



المقارنة بين بعض الطرق الجزائية في وجود التعدد الخطي: دراسة تطبيقية على عينة من مرضى الفشل الكلوي المزمن بمركز خدمات الكلى بمدينة الخمس

سالمة بالعيدة^{a*} وحنان الرشي^b

^aقسم الأحصاء، كلية العلوم، الجامعة الاسمية الإسلامية، زليتن، ليبيا.
^bقسم الأحصاء، كلية العلوم، جامعة المرقب، الخمس، ليبيا.

الكلمات المفتاحية:

التعدد الخطي.
انحدار الشبكة المرن.
انحدار الحرف.
انحدار الحرف القوي.
انحدار لاسو.

الملخص

تناولت هذه المقالة بعض طرق الجزء لمعالجة مشكلة التعدد الخطي والمتمثلة في (انحدار لاسو، انحدار الحرف، انحدار الحرف القوي، انحدار الشبكة المرن)، ومقارنتها بطريقة المربعات الصغرى. حيث تم الاعتماد على أسلوب المحاكاة في ظل وجود درجات متفاوتة من مشكلة التعدد الخطي، من خلال توليد عشر متغيرات مستقلة عند حجوم عينات مختلفة تراوحت من 50 إلى 450 مشاهدة، وكذلك تم توليد 50 متغير مستقل عند حجم عينة 1000 مشاهدة، لمقارنة أداء طرق التقدير المقترحة وبيان أفضلها بالاعتماد على معيار متوسط مربعات الخطأ، معامل التحديد، ومعامل التحديد المعدل. كما تم الاعتماد في هذه المقالة على بيانات واقعية لمرضى الفشل الكلوي المزمن، تم جمعها من مركز خدمات الكلى بمدينة الخمس (يناير: أغسطس 2023). بلغ عدد المشاهدات 100 مشاهدة، وتم استبعاد 34 لعدم توفر البيانات عنها. تضمنت البيانات 14 متغيراً (العمر، الهيموجلوبين، اليوريا، البوتاسيوم، الفسفور، الكالسيوم، البروتين، حمض اليوريك، الماغنيسيوم، الكوليسترول، الدهون الثلاثية، فيتامين-د، معدل الترشيح الكبيبي، الكرياتينين). وباستخدام برنامج R اصدار (4.5.1)، بينت النتائج مرونة انحدار لاسو في التعامل مع البيانات ذات الارتباطات المتعددة، وأنها أكثر كفاءة من الطرق المدروسة عند تحليل البيانات الحقيقية، وعند مختلف تجارب المحاكاة عبر أحجام عينات مختلفة. وعلاوة على ذلك، بينت نتائج تحليل البيانات الحقيقية أن معدل الترشيح الكبيبي لمرضى الكلى المزمن يتأثر سلباً بمستوى (الكرياتينين، الماغنيسيوم، العمر، الفسفور، البروتين، الكوليسترول)، بينما يتأثر إيجاباً بمستوى (الهيموجلوبين، حمض اليوريك، فيتامين. د). وأخيراً، توصي المقالة بالاعتماد أسلوب لاسو كخيار مفضل عند معالجة مشكلة التعدد الخطي، خصوصاً عندما تكون البيانات خالية من القيم الشاذة ويتناسب حجم العينة مع عدد المتغيرات، بغض النظر عن درجة الارتباط بينها، وذلك لقدرتها على تحقيق توازن بين الدقة والبساطة في النموذج.

Comparison Between Some Penalized Regression Methods in the Presence of Multicollinearity: An Applied Study on Chronic Renal Failure Patients

Salma Bleed^{a*}, Hanan Al-Rabshi^b

^aDepartment of Statistics, Faculty of Science, Asmarya Islamic University, Zliten, Libya.

^bDepartment of Statistics, Faculty of Science, Elmergib University, Al-Khums, Libya.

Keywords:

Elastic Net regression.
Lasso Regression.
Multicollinearity.
Ridge Regression.
Robust Ridge Regression.

ABSTRACT

This article discusses some penalty methods for addressing the problem of multicollinearity (Lasso . ridge .robust ridge .elastic net), and compares them with the least squares method. Also, it relies on a simulation approach under varying degrees of multicollinearity .ten independent variables were generated at different sample sizes ranging from 50 to 450 observations .and 50 independent variables were generated at a sample size of 1,000 observations, to compare and demonstrate the best proposed methods based on the MSE criteria, the coefficient and the adjusted coefficient of determination. This article also relies on real- data on chronic kidney failure patients .collected from

*Corresponding author.

E-mail addresses: s.ali@asmarya.edu.ly, (H. Al-Rabshi) younushanan848@gmail.com.

Article History : Received 02 October 25 - Received in revised form 20 February 26 - Accepted 10 March 26

the Kidney Services Center in Al-Khums City (January to August 2023). The number of observations reached 100, and 34 were excluded due to unavailability of data. The data included 14 variables (age, hemoglobin, urea, potassium, phosphorus, calcium, protein, uric acid, magnesium, cholesterol, triglycerides, vitamin-D, glomerular filtration rate, and creatinine). Using R version 4.5.1, the results demonstrated the flexibility of the Lasso method in handling data with multiple correlations, and to be more efficient than the other studied methods when analyzing real data, in various simulations across different sample sizes. Furthermore, the results of the real data analysis showed, the GFR of chronic kidney disease patients is negatively affected by levels of creatinine, magnesium, age, phosphorus, protein, and cholesterol, while it is positively affected by levels of hemoglobin, uric acid, and vitamin-D. Finally, the article recommends adopting the Lasso method as the preferred option when dealing with multicollinearity, especially when the data are free of outliers and the sample size is proportional to the number of variables, regardless of the degree of correlation between them, due to its ability to achieve a balance between accuracy and simplicity in the model.

1. المقدمة

عند غياب العلاقة الخطية بين المتغيرات التفسيرية غياباً تاماً يقال عن هذه المتغيرات أنها متعامدة، ولكن في أغلب تطبيقات الانحدار تكون المتغيرات التفسيرية غير متعامدة، ومرتبطة ارتباطاً قوياً حيث يصعب تقدير تأثير كل متغير تفسيري تقديراً منفرداً في النموذج. فعند حدوث هذه المشكلة، ينعكس ذلك على أداء طريقة تقدير المربعات الصغرى (التقليدية)، بحيث تكون غير كفاءة ومتحيزة وذات تباينات كبيرة. ومن هذا المنطلق نسعى في هذه المقالة إلى تصغير مجموع مربعات البواقي من خلال استخدام طرق معالجة مشكلة التعدد الخطي بين المتغيرات التفسيرية، والمتمثلة في (انحدار لاسو - انحدار الحرف - انحدار الحرف القوي - انحدار الشبكة المرنة)، وكيفية تطبيقها والمقارنة بينها وبين الطريقة التقليدية للوصول إلى أفضلها في علاج مشكلة التعدد الخطي بالاعتماد على عدة معايير (متوسط مربعات الخطأ - معامل التحديد - معامل التحديد المعدل) من خلال التطبيق على بيانات تجريبية، وبيانات واقعية خاصة بمرضى الفشل الكلوي بمركز خدمات الكلى بمدينة الخمس .

مؤخراً تناولت عدة دراسات موضوع التقدير باستخدام الطرق الجزائرية لتقدير معالم نموذج الانحدار الخطي المتعدد في ظل وجود مشكلة التعدد الخطي، فقد استخدم [1] انحدار الحرف لتقدير معالم نموذج الانحدار الخطي بعد التخلص من مشكلة التعدد الخطي باستخدام بيانات الأصداء الجوية لمحطة مطار طرابلس الدولي الشهرية للفترة (2000-2010) لدراسة تأثير عدة عوامل مناخية على كمية سقوط الأمطار، وقد تبين أفضلية انحدار الحرف بسبب قدرتها على تحقيق أقل متوسط مربعات الخطأ. قام [2] بالمقارنة بين الطريقة التقليدية وانحدار الحرف وانحدار لاسو وطريقة المكونات الرئيسية بالتطبيق على بيانات الهجرة الداخلية في مصر خلال الفترة من (1996-2022) بيانات سنوية، وأظهرت النتائج أن انحدار لاسو قدم أداءً أفضل من الطرق الأخرى. قام [3] باختيار أفضل نموذج انحدار للجلطة الدماغية باستخدام بعض الطرائق الجزائرية، حيث تناول المقارنة بين (طريقة الحرف، طريقة لاسو التكميلية، طريقة لاسو، طريقة انحدار الشبكة المرنة). ويهدف دراسة هذه الطرائق والمقارنة بينها تم تطبيق ثلاث تجارب محاكاة بحجم عينات تراوحت من 50 إلى 250 مشاهدة وارتباطات مختلفة. وتم تطبيق الدراسة على بيانات حقيقية تم جمعها من مستشفى الناصرية التعليمي (دائرة ذي قار) للمرضى المصابين بالجلطة الدماغية بتاريخ (26/10/2022)، ولقد أظهرت هذه الدراسة أفضلية طريقة انحدار الشبكة المرنة سواء على مستوى الدقة ومستوى القدرة التفسيرية للنموذج.

