Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9930
https://doi.org/10.56705/ijodas.v6i3.237

Research Article

A logarithmic square root regression model for the average blood glucose levels of the drug induced diabetic experimental rats treated with the Cissampelos pareira L. (Menispermaceae) root extract

Lalit Mohan Upadhyaya [1,*]; Himanshu Pandey [2]; Sudhanshu Aggarwal [3]
[1] Municipal Post Graduate College, Mussoorie, Dehradun, Uttarakhand -248179, India,
lmupadhyaya@rediffmail.com
[2] Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur, Uttar Pradesh -273009, India,
himanshu_pandey62@yahoo.com
[3] National Post Graduate College Barhalganj Gorakhpur Uttar Pradesh India,
sudhanshu30187@gmail.com

Correspondence should be addressed to Lalit Mohan Upadhyaya; lmupadhyaya@rediffmail.com

Received 18 November 2025; Accepted 20 December 2025; Published 31 December 2025

© Authors 2025. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes). License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
We formulate a nonlinear regression model for elucidating the average values of the blood glucose levels of the experimental rats categorized under the division Group 1 Normal Control (G1NC) in a recent study conducted by Ankit Kumar et al. (see, Ankit Kumar, Ravindra Semwal, Ashutosh Chauhan, Ruchi Badoni Semwal, Subhash Chandra, Debabrata Sircar, Partha Roy and Deepak Kumar Semwal, Evaluation of antidiabetic effect of Cissampelos pareira L. (Menispermaceae) root extract in streptozotocin-nicotinamide-induced diabetic rats via targeting SGLT2 inhibition, Phytomedicine Plus 2 (2022) 100374, 11pp., https://doi.org/10.1016/j.phyplu.2022.100374). By treating the recorded average blood glucose levels of the rats of the group G1NC (response), as a function of the number of days of the experiment (predictor), our projected model involves a linear relation between the logarithm of the response and the square root of the predictor and it explains about 85.6239% of the variability in the response.

Keywords: Diabetes, Cissampelos pareira, blood glucose, regression, residual probability plot, leverage, Kolmogorov-Smirnov statistic.

1. Introduction

The worldwide prevalence of diabetes mellitus is a cause of great concern amongst the health care providers and much of their time, energy and efforts are consumed in raising awareness about this disorder among their subjects. Often patients are taught by their physicians and paramedics about the symptoms of diabetes and about the ways to manage them. Among the most general symptoms of diabetes mellitus are thirst, polyuria, weight loss and blurred vision. About 10.5% of the world's entire adult population in the year 2021 was estimated to be suffering from diabetes, out of which about 90% cases were those of the type-2 diabetes. If allowed to go unchecked, in the long run, diabetes is known to damage the blood vessels [1] and about three-fourths of people with diabetes die due to coronary artery diseases [2]. The damage to the kidneys of the patients suffering from diabetes arises mainly due to diabetic nephropathy, which factor is mainly responsible for, more than 50% of the patients undergoing dialysis in the United States [3]. Type 2 Diabetes also causes damage to the liver and is said to be the cause of Nonalcoholic Fatty Liver Disease (NAFLD) among the patients suffering from it [4], [5]. The investigators of [4] studied the common and unique risk factors between Chronic Kidney Disease (CKD) and NAFLD in 1992 patients suffering from type2

diabetes mellitus and found that the age of the patients and their triglycerides and uric acid levels were the common risk factors which contributed to both the CKD and NAFLD and proved the existence of a bidirectional relationship between these two diseases in their subjects of study. The authors of [7] found that NAFLD was present in 50% of their subjects with diabetes, which figure was 23% for those who were free from this ailment. Out of those diabetic patients who also suffered from NAFLD, 76.2% exhibited a deranged serum glutamic oxaloacetic transaminase (SGOT) and 64.3 % had a deranged serum glutamic pyruvic transaminase (SGPT) [5]. This discussion underscores some of the complexities of diabetes, which has necessitated the search for an alternative source of medicine to cure and manage this disease. The alternative systems of medicine like, the Unani, the Ayurvedic and the Homeopathic, etc. are also being investigated simultaneously in search of a possible medicine for the treatment of diabetes. One such study is the recent work of Ankit Kumar et al. [6] in which these authors probed the effectiveness of the ethanolic root extract of the plant Cissampelos Pareira Linn (Menispermaceae) in containing diabetes in experimental rats, which was induced by injecting them with streptozotocin, prior to which these animals were injected with nicotinamide [6]. Cissampelos Pareira is a well-known medicinal plant, whose roots are curved, narrow, long and cylindrical [7]. The root decoction of Cissampelos pareira is said to be used in malaria, pneumonia, and dog and snake bites [8]. In continuation of our ongoing long term project [9] – [13] of studying in depth the mathematical features of the various experimental results of Ankit Kumar et al. [6] from as many diverse angles as possible, which our capabilities allow us at present, we aim to propound a logarithmic square root regression model for the average blood glucose levels of the rats of the group 1 normal control (G1NC) of this paper by treating it as a function of the number of days of the experiment.

The paper is organized as follows. After reproducing the secondary data for this study from the work of Ankit Kumar et al. [6] in section 2 of the paper, we give the details of our formulated nonlinear regression model in section 3. The conclusion of the paper in section 4 finishes the paper.

Abbreviations used in the paper:

For brevity and convenience we shall use the following abbreviations at different places in this paper. For coherence we decided to retain some of the abbreviations used by the authors of [6].

G1NC: group 1 normal control

G1NCm: Mean blood glucose level of the rats of G1NC

1.2345E-03 = 1.2345×10^{-3}

Day: Day of Experiment

1.2345E+03 = 1.2345×10^{3}

ANOVA: Analysis of Variance

Data for the study

Our secondary data of study for this paper is reproduced in **Table 1** below, whose original source is the work of Ankit Kumar et al. [6], **Table 3** on p. 6 of [10].

Table 1. Measurements of the blood glucose levels in mg/dL of the experimental rats showing the effect of the Cissampelos pareira root extract on it.

No.	Day	G1NCm (mg/dL)
1.	0	74.53
2.	7	79.60
3.	14	96.17
4.	21	95.09
5.	28	96.89

Source: A. Kumar et al. [7]

Logarithmic and square root (nonlinear regression) model for the mean blood glucose levels of the rats of the group G1NC

We move ahead to discuss in detail our formulated nonlinear regression model for the dataset of **Table 1**. The precise equation of our model is given by

$$y = \exp\left(a + b\sqrt{x}\right), \tag{1}$$

which can be written alternatively as

$$\log y = a + b\sqrt{x}. \tag{2}$$

It can be seen that (3.2) gives a linear relation between $\log y$ and $\sqrt{x}$, therefore, it justifies the name logarithmic and square root model for the proposed model. In (3.1) and (3.2), the response y denotes the average blood glucose level of the experimental rats, denoted by G1NCm in Table 1 and the predictor x denotes the number of days of the experiment, which is shown by the variable 'Day' in Table 1. The quantities a and b are the regression parameters.

By appealing to the widely known technique of regression analysis (which needs no introduction for the mathematicians and statisticians and for the other interested readers we mention just a few relevant references [14] – [16]), we give below in **Table 2** the details of the model of (3.1).

Table 2. Details of the Nonlinear Model of (3.1)

Simple (Nonlinear) Regression - G1NCm vs. Day
Dependent variable: G1NCm (Mean of Blood glucose (mg/dL))
Independent variable: Day (Day of Experiment)
Logarithmic-Y square root-X model: Y = exp(a + b*sqrt(X))
Number of observations: 5

Coefficients

Parameter	Least Squares Estimate	Standard Error	T Statistic	P-Value
Intercept	4.29715	0.0488262	88.0091	0.0000
Slope	0.0551603	0.0130493	4.22706	0.0242

NOTE: intercept = ln(a)

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0.0520685	1	0.0520685	17.87	0.0242
Residual	0.00874218	3	0.00291406		
Total (Corr.)	0.0608107	4			

Correlation Coefficient = 0.925332
R-squared = 85.6239 percent
R-squared (adjusted for d.f.) = 80.8319 percent
Standard Error of Est. = 0.053982
Mean absolute error = 0.0326116
Durbin-Watson statistic = 3.05566 (P=0.8200)
Lag 1 residual autocorrelation = -0.552784

