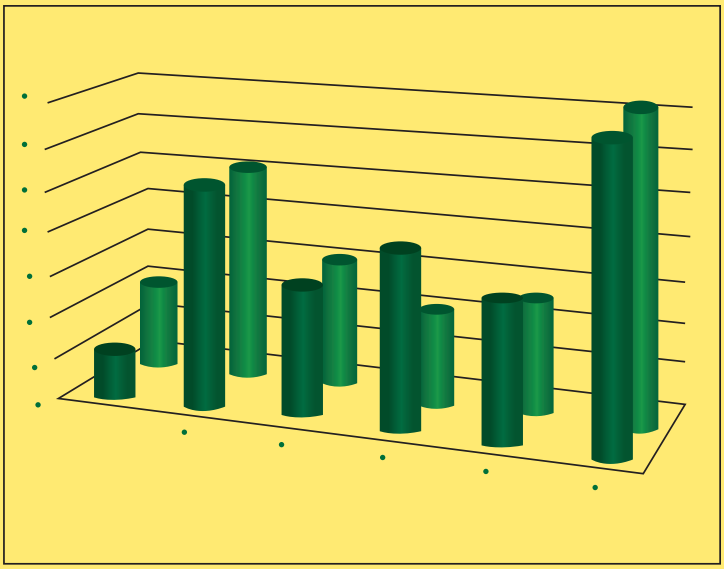
المعهد العربي للتدريب والبحوث الإحصائية

مجلة
العلوم الإحصائية



العدد رقم 22

مجلة علمية محكمة
يصدرها المعهد العربي للتدريب والبحوث الإحصائية
معتمدة في قائمة المجلات العلمية ulrich´s
www.ulrichsweb.com
مصنفة في معامل التأثير والاستشهادات المرجعية العربي (أرسيف)
www.emarefa.net/arcif/
ISSN 2522-64X (Online), ISSN 2519-948X (Print)

مجلة العلوم الإحصائية
March. 2024
No. 22

مجلة العلوم الإحصائية

مجلة علمية محكمة

معتمدة في قائمة المجلات العلمية **Ulrich’s**

www.ulrichsweb.com

مصنفة في معامل التأثير والاستشهادات المرجعية العربي (أرسيف)

www.emarefa.net/arcif/

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

ص.ب: 851104 عمان 11185 الأردن   P.O. Box: 851104 Amman 11185 Jordan

Tel: +96265549805   Fax: +96265549804 www.aitrs.org info@aitrs.org

شروط النشر في مجلة العلوم الإحصائية

1 – تنشر المجلة البحوث والدراسات العلمية في المجالات الإحصائية والمعلوماتية المكتوبة باللغة العربية والانكليزية والفرنسية على أن لا يكون البحث المقدم للنشر قد نشر او قدم للنشر في مجلات او دوريات أخرى او قدم ونشر في دوريات لمؤتمرات أو ندوات.

2 – ترسل البحوث والدراسات الى أمين التحرير على أن تتضمن اسم الباحث او الباحثين وألقابهم العلمية وأماكن عملهم مع ذكر عنوان المراسلة وأرقام الهواتف والبريد الالكتروني. وان يرسل البحث المراد نشره الكترونيا (على قرص او بالبريد الالكتروني) وفق المواصفات أدناه:

أ – أن يكون مطبوعاً على ورق حجم A4 وان يكون على شكل عمود واحد ويستخدم للغة العربية نوع حرف (Simplified Arabic) و(Times New Roman) للإنجليزية والفرنسية وبحجم خط (12). وباستخدام Microsoft Word وعلى وجه واحد للورقة.

ب – الهامش مسافة 2.5 سم لجميع جوانب الورقة.

ج – يرفق الباحث ملخصاً عن بحثه باللغتين العربية والانجليزية والفرنسية بما لا يزيد عن صفحة واحدة.

د - يتم الإشارة الى المصادر العلمية في متن البحث وفي نهايته، مع مراعاة أن لا يتضمن البحث سوى المصادر التي تم الإشارة إليها في المتن ووفق الأصول المعتمدة في ذلك (اسم المؤلف، سنة النشر، عنوان المصدر، دار النشر، البلد).

هـ– ترقم الجداول والرسوم التوضيحية وغيرها حسب ورودها في البحث، كما توثق المستعارة منها بالمصادر الأصلية.

و– أن لا يزيد عدد صفحات البحث اوالدراسة عن (25) صفحة.

