



معالجة مشكلة التعدد الخطي في الانموذج المعلمي باستخدام بعض طرائق الانكماش

Addressing the problem of Multicollinearity in parametric model using some shrinkage methods

م.م. شرين علي حسين
جامعة بغداد

م.م. سكيئة شامل جاسم
جامعة واسط – كلية الادارة والاقتصاد

المستخلص :

ان مشكلة التعدد الخطي (Multicollinearity) هي مشكلة مؤثرة في تحليل الانحدار و ان طريقة المربعات الصغرى الاعتيادية (OLS) قد تؤدي الى تباين كبير في تقديرات معاملات الانحدار في ضل وجود مشكلة التعدد الخطي ولمعالجة هذه المشكلة تم استخدام بعض طرائق الانكماش (Shrinkage Methods) لتقدير انموذج الانحدار الخطي العام و هي (طريقة Lasso وطريقة Elastic- Net) وهذه الطرائق تعمل على اختزال التغير في التقديرات من خلال انكماش المعاملات و في نفس الوقت تنتج نماذج قابلة للتفسير بواسطة انكماش بعض المعاملات وبالضبط الى الصفر . وفي هذا البحث نبين اداء هذه الطرائق في ظل وجود مشكلة التعدد الخطي (Multicollinearity) من خلال تطبيقها على بيانات حقيقية والتوصل الى الطريقة الافضل بالاعتماد على معيار المقارنة متوسط مربعات الخطأ (MSE) اذ حققت طريقة (Elastic Net) افضليتها و تم الحصول على كافة النتائج بالاعتماد على البرنامج الاحصائي (SPSS).

المصطلحات : طرائق الانكماش , مشكلة التعدد الخطي , طريقة لاسو , طريقة الستيك- نت .

Abstract

Multicollinearity has been a serious problem in analysis of Regression , The ordinary least squares method (OLS) may result in high variability in the estimates of the regression coefficients in the presence of multicollinearity .To address this problem using some shrinkage methods for estimation general linear model (GLM) it's (Lasso and Elastic- Net methods) and this methods reduces the variability of the estimation by shrinkage the coefficients and at the

same time produces interpretable models by shrinkage some coefficients to exactly zero. In this research show performance these methods in serious multicollinearity by Application on real data and reach to best method based on mean squares error (MSE) and it's (Elastic- Net) method . All results were obtained depend on (SPSS) program .

keywords: Shrinkage Methods , Multicollinearity , Lasso method, Elastic- Net method

المقدمة :

ان طرائق الانكماش (Shrinkage methods) تحظى باهتمام كبير من قبل الباحثين لخصائصها الجيدة كقدرتها على اختيار المتغير المعنوي واستبعاد المتغيرات غير المهمة من النموذج الانحدار. كما تعمل على دمج بين دقة التنبؤ وتقليل بناء النماذج . ان اساليب التقدير الاعتيادية مثل طريقة المربعات الصغرى (OLS) تميل الى اداء ضعيف في حالة مشكلة الابعاد العالية و مشكلة التعدد الخطي (Multicollinearity) على رغم من ان مقدرات (OLS) تعطى تحيز (Baise) منخفض وتباين عالي للتنبؤ و غالباً يكون من الصعب تفسيرها.

ففي مثل هذه الحالات غالباً ما يلجأ الباحث الى استعمال طرائق الانكماش (Shrinkage methods) لمعالجة هذه المشاكل , اي انكماش المقدر نحو الصفر وهذا في الواقع يتضمن ادخال بعض التحيز من اجل تقليل تباين التنبؤ نتيجة تقليل متوسط مربعات الخطأ و هذا يؤدي الى تحسين دقة التنبؤ بشكل عام فضلاً عن سهولة التفسير اذ ان في كثير من الاحيان نرغب في تحديد اصغر مجموعة جزئية والتي تؤدي الى اقوى التأثيرات.

لذا تم استخدام في هذا البحث بعض الاساليب الاحصائية التي تعمل على انكماش بعض المعاملات ويضع الاخرى مساوية الى صفر من اجل ابقاء على ميزات جيدة لكل من اختيار افضل مجموعة جزئية وعملية الاستمرار بانكماش (Shrinkage) المعاملات و بالتالي تكون هذه الطرائق اكثر استقراراً و من ثم تعطي انموذجاً قابل للتفسير بسهولة .

