STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **x**, Month 202x, pp 0–6.
Published online in International Academic Press (www.IAPress.org)

# Variable selection in beta regression model using firefly algorithm

Zahraa Tariq Mohammed Taher [*]

*Medicine college- Family and community medicine-Ninevah University- Mosul – Iraq*

**Abstract** The Beta regression model presents widespread scientific interest when used for modeling both proportions and rates data. Creating a predictive regression model requires the identification of select important variables from abundant available options. This work introduces the use of a firefly algorithm for selecting variables when applying the beta regression model featuring varying dispersion parameters. Evaluation of the proposed method's performance takes place through simulations and real data implementation. The proposed method demonstrates better performance than corrected Akaike information criterion, corrected Schwarz information criterion, and corrected Hannan and Quinn criterion in results analysis. The proposed method functions effectively to select variables in beta regression models which have varying dispersion levels.

**Keywords** Firefly algorithm; beta regression model; variable selection; varying dispersion.

**AMS 2010 subject classifications** 62Exx, 03E72

**DOI:** 10.19139/soic-2310-5070-2648

## 1. Introduction

basic statistical method for examining and measuring the connection between a dependent variable and one or more independent variables is regression modeling. With this approach, a model is fitted to the data in order to forecast the dependent variable's value based on the independent variables' values. Regression models are useful for both prediction and inference. The statistical model Beta regression exists to process continuous variables when their range lies between 0 and 1. The beta regression technique is optimal for modeling data situations involving proportions alongside percentages and fractions across scientific disciplines that include economics and medicine with social sciences [5, 4, 11, 14, 15, 16] In these circumstances, the traditional linear regression approach, which is predicated on the ordinary least squares method, is not suitable [6, 17, 18]. Consequently, [4] introduced beta regression model in which the response variable is distributed from the beta distribution.

Many real applications see a rise in the practice of obtaining and measuring various standardized variables. The datasets generally contain numerous irrelevant and redundant variables that negatively impact model prediction accuracy. The presence of numerous variables affects the regression model negatively. Building predictive regression models requires the essential task of choosing relevant variables from within excess numbers of variables [1, 13, 19, 20, 21].

The best variable subset identification process presents an NP-hard computational difficulty which demands substantial time and expensive resources for processing. The traditional variable selection methods including stepwise selection and information criteria as well as backward elimination demonstrate higher computational costs when used. In recent years, the meta-heuristics algorithms, such as firefly algorithm, are widely applied as variable selection methods [8, 22, 23]. This is due to the fact that variable selection is regarded as an optimization

[*] Correspondence to: Zahraa Tariq Mohammed Taher (Email: zahraa.mohammed@uoninevah.edu.iq). Medicine college- Family and community medicine-Ninevah University- Mosul – Iraq

ISSN 2310-5070 (online) ISSN 2311-004X (print)

problem, where the goal is to minimize the number of variables chosen while preserving the highest possible prediction accuracy [10, 24].

This paper targets the development of firefly algorithm for selecting variables within beta regression models. The proposed algorithm would successfully determine important variables in the beta regression model along with producing accurate predictions. Simulation tests along with real data analysis demonstrate that the proposed algorithm delivers its advantages.

## 2. Beta regression model

The beta regression model (BRM) finds extensive use across many fields especially economic and medical research including income share analysis and unemployment rate assessment in countries and regional Gini index calculations and university graduation metrics and medical body fat evaluation. In beta regression modeling we examine the relationship between certain explanatory variables and a non-normal response variable just like other GLM regression methods. “The beta regression model restricts its response component within the (0,1) interval to analyze proportions along with percentages and fractional data types.

In BRM, the response variable, $y$, is assumed to follow beta distribution. The probability density function of beta distribution is given by

$$f\left(y;\alpha_1,\alpha_2\right)=\frac{\Gamma(\alpha_2)}{\Gamma(\alpha_1\alpha_2)\Gamma((1-\alpha_1)\alpha_2)}\left(y\right)^{\alpha_1\alpha_2-1}\left(1-y\right)^{((1-\alpha_1)\alpha_2)-1},\quad 0<y<1, \tag{1}$$

where $0<\alpha_1<1$ and $\alpha_2>0$. The mean and variance are $E(y)=\alpha_1$ and $V(y)=\alpha_1(1-\alpha_1)/(1+\alpha_2)$ where $\alpha_2$ is a dispersion parameter.