3 – يتم إشعار الباحث باستلام بحثه خلال مدة لا تتجاوز يومين عمل من تاريخ استلام البحث.

4 – تخضع كافة البحوث المرسلة الى المجلة للتقييم العلمي الموضوعي ويبلغ الباحث بالتقييم والتعديلات المقترحة إن وجدت خلال مدة لا تتجاوز اسبوعان من تاريخ استلام البحث.

5 – لهيئة تحرير المجلة الحق في قبول او رفض البحث ولها الحق في إجراء أي تعديل او إعادة صياغة جزئية للمواد المقدمة للنشر بما يتماشى والنسق المعتمد في النشر لديها بعد موافقة الباحث.

6 – يصبح البحث المنشور ملكاً للمجلة ولا يجوز إعادة نشره في أماكن أخرى.

7 – تعبر المواد المنشورة بالمجلة عن آراء أصحابها، ولا تعكس وجهة نظر المجلة او المعهد العربي للتدريب والبحوث الإحصائية.

8 – ترسل البحوث على العنوان الالكتروني للمجلة:

journal@aitrs.org **/** Info@aitrs.org

المحتويات

Comparison of some partial methods of the logistic regression model with the application

Prof. Dr. Haifa Taha Abd Soadad Rashied Hameed AL-Zuhairi
Statistics department / College of Administration & Economics
Mustansiriyah University

تاريخ استلام البحث: 14/ 2023/09
تاريخ قبول البحوث: 25/ 2023/10
نشر البحث في العدد الثاني والعشرين: اذار / مارس 2024

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/519.5
رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/519.5

Comparison of some partial methods of the logistic regression model with the application

Prof. Dr. Haifa Taha Abd Soadad Rashied Hameed AL-Zuhairi
Statistics department / College of Administration & Economics
Mustansiriyah University

Abstract:
The Logistic regression is a type of regression analysis that is used in the event that the dependent variable is discrete, either 1 or zero, such as (true or false, white or black, colored or uncolored, infected or uninfected, ... etc.) and used to predict The probability of a particular event occurring by adding the values of the variables explained or associated with the event. It is a more powerful tool because it provides a test for the significance of the parameters, and it can also include the qualitative independent variables as well as the effect of the interaction between the independent variables in the dependent variable with a two-value value, and one of the advantages of using logistic regression is that it is less sensitive to deviations from the normal distribution of the study variables, compared to other methods. And that the problem of multilinearity appears when there is a relationship between two or more of the explanatory variables, which leads to estimating the logistic regression model differently due to the presence of some non-significant variables, so we need a way to estimate the parameters of a better model through a comparison between the usual methods and the reduction methods. They are methods through which the sum of the squares of the residuals bound by the penalty constraint of the general logistic linear model is minimized.

Where the shrinkage parameter is assumed so that some coefficients approach zero or may be equal to zero, and if the parameter is zero, this means that the capabilities of the model before shrinkage (usual least squares method) are the same as the capabilities of the model after shrinkage (which has more than one method). Among the most important of these methods used is the method of the normal lasso, the adaptive lasso, the Scad, and the method proposed by the researcher is the method of the Bayesian lasso - with an exponential natural Gama distribution. By conducting the simulation process for samples with sizes (250,200,150,75), the comparison was made by calculating the mean squared errors and the mean squared absolute errors to get the best estimate for the parameter β. In order to apply the simulation in reality, data were taken from the Ministry of Health regarding pregnant women (response variable) and several factors affecting miscarriage such as age and others (explanatory variables), 35 variables, with a frequency of 354 (cases). In addition, calculating the shrinkage parameter and finding the results for the two methods specifically, which is better.

Keywords: logistic regression, downscaling methods, downscaling estimator, Lasso method, maximum likelihood method, Lasso Bayesian method, comparison criteria, testing the goodness of estimators.

