Journal of Economics and Administrative Sciences Vol.29 (NO. 135) 2023, pp. 56-67

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison of Some Methods for Estimating Nonparametric Binary Logistic Regression

Narjes Bassem Khalaf (1)
University of Baghdad/ College of Administration and Economics/ Baghdad, Iraq
narjes.bassem1201a@coadec.uobaghdad.edu.iq

Lekaa Ali Mohammed(2)
University of Baghdad/ College of Administration and Economics/ Baghdad, Iraq
lekaa.a@coadec.uobaghdad.edu.iq

Received: 28/2/2023 **Accepted:19/3/2023** **Published: March / 2023**

Abstract:

In this research, the methods of Kernel estimator (nonparametric density estimator) were relied upon in estimating the two-response logistic regression, where the comparison was used between the method of Nadaraya-Watson and the method of Local Scoring algorithm, and optimal Smoothing parameter λ was estimated by the methods of Cross validation and generalized Cross validation, bandwidth optimal λ has a clear effect in the estimation process. It also has a key role in smoothing the curve as it approaches the real curve, and that the goal of using the Kernel estimator is to modify the observations so that we can obtain estimators with characteristics close to the properties of real parameters, and based on medical data for patients with chronic lymphocytic leukemia and through the use of the Gaussian function and based on the comparison criterion (MSE) it was found that the Nadaraya -Watson method is the best because it obtained the lowest value for this criterion.

Keywords: Binary logistic regression model, Nadaraya–Watson method, Local Scoring algorithm, Cross validation method, generalized Cross validation, Plug in method.

(1)
(2)

1. Introduction:

Regression analysis is a branch of statistics that examines the relationship between the response variable (Y) and the explanatory variables $(X_1, \ldots, X_k)$ for a set of data, equation that is used to estimate and predict the cases to be studied. Regression is divided into two types, the first is linear and the second is not linear, and in this research, the second type will be dealt with, i.e. the nonlinear regression model, represented by the logistic regression model, and its importance is through its representation of qualitative phenomena within the medical and social fields,…etc, and many studies dealt with logistic regression and non-parametric methods:

Berkson (1944) tested the hypothesis that (that the response of people to a certain dose of a drug follows the normal distribution where he used the function of the natural curve and the logistics function on a set of data after converting it to a linear function through the transformation logit, he concluded that the logistic function is better, especially in the case of binary data and when the sample size is large. Hmood (2000) Made comparisons among three estimators (Nadaraya- Watson, Gasser-Muller and Local Liner Smother). Hmood (2005) presented non-parametric and semi-parametric methods with several proposed methods for estimating the probability density, and also reviewed the methods for estimating the Bandwidth. Hamza (2009, 2012) presented a review of some the methods of the Kernel estimate the state of regression non-parametric either in the case of complete data (FNW, FLLS, VNW, VLLS) and incomplete (FSNW, VSNW, FINW, VINW, FSLLS, SILLS, VILLS) and comparing them through simulation as well as using different models, sample sizes and variations. Welin (2013) introduced a nonlinear logistic regression model for classification, the main idea is to assign data based on the kernel density estimation space feature, then a education model is taught to improve the feature weights in addition to Bandwidth for the Nadaraya-Watson kernel density estimator. Rafada et al (2018) developed the risk model from type 2 diabetes in the disease General Hajj Hospital Surabaya using a regression model Nonparametric logistic by Local scoring algorithm.

2. Materials and Methods

2-1 Logistic regression model :

The Logistic regression model is based on a basic and important assumption that dependent variable is a two-response variable that follows the Bernoulli distribution assumed value (1) with probability of (success) and value (0) with probability of (failure), i.e. the occurrence of the response and its occurrence and non-occurrence, and the model can be written as follows (Al-Saffar, 2013):

$$\text{Pi} = \frac{e^{B0+\sum Bjxij}}{1+e^{B0+\sum Bjxij}} \qquad (1)$$

$$Pi = (1 - qi)$$

$$\text{qi} = \frac{1}{1 + e^{B0+\sum Bjxij}} \qquad (2)$$

i=1,2,3…,n sample observation

k= 1,2,3,…, : Number of explanatory variables.

B_0 : Intercept

B_j : j Slope parameters of explanatory variables.

