Journal of Information Systems Engineering and Management
2025, 10(22s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Analysis of Parameter Estimation Methods in Logistic Regression for LD50 and Survival Function Estimation

Prof. Dr. Intisar Araibi Fadaam

Researcher: Batoul Saad Faisal

[1,2]College of Administration and Economics / University of Baghdad

ARTICLE INFO

Received: 24 Dec 2024

Revised: 06 Feb 2025

Accepted: 20 Feb 2025

ABSTRACT

The study aims to evaluate the efficiency of the Logistic regression model in analyzing dose-response data to estimate the effects of different doses on the survival rates of organisms. Four statistical methods were employed to estimate the model parameters: the Maximum Likelihood Estimation using the Newton-Raphson algorithm (MLE N-R) and the Downhill algorithm (MLE D-H), in addition to the Chi-Square Minimization method and the Bayesian method. The results showed that the MLE N-R and MLE D-H methods were the most accurate and stable, providing identical estimates for the median lethal dose (LD50) with consistent and reliable performance across various sample sizes. While the Chi-Square method demonstrated acceptable performance, it was less accurate, particularly when dealing with data that did not conform to the assumed distribution. Conversely, the Bayesian method exhibited variable and unstable performance when estimating median doses for certain variables, making it less reliable. When comparing the survival function, the MLE N-R method produced the highest estimates for survival probabilities, indicating a slower and more gradual decline in survival rates compared to the other methods. On the other hand, the Bayesian method showed the least accuracy, with survival estimates demonstrating a sharp and unexpected decrease. These findings highlight the superiority of MLE N-R as a robust analytical tool for dose-response data analysis.

Keywords: Survival function- Dose-Response, Logistic Model, Logistic Model Transformation, Partial maximum likelihood Estimation, Chi-Square Minimization Method, Bayesian Method, Estimation (LD50), Calculation of the Survival Function, (NR) Algorithm, (DH) Algorithm

Introduction

This study aims to evaluate the efficiency of the Logistic regression model in analyzing dose-response data to estimate the effects of different doses on the survival rates of organisms. This field holds significant importance across various domains, including toxicology, public health, agriculture, and medicine, where such experiments are used to determine safe and lethal doses, aiding in risk assessment and decision-making related to health and the environment.

Later, **C. Ritz (2010)** unified the dose-response modeling approach in environmental toxicology, utilizing four main model categories, including log-logistic and Weibull models, to ensure consistency in data analysis using software like R.

In public health, a study by **Hamer et al. (2008)** demonstrated the positive relationship between physical activity and mental health. It showed that even short durations of physical activity reduce the

risk of mental disorders, supporting recommendations to promote physical activity for better mental well-being .

In the pharmaceutical domain, **Kaneb et al. (2011)** conducted a study on the effects of metformin on Amyotrophic Lateral Sclerosis (ALS). The results revealed that metformin had no positive impact on disease progression in male mice and could be harmful to females at higher doses, emphasizing the need for further research to understand its effects .

Additionally, **Jason Aungst (2012)** provided a comprehensive analysis of dose-response relationships in toxicology, highlighting the importance of addressing uncertainty and biological variability to improve model accuracy and make more informed regulatory decisions .

The Logistic regression model is known for its effectiveness in survival data analysis, particularly when estimating survival probabilities at different dose levels. This model transforms data into a logistic distribution, enabling smooth and accurate estimation of the relationship between dose and response probability (survival or death) for biological data. It also allows for the calculation of the median lethal dose (LD50)—the dose causing death in half of the studied sample serving as a key indicator in dose-response experiments.

Furthermore, the Logistic model excels in handling nonlinear data and determining the effects of varying doses, making it an essential and reliable tool for analyzing experimental results and comparing different factors. This model empowers researchers to draw evidence-based conclusions to support effective decision-making in health and environmental contexts.

1. Methodology

The study was designed to analyze the effects of different doses of substances on survival rates using the Logistic regression model. Simulated data, generated with a normal distribution and various sample sizes, were utilized. Four methods were employed to estimate the model parameters: Maximum Likelihood Estimation using the Newton-Raphson (MLE N-R) and Downhill (MLE D-H) algorithms, the Chi-Square Minimization method, and the Bayesian method. The survival function S(t) was estimated using all four methods to determine the median lethal dose

LD50, with model accuracy assessed through metrics such as Mean Squared Error (MSE) and the Chi-Square test. As an applied component, an agricultural experiment was conducted to examine the effects of three doses of rooting stimulants, NPK fertilizers, and secondary fertilizers on plant growth. Data were analyzed using the Logistic model to identify optimal doses that enhance growth and minimize lethal effects. Simulation and analysis were performed using MATLAB, comparing the statistical methods to identify the most accurate and stable approach for estimating the lethal dose and survival function.

Survival function[13]:

The survival function S(t) represents the probability of an organism surviving beyond a specific time t and is mathematically expressed as:

$S(t) = \Pr(T>t)$

$S(t) = \Pr(T>t) = \int_t^{\infty} p(t)\, dt$

Where T is the random variable representing the time to an event

This function is used to provide a deeper understanding of the distribution of event times and to calculate other metrics such as the hazard rate and the mean survival time. It is widely used in biostatistics and genetics to analyze survival and growth rates.

Dose-Response[8]:

In the fields of toxicology and medicine, the concept of dose-response represents the relationship between the quantity of a drug and its effect on an organism. It aims to determine the optimal dose that achieves therapeutic efficacy with minimal side effects.