Consider that we have a data set $\{(y_i,z_i)\}_{i=1}^{n}$ where $y_i\in\mathbb{R}$ is a response variable belongs to Eq. (1), $z_i=(z_{i1},z_{i2},...,z_{ip})\in\mathbb{R}^p$ is a $p\times 1$ known explanatory variable vector, then in BRM, the mean is related to the explanatory variables as

$$g(\alpha_{1i})={z_i}^T\boldsymbol{\beta}=\eta_i, \tag{2}$$

where $\boldsymbol{\beta}=(\beta_0,\beta_1,...,\beta_p)$ is a $(p+1)\times 1$ vector of unknown regression coefficients.
Depending on Ferrari and Cribari-Neto (2004) the

$$\begin{aligned}\ell(\boldsymbol{\beta},\theta)&=\textstyle\sum_{i=1}^{n}\ell_i(\alpha_{1i},\ \alpha_{2i})\\&=\ln\Gamma(\alpha_{2i})-\ln\Gamma((1-\alpha_{1i})\,\alpha_{2i})+(\alpha_{1i}\,\alpha_{2i}-1)\ln y_i\\&\quad+\{((1-\alpha_{1i})\,\alpha_{2i})-1\}\ln\left(1-y_i\right),\end{aligned} \tag{3}$$

where $\alpha_{1i}=g^{-1}(\eta_i)$ and $\alpha_{2i}=\mathrm{h}^{-1}(\vartheta_i)$. Differentiation of Eq. (3) with respect to the $\boldsymbol{\beta}$ and $\theta$ , respectively, is defined as

$$\mathrm{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\theta)=\sum_{i=1}^{n}\alpha_{2i}(\tilde{y}_i-\tilde{\alpha}_{1i})\,\frac{\mathrm{d}\alpha_{1i}}{\mathrm{d}\eta_i}\frac{\partial\eta_i}{\partial\beta_p}, \tag{4}$$

$$\mathrm{U}_{\theta}(\boldsymbol{\beta},\theta)=\sum_{i=1}^{n}\left\{\begin{array}{l}\theta_{1i}(\tilde{y}_i-\tilde{\alpha}_{1i})+\psi(\alpha_{2i})\ -\psi(1-\alpha_{1i})\,\alpha_{2i}\\+\ln(1-y_i)\end{array}\right\}\frac{\mathrm{d}\alpha_{2i}}{\mathrm{d}\vartheta_i}\frac{\partial\vartheta_i}{\partial\theta_k}, \tag{5}$$

where $\tilde{y}_i=\ln(y_i/(1-y_i))$, $\tilde{\alpha}_{1i}=\psi(\alpha_{1i}\alpha_{2i})\ -\psi((1-\alpha_{1i})\,\alpha_{2i})$, $\psi(.)$ represents the digamma function, $\mathrm{d}\alpha_{1i}/\mathrm{d}\eta_i=1/g'(\alpha_{1i})$, and $\mathrm{d}\alpha_{2i}/\mathrm{d}\vartheta_i=1/h'(\alpha_{2i})$[9].

## 3. Firefly Algorithm

The variable selection procedure serves as a common practice in diverse applications. The analysis typically contains two types of variables: redundant information and useless data. The increase in computational time and lowered prediction accuracy results from additional variables included in the analyses. The selection of significant variables from an entire variable set improves model prediction accuracy together with better interpretability results [6, 25]. Many algorithms inspired by nature have been put out recently as effective methods for resolving continuous optimization issues. An optimization problem involves minimizing the number of variables while increasing forecast accuracy [2, 10, 26, 27, 28]. The Firefly Algorithm (FA) functions as a nature-inspired metaheuristic optimization technique that demonstrates different benefits and drawbacks related to other swarm intelligence methods like Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) and Genetic Algorithm (GA).

Firefly optimization algorithm (FA) is one of the recently efficient proposed nature- inspired algorithms, which is firstly introduced by [12]. Compared to other algorithms, FA is a simple approach to use when solving optimization problems. FA draws inspiration from the way fireflies use flashing lights to interact with one another. In order to solve optimization difficulties, FA allows a swarm of fireflies with low light intensities to migrate toward their neighboring brighter fireflies with better search capabilities.