If the logistic regression model is one of the regression models in which the relationship between the response variable and the explanatory variables is non-linear and often takes the form of the letter S. The logistic function can be assigned to a linear function as follows (Al-Azzawi, 2005):

$$Zi = ln(pi/(1 - pi)) \qquad (3)$$

the following formula:

$$Zi = B0 + B1Xi \tag{4}$$

i=1,2,…,n [pi/(1-pi)] The logarithm will be taken for the ratio

2-2 Methods of estimating the logistic regression model:

2-2-1 Nadarya- Watson smoothing

Nadarya-Watson estimator is one of the most important estimators used to estimate the nonparametric regression function using the average weights for the observations, as it was proposed by the researchers (Nadarya- Watson) and the weight function is calculated as follows (Hassan and Nabi, 2018):

$$\hat{f}_{NW}(x) = \frac{\sum_{i=1}^{n} k\left(\frac{(x-xi)}{h}\right) Y_i}{\sum_{i=1}^{n} k\left(\frac{(x-xi)}{h}\right)} \tag{5}$$

$$\hat{f}_{NW}(x) = \sum_{i=1}^{n} w_{hi}(x) Y_i \tag{6}$$

Equation (6) can be reformulated in the form of matrices as follows (Demir and Toktamis, 2010):

$$\hat{f}_{NW}(x) = WY \tag{7}$$

So that :

$W = (\acute{w}_1, \acute{w}_2, \dots, \acute{w}_n)$; $Y = (Y_1, Y_2, \dots, Y_n)$

When choosing Bandwidth must be select carefully, so that if it was too large, it will affect the smoothing curve so we have a high smoothing curve (over smoothing curve) and if the value is too small it will effect on the smoothing curve and we have a low smoothing curve (under smoothing curve). The cross validation method is one of the commonly used methods to select bandwidth if this parameter is estimated through the use of the (NW) method with that depend on exclusion one observation, so that this method called Leave-one-out· (Hmood, 2000).

2-2-2 Local Scoring Algorithm

The binary nonparametric regression model has a categorical binary response variable (0,1) and a Bernoulli distribution. We estimate the nonparametric binary regression model based on the Kernel estimator using the Local scoring algorithm because this algorithm is very suitable in its use of the nonparametric logistic regression model and that the response variable follows the Bernoulli distribution and belongs to the exponential family, This algorithm consists of two loops that are Scoring step (outer loop) is iterated until the average value of deviance convergent and weighted Back fitting step (inner loop) is iterated until the average value of the Residual Sum of Squares (RSS) convergent. scoring step is done iteratively by determining the adjusted value of the response variable (z) which is formulated as follows (Rifada, 2018):

$$z_i = m_i^{(s)} + \left(y_i - \pi_i{}^{(s)}\right)\left(\frac{\partial \pi_i}{\partial m_i}\right)_{(s)} \; ; i = 1,2,\dots,n \tag{8}$$

And :

$$m_i^{(s)} = \sum_{j=1}^{p} \hat{f}_j^{(s)}(x_{ji}) = \ln\left[\frac{\pi_i^{(s)}}{1-\pi_i^{(s)}}\right] \quad ; s = 0,1,2,\dots,n \tag{9}$$

The steps of the algorithm can be displayed where the algorithm is divided into three sub-algorithms: The first sub-algorithm: An algorithm to determine the value of the optimal bandwidth for each explanatory variable and its steps are as follows:

Entering observations (response variable and explanatory variables) adopted in the logistic regression model, i.e .$(y_i, x_{1i,} x_{2i,} \ldots \ldots x_{ji})$ $i = 1,2, \ldots, n$

Enter Kernel function (Gaussian) function based on the following step:

We assume an initial value for the bandwidth (λ_j) .

The diagonal weights matrix is calculated $W_{ji}\left(\lambda_j\right)$، where

$$W_{ji}(\lambda_j) = diag\left[K_{\lambda_j}(x_{ji} - x_{j1}), K_{\lambda_j}(x_{ji} - x_{j2}), \ldots, K_{\lambda_j}(x_{ji} - x_{jn})\right] \quad (10)$$

a)We determine a matrix $A(\lambda_j)$ ، where

$$A(\lambda_j) = \begin{bmatrix} \frac{L'W_{j1}(\lambda_j)}{L'W_{j1}(\lambda_j)L} \\ \frac{L'W_{j2}(\lambda_j)}{L'W_{j2}(\lambda_j)L} \\ \vdots \\ \frac{L'W_{jn}(\lambda_j)}{L'W_{jn}(\lambda_j)L} \end{bmatrix} \quad ; L = (1,1, \ldots,1)' \qquad (11)$$