From **Table 2**, we obtain that the precise form of (3.1) for the sample of **Table 1** can be rewritten in the form

$$\text{G1NCm} = \exp(4.29715 + 0.0551603*\text{sqrt(Day)}), \tag{3}$$

or,

$$y = \exp\left(4.29715 + 0.0551603\sqrt{x}\right). \quad (4)$$

We observe that **Table 2** gives the P-Value of 0.0242 from the ANOVA table, which value being less than the critical value of 0.05, therefore, we conclude that at the 95% confidence level there is a statistically significant relationship between the response G1NCm and the predictor Day. Our fitted model of (3.1) can explain 85.6239% of the variability of G1NCm as shown by the value of the R-Squared statistic. We may point out here that, based on the coefficient of determination (i.e., the R-Squared) statistic, our four previously proposed nonlinear models [10, 11, 13, 14] (see, [10, **Table 3**, p. 8], [11, Table 2(a), p. 106], [13, Table 2, p. 73] and [14, Table 2(a), p. 115]) could explain respectively 99.63%, 90.45%, 95.707% and 99.66% of the variability in G1NCm, yet we prefer to discuss the present model alongside these and other possible models for the dataset of **Table 1** because we follow a different approach to discuss the present model. We find a very strong relationship between the variables G1NCm and Day as is shown by the high value of the correlation coefficient at 0.925332. We shall utilize the value of the standard error of the estimate of the model (i.e., 0.053982 mg/dL), which is in fact the standard deviation of the residuals, in the sequel to construct the prediction limits for new forecasted possible values of the response G1NCm based on this model. The mean absolute error (MAE) of 0.0326116 mg/dL shows the average value of the residuals. The value of Durbin-Watson statistic [18] for the model is 3.05566 with a high P-Value of 0.82. The Durbin-Watson statistic measures the serial correlation in the residuals. For the cases where the residuals vary randomly, the value of this statistic is close to 2. Since in our case the P-Value of this statistic is 0.82, which being much higher than the threshold of 0.05, therefore, we infer that there is no serial autocorrelation in the residuals of our dataset of Table 1 at the 95% confidence level. The value of the Lag 1 residual autocorrelation, which signifies the estimated correlation between consecutive residuals, at -0.552784 shows that our proposed model cannot account for the significant structure of the data, that explains why the variability of G1NCm based on the R squared statistic is only 85.624% for this model. We could not perform the Analysis of Variance with Lack-of-Fit test for the sample of Table 1 because the experiments of Ankit Kumar et al. [7] have not provided replicated observations of G1NCm at the same values of Day.

In **Figure 1** below we present the plot of the logarithmic - y square root - x model, written in the alternative form of (3.3) above. The solid blue colored line in **Figure 1** corresponds to the plot of (3.3), the green colored inner region corresponds to the 95% confidence limits for the response G1NCm at given values of Day and the outer yellow colored region represents the 95% prediction limits for the same.

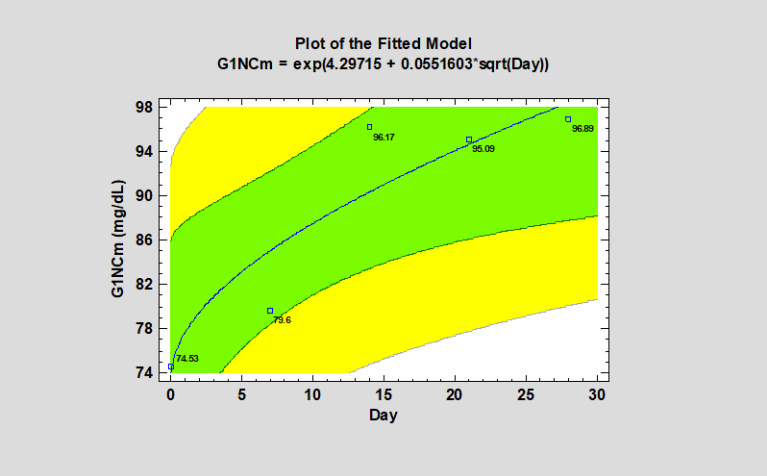


Figure 1. Plot of the fitted model of (3.3) for the sample of Table 1.

The values of G1NCm predicted by the model of (3.3) for the sample of **Table 1** are shown in Table 3(a) along with their respective 95% confidence intervals and 95% prediction intervals, which respectively correspond to the 95% confidence band and the 95% prediction band in the graph (**Figure 1**) of the fitted model. In Table 3(b) we

tabulate the residuals and percentage errors for the sample of Table 1 for the proposed model, which shows that the percentage error varies in this case from a minimum of 0.4887 to a maximum of 6.8308.

Table 3. Predicted Values by the Model of (3.3) for the sample of **Table 1**.

Predicted Values

X (Day)	Predicted Y (G1NCm in mg/dL)	Lower 95% Pred. Limit (mg/dL)	Upper 95% Pred. Limit (mg/dL)	Lower 95% Conf. Limit (mg/dL)	Upper 95% Conf. Limit (mg/dL)
0	73.49	58.2944	92.6466	62.9136	85.8443
7.0	85.0373	70.3314	102.818	78.4323	92.1984
14.0	90.3364	74.7578	109.161	83.435	97.8085
21.0	94.6253	77.7728	115.13	86.0816	104.017
28.0	98.3989	80.0508	120.953	87.7671	110.319

Table 4. Residuals and Percentage Error by the Model of (3.3) for the sample of **Table 1**.

Day x	**Observed G1NCm** y **(mg/dL)**	**Predicted G1NCm** $\hat{y}$ **(mg/dL)**	**Residual** $y-\hat{y}$ **(mg/dL)**	**% Error** $\frac{100\times\lvert y-\hat{y}\rvert}{y}$
0	74.53	73.49	1.04	1.3954
7	79.60	85.0373	-5.4373	6.8308
14	96.17	90.3364	5.8336	6.0659
21	95.09	94.6253	0.4647	0.4887
28	96.89	98.3989	-1.5089	1.5573

In **Table 5** below we predict the values of G1NCm on a daily basis for a complete month from the fitted model of (3.3) along with their 95% confidence intervals and the 95% prediction bands. The reader may easily note that the entire Table 3(a) (with the exception of its first row) is exactly reproduced again in Table 3(c), yet we make no attempt to delete these repeated values from **Table 6**, because in **Table 4** we made a segregated presentation of these values so that an interested reader may easily compare the predictions of G1NCm from the model of (3.3) with the corresponding observed values of G1NCm from **Table 1** for the specifically mentioned weekly values of the predictor Day and their corresponding percentage errors from **Table 5**.

Table 5. Predictions for G1NCm made on daily basis by the Model of (3.3).

Predicted Values

X	Predicted Y	Lower 95% Pred. Limit	Upper 95% Pred. Limit	Lower 95% Conf. Limit	Upper 95% Conf. Limit
1.0	77.6576	62.9383	95.8193	68.8042	87.6501
2.0	79.4523	64.8498	97.343	71.297	88.5405
3.0	80.8576	66.3	98.6116	73.2038	89.3116
4.0	82.0615	67.5065	99.7548	74.7907	90.0392
5.0	83.1371	68.5542	100.822	76.1598	90.7536
6.0	84.1216	69.4871	101.838	77.3633	91.4703
7.0	85.0373	70.3314	102.818	78.4323	92.1984
8.0	85.8985	71.1046	103.77	79.3879	92.9429
9.0	86.7153	71.8188	104.701	80.2454	93.7067
10.0	87.495	72.483	105.616	81.0172	94.4907
11.0	88.2431	73.1041	106.517	81.7135	95.2944
12.0	88.9638	73.6874	107.407	82.3434	96.1165
13.0	89.6607	74.2375	108.288	82.9149	96.9552
14.0	90.3364	74.7578	109.161	83.435	97.8085
15.0	90.9931	75.2514	110.028	83.9101	98.674
16.0	91.6329	75.7209	110.889	84.3457	99.5497
17.0	92.2572	76.1683	111.745	84.7466	100.433
18.0	92.8676	76.5957	112.596	85.1172	101.324
19.0	93.465	77.0046	113.444	85.4612	102.218
20.0	94.0506	77.3966	114.288	85.7817	103.117
21.0	94.6253	77.7728	115.13	86.0816	104.017
22.0	95.1899	78.1345	115.968	86.3632	104.919
23.0	95.745	78.4828	116.804	86.6286	105.821
24.0	96.2913	78.8184	117.638	86.8796	106.723
25.0	96.8294	79.1423	118.469	87.1177	107.624
26.0	97.3597	79.4552	119.299	87.3442	108.524
27.0	97.8828	79.7578	120.127	87.5603	109.422
28.0	98.3989	80.0508	120.953	87.7671	110.319
29.0	98.9086	80.3347	121.777	87.9654	111.213
30.0	99.4122	80.61	122.6	88.1559	112.106

We further remark here that in one of our very recent paper [12] we investigated the question as to what possible probability distribution could best explain the trend of the values of G1NCm in the sample of Table 1. We found that by using the Kolmogorov-Smirnov D statistic [19] as our distinguishing test, the Johnson SB distribution [20, 21] could best explain the observed variation of the response G1NCm in Table 1. In that paper we also generated a set of thirty random values of G1NCm from the fitted Johnson distribution (see [12, Table 12, p. 28]). By arranging those thirty values of G1NCm in an ascending order we recover the values of the predictor Day from the model of (3.3) in **Table 7** below. The reader may at once also observe from **Table 6** and **Table 7** that the maximum observed value of the response G1NCm which is 96.89 mg/dL and is reported on the Day 28 of the experiment of Ankit Kumar et al. [7] (see Table 1), is attained just after the Day 25 according to our proposed model of (3.3); still the agreement between our proposed model and the observed value of G1NCm on the last day (Day 28) of the experiment of Ankit Kumar et al. [7] is not very bad because the observed percentage error is 1.56% (see Table 3(b)), which is within the manageable

limits. Again it would be intuitive for the reader to have a look at the predicted (extrapolated) value of G1NCm on the Day 30 of the experiment, which from **Table 6** is 99.4122 mg/dL.