Let $q$ represents the dimension of the object function that will optimized, $N_f$ represents the number of fireflies, $\mu$ refers the light absorption coefficient, $I_i$ is the light intensity, and $\nu$ is the distance between any two firefly locations $i$ $(c_i)$ and $j$ $(c_j)$. This Cartesian distance can be defined as

$$v(c_i, c_j) = \sqrt{\sum_{m=1}^{q} (c_{i,m} - c_{j,m})^2}. \tag{6}$$

The $I_i$ can be as

$$I(\nu) = I_0 \, e^{-\mu\nu^2}, \tag{7}$$

The attractiveness $\xi$ of a firefly is defined as

$$\xi(\nu) = \xi_0 \, e^{-\mu\nu^2}, \tag{8}$$

where $\xi_0$ represents the attractiveness at $\nu = 0$. Any firefly that moves to its ideal location will be drawn to another firefly as

$$c_i^{(t+1)} = c_i^{(t)} + \xi_0 \, e^{-\mu\nu_{i,j}^2} \left( c_j^{(t)} - c_i^{(t)} \right) + \pi \, (\gamma_1 - 0.5), \tag{9}$$

where $\pi$ and $\gamma_1$ is a random number generated from uniform distribution with [0, 1].

To performing the variable selection, the position in FA is binary in which the value 1 represents that the variable is important and 0 otherwise. That is meaning: the $i^{th}$ variable is included in the model, then $\mathbf{x}_i = 1$, otherwise, $\mathbf{x}_i = 0$.

Consequently, our proposed method setting is as follows:
Step 1: The number of fireflies is $N_f = 30$, $\xi_0 = 1$, $\mu = 0.3$, $\pi = 0.2$, and the maximum number of iterations is $\mathrm{t}_{\max}$= 1000.
Step 2: The original binary firefly algorithm uses a uniform distribution with 0 and 1 to randomly create each firefly's position.
Step 3: The fitness function is defined as

$$\text{fitness} = \min \left[ \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right]. \tag{10}$$

Step 4: Using Eq.(9), the firefly' locations are updated.
Step 5: Steps 3 and 4 are repeated until a $t_{\max}$ is reached.

## 4. Simulation results

In this section, the performance of FA is evaluated. The sample size is considered with $n \in \{50, 150, 250\}$ and $y$is generated as

$$y_i \sim \text{beta}(\alpha_{1i}\,\alpha_{2i}, (1 - \alpha_{1i})\,\alpha_{2i}), \tag{11}$$

where $\alpha_{1i}$ and $\alpha_{2i}$ are generated as

$$\begin{aligned} \alpha_{1i} &= \tfrac{\exp(\mathbf{x}_i{}^T z)}{1+\exp(\mathbf{x}_i{}^T z)}, \\ \alpha_{2i} &= \tfrac{\exp(c_i{}^T \theta)}{1+\exp(c_i{}^T \theta)}, \end{aligned} \tag{12}$$

where the variables $\mathbf{x}_i$ and $\mathbf{s}_i$ are generated from the uniform distribution with 0 and 1. The true parameter vector $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are set as $\boldsymbol{\beta} = (1, 1, -0.5, 1.5, \underbrace{0, ..., 0}_{p-4})^T$ and $\theta = (1, 1, -1.5, 0.5, \underbrace{0, ..., 0}_{k-4})^T$. Here, two cases are considered:

**Case 1:** In this case $p = k = 10$.
**Case 2:** In this case $p = k = 25$.

The four performance metrics for FA include the Mean Squared Error (MSE) per Eq. (13) and the True Zero coefficient count (TZ) and the Incorrect Nonzero coefficient count (INZ) while also determining the correct estimation percentage (PC). A better variable selection performance relates to high PC and TZ metrics and low INZ and MSE values.