b) value is calculated $\widehat{f}_j(x_j) = A(\lambda_j)Y$وان ،

$$\hat{f}_j(x_j) = \begin{bmatrix} \hat{f}_j(x_{j1}) \\ \hat{f}_j(x_{j2}) \\ \vdots \\ \hat{f}_j(x_{jn}) \end{bmatrix} \quad ; Y = (y_1, y_2, \ldots, y_n)' \qquad (12)$$

c)value is calculated $MSE(\lambda_j) = n^{-1} \sum_{i=1}^{n}(y_{i-}\widehat{f}_{\,j}(x_{ji}))^2$ ، where the lowest value $MSE(\lambda_j)$ It is the one that denotes the optimal bandwidth$(\lambda_j)$and that:

$$\hat{f}_j(x_{ji}) \frac{L'W_{ji}(\lambda_j)Y}{L'W_{ji}(\lambda_j)L} \qquad (13)$$

d) value is calculated $GCV = \frac{MSE(\lambda_j)}{\frac{1}{n}t_r\,[I-A(\lambda_j)])^2}$

e)Repeat steps from (b) to (e) to get the lowest value لـ $GCV(\lambda_j)$ and represents the optimal bandwidth (bandwidth (λ_j))

a-The second sub-algorithm: is the initial estimation algorithm for the nonparametric regression function$\widehat{f}_j(x_j)$ For all explanatory variables as follows (Rifada et al, 2018) :

i. Input pair data $(y_i, x_{1j} \ldots x_{ij})$, i=1,2,…,n
ii. Defining the kernel function used is Gaussian kernel function.
iii. We enter the optimal bandwidth obtained from the first sub algorithm.
iv. We determine the matrix of diagonal weights$W_{ji}(\lambda_j)$
v. Define the matrix $A(\lambda_j)$
vi. Calculate the value $\widehat{f}_{\,j}(x_i) = A(\lambda_j)Y$

b- Sub-Algorithm III: Positional Scoring Algorithm for Estimating the Binary Nonparametric Model.

i. Enter the data in an ordered pairs $(y_i, x_{1i}, \ldots x_{ji})$; $i = 1,2, \ldots., n$
ii. Entering the initial estimates for x_j $(\widehat{f}_{\,j}^{\,(0)})$ where $j = 1,2, \ldots., p$ Optimized for the second sub-algorithm
iii. Repeat the scoring steps, (outer ring) as follows:

a. Determining the adjusted value of the response variable (z) with elements of row i-th is:

$$Z_i = m_i^{(s)} + \frac{(y_i - \pi_i{}^{(s)})}{\pi_i{}^{(s)}(1-\pi_i{}^{(s)})} \quad , i = 1,2, \dots , n \qquad (14)$$

With:

$$\pi_i^{(s)} = \frac{\exp(m_i^{(s)})}{1+\exp((m_i^{(s)})} \quad , m_i^{(s)} = \sum_{j=1}^{p} \hat{f}_j^{(s)}(x_{ji})$$

and determine the matrix of weights B, where B is a diagonal matrix (i.e. its elements represent the main diagonal), and it is obtained as follows:

$$b_i = \left(\frac{\partial \pi_i}{\partial m_i}\right)_{(s)}^2 \left(V_i^{(s)}\right)^{-1} \; ; i = 1,2, \dots , n \qquad (15)$$

Since the response variable Y follows Bernoulli's distribution:

$$V_i = Var(Y_i) = \pi_i(1-\pi_i) \qquad (16)$$

Thus, the diagonal matrix for B is:

$$b_i = \pi_i^{(s)} \left(1-\pi_i^{(s)}\right) \quad , i = 1,2, \dots . , n \qquad (17)$$

b. Repeating the weights of the step of reversing, (inner ring) as follows

i. Initial iteration (s=0) and defines:

$\hat{m}_j^{(s)}(X_j) = \hat{m}_j^{(r)}(X_j) \; and \; Z = (z_1, z_2, \dots z_n)^T$

ii. Estimation of nonparametric regression functions in the model- $j = 1,2, \dots . , p$ which:

$$\hat{f}_j^{(s+1)}(X_j) = A(\lambda_j)\left\{z - \sum_{k=1}^{j-1} \hat{f}_k^{(s+1)}(X_k) - \sum_{k=j+1}^{p} \hat{f}_k^{(s)}(X_k)\right\} \quad (18)$$

iii. calculate the average value of the square of the sum of weighted remainder RSS:

$$Avg\,(Rss)^{(s+1)} = \frac{1}{n}\left\{\left(z^{(s+1)} - m^{(s+1)}\right)^T B^{(s+1)}\left(z^{(s+1)} - m^{(s+1)}\right)\right\} \quad (19)$$

iv. Repeat steps (ii) to (iii) when S=S+1 is below the convergent RSS value, i.e.