1. طريقة لاسو Lasso Method: و هذه الطريقة مقترحة من قبل الباحث (Tibshirani) عام (1996) Lasso (absolute shrinkage and selection operator .

2. طريقة *elastic – net* : وهي طريقة مقترحة من قبل الباحثان (Hastic & Zou) في عام (2005)

2. مشكلة البحث :

ان ادخال عدد كبير من المتغيرات التوضيحية في معادلة الانحدار لأسى دراسة غالباً ما تحتوي على متعدد خطي Multicollinearity كما ان بعض المتغيرات تكون غير اساسية في تأثيرها على المتغير المعتمد او يكون تأثيرها مماثل لتأثيرات متغيرات اخرى مما يؤدي الى جعل تأثيرها غير معنوي مما يستوجب استبعاد مثل هذه المتغيرات و هنا تكمن

صعوبة مسألة تحليل الانحدار التي تجعل الباحث يحتاج الى الموازنة بين تقليل المتغيرات التوضيحية تجنباً لزيادة تكاليف الحصول على المعلومات المطلوبة , وبين زيادة المتغيرات التوضيحية من اجل الوصول على نتائج تنبؤية افضل .

3-هدف البحث

يهدف هذا البحث الى معالجة مشكلة التعدد الخطي في البيانات من خلال استخدام بعض طرائق الانكماش (Shrinkage Methods) للتخلص من هذه المشكلة عن طريق اختار المتغيرات المعنوية فقط لغرض الحد من المتغيرات التوضيحية التي لا تحتوي على المعلومات ذات الصلة وبالتالي تحسين النمذجة الاحصائية فضلاً عن زيادة دقة التنبؤ واختزال الكثافة الحسابية , والتوصل الى افضل طريقة من بين الطرائق المستعملة في معالجة هذه المشكلة .

4-فرضيات البحث

يستند هذا البحث على الفرضيات التالية :

الفرضية الاولى (H_0) : لا توجد علاقة ايجابية بين (X_1 : نسبة اليوريا في الدم و نسبة الكلوسترول في الدم Y_i) عند مستوى المعنوية $\alpha = 0.05$.

الفرضية الثانية (H_0) : لا توجد علاقة ايجابية بين (X_2 : نسبة الكرياتين في الدم و نسبة الكلوسترول في الدم Y_i) عند مستوى المعنوية $\alpha = 0.05$.

الفرضية الثالثة (H_0) : لا توجد علاقة ايجابية بين (X_3 : نسبة الدهون الثلاثية في الدم و نسبة الكلوسترول في الدم Y_i) عند مستوى المعنوية $\alpha = 0.05$.

الفرضية الرابعة (H_0) : لا توجد علاقة ايجابية بين (X_4 : عمر المريض و نسبة الكلوسترول في الدم Y_i) عند مستوى المعنوية $\alpha = 0.05$.

الفرضية الخامسة (H_0) : لا توجد علاقة ايجابية بين (X_5 : نسبة السكر في الدم و نسبة الكلوسترول في الدم Y_i) عند مستوى المعنوية $\alpha = 0.05$.

5.الجانب النظري

Shrinkage Methods

5.1- طرائق الانكماش [1],[2]

ان جميع طرائق الانكماش تعمل على تقدير متجه معاملات الانحدار β للمعادلة الاتية :

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i \quad , I = 1,2,3, \dots, n \quad \text{-----} (1)$$

من خلال تقليل دالة الهدف F والتي تتكون من مجموع دالة الخسارة ($Loss function$) والتي تمثل مجموع مربعات الاخطاء Sum و دالة الجزاء ($Penalty function$) اي ان ;

$$F_{\lambda,r}(\beta) = \underset{\beta}{\operatorname{argmin}} \left\{ \underbrace{\sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2}_{\text{loss function=SSE}} + \underbrace{\sum_{j=1}^p p_{\lambda,y}(\beta)}_{\text{penalty function}} \right\} \text{----- (2)}$$

اذ ان :

$$\beta = (\beta_1, \beta_2, \dots, \beta_p)^T \text{ تمثل دالة لمتجه المعلمات } p_{\lambda,r}(\cdot)$$

λ : تمثل معلمة الجزاء (التنظيم) Tunning (penalty) parameter و قيمتها $\lambda > 0$ تعمل على السيطرة (المفاضلة بين تقليل دالة الخسارة و حدود الجزاء) .

r : تمثل معلمة الانكماش Shrinkage parameter و قيمتها اكبر من الصفر لتحديد رتبية (order) دالة الجزاء .

