مجلة الادارة والاقتصاد / السنة 42 – العدد 118 /2019
ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

اختيار المتغيرات في نموذج انحدار بواسون باستخدام طرائق الإمكان الجزائية

أ.م.د. زكريا يحيى نوري* احمد مطلك عبداللطيف**
zakariya.algamal@uomosul.edu.iq Ahmedmotlak10@yahoo.co

المستخلص

يعد أنموذج انحدار بواسون أحد أهم نماذج الانحداراللوغاريتمية الخطية ، وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. وكغيره من سائر نماذج الانحدار، قد يحتوي النموذج على متغيرات مستقلة كثيرة ما يؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج. تهدف هذه الدراسة إلى استعراض ومقارنة طرائق إختيار المتغيرات في نموذج انحدار بواسون عبر طرائق الامكان الجزائية باستخدام المحاكاة والبيانات الحقيقة. تم استخدام أسلوب مونت – كارلو في المحاكاة لتوليد بيانات تتبع نموذج انحدار بواسون تبعا لعوامل مختلفة كحجم العينة وقيمة معامل الارتباط البسيط وعدد المتغيرات المستقلة. تم الاعتماد على جانبين من جوانب تقييم اداء الطرائق الجزائية: الاول هو تقييم دقة التنبؤ والثاني هو تقييم اختيار المتغيرات كمعيار للمقارنة، فقد أظهرت نتائج المحاكاة تفوق طريقة (SCAD) مقارنةً بطرائق التقدير الأخرى. اضافة الى ذلك، طبقت طرائق الإمكان الجزائية على بيانات حقيقية تم جمعها من مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وقد شخصت حالة المرضى من قبل اطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي – وحدة الكلية الاصطناعية.

Variable selection in Poisson regression model using penalized likelihood methods

Ahmed Motlak Abdalteef **Dr. Zakariya Yahya Algama**

Abstract

The Poisson regression model is one of the most important models of linear logarithmic regression it is the tool by which the dependent variable is modeled when the values of that variable are in the form of count values. As with other regression models, the model may contain many explanatory variables, which negatively affect the accuracy of the model and its simplicity in interpreting the results. The aim of this study is to review and compare the methods of selecting variables in the Poisson regression model through the methods of penalization using simulations and real data. The Monte-Carlo method has been used in simulations to generate data following the Poisson regression model according to factors such as sample size, simple correlation coefficient value and number of independent variables. Two aspects of the evaluation of the performance of penal methods has been based on: the first is the evaluation of the accuracy of the prediction and the second is the assessment of the choice of variables as a benchmark for comparison. In addition, the methods of penal potential were applied to real data collected from patients with chronic renal insufficiency and who are treated with continuous dialysis. Patients were diagnosed by specialized doctors in collaboration with Ibn Sina Medical Center - Synthetic College Unit.

* جامعة الموصل / كلية العلوم الحاسوب والرياضيات .
** باحث .
مقبول للنشر بتأريخ 2018/2/21
مستل من رسالة ماجستير

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

1- المقدمة

تحليل الانحدار هو اداة احصائية تقوم ببناء أُنموذج احصائي وذلك لتقدير العلاقة بين متغير واحد يدعى المتغير التابع ومتغير اخر او عدة متغيرات اخرى تدعى المتغيرات التوضيحية (التفسيرية)، بحيث ينتج معادلة احصائية توضح العلاقة بين المتغيرات . لقد إحتل تحليل الانحدار بنماذجه المختلفة مكانة متميزة في توجهات العديد من علماء الإحصاء، ونالت نصيبها الوافر عبر المؤلفات الاحصائية المختلفة ، واصبح دورها مهم جداً في تطبيقات علوم الحياة المتنوعة ولاسيما في المجال الاقتصادي الذي أخذ على عاتقه اعتماد نماذج الانحدار بالدرجة الاساس لتكون ابرز وسائل الدعم العملي للنظريات الاقتصادية، فضلاً عن العلوم الاخرى كالصحية والحياتية والاجتماعية وغيرها (الراوي،1978) .

يفترض أنموذج الانحدار الخطي الكلاسيكي ان متغير الاستجابة يعتمد على مجموعة من المتغيرات التوضيحية، حيث يمكن أن تكون هذه المتغيرات عبارة عن متغيرات مستمرة أو متغيرات قابلة للعد، ومع ذلك، عندما يكون متغير الاستجابة بشكل متغيرات قابلة للعد مثل عدد المرضى، فأنه سوف لن تتحقق افتراضات الانحدار الخطي. لذلك تم اقتراح أنموذج انحدار بواسون كأحد نماذج الانحدار التي تتوافق مع هكذا حالات.

اختيار المتغيرات في بيانات العد باستعمال أنموذج انحدار بواسون الجزائي (Penalized Poisson Regression Model (PPR)) هي واحدة من التحديات في تطبيق أنموذج انحدار بواسون عندما يكون عدد المتغيرات التوضيحية كبير، حيث اصبحت الاساليب التقليدية لاختيار المجموعات الجزئية مثل طريقة الاختيار الامامية (Forward selection) و طريقة الاختيار الى الخلف (Backward elimination) و طريقة الاختيار التدريجية (Stepwise selection) غير جيدة في اداء وظيفتها حيث اصبحت اكثر تكلفة في حسابها ، فضلاً عن ذلك فان معايير المعلومات لاختيار المتغيرات مثل معيار أكاكي للمعلومات (Akaike information criterion (AIC)) ومعيار بيز للمعلومات (Bayesian information criterion (BIC)) اصبحت غير عملية في اختيار المتغيرات التوضيحية وذلك بسبب تعقيدها الحسابي الذي ينمو بشكل طردي مع ازدياد عدد المتغيرات التوضيحية (Algamal,2016).