استخدم [4] أسلوب انحدار الحرف وذلك بالمقارنة مع الطريقة التقليدية على بيانات خاصة بحجم الهجرة الداخلية في جمهورية مصر العربية خلال الفترة (1996-2015)، حيث تبين أفضلية انحدار الحرف على الطريقة التقليدية. قام [5] بدراسة تأثير تكنولوجيا المعلومات والاتصالات على النمو الاقتصادي في مصر خلال الفترة (2000-2020) باستخدام النموذج الخطي المعمم، وتمت المقارنة بين الطرق الجزائرية (انحدار لاسو، انحدار الحرف، وانحدار الشبكة المرنة)، وأظهرت النتائج أن انحدار الشبكة المرنة، وانحدار لاسو هما الأفضل من بين طرق الانكماش. قام [6] بتقدير معالم نموذج الانحدار الخطي المتعدد في ظل وجود التعدد الخطي مستخدماً عينة عشوائية من عدة مراكز صحية تابعة لدائرة صحة بابل (100) مفردة تخص خصوبة المرأة، بهدف دراسة العوامل المؤثرة على عدد الأطفال المولودين. وتم التوصل إلى أن انحدار لاسو أفضل من الطريقة التقليدية، وانحدار الحرف البيزية الغير متحيزة. قام [7] بتقييم أداء انحدار التلال، والشبكة المرنة، وأساليب لاسو في معالجة مشكلات التعدد الخطي، باستخدام بيانات معدل وفيات الرضع في مقاطعة جنوب سولاويزي. وقد بينت النتائج أن أسلوب الشبكة المرنة يتفوق على كل من أسلوب ريدج و لاسو. وقد تم الحصول على النموذج الأفضل أداءً من خلال الشبكة المرنة بقيمة معامل تحديد 60.81%، بينما حققت طريقتا ريدج و لاسو قيم معامل تحديد 54.11% و 58.18% على التوالي. وهذا يُظهر أن تطبيق طريقة الشبكة المرنة قادر على تحقيق نتائج أكثر دقة في نمذجة المتغيرات ضمن تحليل بيانات معدل وفيات الرضع مقارنةً بالطرق الأخرى. قام [8] بمقارنة أداء المربعات الصغرى العادية (OLS) ومشغل الانكماش والاختيار المطلق الأدنى (LASSO) والانحدار التلافي (RR) والشبكة المرنة في التحكم في مشاكل التعدد الخطي بين المتغيرات المستقلة في تحليل الانحدار المتعدد باستخدام بيانات المحاكاة وبيانات الحالة. استخدمت محاكاة البيانات نموذج انحدار متعدد مع 6 متغيرات بمستوى عالٍ من التعدد الخطي عند أحجام عينات متعددة. تم قياس أفضل طريقة بناءً على أصغر (AMSE) وقيم AIC، وقد بينت نتائج البحث أن الشبكة المرنة هي أفضل طريقة للبيانات المحاكاة مقارنةً بـ LASSO و Ridge AIC لكل حجم عينة تمت دراستها. إضافة لذلك، تم الحصول على نتائج مماثلة أيضاً عند تطبيق هذه الطرق الثلاث على البيانات المتعلقة بحالات التقرن لدى الأطفال الصغار في إندونيسيا والتي كانت ذات تعدد خطي مرتفع، وباستخدام أفضل طريقة، بينت نتائج البيانات الحقيقية أن حالات الأطفال الصغار المتقرنين في إندونيسيا تتأثر بنسبة الأطفال الصغار الذين يعانون من سوء التغذية، ونسبة الأطفال الصغار الذين يتلقون حليب الأم الخالص،

$\hat{B} = \min \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$
وتكون معادلة التقلص أو الانكماش كالتالي:

$$\hat{B} = \min \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + k \sum_{j=1}^p \beta_j^2$$

$$\hat{B} = \min \left\{ \text{SSE} + k \sum_{j=1}^p \beta_j^2 \right\}$$

وبتصغير مجموع مربعات الخطأ نحصل على مقدرات انحدار الحرف كالتالي:

$$\hat{\beta}_{RR} = [XX + KI_p]^{-1} (XY)$$

حيث أن

\hat{B}_{RR} : متجه المعلمات المقدرة من انحدار الحرف، K معلمه التحيز وبتزايد قيمة (K) فإن مقدار التحيز يزيد والتباين يقل وبالتالي يجب اختيار (K) بحيث يكون الانخفاض في قيم التباين أكثر من الارتفاع في مقدار مربع التحيز، عند ذلك فإن متوسط مربعات الخطأ لانحدار الحرف يكون أقل من تباين مقدرات الطريقة التقليدية. كما أن زيادة (K) تقلل من قيمة معامل التحديد (R^2) وبالتالي فإن مقدرات انحدار الحرف ليس من الضروري أن تعطي أفضل نموذج ملائم للبيانات بل تبحث عن أفضل معادلة ذات مقدرات ثابتة [2]. والافتراضات المستخدمة في انحدار الحرف تشابه افتراضات الطريقة التقليدية إلا أن انحدار الحرف لا يوفر حدوداً للثقة، [12]، [11].

انحدار الحرف القوي: للتغلب على مشكلة القيم الشاذة، تم تطوير انحدار الحرف القوي عن طريق دمج تقنيات الانحدار المقاوم (**Robust Regression**) مع عقوبة ريدج. فبدلاً من استخدام معيار المربعات الصغرى يتم استبداله بدالة خسارة (Loss function) مثل **Huber** أو **Tukey's Biweight**، هذه الدوال تعطي وزناً أقل للأخطاء الكبيرة الناتجة عن القيم الشاذة. ونحصل على مقدرات انحدار الحرف القوي كالتالي:

$$\hat{B} = \min \left\{ p \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right) + k \sum_{j=1}^p \beta_j^2 \right\}$$

حيث:

$p(\cdot)$: دالة خسارة مقاومة (Robust loss function)

مثلاً: دالة هوبر (Huber) تُعرف كالتالي:

$$p(u) = \begin{cases} 0.5 u^2 & \text{if } |u| \leq \delta \\ 0.5 \delta - |u| \delta^2 & \text{if } |u| > \delta \end{cases}$$

وهذا الشكل، النموذج يجمع بين الانتظام (**Regularization**) عبر عقوبة ريدج، والمقاومة للقيم الشاذة عبر دوال خسارة رويست [13]، [14].

انحدار لاسو: يعد من أنواع الانحدار الخطي وصمم اساساً لنماذج الطريقة التقليدية، ويقوم على فكرة الانكماش (Shrinkage) لمعاملات نموذج الانحدار. ويعد نموذج لاسو من أفضل النماذج ملائمة في حالة تعرض نموذج الانحدار الخطي إلى درجة عالية من الأزواج الخطي. حيث يعمل على تصغير مجموع مربعات الأخطاء وذلك بناء على قيد أو شرط معين يتم فرضه على المعلمات. ومن خلال هذا القيد فإن مقدر لاسو يعمل على جعل عدد من المعلمات لنموذج الانحدار مساوية للصفر وتقليص باقي المعلمات بمقدار معين. بالإضافة إلى اختيار وتنظيم المتغيرات الداخلة في النموذج. وفي هذه الطريقة هناك معلمة تسمى معلمة ضبط وهي التي تتحكم في معاملات نموذج الانحدار، فعندما تكون المعلمة كبيره تضطر بعض المعاملات أن تكون مساوية للصفر وعندما تكون معلمة الضبط

ونسبة الأطفال الصغار الذين يتم مراقبة نموهم ، وتغطية الخدمات الصحية للنساء الحوامل ، وعدد العاملين في مجال التغذية ، ونسبة الأسر التي لديها مياه شرب كافية ، ونسبة الأسر التي لديها صرف صحي مناسب ، ومؤشر التنمية البشرية، والكثافة السكانية. قام [9] بأجراء مقارنة بين طرق علاج التعدد الخطي، حيث تم المقارنة بين (الطريقة التقليدية، الطريقة التقليدية ثنائية المرحلتين، انحدار التلال، انحدار لاسو) مستخدماً أسلوب المحاكاة، وقد أظهرت النتائج أن انحدار الحرف ولاسو الأكثر فعالية في التعامل مع التعدد الخطي.