ان تقليل دالة الهدف F يؤدي الى (مجموعة) من الحلول التي تعتمد على قيمة معلمة الجزاء (λ) .

ان المشتقة الاولى لدالة الجزاء تحدد كيف تؤثر على الحل في دالة الهدف , ففي طريقة لاسو (Lasso) تكون ($r=1$) معلمة الانكماش و نسبة الجزاء تساوي $\hat{P}_{\lambda,y}(\beta) = \lambda$ و هي قيمة ثابتة . و في طريقة الجسر bridge ممكن ان تكون ($r = \frac{1}{2}$) معلمة الانكماش و نسبة الجزاء لها تساوي $\hat{P}_{\lambda,y}(\beta) = \frac{\lambda}{2} \sqrt{\beta}$ و هيه قيمة عالية جداً فان قيمة β تقترب من الصفر ولكن تنخفض بسرعة لتصحيح β كبيرة .

لقد ظهرت العديد من اساليب و تقنيات التقدير في العقدين الماضيين الى الذي ادى الى تحسن عالي في كل من دقة النموذج التنبؤ والكفاءة الحسابية و في هذا البحث تم استخدام بعض الطرائق الانكماش (Shrinkage Methods) و كما موضحة في ادناه:

Lasso Method

5.1.1 طريقة لاسو [8],[6],[3]

اقترحت هذه الطريقة من قبل الباحث (Tibshirani) عام (1996) حيث بين ان طريقة (Lasso) هي جزء من عائلة المربعات الصغرى الجزائية و التي تعمل على اختيار المتغير و تقدير النموذج الانحدار في ان واحد . و وضح ان طريقة (Lasso) ترتبط ارتباط وثيق بطريقة انحدار الحرف (Ridge Regression) و ان حلول (Lasso) لها تعريف مشابه الى حلول (Ridge Regression) من خلال ابدال مربع الجزاء $\sum_{j=1}^p \beta_j^2$ بالجزء المطلق $\sum_{j=1}^p |\beta_j|$ في (Lasso) .

اذ تعمل هذه الطريقة على انكماش بعض المعاملات و يضع الاخرى مساوية بالضبط الى الصفر و يحسب مقدر (Lasso) الى (b) من خلال الصيغة التالية :

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 \text{ subject } \sum_{j=1}^p |\beta_j| \leq S \right\} \text{----- (3)}$$

Or equivalently

$$\hat{\beta}^{Lasso} = argmin_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \text{-----} (4)$$

اذ ان :

S : يمثل عامل الانكماش (مستوى القيمة) (The soft – thresh lading)

λ : تمثل معلمة الجزاء .

و كحالة خاصة لمصفوفة التصميم X (design matrix) اذا كانت (Orthonormal) فمن الممكن الحصول على حلول طريقة (Lasso) وفق الصيغة التالية:

$$\hat{\beta}_j^{Lasso} = S(\hat{\beta}_j^{ols}, \lambda) = \begin{cases} \hat{\beta}_j^{ols} - \lambda & \text{if } \hat{\beta}_j^{ols} > \lambda \\ 0 & \text{if } |\hat{\beta}_j^{ols}| \leq \lambda \\ \hat{\beta}_j^{ols} + \lambda & \text{if } \hat{\beta}_j^{ols} < -\lambda \end{cases} \text{-----} (5)$$

وهذه الطريقة تعمل على تقليل مجموع مربعات الخطأ (MSE) اضافة الى مجموع القيم المطلقة لمعاملات الانحدار. و كلما زادت قيمة معلمة الجزاء (λ) فان معاملات الانحدار تنكمش الى الصفر.