في السنوات الأخيرة، تم اقتراح اساليب جزائية اكتسبت شعبية كبيرة بين الإحصائيين كمفتاح لأداء اختيار المتغيرات وتقدير الأنموذج في وقت واحد، وفقاً لذلك، اقترحت عائلة من طرائق الجزاء وذلك من خلال اضافة قيد جزائي إلى دالة الإمكان الاعظم، وان الهدف وراء اضافة القيد الجزائي هو للسيطرة على تعقيد الأنموذج وتقديم معيار لاختيار متغير عن طريق إدخال بعض القيود على المعلمات، وهذه القيود تجبر بعض المعلمات لتكون قيمها مساوية للصفر مما يؤدي الى تحسين دقة تنبؤ الانموذج وتقديم أنموذج قابل للتفسير بسهولة (James et al., 2013).

لقد تناولت الدراسة الحالية أنموذج انحدار بواسون (Poisson Regression Model(PRM)) الذي يعد أحد النماذج الأكثر شعبية بين النماذج التي لديها متغير استجابة قابل للعد، وقد تم وصفه أولا عن طريق الباحثَين Nelder and Wedderburn (1972)، كحالة خاصة من النماذج الخطية المعممة (Generalized linear models GLMs)، وللوقوف على مدى اهمية المنهجية مقارنة بالطرائق التقليدية الاخرى سيتم اخضاع الأنموذج المستعمل ومن ثم توظيف معايير تقييم المعنوية لنتائج كل طريقة.

يهدف هذا البحث إلى استعراض ومقارنة طرائق إختيار المتغيرات التوضيحية في أنموذج انحدار بواسون عبر طرائق الامكان الجزائية باستخدام المحاكاة والبيانات الحقيقة، من خلال تسليط الضوء على عدد من العوامل التي قد تؤثر على جودة هذه الطرائق ووجوب استعمالها ضمن شروط معينة دون غيرها من الطرائق.

2- نموذج انحدار بواسون

يعد نموذج انحدار بواسون أحد أهم نماذج الانحدار اللوغاريتمية الخطية، وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. وكغيره من سائر نماذج الانحدار، قد يحتوي النموذج على متغيرات مستقلة كثيرة ما يؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج. يفترض هذا الأنموذج أن المتغير المعتمد y_i هو متغير استجابة يتبع توزيع بواسون وبمعلمة قدرها (μ) ، كما تتبع الأخطاء العشوائية في الأنموذج توزيع بواسون بمعلمة قدرها (μ) (Hossain And Ahmed,(2012)) ويعرف وفق الدالة الاحتمالية المعرفة بالصيغة الاتية.

$$y_i = e^{XB+U} \quad \dots (1)$$

ويمكن التعبير عنه ايظاً بصيغة المصفوفات وكالتالي:

$$\mathrm{y} = Exp(X\beta + \mathrm{U}) \quad \dots (2)$$

إذ أن:

y: موجه المتغير التابع ذي درجة $(n\times 1)$

$\mathbf{X}$: مصفوفة المتغيرات المستقلة (التوضيحية) ذات الدرجة $((n\times(p+1))$

β: موجه المعلمات ذو الدرجة $((p+1)\times 1)$

U: موجه الأخطاء العشوائية ذي الدرجة $(n\times 1)$

n : حجم العينة

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

P : عدد المتغيرات المستقلة (التوضيحية).

لأجل تقدير معلمات أنموذج انحدار بواسون باستخدام طرائق الإمكان الجزائية سيتم اللجوء الى تعظيم المشاهدات لتوزيع المتغير المعتمد (y_i)، إذا كان المتغير المعتمد (y_i) يتبع توزيع بواسون بمعلمة قدرها (μ_i) فتكون دالة التوزيع كما في الصيغة (1) والمعرفة سلفا بالشكل التالي :

$$f(y_i/\mu_i) = \frac{e^{-\mu i}\,\mu_i^{y_i}}{y_i!}$$

ومن خلال تعظيم المشاهدات لتوزيع المتغير المعتمد (y_i) الوارد في الصيغة أعلاه تكون دالة الإمكان الأعظم بالشكل الآتي :

$$L(y_1, y_2, \dots, y_n; \mu_i) = \frac{Exp\{-\sum_{i=1}^{n}\mu_i\}\,\mu_i^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i\,!} \qquad \dots (3)$$

وبأخذ اللوغاريتم الطبيعي لدالة الإمكان الأعظم للمشاهدات أعلاه نحصل على:

$$LogL(\mathrm{y_i}|x_i, \beta) = -\sum_{i=1}^{n}\mu_i + \sum_{i=1}^{n} y_i\,(\mathrm{Log}\{\mu_i\}) - \mathrm{Log}\left\{\prod_{i=1}^{n} y_i\,!\right\} \quad \dots (4)$$

وبالاعتماد على الافتراض الثاني من الفروض الأساسية لأنموذج انحدار بواسون $\mu_i = \mathrm{Exp}\{x_i^T\beta\}$ ، يتم تعويض هذا الافتراض بالدالة (4) أعلاه وكما يلي:

$$LogL(\mathrm{y_i}|x_i, \beta) = -\sum_{i=1}^{n}\left(\mathrm{Exp}\{x_i^T\beta\}\right) + \sum_{i=1}^{n} y_i\left(\mathrm{Log}\left\{\mathrm{Exp}\{x_i^T\beta\}\right\}\right) - \mathrm{Log}\left\{\prod_{i=1}^{n} y_i\,!\right\}$$

$$= \sum_{i=1}^{n}(y_i x_i^T\beta - \exp(x_i^T\beta) - \log y_i!) \qquad \dots (5)$$