Dose-response curves typically take an "S" shape, where the response gradually increases with the dose until reaching a maximum effect (Emax). The median effective dose (ED50) is defined as the dose that achieves 50% of the maximum response, while the median lethal dose (LD50) refers to the dose that causes the death of half the sample within a specified time frame. These measures help in understanding safe and effective dosages.

Logistic Model[1]

The logistic model is one of the primary models used in life sciences research to describe the relationship between the probability of response and the dose level. Mathematically, this model can be expressed as follows:

$$p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 xi)}}$$

Where P(y=xi) represents the probability of a response at a given dose x, β0 is the intercept, and β1 represents the slope or the effect of the dose on the probability of response.

Logistic Model Transformation[3]

The logistic model is used to describe the relationship between the logarithm of the dose (X) and the percentage of response (P). This relationship is represented by an S-shaped curve. The response variable (Y) follows a Bernoulli distribution, where P(xi) represents the probability of success, and 1−P(xi) represents the probability of failure.

The logistic regression model includes two forms:

1. **Model with a single variable:**

The model can be expressed as:

$$\mathbf{P(x)} = \frac{1}{1+e^{-(\beta_0+\beta_1 xi)}}$$

$$\mathbf{1 - P(x)} = \frac{\mathbf{1}}{\mathbf{1}+e^{(\beta_0+\beta_1 xi)}}$$

2. **Model with multiple explanatory variables[4]**

The model can be represented as:

$$P(xi) = \frac{e^{-(\boldsymbol{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p})}}{1+e^{-(\boldsymbol{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p})}}$$

The probability of the event not occurring is:

$$1\text{-}p(xi) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}}$$

The logistic transformation, known as Logit, linearizes the relationship as follows[2]

$$L = \text{logit } P(xi) = \text{Ln}\left[\frac{P(xi)}{1-P(xi)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Thus, the final form of the logistic regression model is:

$$P(xi) = \frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}$$

$$1\text{-}P(xi) = \frac{1}{1+e^{\underline{x'i}\,\underline{\beta}}}$$

Partial Maximum Likelihood Estimation Method[18]

This method involves estimating the model parameters that maximize the likelihood function. For a single observation, assuming the binary response random variable follows a Bernoulli distribution (1,p), the equation can be expressed as follows:

$$P(y/x) = (P(xi))^{yi}\ (1-P(xi))^{1-yi}$$

Where P(y|x) is the probability of the event y when the dependent variable is binary, such as (0,1)

To Estimate Parameters Using the Partial Maximum Likelihood Method, the Following Steps Are Followed:

1. The likelihood function is expressed as[14]

$$\mathbf{L(\boldsymbol{\beta})}=\prod_{i=1}^{n}\left[\frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}\right]^{yi}\left[1-\frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}\right]^{1-yi}$$

2. Taking the natural logarithm of the likelihood function, we get:

$$\mathrm{LnL}(\boldsymbol{\beta})=\sum_{i=1}^{n}\left[yi\ Ln\left(\frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}\right)+(1-yi)Ln\left(1-\frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}\right)\right]$$

Simplifying further:

$$=\sum_{i=1}^{n}\left[yi\ \left(\underline{\boldsymbol{x'i}}\,\underline{\boldsymbol{\beta}}\right)-\ Ln\left(1+e^{\underline{x'i}\,\underline{\beta}}\right)\right]$$

3. The partial derivatives of the log-likelihood function with respect to each parameter are calculated[7]

$$\frac{\partial LnL(\beta)}{\partial \beta_j}(\underline{\boldsymbol{x'i}}\,\underline{\boldsymbol{\beta}})=\frac{\partial}{\partial \beta_j}(\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})$$

$$=x'_{ij}\quad\text{and}\quad x_{i0}=1$$

For the second term:

$$\frac{\partial}{\partial \beta_j}\mathrm{Ln}\left(1+e^{\underline{x'i}\,\underline{\beta}}\right)=\frac{1}{1+e^{\underline{x'i}\,\underline{\beta}}}\frac{\partial}{\partial \beta_j}\left(1+e^{\underline{x'i}\,\underline{\beta}}\right)$$

$$=\mathrm{P(xi)}\ x'_{ij}$$

4. Substituting P(xi) , the equation becomes:

$$\mathrm{L}(\beta)=\sum_{i=1}^{n}\left[\left(yi\ -\frac{e^{\underline{x'i}\,\underline{\beta}}}{1+e^{\underline{x'i}\,\underline{\beta}}}\right)x'_{ij}\right]=0$$

Chi-Square Minimization Method

This method is used to find estimates that minimize Pearson's Chi-Square statistic. The estimated value represents the quantity obtained from comparing the observed frequencies with the expected frequencies calculated using the following mathematical formula[15]

$$\chi^2=\mathrm{R}(\beta)=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{Ei}$$

The Chi-Square statistic R(β) for the binary logistic regression model can be expressed as follows:

$$\mathrm{R}(\underline{\beta})=\sum_{i=1}^{n}\frac{(yi-pi)^2}{pi}+\frac{[(1-yi)-(1-pi)]^2}{1-pi}$$

After simplification, the equation becomes:

$$\mathrm{R}(\underline{\beta})=\sum_{i=1}^{n}\frac{(yi-pi)^2}{pi\,(1-pi)}$$

Substituting the values of pi and (1-pi) , the equation transforms into[14]

$$\mathrm{R}(\underline{\beta})=\sum_{i=1}^{n}\left[{y_i}^2\ e^{-\underline{x'i\beta}}+\ (1-Yi)^2 e^{\underline{x'i}\,\underline{\beta}}-2Yi(1-Yi)\right]$$