وقد اقترح الباحث (Zou) عام (2006) اصدار جديد لطريقة (Lasso) يدعى (Adaptive Lasso) من خلال استعمال اوزان التكيف التي تؤدي الى جزاء مختلف للمعاملات التي تظهر في حد الجزاء (L_1) و يمكن حساب طريقة (Adaptive Lasso) وفق الصيغة التالية :

$$\hat{\beta}_{alasso} = argmin_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta)^2 + \lambda \sum_{j=1}^p w_j |\beta_j| \right\} \text{-----} (6)$$

اذ ان :

$W_j = (w_1, w_2, \dots, w_p)$: و هي تمثل اوزان التكيف Adaptive Weight .

وقد اظهر الباحث (Zou) اذا تم اختيار الاوزان بكفاءة و بطريقة تعتمد على البيانات فان طريقة (Adaptive Lasso) يمكن ان تحقق خاصية الكفاءة (Efficient) .

واقترح الباحث (Zou) استعمال الاوزان المقدره وفق الصيغة التالية :

$$\hat{w}_j = |b_j|^{-r} \text{-----} (7)$$

اذ ان :

$b = \{b_j, j = 1, 2, \dots, p\}$ و هو مقدر مشتق \sqrt{n} لـ β و ان قيمة معلمة الانكماش r اكبر من الصفر .

ان اختيار الاوزان (\hat{W}_j) مهم جداً لذا اقترح الباحث (Zou) استعمال تقديرات المربعات الصغرى الاعتيادية لحسابها $(b_{j\ ois})$ بينما يتم اختيار قيمة (r) من خلال $(K\text{-fold Cross – Validation})$.

Elastic – Net Method

5.1.2 طريقة Elastic – Net [7],[4]

اقترح الباحثان (Hastic & Zou) في عام (2005) طريقة $elastic – net$ كنسخة محسنة لطريقة لاسو (Lasso) اذ بينا الباحثين ان طريقة (Lasso) غير مسقرة عندما يكون بين المتغيرات التوضيحية ارتباط عالي , وتعرف طريقة (Elastic – Net) بالشكل التالي :

$$\hat{\beta}_{E.net} = \left(1 + \frac{\lambda_2}{n}\right) \operatorname{argmin}_{\beta} \{ \|y - X\beta\|_2^2 + \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \} \text{----- (8)}$$

اذ ان :

$$y = (y_1, y_2, \dots, y_n)$$

X: is $n \times p$ regression matrix

$$\beta = (\beta_1, \beta_2, \dots, \beta_p)$$

$$\|\beta\|_q = \left\{ \sum_{j=1}^p |\beta_j|^q \right\}^{1/q}$$

λ_1, λ_2 : تمثل معلمات الجزاء (الضبط) .

و عندما تكون المتغيرات التوضيحية توزيعها قياسي فان $\left(1 + \frac{\lambda_2}{n}\right)$ ينبغي استبدالها بـ $(1 + \lambda_2)$.

علماً ان الجزء (L_1) من طريقة $elastic – net$ ينفذ عملية اختيار المتغير بشكل تلقائي . في حين تستقر الطريقة عند الجزء (L_2) و هي تمثل مسارات الحل لغرض تحسين عملية التنبؤ .

6. مفهوم معلمة الجزاء والتبعثر

Tuning (penalty) parameter

6.1 معلمة الضبط (الجزاء) [5]

في تحليل بيانات الجانب العملي فان القيمة المناسبة لمعلمة الضبط (الجزاء) (λ) هي غير معلومة مسبقاً لذا اقترح الباحث (Anderas) واخرون) في عام (2012) معيار معلومات بيز Bayse information criterion لغرض اختيار معلمة الجزاء (λ) المثلى , حيث بين ان هذه المعلمة تلعب دوراً فعالاً في اختيار المتغير المعنوي لأنها تسيطر على درجة انكماش المقدر و تم استخدام هذا المعيار لأنه يتطلب اقل جهد حسابي و يحسب وفق الصيغة التالية :

$$\text{BIC}(\lambda) = \text{Log SSE} + \frac{df_{\lambda} \log(n)}{n} \text{----- (9)}$$

يستخدم هذا المعيار لطريقة (Lasso) و (Adaptive Lasso) و طريقة (Elastic . Net)

Sparsity

6.2 التبعر [5]

ان طرائق الانكماش تضع المعاملات مساوية الى الصفر بالضبط اذا كانت المعلمة غير مهمة و معلمة الضبط (الجزء) (λ) كبيرة بما فيه الكفاية . وهذه الطرائق تعود الى نماذج التبعر (Sparse models) اي النماذج التي تتضمن فقط على المجموعة الجزئية من المتغيرات (المعنوية فقط) و هذا ما يدعى بالتبعر (Sparsity) .