3- انموذج انحدار بواسون الجزائي Penalized Poisson Regression Model

كما ذكرنا سابقا بأهمية نماذج انحدار بواسون كأنموذج لوصف البيانات القابلة للعد التي تحمل قيم عددية لعدد من الأحداث التي وقعت في فترة معينة. ان اختيار المتغيرات التوضيحية في بيانات العد باستخدام انحدار بواسون الجزائي هي واحدة من الاساليب الجديدة والتي تعمل على تحديد مجموعة جزئية من المتغيرات الهامة من بين عدد كبير من المتغيرات التوضيحية. وان الفائدة من اختيار مجموعة جزئية من المتغيرات التوضيحية المهمة، هي للحد من تلك المتغيرات التوضيحية التي لا تحتوي على المعلومات المهمة(المؤثرة)، وبالتالي زيادة كفاءة انموذج انحدار بواسون في التنبؤ وتسهيل عملية تفسير الانموذج.

وذلك من خلال ربط حد الجزاء $\lambda P(\beta)$ مع دالة الامكان الاعظم اللوغاريتمية لانموذج انحدار بواسون، يمكن كتابة انموذج انحدار بواسون الجزائي Penalized Poisson regression model (PPR) بالشكل التالي:

$$PPR = \ell(\beta) + \lambda P(\beta) \qquad \dots (6)$$

حيث يتم تعريف λ كمعلمة ضبط (λ ≥ 0). فهي تسيطر على قوة تقليص المتغيرات التوضيحية، عندما تكون قيمة λ كبيرة، فانها ستعطى وزنا أكبر للحد الجزائي. حيث ان قيمة λ تعتمد على البيانات، و يمكن حسابها باستخدام طريقة العبور الشرعي (cross-validation) . من الجدير بالامر ان قبل الشروع بحل PPR، يفضل اجراء تحويل قياسي للمتغيرات التوضيحية لازالة تاثير وحدات القياس الخاصة بهذه المتغيرات.

ويمكن كتابة طرائق الجزاء، التي ذكرت سابقا، بدلالة دالة الامكان الاعظم اللوغاريتمية لانموذج انحدار بواسون والمعرفة بالمعادلة (5) نحصل على مجموعة من نماذج انحدار بواسون الجزائي والتي تضم طريقة LASSO ، SCAD ، ALASSO ، Elastic Net والمعرفة بالشكل الرياضي التالي وعلى الترتيب:

1- طريقة (LASSO)

تعد دالة الجزاء (LASSO) والتي تمثل مختصر (Least Absolute Shrinkage and Selection Operator) من اشهر واكثر الطرائق استخداما والتي تم اقتراحها من قبل Tibshirani (1996) ، ويتم الحصول على انحدار بواسون الجزائي باستعمال طريقة (LASSO) على النحو التالي:

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

$$\mathrm{PPR}(\beta,\lambda)^{\mathrm{LASSO}} = \left\{\sum_{i=1}^{n}(y_i x_i^T\beta - \exp(x_i^T\beta) - \log y_i!) + [\lambda\sum_{j=1}^{p}|\beta_j|\right\} \quad \dots(7)$$

2- طريقة (*SCAD*)

تعد احد الطرائق الجزائية المستخدمة في تقدير معلمات نماذج الانحدار الخطية وقد تم اقتراحه من قبل Fan And Li (2001)، يمكن كتابة معادلة انحدار بواسون الجزائي باستخدام طريقة (SCAD) وكالاتي :

$$\mathrm{PPR}(\beta,\lambda)^{\mathrm{SCAD}} = \left\{\sum_{i=1}^{n}(y_i x_i^T\beta - \exp(x_i^T\beta) - \log y_i!) + \lambda\sum_{i=1}^{p}p_\lambda(|\beta_i|)\right\} \quad \dots(8)$$

حيث ان :

$$\hat{\beta}_j^{\mathrm{SCAD}} = \begin{cases}(|\hat{\beta}_j| - \lambda) + \mathrm{sign}(\hat{\beta}_j) & if\ |\hat{\beta}_j| < 2\lambda;\\ \{(a-1)\hat{\beta}_j - \mathrm{sign}(\hat{\beta}_j)a\lambda\}/(a-2) & if\ 2\lambda < |\hat{\beta}_j| < a\lambda;\\ \hat{\beta}_j & if\ a\lambda < |\hat{\beta}_j|;\end{cases} \quad \dots(9)$$

3- **طريقة (*ALASSO*)**

اقترحت طريقة Adaptive Lasso (ALASSO) من قبل Zou (2006)، ويمكن تعريف معادلة انحدار بواسون الجزائي باستخدام طريقة (ALASSO) كالاتي:

$$\mathrm{PPR}(\beta,\lambda)^{\mathrm{ALASSO}} = \left\{\sum_{i=1}^{n}(y_i x_i^T\beta - \exp(x_i^T\beta) - \log y_i!) + \lambda\sum_{j=1}^{p}w_j|\beta_j|\right\} \dots(10)$$

4- طريقة (*Elastic Net*)

