MSE للطرائقRSMAVE-AdEN، RSMAVE-EN، SMAVE-AdEN

الشكل(7) يبين عدد المعاملات الصفرية للطرائقRSMAVE-AdEN، RSMAVE-EN ، SMAVE-AdEN



الشكل(7)

يبين عدد المعاملات الصفرية للطرائقRSMAVE-AdEN، RSMAVE-EN، SMAVE-AdEN

الفصل الرابع

Chapter four

الاستنتاجات والتوصيات والعمل المستقبلي

Conclusions , Recommendations and future works.

الفصل الرابع

الاستنتاجات والتوصيات والعمل المستقبلي

1-4 الاستنتاجات Conclusions

من خلال ما تم عرضه في الجانب التجريبي والجانب التطبيقي ومن خلال نتائج تحليل البيانات الحقيقية التي تم الحصول عليها نبين أهم الاستنتاجات وهي ما يأتي:

1-ان الطريقة المقترحة RSMAVE-AdENفي هذة الرسالة هي طريقة حصينة لاختيار المتغيرات وتقليل الابعاد في وقت واحد.

2- تتمتع هذه الطريقة بالكفاءة عندما تكون الارتباطات عاليه بين المتغيرات ضمن إعدادات SDR.

3- أظهرت النتائج لكل من عمليات المحاكاة وتحليل البيانات الحقيقية أن طريقة RSMAVE-AdEN المقترحة لها أداء جيد في اختيار المتغير ودقة تقدير حتى مع وجود القيم الشاذة في المتغيرxومتغير الاستجابةy، فقد حلت طريقتنا بالمرتبة الاولى من حيث دقة التقدير بناءً على معيار المقارنة متوسط مربعات الخطأ MSE و اعطت أقل MSEلجميع حالات التلويث(5%،10%،15%،20%)ولأحجام العينات (50،100،200) ، لكن كلما ارتفعت نسبة التلويث ارتفعت قيمة MSE.

4-2 التوصيات والعمل المستقبلي Recommendations and future work

من خلال ماتم عرضه في الجانب التجريبي والجانب التطبيقي ومن خلال النتائج التي تم الحصول عليها والاستنتاجات نضع التوصيات الآتية:-

1- توصي الدراسة باستخدام طريقة RSMAVE-AdEN المقترحة في تحليل البيانات خاصة عند وجود قيم شاذة في المتغيرات المستقلةx والمتغير المعتمدy.

2- توصي الدراسة بأستخدام الطريقة المقترحةRSMAVE-AdEN في حالة التعامل مع البيانات عالية الابعاد وفي مختلف المجالات وذلك بسبب كفاءتها في الحصول على تقديرات موثوقة وتنبؤ دقيق.

3- توصي الدراسة بضرورة اعتماد التوثيق الالكتروني لتسجيل نتائج تحاليل المرضى في المستشفيات من اجل توفير قاعدة بيانات لتسهيل عملية جمع البيانات من قبل الباحثين.

4- توصي الدراسة بأهمية تزويد حواسيب مركزية بقدرات تشغيلية عالية جدا للمساعدة في تطبيق هكذا طرائق، لاسيما مع زيادة عدد المتغيرات التوضيحية وزيادة أحجام العينات والتكرارات لتنفيذ العملية ، إذ ان هذه الطرائق تستغرق الكثيرمن الوقت للتنفيذ مع أجهزة الكمبيوتر العادية .

5- يمكن توسيع فكرة الطريقة المقترحة لتشمل طرقاً أخرى ضمن SDRمثل SAVEوSIR.

المصـادر

References

المصادر (References)

اولاً:- المصادر العربية

[1]الراوي، خاشع محمود(1987)"المدخل الى تحليل الانحدار" مديرية دار الكتب للطباعة والنشر، جامعة الموصل العراق.

[2] الربيعان، خالدعلي . (2022). داء السكري يكشف التفوق الطبي للحضارة الإسلامية على الحضارات الإنسانية القديمة: دراسة تاريخية لأربعة آلاف سنة Arabian Journal of Scientific . Research, 2022(1).

ثانياً:- المصادر الاجنبية

.

[3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In second International Symposium on Information Theory. Akademia Kiado, Budapest, 267-281.

[4] Alkenani, A. (2020). Robust variable selection in sliced inverse regression using Tukey biweight criterion and ball covariance. Journal of Physics Conference Series, 1664, 012034.

[5] Alkenani, A. (2021). Robust group identification and variable selection in sliced inverse regression using Tukey's biweight criterion and ball covariance. Gazi University Journal of Science 35 (2).

[6] Alkenani, A. and Abdulkadhim, M. (2020). Regularized sliced inverse regression through the elastic net penalty. Journal of Physics Conference Series. Submitted.

[7] Alkenani, A. and Aljobori, N. (2021). Robust sparse MAVE through elastic net penalty. International journal of Agricultural and Statistical Sciences, Vol.17, Supplement 1, 2039 - 2046.

[8] Alkenani, A. and Dikheel, T. (2017). Robust Group Identification and Variable Selection in Regression. Journal of Probability and Statistics 2017, Article ID 2170816, 8 pages.

[9] Alkenani, A. and Rahman, E. (2020). Sparse minimum average variance estimation via the adaptive elastic net when the predictors correlated, Journal of Physics Conference Series, 1591, 012041.

[10] Alkenani, A. and Rahman, E. (2021). Regularized MAVE through the elastic net with correlated predictors, Journal of Physics Conference Series, 1897, 012018.

[11] Alkenani, A. and Reisan, T (2016). Sparse sliced inverse quantile regression. Journal of Mathematics and Statistics. Volume 12, Issue 3.

[12] Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34, 85–105.

[13] Bellman, R. E. (1961). Adaptive Control Processes. Princeton University Press, Princeton, New Jersey.

[14] Bondell, H. D. and Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR," Biometrics, 64, 115-123.

[15] Breiman, L. (1996). Heuristics of instability and stabilization in model selection. The Annals of Statistics, 24(6), 2350–2383.

[16] Brillinger, D. R. (1983). A generalized linear model with (Gaussian) regression variables. In A Festschrift for Erich L. Lehmann (eds P. J. Bickel, k. A. Doksum and J. L. Hodges, Jr), 97-114 Belmont: Wadsworth.

[17] Carlos, A. M. and Sergioc, C. S. (2012). Does BIC Estimate and Forecast Better than AIC. Available at (https://mpra.ub.uni-muenchen. de/42235/).

[18] Chand , S., and Kamal , S .(2011) , " variable selection by lasso – type method " . Pakistan Journal of statistics and operation research , 451-464.