المستخلص

ان الانحدار اللوجستي Logistic Regression هو نوع من انواع تحليل الانحدار الذي يستخدم في حالة كون المتغير التابع منفصل اما 1 او صفر مثل (صح او خطأ , ابيض او اسود , ملون او غير ملون , مصاب او غير مصاب , ... الخ) ويستخدم للتنبؤ باحتمالية وقوع حدث معين بأضافة قيم متغيرات مفسرة او مرتبطة بالحدث , وهو اداة اكثر قوة لانه يقدم اختبارا لمعنوية المعلمات ، كما يمكنه ان يتضمن المتغيرات المستقلة النوعية وكذلك تأثير التفاعل بين المتغيرات المستقلة في المتغير التابع ثنائي القيمة ، كما ان من مزايا استخدام الانحدار اللوجستي انه اقل حساسية تجاه الانحرافات عن التوزيع الطبيعي لمتغيرات الدراسة. وان مشكلة التعدد الخطي تظهر عند وجود علاقة بين اثنين او اكثر من المتغيرات التوضيحية مما يؤدي الى تقدير أنموذج أنحدار لوجستي مختلف لوجود بعض المتغيرات غير المعنوية ، لذا نحتاج الى طريقة تقدير معلمات أنموذج أفضل عن طريق مقارنة بين الطرق الاعتيادية وطرق التقليص . وهي طرائق يتم من خلالها تصغير مجموع مربعات البواقي المقيدة بقيد الجزاء للانموذج الخطي العام اللوجستي ، حيث يتم افتراض معلمة التقلص لتقترب بعض المعاملات نحو الصفر او قد تساوي صفراً ، وفي حال كون المعلمة صفراً هذا يعني ان مقدرات الانموذج قبل التقليص (طريقة المربعات الصغرى الاعتيادية) هي نفسها مقدرات الانموذج بعد التقليص (والتي لها اكثر من طريقة) . ومن أهم تلك الطرائق أستخداما" هي طريقة LASSO , LASSO التكييفي , سكاد , والطريقة التي تم توظيفها من الباحثة هي طريقة LASSO بيزي – بتوزيع كاما طبيعي أسي . وبأجراء عملية المحاكاة للعينات بحجوم (250,200,150,75) وتمت المقارنة بحساب متوسط مربعات الاخطاء , ومتوسط مربعات الاخطاء المطلق للحصول على أفضل مقدر للمعلمة β . ولتطبيق المحاكاة في الواقع أخذت بيانات من وزارة الصحة تخص نساء حوامل (متغير أستجابة) وعدة عوامل مؤثرة على أسقاط الحمل مثل العمر وغيرها (متغيرات توضيحية) عدد 35 متغيرا" وبتكرار 354 (حالة) . وحساب معلمة التقليص وايجاد النتائج للطرائق الاربعة تحديدا" ايهم افضل.

الكلمات الافتتاحية : الانحدار اللوجستي , طرائق التقليص , مقدر التقليص , طريقة لاسو , طريقة الامكان الاعظم , طريقة لاسو بيز , معايير المقارنة , اختبار جودة المقدرات .

Introduction

Statistics is an important social science that contributes to collecting and analyzing data and extracting results in many areas of medical, agricultural, industrial and other life. One of the important topics in statistics is regression analysis, which is a function of analyzing the relationship between two or more variables. Estimation or prediction is make through a probabilistic mathematical model that contains only one response variable Yi, and one or more explanatory variables Xi.

1- The Theoretical side

1-1- Logistic Regression Model:

It is a statistical model that belongs to the general linear regression models, as the response variable is binary or multiple, and it is a type of regression analysis that is used in the event that the dependent variable is separate, either 1 or zero, such as (true or false, white or black, colored or uncolored, infected). Alternatively, uninfected ... etc.) It is use to find the relationship between two variables and to predict the probability of a particular event occurring by adding the values of the variables explained or related to the event. It is a more powerful tool because it provides a test for the significance of the parameters. It can also include the qualitative independent variables as well as the effect of the interaction between the independent variables in the dependent variable of two values.

One of the advantages of using logistic regression is that it is less sensitive to deviations from the normal distribution of the study variables, compared to other statistical methods such as discriminant analysis and linear regression. Logistic regression is important in the field of artificial intelligence and machine learning, which is performing large and complex data processing tasks without human intervention, and this, helps organizations to carry out their data work efficiently and quickly.

The probability function for the variable y is:

$$\pi(t) = Log\,\frac{\pi(x)}{1-\pi(x)} = \beta_0 + x\,.\,\beta\pi_i = \frac{\exp\,(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp\,(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}$$

The logistic regression model with predictor variable P is expressed as follows:

It is assumed that Y1, …, Yn are response variables with values of 0, 1

And that X1, …, Xn are independent variables, so that Xi = (1,Xi1,…, Xip)

$$\pi(t) = e^{t} \,/\, (1 + e^{t})$$

$$\pi_i = \frac{\exp\,(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp\,(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}$$

The logistic regression model is transformed using logarithm to get:

$$\log\left(\frac{\pi(x)}{1+\ \pi(x)}\right) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}{1+\ e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}$$

$$\text{Log}\left(\frac{\pi(x)}{1+\ \pi(x)}\right) = \frac{1}{1+\ e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}}$$

1-2- Shrinkage Methods:

The problem of multilinear appears when there is a relationship between two or more of the explanatory variables, which leads to a different estimation of the logistic regression model due to the presence of some non-significant variables, so we need a way to estimate the parameters of a better model through a comparison between the usual methods and the reduction methods.

