Warith Scientific Journal

اختيار أهم المتغيرات ألمؤثرة على نقص الحديد في الدم باستخدام طريقة الشبكة المرنة البيزية

Selecting the most important variables affecting iron deficiency in the blood using the Bayesian elastic network method

أ.م. د. سيف حسام رحيم
Dr . Saif Hossam Rahim
جامعة القادسية
كلية الادارة و الاقتصاد
قسم الاحصاء
saif.hosam@qu.edu.iq

أ. د. احمد نعيم فليح
A. D. Ahmed Naeem Falih2
جامعة القادسية
كلية الادارة و الاقتصاد
قسم الاحصاء
ahmad.flaih@qu.edu.iq

المستخلص

تحليل الانحدار يُعتبر أحد أهم المواضيع في الإحصاء ويتمتع بانتشار واسع لما يحققه من غرضين أساسيين. الغرض الأول منه هو استخدام نموذج الانحدار كأداة لتفسير وشرح الظاهرة المدروسة بشكل فعّال، بينما يُعتبر الغرض الثاني منه استخدام نموذج الانحدار كأداة تنبؤيه لهذه الظاهرة. سيتم في هذه الدراسة التعرف الى الجانب النظري لمشكلة اختيار المتغيرات من خلال طرق التنظيم المتمثلة بطريقة الشبكة المرنة Elastic Net وتقدير معالم نموذج الانحدار الخطي المتعدد وفق أسلوب Bayes . تُعتبر طريقة الشبكة المرنة احد طرق اختزال المتغيرات و تقدير المعالم في نفس الوقت . سيتم استعمال نظرية Bayes لتقدير معالم نموذج الانحدار المتعدد وفقاً لطريقة الجزاء الشبكة المرنة Elastic Net. تم اقتراح نموذج هرمي للتوزيعات الاولية .أجريت ثلاث تجارب محاكاة تحت أحجام عينات مختلفة ، حيث تم مقارنة الطريقة المقترحة مع طريقة لاسو التكيفي البيزي Bayesian adaptive lasso .تم اعتماد المعيارين MMAD وMSE في المقارنة بين الطريقة المقترحة و طريقة المقارنة ، حيث أظهرت النتائج افضلية الطريقة المقترحة. كما تم إجراء التطبيق العملي للبيانات والذي يمثل دراسة نموذجية انحداريه لمجموعة من المتغيرات التفسيرية على متغير نقص الحديد في الدم لمجموعة من المرضى. ومن خلال استخدام الطريقة المقترحة، اتضح أنها قدمت طريقة لاختيار المتغيرات، بناء على القيم المقدرة ومقارنة النتائج مع الطرق المشار إليها في الجانب التجريبي.

الكلمات المفتاحية : الانحدار المتعدد , بيز , الشبكة المرنة , الشبكة المرنة البيزية , نقص الحديد**Abstract**

Regression analysis is considered one of the most important topics in statistics and enjoys wide spread because it achieves two basic purposes. Its first purpose is to use the regression model as a tool to effectively interpret and explain the studied phenomenon, while its second purpose is to use the regression model as a predictive tool for this phenomenon. In this study, the theoretical aspect of the problem of selecting variables will be identified through organization methods represented by the Elastic Net method Estimating the parameters of the multiple linear regression model according to the Bayes method. The flexible network method is considered one of the methods of reducing variables and estimating parameters at the same time. Bayes' theory will be used to estimate the parameters of the multiple regression model according to the Elastic Net penalty method. A hierarchical model for initial distributions was proposed. Three simulation experiments were conducted under different sample sizes, where the proposed method was compared with the Bayesian adaptive lasso method. The MMAD and MSE standards were adopted in the comparison. Between the proposed method and the comparison method, where the results showed the superiority of the proposed method. The practical application of the data was also carried out, which represents a regression model study of a set of explanatory variables on the variable of iron deficiency in the blood for a group of patients. Through the use of the proposed method, it became clear that it provided a method for selecting variables, based on the estimated values and comparing the results with the methods referred to in the experimental aspect.

Keywords: multiple regression, Bayesian, elastic network, Bayesian elastic network, iron deficiency. Abstract