$$abs(Avg\,(RSS))^{(s+1)} - Avg(RSS)^{(s)} < \varepsilon\,, \; for\; \varepsilon = 0.0001 \qquad (20)$$

c. Is defined $\hat{f}_j^{(r+1)}(x_j) = \hat{f}_j^{(s+1)}(x_j)$ Hence, we determine the value of the rate of deviation, that is:

$$Avg(D\;\;(y_i;\pi_i))^{(r+1)} \approx \frac{-2}{n}\sum_{i=1}^{n} y_i \ln \pi_i{}^{(r+1)} + (1-y_i)\ln\left(1-\pi_i{}^{(r+1)}\right) \quad (21)$$

d. Repeat steps (a) to (c) when r=r+1 to get the average value of the convergent deviation, i.e.

$$abs(Avg\left(D(y_{i;}\pi_i\right)^{(r+1)} - Avg\left(D(y_{i;}\pi_i\right)^{(r)} < \varepsilon \; with \; \varepsilon = 0.0001 \qquad (22)$$

3. Comparison Criterion:

Estimation methods are compared by the mean squares of error (MSE) of the model:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} \qquad (23)$$

3-1 Simulation

The concept of analysis using simulation is a natural and logical extension of analytical and mathematical models, where there are many cases that cannot be represented mathematically and the reason may be the random nature of the studied problem or because of its complexity or due to the lack of description of the problem, the method that is often used when all methods fail to obtain data is the method of simulation, the recent progress and the availability of software and technical developments in the field of programming and computer made the simulation method to achieve an analytical solution to the problem What through the mathematical formulation of data (Shehab, 2017).

3-1-1 Description of the stages of the simulation experiment

The first stage:

It is an important stage, which is the stage of determining the default values, on which the rest of the stages depend, as these values were chosen as follows.

i. Assume the required sample size n=(50,100,150)

ii.Assume standard deviation values σ=(0.1,1,3)

iii. Assuming Kernel function (Gaussian)

iv. Determine the Replication of the simulation experiment (500) times

The second stage:

At this stage, variables are generated as follows.

i. Generate p from illustrative variables where p=(4,6,8) follows the normal distribution with mean (0) and variance (σ) using the (Monte Carlo method)

ii.The variables generated in the above steps are linked according to the following formula and according to the number of variables included in the model: The model that is being worked on:

$$f_{(x)} = 1 - 48\text{x} + 18x^2 - 15x^3 + 45x^4 - x^5 \qquad (24)$$

The p-value is calculated $\pi(x_i)$ as:

$\pi(x_i) = 1/(1 + \exp(-f_{(x)})$

Where

$f_{(x)}$ (Equal to model (24) above based on the number of variables included in the model)

iii. Generate values of the dependent variable y according to a binomial distribution with a sample size equal to n

The third stage:

At this stage, the features are estimated as follows:

Estimating the above models according to the following methods:

- Nadarya- Watson smoothing (NW).
- Algorithm Local Scoring.

i. For the estimate the preamble parameter Three methods have been used to estimate it:

1. Estimate bandwidth by using three methods.
2. Cross validation.
3. Generalized Cross Validation.

ii.Comparison of methods to determine the best method based on the MASE criterion.

3-1-2 Results of the simulation experiment

The results will be analyzed and interpreted by comparing the methods with the (MASE) Comparison Criterion with different sample sizes, standard deviations and values of bandwidth as follows:

Table (1) shows MASE for Compare results of the logistic regression function for n= 50, 100, 150 and σ=0.1, 1, 3 with CV method

	n=50		**n=100**		**n=150**	
σ	**LSA**	**NW**	**LSA**	**NW**	**LSA**	**NW**
0.1	0.5158972	0.2678695	0.4963892	0.2713597	0.4946409	0.2698858
1	0.5018234	0.3101162	0.4904191	0.3203379	0.4810336	0.3259145
3	0.4945971	0.3838615	0.4845277	0.3989768	0.4804463	0.4102643