To find the estimator β that minimizes R(β), the first derivative is computed and set to zero as follows:

$$\frac{\partial \mathrm{R}(\underline{\beta})}{\partial \underline{\beta}}=\sum_{i=1}^{n}x'_{ij}\left[(1-yi)^2\left(\frac{Pi}{1-Pi}\right)-\ (yi)^2\left(\frac{1-Pi}{Pi}\right)\right]$$

Bayesian Method:

The Bayesian estimation methodology relies on using prior information about the parameters to be estimated, treating them as random variables rather than fixed quantities, as in the previous classical estimation methods. This prior information follows a specific probability distribution, known as the

prior distribution. To determine the prior distribution for the parameter β, the following steps should be followed[11]

The parameter β is considered a random variable following a uniform distribution, which is a special case of the Beta distribution. Hence[12]

$$f(\beta) = \frac{1}{b-a} = \frac{1}{1-0} = 1 \qquad 0 \leq \beta \leq 1$$

The values of y represent the discrete random variable (the dependent variable), which indicates the number of successes and follows a Binomial distribution as expressed below:

$$f(y/\beta) = \frac{n!}{y!(n-y)!} P^y (1-p)^{n-y} \qquad y = 0,1,2,\ldots,n$$

The likelihood function for the Binomial distribution is given as:

$$L(\beta/y) \propto \beta^y (1-\beta)^{n-y}$$

Using Bayes' theorem, the posterior distribution of the parameter β can be determined as:

$$h(\beta/y) = \frac{L(\beta/y).f(\beta)}{\int L(\beta/y).f(\beta)\, d\beta}$$

The posterior distribution of the parameter β is a Beta distribution, expressed as:

$$\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)} \beta^y (1-\beta)^{n-y}$$

The mean of the posterior distribution for the Beta distribution is calculated as:

$$\hat{\beta} = \frac{y+1}{n+2}$$

Where:
$\hat{\beta}$: Represents the Bayesian estimator for the parameter β.

Estimation of the Median Lethal Dose (LD50)[5]

This method represents another type of regression approach where a different transformation of binary data is used, similar to the Probit method. Finney proposed using this method instead of the Probit method when the data distribution is non-normal. In this method, the proportions are transformed using the following function:

$$P(xi) = \frac{e^{(\beta_0+\beta_1 x_i)}}{1+e^{(\beta_0+\beta_1 x_i)}}$$

To determine the effect of the median lethal dose (LD50), we need the dose that makes P(xi)=0.5. The LD50 is calculated using the following steps:

$$0.5 = \frac{e^{(\beta_0+\beta_1 x_i)}}{1+e^{(\beta_0+\beta_1 x_i)}}$$

By solving the equation mathematically, we obtain:

$$Xi = -\frac{\beta_0}{\beta_1}$$

This value represents the dose at which the probability of death is 50%.

Calculation of the Survival Function [6]

The logistic model is commonly used for binary classification:

$$P(xi) = \frac{1}{1+e^{-(\beta_0+\beta_1 xi)}}$$

The survival function S(t) is defined as :

$$S(t) = Pr(T>t)$$

The survival function S(t) can be calculated from the cumulative distribution function F(t). Assuming P(t)=F(t) , meaning the cumulative distribution function is equivalent to the probability density function (p.d.f.), F(t) is given by :

$F(t) = \frac{1}{1+ e^{-(\beta_0+\beta_1 t)}}$

$S(t) = 1- F(t) = 1 - \frac{1}{1+ e^{-(\beta_0+\beta_1 t)}}$

$S(t) = \frac{e^{-(\beta_0+\beta_1 t)}}{1+ e^{-(\beta_0+\beta_1 t)}}$

Newton-Raphson Algorithm[16][17]

The Newton-Raphson algorithm is a well-known technique in numerical mathematics for approximating the roots of a real-valued function. It is used in numerical computation to solve nonlinear equations.

Estimation Steps:

1. Choose initial values for β .

   Begin with initial estimates of the parameter vector β.

2. Calculate the first and second derivatives**:**

   Compute the first derivative (gradient) and the second derivative (Hessian matrix) at each iterative step.

3. Update the estimates**:**

   Update the parameter estimates using the following formula:

   $\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} - F^{(r)}(P^{(r)})^{-1}$

   Here, F represents the gradient vector, and P is the Hessian matrix.

4. Repeat steps 2 to 3:

   Repeat the process until the change in parameter estimates between iterations is smaller than a predefined threshold. The stopping condition is given as:

   $\left|\hat{\beta}^{(r+1)} - \hat{\beta}^{(r)}\right| < d$

   Where d is a very small value (e.g. in the log-likelihood function, less than 10^{-6}).

Downhill Algorithm [10][14]

The Downhill Simplex Algorithm, also known as the Nelder-Mead method, is a robust optimization technique used to find the optimal solution for nonlinear problems without requiring derivative calculations. This method adjusts and moves the geometric shape "Simplex," which consists of n-dimensions and n+1 points in space, to evaluate and improve the objective function.

Basic Steps in Parameter Estimation Using the Downhill Algorithm[9]

1. Define the objective function for the algorithm, represented by the partial likelihood function

   f_z = - Ln L(z) where z =(β)

2. Set the four algorithm parameters**.** In most studies, these parameters are set to (σ=−0.5), (γ=0.5), (ϵ=2), and (α=1).