المقدمة

تحليل الانحدار ونماذج التنبؤ هي أدوات حيوية لفهم العلاقة بين متغير الاستجابة ومجموعة من المتغيرات التوضيحية. يُسهم هذا الفهم في تفسير سلوك الظاهرة المدروسة من خلال تحديد القيم المتوسطة لمتغير الاستجابة، وبالتالي يسهل توقع القيم المستقبلية للظاهرة. بالإضافة إلى ذلك، يمكن لنماذج التنبؤ توفير معلومات حول المتغيرات التفسيرية التي تؤثر في قيمة متغير الاستجابة وعلاقتها بذلك المتغير. نماذج الانحدار لها تطبيقات غير محددة في مجالات مختلفة، بما في ذلك المجالات الطبية، حيث يستخدم نموذج الانحدار المقدر للمساعدة في الكشف عن المخاطر العالية لبعض الأمراض، مما يساعد على منع تطور مرض معين في وقت مبكر من خلال التدخلات اللازمة. ويمكن لهذه النماذج أيضًا التنبؤ بالمرض في المستقبل، مما يسهل التواصل بين المريض والطبيب المختص بناءً على معلومات أكثر واقعية، وبالتالي يساعد النموذج المتوقع في تقديم المساعدة للأطباء والمرضى فيما يتعلق بإجراء اختبارات أفضل، والمساعدة في الرعاية والخدمات الصحية. وبالتالي، ينعكس ذلك في اتخاذ القرار المناسب فيما يتعلق بالتخطيط وإدارة الجودة للمؤسسات الصحية. اختيار المتغيرات المناسبة في نموذج الانحدار يُعتبر أحد أهم الخطوات التي يحتاجها الباحث، حيث يتعين عليه اختيار المتغيرات التوضيحية المهمة أو المتعلقة بمتغير الاستجابة وتضمينها في النموذج. إن هذه العملية تُعد من أصعب مراحل بناء النموذج، وتُعرف باسم طريقة اختيار المتغيرات. تقتصر هذه الطريقة ببساطة على اختيار المتغيرات التوضيحية من بين مجموعة من المتغيرات، ثم إدراجها في النموذج المقدر للانحدار. يتم ذلك عن طريق حذف المتغيرات التي لا تؤثر على متغير الاستجابة، والاحتفاظ بتلك التي تؤثر عليه. ومن ثم يتم تكوين نموذج الانحدار الذي يعكس العلاقة بين المتغيرات بشكل أفضل ويتماشى مع بيانات الدراسة، مما يؤدي إلى تحقيق أفضل قدرة تنبؤيه لنموذج الانحدار. مشكلة البحث تتطلب مجموعات البيانات الطبية الكبيرة والمتعددة المتغيرات تطبيقًا فعّالًا للنماذج الإحصائية. ومع ذلك، ينتج استخدام النماذج التقليدية عندما يكون لدينا متغيرات تفسيرية كبيرة نسبياً مقارنة بحجم العينة، عن ضعف في التفسير. لذلك، تطوير أساليب إحصائية جديدة التي تختار المتغيرات بحرص، أصبح ضرورياً لتحسين كفاءة النماذج وسهولة فهم البيانات المدروسة. يهدف البحث إلى تحديد العوامل التي لها تأثير مباشر على مرض نقص الحديد في الدم في المجال الطبي. ويهدف الى اختبار كفاءة طريقة Bayesian Elastic Net ومقارنتها مع طريقة Bayesian adaptive lasso في الجانبين التجريبي والعملي.

طرائق اختيار المتغيرات

في عام 1996، أشار الباحث Tibshirani [1] إلى أن نتائج طريقة المربعات الصغرى العادية لا يمكن الاعتماد عليها عند تحليل البيانات، وذلك لسببين. السبب الأول يتعلق بدقة التنبؤ، حيث أن مقدرات طريقة المربعات الصغرى تعتبر متحيزة قليلاً ولكن لها تباين كبير، وبالتالي يتم تقليل بعض المتغيرات وجعلها تساوي الصفر لتحسين دقة التنبؤ، ورغم أن هذا الإجراء يزيد من انحياز المقدر إلا أنه يقلل من تباين القيم المتوقعة مما يؤدي إلى تحسين دقة التنبؤ. السبب الثاني يتعلق بالقوة التفسيرية للنموذج، حيث يتطلب عادة اختيار بعض المتغيرات ذات التأثير القوي على متغير الاستجابة لإنتاج نموذج الانحدار ذو قوة تفسيرية عالية لتحسين أداء طريقة المربعات الصغرى، تم اقتراح طريقتين بديلتين، وهما طريقة ridge [6] وطريقة اختيار المجموعة الفرعية subset selection. ومع ذلك، تعاني هذه الأساليب من بعض العيوب. ولذا قدّم الباحث Tibshirani عام 1996 تقنية جديدة وهي إحدى طرق الجزاء تُسمى لاسو. تعمل هذه التقنية على تقليل بعض معاملات المتغيرات وجعل بعض المعاملات تساوي الصفر [9]، وبالتالي تجمع بين خصائص طرق ridge وطريقة اختيار المجموعة الفرعية subset selection .