Moreover, by them we mean methods through which the sum of the squares of the residuals bound by the penalty restriction of the general linear logistic model is minimize, where the shrinkage parameter or the penalty is assume so that some coefficients approach zero or may be equal to zero.

Ordinary minimum) is the same as the capabilities of the model after shrinkage (which has more than one method).

In other words, the parameter estimates shrink towards zero for least squares estimates. Shrinkage methods do not use the least squares method, but rather use a different criterion in that there is a penalty function that penalizes the model for having a large number of parameters or a large size of parameters that will typically shrink those parameters towards zero.

1-3- Shrinkage Estimator:

A new estimate produced by shrinking the initial estimate (such as the sample mean). For example, if two extreme mean values can be combined to create a more central mean value, repeating this for all means in the sample will adjust the sample mean, which has shrunk towards the true population mean. In addition, assuming the penalty parameter λ. Moreover, by using a different criterion represented by the presence of a penalty function penalizing the model, it will lead to the reduction as the parameters approach towards zero and get rid of the variables that have no effect on the model. We will explain in detail the

four methods and then compare them to find out which one is more efficient for estimation:
Dozens of shrinkage estimates have been developed by different authors since Stein first introduced the idea in the 1950s.

Among the most famous:

- LASSO estimator.
- Adaptive LASSO estimator.
- SCAD Estimator
- Baysian LASSO estimator.

1-4- The iterated Lasso for logistic regression

abbreviated to (Least Absolute Shrinkage and Selection Operator) It is one of the penalty methods of logistic regression and one of the most famous and most used methods, which appeared for the first time in the geophysical literature in the year (1982), then it was rediscovered by the researcher Robert Tibshirani (1996), where he formulated it and how to perform it.
The lasso is a penalty function of the linear regression model, and it is a method for estimating the parameters of the regression model and selecting and organizing variables to increase accuracy in analyzing the studied phenomenon. Zero so that non-zero coefficients are chosen after the reduction process to be the most accurate prediction model for estimation.

$$f = \sum_{i=1}^{n}\left(\mathrm{Y}_i - \sum_{j=1}^{p}\beta_j\ x_{ij}\right)^2 + \lambda \sum_{j=1}^{p}|\beta_j|$$

$$\mathrm{L}_1\,(\beta\ ;\ \lambda) = \ell\mathrm{n}\,(\ \beta) + \ \lambda \textstyle\sum_{j=1}^{p}|\beta_j|$$

The penalty parameter $\sum_{j=1}^{p}|\beta_j| \ \le\ \lambda \quad , \quad \lambda\ \ge\ 0$

Keep in mind that we have a sample consisting of N cases of covariates Y, which is the result, and explanatory variables $x_i = \left(x_1, x_2, \dots, x_p\right)_i^T$
The purpose of lasso function is to solve:

$$\min_{\beta_0\beta}\left\{\sum_{i=1}^{N}(y_i - \beta_0\ -\ x_i^T\beta\ )^2\right\}$$

Subject to $\sum_{j=1}^{p}|\beta_j|\ \le \lambda$

Here β_0 is the constant coefficient, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ is parameter vector,

And $\boldsymbol{\lambda}$ a free parameter determines the degree of regulation, and different values of it lead to different penalty techniques.

Lasso solution

$$\begin{cases} \sum_{i=1}^{n}\left[Y_i - \pi(x'_i\hat{\beta})\right] = 0 \\ \sum_{i=1}^{n}\left[Y_i - \pi(x'_i\hat{\beta})\right]x_{ij} = \lambda_1\, sgn\left(\hat{\beta}_j\right), \quad \hat{\beta}_j \neq 0, j \geq 1\,, \\ \left|\sum_{i=1}^{n}\left[Y_i - \pi(x'_i\hat{\beta})\right]x_{ij}\right| \leq \lambda_1\,, \quad \hat{\beta}_j = 0, j \geq 1\,, \end{cases}$$

