

A Comparison To Estimate The Parameters Of The Logistic Regression Model By The Maximum Likelihood Estimation And The Firefly Algorithm" A Field Study About Projections In Pregnant Women"

Waleed Ahmed Hassen

University of Diyala \ College of Education for Pure Sciences.
purecomp.waleed.hassan@uodiyala.edu.iq

Received 7 May.2020. Accepted 28 May 2020. Published 21 June 2020.

Abstract

Frequent miscarriages in pregnant women are considered one of the reasons for losing fetuses in the first trimester of pregnancy and at the end of pregnancy. Many studies have tried to find the real reasons associated with this case and to find the real reason behind it. The researchers wanted to study the case to find out one of the real reasons. A sample (size 110) was taken from pregnant women as follows They went for the purpose of examination to medical laboratories, and the following was done by the vitamin D deficiency test. The estimate was done in two ways: the classical method, the greatest possibility, and the second method, the Gravity Search Algorithm (GSA).

Keywords:

Introduction

Recent studies in the health field that emphasize prevention is better than treatment have begun to search for the causes that lead to diseases. One of these cases, which are present and prevalent among pregnant women, are the frequent miscarriages of the pregnant woman, which poses a risk to pregnant women and the loss of the fetus that stops growing in the first three months of pregnancy. Many studies Local and global, which examined the causes that lead to the situation.

In this study, the researchers tried to study the relationship of vitamin D deficiency with frequent miscarriages of women, knowing that there are several reasons for this situation.

- 1- anatomical reasons,
- 2- Glands dysfunction and hormonal imbalance
- 3- The mother's age is over 35 years old
- 4- Causes related to the immune system,

There are many other causes, some of which are known and others are unknown.

In this study, the researchers estimated the parameters of the binary logistic regression model in two ways: the first method is a classical method, which is the greatest possibility (mle) and the (GSA) algorithm in order to compare the two methods and find out which of the two methods is more efficient.

The study includes

- 1- What is the best possible way?
- 2- What is the GSA algorithm?
- 3- The practical side
- 4- Discussing the results
- 5- Recommendations and conclusions

Previous Studies

Including a study in 2011 conducted by (Nawas Al-Salihi and Saad Al-Araji) from the University of Babylon for a sample of pregnant women, where the results showed that (31%) of the sample was the result of an antibody (toxoplasma) and that (17%) of the same sample was due to cytomegalovirus, and that 6.8% of women have diabetes.

In 2016, the researcher (Nada Mahouder Al-Khalili) from the University of Al-Qadisiyah, the researcher found that there is a relationship between anti-cardiolipin and recurrent miscarriages for women, as anti-cardiolipins were present in 11.9% of the patients in the study sample, and the highest rates were found in women who had recurrent miscarriages

In 2016 (Nada Al-Ibrahimi and Bushra Abdel-Amir) from the University of Al-Qadisbeh published a research entitled (The Effect of Insulin Level on Recurrent Pregnancy Loss) and their results were a high level of insulin in the blood for women who had previous recurrent miscarriages compared to women who did not have previous recurrent miscarriages

(A. Schreiber Courtney) and others in 2018 studied the effect of some medications on recurrent miscarriages and found that there are some medications whose side effects are the occurrence of recurrent miscarriages.

(Michael R. Strug) et al. in 2018 studied early shedding and its relationship to genetic genes and found a simple relationship between genotypes and recurrent miscarriages.

Concept Regression Logistic

Logistic regression hypothesizes that the dependent variable is a binary response variable (0,1) that is, when the response occurs, the dependent variable will have its value (1) with a probability (P) and when no response occurs, the value of the dependent variable will be (0) with a probability (q) and the simple logarithmic regression contains one independent variable and the relationship is

$$\log(P/q) = b_0 + b_1 x$$

This model is called the logistic regression model, and the logistic function is a continuous function that takes the values (1-0) and approaches (y) from zero as the right side of the logistic function approaches from (∞ -) and (y) approaches one as the right side of this function approaches from (∞), This function when the right side of this function is equal to zero, and the ratio (q/p) is called the success rate of success or the priority of the desired event, and the ratio (q/p) is called the failure priority (Failure of (Odd) and that the expression (q/pLog) is called the logarithm of the ratio of preference (Ratio Odds Log) In short, the logistic regression model is simply a logarithmic transformation of linear regression and therefore it is appropriate to use the properties of the logistic distribution that constrains the estimated probabilities It makes it confined between (zero and one). It is necessary to point out that the estimation of the parameters of the logit model is done by the probability maximum, which is one of the most famous methods of estimation in statistics and the measurement of the possibility function. The greatest (L.M) probabilities observed for a number (n) of the independent variables (Pn,, P2, P1) that fall in the sample and the product of these probabilities represents the function of greatest possibility.

Maximum Likelihood Estimation Method

Estimated maximum likelihood estimation of Poisson regression parameters (MLE) method. Maximum likelihood estimation data result It seeks an answer to the question of what values the regression coefficients can take.

The maximum likelihood estimation depends on a likelihood function. This function describes the probability of seeing data as a function of the parameter set., poisson regression uses the poisson distribution as the probability model and the regression coefficients defines the parameters that determine the average structure of the data. maximum likelihood The purpose of the method is to find the regression coefficients that

maximize the likelihood function.

is to guess. Equalizing the first derivative of this probability equation to zero and solving for the regression coefficients becomes possible by.

