A COMPARISON OF SOME SHRINKAGE METHODS FOR LOGISTIC REGRESSION MODEL

[1] **Soadad Rashied Hameed**
soadadrashied@gmail.com

[2] **Haifa Taha Abd**
haefaa_adm@uomustansiriyah.edu.iq
[1],[2] Al-Mustansiriya University/ administration and economic college/ statistics department

Abstract:

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. And assuming the penalty parameter λ . And 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.

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, It was concluded that the adaptive Lasso method was better Reduction methods: The Bayesian lasso method also showed good results.

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 method of analyzing the relationship between two or more variables, and because it is a multi-use statistical tool for analyzing data, estimating parameters, and representing the relationship between phenomena. Estimation or prediction is made through a probabilistic mathematical model.

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

Logistic Regression Model:

The probability function for the variable y is:

$$\pi(t) = Log\,\frac{\pi(x)}{1-\pi(x)} = \beta_0 + x\,.\,\beta$$

$$\pi_i = \frac{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)} \qquad ; \; 1 \le i \le p$$

The logistic regression model is transformed using logarithm to get:

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

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

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. And assuming the penalty parameter λ . And 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
- Bayesian LASSO estimator.

Simulation results

Initial values were determined to conduct the simulation experiment for 34 variables, as shown in the two tables(1,2).

Table (1)

Generate initial variables

Mean	Sigma	Explanatory variables	Sizes of samples			
0	0.5 , 1 , 2	35	75	150	200	250

Table (2)
The value of the initial parameters

Variable	Value	Variable	Value
constant	5.00	x20	2.10
x1	2.30	x21	2.60
x2	1.50	x22	1.20
x3	1.90	x23	1.40
x4	3.10	x24	3.20
x5	4.00	x25	2.40
x6	3.50	x26	1.30
x7	1.30	x27	1.90
x8	3.20	x28	2.10
x9	2.40	x29	1.90
x10	1.80	x30	3.30
x11	2.70	x31	4.20
x12	3.40	x32	2.00
x13	1.30	x33	3.80
x14	2.70	x34	2.90
x15	4.10	x35	4.60
x16	3.20		
x17	2.80		
x18	1.10		
x19	2.30		

- **Calculating the lambda value using the bootstrap method**

Table (3)

results between the methods at a mean value of 0 and a variance of 0.5 with 1000 repetitions

Methods		**LASSO**	**Ad LASSO**	**SCAD**	**Bayesian lasso**
N=75	**λ**	0.0085	0.0001	0.0002	0.0004
	AIC	175.4669	490.1311	325.0953	297.2949
	MSE	0.270398	0.048241	0.162762	0.125242
	MAE	0.092065	0.028439	0.071334	0.056295
N=150	**λ**	0.0023	0.0001	0.0003	0.0987
	AIC	279.4449	953.7731	532.8052	793.856
	MSE	0.264317	0.029014	0.009596	0.056042
	MAE	0.090972	0.015927	0.016805	0.035691

N=200	**λ**	0.00024	0.00021	0.0003	0.0232
	AIC	357.1948	1381.544	718.831	1995.097
	MSE	0.265575	0.027157	0.122756	0.096973
	MAE	0.091111	0.01302	0.061877	0.041786
N=250	**λ**	0.0002	0.0004	0.0001	0.0049
	AIC	426.0112	1601.711	866.4647	3269.661
	MSE	0.263493	0.027692	0.143857	0.130391
	MAE	0.090764	0.014074	0.067031	0.042997

Based on the ACI value, it turns out that the LASSO method is the best method at a sample size of 75, 150, 200, and 250. As for the MSE values, it turns out that the Bayesian lasso method is the best at a sample size of 200, the SCAD method is the best at a sample size of 150, and the ALASSO method is the best at a sample size of 150. Sample 75, 250 based on the lowest value of the standard.