طريقة جزاء الشبكة المرنة Elastic Net

في عام 2005، قدم الباحثان (Hastie & Zou) [2] طريقة جديدة للجزاء تُسمى الشبكة المرنة، وهي بديل مقترح لطريقة Lasso في تحليل بيانات الانحدار بدراسة الجينات genes الوراثية. تعمل طريقة الشبكة المرنة كطريقة لاختيار المتغيرات، حيث تقوم بتقليل بعض المعاملات واختيار مجموعات من المتغيرات التفسيرية المترابطة، مما يساعد في التغلب على مشكلة عدم توافر معلومات كافية بالنسبة للمتغيرات التفسيرية. يعتمد عمل طريقة الشبكة المرنة بشكل أساسي على مزج دالة الجزاء ridge و دالة الجزاء Lasso ، مما يتيح ترتيب مقدر الشبكة المرنة بشكل فعال.

$$\hat{\beta}_{\text{elastic net}} = [\|Y - X\beta\|_2^2 + \lambda_1\|\beta_1\| + \lambda_2\|\beta\|_2^2]\} \qquad (1)$$

حيث ان $\lambda_1 \geq 0$ و $\lambda_2 \geq 0$ هي معالم دوال التنظيم او التقليص [7] ، حيث ان λ_1 تتحكم بكمية التقليص واختيار المتغيرات و λ_2 تتحكم بكمية grouping . وبهذا يمكن القول ان طريقة الشبكة المرنة هي طريقة تنظيم تستخدم لتحسين أسلوب اختيار المتغيرات في حالة كون ان هناك ارتباطات قوية بين مجاميع من المتغيرات التوضيحية .

أشار الباحثان (Hastie & Zou) في عام 2005 إلى أن الدراسات في مجال علم الأحياء، وخاصة تلك المتعلقة بتحليل الجينات الجينية، غالبًا ما تشمل آلاف الجينات، في حين أن حجم العينة يكون أقل من 100 مشاهده . بسبب ارتباط الجينات ببعضها البعض وتشكيلها لمجموعات، يعني اختيار الجين اختيار المجموعة التي ينتمي إليها. في ظل هذه الظروف، تكون طريقة Lasso غير مثالية، حيث تختار فقط عددًا من المتغيرات من بين العديد، وبالتالي تفتقر إلى القدرة على استخلاص المعلومات المتوفرة في

مجموعات المتغيرات [8]. على الرغم من تحسين دقة التنبؤ لطريقة Lasso من خلال الشبكة المرنة بسبب وجود ارتباطات قوية بين المتغيرات التوضيحية، إلا أنها لا تتمتع بخصائص الأوراكل، مثل الاتساق والتوزيع الطبيعي الموازي للإمكانيات.

النموذج الهرمي البيزي

نفترض النموذج البيزي باستخدام التوزيع الاولي لمعلمات الانحدار

$$\pi(\beta_j|\lambda_1,\lambda_2) = c(\lambda_1,\lambda_2)\frac{\lambda_1}{2}\exp\{-\lambda_1|\beta_j| - \lambda_2\beta_j^2\} \quad (2)$$

حيث ان $c(\lambda_1,\lambda_2)$ تمثل normalizing constant .

$$c(\lambda_1,\lambda_2) = (\frac{\lambda_1}{2\sqrt{\lambda_2}})^{-1}\Gamma^{-1}\left(\frac{1}{2},\frac{\lambda_1^2}{4\lambda_2}\right)\exp\{-\frac{\lambda_1^2}{4\lambda_2}\} \quad (3)$$

سيتم استعمال توزيع Gamma الاولي للمعلمات (λ_1,λ_2) وكذلك اضافة مقياس ثابت scale-invariant للتوزيع الاولي الحدي لـ(σ^2) أي ان $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$. لاخذ خوارزمية Gibbs لطريقة الشبكة المرنة البيزية سيتم استعمال تمثلا جديد لتوزيع الاولي للمعادلة رقم (2) لغرض تسهيل عملية الحساب . و للتوضيح اكثر سيتم اتباع المثال التالي .

التوزيع الاولي للمعادلة رقم (2) هو $\pi(\beta|\lambda_1,\lambda_2)$ و الذي يمكن كتابته بالشكل التالي :

$$\pi(\beta|\lambda_1,\lambda_2) = \prod_{j=1}^{k} c(\lambda_1,\lambda_2)\frac{\lambda_1^2}{2}\int_{u_j>|\beta_j|}\exp\{-(\lambda_1 u_j + \lambda_2\beta_j^2)\}du_j \quad (4)$$

حيث يكون mixing variable u_j يمتلك توزيع Gamma .