Table 6 Recovering the values of the predictor Day for given values of the response G1NCm from the Model of (3.3).

S. No.	$y =$ G1NCm (mg/dL)	Day $x = \left(\frac{\ln y - 4.29715}{0.0551603}\right)^2$	S. No.	$y =$ G1NCm (mg/dL)	Day $x = \left(\frac{\ln y - 4.29715}{0.0551603}\right)^2$
1.	74.4789	0.058714	16.	94.0342	19.9715
2.	74.6845	0.085431	17.	94.8573	21.40862
3.	74.8733	0.114283	18.	95.2849	22.16981
4.	76.0764	0.393188	19.	95.4347	22.43881
5.	76.8592	0.660375	20.	96.1989	23.82956
6.	80.4258	2.673102	21.	96.3233	24.05884
7.	80.8209	2.971503	22.	96.3729	24.15048
8.	84.5873	6.500223	23.	96.8295	25.00004
9.	84.6249	6.54137	24.	96.845	25.02906
10.	85.1435	7.120196	25.	96.8691	25.07422
11.	85.3867	7.398828	26.	96.8768	25.08865
12.	87.6584	10.21501	27.	96.892	25.11715
13.	88.6579	11.57111	28.	96.896	25.12465
14.	89.6397	12.9693	29.	96.897	25.12653
15.	90.0592	13.58611	30.	96.8972	25.1269

Replying to the queries of the referees on the original version of this paper, we elaborate here that this is not at all surprising that an exponential function with a positive argument (as is the case with (3.3)) has an inherent increasing character, therefore, we add an explanatory note here that an inquiring reader will obviously ask the natural question that with the increasing number of the predictor Day in (3.3), the response G1NCm would increase indefinitely beyond limits, which would not have been true in the actual practice, if the experimenters Ankit Kumar et al. [7] would have carried on their experiment of [7] beyond the reported period of four weeks. In order to resolve this issue, we checked the rate of growth of the exponential function in (3.3) and we point it out here that the function does not grow at a very fast rate. Therefore, it suffices for our needs of interpretation of growth of G1NCm with Day according to the dataset of **Table 1**, since we have checked that G1NCm reaches a value of 198.56 mg/dL on Day 324.6895 as per (3.3), which is the average value of blood glucose level of the rats of the group 4 test (G4T) attained on the 28th day of the experiments of [7, p. 6 of 11] after this group of rats were administered orally with the alcoholic root extract of Cissampelos pareira at 500 mg/kg/day for twenty eight days, which is said to have brought down their observed average blood glucose level from 303.9 mg/dL on the basal (zero) day to 198.56 mg/dL on the twenty eighth day. We have tried to capture a snapshot of this situation in Fig. 2(a), which shows an approximate visualization of (3.3) roughly revealing that G1NCm is 198.838 mg/dL at a value 325.031 Day. We also draw in **Figure 3** a plot of (3.4) in the range of $0 \le x \le 400$, which clearly justifies our above-mentioned stand. The first derivative of G1NCm with respect to the predictor Day after simplification is given by

$$y' = \frac{2.026866556e^{0.0551603\sqrt{x}}}{\sqrt{x}}. \tag{5}$$

The plot of (3.5) is shown in **Figure 5**, which shows that this function decreases sharply after the first day and we studied the behavior of this first derivative for some specific values of the Day up to four hundred days and we present our calculated values of y' in **Table 8**, which shows that the function of (3.4) (hence (3.3)) increases at a very slow rate after the seventh day up to the four hundredth day.

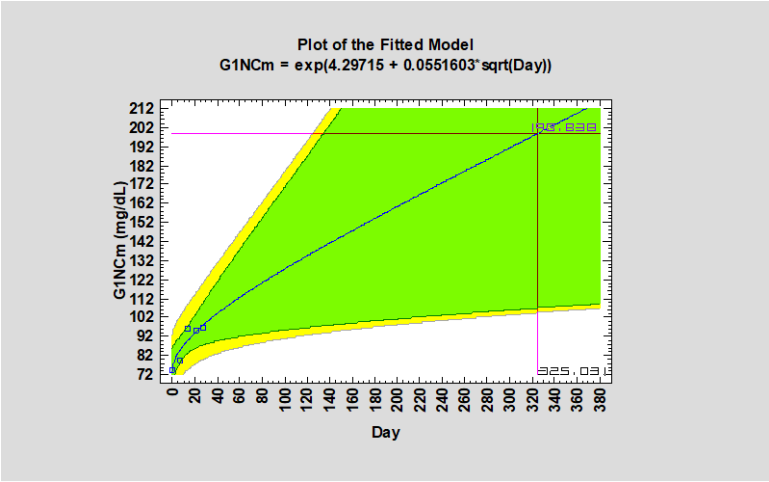


Figure 2. Plot showing the behavior of the response G1NCm according to (3.3) with increasing number of the predictor Day.

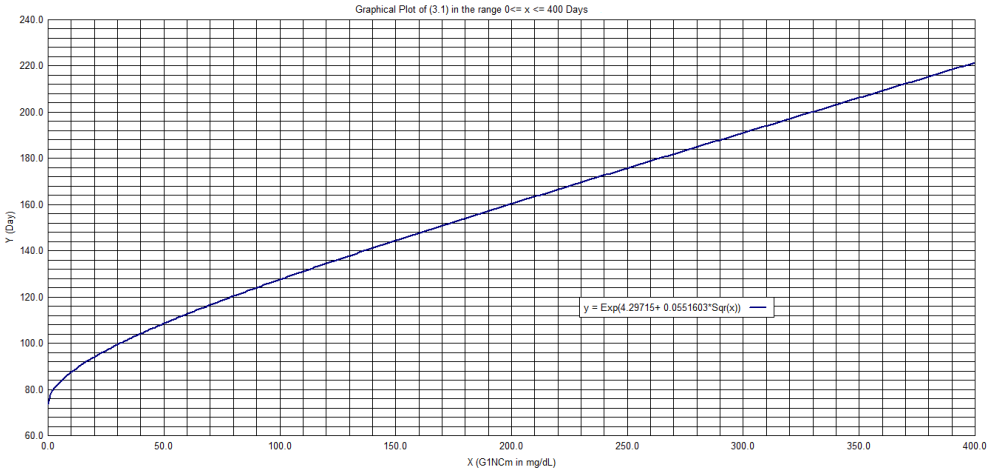


Figure 3. Plot showing the behavior of the response G1NCm (y) according to (3.4) with increasing number of the predictor Day (x) .

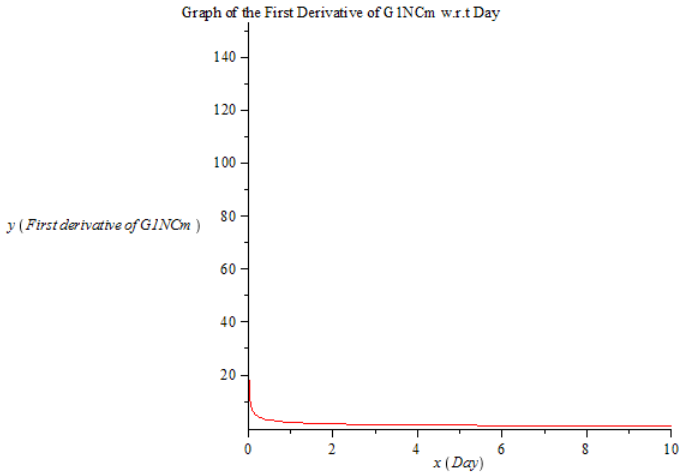


Figure 4. Plot of the first derivative of G1NCm with respect to Day.

Table 7. Values of the first derivative of G1NCm with respect to Day for specified values of the latter.