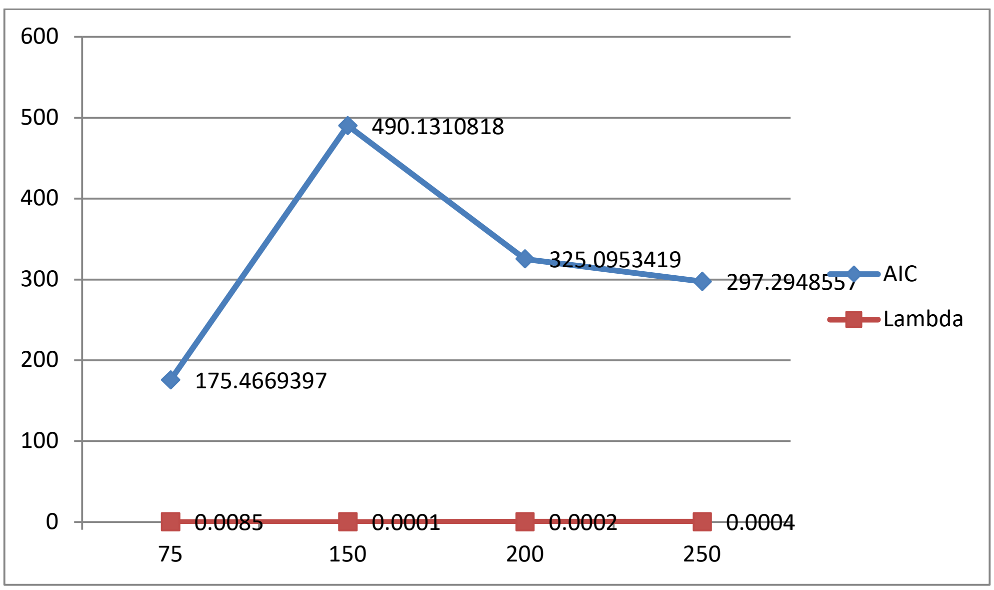


Figure (1): methods for range with AIC at a sample size of 75

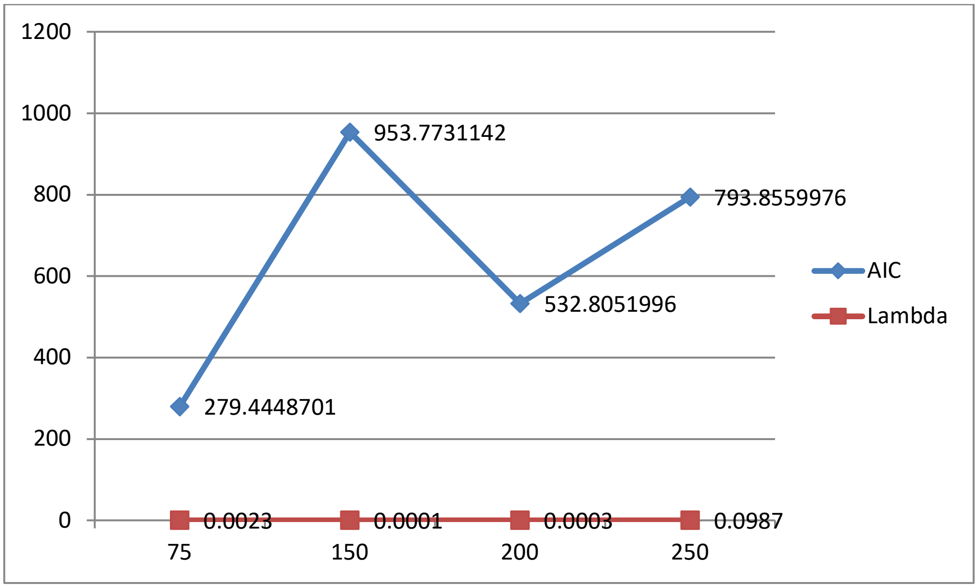


Figure (2): methods for range with AIC at a sample size of 150

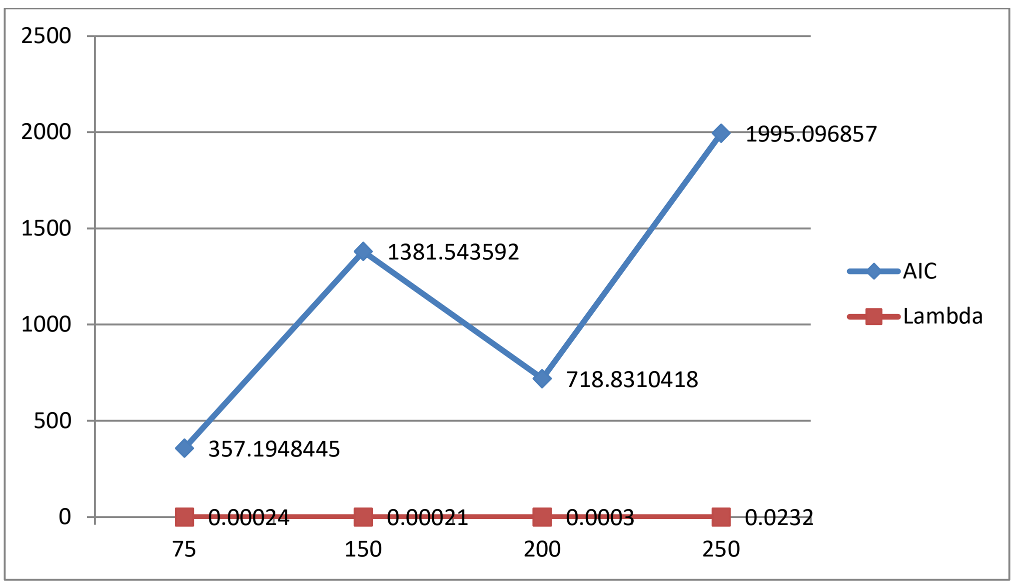