اقترح الباحثان (Alhamzawi & et) [3] النموذج الهرمي التالي :

$$y|\beta,\sigma^2 \sim N(X\beta,\sigma^2 I_n)$$

$$\beta,u|\lambda_1,\lambda_2 \sim \prod_{j=1}^{k} c(\lambda_1,\lambda_2)\frac{\lambda_1^2}{2}\exp\{-(\lambda_1 u_j + \lambda_2\beta_j^2)\}$$

$$\sigma^2 \sim \frac{1}{\sigma^2}$$

$$\lambda_1 \sim \lambda_1^{a-1}\exp\{-b\lambda_1\}, \lambda_2 \sim \lambda_2^{a-1}\exp\{-d\lambda_2\}$$

من خلال تقديم mixing variable u_j يمكن اشتقاق التوزيعات الاولية الشرطية لمعلمات الانحدار و كما يلي .

$$\beta\,|y,X,u,\sigma^2,\lambda_1,\lambda_2 \sim N(\Sigma^{-1}X'y,\Sigma^{-1})\prod_{j=1}^{k} I\{|\beta_j| < u_j\}$$

حيث ان $\Sigma = \frac{1}{\sigma^2}X'X + 2\lambda_2 I_k$ و ان $I\{.\}$ تمثل دالة المؤشر indicator function . وان التوزيع الشرطي الاولي لمتغيرات الخلط u_j mixing variables هو

$$u|\beta,\lambda_1 \sim \prod_{j=1}^{k}\exp(\lambda_1)\, I\{u_j > |\beta_j|\}$$

لرسم قيم u_j من التوزيع الاسي المقطوع من اليسار باستعمال طريقة الانعكاس وعلى النحو التالي[4] .

1- رسم $u_j^* \sim Exp(\lambda_1)$.

2- $u_j = u_j^* + |\beta_j|$

حيث ان التوزيع الشرطي لـ(σ^2) هو .

$$\sigma^2|y,X,\beta \sim Inverse\ Gamma\ (\frac{n}{2},\frac{1}{2}(y-X\beta)'(y-X\beta))$$

اختيار قيمتي المعلمات (λ_1,λ_2)

في هذا الجزء سيتم تقدير قيمة المعلمتين (λ_1, λ_2) باستخدام طريقة Monte Carlo حسب خوارزمية الشبكة المرنة البيزية [5] , في كل خطوة يتم معالجة كبيانات مفقودة , و متوسطهم لتقدير قيمتي المعلمتين (λ_1, λ_2) من خلال تعظيم توقع لوغاريتم التوزيع اللاحق الشرطي . اكمال بينات لوغاريتم الامكان الاعظم من خلال[3] .

$$\log p\,(\beta, \sigma^2, u, \lambda_1, \lambda_2 | y, X) \propto \sum_{i=1}^{n} \log p(y_i \,|X_i\beta, \sigma^2) + \sum_{j=1}^{k} \log p(\beta_i \,|u_i, \lambda_2)$$
$$+ \sum_{j=1}^{k} \log p(u_i \,|\lambda_1) + \log p(\lambda_1) + \log p(\lambda_2)$$

المحاكاة

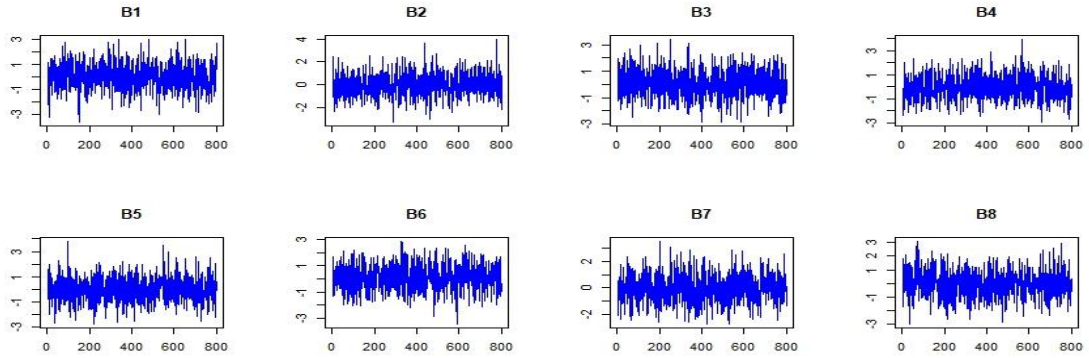
سيتم عمل ثلاثة تجارب محاكاة باستخدام تقنية MCMC وذلك وفق طريقة الشبكة المرنة البيزية BEN , وكذلك مقارنة نتائج الطريقة قيد الدراسة مع طريقية لاسو التكيفي البيزي BALsso وذلك بالاعتماد على معيارين للمقارنة وهي وسيط متوسط الانحرافات المطلقة MMAD , و معيار متوسط مربعات الاخطاء MSE .وباستعمال اربعة احجام للعينات (25,50,75,100) , وكذلك تم افتراض قيمتين لتباين الاخطاء (4,6). تم تكرار كل مثال 1000 مرة حيث تم استبعاد او 200 تكرار وذلك لغرض ضمان استقرار خوارزمية Gibbs .