From Table (1) it is clear that**:**

− At the sample size (50,100,150) we note that the best method with different standard deviations is NW method **.**

− We notice that the value of (MASE) increases when the value of standard deviations increases with different sample sizes

Size 150 NW	Size 100 NW	Size 50 NW

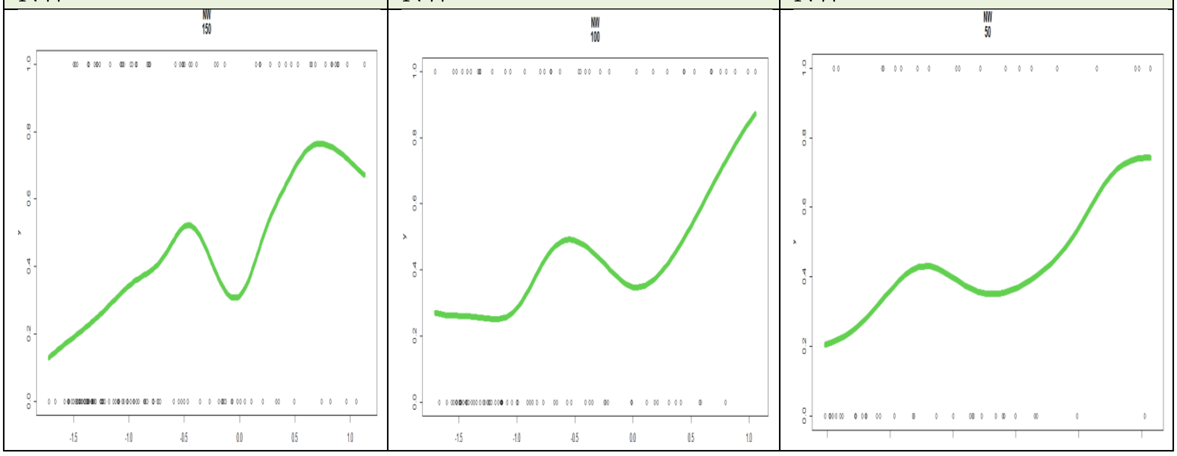


Figure (1) estimated values of the dependent variable Y CV-method estimation for bandwidth with n=50,100,150 and σ=0.1

Table (2) shows the MASE criterion for comparing estimates of the logistic regression function for sample sizes n=150 n=100, n=50, σ=0.1, σ=1,σ=3 and GCV bandwidth

	n=50		n=100		n=150	
σ	LSA	NW	LSA	NW	LSA	NW
0.1	0.5171557	0.2799637	0.4880042	0.2757808	0.4888043	0.2783195
1	0.5101642	0.3107641	0.4826771	0.3208041	0.4769862	0.3282777
3	0.4885546	0.3875923	0.4856675	0.4047017	0.4779822	0.4100840

From Table (2) it is clear that**:**

-At the sample sizes (50,100,150), we notice the superiority of the NW method specifically at $\sigma = 0.1$

-We note that the value of (MASE) increases when the value of standard deviations increases and according to the values of the sample sizes**.**

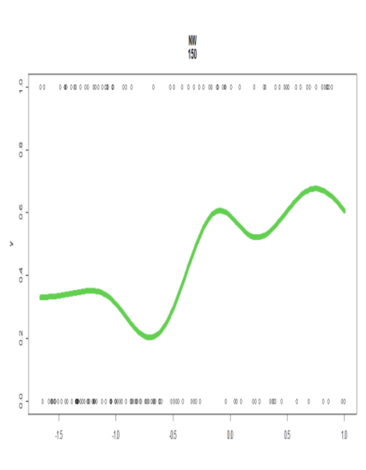
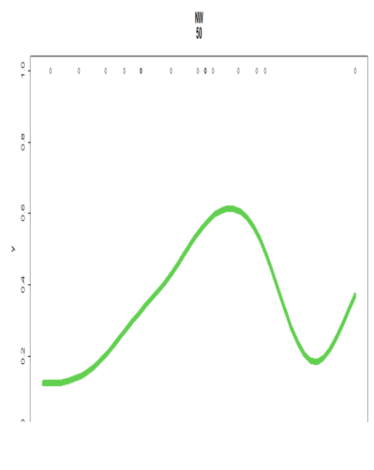
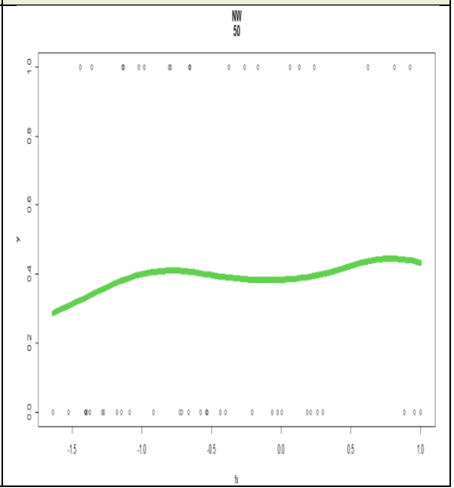
Size 150 NW	Size 100 NW	Size 50 NW
		