Figure (3): methods for range with AIC at a sample size of 200

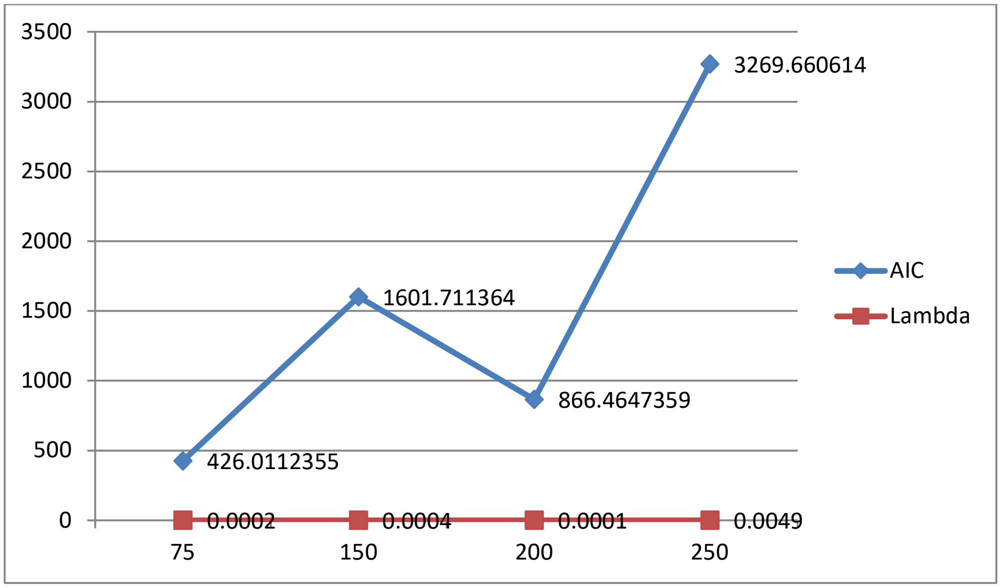


Figure (4): methods for range with AIC at a sample size of 250

Table (4)

results between the methods at a mean value of 0 and a variance of 1 with 1000 repetitions