The performance of the FA was compared with following variable selection criteria as in[3]. They are:
The corrected Hannan and Quinn criterion (CHQ)

$$\text{CHQ} = -2(\hat{\boldsymbol{\beta}}, \hat{\theta}) + \frac{2n\,(p+k)\ln(\ln(n))}{n-(p+k)-1}. \tag{13}$$

The corrected Schwarz information criterion (CSIC)

$$\text{CSIC} = -2(\hat{\boldsymbol{\beta}}, \hat{\theta}) + \frac{n\,(p+k)\ln(n)}{n-(p+k)-1} \tag{14}$$

The corrected Akaike information criterion (CAIC)

$$\text{CAIC} = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\theta}) + \frac{2n\,(p+k)}{n-(p+k)-1} \tag{15}$$

Depending on 1000 times of generated the data, the averaged MSE, TZ, INZ, and PC are reported in Tables 1 and 2, respectively", for case 1 and case 2.

A few concluding observations can be derived from the presented tables. Every experimental condition demonstrates that the FA reaches lower MSE results than the other competing procedures including CAIC, CHQ, and CSIC. The MSE level of CAIC remains the highest in comparison to other methods. The MSE reduction from FA reached 42.69% and 37.85% and 51.11% when compared against CHQ, CSIC and CAIC when n=150. The FA maintains the lowest MSE value compared to competing methodologies for all considered n values. Sample size increases improve both MSE performance and general effectiveness for the CAIC and CHQ methods as well as for the CSIC method.

With regard to the TZ criterion FA demonstrates an optimal precision in identifying actual nonzero in both components of the sub-model mean and sub-model dispersion. The FA process picked more than 7 essential variables from among the 8 actual variables in case 1 before picking more than 20 out of 22 actual variables in case 2. The FA produces superior performance when compared to CAIC, CHQ and the CSIC according to TZ criterion. The designed research used FA to identify an average of 21 substantial variables from a pool of 22 essential variables based on the dataset with n=150. While CAIC, CHQ, and CSIC select no more than 18 relevant variables. Further, the FA procedure selects extremely few unimportant variables from the mean sub-model and dispersion sub-model compared to CAIC, CHQ, and CBIC when examining the INZ criterion since it correctly sets to zero only a low number of truly nonzero coefficients on average.

### *4.1. real application results*

In this section, a real application is considered for testing our proposed method to a data from a body fat study, which had been analyzed by Zhao et al. (2014). In this data, there are 252 observations for body fat patients on

Table 1. Case 1 results for the used methods

| Methods | MSE | TZ | INZ | PC |
|---|---|---|---|---|
| | $n = 50$ | | | |
| **FA** | **4.392** | **7.598** | **0.487** | **0.955** |
| CHQ | 7.626 | 5.073 | 2.945 | 0.823 |
| CSIC | 6.982 | 6.022 | 2.139 | 0.88 |
| CAIC | 8.232 | 4.183 | 3.314 | 0.722 |
| | $n = 150$ | | | |
| **FA** | **4.273** | **7.628** | **0.293** | **0.978** |
| CHQ | 7.456 | 5.186 | 3.001 | 0.829 |
| CSIC | 6.722 | 5.742 | 2.205 | 0.898 |
| CAIC | 8.745 | 4.477 | 3.27 | 0.715 |
| | $n = 250$ | | | |
| **FA** | **4.236** | **7.682** | **0.368** | **0.982** |
| CHQ | 7.417 | 5.283 | 3.576 | 0.835 |
| CSIC | 6.876 | 6.086 | 1.785 | 0.906 |
| CAIC | 8.596 | 4.544 | 3.245 | 0.724 |

Table 2. Case 2 results for the used methods

| Methods | MSE | TZ | INZ | PC |
|---|---|---|---|---|
| | $n = 50$ | | | |
| **FA** | **6.672** | **21.786** | **0.068** | **0.959** |
| CHQ | 9.657 | 16.597 | 2.826 | 0.817 |
| CSIC | 9.219 | 17.172 | 1.14 | 0.875 |
| CAIC | 10.949 | 14.582 | 3.16 | 0.806 |
| | $n = 150$ | | | |
| **FA** | **6.305** | **21.005** | **0.179** | **0.951** |
| CHQ | 9.476 | 16.859 | 2.579 | 0.822 |
| CSIC | 8.948 | 18.23 | 1.166 | 0.895 |
| CAIC | 10.565 | 15.646 | 3.12 | 0.801 |
| | $n = 250$ | | | |
| **FA** | **6.23** | **21.365** | **0.142** | **0.967** |
| CHQ | 9.413 | 16.87 | 2.692 | 0.826 |
| CSIC | 8.87 | 18.941 | 1.468 | 0.907 |
| CAIC | 10.59 | 15.655 | 3.258 | 0.814 |