[19] Cizek,P. and Hardle, W. (2006). Robust estimation of dimension reduction space. Computational Statistics and data analysis, 51, 545-555.

[20] Common, P. (1994). Independent component analysis, a new concept?. Signal Processing, 36(3), 287–314.

[21] Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York, Wily.

[22] Cook, R. D. and Li, B. (2002). Dimension reduction for the conditional mean in regression. The Annals of Statistics 30, 455–474.

[23] Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). Journal of the American Statistical Association 86, 328–332.

[24] Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. Econometrics, 6(4), 45

[25] Donoho, D. L., and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425–455.

[26] Efroymson, M. A. (1960). Multiple regression analysis. Mathematical Methods for Digital Computers, 191–203.

[27] Efron, B. et al. (2004). Least angle regression. The Annals of Statistics 32, 407–499.

[28] Fan, J. and Li, R. Z. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96,1348–1360.

[29] Gorsuch, R. L. (1983). Factor Analysis, Hillsdale, New Jersey, L. Erlbaum Associates.

[30] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157 – 1182.

[31] Hassel, M. (2021). Sparse sliced inverse regression via elastic net penalty with an application. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

[32] Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and ℓ1 penalized regression: A review. Statistics Surveys, 2, 61-93.

[33] Härdle, W. and Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 84(408), 986–995.

[34] Horowitz ,J.L., and Lee ,S . (2002) , " semi-parametric methods in applied econometrics " . statistical modeling , 2 , 3-22.

[35] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58(1–2), 71–120.

[36] Jabbar, E. (2020). A non-linear multi-dimensional estimation and variable selection via regularized MAVE method. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

[37] Jolliffe, I. T. (2002). Principal components in regression analysis. Principal Component Analysis, 167–198.

Kong , E ., Xia , Yi . (2007) , " variable selection for the single index model " . Biometrika 94 , 217-229.

[38] Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–327.

[39] Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein‟s lemma. Journal of the American Statistical Association 87, 1025–1039.

[40] Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94, 603–613.

[41] Li, L., Cook, R. D. and Nachtsheim, C. J. (2005). Model-free variable selection. Journal of the Royal Statistical Society Series B, 67, 285–299.

[42] Li, L., Li, B. and Zhu, L.-X. (2010). Groupwise dimension reduction. Journal of the American Statistical Association. 105, 1188–1201.

[43] Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics 48, 503–510.

[44] Li, L. and Yin, X. (2008). Sliced Inverse Regression with regularizations. Biometrics 64, 124–131.

[45] Lehmann, I. R. (2013). The 3σ-rule for outlier detection from the viewpoint of geodetic adjustment.

[46] Malik, D. (2019). Sparse dimension reduction through penalized quantile MAVE with application. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

[47] Meier, L., Van De Geer, S., Bühlmann, P. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1), 53-71.

[48] Ni, L. et al. (2005). A note on shrinkage sliced inverse regression. Biometrika 92, 242–247.

[49] Powell, J. et al. (1989). Semiparametric estimation of index coefficients. Econometrica: Journal of the Econometric Society, 1403–1430.

[50] Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of s-estimators. In Robust and Nonlinear Time Series Analysis, 256-272.

[51] Salman, D. (2021). Sparse dimension reduction via regularized sliced inverse regression with application. Thesis submitted to college of administration and economics. University of Al-Qadisiyah. Iraq.

[52] Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464.

[53] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, 58, 267–288.

[54] Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and statistics, 2:448-485.

[55] Wand, M. P. and Jones, M. C. (1995) Kernel Smoothing. Chapman and Hall, London. http://dx.doi.org/10. 1007/978-1-4899-4493-1.

[56] Wang, Q. and Yin, X. (2008). A Nonlinear Multi-Dimensional Variable Selection Method for High Dimensional Data: Sparse MAVE. Computational Statistics and Data Analysis 52, 4512–4520.

[57] Wang, Q. and Yao, W. (2013). Robust Variable Selection through MAVE. Computational Statistics and Data Analysis 63, 42-49.

[58] Wang, T. et al. (2013). Penalized minimum average variance estimation. Statist. Sinica 23, 543–569.

[59] Wang, T. et al. (2015). Variable selection and estimation for semi parametric multiple-index models. Bernoulli 21 (1), 242–275.10

[60] Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics, 35(6), 2654–2690.

[61] Xia, Y. (2008). A multiple-index model and dimension reduction. Journal of the American Statistical Association, 103(484), 1631– 1640.

[62] Xia, Y. et al. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society Series B 64, 363–410.

[63] Yin, X. and Cook, R. D. (2005). Direction estimation in single index regressions. Biometrika, 92(2), 371–384. ˇ

[64] Yin, X. et al. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. Journal of Multivariate Analysis, 99(8), 1733–1757.

[65] Yu, Z. and Zhu, L. (2013). Dimension reduction and predictor selection in semi parametric models. Biometrika, 100, 641-654.

[66] Zhang, C. H. (2010). Nearly unbiased variable selection under Minimax Concave Penalty. Annals of Statistics 38, 894–942.

[67] Zhang, J. and Olive, D. J. (2009). Applications of a robust dispersion estimator. Southern Illinois University Carbondale.

[68] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418–1429.

[69] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society, Series B 67, 301–320.

[70] Zou, H., and Zhang, H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of Statistics, 37(4), 1733.

ABSTRACT

If the number of variables is large, then the regression analysis is a difficult process. In other words, increasing the number of variables leads to an increase in the complexity of the model, and this may lead to the problem of dimensionality, which may appear by increasing the number of variables and increasing the size of the data.

It is difficult to formulate a parametric model for a large number of variables. This problem led researchers to work on reducing these high dimensions of the data. Some explanatory variables do not have a significant effect on the dependent variable, as well as some of these variables have an internal correlation with each other, and this requires the exclusion of such variables in order to increase the accuracy of the model.

There are two ways to reduce the dimensions, namely the method of selecting variables (v.s) variable selection and variables extractions. under the assumptions of the theory of SDR (Sufficient dimension reduction) the researchers worked on proposing methods to reduce the dimensions, including the integration of SDR methods with regularization methods (Regularization method) And the methods of regulation mean adding a penalty limit to control the complexity of the model as it greatly reduces the variance of the model, and among these methods SMAVE-AdEN (Alkenani and Rahman,2020) is a method for selecting a variable under the assumptions of SDR theory.