Methods		LASSO	Ad LASSO	SCAD	Baysian lasso
N=75	λ	0.0018	0.0002	0.0001	0.0173
	AIC	180.2763	445.856	188.8781	122.3132
	MSE	0.260806	0.03148	0.210886	0.112894
	MAE	0.090076	0.017056	0.081274	0.051693
N=150	λ	0.00026	0.00024	0.0003	0.0005
	AIC	282.6247	638.8852	229.0194	147.7074
	MSE	0.25799	0.026802	0.044186	0.066576
	MAE	0.089878	0.012351	0.036886	0.040797
N=200	λ	0.00023	0.00027	0.0001	0.0076
	AIC	352.7347	827.1732	276.7883	561.8105
	MSE	0.258055	0.026709	0.008233	0.046591
	MAE	0.089819	0.012083	0.015505	0.033475
N=250	λ	0.0002	0.0003	0.0004	0.2972
	AIC	422.1967	1046.379	361.1637	1767.26
	MSE	0.259808	0.026552	0.051655	0.108029
	MAE	0.090197	0.011743	0.039933	0.041654

Based on the ACI value, it turns out that the Bayesian LASSO method is the best method at a sample size of 75, 150, followed by the SCAD method, the best method at a sample size of 200, 250. As for the MSE values, it shows that the ALASSO method is the best at a sample size of 75, and the SCAD method is the best at a sample size of 200, 250. The Bayesian lasso method is best at a sample size of 150, based on the lowest value of the standard.