7- الجانب التطبيقي

تناول الجانب التطبيقي دراسة و تحليل العلاقة بين العوامل المؤثرة على ارتفاع نسبة السكر في الدم لبعض المرضى في مستشفى الكرامة التعليمي باستعمال بعض طرائق الانكماش (طريقة لاسو (*Lasso*) و طريقة (elastic- net) و تمت المقارنة بين الطرائق من خلال المعيار متوسط المربعات الخطأ (*MSE*) و تم الحصول على جميع النتائج بالاعتماد على البرنامج الاحصائي (*SPSS*) .

7.1 : البيانات الحقيقية

تم الحصول على عينة البحث من السجلات الخاصة بمستشفى الكرامة التعليمي لعام 2019 و المتمثلة بخمسة متغيرات توضيحية لدراسة مدى تأثيرها على ارتفاع نسبة السكر في الدم باعتباره كمتغير معتمد ل (25) مريض. وكما موضح في الجدول رقم (1) .

جدول رقم (1)

يمثل البيانات الحقيقية للعوامل المؤثرة على ارتفاع نسبة السكر في الدم لبعض المرضى في مستشفى الكرامة (الكوت) .

n	y نسبة السكر في الدم	x1 عمر المريض	x2 نسبة اليوريا	x3 نسبة الكرياتينين	x4 نسبة الدهون الثلاثية	x5 نسبة الكلسترول
1	142	63	0.59	26.2	184	180
2	139	45	0.77	36.8	470	233
3	114	30	0.45	28.7	35	125
4	102	34	0.69	31	290	202
5	80	57	0.4	23	171	203
6	95	42	0.6	24	100	242
7	97	50	0.4	28	170	155
8	85	33	0.6	32	95	126
9	163	35	0.45	31.3	134	189
10	94	42	0.32	28.3	190	112
11	100	35	0.4	16.5	104	140
12	151	56	0.6	39	85	144
13	75	40	0.54	17.1	133	187
14	175	50	0.76	51.7	148	806
15	100	43	0.8	35.3	132	130
16	119	60	0.5	20	111	199
17	156	71	0.9	61	267	128
18	92	43	0.8	32	277	195
19	94	26	0.5	24	157	142
20	95	30	0.97	45	209	168
21	94	29	0.6	18	62	125
22	138	50	0.4	24	152	261
23	96	50	0.7	33	152	168
24	99	60	0.7	28	192	244
25	83	40	0.7	24	66	187

و يمكن توضيح متغيرات الدراسة بالشكل الآتي:

المتغير المعتمد (Y_i): يمثل نسبة السكر في الدم: ويتراوح المستوى الطبيعي له من (80 الى 100) مليكرام / دسل في الصباح و بعد الاكل يجب ان يكون السكر في الدم اقل من (140) مليكرام دسل

والمتغيرات التوضيحية تتمثل بالعوامل الآتية:

X1: يمثل عمر المريض

X2: و يمثل نسبة اليوريا في الدم: و هو من وظائف الكلى و المستوى الطبيعي له من (20 الى 40) مليكرام لكل 100 مليلتر دم.

X3: يمثل نسبة الكرياتينين في الدم: و هو من وظائف الكلى والمستوى الطبيعي له من (1.5 الى 5) مليكرام لكل 100 مليلتر دم.

X4: يمثل نسبة الدهون الثلاثية في الدم: تكون النسبة الطبيعية في الجسم اقل من 150 ملليغرام لكل ديسيلتر .

X5: يمثل نسبة الكولسترول في الدم: تكون النسبة الطبيعية في الجسم 170 او اقل ملليغرام لكل ديسيلتر .