The SMAVE-AdEN method is a combination of Adaptive elastic net with MAVE (Minimum average variance estimator) method for estimating minimum average variance. This method is effective when the variables are highly correlated under SDR assumptions. But the SMAVE-AdEN method is not immune and it is a sensitive method that is affected when there are outliers in the data, because we use the least squares criterion.

Here we propose a method for selecting a vulnerable variant under SDR assumptions called (RSMAVE-AdEN). It is not affected by outliers found in both the explanatory and response variables.

The efficiency of the proposed method was verified by simulation and using real data.

Republic of Iraq
Ministry of Higher Education
And Scientific Research
University of Al-Qadisiyah
College of Administration and Economics
Department of statistics

Robust sparse minimum average variance estimation with application

A thesis submitted to
The council of the college of Administration and Economics at University of Al-Qadisiyah as partial Fulfillment of the requirements for the M.S.c in Statistics

by

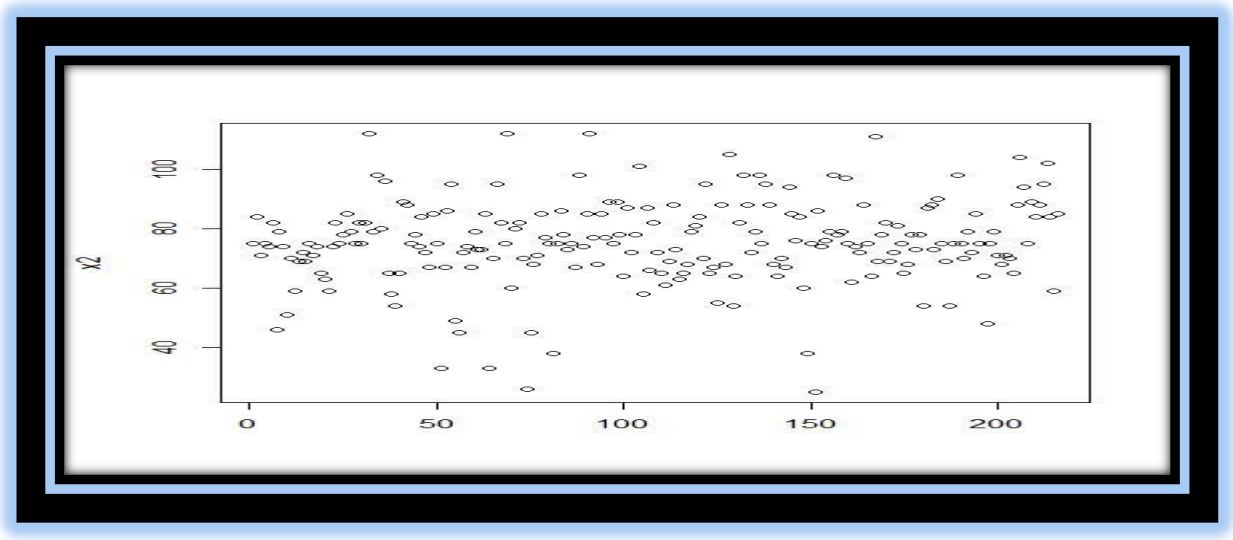
Sanaa Jabbar Tuama

Supervised by
Prof. Dr. Ali J. Kadhim Alkenani

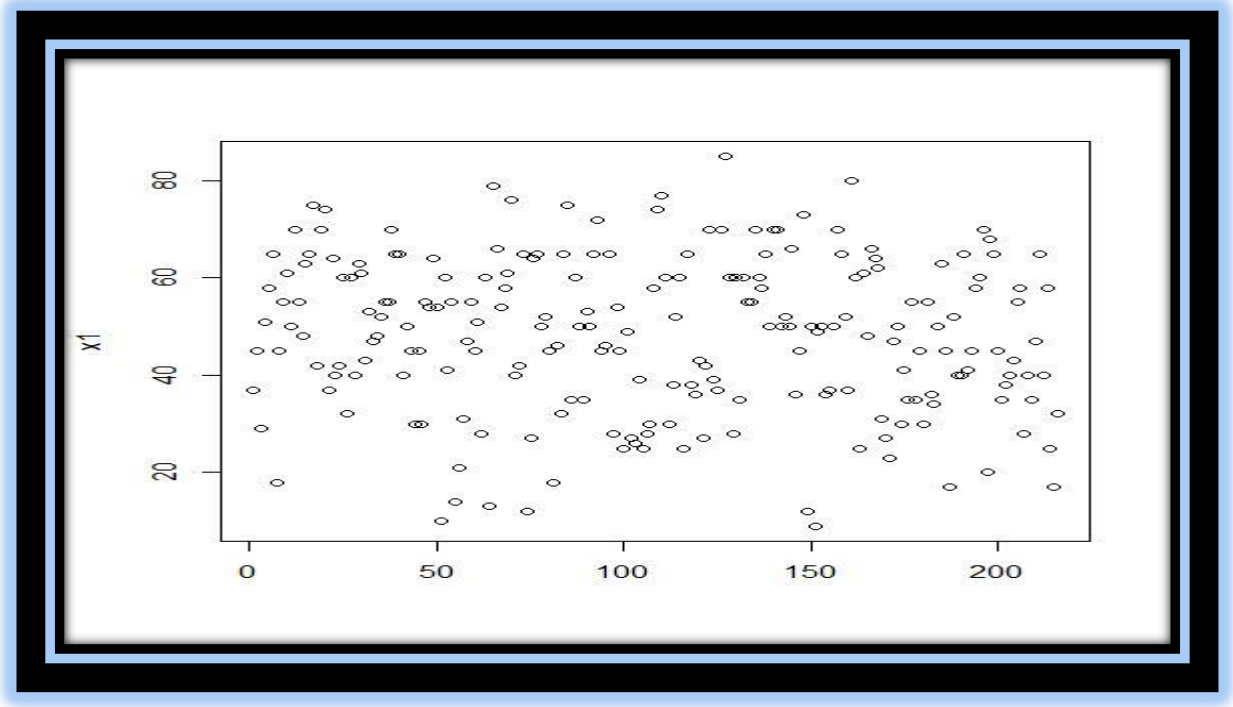
1444 A.H 2023 A.D

الملاحق

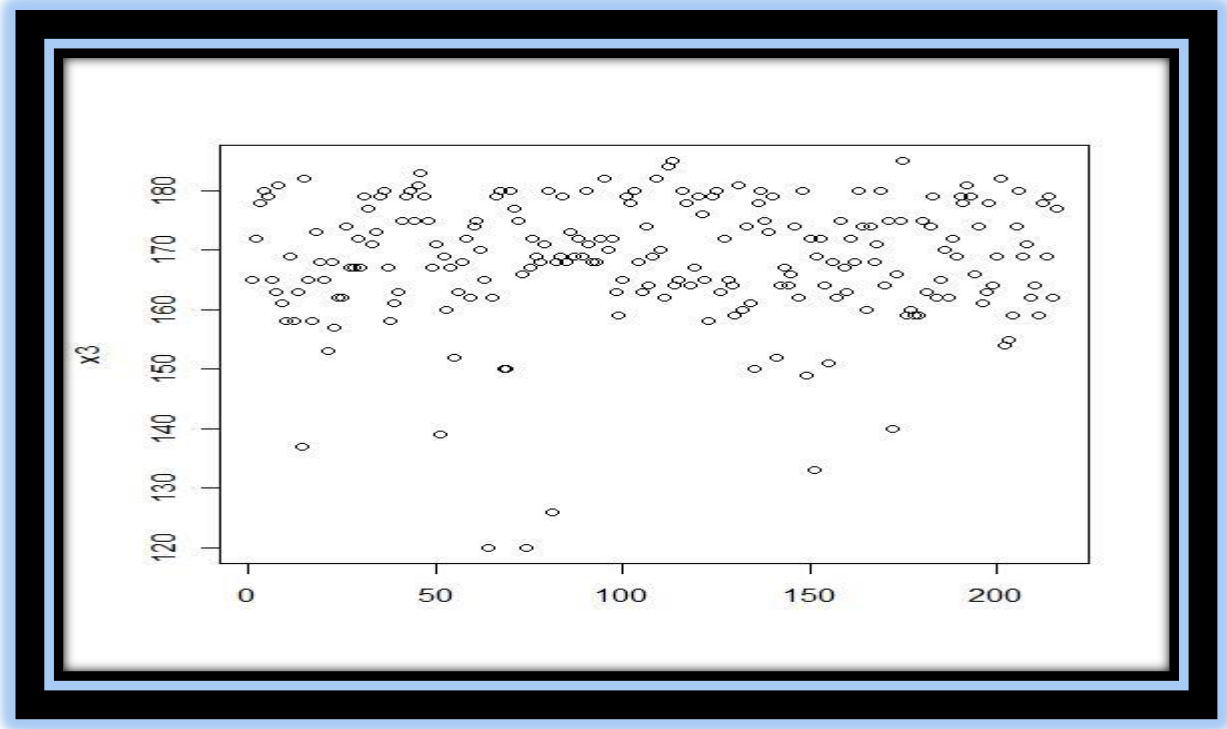
الاشكال الآتية تبين المتغيرات التي لا تضم قيم شاذة وكا موضح ادناه:-