تستمد هذه المقالة أهميتها من خلال إيجاد حل لخطورة مشكلة التعدد الخطي وتأثيرها على نموذج الانحدار الخطي المتعدد، وعلى مقدرات طريقة المربعات الصغرى، وخطورة استخدام الطرق التقليدية في علاج مشكلة التعدد الخطي (حذف المتغيرات الأكثر ارتباطاً)، وقد يكون منها ما هو ذات أهمية تنبؤية عالية في تفسير الظاهرة محل الدراسة، وحذفها يضعف من قدرة النموذج التفسيرية، بالإضافة للوقوع في خطأ توصيف النموذج. كما تستمد هذه المقالة أهميتها من خلال استخدام انحدار لاسو لعلاج مشكلة التعدد الخطي إضافة إلى استخدام انحدار الحرف، وانحدار الحرف القوي، وانحدار الشبكة المرنة. كطرق حديثة في علاج مشكلة التعدد الخطي وذلك بالتطبيق على بيانات تجريبية، وبيانات حقيقية خاصة بمرضى الفشل الكلوي بمركز خدمات الكلى بمدينة الخمس.

2. المواد وطرق العمل

يعتبر تحليل الانحدار أسلوب إحصائي لتحليل البيانات، يدرس العلاقة بين المتغيرات عن طريق بناء نموذج رياضي يفسر العلاقة بين متغير تابع ومتغير مستقل واحد أو عدة متغيرات مستقلة.

حيث يصف الانحدار الخطي البسيط العلاقة بين متغير تابع واحد ومتغير مستقل واحد. بينما يصف الانحدار الخطي المتعدد العلاقة بين متغير تابع واحد وعدة متغيرات مستقلة. ويعتبر من التقنيات الإحصائية الأكثر استخداماً في مختلف المجالات، وهو عبارة عن معادلة رياضية تعبر عن العلاقة بين متغيرين أو أكثر من المتغيرات المستقلة مع متغير تابع واحد، أو بمعنى آخر هو عبارة عن انحدار المتغير التابع Y على العديد من المتغيرات المستقلة (X_1, X_2, \dots, X_p) ، حيث تعرف الصيغة الرياضية للنموذج الخطي العام على النحو التالي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + e_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + e_i, \quad i=1,2,3,\dots,n$$

حيث:

Y_i : المتغير التابع، β_0 : الحد الثابت، β_j : ميل انحدار المتغير التابع Y على المتغير المستقل، X_j : المتغيرات المستقلة، e_i : حد الخطأ العشوائي

انحدار الحرف: إحدى الطرق المستخدمة في تحليل الانحدار المتعدد عندما يعاني من مشكلة الأزواج الخطي. حيث تقوم فكرة انحدار الحرف عن طريق إضافة مقدار صغير موجب ($k \geq 0$) يعرف بمعلمة التحيز إلى عناصر القطر الرئيسي في المصفوفة (XX) ، والذي يعمل على تقليص مقدرات الانحدار وتخفيض تباين المعلمات المقدرة والحصول على مجموع مربعات خطأ أقل. تقوم انحدار الحرف بتقدير المعلمات بحيث تقلل مجموع مربعات الخطأ (SSE) كالتالي:

$$\widehat{\beta}^{\text{ridge}} = \beta^{\text{Argmin}} \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 / \sum_{j=1}^p \beta_j^2 \leq t$$

وتكتب حسب دالة لاجرانج بالصيغة التالية

$$\widehat{\beta}^{\text{ridge}} = \beta^{\text{Argmin}} \left(\sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right)$$

وعلى النقيض من انحدار لاسو، نجد أن انحدار الحرف لا تختار أفضل المتغيرات المستقلة، بل هي طريقة انكماش فقط وكلما كبرت قيمة (λ) يحدث انكماش في المتغيرات المستقلة إلى أن تصل قيمة المقدرات إلى الصفر. وعليه قام [13] باقتراح انحدار الشبكة المرنة المعرفة كما يلي:

$$\widehat{\beta}^{\text{E.Net}} = \beta^{\text{Argmin}} \left(\sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^p \left((1-\alpha) \frac{1}{2} \beta_j^2 + \alpha |\beta_j| \right) \right)$$

أي أنه قام بمزج كلا من انحدار لاسو وانحدار الحرف من خلال معاملي التقليل (L₁) و (L₂). ويتم الحصول على تقديرات المعلمات حسب قيم الضبط λ، ومراقبة أوزان كلا من (L₁, L₂) حسب قيمة (λ)، فإذا كانت قيمة 0λ = نحصل على انحدار ريدج، وإذا كانت قيمة 1λ = نحصل على انحدار لاسو، أما إذا كانت قيمة (0 < λ < 1) نحصل على انحدار الشبكة المرنة [15].

3. النتائج والمناقشة

يتضمن هذا البند المحاكاة التجريبية والتي تهدف إلى تقييم فعالية بعض طرق التقدير المختلفة في ظل أوضاع بيانات متباينة من حيث حجم العينة ومستويات الارتباط بين المتغيرات المستقلة. وكذلك تضمن تطبيق بعض النماذج الإحصائية على بيانات مرضى الفشل الكلوي بمرکز خدمات الكلى مدينة الخمس خلال الفترة (يناير 2023 - أغسطس 2023).

1.3. المحاكاة التجريبية

تم الاعتماد على أسلوب المحاكاة (Simulation) لتوليد بيانات بهدف مقارنة أداء خمس طرق للتقدير (الانحدار التقليدي (OLS)، انحدار لاسو (LASSO)، انحدار الحرف (Ridge)، انحدار الحرف القوي (Robust)، انحدار الشبكة المرنة (Elastic NET))، وقد تم استخدام طريقة Cross-Validation لتحديد قيم المعلمة λ. حيث استندت المقارنة إلى مجموعة من المؤشرات الإحصائية (متوسط مربع الخطأ (MSE)، معامل التحديد (R²)، معامل التحديد المعدل (R²_{adj})). وفي ظل وجود درجات مختلفة من مشكلة التعدد الخطي، تم توليد عشر من المتغيرات المستقلة عند حجوم عينات (50، 150، 250، 450)، وتم توليد 50 من المتغيرات المستقلة عند حجم عينة (1000). حيث توجد ثلاث متغيرات مؤثرة فقط، بينما باقي المتغيرات غير مؤثرة، وذلك لتمكين اختبار قدرة لاسو على اختيار المتغيرات المهمة والتخلص من غير المؤثرة. تم توليد المتغيرات المستقلة من التوزيع الطبيعي متعدد المتغيرات بمتوسط صفر وتباين متكافئ يسمح بالتحكم في درجة الارتباط بين المتغيرات، بافتراض أربع حالات مختلفة لمعامل الارتباط (0.50، 0.65، 0.70، 0.95) لتمثيل درجات مختلفة من التعدد الخطي. تم توليد المتغير التابع بافتراض وجود نموذج الأنحدار الخطي المتعدد، وقد تم الحصول على النتائج الموضحة في الجداول من 1 إلى 5.