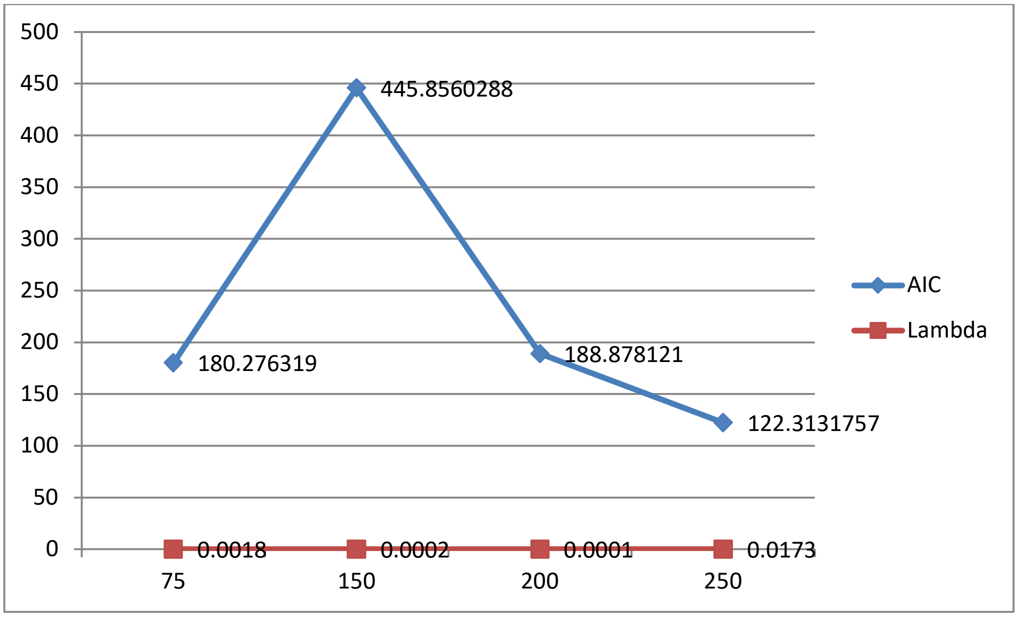


Figure (5):methods for range with AIC at a sample size of 75

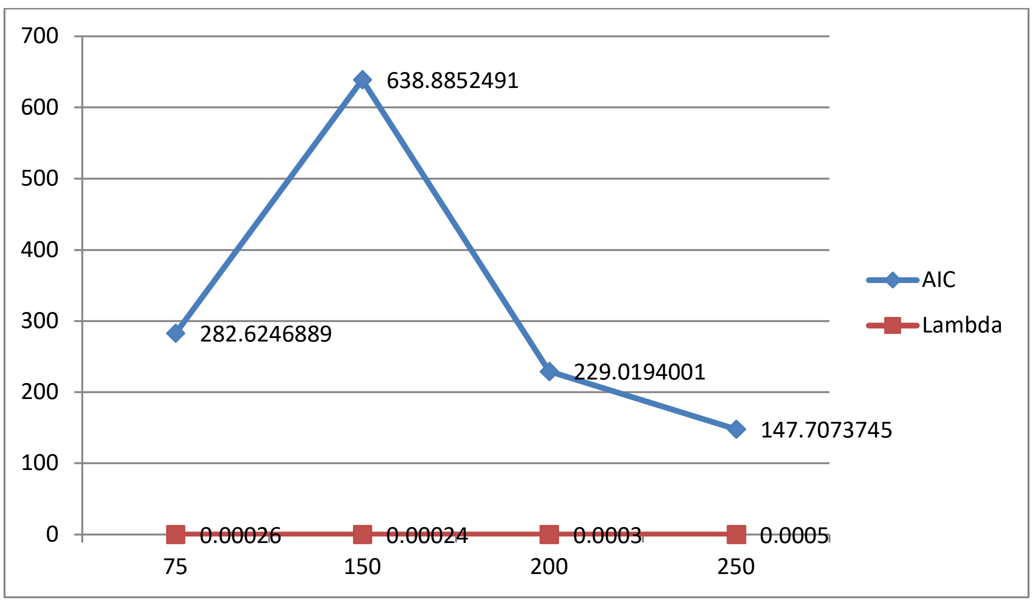


Figure (6):methods for range with AIC at a sample size of 150

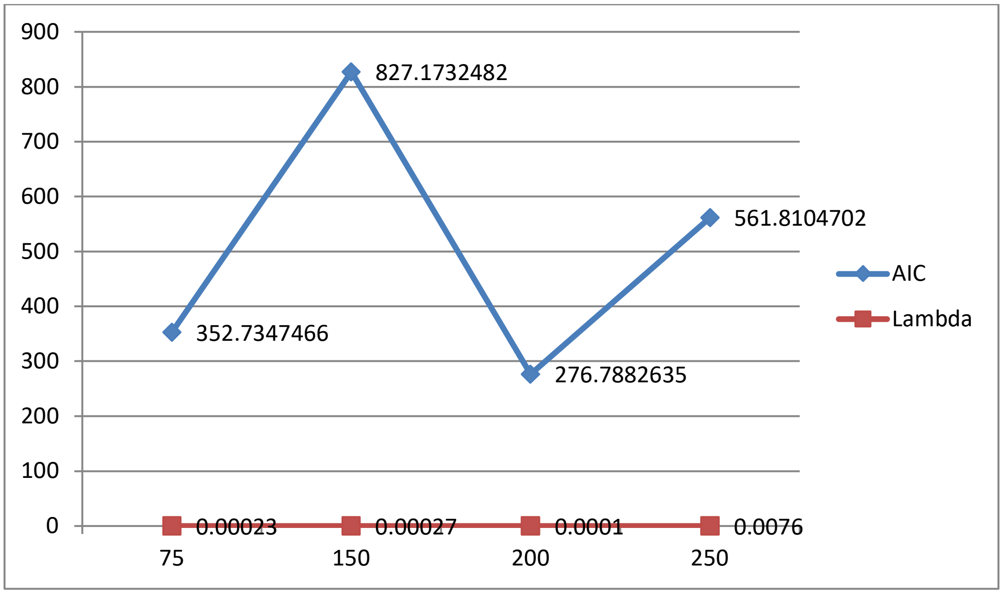


Figure (7): methods for range with AIC at a sample size of 200

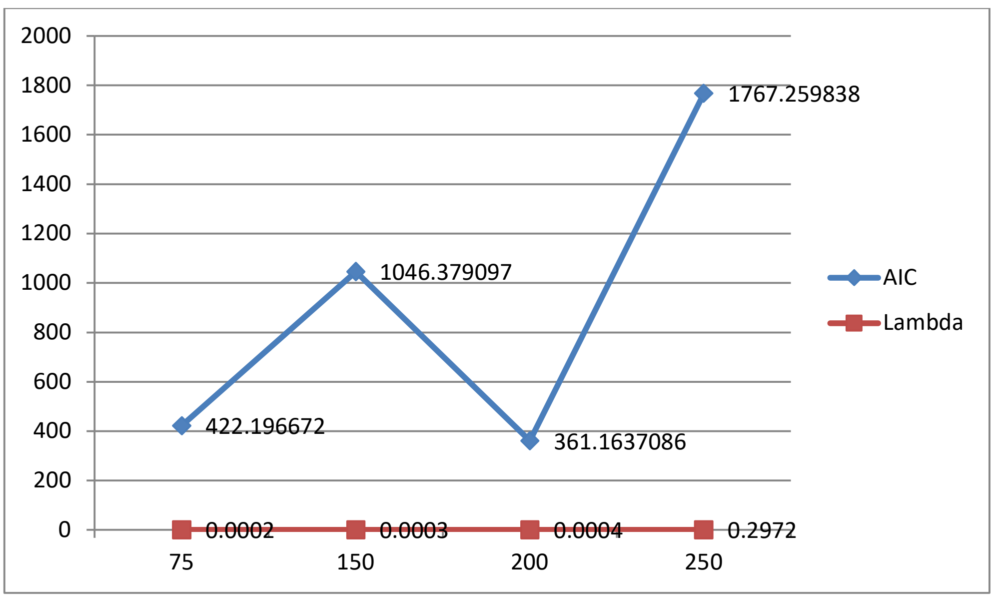


Figure (8): methods for range with AIC at a sample size of 250

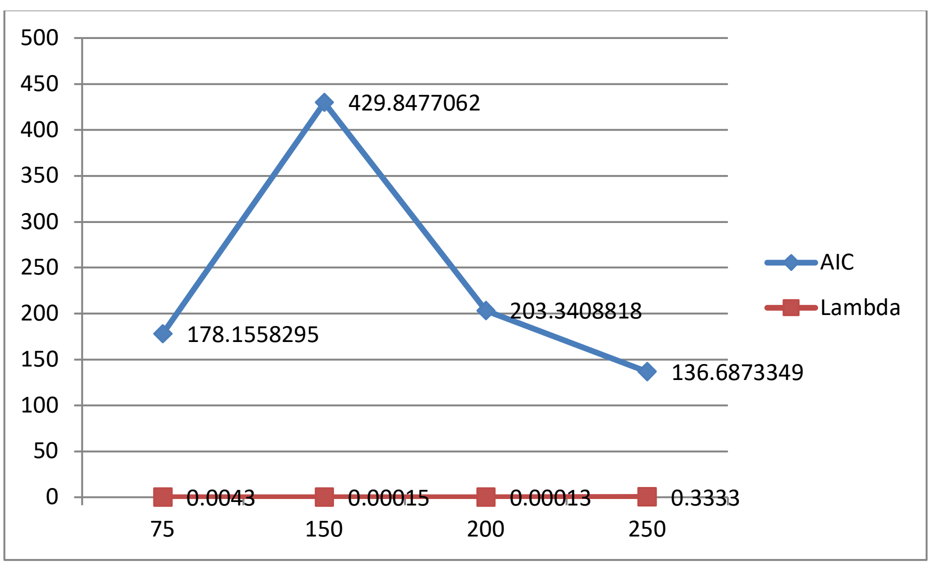
Table (5)

results between the methods at a mean value of 0 and a variance of 2 with 1000 repetitions

Methods		**LASSO**	**Ad LASSO**	**SCAD**	**BAYES**
N=75	**λ**	0.0043	0.00015	0.00013	0.3333
	AIC	178.1558	429.8477	203.3409	136.6873
	MSE	0.266089	0.033355	0.262414	0.171661

	MAE	0.090871	0.019229	0.090723	0.069262
N=150	**λ**	0.00024	0.00023	0.0001	0.0017
	AIC	278.5725	684.0223	312.171	301.7659
	MSE	0.264709	0.027831	0.102023	0.072028
	MAE	0.090939	0.01418	0.05636	0.044488
N=200	**λ**	0.00025	0.0002	0.0001	0.0040
	AIC	349.9496	1007.239	402.6138	1080.909
	MSE	0.262677	0.026525	0.055454	0.08675
	MAE	0.090524	0.011699	0.041397	0.042424
N=250	**λ**	0.00027	0.0002	0.00014	0.0215
	AIC	419.9862	1191.77	499.7032	2256.898
	MSE	0.263537	0.026948	0.000779	0.115463
	MAE	0.090841	0.012742	0.003857	0.043581

Based on the ACI value, it turns out that the Bayesian LASSO method is the best method at a sample size of 75.150, followed by the LASSO method as the best method at a sample size of 200.250. As for the MSE values, it shows that the ALASSO method is the best at a sample size of 75 and the Bayesian lasso method is the best at a sample size of 150.200. The SCAD method is best at a sample size of 250, based on the lowest value of the standard.