3. Generate a matrix consisting of **n+1** test points for each parameter in the function.

4. Substitute the values of each row from the above matrix into the objective function, calculate its values, and then sort the resulting objective functions (, where the total number is n:

   $f_{z1} \leq f_{z_2} \leq \cdots \leq f_{z_{n+1}}$

5. Calculate the mean of the solution matrix:

$\overline{w}=\frac{\sum_{i=1}^{n}\beta_i}{n}$ where i = 1,2,...,n

6. Find a new test point called the Reflection Point using the following formula:

   $Z_r = \overline{w} + \alpha\ (\overline{w} - Z_{n+1})$

   After calculating the reflection, compute the objective function (f_{z_r}) . If:

   $f_{z_1} < f_{z_r} < \cdots < f_{z_n}$

   set $z_{n+1} = z_r$ and move to step (10). Otherwise, proceed to the next step

7. Find a new test point called the(Expansion Point : e) using the following formula :

   $Z_e = \overline{w} + \varepsilon\ (Z_r - \overline{w}\ )$

   After calculating the expansion, compute the objective function (f_{z_e}). If $f_{z_e} < f_{z_r}$, set $Z_e = Z_{n+1}$ and move to step (10). Otherwise, proceed to the next step.

8. Find a new test point called the (Contraction Point: c) using the following formula:

   $Z_c = \overline{w} + \gamma(Z_{n+1} - \overline{w}\ )$

   After calculating the contraction, compute the objective function (f_{z_c}). If $f_{z_c} < f_{z_n}$, set $Z_c = Z_{n+1}$ and move to step (10). Otherwise, proceed to the next step.

9. Find a new test point called (the Shrink Point: sh) using the following formula:

   $Z_{sh} = Z_1 + \sigma\ (Z_i - Z_1)$

10. Check the stopping condition. This step is applied when the stopping condition is satisfied, defined by the following formula

    $\left|\frac{\max(f)-\min(f)}{\max(f)}\right| < \in$

    where $\in$ is a very small number.

    If the stopping condition is met, the optimal solution is printed; otherwise, return to step 6.

Experimental Side:

Simulation is an effective tool for mimicking real-world systems, facilitating the study of complex processes and providing approximate solutions that closely resemble reality. Simulation relies on generating samples of varying sizes to enhance understanding of the systems under investigation. In this study, the Monte Carlo method was used to generate data, focusing on comparing estimation methods for the Logistic model, including: the Maximum Likelihood Estimation (using Newton-Raphson and Downhill algorithms), the Chi-Square Minimization method, and the Bayesian Estimator, with the aim of identifying the most suitable method for estimating the lethal dose (LD50).

The simulation was implemented using (MATLAB 2022a), where doses were randomly generated from a uniform distribution (0,1), and the data was then modeled based on the normal distribution using assumed mean and standard deviation values.

Estimation of Data in the Case of Fewer Than True Data

The assumed parameter values (β) for the first model are as follows: (0.1100; 0.1; 0.1300; 0.0108), which are used as starting points for estimation in the Logistic model.

Table (1): Simulation Results for Sample Size 34 (Less Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.1100	0.1000	0.1300	0.0108	1.1000	0.8462	10.1852	0.0130	0.7368
	'MLE D-H'	0.1130	0.1027	0.1335	0.0111	1.1000	0.8462	10.1852	0.0130	0.7364
	'Chi2 '	0.4611	0.1367	0.2585	0.1536	6.3623	5.7557	4.5477	0.0416	1.7724
	'Bayes'	0.1135	0.1031	0.1352	0.0308	1.2083	0.8651	16.8498	0.0144	0.8037

Note: The value of Chi² table = 47.3999, and sample size n=34.

Interpretation:
The results of MLE N-R and MLE D-H are the most accurate, with the lowest MSE (0.0130) and low Chi-square values (0.7368 and 0.7364). The Bayes method performed relatively well (MSE = 0.0144), while the Chi² method was the least accurate (MSE = 0.0416 and Chi-square = 1.7724)

Table (2): Simulation Results for Sample Size 54 (Less Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.1100	0.1000	0.1300	0.0108	1.1000	0.8462	10.1852	0.0129	1.1625
	'MLE D-H'	0.1153	0.1048	0.1363	0.0113	1.1000	0.8462	10.1852	0.0130	1.1694
	'Chi2 '	0.3115	0.1457	0.1898	0.0830	3.1756	3.9721	17.2265	0.0259	1.9329
	'Bayes'	0.1139	0.1096	0.1342	0.0251	1.1441	0.8672	13.6451	0.0138	1.2241

Note: The value of Chi² table = 70.9935, and sample size n=54.

Interpretation:
MLE N-R and MLE D-H are the most accurate and stable methods, with the lowest MSE (0.0129 and 0.0130) and low Chi-square values (1.1625 and 1.1694). MLE N-R is slightly preferred due to the smallest MSE. The Bayes method performed well (MSE = 0.0138 and Chi-square = 1.2241), while the Chi² method was the least accurate (MSE = 0.0259 and Chi-square = 1.9329).