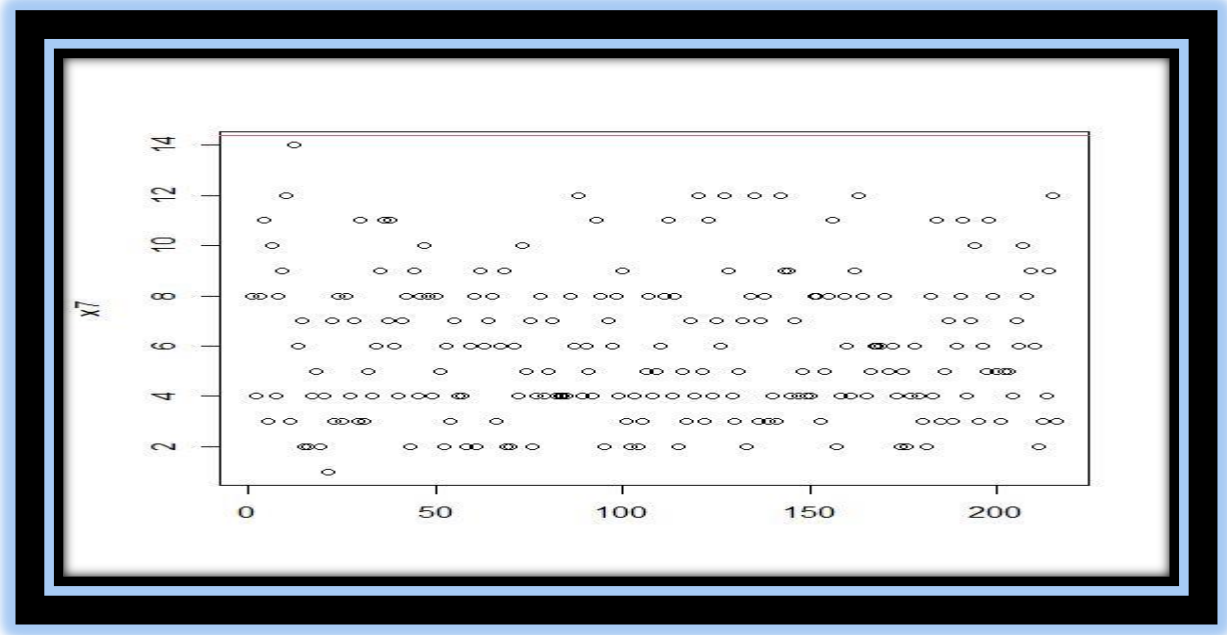
الشكل (1) يوضح البيانات في المتغير x_2



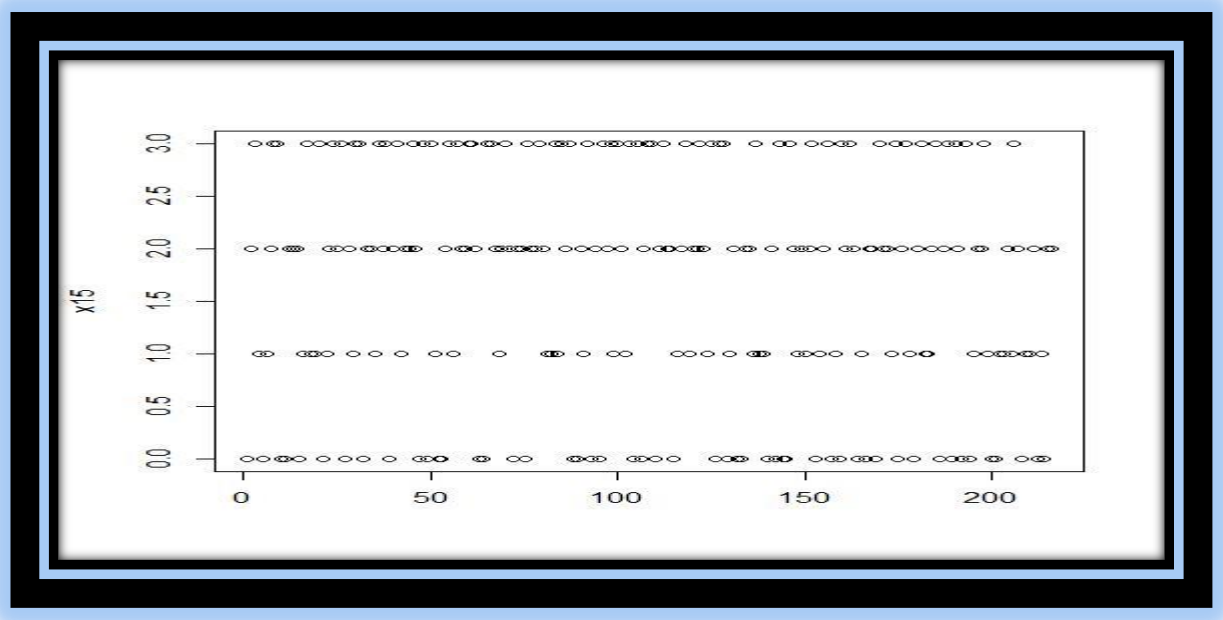
الشكل (2) يوضح البيانات في المتغير x_1



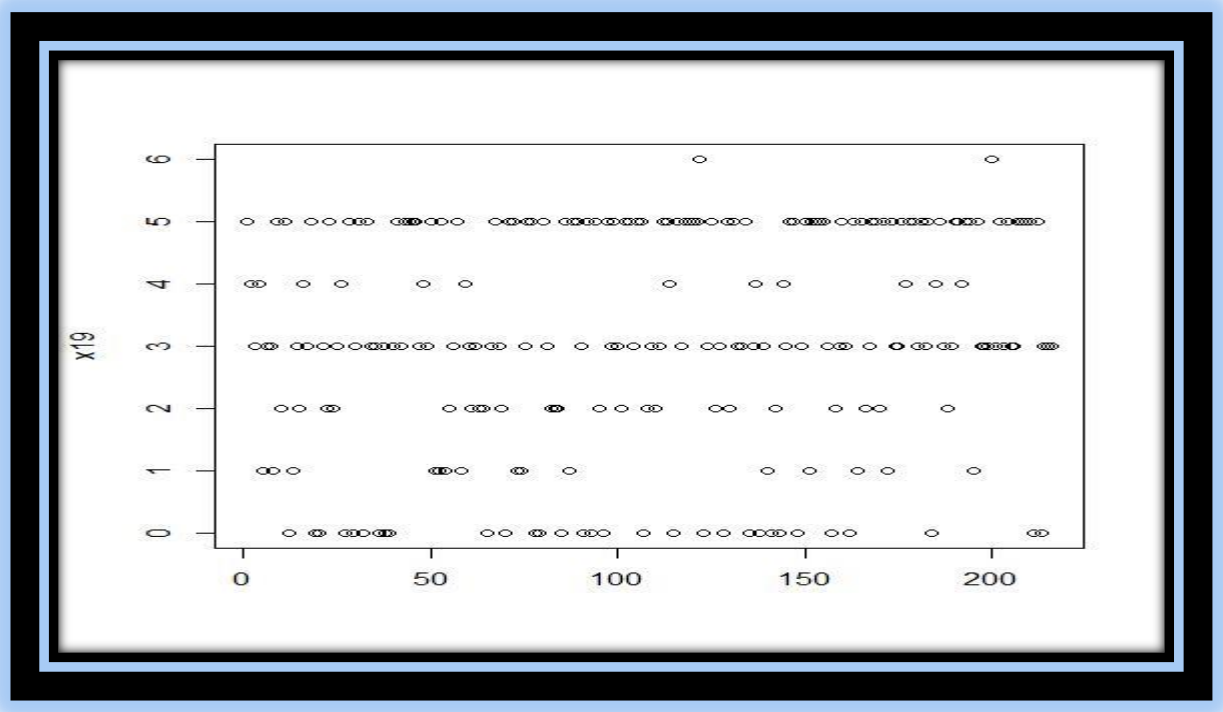
الشكل (3) يوضح البيانات في المتغير x_3