1-5- In Maximum Likelihood Method :

The parameter vector $\hat{\beta}$ with dimensions 1 × (p+1) can be obtained from taking the logarithm of the function:

$$\sum_{i=1}^{n}\left[-Y_i\ \left(\beta_0\,, \beta_1 x_{i1}, \dots, \beta_p x_{ip}\right) + \ Log\left(1 + \exp\left(\beta_0\,, \beta_1 x_{i1}, \dots, \beta_p x_{ip}\right)\right)\right]$$

$$\hat{\beta}_{Lasso} = \min\left\{U'U + \ \lambda \sum_{j=1}^{p}|\beta_j|\right\}$$

1-6- The Bayesian Lasso for logistic regression :-

Abbreviated to (Bayesian Least Absolute Shrinkage and Selection Operator) **,** It is also "one of the partial methods of logistic regression, which was proposed by Bhattacharyya (2022), and is based on the use of an initial distribution and a posteriori distribution and its combination with the logistic regression function. Most of the used reduction methods such as Lasso use the penalty parameter to reduce the coefficients of irrelevant variables,

Bayesian regression methods that include shrink properties as prior information are a reasonable alternative approach. Health researchers often face the problem of predicting a categorical response, depending on the patient's clinical results, and the initial outcome must be expected, for example, the patient's discharge, death, or hospitalization. Logistic regression models allow evaluating the association between the independent variable(s) and the response variable,

When the response variable contains more than two categories, such as polynomial logistic regression models, it can deal with a large number of main and defining effects, environmental effects and the effects of interactions between the environment and genes.

This method adds penalty terms to the probability function of the parameter inference and the minimization leads to a desirable downsizing behavior of the estimates, namely it "draws the small coefficients towards zero while slightly affecting the large coefficients," thus selecting only the most relevant variables. Assuming
Y1, …, Yn are dependent variables with values of 0 , 1
X1, …, Xn are independent variables , so that Xi = (1,Xi1 ,…, Xip) ,
λ is the penalty parameter that modifies the amount of shrinkage, and when equal to zero leads to the logistic probability without shrinkage.

$$L_\lambda(\beta) = \sum_{i=1}^{n} y_i x_i \beta \ \log\left(1 + e^{x_i^{\mathrm{T}} \beta}\right) + \lambda \left(\sum_{j=1}^{p} \left|\beta_j\right|^r\right)^{\frac{1}{r}} \quad , \ r > 0 \, .$$

$$L_\lambda(\beta) = \sum_{i=1}^{N} y_i \ \sum_{k=0}^{K} x_{ik}\beta_k - n_i \log\left(1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}\right) + \lambda \left(\sum_{k=1}^{K}\sum_{j=1}^{p} \left|\beta_{kj}\right|^r\right)^{\frac{1}{r}}$$

Different values of r lead to different penalties.

$$Log\ (P_i) = X_i^T \ \beta$$

Where: $Pi = \frac{1}{1 + e^{-X_i^T \ \beta}}$

Prior dis.

We have:

1. The parameter β_{i} $\mathrm{i} = 1,2, \dots, \mathrm{k}$, $\beta_{\mathrm{i}} \sim \mathrm{N}\ (0\ , \sigma_{\mathrm{i}}^2)$.
2. σ_{i}^2 $\mathrm{i} = 1,2, \dots, \mathrm{k}$, $P\ (\sigma_{\mathrm{i}}^2) = \lambda \exp(-\lambda\, \sigma_{\mathrm{i}}^2\ ) \, , \lambda > 0$.
3. The parameter $\lambda \sim Gamma\ (\ a\ , b)$, We name the logistic Regression model with this normal-exponential-gamma (BLASSO-NEG).

$$P\left(\sigma_i^2\right) = \int_0^\infty \left(P(\ \sigma_i^2\ \left|\lambda\right) P(\lambda) d(\lambda) = \frac{a}{\left(b(\sigma_i^2 \middle| b + 1\right)^{a+1}}$$

$$P(\beta\ , ; \sigma^2 | y) \propto p(y|\beta) p(\beta|\sigma^2) p(\sigma^2)\ )$$

$$p(y|\beta) = \prod_{i=1}^{n} (1 - p_i)^{1-yi} \quad and \quad p(\beta|\sigma^2)$$

Let is A is Diagonal matrix $\boldsymbol{Y}$, $\boldsymbol{A} = \widehat{\boldsymbol{A}}$