ومن خلال نتائج المحاكاة، أظهرت النتائج أن انحدار لاسو (LASSO) حقق أفضل أداء مقارنة بطرق التقدير (الانحدار التقليدي (OLS)، انحدار الحرف (Ridge)، انحدار الحرف القوي (Ridge Robust)، انحدار الشبكة المرنة

مساوية للصفر سنحصل على مقدرات الطريقة التقليدية. ويتميز انحدار لاسو في حالة وجود عدد من المشاهدات أقل وعدد كبير من المتغيرات، حيث أن تقلص أو إزالة المتغيرات تخفض التباين دون الزيادة الكبيرة في درجة التحيز. كما أنه يسهل عملية تفسير النموذج من خلال استبعاد المتغيرات غير المرتبطة بالمتغير التابع، ويساعد على زيادة إمكانية تفسير النموذج من خلال القضاء على المتغيرات التي لا ترتبط بمتغير الاستجابة. فإذا كان هناك عينة حجمها n من الحالات كل حاله تحتوي على P من المتغيرات المستقلة ومتغير تابع واحد Y_i، وليكن X متجه المتغيرات المستقلة للحالة ith، فيكون هدف انحدار لاسو هو الوصول لحل المعادلة الآتية:

$$\min \left(\frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - x_i' \beta)^2 \right), \text{ subject to } \sum_{j=1}^p |\beta_j| \leq t$$

t: معلمة تحدد مسبقاً وتمثل مقدار الانكماش (التقليل)

x: مصفوفة المتغيرات المستقلة

X_i: هي الصف ith من المصفوفة X

ويمكن صياغة لاسو كالتالي:

$$\min_{\beta_0, \beta} \left(\frac{1}{n} \|y - \beta_0 I_n - x\beta\|_2^2 \right), \text{ subject to } \|\beta\| \leq t$$

حيث

$$\|\beta\|_p = \left(\sum_{i=1}^n |\beta_i|^p \right)^{\frac{1}{p}}$$

وعندما p=1 يصبح $\|\beta\|_1$ الطول القياسي L¹، وتحويل المتغيرات إلى قيم معيارية نحصل على الصيغة الآتية:

$$\min_{\beta_0, \beta} \left(\frac{1}{n} \|y - x\beta\|_2^2 \right), \text{ subject to } \|\beta\| \leq t$$

ويكون بصيغة مضاعف لاجرانج كالتالي:

$$\min_{\beta \in R^p} \left(\frac{1}{n} \|y - x\beta\|_2^2 + \lambda \|\beta\|_1 \right)$$

حيث أن (λ) معلمة الضبط والتحكم والانكماش لمعاملات الانحدار. وبالتالي فإن انحدار لاسو يتميز عن انحدار الحرف، في أنه يحول القيم إلى صفر (جعلها صفر تماماً إذا كانت صغيرة بما فيه الكفاية) بدلاً من وضع قيم أقرب إلى الصفر وترك القيم الأكبر دون مساس. أي أن تقديرات لاسو تجمع بين مميزات انحدار الحرف وانحدار أفضل اختيار جزئي والتي تحول جميع المعاملات إلى صفر بقيمة ثابتة وتضبطها إلى الصفر إذا وصلت إليها [10]. [6]

انحدار الشبكة المرنة: أن انحدار الشبكة المرنة يعمل كأسلوب لاختيار المتغيرات، وتقليل المعالم الأخرى واختبار مجموعة من المتغيرات التفسيرية المرتبطة، وبالتالي يمكن أن تتغلب على مشكلة عدد المتغيرات التوضيحية أكبر من عدد المشاهدات. حيث تعتمد على دمج دالة الجزء (Ridge) ودالة جزء (Lasso)، وبالتالي يمكن القول إن انحدار الشبكة المرنة هو طريقة تنظيم تستخدم لتحسين أسلوب اختيار المتغيرات في حالة هناك ارتباطات قوية بين مجاميع من المتغيرات التوضيحية. وتعتبر انحدار الشبكة المرنة مشابهة لانحدار لاسو حيث تضيق إلى معامل التقليل (L₁) معامل تقلص آخر (L₂)، حيث

$$L_1 = \sum_{j=1}^p |\beta_j|, \quad L_2 = \sum_{j=1}^p \beta_j^2$$

حيث

جدول 3: نتائج المقارنة بين طرق التقدير عند حجم عينة 250 مشاهدة

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.50	OLS	5.486918e-136	0.9353171	0.9326107	0.9912097
	LASSO =0.01921753 λ	2.892784e-140	0.9349268	0.9330522	0.9807764
	Ridge =0.0205677 λ	3.760311e-129	0.9253888	0.92228	1.138589
	Ridge robust	1.814094e-136	0.9351841	0.9324834	0.9891098
	Elastic Net =0.02199395 λ	2.166456e-136	0.9350879	0.9323832	0.9905774

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.70	OLS	1.248108e-130	0.9282658	0.9252644	0.9912097
	LASSO =0.01446009 λ	1.912758e-133	0.9277946	0.9254076	0.9853522
	Ridge =0.0188639 λ	1.699599	0.9135467	0.9099445	1.189618
	Ridge robust	445329e-131	0.9281183	0.9251232	0.9891098
	Elastic Net =0.02187709 λ	128698e-131	0.9278386	0.9251325	0.9929583

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.95	OLS	1.204574e-127	0.924008	0.9208284	0.9912097
	LASSO =0.00785907 λ	3.089269e-130	0.9232398	0.9207022	0.9888177
	Ridge =0.0323794 λ	7.551838e-112	0.895898	0.8915604	152207
	Ridge robust	416254e-128	0.9238517	0.9206788	0.9891098
	Elastic Net =0.00987145 λ	1.451734e-128	0.9234959	0.920627	0.9937305

جدول 4: نتائج المقارنة بين طرق التقدير عند حجم عينة 450 مشاهدة

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.50	OLS	2.046379e-254	0.9356876	0.932226	0.9477279
	LASSO =0.01191301 λ	1.093222e-257	0.9355365	0.9343697	0.9435064
	Ridge =0.01642 λ	1.569526e-241	0.9259312	0.9242478	1.08902
	Ridge robust	6.87539e-255	0.9356078	0.9341444	0.9467465
	Elastic Net =0.01978078 λ	1.716584e-256	0.935529	0.9343568	0.9479055

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.65	OLS	3.687052e-247	0.9305946	0.9290136	0.9477279
	LASSO =0.01765765 λ	5.646249e-253	0.9301414	0.9291973	0.9431739
	Ridge =0.0232653 λ	8.241499e-230	0.9162487	0.9143453	1.141021
	Ridge robust	1.286847e-247	0.9305085	0.9289291	0.9467465
	Elastic Net =0.01096494 λ	1.428549e-250	0.9300686	0.929115	0.9527393

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.95	OLS	144677e-246	0.9301837	0.9285933	0.9477279
	LASSO =0.00618852 λ	7.427628e-251	0.929701	0.9285902	0.9456632
	Ridge =0.0459886 λ	8.511451e-222	0.9089076	0.9068373	1.233731
	Ridge robust	4.707025e-247	0.9300971	0.9285084	0.9467465
	Elastic Net =0.01027564 λ	4.277274e-249	0.929673	0.9285541	0.9524906

(Elastic NET))، حيث حقق أقل قيمة متوسط مربعات الخطأ (MSE) عند مختلف قيم الارتباطات، وعملت كأسلوب اختيار المتغيرات مما يعكس طبيعتها إذ أنها تعمل على تصفير معالم المتغيرات التي ليس لها تأثيراً على المتغير التابع وتقليص المعالم مما يعكس تفوقها في معالجة التعدد الخطي وقدرتها على التعامل مع جميع مستويات الارتباط. ثم يأتي انحدار الحرف القوي، ويلهما انحدار الشبكة المرن، وأخيراً انحدار الحرف، على الترتيب من ناحية جودة الأداء عند قيم ارتباط مختلفة، وعند حجوم عينات مختلفة.

جدول 1: نتائج المقارنة بين طرق التقدير عند حجم عينة 50 مشاهدة

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.50	OLS	1.464292	0.9521408	0.9398692	0.8649478
	LASSO =0.07576142 λ	7e-21.077663	0.9479151	0.9421279	0.8158084
	Ridge =0.0130471 λ	1e-21.122145	0.9429793	0.9287241	1.004759
	Ridge robust	e-22434656	0.9516245	0.9395306	0.852422
	Elastic Net =0.07900413 λ	e-246.025228	0.9498144	0.941032	0.8843169

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.7	OLS	1.526728e-21	0.9459585	0.9321017	0.8649478
	LASSO =0.04832221 λ	1.23436 e-25	0.9425788	0.9347486	0.8146044
	Ridge =0.0179279 λ	2.761048e-20	0.9329396	0.9161745	1.046485
	Ridge robust	4.801026e-22	0.9453755	0.9317193	0.852422
	Elastic Net =0.0880588 λ	1.219216e-22	0.941611	0.9313929	0.9111672

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.95	OLS	1.450458e-20	0.9392609	0.9236868	0.8649478
	LASSO =0.01187852 λ	9.806296e-23	0.9371397	0.9251663	0.8312146
	Ridge =0.0155039 λ	6.064732e-18	0.9118058	0.8897573	1.224521
	Ridge robust	4.833914e-21	0.9386056	0.923257	0.852422
	Elastic Net =0.01972351 λ	1.495531e-21	0.9368506	0.9242208	0.8767888