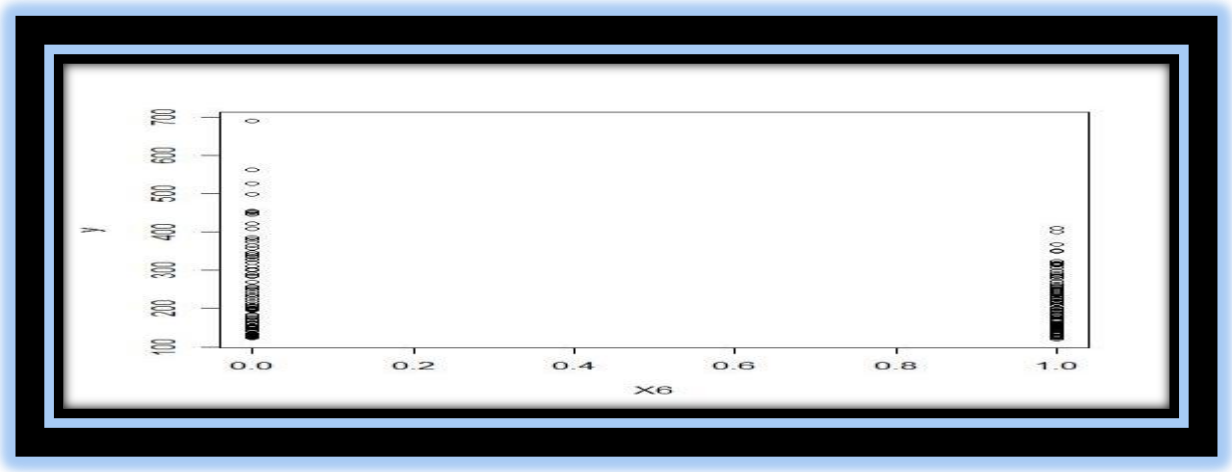
الشكل (4) يوضح البيانات في المتغير x_7



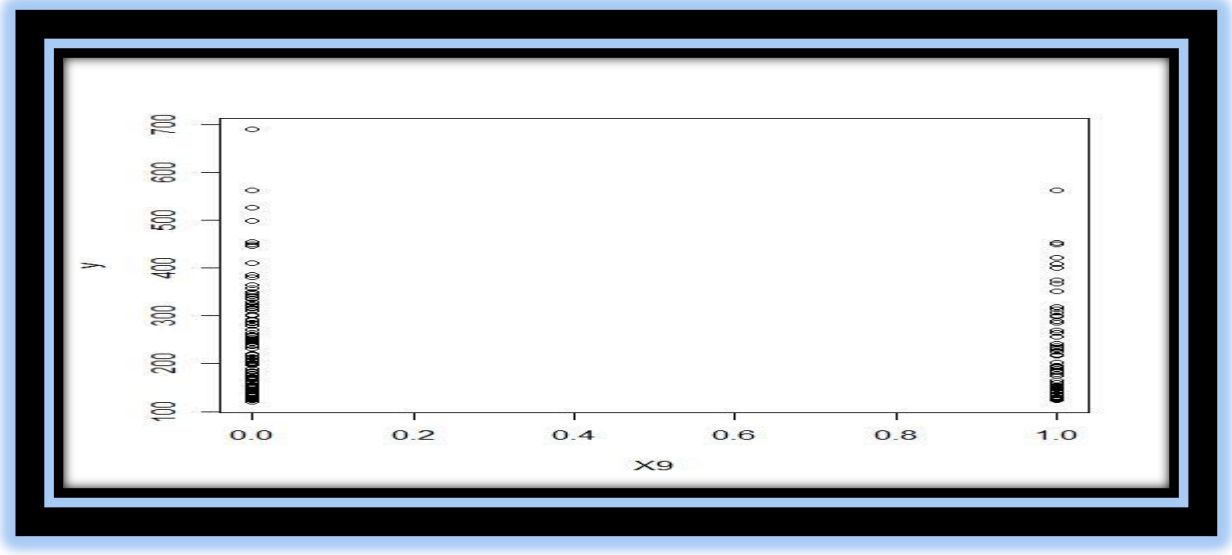
الشكل (5) يوضح البيانات في المتغير x_{15}



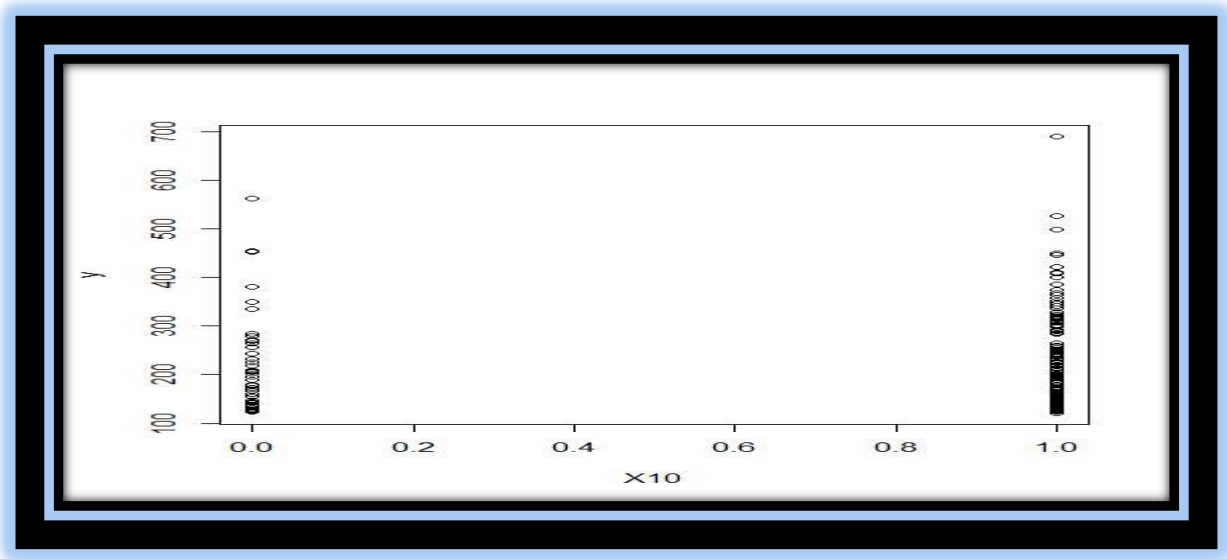
الشكل (6) يوضح البيانات في المتغيرx_{19}