Table (3): Simulation Results for Sample Size 74 (Less Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.1100	0.1000	0.1300	0.0108	1.1000	0.8462	10.1852	0.0129	1.5981
	'MLE D-H'	0.1172	0.1066	0.1385	0.0115	1.1000	0.8462	10.1852	0.0131	1.6021
	'Chi2 '	0.2783	0.1364	0.1739	0.0843	3.6442	5.0898	8.1830	0.0229	2.4229
	'Bayes'	0.1171	0.1015	0.1274	0.0322	1.3415	0.9443	9.4854	0.0141	1.7184

Note: The value of Chi² table = 93.9453, and sample size n=74.

Interpretation:
MLE N-R and MLE D-H are the most accurate, with the lowest MSE (0.0129 and 0.0131) and low Chi-square values (1.5981 and 1.6021). MLE N-R is slightly preferred due to the smallest MSE. The Bayes method performed well (MSE = 0.0141 and Chi-square = 1.7184), while the Chi² method was the least accurate (MSE = 0.0229 and Chi-square = 2.4229).

Estimation of Data in the Case of True Data

Assuming the parameter values (β) for the first model, defined as follows: (0.2100; 0.1292; 0.2300; 0.1108), which serve as starting points for the estimation process in the Logistic model.

Table (4): Simulation Results for Sample Size 34 (Equal to True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.2100	0.1292	0.2300	0.1108	1.6254	0.9130	1.8953	0.0094	0.4342
	'MLE D-H'	0.1573	0.0968	0.1723	0.0830	1.6254	0.9130	1.8953	0.0118	0.6022
	'Chi2 '	0.3398	0.1425	0.1964	0.1264	9.0593	6.7676	11.7109	0.0114	0.5066
	'Bayes'	0.2158	0.1385	0.2352	0.1136	1.7667	0.9255	2.1250	0.0110	0.5061

Note: The value of Chi² table = 47.3999, and sample size n=34

Interpretation:
MLE N-R is the most accurate, with the lowest MSE (0.0094) and Chi-square (0.4342), followed by Bayes (MSE = 0.0110, Chi-square = 0.5061). MLE D-H and Chi² methods showed less accuracy, with higher MSE values (0.0118 and 0.0114) and relatively higher Chi-square values.

Table (5): Simulation Results for Sample Size 54 (Equal to True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.2100	0.1292	0.2300	0.1108	1.6254	0.9130	1.8953	0.0096	0.7107
	'MLE D-H'	0.1638	0.1008	0.1794	0.0864	1.6254	0.9130	1.8953	0.0112	0.8775
	'Chi2 '	0.3012	0.1420	0.2598	0.1185	4.2343	1.8292	8.7036	0.0142	0.9787
	'Bayes'	0.2178	0.1405	0.2373	0.1250	1.7032	0.9231	1.8591	0.0111	0.8028

Note: The value of Chi² table = 70.9935, and sample size n=54.

Interpretation:
MLE N-R is the most accurate, with the lowest MSE (0.0096) and Chi-square (0.7107), followed by Bayes (MSE = 0.0111, Chi-square = 0.8028). MLE D-H showed acceptable performance (MSE = 0.0112), while Chi² was the least accurate (MSE = 0.0142, Chi-square = 0.9787).

Table (6): Simulation Results for Sample Size 74 (Equal to True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.2100	0.1292	0.2300	0.1108	1.6254	0.9130	1.8953	0.0094	0.9536
	'MLE D-H'	0.1604	0.0987	0.1757	0.0846	1.6254	0.9130	1.8953	0.0107	1.1603
	'Chi2 '	0.3162	0.1063	0.1945	0.0842	5.8423	2.5785	17.0344	0.0095	0.9725
	'Bayes'	0.2275	0.1538	0.2510	0.1318	1.5220	0.9148	1.7716	0.0109	1.0654

Note: The value of Chi² table = 93.9453, and sample size n=74

Interpretation:
MLE N-R is the most accurate, with the lowest MSE (0.0094) and Chi-square (0.9536), followed closely by Chi² (MSE = 0.0095, Chi-square = 0.9725). MLE D-H and Bayes showed lower accuracy, with relatively higher MSE and Chi-square values.

Estimation of Data in the Case of More Than True Data

The assumed parameter values (β) for the third model are as follows: (0.3100; 0.2292; 0.3300; 0.2108), which are used as starting points for the estimation process in the Logistic model.

Table (7): Simulation Results for Sample Size 34 (Greater Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.2198	0.2292	0.2003	0.2108	0.9591	1.1072	1.0428	0.0059	0.2581
	'MLE D-H'	0.1714	0.1778	0.1563	0.1636	0.9591	1.1072	1.0428	0.0116	0.5362
	'Chi2 '	0.3429	0.2897	0.2246	0.1224	1.1457	1.4464	20.7701	0.0042	0.1795
	'Bayes'	0.3034	0.2500	0.3746	0.2438	1.2154	0.8099	1.2461	0.0061	0.2422

Note: The value of Chi² table = 47.3999, and sample size n=34.

Interpretation:
Chi² is the most accurate, with the lowest MSE (0.0042) and Chi-square (0.1795), followed by MLE N-R (MSE = 0.0059), and then Bayes (MSE = 0.0061). MLE D-H was the least accurate, with the highest MSE (0.0116) and a relatively high Chi-square value (0.5362).

Table (8): Simulation Results for Sample Size 54 (Greater Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.1871	0.2239	0.1554	0.2101	0.8314	1.2246	0.8896	0.0074	0.5153
	'MLE D-H'	0.1359	0.1629	0.1128	0.1529	0.8314	1.2246	0.8896	0.0180	1.3490
	'Chi2 '	0.3186	0.2120	0.3126	0.2100	1.9838	2.6463	1.7663	0.0040	0.2571
	'Bayes'	0.3154	0.2438	0.3541	0.2249	1.3062	0.8959	1.4322	0.0046	0.2938

Note: The value of Chi² table = 70.9935, and sample size n=54.