13 explanatory variables, of which the y is a quantitative measurement of the percentage of body fat. The 13 explanatory variables include age (years) (x1); weight (pounds) (x2); height (inches) (x3); neck circumference (cm) (x4); chest circumference (cm) (x5); abdomen circumference (cm) (x6); hip circumference (cm) (x7); thigh circumference (cm) (x8); knee circumference (cm) (x9); ankle circumference (cm) (x10); extended biceps circumference (x11); forearm circumference (cm) (x12) and wrist circumference (cm) (x13).

Related to Zhao et al. (2014), The beta distribution indicates appropriateness for BRM analysis when using the logit link function for the mean sub-model combined with the identity link function for dispersion sub-models.

The mean sub-model demonstrates significant connections between five main explanatory variables, x2, x6, x7, x12, and x13, with response variables when removing the three outlier observations. The used methods yield their results regarding performance as presented in Table 3. The mean model variables receive appropriate selection from FA as per Table 3 findings while x6 fails to fulfill these criteria. The FA produces higher prediction accuracy through MSE reduction than CAIC and CHQ and CSIC. The result shows that CAIC and CHQ and CSIC selected variables which did not achieve statistical significance. CAIC selected the variables x4 and x9 while these variables did not appear as statistically significant.

The important variables x2 and x7 appear in all selections made by the utilized methods.

Table 3. The results of the application data

| Methods | Selected variables | MSE |
|---|---|---|
| FA | X2, x12, x13 | 248.521 |
| CSIC | X2, x4, x7, x13 | 362.114 |
| CHQ | X2, x4, x7, x12 | 418.749 |
| CAIC | X2, x4, x7, x9, x12 | 463.109 |

## 5. Conclusion

The BRM serves as an effective instrument to analyze continuous data between specific boundaries since it outperforms traditional methods by offering better flexibility along with improved interpretability. Compared to other algorithms, the FA, a metaheuristic optimization technique inspired by nature, offers unique benefits and drawbacks. In this work, an assessment of variable selection problems in BRM takes place within this research. A FA operated as the variable selection method. Both simulation tests and real-data evaluation are performed for the FA and alternative approach selection techniques. The research outcomes demonstrate that FA accomplishes superior performance over CAIC, CHQ and CSIC based on MSE scores together with TZ, INZ and PC measurements.

REFERENCES

1. Z. Y.Algamal, *A particle swarm optimization method for variable selection in beta regression model*, Electronic Journal of Applied Statistical Analysis, 12(2), 2019. .
2. Ş.Ay , E.Ekinci , and Z.Garip , *A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases*, The Journal of Supercomputing, 79(11), pp. 11797-11826, 2023.
3. F. M.Bayer , and F.Cribari-Neto , *Model selection criteria in beta regression with varying dispersion*, Communications in Statistics - Simulation and Computation, 46(1), pp. 729-746. https://doi.org/10.1080/03610918.2014.977918 , 2015.
4. S.Ferrari , and F.Cribari-Neto , *Beta regression for modelling rates and proportions*, Journal of Applied Statistics, 31(7), pp. 799-815. https://doi.org/10.1080/0266476042000214501 ,2004.
5. Y. S.Maluf, S. L. Ferrari, , and F. F.Queiroz , *Robust beta regression through the logit transformation*, Metrika, 88(1), pp. 61-81,2025.
6. E. O. Ogundimu, , and G. S.Collins, *Predictive performance of penalized beta regression model for continuous bounded outcomes*, Journal of Applied Statistics, 45(6), pp. 1030-1040 ,2018.
7. R.Ospina , and S. L. P.Ferrari , *A general class of zero-or-one inflated beta regression models*, Computational Statistics and Data Analysis, 56(6), pp. 1609-1623, 2012. https://doi.org/10.1016/j.csda.2011.10.005
8. Z.Sadeghian , E.Akbari , H.Nematzadeh , and H.Motameni , *A review of feature selection methods based on meta-heuristic algorithms*, Journal of Experimental and Theoretical Artificial Intelligence, 37(1), pp. 1-51 , 2025.
9. A. B.Simas , W.Barreto-Souza , and A. V.Rocha, *Improved estimators for a general class of beta regression models*, Computational Statistics and Data Analysis, 54(2), pp. 348-366 , 2010 . https://doi.org/10.1016/j.csda.2009.08.017
10. R.Sindhu , R.Ngadiran, Y. M.Yacob , N. A. H.Zahri, and M.Hariharan, *Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism*, Neural Computing and Applications, 28(10), 2947-2958, 2017. https://doi.org/10.1007/s00521-017-2837-7