جدول 2: نتائج المقارنة بين طرق التقدير عند حجم عينة 150 مشاهدة

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.50	OLS	1.977755e-87	0.9535914	0.9502526	0.927533
	LASSO =0.02673099 λ	3.032989e-90	0.9529801	0.9503311	0.9198957
	Ridge =0.0701938 λ	3.53313e-83	0.9454363	0.9415389	1.082733
	Ridge robust	9.17792e-88	0.953097	0.9497468	0.9307181
	Elastic Net =0.04044202 λ	5.257033e-89	0.9530899	0.9504093	0.9308595

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.70	OLS	2.056923e-73	0.9260387	0.9207177	0.927533
	LASSO =0.02847201 λ	6.46478e-78	0.9239314	0.9207619	0.9208358
	Ridge =0.2982773 λ	3.758559e-68	0.9103609	0.9039581	1.116115
	Ridge robust	1.203424e-73	0.9252508	0.9199115	0.9307181
	Elastic Net =0.04044202 λ	4.81753e-75	0.9240188	0.9202197	0.9460578

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.95	OLS	3.237619e-76	0.9326249	0.9277778	0.927533
	LASSO =0.00803117 λ	5.492469e-80	0.9316691	0.9283243	0.914378
	Ridge =0.039658 λ	1.419022e-69	0.9144819	0.9083735	1.168893
	Ridge robust	1.808482e-76	0.9319072	0.9270434	0.9307181
	Elastic Net =0.01107116 λ	1.956345e-76	0.9318304	0.9269611	0.931768

وصحة النتائج الإحصائية. تم اعتماد معدل الترشيح الكبيبي كمتغير تابع، لدراسة تأثير المتغيرات المستقلة عليه باستخدام طرق التقدير الخمسة. وعند دراسة تأثير المتغيرات المستقلة على معدل الترشيح الكبيبي، بينت نتائج جدول (6) معنوية النموذج المقدر ككل وذلك عند مستوى معنوية 5% لجميع طرق التقدير تحت الدراسة، وسنعرض النتائج كالتالي:

طريقة المربعات الصغرى: بينت نتائج طريقة المربعات الصغرى أن متغير العمر كان معنوي عند مستوى معنوية 5% بمعامل سالب (0.0369)، واتضح أيضا معنوية الكرياتينين (s.creat) بمعامل سالب (0.6477) وكذلك معنوية الكوليسترول (chol) بمعامل سالب (0.0076)، ومعنوية مستوى الهيموجلوبين بمعامل موجب (0.1458)، أما بقية المتغيرات اتضح عدم معنويتها أي أنها لم تصل إلى مستوى الدلالة الإحصائية عند مستوى المعنوية 5%.

انحدار لاسو: نلاحظ أن بعض المتغيرات بقيت في النموذج، وفي المقابل هناك بعض المتغيرات تم حذفها وتصغيرها. أي أنه تم اختيار المتغيرات الأكثر تأثيراً باستخدام خاصية تقليص المعاملات، وتم تحديد القيمة المثلى لمعامل التقلص باستخدام تقنية التحقق المتقاطع (Cross-Validation)، حيث اختيرت القيمة التي أعطت أقل متوسط خطأ. وهذه المتغيرات تمثلت في (الكرياتينين - الماغنيسيوم - مستوى الهيموجلوبين - حمض اليوريك - العمر - الفسفور - البروتين - الكوليسترول - فيتامين د) مرتبة حسب الأكثر أهمية والتي لها تأثير على معدل الترشيح الكبيبي لدى المرضى.

انحدار الحرف: تقوم هذه الطريقة باستخراج جميع معاملات النموذج المقدر، ونظراً لأن انحدار الحرف لا يصغر المعاملات كما هو الحال في انحدار لاسو، فقد تم تحديد أهم المتغيرات تأثيراً باستخدام معيار القيمة المطلقة للمعامل، حيث تم اختيار المتغيرات التي كانت القيمة المطلقة لمعاملها أكبر مقارنة بقيمة المتغيرات الأخرى كمؤشر على الأثر العملي القوي. ويلاحظ من خلال النتائج أن بعض المتغيرات ذات الارتباط القوي مع المتغير التابع احتفظت بأثر كبير حتى بعد تطبيق العقوبة، ما يدل على أهميتها النسبية في التنبؤ بقيم المتغير التابع. وقد تبين أن المتغيرات (الكرياتينين - الماغنيسيوم - البروتين - مستوى الهيموجلوبين - البوتاسيوم) مرتبة حسب الأكثر أهمية لها تأثير على معدل الترشيح الكبيبي لدى المرضى.

انحدار الحرف القوي: تم تقدير معاملات النموذج باستخدام انحدار الحرف القوي بالتوازي مع تقنية الحرف لتقليل أثر التعدد الخطي. وقد أظهرت النتائج المستخلصة من المخرجات أن معاملات بعض المتغيرات حافظت على تأثيرها القوي (قيم المعاملات مرتفعة نسبياً). وقد بينت النتائج أن المتغيرات المتمثلة في (الكالسيوم - الدهون الثلاثية - الكوليسترول - مستوى الهيموجلوبين - البوتاسيوم - الفسفور) مرتبة حسب الأكثر أهمية والتي لها تأثير على معدل الترشيح الكبيبي لدى المرضى.

انحدار الشبكة المرنة: اتضح من النتائج أن جميع المتغيرات بقيت في النموذج، ولكن تم موازنة التأثير بين الانكماش (Ridge) والاختيار (LASSO)، حيث تم تحديد أكثر المتغيرات تأثيراً على معدل الترشيح الكبيبي حسب حجم المعاملات، وقد تمثلت في (الكرياتينين - الماغنيسيوم - مستوى الهيموجلوبين - حمض اليوريك - البروتين - الفسفور - العمر - البوتاسيوم - الكوليسترول - فيتامين د - اليوريا) مرتبة حسب الأكثر أهمية والتي لها تأثير على معدل الترشيح الكبيبي لدى المرضى.

جدول 5: نتائج المقارنة بين طرق التقدير عند حجم عينة 1000 مشاهدة

ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.50	OLS	0.0000	0.9278037	0.9239999	1.040972
	LASSO = 0.02255213λ	0.0000	0.9262911	0.9253203	1.021864
	Ridge = 0.123213λ	0.0000	0.9158327	0.9114028	1.212301
	Ridge robust	0.0000	0.9274762	0.9236591	1.044595
	Elastic Net = 0.04109731λ	0.0000	0.9261868	0.9251767	1.063166
ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.65	OLS	0.0000	0.9264341	0.9225581	1.040972
	LASSO = 0.03043082λ	0.0000	0.923106	0.9224859	1.040901
	Ridge = 0.187981λ	0.0000	0.9104148	0.9056997	1.266314
	Ridge robust	0.0000	0.9261003	0.9222108	1.044595
	Elastic Net = 0.04194957λ	0.0000	0.9237354	0.9223706	1.078024
ρ	Methods	p-value	R^2	$R^2_{adj.}$	MSE
0.95	OLS	0.0000	0.9265099	0.9226379	1.040972
	LASSO = 0.009872328λ	0.0000	0.9237843	0.9229366	1.035915
	Ridge = 0.466364λ	0.0000	0.9012312	0.8960328	1.97567
	Ridge robust	0.0000	0.9261764	0.922291	1.044595
	Elastic Net = 0.01360923λ	0.0000	0.924323	0.9229688	1.070821