Interpretation:
Chi² is the most accurate, with the lowest MSE (0.0040) and Chi-square (0.2571), followed closely by Bayes (MSE = 0.0046, Chi-square = 0.2938). MLE N-R showed acceptable performance, while MLE D-H was the least accurate.

Table (9): Simulation Results for Sample Size 74 (Greater Than True Data) with 1000 Replications

Models	Methods	est B0	est B1	est B2	est B3	X1LD50	X2LD50	X3LD50	MSE	Chi2 cal.
Logistic	'MLE N-R'	0.2074	0.2292	0.1781	0.2108	0.9048	1.1760	0.9838	0.0062	0.5833
	'MLE D-H'	0.1573	0.1736	0.1353	0.1596	0.9048	1.1760	0.9838	0.0134	1.3460
	'Chi2 '	0.1774	0.2598	0.3195	0.2133	0.7653	0.6645	0.8625	0.0049	0.4420
	'Bayes'	0.3245	0.2853	0.3397	0.2360	1.1371	0.9671	1.3771	0.0063	0.5459

Note: The value of Chi² table = 93.9453, and sample size n=74.

Interpretation:
Chi² is the most accurate, with the lowest MSE (0.0049) and Chi-square (0.4420), followed by MLE N-R (MSE = 0.0062) and Bayes (MSE = 0.0063), both showing acceptable performance. MLE D-H was the least accurate, with the highest MSE (0.0134) and a relatively high Chi-square value.

Application Side:

An agricultural experiment was designed to study the effect of three different doses of a rooting stimulant, NPK fertilizer, and secondary fertilizer on plant growth using the Logistic model. The analysis focused on the dry weight of the vegetative parts as the response variable(y)

to estimate the optimal doses that enhance growth and determine the lethal dose (LD50).

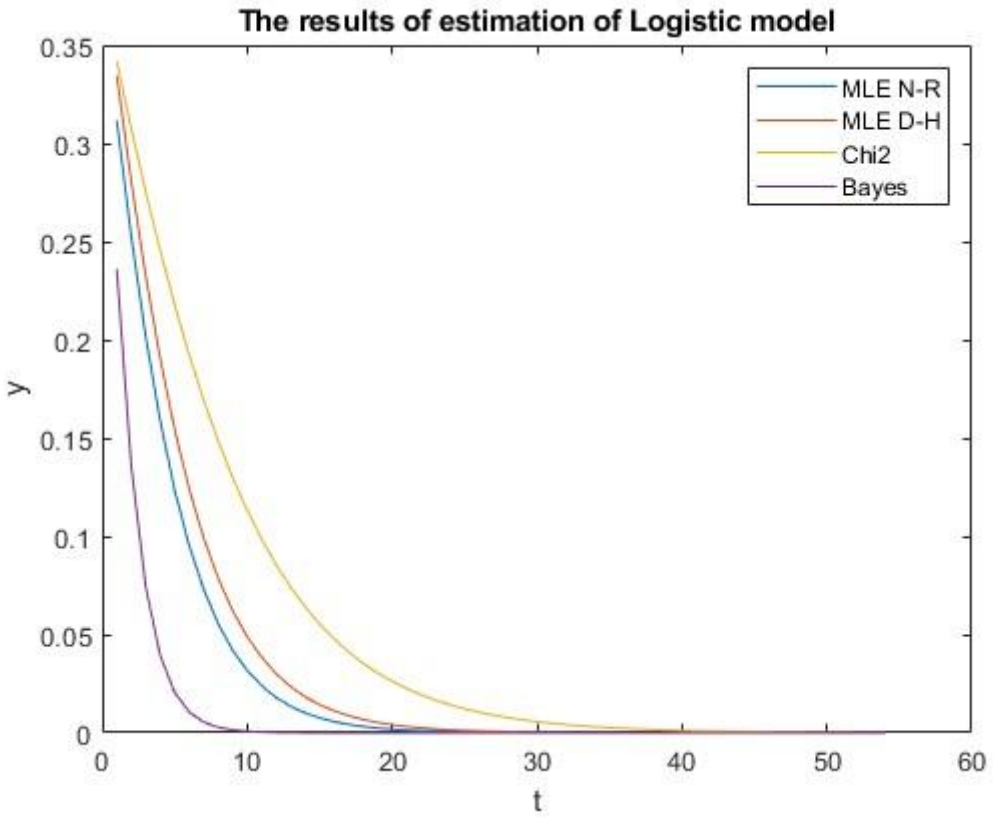
Data were collected and analyzed to determine the relationship between doses and response, emphasizing the doses that achieve maximum growth and prolonged plant survival. Mean Squared Error (MSE) and the Chi-Square test were used to evaluate the accuracy of the model and select the optimal doses.

Logistic Model Estimation

In the logistic regression model, the model parameters were estimated using various methods to identify the optimal approach, as shown in the table.