The parameter dis. $p(\beta|y) \propto p(y|\beta) p(\beta)$

$$\boldsymbol{Log\ p(\beta|y)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} [\boldsymbol{y_i Log\ p_i} + (\boldsymbol{1} - \boldsymbol{y_i})\ \boldsymbol{Log\ (1 - p_i)}] - \frac{\boldsymbol{1}}{\boldsymbol{2}}\ \boldsymbol{\beta^T\ A\beta} + \textbf{constant}$$

1-7- Estimates quality standards:

Among the criteria, that we use to compare the results of the four methods used to estimate the parameters of the model: -

1. Mean squares of MSE errors
2. Absolute value of mean squared errors |MSE| , AMSE
3. Akaike's Information Criteria AIC
4. Bayesian Information Criterion BIC

2- Test the significance of the logistic regression model coefficients for the four methods:

To test the significance of the regression model coefficients of the Lasso, and Bayesian Lasso methods, the Wald test statistics are used for each parameter of the logistic regression coefficient corresponding to each explanatory variable in order to test the null hypothesis (the null hypothesis), which states that the effect of the logit coefficient is equal to zero" according to Wald's statistic formula:

$$Wald = \frac{b}{SE_b}$$

b The value of the coefficient of the logistic regression model for the explanatory variable

〚SE〛 b is the value of the standard error of the coefficient of the explanatory variable

When Wald test tabular value = 3.8415

Sig =1 The variable is Sig =0 The variable is

Note :- that the test statistic follows the chi-square distribution of X^2 and by comparing the calculated value with the tabular value, in case it is less, the null hypothesis is accepted, and vice versa, the alternative hypothesis is accepted at the confidence limits of 0.01, 0.05, meaning that the effect of the coefficient is not equal to zero in the community from which the sample was drawn, which is Moral.

Data :- Xi , i = 1,2,….,n were n = 35 Case = 354

3- Simulation results / Experimental aspect

Table No. (1) Mean value of 0 and variance of 1 with 1000 repetitions

Methods		**LASSO**	**BAYES**
N=75	λ	0.0018	0.0173
	AIC	180.2763	122.3132
	MSE	0.260806	0.112894
	MAE	0.090076	0.051693
N=150	**λ**	0.00026	0.0005
	AIC	282.6247	147.7074
	MSE	0.25799	0.066576
	MAE	0.089878	0.040797
N=200	**λ**	0.00023	0.0076
	AIC	352.7347	561.8105
	MSE	0.258055	0.046591
	MAE	0.089819	0.033475
N=250	**λ**	0.0002	0.2972
	AIC	422.1967	1767.26
	MSE	0.259808	0.108029
	MAE	0.090197	0.041654

Table No. (2) Mean value of 0 and variance of 2 with 1000 repetitions

Methods		**LASSO**	**BAYES**
N=75	λ	0.0043	0.3333
	AIC	178.1558	136.6873
	MSE	0.266089	0.171661
	MAE	0.090871	0.069262
N=150	λ	0.00024	0.0017
	AIC	278.5725	301.7659
	MSE	0.264709	0.072028
	MAE	0.090939	0.044488
N=200	λ	0.00025	0.0040
	AIC	349.9496	1080.909
	MSE	0.262677	0.08675
	MAE	0.090524	0.042424
N=250	λ	0.00027	0.0215
	AIC	419.9862	2256.898
	MSE	0.263537	0.115463
	MAE	0.090841	0.043581

Simulation results / the applied aspect

Table No. (3) Practical results / Real data chart

	LASSO B	ADLASSO B	SCAD B	BAYES B
MSE	0.2327	0.2538	0.2503	**0.2134**
MAE	0.4741	0.4253	0.5003	**0.4144**
AIC	552.5064	339.7298	570.959	**72**
BIC	269.3136	56.537	287.7661	**211.1929-**
Lamda	0.0445	0.000045	0.0433	**0.000031**