Figure (2) Estimated values of the dependent variable Y using (GCV), sample sizes (50, 100, 150) and σ=0.1

Table (3) shows the MASE criterion for comparing estimates for the logistic regression function for sample sizes n=150 ,100, 50, and σ=0.1,1, 3 With the plug in

	n=50		n=100		n=150	
Σ	LSA	NW	LSA	NW	LSA	NW
0.1	0.5167924	0.2684418	0.5065730	0.2710181	0.4808937	0.2697185
1	0.5067308	0.3097728	0.4896204	0.3197671	0.4838743	0.3306209
3	0.4946868	0.3907838	0.4793816	0.4052036	0.4783190	0.4124434

From Table (3) it is clear that**:**

− At the sample sizes (50, 100, 150), we Note the superiority of the NW method at $\sigma = 0.1$

-We notice that the value of (MASE) increases when the value of standard deviation increases with different sample sizes .

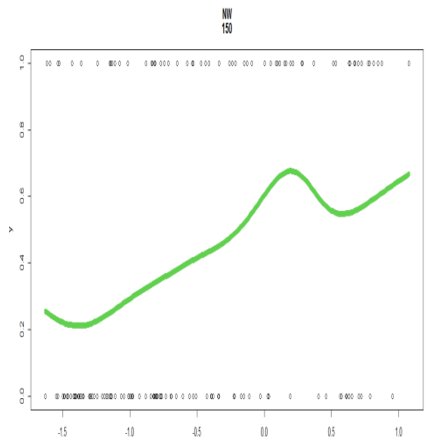
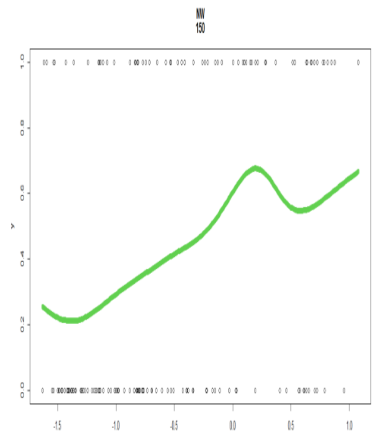
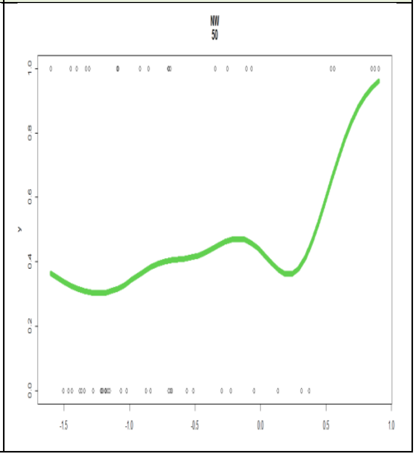
Size 150 NW	Size 100 NW	Size 50 NW
		

Figure (3) Estimated values of the dependent variable Y for the best method using the value of the bandwidth (plug in), n=50,100,150 and $\sigma = 0.1$

4. Real Practical Example

Lymphocytic leukemia is one of the types of lymphomas that are of low degree of malignancy and differ from the types of tumors that the carcinogenic cells in this disease originate in the bone marrow and that they lose the elements that bind them to the bone marrow and thus move from the bone marrow to the blood circulation and appear in the blood and the fact that this disease can coexist with a person for a period of 5-15 years with health problems that require treatment only in the last years of the disease. There are many factors affecting the reduction of chronic lymphocytic leukemia diseases Most medical studies that dealt with chronic lymphocytic leukemia mentioned several factors that affect the incidence of the disease as well as the variables that have been proposed, as follows: Y (response variable): It is an approved variable that takes only two values: (0) if the person has the disease, and (1) if the person does not have the disease. The explanatory variables are:

X_1 : represents gender (Sex) where (1) stands for males and (2) for females.
X_2 : (age).
X_3 : (WBC)White Blood Cells.
X_4 : (HGB) Hemoglobin Blood
.X_5 : (HCT) Haematocrit
X_6 : (PLT) Blood Platelets .