اولاً : التجربة الاولى تم افتراض قيم المعلمات الحقيقة وهي كما يلي: $\beta = (1,1,1,1,2,2,2,2)$

جدول رقم (1) يبين قيم معايير المقارنة للمثال الاول

sample size	Methods	σ^2	MMAD	MSE
25	BALasso	4	0.33915	0.11139
	BEN		0.33087	0.09025
	BALasso	6	0.39513	0.12104
	BEN		0.39088	0.11945
50	BALasso	4	0.24882	0.04978
	BEN		0.1493	0.03884
	BALasso	6	0.05063	0.07417
	BEN		0.03742	0.02453
75	BALasso	4	0.17417	0.04105
	BEN		0.12453	0.03845
	BALasso	6	0.0771	0.05003
	BEN		0.04316	0.03348
100	BALasso	4	0.23043	0.11049
	BEN		0.22803	0.10973
	BALasso	6	0.21326	0.08729
	BEN		0.25058	0.07131

من خلال نتائج الجدول رقم (1) يتبين لنا عند حجم عينة (25) و بتباين (4) بان طريقة BEN حققت الافضلية من خلال حصولها على اقل قيمة لمعايير المفاضلة .اما عند زيادة قيمة التباين نجد زيادة بقيمة معايير المقارنة وذلك لزيادة تشتت البيانات لكن بقي لطريقة قيد الدراسة الافضلية بالمقارنة مع طريقة BALasso . اما في احجام العينات الاخرى (50,75,100) و بتباين (4,6) نلاحظ افضلية طريقة BEN وذلك لان كلما يزداد حجم العينة يزداد وزن دالة الامكان الاعظم المضروب بدالة التوزيع الاولي .



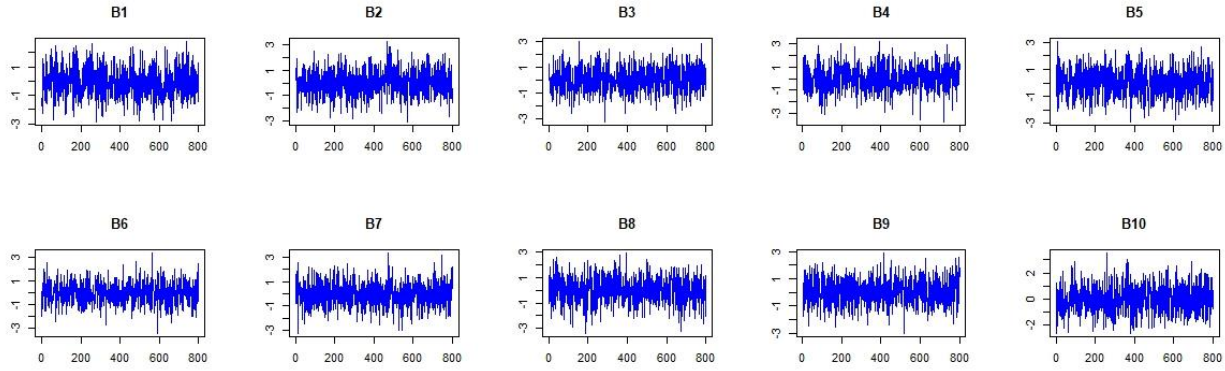
شكل رقم (1) يبين رسم trace plot لمعالم المثال الاول

ثانياً: : التجربة الثانية تم افتراض قيم المعلمات الحقيقة وهي كما يلي :$\beta = (1,0,1,0,1,0,2,0,2,0,2,0)$

جدول رقم (2) يبين قيم معايير المقارنة للمثال الثاني

sample size	Methods	σ^2	MMAD	MSE
25	BALasso	4	0.28674	0.12104
	BEN		0.27872	0.1205
	BALasso	6	0.16652	0.36387
	BEN		0.16379	0.35645
50	BALasso	4	0.23144	0.16992
	BEN		0.23104	0.0966
	BALasso	6	0.25418	0.13797
	BEN		0.21422	0.11391
75	BALasso	4	0.28239	0.09079
	BEN		0.23249	0.08875
	BALasso	6	0.2632	0.16559
	BEN		0.13943	0.15631
100	BALasso	4	0.26855	0.15165
	BEN		0.20342	0.13547
	BALasso	6	0.17634	0.09374
	BEN		0.14376	0.07017

من خلال نتائج الجدول رقم (2) يتبين لنا عند حجم عينة (25) و بتباين (4) بان طريقة BEN حققت الافضلية من خلال حصولها على اقل قيمة لمعايير المفاضلة .اما عند زيادة قيمة التباين نجد زيادة بقيمة معايير المقارنة وذلك لزيادة تشتت البيانات لكن بقي لطريقة قيد الدراسة الافضلية بالمقارنة مع طريقة BALasso . اما في احجام العينات الاخرى (50,75,100) و بتباين (4,6) نلاحظ افضلية طريقة BEN وذلك لان كلما يزداد حجم العينة يزداد وزن دالة الامكان الاعظم المضروب بدالة التوزيع الاولي .