2.3. الجانب التطبيقي

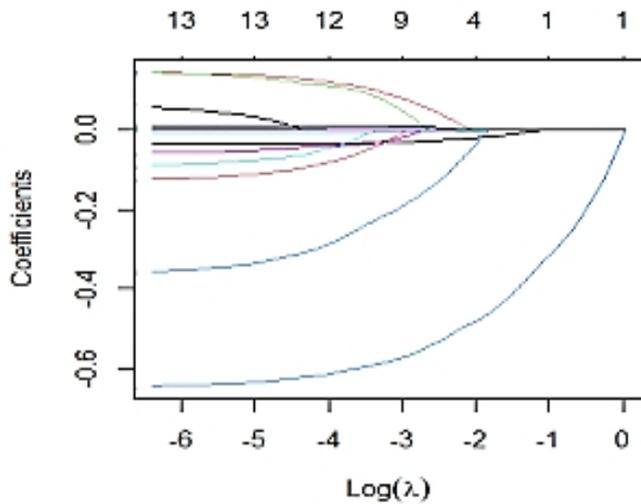
تم جمع بيانات من مرضى الفشل الكلوي بمركز خدمات الكلى في مدينة الخمس، وقد تضمنت البيانات على 100 مشاهدة خلال الفترة (يناير 2023 - أغسطس 2023). وقد تم استبعاد 34 حالة لعدم توفر البيانات عنها. ويجدر الإشارة هنا، إلى أن نقص المعلومات عن مرضى الكلى يعود إلى ضعف نظم السجلات الطبية وعدم وجود قواعد بيانات كاملة. ولذلك تم استبعاد حوالي 34% من مرضى الكلى من العينة الأصلية بسبب عدم توفر بيانات كاملة عن المتغيرات الأساسية تحت الدراسة، والتي تُعد ضرورية لإجراء تحليل موثوق للمتغير التابع الكمي والمتغيرات المستقلة. وقد تم اختيار عدم تعويض هذه القيم للحفاظ على واقعية البيانات وطبيعتها الحقيقية دون إدخال تقديرات قد تؤثر على النتائج. وعلى الرغم من هذا الاستبعاد، أظهر اختبار KMO كفاية حجم العينة لإجراء التحليل، كما أظهرت نتائج تحليل الحساسية عبر مجموعات فرعية ثبات العلاقات والمعنويات، مما يشير إلى أن الاستبعاد لم يؤثر جوهرياً على دقة الاستنتاجات، مع التنويه إلى أن النتائج تعكس فقط المرضى الذين تتوفر بياناتهم بالكامل.

تضمنت البيانات 14 متغير من المتغيرات البيوكيميائية والسريية التي قد تؤثر على بعض الوظائف الكلوية لمرضى الفشل الكلوي والتي لها علاقة كبيرة بموضوع المقالة بالاعتماد على الدراسات السابقة وعلى آراء بعض الأطباء بمركز خدمات الكلى في مدينة الخمس، وهذه المتغيرات تمثلت في (معدل الترشيح الكبيبي (eGFR) - الكرياتينين (S.creat) - مستوى الهيموجلوبين (Hb) - اليوريا (BL.urea) - البوتاسيوم (k) - الفسفور (Ph) - الكالسيوم (Ca) - البروتين (Prot) - تحليل مكونات البول (Ua) - الماغنيسيوم (Mg) - الكوليسترول (Chol) - الدهون الثلاثية (TG) - فيتامين د (Vit-D) - العمر (Age)). تم اختبار فرضيات الانحدار الخطي المتعدد للتحقق من مدى ملاءمتها للبيانات، وذلك من خلال التحقق من خطية المتغيرات، فرض التوزيع الطبيعي للبواقي، ثبات التباين، وجود قيم شاذة لضمان ملاءمة النموذج

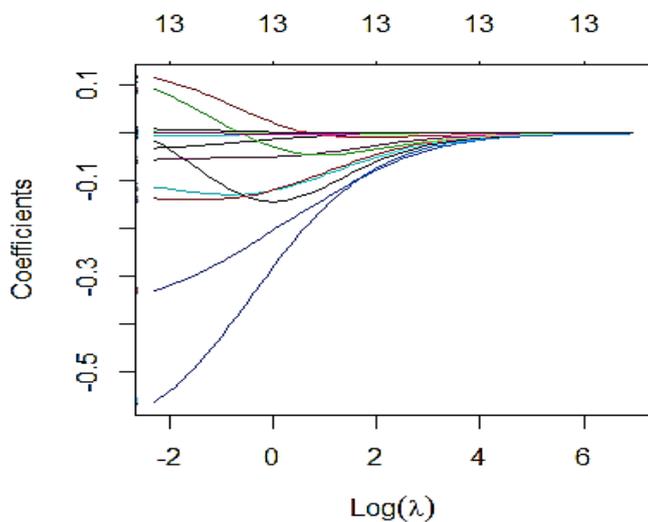
جدول 6: نتائج طرق التقدير لتأثير المتغيرات المستقلة على معدل الترشيح

variables	الكبيي				
	OLS	LASSO = λ 0.0437	Ridge = $1032 \cdot \lambda$	Ridge robust	Elastic Net = λ 0.0548
(Intercept)	13.5621	12.9595	14.0027	5.6621	13.2942
Age	** -0.0369	-0.0324	-	-	-0.0328
Hb	**0.1458	0.0870	0.1154	1.4201	0.1008
BL.urea	0.0033	-	-	-	0.0002
s.creat	** -0.6477	-0.5810	-0.5617	-	-0.5806
Ua	0.1469	0.0633	-	-	0.08753
K	-0.0930	-	-0.1133	1.2776	-0.02696
Ph	-0.0604	-0.0257	-	1.2265	-0.0396
Ca	0.0633	-	-	-6378	-
Prot	-0.1297	-0.0233	-0.1368	-	-0.0648
Mg	-0.637	-0.2075	-0.282	-	-0.2467
Chol	** -0.0076	-0.0055	-	1.4667	-0.0059
TG	0.0009	-	-	5.0059	-
vit.D	0.0103	0.0080	-	-	0.0083
F-statistic	12.1118	17.5290	11.6691	11.7808	13.7118
p-value	e-1121.648	e-1361.873	e-1192.544	2.1373e-11	6.4775e-12

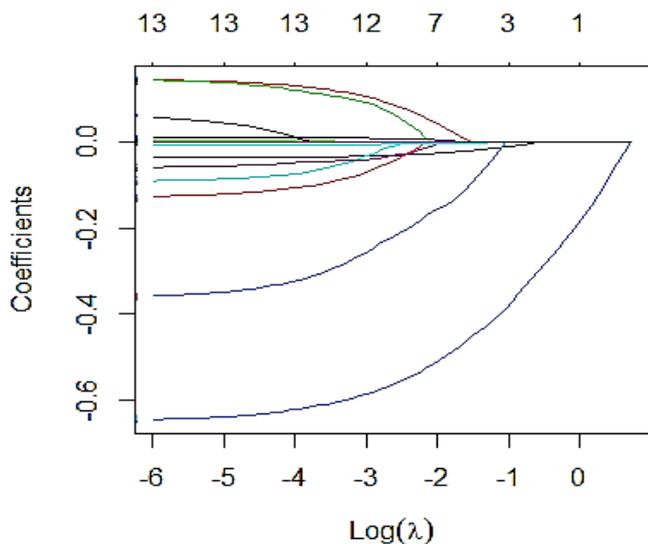
**دال احصائيا عند مستوى معنوية 1% - *دال احصائيا عند مستوى معنوية 5%



شكل 1: سلوك حل معالم النموذج لانحدار لاسو



شكل 2: سلوك حل معالم النموذج لانحدار الحرف



شكل 3: سلوك حل معالم النموذج لطريقة انحدار الشبكة المرن

أما الأشكال من (4) إلى (6)، تبيين أن المحور الأفقي يمثل معكوس معالم الانكماش والمحور الرأسي يمثل متوسط خطأ التحقق المتقاطع والنقاط الحمراء تمثل متوسط خطأ التحقق المتقاطع عند كل قيمة λ والخط

كما نلاحظ من الشكل (1) في انحدار لاسو، أنه مع زيادة قيمة λ اتجاه اليسار تبدأ المعاملات بالانكماش نحو الصفر تدريجياً، وعند قيم λ العالية جداً يتم تصفير بعض المعاملات تماماً، مما يعكس أثر الانتظام (regularization) الذي يفرضه انحدار لاسو وميزتها في اختيار المتغيرات. وبالتالي فإن عدد المعاملات الغير صفيرية يتناقص كلما زادت قيمة λ وهذا يستخدم لاختيار نموذج أبسط بأقل عدد من المتغيرات دون فقد كبير في الدقة. وفي انحدار الحرف، يوضح الشكل (2)، مسارات المعاملات المقدر في نموذج انحدار الحرف القوي عند قيم مختلفة لمعكوس معامل الانكماش، حيث أن كل خط ملون في الشكل يمثل معاملاً خاصاً بأحد المتغيرات التفسيرية في النموذج. ومن الشكل نلاحظ أن قيمة معامل الانكماش تزداد نحو اليسار، حيث تبدأ معاملات الانحدار في الانكماش تدريجياً نحو الصفر ولكنها لاتصل إلى الصفر تماماً، بل تظل المعاملات صغيرة ولكن غير معدومة. فهذه الخاصية تعد من أهم خصائص انحدار الحرف، حيث تعمل على تقليل حجم المعاملات دون استبعاد المتغيرات، مما يساعد على معالجة مشكلة التعدد الخطي بين المتغيرات المستقلة، مع الحفاظ على جميع المتغيرات داخل النموذج. وبالتالي فإن الشكل يوضح كيف تساهم هذه الطريقة في التحكم بقيم المعاملات من خلال فرض قيود انكماشية عليها، مما يحسن من استقرار التقديرات ويقلل من تباين النموذج دون التضحية بالمتغيرات. وفي طريقة انحدار الشبكة المرن، نلاحظ من الشكل (3)، أن مع زيادة قيمة λ اتجاه اليسار تبدأ المعاملات بالانكماش نحو الصفر تدريجياً، مما يعكس أثر الانتظام (regularization) الذي يفرضه لاسو حيث نلاحظ أيضاً عند قيم λ العالية جداً، يتم تصفير بعض المعاملات تماماً، وهذا يوضح ميزة لاسو في اختيار المتغيرات وبالتالي فإن عدد المعاملات الغير صفيرية يتناقص كلما زادت قيمة λ وهذا يستخدم لاختيار نموذج أبسط بأقل عدد من المتغيرات دون فقد كبير في الدقة. ونلاحظ تغير متوسط مربعات الخطأ عند تغير قيمة λ حيث يساعدنا ذلك في اختيار قيمة التنظيم الأمثل التي تحقق توازناً بين دقة النموذج والتعميم وبالتالي تحسين الأداء عند استخدام بيانات جديدة.