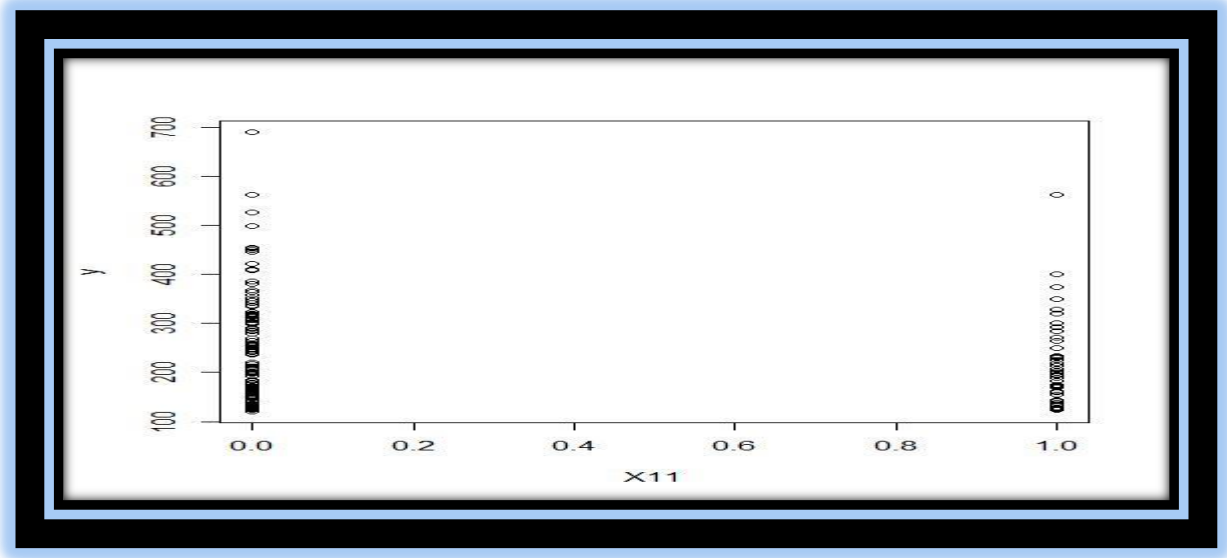
الشكل (7) يوضح البيانات في المتغير x_6



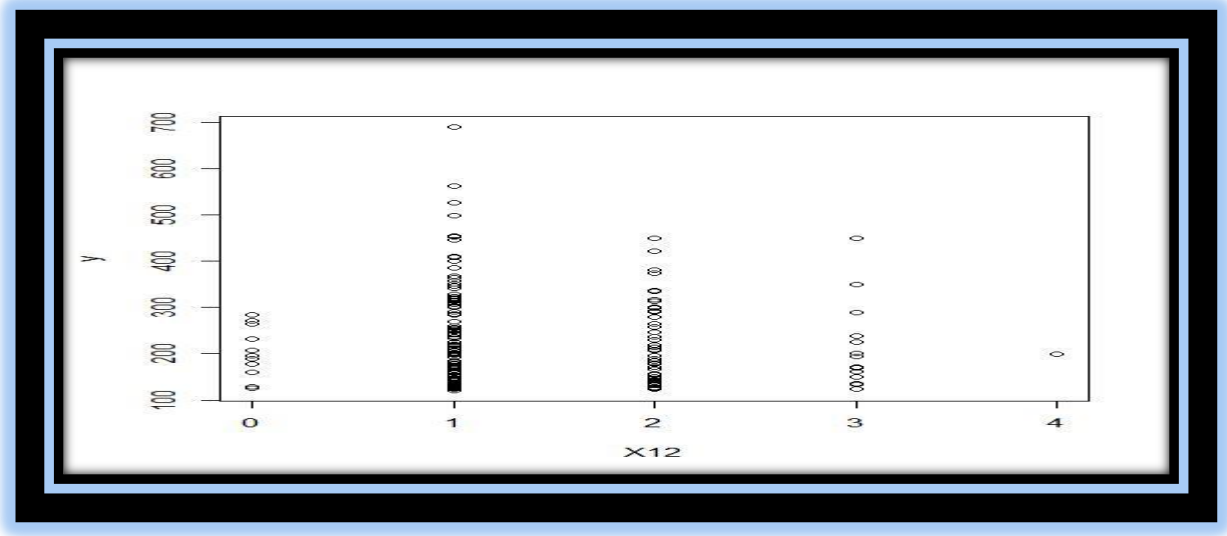
الشكل (8) يوضح البيانات في المتغير x_9



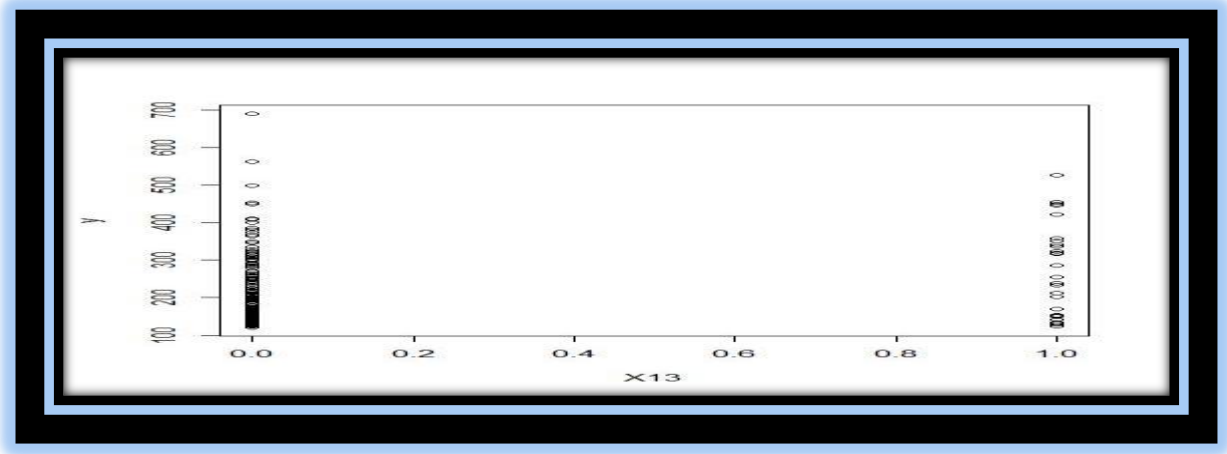
الشكل (9) يوضح البيانات في المتغير x_{10}



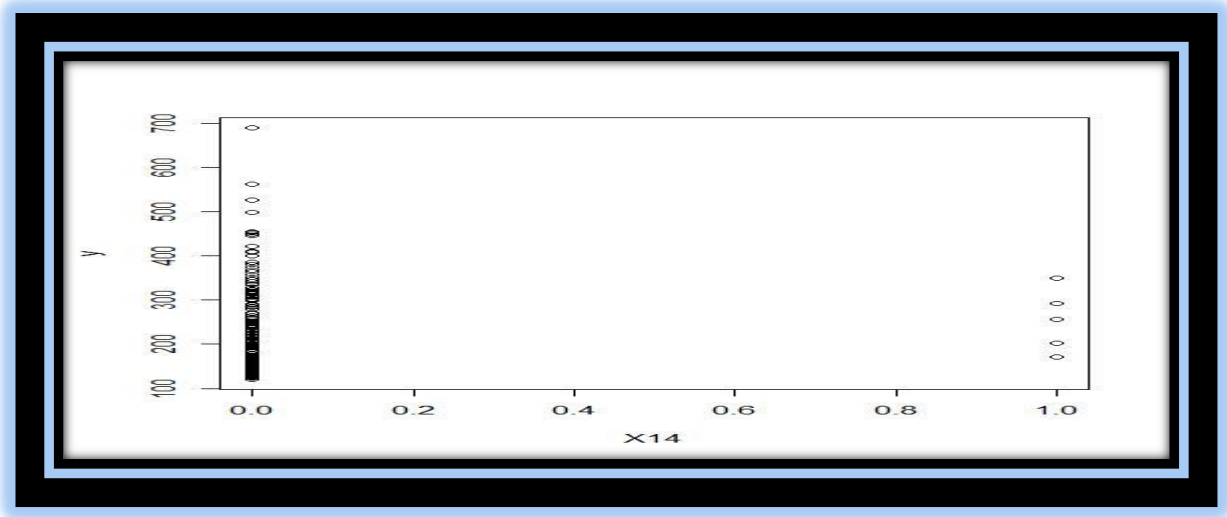
الشكل (10) يوضح البيانات في المتغير x_{11}