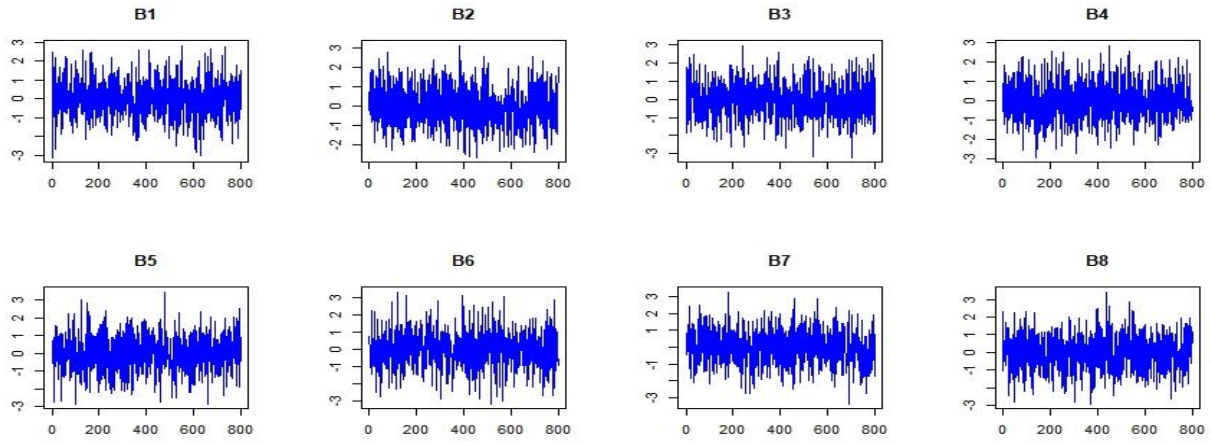
شكل رقم (2) يبين رسم trace plot لمعالم المثال الثاني

ثالثاً : التجربة الثانية تم افتراض قيم المعلمات الحقيقة وهي كما يلي :$\beta = (1,0,0,0,0,0,0,2)$

جدول رقم (3) يبين قيم معايير المقارنة للمثال الثالث

sample size	Methods	σ^2	MMAD	MSE
25	BALasso	4	0.04807	0.02734
	BEN		0.0317	0.02576
	BALasso	6	0.04874	0.0252
	BEN		0.04017	0.02412
50	BALasso	4	0.07618	0.09611
	BEN		0.07565	0.09218
	BALasso	6	0.07357	0.08452
	BEN		0.05614	0.05734
75	BALasso	4	0.07675	0.05526
	BEN		0.06322	0.04354
	BALasso	6	0.08163	0.05461
	BEN		0.06991	0.03843
100	BALasso	4	0.08142	0.07635
	BEN		0.07237	0.05566
	BALasso	6	0.0789	0.0847
	BEN		0.06369	0.07855

من خلال نتائج الجدول رقم (3) يتبين لنا عند حجم عينة (25) و بتباين (4) بان طريقة BEN حققت الافضلية من خلال حصولها على اقل قيمة لمعايير المفاضلة .اما عند زيادة قيمة التباين نجد زيادة بقيمة معايير المقارنة وذلك لزيادة تشتت البيانات لكن بقي لطريقة قيد الدراسة الافضلية بالمقارنة مع طريقة BALasso . اما في احجام العينات الاخرى (50,75,100) و بتباين (4,6) نلاحظ افضلية طريقة BEN وذلك لان كلما يزداد حجم العينة يزداد وزن دالة الامكان الاعظم المضروب بدالة التوزيع الاولي .



شكل رقم (3) يبين رسم trace plot لمعالم المثال الثاني

البيانات الحقيقية

نقص الحديد في الدم هو حالة شائعة تحدث عندما يكون مستوى الحديد في الجسم أقل من المستوى الطبيعي. يُعتبر الحديد من المعادن الأساسية التي يحتاجها الجسم لأداء وظائفه الحيوية مثل نقل الأكسجين في الدم ودعم وظائف الجهاز المناعي. يمكن أن يؤثر نقص الحديد في الدم على الصحة والعافية بشكل شديد إذا لم يُعالج، مما يجعله موضوعاً هاماً يستحق الاهتمام. تتمثل أعراض نقص الحديد في الدم في تعب شديد، وشحوب البشرة، وخفقان القلب، وضيق التنفس، وآلام في الصدر، والدوخة، وتغيرات في

الحالة الإدراكية، وبرودة اليدين والقدمين، والصداع. يُمكن تشخيص نقص الحديد في الدم من خلال فحص الدم وعلاجه عادةً بتناول مكملات الحديد وتغيير النظام الغذائي.