تعتبر Elastic Net احد طرائق الامكان الجزائية المهمة، الذي تم اقتراحها من قبل Zou and Hastie (2005)، وتعرف معادلة انحدار بواسون الجزائي باستخدام طريقة (Elastic Net) كالاتي :

$$\mathrm{PPR}(\beta;\lambda_1,\lambda_2)^{\mathrm{Elastic}} = \left\{\sum_{i=1}^{n}(y_i x_i^T\beta - \exp(x_i^T\beta) - \log y_i!) + \lambda_1\sum_{j=1}^{p}|\beta_j| + \lambda_2\sum_{j=1}^{p}\beta_j^2\right\} \quad \dots(11)$$

4- وصف تجربة المحاكاة Description Simulation Experiment

لقد تم تصميم تجربة ومحاكاتها باستعمال لغة البرمجة (R) حيث تم توليد المتغير (y_i) في انموذج انحدار بواسون الذي يتبع توزيع بواسون بمعدل مقداره (μ_i) ، حيث تم استخدام اسلوب مونت كارلو (Mont Carlo) في المحاكاة وكالاتي:

حيث تم تعيين قيم عدد المتغيرات التوضيحية (p) وهي (10,15) ، وحجم العينات (n) حيث تم استعمال احجام من العينات وهي (50,100) وذلك لأجل دراسة المقارنة وفق العينات باختلاف أنواعها ، ومعاملات الارتباط (r) التي تبين قوة العلاقة بين المتغيرات التوضيحية ثلاث قيم ارتباط وهي (0.9 , 0.7 , 0.5) ، وكما مبين بالجدول (1).

جدول (1)

يوضح اعداد وقيم العوامل p , n , r

العامل	القيم
P	10 15
N	50 100
R	0.7 0.9 0.5

5- دراسات المحاكاة Simulation Studies

اولاً : تم توليد بيانات المتغير y التي تتبع انموذج انحدار بواسون وكالاتي :

$$y \sim P(\exp(X\beta))$$

ثانياً : تم توليد مصفوفة المتغيرات التوضيحية X ذات ابعاد $(n \times p)$ التي تتبع التوزيع الطبيعي المتعدد (Multivariate Normal Distribution) كالاتي :

$$X \sim \mathrm{MN}(\mathrm{n},\mathrm{M})$$

حيث ان M هي مصفوفة التباين المشترك، حيث ان $M_{ij} = r^{|i-j|}$، عندما $(i,j = 1,2,\dots,p)$ حيث ان المتغيرات التوضيحية تكون مرتبطة.

ثالثاً : تم تكرار التجربة (100) مرة وذلك لغرض تقليل التحيز في تجارب مونت كارلو (Mont Carlo).

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

رابعاً : تم توليد بيانات نموذج انحدار بواسون تبعا لقيم متجه معلمات الانحدار β الذي ابعاده $(1 \times p)$ وكانت قيم متجه معلمات الانحدار β كالاتي $\beta = (1.2, -0.6, 0.8, -0.4, 0, \dots, 0)^{T}$ ، حيث ان المعلمات غير الصفرية عددها $q = 4$ ، وان المعلمات الصفرية تساوي $p - q$.

خامسا : تم اعتماد قيم معلمة الضبط لكل طريقة حسب الجدول التالي :

جدول (2)

يوضح قيم معلمة الضبط لكل طريقة

الطريقة	قيمة المعلمة
LASSO	$0 < \lambda < 100$
ALASSO	$0 < \lambda < 100$, $\omega = 1$
SCAD	$0 < \lambda < 100$, $a = 3.7$
Enet	$0 < \lambda_1 < 100$, $0 < \lambda_2 < 100$

وباستخدام طريقة الـ(CV) بـ(K=10) سوف يتم تقدير معلمة الضبط لكل طريقة.

6- تفسير نتائج المحاكاة

سيتم تحليل وتفسير نتائج تجربة المحاكاة تبعا لمعايير دقة التنبؤ ومعاير دقة اختيار المتغيرات من حيث حجم العينة ومعامل الارتباط وعدد المتغيرات التوضيحية عندما كانت قيم معلمات الانحدار :-

$\beta = (1.2, -0.6, 0.8, -0.4, 0, \dots, 0)^{T}$

من خلال ملاحظة الجدول (3) الذي يوضح قيم معايير كل من (EE , PE, C, I) للطرائق الجزائية (LASSO, ALASSO, SCAD, Enet)، يمكن استخلاص ما يأتي :

1. عندما يتغير معامل الارتباط بين المتغيرات من (0.5) الى (0.7)، يتبين ان طريقة (SCAD) اعطت اقل قيم (EE , PE) حيث بلغ مقدار التحسن بالتنبؤ بالاعتماد على المعيار (PE) بمقدار 69.36% و 2.3% و 10.90% عند (r=0.5) و 64.63% و 1.83% و 11.20% عند (r=0.7) مقارنة بـ(LASSO و ALASSO وEnet) على الترتيب، كما وبلغ التحسن بخطأ التقدير بالاعتماد على المعيار (EE) بمقدار 99.01%و 45.22% و46.87% عند (r=0.5) و 97.66% و 39.34% و 46.80% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على الترتيب.
2. عندما يكون معامل الارتباط مساوي الى (0.9) اعطت طريقة (Enet)افضل النتائج مقارنة بالطرائق الاخرى حيث تحسن التنبؤ بالاعتماد على المعيار (PE) بمقدار 52.59% و 7.28% و 3.70% مقارنة بـ(LASSO وALASSO وSCAD) على الترتيب ،كما وبلغ التحسن في خطأ التقدير بالاعتماد على المعيار (EE) بمقدار 94.90% و 78.77% و62.00 % مقارنة بـ(LASSO وALASSO وSCAD) على الترتيب.
3. بالاعتماد على معايير اختيار المتغيرات، فقد امتلكت طريقة (SCAD) اعلى قيم (C) الذي هو عدد المعاملات الحقيقية ذات القيم الصفرية والتي تم تقديرها بشكل صحيح على انها ذات قيم صفرية، واعطت اقل قيم (I) الذي يعرف انه عدد المعاملات الحقيقية ذات القيم غير الصفرية والذي تم تقديرها بشكل غير صحيح على انها ذات قيم صفرية عند قيم معامل الارتباط (0.5) و (0.7). في حين اظهرت طريقة (Enet) تباينا في معايير اختيار المتغيرات عند قيمة معامل الارتباط (0.9).
4. ظهرت طريقة (LASSO) كاسوأ طريقة في التقدير لأنها تعطي أعلى قيم لـ (PE و EE) وكذلك كاسوأ طريقة في اختيار المتغيرات كونها تميل الى اختيار متغيرات توضيحية غير مهمة.