S. No.	Day	y' (mg/(dL Day))	S. No.	Day	y' (mg/(dL Day))
1.	50	0.4233829250	5.	250	0.3066407215
2.	100	0.3518708547	6.	300	0.3042225469
3.	150	0.3252199133	7.	350	0.3040628593
4.	200	0.3126774258	8.	400	0.3054298222

While searching for the model of (3.3) (or say, (3.1)), we also looked into the possibilities of other feasible contending nonlinear models, which could explain reasonably well the variation of the response G1NCm in **Table 1** with that of the predictor Day. The comparative analysis of our search showed us that based on the R-Squared statistic, the model of (3.1) (i.e., (3.3)) excelled the other available contenders, as is shown underneath in **Table 9**, therefore, we undertook to describe the model of (3.1) (i.e., the logarithmic y - square root x model) in full detail in this paper.

Table 8. Relative comparison of the other alternative possible models.

Comparison of Alternative Models

Model	Correlation	R-Squared
Logarithmic-Y square root-X	0.9253	85.62%
Double square root	0.9226	85.13%
Square root-X	0.9200	84.63%
Squared-Y square root-X	0.9146	83.66%
Reciprocal-Y	-0.9011	81.19%
Exponential	0.9009	81.17%
Square root-Y	0.9008	81.14%
Linear	0.9006	81.10%
Squared-Y	0.9000	81.00%
Double squared	0.7800	60.84%
Squared-X	0.7768	60.35%
Square root-Y squared-X	0.7751	60.08%
Logarithmic-Y squared-X	0.7734	59.81%
Reciprocal-Y squared-X	-0.7696	59.23%

From Table 10 we see that there are no Unusual Residuals in the fitted model of (3.3), meaning thereby that there are no observations in the sample of Table 1 for which the Studentized residuals for the fitted model are greater than 2 in the absolute value. For the purpose of information only we recall here that the Studentized residuals indicate how many standard deviations far away each observed value of the response G1NCm lies from the fitted model, using all the data points of the sample, except the observation being talked about. It is quiet well known that in the process of fitting a regression model to a given dataset; all the observations of the dataset do not exert equal influence on parameter estimates of the fitted model. While fitting a simple regression by the method of ordinary least squares only, the points located at the extreme values (very high or very low) of the predictor have much greater influence on the parameter estimates of the fitted model in comparison to those occurring in the vicinity of the mean of the predictor. The influence of an observation of a dataset in the determining the values of the parameters of a regression model are measured by a statistic called the leverage [22 - 24]. Table 11 underneath shows that there are no Influential Points in our fitted model of (3.3) to the sample of Table 1. We recapitulate that the Influential Points of a sample are those points which have leverage values greater than three times the leverage of an average data point. In our case, the leverage of an average data point of the sample of Table 1 is 0.4 (see Table 11) under the fitted model of (3.3), thus there are no observations in the sample of Table 1 which show a leverage of 1.2 or more.

Table 9. Unusual Residuals of the fitted model of (3.3).

Unusual Residuals

Row	X	Y	Predicted Y	Residual	Studentized Residual
Nil	Nil	Nil	Nil	Nil	Nil

Table 10. Influential Points of the fitted model of (3.3).

Influential Points

Row	X	Y	Predicted Y	Studentized Residual	Leverage
Nil	Nil	Nil	Nil	Nil	Nil

Average leverage of single data point = 0.4

Figure 5 shows the plot of the observed values of the response G1NCm against those of its predicted by the model of (3.3). We see from Fig. 3(a) that three of the five values of Table 1 lie close to the blue colored diagonal reference line, while the two remaining values of 79.6 mg/dL and 96.17 mg/dL (i.e., the second and the third observations of Table 1) lie somewhat farther from this line, this is not unexpected because we have already seen in Table 5 that the percentage error corresponding to these two observations is maximum at 6.8308 and 6.0659 respectively. Within the range of the dataset of Table 1, we further observe from Fig. 3 that with increasing values of the predictor day, the values of G1NCm do not show an abnormal increase in variability around the diagonal reference line, thus no heteroscedasticity is apparent in the sample of Table 1. From the box plot of G1NCm drawn in Figure 6 we find that there are no outside points and no far outside points in this sample, thus there are no outliers present in the sample of Table 1. It is also instructive to point out that in Fig. 3(b) the sample mean is marked by a plus sign and the vertical center line within the box shows the sample median.

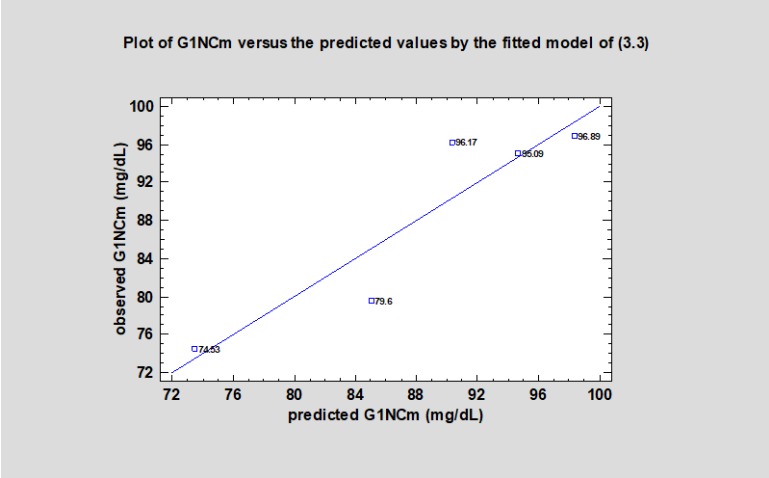


Figure 5. Plot of G1NCm versus the predicted values by the fitted model of (3.3).

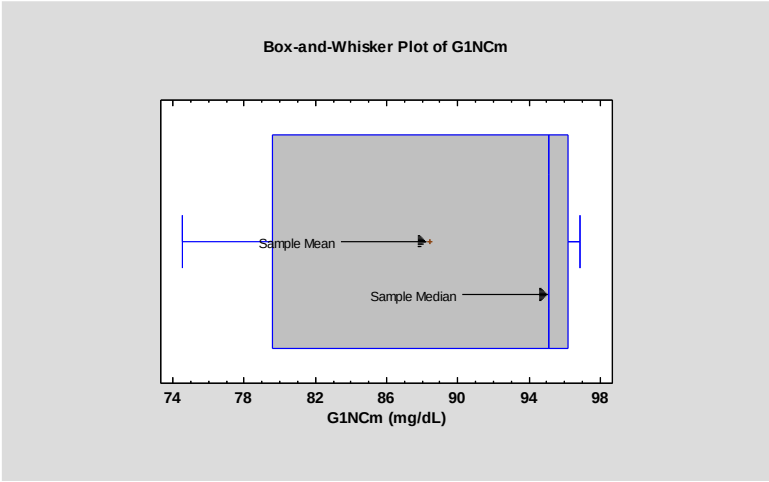


Figure 6. Box and Whisker Plot of G1NCm.

The plot of the ordinary residuals (see, column four, Table 3(b)) versus the predictor Day for our fitted model of (3.3) is shown in **Figure 7**. A random pattern of the residuals around the blue colored zero line indicates a good fit. **Figure 8** shows the plot of the Studentized residuals against Day, as is evident from there that there are no observations which lie beyond the limits of ±2 mg/dL, thus this confirms that there are neither any unusual points (see also **Table 10**) nor any outliers in the sample of **Table 1**.

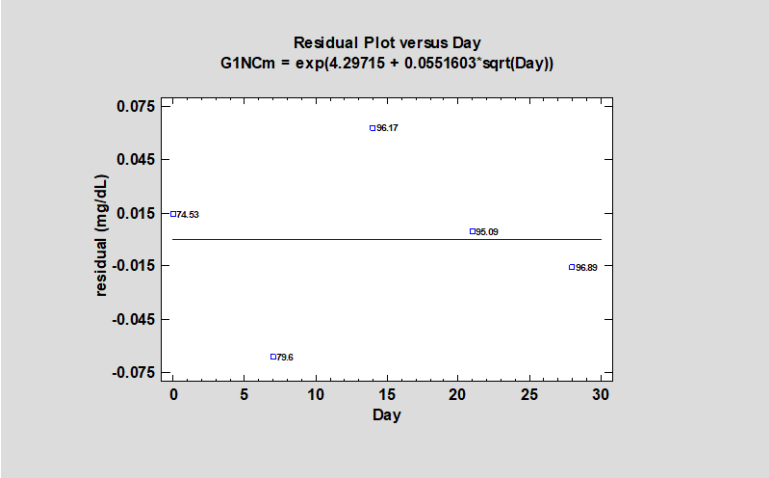


Figure 7. Plot of (ordinary) Residuals versus Day for the model of (3.3).