نقص الحديد في الدم يمكن أن يترافق مع مجموعة من الأعراض المتنوعة التي تشمل:

- التعب: إذا كانت لديك مستويات منخفضة من الحديد في الجسم، قد تشعر بالتعب الشديد ونقص الطاقة.
- شحوب البشرة: يمكن أن يتسبب نقص الحديد في الدم في شحوب البشرة، حيث تصبح البشرة باهتة اللون.
- خفقان القلب السريع وغير المنتظم: يعتبر هذا الشعور بالخفقان في القلب واحدًا من الأعراض الشائعة لنقص الحديد في الدم.
- ضيق التنفس: قد تواجه صعوبة في التنفس أو تشعر بضيق التنفس بسبب نقص الحديد في الدم.
- آلام في الصدر: قد تشعر بآلام في الصدر نتيجة لنقص الحديد في الدم، وقد يتطلب ذلك اهتمامًا فوريًا.
- الدوخة: قد تشعر بالدوخة وعدم الاستقرار بسبب نقص الحديد في الدم.
- تغيرات في الحالة الإدراكية: قد تصاب بتشتت في الانتباه أو صعوبة في التركيز بسبب نقص الحديد في الدم.
- برودة اليدين والقدمين: قد تشعر ببرودة غير عادية في الأطراف مثل اليدين والقدمين نتيجة لنقص الحديد في الدم.
- الصداع: يمكن أن يكون الصداع واحدًا من الأعراض التي تصاحب نقص الحديد في الدم، وقد يكون شديدًا في بعض الحالات.

تم تحليل نموذج الانحدار الخطي المتعدد لظاهرة نقص الحديد في الدم، مع افتراض وجود المتغيرات التالية:

جدول رقم (4) يبين المتغير المعتمد و المتغيرات المستقلة لنموذج الانحدار الخطي المتعدد

المتغير	العامل	المتغير	العامل
y	حجم الكريات الحمراء	x10	نسبة كريات الدم البيضاء
x1	مستويات الحديد في الدم	x11	نسبة الخلايا اللمفاوية في الدم
x2	مستويات الهيموغلوبين	x12	عدد كريات الدم البيضاء في الدم
x3	عمر المريض	x13	الوزن ومؤشر كتلة الجسم
x4	عدد خلايا الدم الحمراء	x14	متوسط تركيز الهيموجلوبين في كريات الدم الحمراء
x5	عدد الصفائح الدموية	x15	نسبة الصفائح الدموية
x6	معدل التنفس	x16	هرمون البروكالستين
x7	مستويات الأكسجين في الجسم	x17	نسبة السكر في الدم
x8	وظائف القلب والشرايين	x18	نسبة اليوريا في الدم
x9	وظائف الجهاز المناعي	X19	الكرياتين في الدم

سيتم تقدير معالم النموذج الخطي المتعدد التالي :

$$y_i = \sum_{i=1}^{19} X_i \hat{\beta}$$

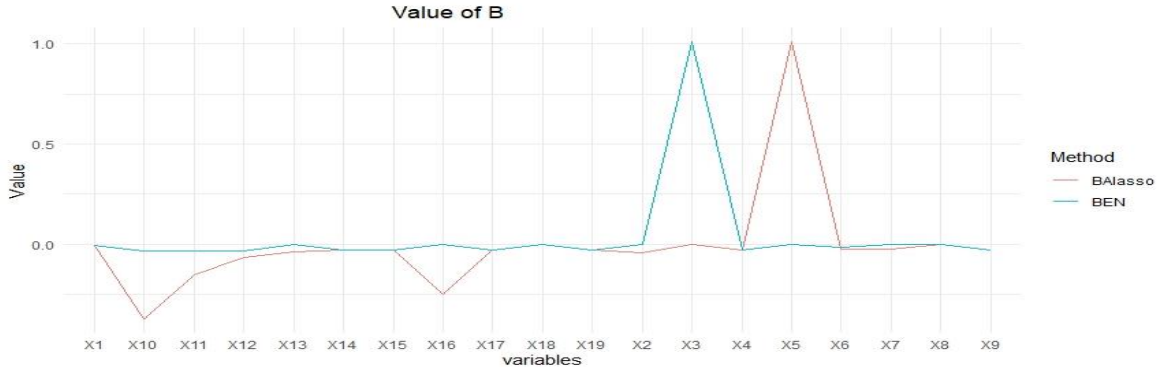
الجدول ادناه يبين قيم المعالم لنموذج الانحدار المتعدد اعلاه باستعمال طريقة (BEN) و كذلك طريقة (BALasso) .