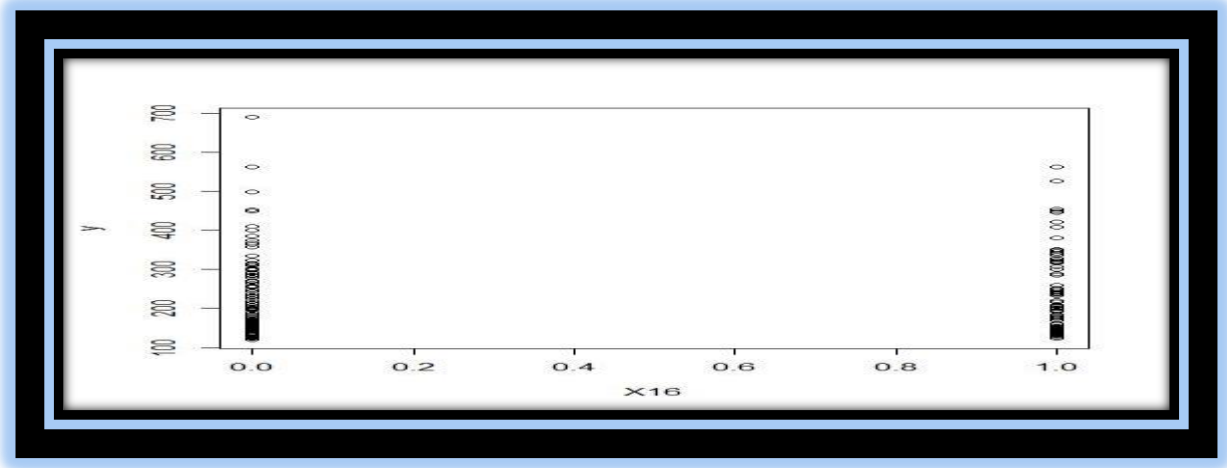
الشكل (11) يوضح البيانات في المتغير x_{12}



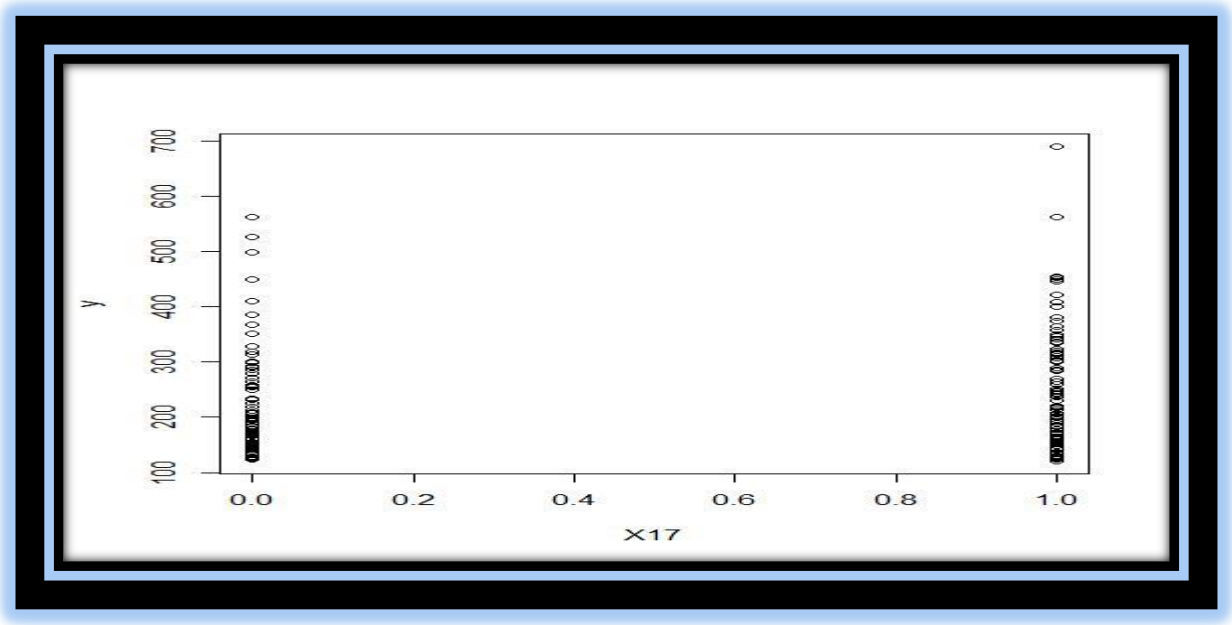
الشكل (12) يوضح البيانات في المتغير x_{13}



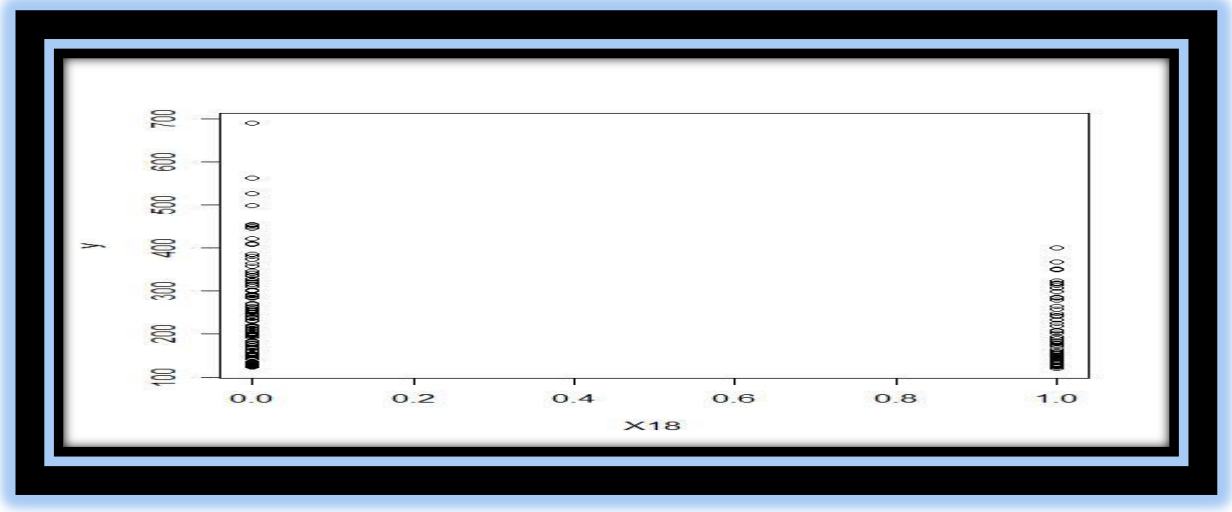
الشكل (13) يوضح البيانات في المتغير x_{14}



الشكل (14) يوضح البيانات في المتغير x_{16}



الشكل (15) يوضح البيانات في المتغيرx_{17}



الشكل (17) يوضح البيانات في المتغيرx_{18}