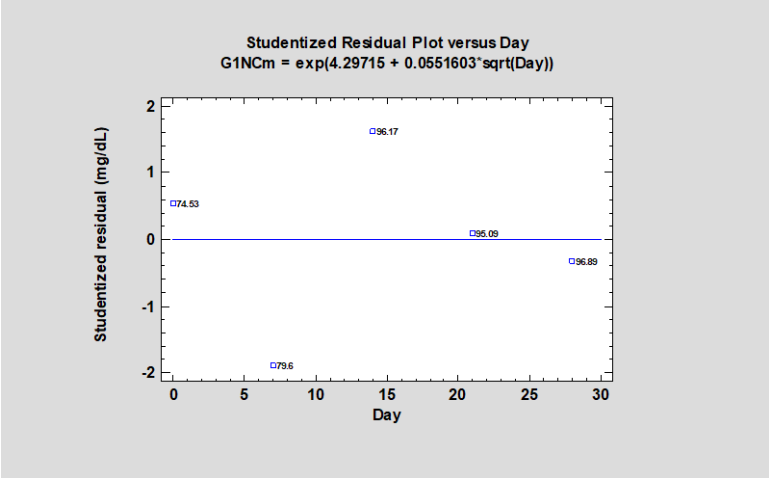


Figure 8. Plot of Studentized Residuals versus Day for the model of (3.3).

Figure 9 shows the plot of the ordinary residuals versus the predicted values of G1NCm and **Figure 10** displays the plot of the Studentized residuals versus the predicted values of G1NCm, from both these plots we see that as the dependent variable G1NCm changes, the variability of the residuals also changes, but with increasing values of G1NCm towards the right of the figure, the increase in variability of the residuals approaches the more or less towards the central zero line in contrast to the trend exhibited in the middle of the figure, thus we infer that there is no heteroscedasticity in the sample of **Table 1**.

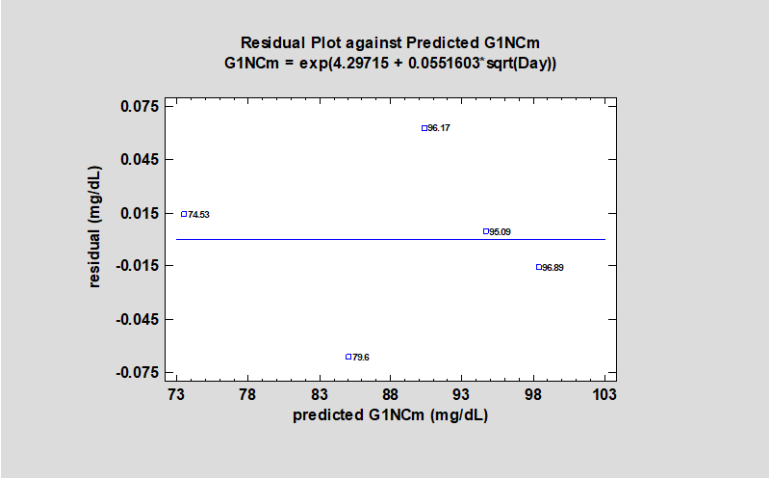


Figure 9. Plot of (ordinary) Residuals versus predicted G1NCm for the model of (3.3).

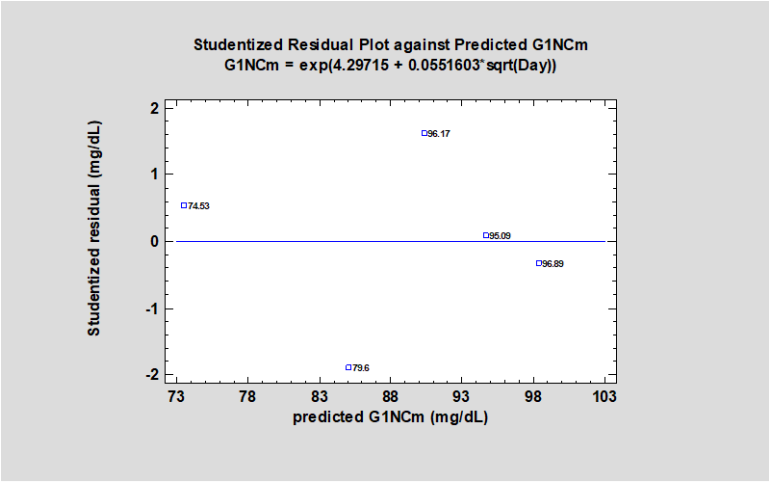


Figure 10. Plot of Studentized Residuals versus predicted G1NCm for the model of (3.3).

The ordinary Residuals and the Studentized Residuals of the fitted model versus the Row Number are plotted respectively in **Figure 11** and **Figure 12**, the pattern of the variation of the residuals with the row number seems to be random, thus we infer that no serial correlation is present in our sample of **Table 1**, where it is very much evident from **Table 1** that the successive observed values of G1NCm were collected in the sequential order of the row number. This also confirms the finding that we have recorded above in **Table 2** that the P-Value of the Durbin-Watson statistic being 0.82 shows no signs of serial autocorrelation in the residuals at the 95% confidence level.

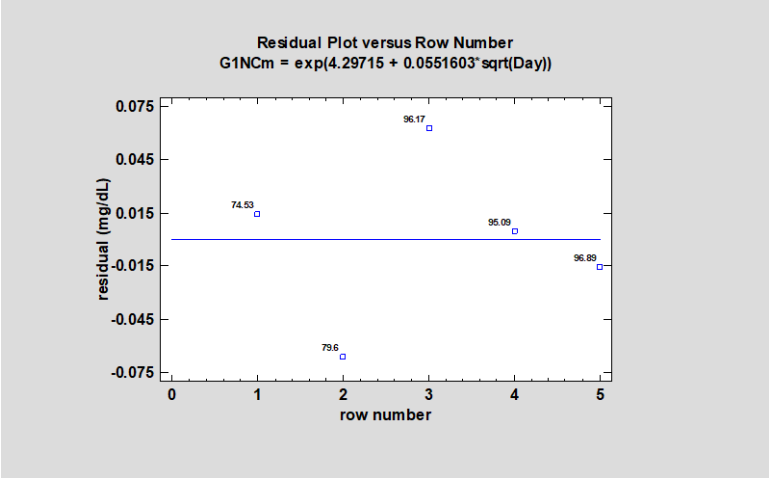


Figure 11. Plot of (ordinary) Residuals versus Row Number for the model of (3.3).

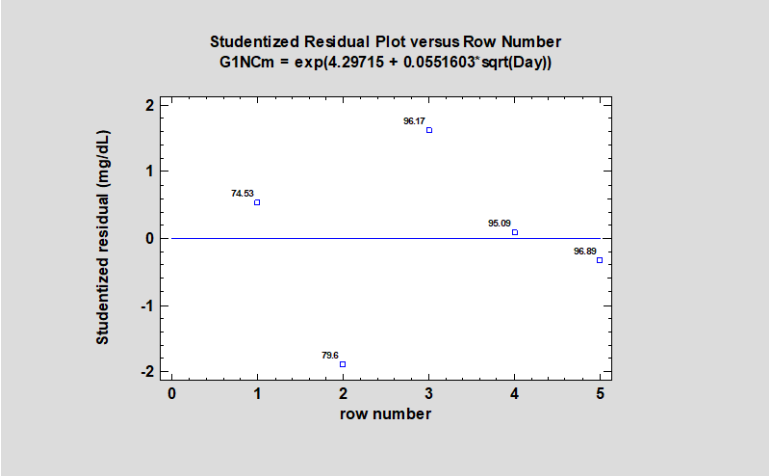


Figure 12. Plot of Studentized Residuals versus Row Number for the model of (3.3).

After arranging the residuals in an increasing order and then determining the order statistic (keeping in mind that, by definition, the i^{th} order statistic is the i^{th} smallest observation in the sample), which we denote by, say, x_i and then plotting the data at the points with coordinates $\left(x_i, \Phi^{-1}\left(\frac{i-0.375}{n+0.25} \right) \right)$, where, $n=5$, the number of residuals, $1 \le i \le 5$ and $\Phi^{-1}(k)$ represents the inverse standard normal distribution evaluated at the point k, we obtain the various Residual and the Studentized Residual Probability Plots shown in the **Figure 13** - **18**. We also draw the blue colored reference line in all these plots using three different methods. When using the method of Quartiles, the reference line passes through the median when the Percentage (i.e., the abscissa of these plots) equals 50 and the slope of this line is determined by the interquartile range of this data of residuals of the fitted model of (3.3), which are shown in the fourth column of Table 3(b) above. From **Table 11** below, it is obvious that the interquartile range of the data of residuals of the fit of (3.3) is $Q_3 - Q_1 = 1.04 - (-1.5089) = 2.5489$ mg/dL. As a better aid to the understanding of the Residual Probability Plots of the **Figure 13**, **15** and **18** we give in Table 6 the summary statistics of the data of the Residuals (column four, Table 3(b)) and their Percentiles.

Table 11. Summary Descriptive Statistics and Percentiles of the data of Residuals (column four of Table 3(b)) of the model of (3.3) fitted to the sample of **Table 1**.