اذ تم تحليل هذه البيانات باستعمال انموذج الانحدار الخطي العام والحصول على كافة النتائج من خلال استعمال البرنامج الاحصائي (SPSS) لجميع طرائق التقدير المستعملة في هذا البحث .

7.2: اختبار البيانات

7.2.1: اختبار وجود مشكلة التعدد الخطي في البيانات الحقيقية :

يمكن قياس مشكلة التعدد الخطي من حيث وجودها او عدم وجودها في بيانات الدراسة من خلال نتائج تحليل الانحدار والمبينة في الجدول ادناه والخاص بتشخيص العلاقة الخطية المتداخلة (collinearity diagnostics) بين المتغيرات المستقلة .

جدول رقم (2)

يوضح تشخيص العلاقة الخطية المتداخلة (collinearity diagnostics) بين المتغيرات المستقلة

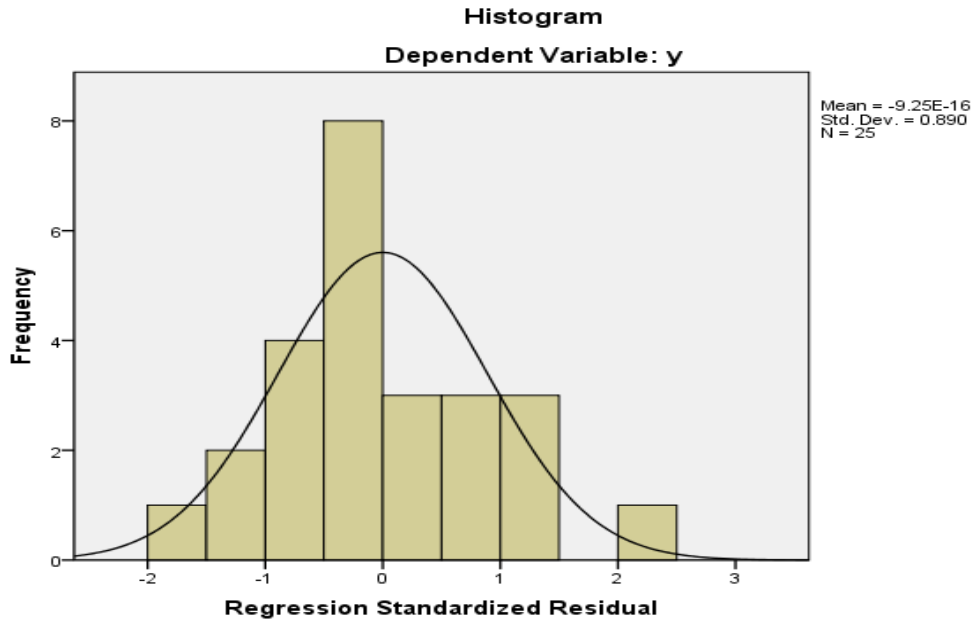
يمثل الجدول اعلاه حسابات مرتبطة بصفوفة المتغيرات حيث يرتبط الصف الاول بالقيمة الثابتة ويرتبط الصف الثاني بالمتغير المستقل الاول ويرتبط الصف الثالث بالمتغير المستقل الثاني ويرتبط الصف الرابع بالمتغير المستقل الثالث ويرتبط الصف الخامس بالمتغير المستقل الرابع ويرتبط الصف السادس بالمتغير المستقل الخامس . وتعتبر مشكلة التعدد الخطي مؤثرة اذا كانت قيمة دليل الحالة للمتغير كبيرة فاذا زادت القيمة عن 15 فهذا مؤشر على وجود مشكلة التعدد الخطي ونلاحظ ان قيمة دليل الحالة للمتغير الخامس والمساوية الى 16.939 وهي اكبر من القيمة 15 فهذا يعني وجود مشكلة التعدد الخطي في بيانات الدراسة .

7.2.2: اختبار التوزيع الطبيعي للبيانات الحقيقية :

للتحقق من ان البيانات المستعملة في هذا البحث تتبع التوزيع الطبيعي وهناك عدة اختبارات تستعمل لهذا الغرض اذ تم استعمال اختبار المدرج التكراري ومن الشكل البياني رقم (1) نلاحظ ان البيانات تتوزع توزيعا طبيعيا .