جدول (3)

معدل معايير تقييم طرائق الجزاء عندما p=10 , n=50

r	Method	PE	EE	C	I
0.5	LASSO	32.3507	2.1488	1	0
	ALASSO	10.1541	0.0387	4	0
	SCAD	9.9122	0.0212	5	0
	Enet	11.1251	0.0399	4	0
0.7	LASSO	29.9037	2.0302	3.5	0
	ALASSO	10.7742	0.0783	4	0
	SCAD	10.5771	0.0475	5	0
	Enet	11.9122	0.0893	4	0
0.9	LASSO	24.1644	1.9384	5	1
	ALASSO	12.3546	0.4654	4.5	0
	SCAD	11.8954	0.2600	6	0
	Enet	11.4556	0.0988	4.5	0

من خلال ملاحظة الجدول (4) الذي يبين قيم معايير دقة التنبؤ ومعاير دقة اختيار المتغيرات للطرائق الجزائية (LASSO, ALASSO, SCAD, Enet)، فقد تم الحصول على ما يأتي :

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

1. عندما تغيرت قيم معامل الارتباط بين المتغيرات من (0.5) الى (0.7)، تبين ان طريقة (SCAD) تغلبت على باقي طرائق التقدير بامتلاكها اقل قيم (EE , PE) اذ بلغ مقدار التحسن بالتنبؤ بالاعتماد على المعيار (PE) بمقدار 69.44% و 4.68% و 13.39% عند (r=0.5) و 63.05% و 3.21% و 3.46% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على الترتيب، كما وبلغ التحسن بخطأ التقدير بالاعتماد على المعيار (EE) بمقدار 98.71% و 51.25% و53.16% عند (r=0.5) و 97.41% و 54.77% و 75.50% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على الترتيب.
2. عندما كان معامل الارتباط مساوي الى (0.9) اعطت طريقة (Enet)افضل النتائج مقارنة بالطرائق الاخرى اذ تحسن التنبؤ بالاعتماد على المعيار (PE) بمقدار 51.92% و 7.99% و 4.01% مقارنة بـ(LASSO و ALASSO وSCAD) على الترتيب، كما وبلغ التحسن في خطأ التقدير بالاعتماد على المعيار (EE) بمقدار 89.74% و 63.51% و 36.06% مقارنة بـ(LASSO وALASSO وSCAD) على الترتيب.
3. بالاعتماد على معايير اختيار المتغيرات، فقد امتلكت طريقة (SCAD) اعلى قيم (C)، واعطت اقل قيم (I)، عند قيم معامل الارتباط (0.5) و (0.7)، وهذا يعطيها افضلية كون جودة طرائق الجزاء من ناحية معايير تقييم دقة اختيار المتغيرات تعتمد على من يعطي اعلى قيمة لـ (C) واقل قيمة لـ (I)، في حين اظهرت طريقة (Enet) تباينا في معايير اختيار المتغيرات عند قيمة معامل الارتباط (0.9).
4. ظهرت طريقة (LASSO) كاسوأ طريقة في التقدير لأنها تعطي أعلى قيم لـ (PE و EE) وكذلك كاسوأ طريقة في اختيار المتغيرات كونها تميل الى اختيار متغيرات توضيحية غير مهمة في جميع حالات معامل الارتباط.

جدول (4)

معدل معايير تقييم طرائق الجزاء عندما n=50 , p=15

r	Method	PE	EE	C	I
0.5	LASSO	32.3806	2.1308	3.5	0
	ALASSO	10.3822	0.0562	9	0
	SCAD	9.8959	0.0274	10	0
	Enet	11.4257	0.0585	9	0
0.7	LASSO	28.7772	1.9957	8	0
	ALASSO	10.9878	0.1143	8	0
	SCAD	10.6342	0.0517	10	0
	Enet	11.0153	0.2110	8	0
0.9	LASSO	24.1561	1.9128	10	0
	ALASSO	12.6240	0.5380	9	1
	SCAD	12.1008	0.3070	10	1
	Enet	11.6151	0.1963	9	1

من خلال ملاحظة الجدول(5) الذي يبين قيم معايير دقة التنبؤ ومعاير دقة اختيار المتغيرات للطرائق الجزائية (LASSO, ALASSO, SCAD, Enet)، فقد اظهرت النتائج ما يلي:

1. عندما تغيرت قيم معامل الارتباط بين المتغيرات من (0.5) الى (0.7)، برزت طريقة (SCAD) وتفوقت على باقي طرائق التقدير الاخرى فقد اعطت اقل قيم (EE , PE) اذ بلغ مقدار التحسن بالتنبؤ بالاعتماد على المعيار (PE) بمقدار 65.26% و 4.02% و 6.08% عند (r=0.5) و 60.65% و 3.42% و 9.56% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على الترتيب، كما بلغ التحسن بخطأ التقدير بالاعتماد على المعيار (EE) بمقدار 97.88% و38.20% و45.25% عند (r=0.5) و 95.31% و 41.45% و 53.06% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على التوالي.
2. تبين ان طريقة (Enet) هي أفضل طريقة للتقدير اذ اعطت اقل قيم (EE, PE) عندما كانت قيمة معامل الارتباط عالية (r= 0.9) اذ تحسن التنبؤ بالاعتماد على المعيار (PE) بمقدار 49.63% و 11.96% و 9.93% مقارنة بـ(LASSO وALASSO وSCAD) على الترتيب، كما وبلغ التحسن في خطأ التقدير بالاعتماد على المعيار (EE) بمقدار 93.42% و 83.21% و 78.72% مقارنة بطريقة (LASSO و ALASSO و SCAD).
3. بالنظر الى قيم معايير اختيار المتغيرات تبين ان طريقة (SCAD) اعطت اعلى قيم (C) و اقل قيم (I)، عندما كانت قيم معامل الارتباط (0.5) و (0.7)، وهذا يميزها عن الطرائق الاخرى كون جودة طرائق الجزاء من ناحية معايير تقييم دقة اختيار المتغيرات تعتمد على من يعطي اعلى قيمة لـ (C) واقل قيمة لـ (I)، في حين اظهرت طريقة (Enet) تباينا في معايير اختيار المتغيرات عند قيمة معامل الارتباط (0.9).
4. ظهرت طريقة (LASSO) كاسوأ طريقة في التقدير لأنها تعطي أعلى قيم لـ (PE و EE) وكذلك كاسوأ طريقة في اختيار المتغيرات كونها تميل الى اختيار متغيرات توضيحية غير مهمة .

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

جدول (5)

معدل معايير تقييم طرائق الجزاء عندما n=100 , p=10

r	Method	PE	EE	C	I
0.5	LASSO	19.3353	2.0341	2	0
	ALASSO	6.9986	0.0699	4	0
	SCAD	6.7172	0.0432	5	0
	Enet	7.1520	0.0789	4	0
0.7	LASSO	18.6220	1.8992	3	0
	ALASSO	7.5870	0.1520	4	0
	SCAD	7.3276	0.0890	5	0
	Enet	8.1021	0.1896	4	0
0.9	LASSO	14.8017	1.7818	5	1
	ALASSO	8.5148	0.6986	5	1
	SCAD	8.3226	0.5511	6	1
	Enet	7.4962	0.1173	5	1

من خلال ملاحظة الجدول (6) الذي يبين قيم معايير دقة التنبؤ(PE , EE) ومعاير دقة اختيار المتغيرات (C, I) للطرائق الجزائية (LASSO, ALASSO, SCAD, Enet)، يمكن ان نستنتج ما يلي:

1. عندما يكون التغير في قيم معامل الارتباط بين المتغيرات من (0.5) الى (0.7)، اظهرت طريقة (SCAD) افضلية على باقي طرائق التقدير المستخدمة فقد نتجت اقل قيم (PE , EE) اذ بلغ مقدار التحسن بالتنبؤ بالاعتماد على المعيار (PE) بمقدار 65.02% و 5.92% و 15.58% عند (r=0.5) و 58.86% و 4.25% و 8.88% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على الترتيب، كما بلغ التحسن بخطأ التقدير بالاعتماد على المعيار (EE) بمقدار 97.68% و49.00% و63.53% عند (r=0.5) و 95.32% و 52.67% و 55.63% عند (r=0.7) مقارنة بـ(LASSO وALASSO وEnet) على التوالي.
2. اما عندما كانت قيمة معامل الارتباط (r = 0.9)، برزت طريقة (Enet) كأفضل طريقة للتقدير مقارنة بالطرائق الجزائية الاخرى المستخدمة اذ اعطت اقل قيم (EE, PE) اذ تحسن التنبؤ بالاعتماد على المعيار (PE) بمقدار 45.15% و 10.42% و 8.12% مقارنة بـ(LASSO وALASSO وSCAD) على الترتيب، كما وبلغ التحسن في خطأ التقدير بالاعتماد على المعيار (EE) بمقدار 82.05% و 64.95% و 31.91% مقارنة بطريقة (LASSO و ALASSO و SCAD).
3. اما من خلال المقارنة بواسطة معايير اختيار المتغيرات، فقد امتلكت طريقة (SCAD) اعلى قيم (C)، و اقل قيم (I)، عندما كانت قيم معامل الارتباط (0.5) و (0.7)، وهذا يعطيها افضلية جيدة على باقي الطرائق كون جودة طرائق الجزاء من ناحية معايير تقييم دقة اختيار المتغيرات تعتمد على من يعطي اعلى قيمة لـ (C) واقل قيمة لـ (I)، في حين اظهرت طريقة (Enet) تباينا في معايير اختيار المتغيرات عند قيمة معامل الارتباط (0.9).
4. كما كانت طريقة (LASSO) اسوأ طريقة في التقدير اذ انها اعطت أعلى قيم لـ (PE و EE) وكذلك اسوأ طريقة في اختيار المتغيرات كونها تميل الى اختيار متغيرات توضيحية غير مهمة .