Summary Statistics for Residuals	
Count	5
Average	0.07842
Standard deviation	4.09691
Coeff. of variation	5224.32%
Minimum	-5.4373
Maximum	5.8336
Range	11.2709
Stnd. skewness	0.109328
Stnd. kurtosis	0.459364

Percentiles for Residuals	
	Percentiles
1.0%	-5.4373
5.0%	-5.4373
10.0%	-5.4373
25.0%	-1.5089
50.0%	0.4647
75.0%	1.04
90.0%	5.8336
95.0%	5.8336
99.0%	5.8336

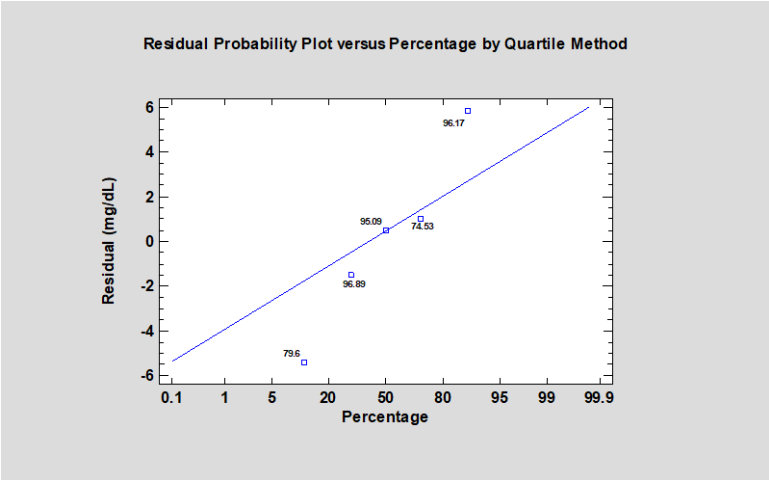


Figure 13. Residual Probability Plot versus Percentage by Quartile Method for the model of (3.3).

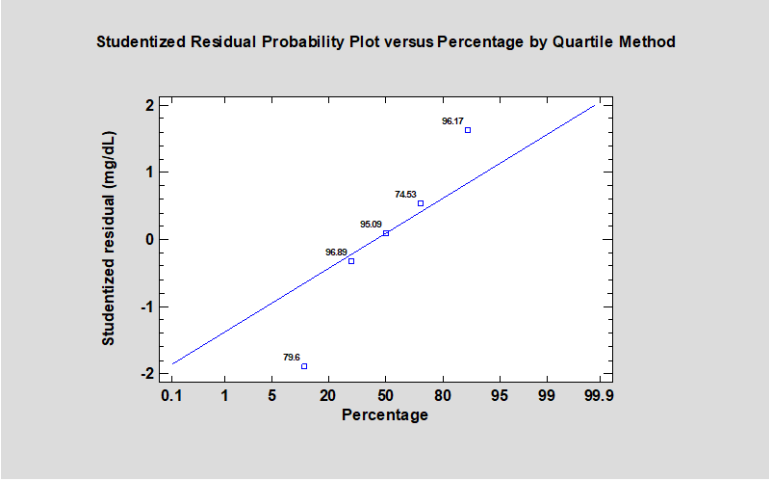


Figure 14. Studentized Residual Probability Plot versus Percentage by Quartile Method for the model of (3.3).

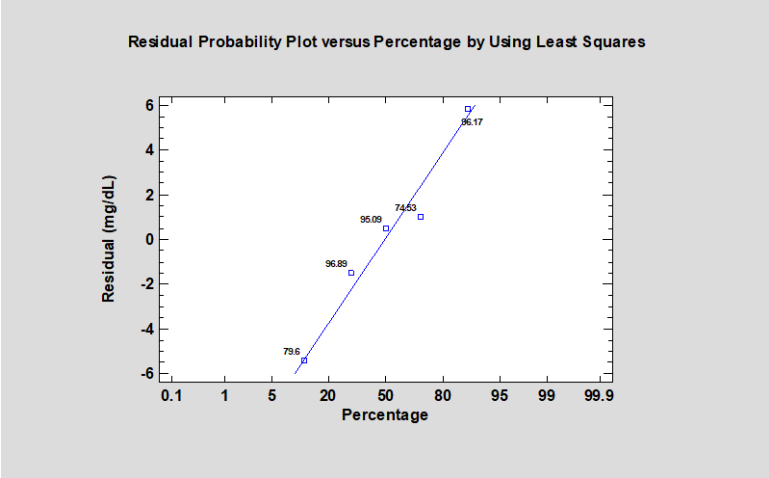


Figure 15. Residual Probability Plot versus Percentage by Using Least Squares for the model of (3.3).

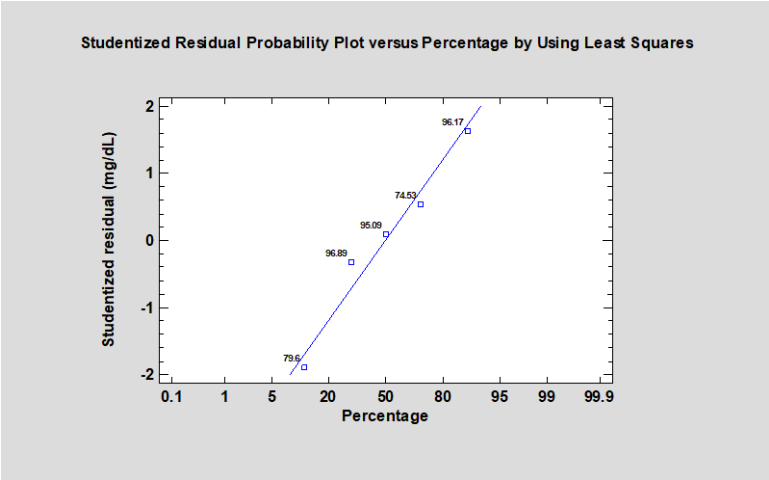


Figure 16. Studentized Residual Probability Plot versus Percentage by Using Least Squares for the model of (3.3).

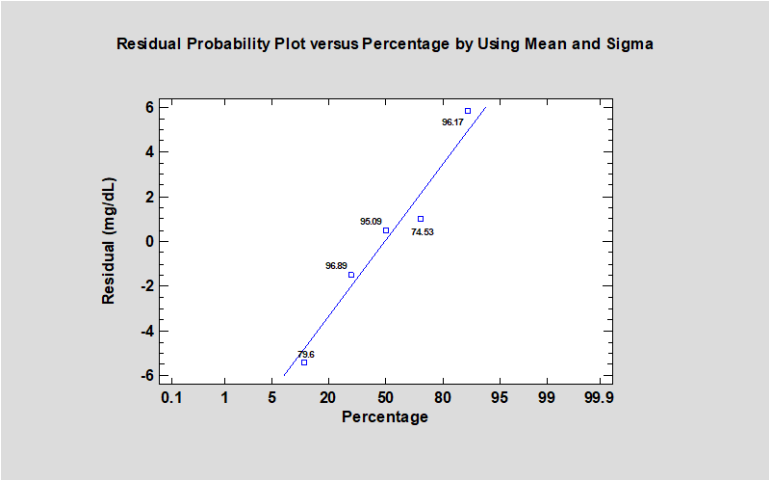


Figure 17. Residual Probability Plot versus Percentage by Using Mean and Sigma for the model of (3.3).

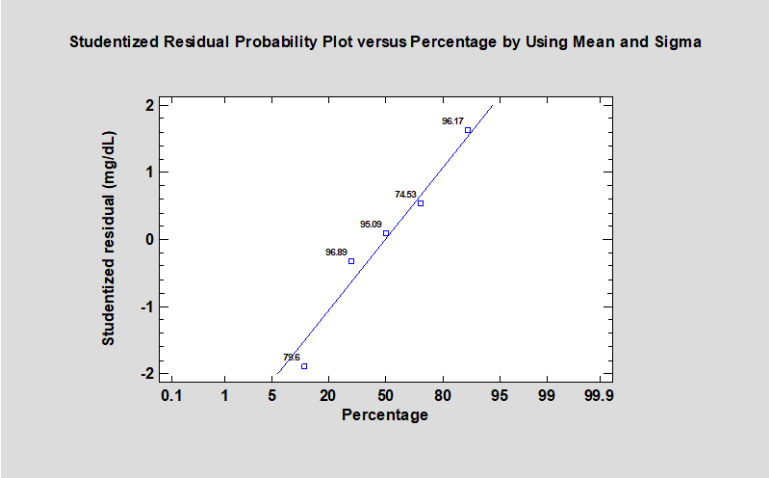


Figure 18. Studentized Residual Probability Plot versus Percentage by Using Mean and Sigma for the model of (3.3).