3.3. المقارنة بين طرق التقدير تحت الدراسة

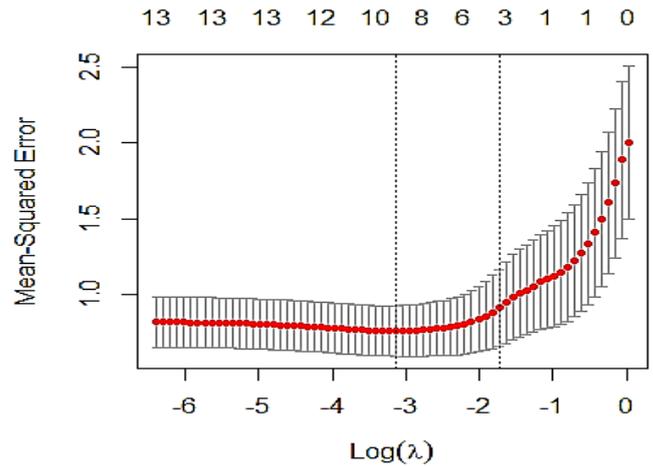
عند المقارنة بين طرق التقدير لتأثير متغيرات الدراسة على معدل الترشيح الكبيبي، بينت نتائج جدول (7)، إن طرق التقدير الخمسة قدّمت مستويات متقاربة من حيث معامل التحديد ومعامل المعدل، إلا أن هناك فروقاً طفيفة من حيث متوسط مربعات الخطأ. فقد أظهر انحدار لاسو (LASSO) أفضل أداء من حيث تقليل متوسط مربعات الخطأ مع المحافظة على نفس القدرة التفسيرية مقارنةً ببقية النماذج، مما يشير إلى أنها أكثر كفاءة في التنبؤ وقابلية التعميم على بيانات جديدة ومرونتها في التعامل مع البيانات ذات الارتباطات المتعددة، خصوصاً مع ميزة لاسو في اختيار المتغيرات المهمة. بينما كان أداء الانحدار التقليدي (OLS) وانحدار الحرف القوي (Ridge) (robust) مقبولاً، لكنه أقل تميزاً مقارنةً بانحدار لاسو. أما انحدار الحرف (Ridge) وانحدار الشبكة المرنة (Elastic-net) جاءت بأضعف النتائج، حيث انخفض معامل التحديد وارتفع متوسط مربعات الخطأ، مما يعكس محدودية فعاليتها مع هذه البيانات. وبناءً على ذلك يمكن القول إن انحدار لاسو هي الأكثر ملاءمة للبيانات المدروسة نظراً لقدرتها على التعميم وتقليل الخطأ على البيانات. وبناءً على نتائج النموذج الأمثل (نموذج لاسو)، تبين أن معدل الترشيح الكبيبي يتأثر بشكل سلبى ببعض المؤشرات الكيميائية الحيوية (الكرياتينين - الماغنيسيوم - العمر - الفسفور - البروتين - الكوليسترول). وتأثر بشكل إيجابي بمستوى (الهيموجلوبين - حمض اليوريك - فيتامين د)، مما يعكس أن ارتفاع أو انخفاض هذه المتغيرات يرتبط غالباً بضعف وظائف الكلى أو اضطراب التوازن الأيضي والكهربي. بينما لم يظهر (مستوى اليوريا - البوتاسيوم - الكالسيوم - الدهون الثلاثية) تأثير معنوي على معدل الترشيح الكبيبي في هذا النموذج.

جدول 7: نتائج المقارنة بين طرق التقدير

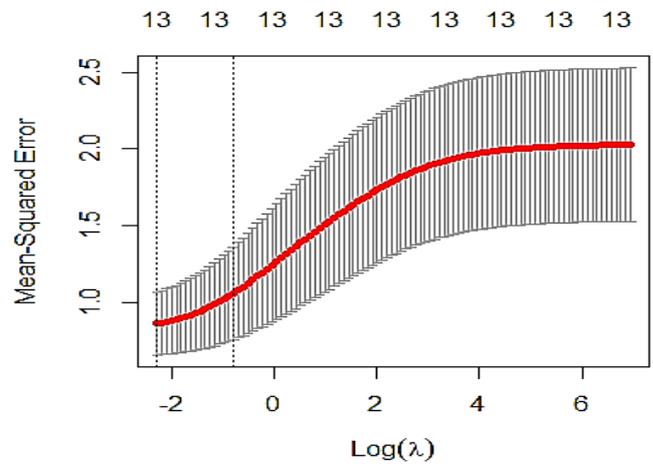
Methods	R^2	R^2_{adj}	MSE	معنوية النموذج
OLS	0.752	0.690	0.617	معنوي
LASSO	0.735	0.693	0.602	معنوي
Ridge	0.741	0.678	0.632	معنوي
Ridge robust	0.743	0.680	0.627	معنوي
Elastic Net	0.740	0.686	0.634	معنوي

وفي ضوء ما سبق عرضه من النتائج، أظهرت النتائج أن طريقة الانحدار بالانكماش (Lasso Regression) تفوقت على بقية الطرق المقارنة في جميع السيناريوهات المدروسة. فقد حققت Lasso أفضل أداء سواء في حالة عدد المتغيرات المحدود 10 متغيرات مع أحجام عينات تراوحت بين 50 و 450 مفردة، أو في حالة زيادة عدد المتغيرات إلى 50 متغيراً مع حجم عينة كبير يبلغ 1000 مفردة، وذلك عبر جميع مستويات الارتباط الضعيف، المتوسط، والقوي بين المتغيرات المستقلة. وعلاوة على ذلك، قدمت لاسو أفضل أداء في تحليل البيانات الحقيقية المدروسة. وعند المقارنة بين طرق التقدير لتأثير متغيرات الدراسة على معدل الترشيح الكبيبي فقد كان أداء انحدار الحرف القوي مقبولاً، لكنه أقل تميزاً مقارنةً بانحدار لاسو. أما انحدار الحرف وانحدار الشبكة المرنة فقد جاءت بأضعف النتائج. وتبين من خلال تحليل البيانات الحقيقية، أن معدل الترشيح الكبيبي يتأثر سلبياً بمستوى (الكرياتينين - الماغنيسيوم - العمر - الفسفور - البروتين - الكوليسترول)، ويتأثر إيجابياً بمستوى (الهيموجلوبين - حمض اليوريك - فيتامين د). ويُعزى هذا التفوق إلى قدرة Lasso على اختيار المتغيرات الأكثر تأثيراً تلقائياً وتصغير معاملات المتغيرات غير المهمة، مما يقلل من تعقيد النموذج ويحد من أثر التعدد الخطي حتى في وجود ارتباطات قوية. كما ساهم كبر حجم العينة