Table 10 : Estimated Parameters of the Logistic Model for True Data

Methods	est B0	est B1	est B2	est B3
'MLE N-R'	0.5000	0.3000	0.4759	0.1000
'MLE D-H'	0.4340	0.2604	0.4131	0.0868
'Chi2 '	0.2075	0.1292	0.2270	0.1113
'Bayes'	0.5251	0.3060	1.6140	0.1032



. Figure(1) illustrates the parameter estimates of the Logistic model

From the table and figure, MLE N-R and MLE D-H showed similar performance, with a slight advantage for MLE N-R in terms of faster convergence and higher accuracy. The Chi-Square method exhibited acceptable performance but was slightly less accurate. Meanwhile, the Bayes method was the least accurate due to its significant deviation from the other methods.

Estimation of the Survival Function $\hat{S}(t)$ for All Cases

The survival function was estimated for the data using four methods, as shown in the following table:

Table (11) : Survival Function Estimation for the Logistic Model

T	St_MLE_NR	St_MLE_DH	St_Chi2	St_BAY	T	St_MLE_NR	St_MLE_DH	St_Chi2	St_BAY
1	0.311744255	0.33461483	0.341673765	0.236065358	28	0.000170711	0.00053663	0.007663159	3.82213E-09
2	0.252758362	0.280738022	0.307530813	0.136020074	29	0.000127499	0.000416552	0.006564567	1.94728E-09
3	0.201663132	0.232504776	0.275372564	0.074253152	30	9.5211E-05	0.000323334	0.005622577	9.92092E-10
4	0.158703117	0.190365154	0.245385439	0.039260133	31	7.11039E-05	0.000250972	0.004815103	5.05446E-10
5	0.123479312	0.154327264	0.217682833	0.020394834	32	5.31003E-05	0.000194801	0.004123113	2.57512E-10
6	0.0951888	0.124066249	0.192310669	0.010495672	33	3.96551E-05	0.0001512	0.003530217	1.31196E-10
7	0.072841332	0.099043973	0.169255962	0.005374936	34	2.96141E-05	0.000117357	0.003022332	6.68412E-11
8	0.055419014	0.078615374	0.148457104	0.002745649	35	2.21156E-05	9.10883E-05	0.002587305	3.40541E-11
9	0.041975147	0.062109861	0.129814709	0.001400728	36	1.65157E-05	7.06991E-05	0.002214764	1.73497E-11
10	0.031683181	0.048885863	0.113202112	0.000714127	37	1.23337E-05	5.48735E-05	0.001895763	8.83915E-12
11	0.023851894	0.038362263	0.098474873	0.000363958	38	9.21067E-06	4.25902E-05	0.001622634	4.50329E-12
12	0.017920481	0.030032532	0.085478932	0.000185461	39	6.87841E-06	3.30564E-05	0.001388801	2.29439E-12
13	0.013443759	0.023467327	0.074057182	9.44963E-05	40	5.13671E-06	2.56567E-05	0.001188625	1.16884E-12
14	0.010073897	0.018310204	0.064054727	4.81458E-05	41	3.83602E-06	1.99134E-05	0.001017272	5.95524E-13
15	0.007542279	0.014269837	0.055322526	2.45297E-05	42	2.86469E-06	1.54557E-05	0.000870599	3.03313E-13
16	0.005643242	0.011110935	0.047720004	1.24974E-05	43	2.13931E-06	1.19959E-05	0.000745059	1.54543E-13
17	0.0042203	0.0086451	0.04111682	6.36716E-	44	1.59761E-	9.31058E-	0.0006376	7.88258E-

	22	84	4	06		06	06	1	14
18	0.003155046	0.006722914	0.035393362	3.24392E-06	45	1.19307E-06	7.22636E-06	0.000545648	4.01901E-14
19	0.002358027	0.005225812	0.030441314	1.6527E-06	46	8.90967E-07	5.6087E-06	0.000466944	2.04281E-14
20	0.001761992	0.004060073	0.026163336	8.42012E-07	47	6.65361E-07	4.35317E-06	0.000399587	1.04361E-14
21	0.001316417	0.003154577	0.022472617	4.28985E-07	48	4.96882E-07	3.37868E-06	0.000341944	5.32907E-15
22	0.000983409	0.002450135	0.019292213	2.18557E-07	49	3.71065E-07	2.62235E-06	0.000292613	2.66454E-15
23	0.000734579	0.0019027	0.016554289	1.1135E-07	50	2.77106E-07	2.03532E-06	0.000250397	1.33227E-15
24	0.000548675	0.001477398	0.014199301	5.67299E-08	51	2.06938E-07	1.5797E-06	0.000214271	6.66134E-16
25	0.000409799	0.001147052	0.012175183	2.89025E-08	52	1.54539E-07	1.22607E-06	0.000183356	4.44089E-16
26	0.000306064	0.000890506	0.010436549	1.47251E-08	53	1.15407E-07	9.5161E-07	0.0001569	2.22045E-16
27	0.000228582	0.000691298	0.008943947	7.50208E-09	54	8.61844E-08	7.38586E-07	0.000134261	0

The survival function was estimated using four methods: MLE N-R, MLE D-H, Chi-Square, and Bayes. MLE N-R demonstrated the best performance with the highest survival estimates (0.3117) and a less steep decline. MLE D-H ranked second with performance close to MLE N-R. The Chi-Square method showed moderate performance, with a faster decline in the survival function, while the Bayes method was the least accurate, showing very low survival rates and a rapid decline, reflecting its limited efficiency compared to the other methods.