جدول (6)

معدل معايير تقييم طرائق الجزاء عندما n=100 , p=15

r	Method	PE	EE	C	I
0.5	LASSO	19.5771	1.9705	8	0
	ALASSO	7.2789	0.0896	9	0
	SCAD	6.8482	0.0457	9	0
	Enet	8.1120	0.1253	9	0
0.7	LASSO	17.7149	1.8755	9	0
	ALASSO	7.6106	0.1855	9	0
	SCAD	7.2872	0.0878	10	0
	Enet	7.9971	0.1979	9	0
0.9	LASSO	14.1911	1.8582	10	1
	ALASSO	8.6894	0.9516	9	1
	SCAD	8.4719	0.4898	10	1
	Enet	7.7841	0.3335	9	1

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

7- معايير تقييم طرائق الجزاء

إن اسلوب تقييم اداء الطرائق الجزائية ومقارنة هذه الطرائق فيما بينها واختيار الطريقة الافضل هو جانب مهم من جوانب تحليل البيانات. بشكل عام هناك جانبين من جوانب تقييم اداء الطرائق الجزائية: الاول هو تقييم دقة التنبؤ والثاني هو تقييم اختيار المتغيرات.

1-7 معايير تقييم دقة التنبؤ

أولاً : خطأ التقدير (Estimation Error) (EE)

ويعرف بانه مربع الفرق بين قيمة المعلمات الحقيقية وقيمة المعلمات المقدرة ويعرف بالشكل الرياضي التالي :

$$EE = \left(\hat{\beta} - \beta\right)^T \left(\hat{\beta} - \beta\right) \quad \ldots (12)$$

حيث ان:

$\hat{\beta}$: هو متجه المعلمات المقدرة وفق الطرائق المستخدمة .

β :هو متجه المعلمات الحقيقية.

ثانياً : خطأ التنبؤ (Prediction Error) (PE)

ويعرف بانه مربع الفرق بين القيمة الحقيقية لمتغير الاستجابة والقيمة التنبؤية المرافقة له، ويعرف رياضيا بالمعادلة التالية :

$$PE = (y - \hat{y})^T (y - \hat{y}) \quad \ldots (13)$$

حيث ان $\hat{y} = \mathrm{Exp}\{X^T \beta\}$

بالاعتماد على هذين المعيارين يتم تحديد الطريقة الافضل التي تعطي اقل قيمة مقارنة بالطرائق الاخرى.

2-7 معايير تقييم دقة اختيار المتغيرات

تعمل طرائق الجزاء بصورة عامة على اختيار المتغيرات، لذلك من المهم تقييم وقياس قدرة هذه الطرائق وجودتها في كيفية اختيار المتغيرات المهمة. ولذلك، تم الاعتماد على معيارين في دراستنا لهذا الغرض وبالشكل التالي:

اولاً : معيار التقييم"C"

هو معيار التقييم الذي يرمز له بـــ(C) والذي يعرف بانه عدد المعاملات الحقيقية ذات القيم الصفرية والتي تم تقديرها بشكل صحيح على انها ذات قيم صفرية.

ثانياً : معيار التقييم "I"

معيار التقييم الذي يرمز له بـــ(I) وهو يعرف على انه عدد المعاملات الحقيقية ذات القيم غير الصفرية والذي تم تقديرها بشكل غير صحيح على انها ذات قيم صفرية.

تعتمد جودة طرائق الجزاء من ناحية معايير تقييم دقة اختيار المتغيرات على من يعطي اعلى قيمة لـــ (C) واقل قيمة لـــ (I) .

8-الجانب التطبيقي

لغرض اتمام الفائدة المرجوة من البحث ، تم التطبيق على بيانات تتبع توزيع بواسون والتي أخذت من بيانات استعملت من قبل (لقاء سعيد واخرون،2011) حول مرض الفشل الكلوي المزمن حيث تم جمع (73) نموذج دم لأشخاص مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وتم سحب نماذج الدم لمجموعة المرضى من قبل اجراء عملية الغسيل الكلوي التي تستغرق (4-3) ساعات، وقد شخصت حالة المرضى من قبل اطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي – وحدة الكلية الاصطناعية، تراوحت اعمارهم بين (80-20) سنة ، وتتضمن (38) نموذجا للذكور و (35) نموذجا للإناث، ودونت المعلومات الخاصة بالمرضى على وفق استمارة استبيان خاصة لكل مريض اعدت لهذا الغرض لسنة 2013، حيث سجلت الدراسة ثمانية متغيرات توضيحية والتي يعتقد بان لها تأثير في متغير الاستجابة الذي يمثل عدد مرات الغسيل الكلوي بالشهر. يوضح الجدول (3) وصف المتغيرات التوضيحية المستعملة في الدراسة.

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

جدول (12)
وصف المتغيرات المستقلة المستخدمة في الدراسة

رمز المتغير التوضيحي	وصف المتغير التوضيحي	وحدة القياس
X1	الجنس	(ذكر =1، انثى =2)
X2	العمر	سنوات
X3	مدة المرض	الايام
X4	الوراثة	(نعم=1، كلا=2)
X5	نسبة اليوريا	(ملي مول/لتر)
X6	نسبة البروتين الكلي	غرام/100 ميليلتر
X7	نسبة الالبومين	غرام/100 ميليلتر
X8	نسبة الكلوبيولين	غرام/100 ميليلتر