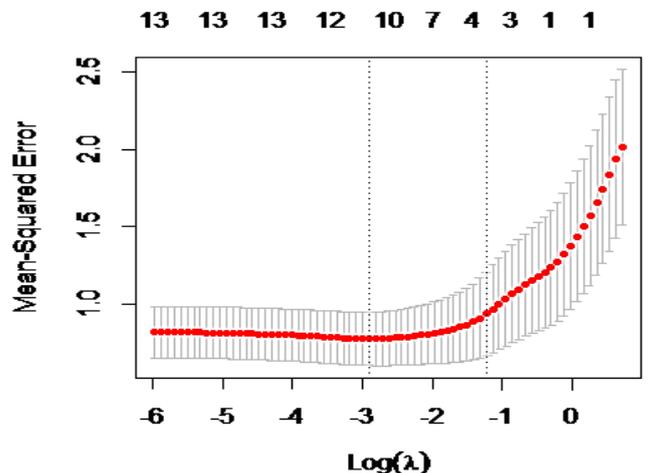
العمودي يحدد القيمة التي تعطي أقل خطأ. وبالتالي نلاحظ الهدف من هذه الأشكال الموازنة بين تعقيد النموذج ودقته، فعند اختيار أقل قيمة للمعلمة λ سوف نتحصل على أعلى دقة ممكنة تساعدنا في اختيار النموذج الأمثل، وتم تحديد القيمة المثلى لمعامل التقلص باستخدام تقنية التحقق المتقاطع (Cross-validation)، حيث اختيرت القيمة التي تعطي أقل متوسط خطأ. حيث وضع شكل (4)، أن هناك استقرارية في الحل بعد قيم $\log(\lambda)$ بين القيم (2,-6). والشكل (5)، وضع أن هناك استقرارية في الحل بعد قيم $\log(\lambda)$ بين القيم (4,6)، أما الشكل (6)، وضع هناك استقرارية في الحل بعد قيم $\log(\lambda)$ بين القيم (2,-6).



شكل 4: سلوك حل معلمة الانكماش لانحدار لاسو



شكل 5: سلوك حل معلمة الانكماش لانحدار الحرف



شكل 6: سلوك حل معلمة الانكماش لانحدار الشبكة المرنة

- والتحويل.
- [3] حيدر، علي ناظم محمد. 2023. دراسة اختيار أفضل نموذج انحدار للجلطة الدماغية باستخدام بعض الطرائق الجزائية. رسالة ماجستير، جامعة القادسية، كلية الإدارة والاقتصاد، قسم الإحصاء.
- [4] الحلواني، ماجي خليل. 2022. استخدام أسلوب انحدار ريدج لتقدير حجم الهجرة الداخلية في مصر. المجلة المصرية للتنمية والتخطيط.
- [5] الغندور خالد محمد، والدواخلي، وائل سعد. 2021. استخدام بعض طرق المربعات الصغرى الجزائية لتقدير واختيار متغيرات نموذج الانحدار الخطي في ظل وجود التعدد الخطي"، المجلة العلمية للاقتصاد والتجارة، جامعة عين شمس.
- [6] الكفيشي، سارة. 2020. تقدير معاملات أنموذج الانحدار الخطي المتعدد في ظل وجود مشكلة التعدد الخطي. مجلة الإدارة والاقتصاد جامعة كربلاء.
- [7] and S. Siswanto. 2024. "Comparative ,A. K. Jaya ,A. R. Nur and Elastic Net Regularization ,LASSO ,Analysis of Ridge Approaches in Handling Multicollinearity for Infant Mortality Data ,Statistika dan Komputasi ,in South Sulawesi," *Jurnal Matematika* pp. 311–319. doi: 10.20956/j.v20i2.31632. ,no. 2 ,vol. 20
- [8] and A. Sutrisno. 2023. "The ,A. Wijayanti ,N. Herawati and Elastic-Net in ,LASSO ,Performance of Ridge Regression Controlling Multicollinearity: A Simulation and Application," Issue 2 . ,Vol. 23 ,*Journal of Modern Applied Statistical Methods*
- [9] Rita. Obikimari. 2023. A comparative study of ,Efiezomor Methods of Remediating Multicollinearity. *American Journal of Theoretical and Applied Statistics*. 124:87–91.
- [10] V. 2017. Feature Selection using LASSO. *Research ,Fonti* VU Amsterdam. ,paper in business Analytics
- [11] M. 2015. Ridge Regression for Solving ,& Shariff .H. ,Duzan the Multicollinearity Problem Review of Methods & Models. *Journal of Applied science*.
- [12] T. 2012. A ,& Verdonck ,P. J. ,Rousseeuw ,M. ,Hubert deterministic algorithm for robust regression and outlier detection. 618-637. ,213 ,*Journal of Computational and Graphical Statistics*
- [13] R. W. 2000. Ridge Regression: Biased ,& Kennard ,E. ,Hoerl Estimation for Non-Orthogonal Problem. *Technometric*. 421: 80-86.
- [14] J. L. 1990. *Robust Regression: ,K. D. & Arthur ,Lawrnce* New York. ,*Analysis & Application*. Marcel Deker
- [15] T. 2005. Regularization and Variable ,& Hastie ,H. ,Zou Selection Via the Elastic Net. *Journal of the Royal Statistical Society: Series B* 672: 301–320.

واستقرار البيانات في تعزيز دقة النموذج وكفاءته التنبؤية مقارنة بالطرق الأخرى مثل Ridge و Elastic Net. وبناءً على ذلك، توصي الدراسة باستخدام Lasso Regression كخيار إحصائي مفضل في الدراسات التي تهدف إلى تقدير تأثير المتغيرات والتنبؤ بالنتائج، خصوصاً عندما تكون البيانات خالية من القيم الشاذة ويتناسب حجم العينة مع عدد المتغيرات، بغض النظر عن درجة الارتباط بينها.

4. قيود وحدود الدراسة

تقتصر هذه المقالة على بيانات مأخوذة من مركز خدمات الكلي بمدينة الخمس خلال فترة زمنية قصيرة (ثمانية) أشهر خلال عام (2023). وذلك لأسباب عملية ومنهجية، من بينها قرب المركز من الباحثين وسهولة الوصول إلى المرضى والسجلات الطبية. كما يُعد هذا المركز الوحيد المتخصص في تقديم خدمات غسيل الكلي داخل المدينة، مما جعله المصدر الأنسب والوحيد المتاح للحصول على بيانات دقيقة تخدم أهداف الدراسة. وعلى الرغم من دقة التحليل للمرضى المشمولين في العينة، فإن الاعتماد على مركز واحد وفترة زمنية محدودة قد يحد من إمكانية تعميم النتائج على نطاق أوسع، إذ تعكس النتائج خصائص المرضى المعالجين في هذا المركز فقط خلال الفترة المذكورة، وقد تختلف الظروف والإمكانات بين المراكز أو عبر فترات زمنية مختلفة. لذلك، يُوصى بأن تشمل الدراسات المستقبلية بيانات من مراكز متعددة وفترات زمنية أطول، بما يعزز القوة التعميمية ويتيح فهماً أكثر شمولاً للظاهرة قيد الدراسة.

كما ركزت على مجموعات بيانات تحتوي على عدد متغيرات مستقل متوسط حوالي 10 متغيرات في بعض السيناريوهات، وعدد كبير من المتغيرات حوالي 50 متغيراً في سيناريوهات أخرى، بهدف دراسة تأثير حجم المتغيرات على نتائج الانحدار. وقد تم استخدام خمس طرق انحدار مختلفة، حيث أظهرت النتائج تفوق طريقة الانحدار بالانكماش (Lasso Regression) في جميع الحالات، بغض النظر عن قوة الارتباط بين المتغيرات. وعلى الرغم من هذه المحددات، فإن حجم العينات الكبير نسبياً بين 50 و 1000 مفردة، وخلو البيانات من القيم الشاذة عززا من دقة النتائج وموثوقيتها، مما يجعل الدراسة إسهاماً أولياً مهماً يمكن البناء عليه في أبحاث مستقبلية أكثر شمولاً. ولتعزيز التعميم والمصدقية في الأبحاث اللاحقة، يُوصى بأن تشمل الدراسات المستقبلية بيانات من مراكز متعددة وفترات زمنية أطول، مع توسيع المقارنة لتشمل طرائق انحدار وتحليل تنبؤي مختلفة لتعزيز القوة التعميمية والاستنتاجية للنتائج.

5. قائمة المراجع

- [1] الديب، نادية عبد الله محمد. 2023. استخدام طريقة انحدار الحافة لمعالجة مشكلة التعدد الخطي. رسالة ماجستير، الأكاديمية الليبية طرابلس-ليبيا -قسم العلوم الرياضية -شعبة الإحصاء.
- [2] بسبوني، عبد الرحيم عوض. 2023. دراسة مقارنة لطرق علاج مشكلة الازدواج الخطي بالتطبيق على الهجرة الداخلية في مصر. مجلة التجارة