شكل رقم (1)

يوضح توزيع البيانات الطبيعي



7.2.3 : تحليل وتفسير نتائج الجانب التطبيقي

تم الحصول على كافة النتائج الخاصة بالبحث و لجميع طرائق التقدير المدروسة بالاعتماد على البرنامج الاحصائي SPSS و كما موضحة ادناه.

اولاً : تحليل البيانات باستعمال طريقة لاسو (Lasso – Method) : بالاعتماد على انموذج الانحدار الخطي العام و كما مبين في ادناه:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + e_i \quad \text{----- (10)}$$

ومن خلال بيانات الجدول رقم (1) و بالاعتماد على انموذج الدراسة اعلاه و باستعمال طريقة لاسو تبين نتائج هذه الطريقة في الجدولين رقم (2) و (3) وكما يلي :

جدول رقم (3)

يوضح قيم معامل التحديد R^2 و R^2_{Adj} و متوسط مربعات الخطأ (MSE) للانموذج (GIM) بطريقة لاسو.

method	R^2	R^2_{adj}	MSE
lasso	0.753	0.688	0.688

جدول رقم (4)

يوضح تقديرات معاملات الانموذج (GIM) و الانحراف المعياري للمعاملات باستعمال طريقة لاسو.

Coefficients	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\hat{\beta}$	0.000	0.000	0.060	0.000	0.299
$SD(\hat{\beta})$	0.086	0.014	0.137	0.042	0.171

ثانياً: تحليل البيانات باستعمال طريقة (Elastic - Net)

نلاحظ من خلال بيانات الجدول رقم (1) و بالاعتماد على انموذج الدراسة الموضح في المعادلة رقم (10) و باستعمال طريقة (Elastic - Net) تبين نتائج هذه الطريقة في الجدولين رقم (5) و (6) وكما يلي :

جدول رقم (5)

يوضح قيم معامل التحديد R^2 و R^2_{adj} و متوسط مربعات الخطأ (MSE) للانموذج (GIM) بطريقة (Elastic - Net)

method	R^2	R^2_{adj}	MSE
Elastic - Net	0.776	0.717	0.581

جدول رقم (6)

يوضح تقديرات معاملات الانموذج (GIM) و الانحراف المعياري (SD) للمعاملات باستعمال طريقة (Elastic - Net)

Coefficients	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\hat{\beta}$	0.000	0.000	0.167	0.000	0.284
$SD(\hat{\beta})$	0.078	0.030	0.107	0.042	0.151

ثالثاً : تحليل البيانات باستعمال طريقة OLS

نلاحظ من خلال بيانات الجدول رقم (1) و بالاعتماد على انموذج الدراسة الموضح في المعادلة رقم (10) و باستعمال طريقة (OLS) تبين نتائج هذه الطريقة في الجدولين رقم (7) و (8) وكما يلي :

جدول رقم (7)

يوضح قيم معامل التحديد R^2 و R^2_{adj} ومتوسط مربعات الخطأ (MSE) للنموذج (GIM) بطريقة OLS.

method	R^2	R^2_{adj}	MSE
OLS	0.560	0.432	457.164

جدول رقم (8)

يوضح تقديرات معاملات الانموذج (GIM) و الانحراف المعياري (SD) للمعاملات باستعمال طريقة OLS .

Coefficients	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\hat{\beta}$	63.933	0.433	-69.328	1.862	0.010	0.061
$SD(\hat{\beta})$	21.866	0.395	35.791	0.621	0.054	0.035

رابعاً : مقارنة نتائج طرائق التقدير

يمكن مقارنة جميع طرائق التقدير المستعملة في هذا البحث بالاعتماد على معيار المقارنة متوسط مربعات الخطأ (MSE) و كما مبين في الجدول رقم (9) الموضح ادناه.

جدول رقم (9)

يوضح قيم متوسط مربعات الخطأ (MSE) لجميع الطرائق

Method	Lasso	Elastic - Net	OLS	Best
MSE	0.688	0.581	457.164	Elastic - Net

نلاحظ من خلال نتائج الجدول اعلاه افضلية طريقة **Elastic - Net** في تقدير انموذج الانحدار الخطي العام (GIM) مقارنة بالطرائق الاخرى لكونها تعطي اقل قيمة لمتوسط مربعات الخطأ MSE تليها في الافضلية طريقة **Lasso** .