9- مقارنة الطرق الجزائية حسب معيار MSE

يتم تقدير معلمات نموذج انحدار بواسون بواسطة مقدر الامكان الاعظم (MLE) بغض النظـــــر عن تقدير(B_0) ثم يتم ايجاد قيم$(\hat{Y})$ لحساب متوسط مربعات الخطأ (MSE) للنموذج، ومن خلال ملاحظة الجدول (4) الذي يوضح نتائج متوسط مربعات الخطأ للنموذج المقدر الذي تم الحصول عليها باستخدام الطرق الجزائية نلاحظ تفوق طريقة (SCAD) على باقي طرائق التقدير المستخدمة الاخرى، حيث انها اعطت اقل قيمة لمتوسط مربعات الخطأ مما يجعلها افضل طريقة للتقدير، ثم تأتي بعده طريقة (Enet) بالمرتبة الثانية من حيث قيمة متوسط مربعات الخطأ، وكذلك كانت طريقتي (MLE , LASSO) أسوأ طريقتين كونها اعطت اعلى قيم لمتوسط مربعات الخطأ، ومن الجدير بالذكر ان النتائج التي تم الحصول عليها جاءت مطابقة الى نتائج تجارب المحاكاة في الجانب التجريبي عند معامل ارتباط 0.5. يعرف معيار متوسط مربعات الخطأ (MSE) بالشكل الاتي:

$$MSE = \frac{\sum_{r=1}^{R}(\hat{\beta}_r-\beta)^T(\hat{\beta}_r-\beta)}{R} \quad \dots (14)$$

إذ أن:
$\hat{\beta}_r$: قيمة المعلمة المقدرة وفق طرائق التقدير المختلفة.
β: قيمة المعلمة الحقيقية.
R : هي عدد مرات تكرار التجربة.

جدول (13)
نتائج الطرق المستخدمة بالاعتماد على معيار MSE في بيانات مرضى العجز الكلوي

Methods	MSE
MLE	9.358487
LASSO	7.877187
SCAD	5.8744
ALASSO	6.9741
Enet	6.8146

10-الاستنتاجات :

اهم الاستنتاجات التي تم التوصل اليها في هذا البحث هي :

1- اظهرت نتائج المحاكاة والتطبيق العملي ان طريقة SCAD هي افضل من الطرق الجزائية الاخرى وتغلبت عليها عندما يكون الارتباط بين المتغيرات (0.5) و (0.7)، اذ امتلكت طريقة SCAD اقل قيم معايير (I ,PE, EE) واعلى قيم (C) لجميع نماذج المحاكاة عندما كان معامل الارتباط بين المتغيرات (0.5) و (0.7).
2- من خلال نتائج المحاكاة تبين ان طريقة (Enet) هي الافضل اداءً بين الطرق الجزائية الاخرى عندما تكون قيمة الارتباط بين المتغيرات عالية (0.9)، حيث ان طريقة (Enet) تتفوق في ادائها في حال كون المتغيرات عالية الارتباط وانها تكون اقل موثوقية إذا كان معامل الارتباط بين المتغيرات أقل من (0.95) (El Anbari and Mkhadri , 2013).
3- اظهرت نتائج المحاكاة والتطبيق العملي ان طريقة LASSO هي أسوأ الطرق الجزائية الاخرى، اذ ان طريقة LASSO اعطت اعلى قيم للمعايير (I ,PE, EE) واقل قيم (C) لجميع النماذج عندما كان معامل الارتباط بين المتغيرات (0.5) و (0.7) و (0.9) .

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.118.19

11-المصادر

1. الراوي، خاشع محمود، (1978)،"مدخل الى تحليل الانحدار"، جامعة الموصل.
2. صبري، حسام موفق (2013)، "مقارنة طرائق تقدير معلمات انموذج انحدار بواسون في ظل وجود مشكلة التعدد الخطي مع تطبيق عملي"، أطروحة دكتوراه، غير منشورة، كلية الإدارة والاقتصاد، جامعة بغداد.
3. عبدالله، لقاء سعيد وعلوش، ذكرى علي و الجراح، إسراء عبد الحق ،(2011)، "دراسة إنزيم ميتالوإندوببتايديز وعلاقته بمرض العجز الكلوي المزمن"، مجلة علوم الرافدين، المجلد (22)، العدد (4)، ص (71-87).
4. Algamal, Z. Y. (2016)." Adaptive Penalized Likelihood Methods In High Dimensionalgeneralized Linear Models", Universiti Teknologi Malaysia.
5. Algamal, Z. Y. and Lee, M. H. (2015). "Penalized Poisson Regression Model using adaptive modified Elastic Net Penalty". Electronic Journal of Applied Statistical Analysis,Vol. 08, Issue 02, 236-245.
6. El Anbari, M. and Mkhadri, A. (2013)." The adaptive gril estimator with a diverging number of parameters. Communications in Statistics - Theory and Methods". 42(14), 2634–2660.
7. Fan, J., & Li, R. (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties". Journal of the American Statistical Association, 96(456), 1348-1360.
8. Hossain, S. and Ahmed, E. (2012). "Shrinkage and penalty estimators of a Poisson regression model". Australian & New Zealand Journal of Statistics. 54(3), 359–373.
9. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). "An introduction to statistical learning". Springer, New York.
10. Månsson, K., Kibria, B. G., Sjolander, P & Shukur, G. (2012), "Improved Liu Estimators for the Poisson Regression Model", International Journal of Statistics and Probability, Vol. 1, No. 1, pp. 1-6.
11. Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society". Series B (Methodological). 58(1), 267–288.
12. Zeng, L. and Xie, J. (2012). "Group variable selection via SCAD-L2". Statistics. 48(1),49–66.
13. Zou, H. (2006)." The adaptive lasso and its oracle properties. Journal of the American Statistical Association". 101(476), 1418–1429.
14. Zou, H. and Hastie, T. (2005). "Regularization and variable selection via the elastic net". Journal of the Royal Statistical Society. Series B (Methodological). 67(2), 301–320.

……
……..
……