Table No. (4) Shows the results

Methods	LASSO	Wald	Sig.	BAYES	Wald	Sig.
Constant	-1.10261	1.373333	0	14.58087	0.751807	0
x1	0	0	0	-2.89975	17.66754	1
x2	0	0	0	-1.86489	17.26774	1
x3	0	0	0	2.075023	134.7509	1
x4	0.467998	87.15934	1	11.07507	651.5928	1
x5	0	0	0	3.964936	-3.47862	0
x6	0	0	0	-0.14785	0.805053	0
x7	0	0	0	-2.67531	-737.434	0
x8	0	0	0	4.901702	29.04195	1
x9	0	0	0	-1.16659	7.099813	1
x10	0.037026	28.17539	1	2.791139	-102.054	0
x11	0	0	0	-0.19396	1.721278	0
x12	0	0	0	32.97147	79.05724	1
x13	0	0	0	-0.0304	0.328499	0
x14	0	0	0	0.555553	-4.63049	0
x15	0	0	0	-0.01306	0.108855	0
x16	0	0	0	2.209889	18.69149	1
x17	-0.00706	-3.71293	0	5.69171	56.26704	1
x18	0	0	0	3.076329	-78.0534	0
x19	0	0	0	-7.36898	13.95081	1
x20	0	0	0	-0.15186	0.256604	0
x21	0	0	0	5.00664	-17.2227	0
x22	0	0	0	50.23825	64.51828	
x23	0	0	0	-4.57261	3.713257	0
x24	0	0	0	-0.05517	0.027698	0

Methods	LASSO	Wald	Sig.	BAYES	Wald	Sig.
x25	0	0	0	2.688829	-4.9486	0
x26	0	0	0	1.002206	5.341749	1
x27	0	0	0	0.205508	-0.08675	0
x28	0	0	0	11.15969	47.64421	1
x29	0	0	0	0.163936	0.748349	0
x30	0	0	0	-33.2652	74.00046	1
x31	0	0	0	0.123379	-3.08149	0
x32	0	0	0	4.170775	-2.48864	0
x33	0	0	0	0.803512	-2.76054	0
x34	0	0	0	5.676594	-16.5728	0
x35	0	0	0	-0.44507	1.371928	0
Sum	2			14		
MSE	0.2327			0.4844		
MAE	0.4741			0.4844		
Lamda	0.0445			0.000031		

4- Conclusions and recommendations

4-1- Conclusions:

1. Employing the Lasso Bayes method (Lasso - Gamma Exponential Normal) is the best efficient and most accurate in the second case for values of mean = 0 and variance = 1, and the possibility of applying it in the case of data that follows a linear or nonlinear distribution, logistic distribution, normal distribution, exponential distribution, Poisson...
2. When calculating the value of λ (the shrinkage parameter) based on the bootstrap method, at a sample size of 75 we find that the BLASSO method is the best based on the lowest AIC value, while for other sample sizes the LASSO method was the best.
3. From the simulation results, it appears that the value of the comparison standard MAE decreases as the sample size increases, while the value of MSE and AIC increases as the sample size increases.
4. 6. From the Wald test, the significance of the explanatory variables was determined so that non-significant variables were deleted and excluded from the logistic model, sig = 0, non-significant, the variable is removed, sig = 1, significant, the variable remains, as 21 variables were excluded by the BLASSO method.

4-2- Recommendations

1. The possibility of applying downscaling methods to other studies in the field of health, agriculture, ... and other areas of life due to their accuracy and efficiency and the exclusion of variables that have no effect.
2. We suggest applying the Bayesian adaptive Lasso method to the data of this thesis or other data, due to the possibility of its application in terms of the closeness of the Lasso model to the adaptive Lasso.

References

1. Pekhimenko, Gennady G. (2008) “Penalizied Logistic Regression for Classification “Department of Computer Science, University of Toronto.
2. Huang, Jian Huang (2008) “The Iterated Lasso for High-Dimensional Logistic Regression”, University of Iowa.
3. Yan, Fang – Rong Yan (2011) “Sparse Logisitic Regression for Diagnosis of Live Fibrosis in Rat by Using SCAD-Penalized Likelihood “Journal of Biomedicine and Biotechnology.
4. Huang, Anhui Huang (2013) "Empirical Bayesian LASSO-Logistic regression for multiple binary locus mapping” Department of Electrical and Computer Engineering, University of Miami, USA.
5. Algamal, Zakariya Yahya (2015) “Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification, Department of Mathematical Sciences, Universiti Teknologi Malaysia.
6. Shofiyah, F Shofiyah (2018) “Split and Conquer Method in Penalized Logistic Regression with Lasso (Application on Credit Scoring Data)”, Journal of Physics.
7. Hassel, Mohammed Abdulkadhim (2021) “Sparse sliced inverse regression via elastic net penalty with an application" College of Administration and Economics, University of Al-Qadisiyah.
8. Mitra, Riten Mitra (2022) ". Applications of Bayesian shrinkage prior models in clinical research with categorical responses", BMC Medical Research Methodology.
9. Chen, Shunjie (2022) “Recent Advances on Penalized Regression Models for Biological Data “, School of Mathematics and Statistics, Henan University, China.