11. Z.Wang, S.Ma, M.Zappitelli, C. Parikh, C. Y.Wang, and P. Devarajan, *Penalized count data regression with application to hospital stay after pediatric cardiac surgery*, Stat. Meth. Med. Res., In press , 2014. https://doi.org/10.1177/0962280214530608.
12. X.-S.Yang, *Multiobjective firefly algorithm for continuous optimization*, Engineering with Computers, 29(2), pp.175-184 , 2013.
13. W.Zhao, R.Zhang, Y.Lv, and J.Liu, *Variable selection for varying dispersion beta regression model*, Journal of Applied Statistics, 41(1), pp. 95-108, 2014.
14. Algamal, Z. Y., Asar, Y. *Liu-type estimator for the gamma regression model* , Communications in Statistics-Simulation and Computation, vol. 49, no . 8, p. 2035-2048, 2020.
15. Algamal, Z. Y., Lee, M. H. *A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives* , SAR and QSAR in Environmental Research, vol. 28, no. 1 , p. 75-90, 2017.
16. Kahya, M. A., Altamir, S. A., Algamal, Z. Y. *Improving whale optimization algorithm for feature selection with a time-varying transfer function* , Numerical Algebra, Control and Optimization, vol. 11, no. 1 , p. 87-98 , 2020.
17. Algamal, Z. Y., Qasim, M. K., Ali, H. T. M. *A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine* , SAR and QSAR in Environmental Research, vol. 28 , no. 5, p. 415-426 , 2017.
18. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M. *High-dimensional quantitative structure–activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression* , Journal of Chemometrics, vol. 30 , no.2 m p. 50-57 , 2016.
19. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., Aziz, M. *High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty* , Journal of Chemometrics, vol. 31 , p.6, p. e2889 , 2017.
20. Algamal, Z. Y., Qasim, M. K., Lee, M. H., Ali, H. T. M. *High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm* , Chemometrics and Intelligent Laboratory Systems, vol. 206, p. 104170 , 2020.
21. Ismael, O. M., Qasim, O. S., Algamal, Z. Y. *Improving Harris hawks optimization algorithm for hyperparameters estimation and feature selection in v-support vector regression based on opposition-based learning* , Journal of Chemometrics, vol 34 , no. 11, e3311, 2020.
22. Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., Taha, I. M. *A new two-parameter estimator for beta regression model: method, simulation, and application* , Frontiers in Applied Mathematics and Statistics, vol. 7,p. 780322 ,2022.
23. Algamal, Z. Y., Abonazel, M. R. *Developing a Liu-type estimator in beta regression model* , Concurrency and Computation: Practice and Experience, vol.34 ,no. 5 , p. e6685.
24. Algamal, Z., Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256 , 2017.
25. Salih, A. M., Algamal, Z., Khaleel, M. A. *A new ridge-type estimator for the gamma regression model* , Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98.
26. Alkhateeb, A., Algamal, Z. *). Jackknifed Liu-type estimator in Poisson regression model* , Journal of the Iranian Statistical Society,Vol .11 ,no . 1, p. 21-37, 2022.
27. Mahmood, S. W., Basheer, G. T., Algamal, Z. Y. *Quantitative Structure–Activity Relationship Modeling Based on Improving Kernel Ridge Regression* , Journal of Chemometrics, vol. 39, no. 5, p. e70027, 2025.
28. Mahmood, S. W., Basheer, G. T., Algamal, Z. Y. *Improving kernel ridge regression for medical data classification based on meta-heuristic algorithms* , Kuwait Journal of Science, vol. 52, no. 3 , p. 100408 , 2025