The box plot of the data of residuals of the fit of (3.3), depicted in **Figure 19**, shows the presence of two outside points, viz. -5.4373 mg/dL and 5.8336 mg/dL (corresponding respectively to the observed G1NCm values of 79.6 mg/dL and 96.17 mg/dL in the sample of **Table 1**), which lie on either side of the box at distances more than 1.5 times the interquartile range of 2.5489 mg/dL. The Density Trace of the sample of Residuals of the fit of (3.3) using the cosine method is shown in **Figure 20**. This density trace basically gives us an idea of the probability density function of the population from which this sample of Residuals of the fit of (3.3) is drawn. We see that with the exception of the two outside points of -5.4373 mg/dL and 5.8336 mg/dL, the middle portion of the sample of Residuals seems to come from a normal distribution. However, before ascertaining the fact that the sample of the Residuals, comes from a normal population, we also performed the Grubb's test, the Dixon's test and the procedure of outlier identification analysis of this sample (the lengthy details of which processes we omit here), we record our findings of that analysis here to mention that the two outside points of -5.4373 mg/dL and 5.8336 mg/dL are not outliers in this sample of Residuals at the 95% confidence level. Thus we conclude with 95% confidence that the sample of Residuals of the fit of (3.3) is drawn from a normal population. This fact is also supported by the high P-Value (=0.9278) of the Shapiro-Wilk test performed by us, whose outcomes are prominently displayed in the legend boxes of the various Normal Probability Plots for the data of Residuals of the fit of the model of (3.3) to the sample of **Table 1** in **Figures 21**, **22** and **23** using respectively the Quartile method, the Least Squares method and the Mean and Sigma method. We also note from **Figure 23** that the two outside points of -5.4373 mg/dL and 5.8336 mg/dL lie well within the 95% confidence band (the curved pink colored lines) of the fitted reference (blue colored) line, which is also an indicator of the fact they are not outliers in this sample of the Residuals of the fit of the model of (3.3). The three plots of **Figures 21**, **22** and **23** help us to understand the Residual Probability Plots of **Figure 13**, **Figure 15** and **17** respectively. In the method based on Quartiles, we can see from **Figure 13** and **Figure 21** that the diagonal reference lines pass exactly through the values of observed G1NCm of 95.09 mg/dL in **Figure 13** and the minimum residual of 0.4647 mg/dL in **Figure 21** both of which correspond to the Percentage (abscissa) of 50 in these plots. In addition to this the plot of **Figure 21** also includes a cross hair showing that the value of the ninetieth percentile of this data of Residuals is 2.88437 mg/dL as estimated using the fitted line. It is also to be noted here that one of the primary benefits of using the Quartile method for plotting the reference line in the plots of **Figure 13** and **Figure 21** is that by this method the shape of the central portion of the data is plotted with more weight due to which we are able to visualize the departures from the normality at the tails of the data, which is not possible with the other two methods (see **Figure 15** and **Figure 22** (both using the Least Squares method) and **Figure 17** and **23** (both using the Mean and Sigma method)). We especially draw the attention of the readers towards the fact that while in **Figure 13** and **22** the fitted reference lines pass almost exactly through the two extreme points (i.e., 79.6 mg/dL and 96.17 mg/dL in **Figure 15** and -5.4373 mg/dL and 5.8336 mg/dL in **Figure 22**); in **Figure 17** and **23** these said extreme points lie relatively

very close to the drawn reference lines. Therefore, the latter two methods of drawing the reference lines, viz., the Least Squares method and the Mean and Sigma method cannot most clearly picture the departures from normality at the tails of the data. The Density Trace plot of the sample of Residuals shown in Fig. 8(b) also supports these observations of ours.

When we fit the reference line in the Residual Probability Plot of the sample of Residuals in **Figure 15** or in the Normal Probability Plot in Fig. 8(d) by the method of Least Squares – this method calculates this line by performing the least squares regression on the normal quantiles of the observed order statistics. Similarly, for the purpose of fitting the reference line in the Residual Probability Plot of the sample of Residuals in **Figure 17**or in the Normal Probability Plot in Fig. 8(e), we employed the method of Mean and Sigma, where the reference lines of these graphs were determined from the mean and standard deviation of the five observations. It is also important to see that the cross hair in Fig. 8(d) shows that the ninetieth percentile of the sample of Residuals using the method of Least Squares is 5.98614 mg/dL and the corresponding value for **Figure 23** (while using the method of Mean and Sigma) is 5.32883 mg/dL with a 95% confidence interval of (1.67037 mg/dL, 17.1473 mg/dL). The reader would also appreciate the fact that the value of the ninetieth percentile (hence, the ninety fifth and the ninety ninth percentiles) of the sample of Residuals as reported in the second sub-table of Table 6 is 5.8336 mg/dL, lies closest to the prediction made for the ninetieth percentile of this sample in **Figure 22** by the method of Least Squares. Thus, each method of fitting the reference line in the Normal Probability Plot has its own advantages and disadvantages.

The Studentized Residual Probability Plots of the sample of Residuals with the corresponding fitted reference lines by the above said three methods for those readers who may also be interested in these frequently used plots.

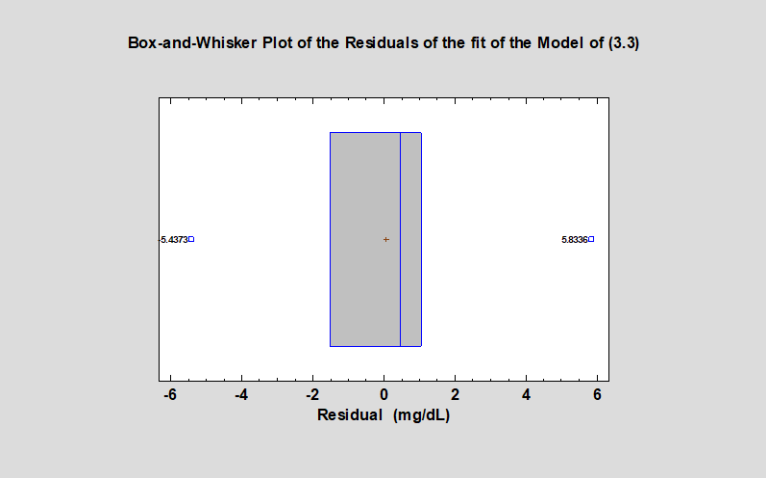


Figure 19. Box and Whisker plot of the Residuals of the fit of the Model of (3.3) to the sample of **Table 1**.

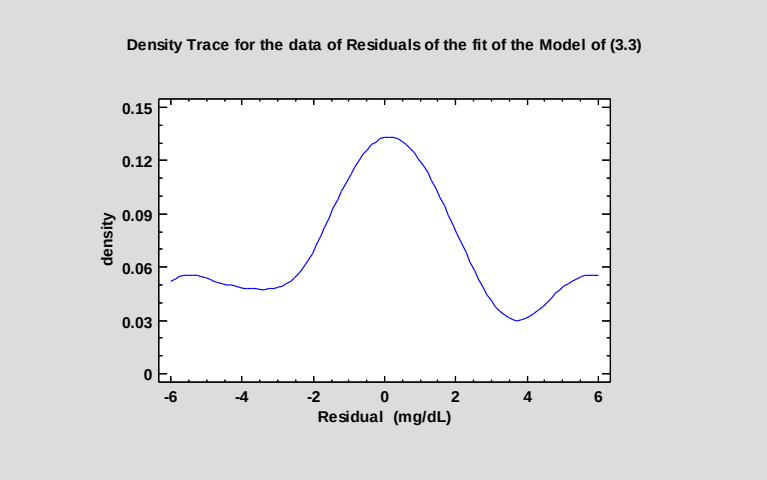


Figure 20. Density trace by cosine method of the Residuals of the fit of the Model of (3.3) to the sample of **Table 1**.

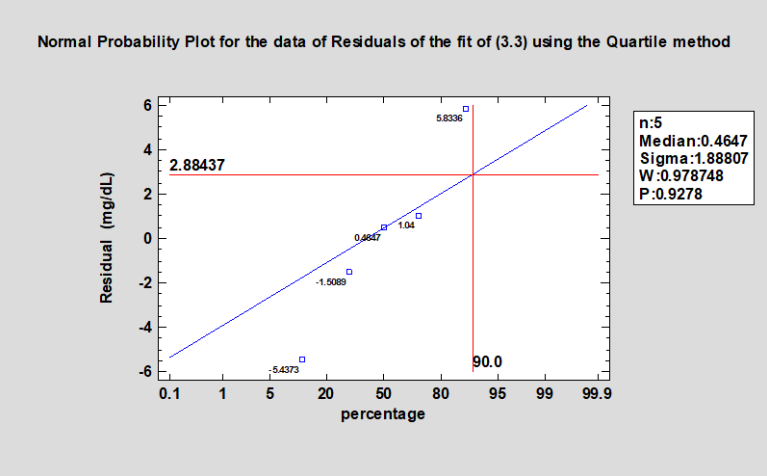


Figure 21. Normal Probability Plot for the data of Residuals of the fit of the Model of (3.3) to the sample of **Table 1** using the Quartile Method with the Shapiro-Wilk W(=0.978748) statistic and its P-Value (=0.9278).