Estimation of the Lethal Dose (LD50)

The lethal dose causing the death of half of the group for the Logistic model is presented in the following table:

Table(12): Estimation of Median Lethal Dose (LD50)

Methods	X1LD50	X2LD50	X3LD50
'MLE N-R'	1.6667	1.0506	5.0000
'MLE D-H'	1.6667	1.0506	5.0000
'Chi2 '	1.6056	0.9138	1.8643
'Bayes'	1.7161	0.3253	5.0901

The analysis of LD50 estimates indicated that the MLE N-R and MLE D-H methods were the most accurate and stable, providing identical and reliable estimates. The Chi-Square method showed less accuracy, particularly when the data did not align with the assumed distribution. The Bayes method

exhibited variability, with unstable estimates for X2, though it was acceptable for X1 and X3. Thus, MLE N-R and MLE D-H are the preferred methods for LD50 estimation.

Conclusions:

1. The Logistic regression model demonstrated a high capability in analyzing dose-response data and accurately estimating median lethal doses (LD50), making it suitable for applications in health, environmental studies, and agriculture.

2. The MLE N-R and MLE D-H methods were the most accurate and stable, providing consistent results and reliable estimates across different sample sizes. The Chi-Square method showed acceptable performance but was less accurate, particularly when dealing with data that did not conform to the assumed distribution. Meanwhile, the Bayes method exhibited significant variability and less stable estimates, making it less effective for complex data.

3. The MLE N-R method achieved the highest survival probabilities with a more gradual decline compared to other methods, while the Bayes method was the least accurate due to steep drops in survival probability.

4. The MLE N-R and MLE D-H methods were the most accurate in estimating median lethal doses, reflecting their superior stability compared to other methods.

Recommendations:

1. It is recommended to use the MLE N-R and MLE D-H methods for estimating the parameters of the Logistic regression model due to their high accuracy and stability in analyzing dose-response data.
2. The performance of the Bayes method can be improved by using precise prior distributions that align with the nature of the data.
3. It is advised to apply the Logistic model in dose-response studies, especially when dealing with nonlinear data, given its accuracy and flexibility in analyzing dose effects.
4. Additional studies involving other models such as Probit and Cox are recommended for a comprehensive comparison of performance.
5. The findings are suggested to be used to enhance agricultural, environmental, and health dose strategies, achieving a balance between effectiveness and risk reduction.

References

a. Abdul-Razzaq, Mohammed Sadiq, and Za'lan, Risan Abdul-Imam (2016). "Using Logistic Regression Method to Analyze the Effect of Psychological Stress on Hypertension," Basra, Gulf Economy Journal, pp. 48–66.

[2] Agresti, A., (2002). "Categorical Data Analysis," 2nd edition, John Wiley & Sons Inc., Hoboken, New Jersey.

[3] Al-Azzawi, Ahmed Diyab (2005). "Comparison Between Some Methods of Estimating the Logistic Regression Model and Robust Methods for Binary Response Life Experiments Using Simulation." Master's Thesis submitted to the College of Administration and Economics, University of Baghdad.

[4] Alrahamneh, A. & Hawamdeh, O., (2017). "The Factors Affecting Eye Patients (Cataract) In Jordan by Using the Logistic Regression Model," MAS, pp. 38-42.

[5] Al-Samarai, Dr. Firas Rasheed, July (2020). "Description of Dose Response Curve," Academia.edu, (P.P.4, 21).

[6] Collett, D., (2015). "Modelling Survival Data in Medical Research," 2nd ed., CRC Press.

[7] Cramer, J. S., (2003). "Logit Models From Economics and Other Fields," Cambridge University Press, Cape Town, New York, ISBN, pp. 33-45.

[8] Crump, K. S.; Hoel, D. G.; Langley, C. H.; Peto, R. (1976). "Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment," Cancer Research 36 (9 Part1): 2973–9. PMID 975067.

[9] Dawood, M., (2022). "Estimating the Reliability of Degradation Data That Follows the Wiener Process with Practical Application," pp. 1-89.

[10] Gal'antai, A. (2021). "A Convergence Analysis of the Nelder-Mead Simplex Method," Acta Polytechnica Hungarica, Vol. 18, No. 5, pp. 1-13.

[11] Gelman, A. & et al., (2014). "Bayesian Data Analysis," Texts in Statistical Science, CRC, LC, Chapman and Hall Book.

[12] Harrison, P. J. & Stevens, C. F., (1971). "A Bayesian Approach to Short-Term Forecasting," Operation Research, Vol. 22, No. 2, pp. 341-362.

[13] Hastings, N. & Evans, M. & Peacock, B., (2000). "Statistical Distributions," 3rd ed., New York: Wiley, p. 13.

[14] Hosmer, D., Lemeshow, S. & Sturdivant, R., (2013). "Applied Logistic Regression," 3rd edition, New York: Wiley, WSIPS, http://ihmsi.org.

[15] Hussain, J. N. & Nassir, A. J., (2015). "Cluster Analysis as a Strategy of Grouping to Construct Goodness-Of-Fit Tests When the Continuous Covariates Present in the Logistic Regression Models," BJMCS, pp. 1-16.

[16] Matilainen, K., (2013). "Employing a Monte Carlo Algorithm in Newton-Type Methods for Maximum Likelihood Estimation of Genetic Parameters," PLOS ONE, pp. 1-7.

[17] Mccullagh, P., & Nelder, J., (1983). "Generalized Linear Models," London: Chapman and Hall.

[18] Menard, S., (2002). "Applied Logistic Regression Analysis," 2nd Edition, Thousand Oaks, CA: SAGE Publications, Series Quantitative Applications in the Social Sciences.