In the most practical situations, maximum likelihood estimation requires iterative processes.

This adds extra complexity to these models. In particular, many parameters and Complex models with small sample sizes prevent the process from converging.

Finally, the results of the maximum likelihood estimation are for the regression coefficient. returns asymptotic standard errors. Calculate the maximum likelihood for the Poisson regression.

For discussion, let (μ_i) be the mean of the i th result variable, with $i = 1, 2, \dots, n$.

The mean of the outcome variable; One of the set of descriptive variables X_1, X_2, \dots, X_k Since it is assumed as a function of $\mu(X_i, \beta)$ notation; mean (μ_i) X_i (i . th value of the explanatory variable) and β (regression coefficients) is used to associate Probability density function (pdf), for random variable y θ $f(y|\theta)$ is shown with a set of parameters. This function is data creation.

defines the process. This process forms the basis of the observed sample data; same It also provides a mathematical explanation of the data that the process will create. This Joint density of n independent variables and ideally distributed observations in the process is the product of individual densities:

$$f(y_1, \dots, y_n | \theta) = \prod_{i=1}^n f(y_i | \theta) = L(\theta | y)$$

This is the joint density likelihood function. With the unknown parameter vector θ is a defined function. Here, y is the collection of sample data.

used to indicate

$$\hat{\beta}_{MLE} = (X' \hat{W} X)^{-1} (X' \hat{W} Z)$$

1.3 Firefly **Algorithm Algorithm** :

Firefly was presented by Rashedi. In(2009) and is planned to tackle improvement issues. The populace put together heuristic calculation is based with respect to the law of gravity and mass connections. The calculation is involved assortment of searcher specialists that associate with one another through the gravity power. The specialists are considered as items and their exhibition is estimated by their masses. The gravity power causes a

worldwide development where all articles move towards different items with heavier masses. (Gu.,et al, 2014).

Firelty is a heuristic stochastic multitude based pursuit calculation in the field of mathematical streamlining, in light of the gravitational law and laws of movement. In the same way as other nature roused calculations, it needs refinements to augment its presentation in tackling different kinds of issues. Notwithstanding the issue encoding that occasionally can be a test, tweaking its boundaries assume a critical part adjusting the inquiry time versus arrangement quality. This calculation is somewhat later and not intensely investigated.

The articles masses are complying with the low of gravity as following:

$$F = G \frac{M_1 M_2}{R^2}$$

EQUATION REPRESENTS the Newton law of gravity where

F is magnitude of the gravitational force

G IS gravitational constant

M₁ IS THE MASS OF THE FIRST OBJECT

M₂ IS THE MASS OF THE second OBJECT

R is the distance between the two objects M₁,M₂

Practical Part

The following model has been evaluated:

$$\log(P/q) = \hat{b}_0 + \hat{b}_1 x$$

The parameters of the two models were tested using both (MSE & MAPE) with the following formulas:

$$MSE = \frac{\sum_{r=1}^R SE_i}{R} = \frac{\sum_{r=1}^R (\hat{\beta} - \beta)'_r (\hat{\beta} - \beta)_r}{R}$$

r=1,2,3,4,5,.....100

$$MAPE = \left\{ \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \right\} * 100$$

where A_t is the actual value and F_t is the forecast value

The difference between A_t and F_t is divided by the Actual value A_t again. The absolute value in this calculation is summed for every forecasted point in time and divided by the

number of fitted points n . multiplying by 100 makes it a percentage error.

In two ways, as shown in the following table:

N	Method		B_0	B_1
110	MLE	Parametre	0.32	0.29
		MSE	0.90	0.52
		MAPE	4.35	2.36
	Firelty	PARAMETER	0.46	0.31
		MSE	0.002	0.007
		MAPE	0.12	0.91

Conclusions:

- 1- The researcher found that the estimation using the algorithms method was more accurate than the classical method, the greatest possibility.
- 2- The results of the mean error of the algorithm method were the lowest

Recommendations:

- 1- Using Algorithms as a method for estimating
- 2 -Using modern techniques in statistics in forecasting and estimating

References:

Myung, J., (2002), Tutorial On Maximum Likelihood Estimation, Journal of Mathematical Psychology, pp.93-97

Rashedi E, Nezamabadi-Pour H, Saryazdi S (2010) BGSA: Binary Gravitational Search Algorithm. Natural Computation, 9:727–745

Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S., (2009). GSA: A Gravitational Search Algorithm, Information Sciences, Vol. 179, No. 13, p. 2232–2248

Wang, W., Fameo, F., (1997). Modeling Household Fertility Decisions With Generalized Poisson Regression” Department of Economics, University of Kentucky, Lexington, USA

Khafa H. Ghali, Naief H. Al-Mutairi: Government Spending in Adynamic model of Endogenous Growth, With Application to the State of Kuwait, Arab Journal of Administrative Sciences, Vol.9, No.3, 2002, Kuwait

Omar A. Ibrahim Anar (Public spending track and its contribution to achieving economic growth in Iraq for the period (2004-2015) Master’s thesis
University of Fallujah - Iraq

Niloy Bose, M. Emranul Heque, Denise. R. Osbern : "Public Expenditure and Economic

Webology (ISSN: 1735-188X)

Volume 17, Number 3, 2020

Growth - A disaggregated Analysis for Developing Countries", Journal Compliation, The Manchester University, Vol. 75, No.5, 2007.