The method of data collection was done by relying on the patient file (infected and non-injured), which included a study on the disease (chronic lymphocytic leukemia) at Diwaniyah Teaching Hospital in Al-Qadisiyah Governorate (Oncology Consulting Unit) The study included 100 male and female views (infected and non-injured) and the observations were divided into two groups as follows:

The first group: It included people with the disease in the size of (53) samples and we coded them with the symbol (0).
The second group: It included people who are not infected with the disease with a size of (47) samples and we symbolized them with the symbol (1).

5. Discussion of Results

Before starting the estimation process, a test (VIF) was conducted to test the existence of a linear correlation problem between the independent variables, as it was found that there is no linear correlation between the independent variables, as all the illustrative variables had a VIF value less than 5, as shown in the table below**:**

Table (4) VIF Test

variable	SEX	Age	WBC	HGB	HCT	PLT
VIF	1.164	1.484	1.128	4.57	4.11	1.049

Table (5)
Shows the values of the mean square Error (MASE) for the incidence rate of lymphocytic leukemia

Estimation Method	NW	LSA
MSE	0.2962321	0.3750245

mean square error (MSE) values calculated where the results showed that the NW method is the best method

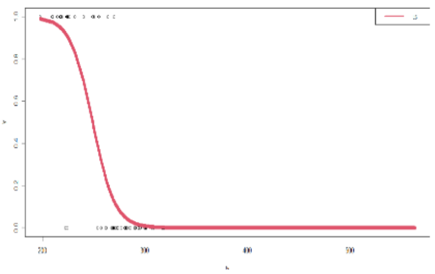
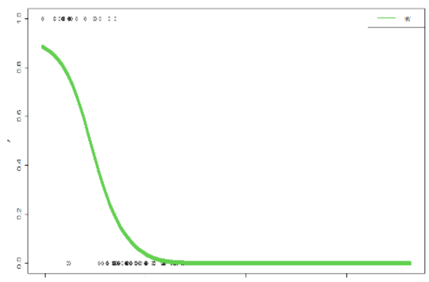
Estimated values	Real data
	

Figure (4) shows the real and estimated data of the dependent variable y (injury rate) not the preference of the estimation methods in the above

6. Conclusions

In this research, we conclude that the nonparametric methods of the Kernel methods in estimating the binary logistic regression, and a comparison was made between method Nadarya - Watson and method the Local scoring algorithm, and the optimal λ bandwidth was estimated using the method of legitimate crossing and generalized cross validation (optimal), and based on medical data with chronic lymphocytic leukemia and through the use of the function (Gaussian) and based on the comparison criterion (MSE), it was found that the Nadarya - Watson method is better than the Local scoring algorithm at sample sizes 50,100,150 and variance values of 0.1, 1, 3 where the lowest value was obtained for this criterion and when applying the real data of lymphocytic leukemia and through the comparison criterion, the NW method was the best.

References

1.Abdul Hassan, Maysam Abdul Nabi, (2018), "Comparison of nonparametric estimates for multiple regression analysis with practical application", Master Thesis, Department of Statistics, College of Administration and Economics, University of Baghdad.

2.Al-Azzawi, Ahmed Diab Ahmed, (2005), "Comparison between Some Methods of Estimating the Logistic Regression Model and the Fortified Methods of Bilateral-Response Life Experiences Using the Simulation Method" Master's Thesis, College of Administration and Economics, University of Baghdad.

3.Al-Saffar, Roaa Saleh Mohammed, (2013), "Nonparametric and modified methods in estimating the reliability function of complete data with practical application", PhD thesis, Department of Applied Statistics, College of Administration and Economics, Al-Mustansiriya University.

4.Demir, S. and Toktamiş, Ö., 2010. On the adaptive Nadaraya-Watson kernel regression estimators. Hacettepe Journal of Mathematics and Statistics, vol 39 no 3, pp.429-437.

5.Hmood, Munaf Yousif, (2000), " Comparing Nonparametric Kernel estimators for Estimating Regression Function", Master thesis, Department of Statistics, College of Administration and Economics, University of Baghdad.