Figure(9): methods for range with AIC at a sample size of 75

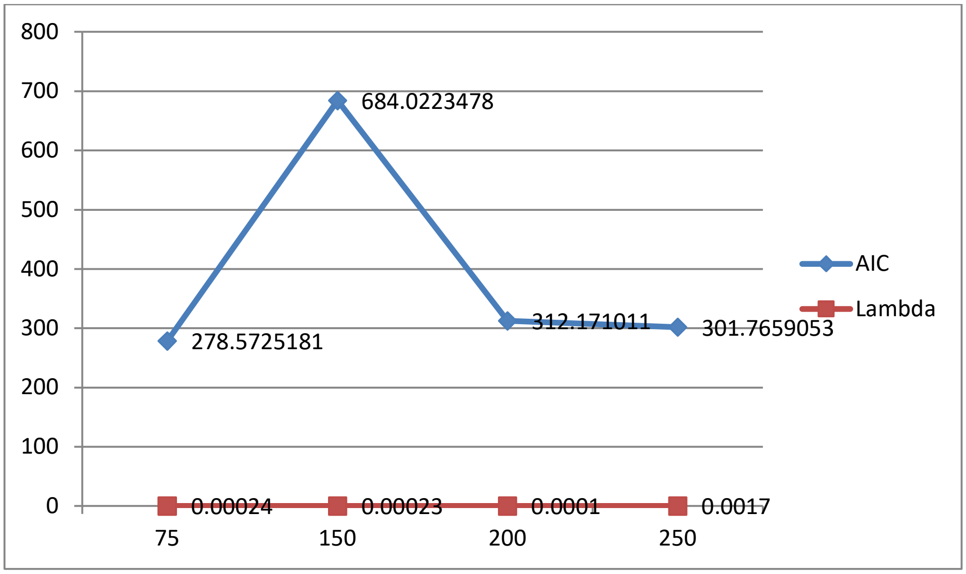


Figure (10):methods for range with AIC at a sample size of 150

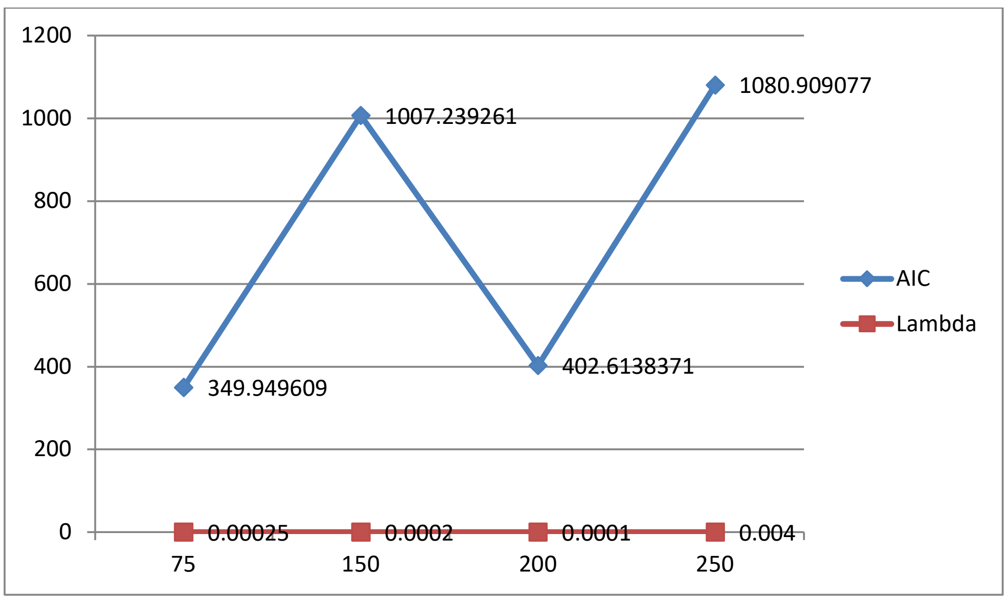


Figure (11): methods for range with AIC at a sample size of 200

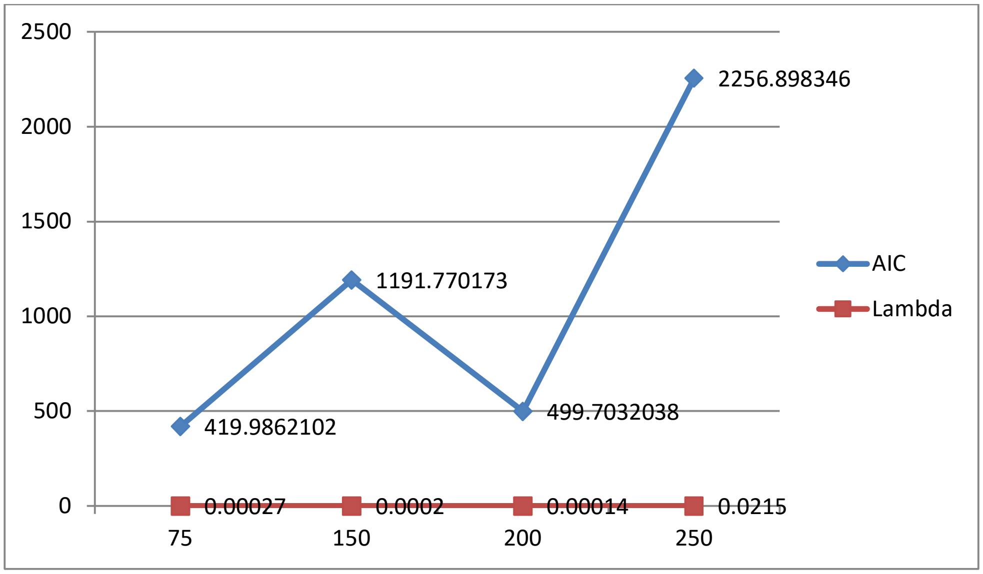


Figure (12): methods for range with AIC at a sample size of 250

Conclusions And recommendations

Conclusions :-

1. It was concluded that the Bayesian lasso method is the best efficient and most accurate, in the second case of mean and variance values, and the possibility of applying it in the case of data following a linear or multiple distribution, logistic distribution. , normal distribution, exponential distribution, Poisson distribution,....

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 Baysian LASSO method is better based on the lowest ACI value, while at 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, AIC, and BIC increases as the sample size increases.

4. From the simulation results, we note the efficiency of the Baysian LASSO method when choosing a mean of 0 and a variance of 1.2 for different samples compared to "when choosing a mean of 0 and a variance of 0.5."

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. Pay attention to how to calculate the value of the shrinkage parameter because it is one of the values that most affects the shrinkage methods.

3. Although the Bayesian Lasso method was not the best method, I nevertheless recommend studying the Bayesian method with each of the downscaling methods because I believe that the Bayesian method is one of the most important estimation methods. Connecting other downscaling methods such as Ridge, Adaptive Lasso to the Bayesian method will give Definitely better results.

Sources :-

1. 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 .
2. Huang , Anhui Huang (2013) "Empirical Bayesian LASSO-Logistic regression for multiple binary locus mapping " Department of Electrical and Computer Engineering, University of Miami, USA.
3. Shofiyah , F Shofiyah (2018) " Split and Conquer Method in Penalized Logistic Regression with Lasso (Application on Credit Scoring Data) " , Journal of Physics .
4. Hassel , Mohammed Abdulkadhim (2021) " Sparse sliced inverse regression via elastic net penalty with an application" College of Administration and Economics , University of Al-Qadisiyah .
5. Mitra, Riten Mitra (2022) ". Applications of Bayesian shrinkage prior models in clinical research with categorical responses", BMC Medical Research Methodology.
6. Chen , Shunjie (2022) " Recent Advances on Penalized Regression Models for Biological Data " , School of Mathematics and Statistics, Henan University, China.