جدول رقم (5) يبين القيم التقديرية لمعالم نموذج الانحدار

Variables	Alasso	BEN
X1	-0.00693	-0.00634
X2	-0.04342	0
X3	0	1.00877
X4	-0.02861	-0.0285

X5	1.00856	0
X6	-0.02535	-0.01505
X7	-0.02416	0
X8	0	0
X9	-0.02987	-0.03023
X10	-0.37189	-0.03288
X11	-0.15232	-0.03219
X12	-0.06432	-0.03041
X13	-0.03674	0
X14	-0.02637	-0.02661
X15	-0.02879	-0.02874
X16	-0.24832	0
X17	-0.02609	-0.02641
X18	0	0
X19	-0.02661	-0.02689

من خلال الجدول رقم (5) نلاحظ بان طريقة (BEN) قامت باستبعاد سبعة متغيرات غير ليس لها تأثير في النموذج وهي (x2,x5,x7,x8,x13,x16,x18) أي انها ليس لها تأثير مباشر على نقص الحديد في الدم .وعلية نجد ان متغير (مستويات الهيموغلوبين) ليس له تأثير مباشر على نقص الحديد بالدم , وكذلك متغير (عدد الصفائح الدموية) ايضا ليس له تأثر على نقص الحديد بالدم , اما المتغيرين(مستويات الأكسجين في الجسم) و (وظائف القلب والشرايين) ايضاً لم يكن لهما تأثير ملحوظ على المرض , ونفس الحال للمتغيرات (الوزن ومؤشر كتلة الجسم) و (هرمون البروكالستين) و (نسبة اليوريا في الدم) ليس لهما تأثير على المرض . اما بالنسبة للمتغيرات التي كان لها تأثير على نقص الحديد بالدم هي (x1,x3,x4,x6,x9,x10,x11,x12,x14,x15,x17,x19) .



شكل رقم (4) يبين قيم المعالم المقدرة لطريقة BEN و BALasso

الاستنتاجات

1- تحديد المتغيرات الرئيسية التي تؤثر على نقص الحديد في الدم وهي (مستويات الحديد في الدم ,عمر المريض ,عدد خلايا الدم الحمراء ,معدل التنفس ,وظائف الجهاز المناعي ,نسبة كريات الدم البيضاء ,نسبة الخلايا اللمفاوية في الدم ,عدد كريات الدم البيضاء في الدم ,متوسط تركيز الهيموجلوبين في كريات الدم الحمراء ,نسبة الصفائح الدموية ,نسبة السكر في الدم ,الكرياتين في الدم)

2- تحديد المتغيرات التي ليست لها تأثير على نقص الحديد في الدم وهي (مستويات الهيموغلوبين ,عدد الصفائح الدموية ,مستويات الأكسجين في الجسم ,وظائف القلب والشرايين ,الوزن ومؤشر كتلة الجسم ,هرمون البروكالستين ,نسبة اليوريا في الدم)

3- طريقة الشبكة المرنة البيزية لها افضلية في اختيار و تقدير معلمات نموذج الانحدار بالمقارنة مع طريقة لاسو التكيفي البيزي

التوصيات

1- اعتماد ما توصل الية الباحثان من استنتاجات بخصوص العوامل التي لها تاثير ملحوظ على نقص الحديد بالدم في المستشفيات للحد من خطورة المرض على المصابين به .

2- توعية المصابين حول خطورة المرض و اهمية التشخيص بوقت مبكر من خلال حملات و ندوات من قبل وزارة الصحة و الجهات المختصة .

3- استعمال طرق حديثة للكشف عن العوامل المؤثرة على مرض نقص الحديد مثل طريقة معكوس لاسو البيزية .

1- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.

2- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.

3- Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian elastic net regression. Communications in Statistics-Simulation and Computation, 47(4), 1168-1178.

4- Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571.

5- Li, Q., & Lin, N. (2010). The Bayesian elastic net. Bayesian Analysis 5(1):151–170.

6- McDonald, G. C. (2009). Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 93-100.

7- De Mol, C., De Vito, E., & Rosasco, L. (2009). Elastic-net regularization in learning theory. Journal of Complexity, 25(2), 201-230.

8- AL-Sabbah, S. A., & Raheem, S. H. (2021). USE BAYESIAN ADAPTIVE LASSO FOR TOBIT REGRESSION WITH REAL DATA. International Journal of Agricultural & Statistical Sciences, 17.

9- Raheem, S. H. (2017). The use of factor analysis to identify the most important factors influencing the migration Of Iraqi youth: A statistical study of the Status the migration of young people in Diwaniyah province. Journal of Al-Qadisiyah for computer science and mathematics, 9(1), 1-11.