6.Hmood, Munaf Yousif, (2005), " Comparing Nonparametric Estimators for Probability Density Functions ", PhD dissertation in Statistics, College of Administration and Economics, University of Baghdad.

7.Hmood, Munaf Yousif and Ashour, Marwan Abdul Hamid, (2012), " Comparing Several Nonlinear Estimators for Regression",Journal of Economics and Administrative Sciences, Vol.18, No.69, PP.359-372

8.Hamza, Saad Kazim, (2009), "Comparing Some Core Methods in Estimating Nonparametric Regression Models with Complete and Incomplete Data", Department of Statistics, Master's Thesis, College of Administration and Economics, University of Baghdad.

9.Hussein, Shireen Ali, (2009), "Fortified Weighted Greatest Potential Capabilities and Their Comparison with Other Methods of the Logistics Model with a Practical Application", Master's Thesis, College of Administration and Economics, University of Baghdad.

10. Khamo, Kholoud Yusuf, (2005), "Comparison of Slide Methods for Estimating Nonparametric Regression Curve", Iraqi Journal of Statistical Sciences.

11. Liu, H., 2008. Generalized additive model. Department of Mathematics and Statistics University of Minnesota Duluth: Duluth, MN, USA, 55812.

12. Rasheed, Hussam Abdul Razzaq, (2014), "Non-parametric Precursors to the Variable and Partially Variable Transactions Model", PhD dissertation, Department of Statistics, College of Administration and Economics, University of Baghdad.

13. Rifada, M., Suliyanto, E.T. and Kesumawati, A., 2018. The Logistic Regression Analysis with Nonparametric Approach based on Local Scoring Algorithm (Case Study: Diabetes Mellitus Type II Cases in Surabaya of Indonesia). Int. J. Adv. Soft Comput. Appl, 10, pp.167-178.

14. Shehab, Dhymeaa Hameed, (2017), " Comparison Some Robust Estimation Methods and Bayesian Method in Estimate the logistic Regression Function with Practical Application ", Master Thesis, College of Administration and Economics, University of Baghdad.

15. H.R. Dhafir, M. Y. Hmood, Saad, K. H. , 2012 "Nadaraya-Watson Estimator a Smoothing Technique for Estimating Regression Function" Journal of Economics and Administrative Sciences, Vol.18 , No.65,PP. 283-291

مقارنه بين بعض الطرق لتقدير الانحدار اللوجستي الثنائي اللامعلمي

الباحث/ نرجس باسم خلف [1]
جامعة بغداد/ كلية الإدارة والاقتصاد/ بغداد، العراق
narjes.bassem1201a@coadec.uobaghdad.edu.iq

أ.د لقاء علي محمد
جامعة بغداد / كلية الإدارة والاقتصاد / بغداد ، العراق
lekaa.a@coadec.uobaghdad.edu.iq

Received: 28/2/2023 Accepted:19/3/2023 Published: March / 2023

مستخلص البحث:
في هذا البحث تم الاعتماد على طرائق المقدر اللبي (مقدر الكثافة اللامعلمي) في تقدير الانحدار اللوجستي ثنائي الاستجابة، حيث تم استعمال المقارنة بين طريقة نادارايا –واتسون وطريقة خوارزمية التهديف الموضعي ، وبطريقه العبور الشرعي والعبور الشرعي المعمم تم تقدير معلمة التمهيدλ (مثلى) وان معلمه التمهيد لها تأثير واضح في عملية التقدير ايضا لها دور اساسي في تقريب وتنعيم المنحني واقترابه من المنحنى الحقيقي، وان الهدف من استعمال المقدرات اللبية هو تعديل المشاهدات لكي نتمكن من الحصول على مقدرات ذات صفات قريبة من خواص المعلمات الحقيقية، وبالاعتماد على بيانات طبية المصابين بأمراض سرطان الدم الليمفاوي المزمن ومن خلال استعمال دالة (Gaussin) وبالاعتماد على معيار المقارنة (MSE) تبين ان طريقة نادارايا –واتسون هي الافضل وذلك لحصولها على اقل قيمة لهذا المعيار.

نوع البحث: ورقة بحثية.

المصطلحات الرئيسة للبحث: انموذج الانحدار اللوجستي الثنائي ، طريقه نادارايا –واتسون ، خوارزمية التهديف الموضعي، طريقة العبور الشرعي، العبور الشرعي المعمم ، طريقة الملئ المباشر .

[1] بحث مستل من رسالة ماجستير