8: الاستنتاجات والتوصيات**8.1: الاستنتاجات**

1. من خلال النتائج التي تم الحصول عليها بواسطة الطرائق المستعملة في معالجة مشكلة التعدد الخطي نستنتج ان طريقة (elastic-net) هي اكثر ملائمة في تحليل العلاقة بين العوامل المؤثرة على ارتفاع نسبة السكر في الدم لانموذج الانحدار الخطي العام .
2. اظهرت النتائج ان طريقة الانكماش (elastic- net) الخاصة بتقدير واختيار المتغير المعنوي لها افضلية واضحة على الطرائق الاخرى تليها في الافضلية طريقة Lasso كونها تعطي اقل قيمة لـ (MSE) .
3. بالاعتماد على نتائج الطريقة الافضل (elastic- net) تم تحديد العوامل التي تؤثر و بشكل معنوي على ارتفاع نسبة السكر في الدم وهي X3(والذي يمثل نسبة الكرياتينين في الدم) و X5(والذي يمثل نسبة الكلسترول في الدم) وهذا التأثير يؤدي الى ارتفاع نسبة السكر في الدم.
4. اظهرت النتائج ان المتغير التوضيحي X5: والذي يمثل نسبة الكلسترول في الدم هي اكثر تأثيراً على ارتفاع نسبة السكر في الدم مقارنة بالمتغيرات الاخرى وبالاعتماد على نتائج طريقة elastic- net.

8.2: التوصيات

- بناءً على الاستنتاجات التي تم التوصل اليها و اعتماداً على نتائج واقع التحليل الاحصائي للبيانات فانه يمكن توضيح اهم التوصيات التي خرج بها البحث.
1. استعمال طريقة الانكماش (elastic- net) في عملية معالجة مشكلة التعدد الخطي في انموذج الانحدار في تحليل البيانات كونها اعطت افضل النتائج مقارنة بالطرائق الاخرى المستعملة في هذا البحث .
 2. نوصي باستعمال طريقة اخرى من طرائق الانكماش كطريقة (SCAD) .
 - 3.نوصي مستشفى الكرامة التعليمي التركيز على معالجة العوامل المؤثرة والمتمثلة بقياسات الدم للمرضى وهي (نسبة الكرياتينين ونسبة الكلسترول في الدم) كونها المسبب الرئيسي في ارتفاع مستوى السكر في الدم وان هذا الارتفاع له نتائج سلبية على المرضى .

1. Al –momani, M. ,(2013), “ shrinkage and penalty estimation strategies in some partial models” . university of Windsor . Canada . these of doctor philosophy in statistics.
2. Gafar . ,M. , O, eyitayo. , O., O. and Adeyinka . , I, F, (2015). “On performance of shrinkage methods – Amontecarlo study ” .International journal of statistic and application , 5(2) , pp . 72-76.
3. Joseph , O. Ogutu and Hans , P,P, (2014) , “ Regularized group regression methods for genomic prediction : Bridge , Mcp, SCAD, group bridge , group Lasso , sparse group Lasso , group Mcp and group SCAD” , Puplished in BMC proceeding volume and , supplement 5.
4. Nan . X , and Qing . S. , X, (2015), “ Multi –step Adaptive elastic , net , reducing false positive in high – dimensional variable selection” . Journal of statistical computation and simulation . volume 85. Issue 18, pp1-11.
5. Peng ., H. and Huang ., T, (2011) , “ Penelized least squares for single index models” .Journal of statistical planning and inference 14 , (2011) , pp. 1362-1379.
6. Saria,I.(2010), “ A comparative study of the Lasso –Type and herristic model selection models”. Jahibucher for national okonomie and statistic , volume 233, Issue 4 , pp. 526-549.
7. Su, L, and ZHANG , Y. ,(2012) , “ variable selection in nonparametric and semiparametric regression models”, school of Economics , Singapore , management university .
8. Tibshirani, R., (1996), “Regression shrinkage and selection Via the Lasso” . Journal of the Royal statistical society . Vol 58 . No.1 , pp267-288.