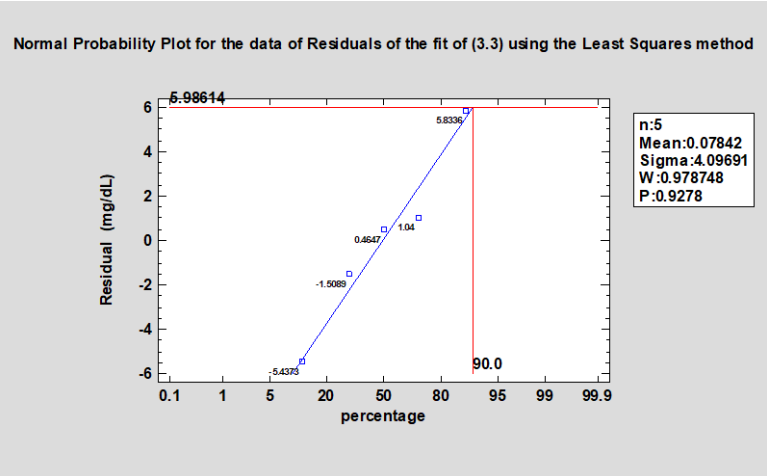


Figure 22. Normal Probability Plot for the data of Residuals of the fit of the Model of (3.3) to the sample of **Table 1** using the Least Squares method.

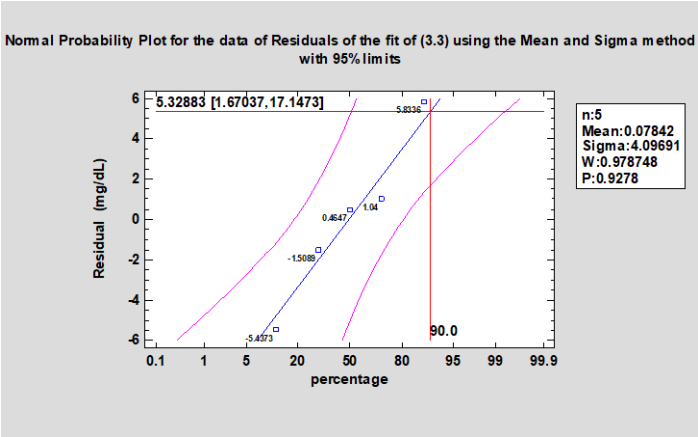


Figure 23. Normal Probability Plot for the data of Residuals of the fit of the Model of (3.3) to the sample of **Table 1** using the Mean and Sigma method.

4. Conclusion

A logarithmic square root (nonlinear) model was fitted to the dataset of Table 1 which shows the average blood glucose levels of the experimental diabetic rats belonging to the group G1NC in the study of Ankit Kumar et al. [7]. Though our model can explain only about 85.6239% of the variability in the response G1NCm treated as a function of the predictor Day, which is less than that of our previously proposed models [10,11,13,14], yet using a different approach we explained in full detail our proposed model of (3.1) (resp. (3.3)) and threw light on its various illuminating features. Our future communications regarding our ongoing study of the experimental results of Ankit Kumar et al. [7] will discuss in detail the other possible investigations that we have undertaken so far in this direction.

Acknowledgements

The authors record their gratitude to all the authors, publishers and copyright holders of the work [7] from where the secondary data of Table 1 of this paper is drawn for the studies made in this work. The authors are also thankful to the anonymous referees and the Editors whose queries have made them to struggle to put more meaningful analysis and information as pointed out above in the paper.

References

[1] "Diabetes – long-term effects," *Better Health Channel*, Victoria: Department of Health

[2] P. T. O'Gara *et al*., "2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines," *Circulation*, vol. 127, no. 4, p. e368, Jan. 2013, doi: 10.1161/CIR.0b013e3182742cf6

[3] E. J. Wing and F. Schiffman, *Cecil Essentials of Medicine*, 10th ed. Pennsylvania: Elsevier, 2022, ISBN: 978-0-323-72271-1.

[4] Y.-J. Lee *et al*., "Common and unique factors and the bidirectional relationship between chronic kidney disease and nonalcoholic fatty liver in type 2 diabetes patients," *Diabetes, Metabolic Syndrome and Obesity*, pp. 1203–1214, 2020, doi: 10.2147/DMSO.S237700.

[5] R. Gupta, S. P. S. Bhat, P. R. Gangadhar Rai, A. K. Kulamarva, and P. S. Prakash, "Study of liver function tests in patients with long standing type 2 diabetes mellitus in comparison to healthy individuals," vol. 10, no. 5, pp. 289+, Feb. 2021.

[6] A. Kumar *et al*., "Evaluation of antidiabetic effect of *Cissampelos pareira* L. (Menispermaceae) root extract in streptozotocin-nicotinamide-induced diabetic rats via targeting SGLT2 inhibition," *Phytomedicine Plus*, vol. 2, p. 100374, 2022, doi: 10.1016/j.phyplu.2022.100374.

[7] K. Shah, S. S. Qureshi, J. K. Gupta, N. Upmanyu, and N. S. Chauhan, "Ethnomedicinal uses, phytochemistry and pharmacology of *Cissampelos pareira*: A review," in *OMICS Group eBooks*, B. Kshetrimayum, Ed., pp. 19.

[8] S. Kumari, Anmol, V. Bhatt, S. S. Patil, and U. Sharma, "*Cissampelos pareira* L.: A review of its traditional uses, phytochemistry, and pharmacology," *J. Ethnopharmacol.*, 2021, doi: 10.1016/j.jep.2021.113850.

[9] L. M. Upadhyaya and S. Aggarwal, "Viewing the blood glucose levels of the drug induced diabetic experimental rats treated with the *Cissampelos pareira* L. (Menispermaceae) root extract," *J. Adv. Res. Appl. Math. Stat.*, vol. 9, no. 1&2, pp. 5–15, 2024.

[10] L. M. Upadhyaya and S. Aggarwal, "Viewing the blood glucose levels of the drug induced diabetic experimental rats treated with the *Cissampelos pareira* L. (Menispermaceae) root extract – II," *Partners Universal Int. Res. J. (PUIRJ)*, vol. 3, no. 2, pp. 102–119, 2024, doi: 10.5281/zenodo.12553641.

[11] L. M. Upadhyaya and S. Aggarwal, "A Johnson distribution for the average blood glucose levels of the drug induced diabetic experimental rats treated with the *Cissampelos pareira* L. (Menispermaceae) root extract," *J. Adv. Res. Appl. Math. Stat.*, vol. 9, no. 3&4, pp. 13–31, 2024.

[12] L. M. Upadhyaya and S. Aggarwal, "A coup d'oeil of the average blood glucose and SGPT levels of drug induced diabetic experimental rats treated with the *Cissampelos pareira* L. (Menispermaceae) root extract," *Partners Universal Multidisciplinary Res. J. (PUMRJ)*, vol. 1, no. 4, pp. 69–95, 2024, doi: 10.5281/zenodo.14203060.

[13] L. M. Upadhyaya and S. Aggarwal, "A power model for the average blood glucose and SGPT levels of drug induced diabetic experimental rats treated with the *Cissampelos pareira* L. (Menispermaceae) root extract," *Partners Universal Multidisciplinary Res. J. (PUMRJ)*, vol. 1, no. 4, pp. 112–130, 2024, doi: 10.5281/zenodo.14210174.

[14] N. H. Bingham and J. M. Fry, *Regression Linear Models in Statistics*. London, Dordrecht, Heidelberg, New York: Springer, 2010, ISBN: 978-1-84882-968-8.

[15] R. B. Darlington and A. F. Hayes, *Regression Analysis and Linear Models: Concepts, Applications, and Implementation*, Methodology in the Social Sciences. New York, NY, USA: The Guilford Press, 2017, ISBN: 978-1-4625-2113-5.

[16] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York, NY, USA: John Wiley & Sons, 1998.

[17] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression II," *Biometrika*, vol. 38, pp. 159–178, 1951.

[18] "Kolmogorov–Smirnov test," *Wikipedia*, The Free Encyclopedia.

[19] N. L. Johnson, "Systems of frequency curves generated by methods of translation," *Biometrika*, vol. 36, pp. 149–176, 1949.

[20] "Johnson Distribution," *Wolfram Language Documentation*.

[21] "Unusual observations, outlier, leverage, and influential points," *The Open Educator*.

[22] "Using leverages to help identify extreme values," *Penn State Eberly College of Science, STAT 501*.