Publication Prerequisites and terms

1- The journal publishes scientific research and studies in statistics and informatics written in Arabic, English and French, to make it clear that research submitted for publication has been published or submitted for publication in magazines or other periodicals or presented and published in periodicals for conferences or seminars.
2- Send electronic copies (word & PDF) of the research and studies to the editor should include the name of the researcher or researchers and their scientific titles and places of work with the address of the correspondence, the numbers of telephones and e-mail. The research to be published should be sent electronically in accordance with the specifications below:
    a. To be printed on A4 paper and be in the form of a single column and use the Type simplified Arabic and Times New Roman for English and French and with a font size (12). Using Microsoft Word and on one face of the paper.
    b. The margin is 2.5 cm for all sides of the paper.
    c. The researcher will attach a summary of his research in Arabic, English, or French in no more than one page.
    d. Place references at the end of the paper and separated page. It is recommended to use the Harvard system of referencing, which (author's name, year of publication, source address, publishing house, country).
    e. Numbered tables, illustrations, and others as they are received in the research, documents as aliases of the original sources.
    f. The number of search or study pages should not exceed (25) pages.
3- Authors will notified of receiving their research within two working days from the date of receipt of the research.
4- Referees will evaluate all submitted research, the Authors will informed of the proposed evaluation and modifications if any within two weeks of receipt of the research.
5- The editorial board of the Journal has the right to accept or reject the research and has the right to make any modification or partial redrafting of the material submitted for publication in accordance with the format adopted in its publication after the approval of the researcher.
6- Published research becomes the property of the Journal and may not republished elsewhere.
7- The articles published in the magazine reflect the opinions of the authors, and do not necessarily reflect the view of the Journal or the Arab Institute for Training and Research in Statistics.
8- The research is sent to the magazine's e-mail address:
    journal@aitrs.org or **Info@aitrs.org**

Journal of Statistical Sciences

Scientific Referred Journal

Listed in Ulrich’s website
www.ulrichsweb.com
Classified in The Arab Citation & Impact Factor (Arcif)
www.emarefa.net/arcif/
ISSN 2522-64X (Online), ISSN 2519-948X (Print)

P.O. Box: 851104- Amman 11185 Jordan ص.ب: 851104 عمان 11185 الأردن
Tel: +96265549805 – Fax: +96265549804- www.aitrs.org , info@aitrs.org

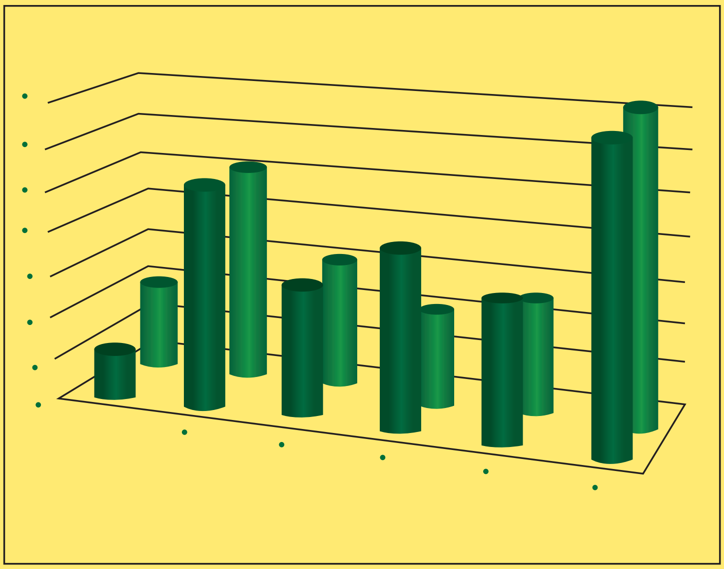
مجلة العلوم الإحصائية

March. 2024

No. 22

Arab Institute for Training and Research in Statistics

Journal of Statistical Sciences



Issue No. 22

Scientific Peer-reviewed Journal issued by

Arab Institute for Training and Research in Statistics

Listed in Ulrich's website

www.ulrichsweb.com

Classified in The Arab Citation & Impact Factor (Arcif)

www.emarefa.net/arcif/

ISSN 2522-64X (Online), ISSN 2519-948X (Print)