RESEARCH ARTICLE

MAXIMUM LIKELIHOOD BASED ON NEWTON RAPHSON, FISHER SCORING AND EXPECTATION MAXIMIZATION ALGORITHM APPLICATION ON ACCIDENT DATA.

Switamy Angnitha Purba, Sutarman and Open Darnius.
Faculty of Math and Science Universitas Sumatera Utara.

………………………………………………………………………………………………………......

Manuscript Info	*Abstract*
……………………..	………………………………………………………………………

The parameter estimate is the value of the parameter based on data or samples taken from a certain popolation. There are several methods to estimate the parameters of one of them is Maximum Likelihood Estimation (MLE). MLE is a distribution approach by maximizing likelihood function. The purpose of this study is to estimate the parameter value of a data distributed with Maximum Likelihood based on the iteration algorithm. The iteration algorithm that will be used is Newton Raphson, Fisher Scoring and Expectation Maximization Algorithm with the help of Matlab 2016a. The purpose of this paper is to look at the parameter values of three algorithms that have the same results or have great results and with regard to the number of iterations performed by the three algorithms. In this paper the three algorithms will be applied to the accident data.

………………………………………………………………………………………………………......

Introduction:-
Parameter estimation is an assessment of population parameter values such as mean, standard deviation, proportion, etc. Parameter estimation is based on data or samples taken from a population. Estimating the parameters of a population is one of the important things in statistical inference. There are several methods to estimate generalized linear model parameters, such as Maximum Likelihood Estimation (MLE) using a distribution approach that maximizes likelihood function.

Maximum Likelihood Estimator (MLE) is a method of estimating the parameters of the data clusters following a particular distribution distribution. In this case MLE is a method applied to maximize the likelihood function and least squares method using a geometric approach by minimizing the error so that it can generate the probable parameters with maximum likelihood. In general the maximum of a function can not be solved analytically because if it is obtained implicit and nonlinear form so that it can be solved using Newton Raphson Algorithm, Fisher-Scoring Algorithm and Expectation Maximization Algorithm.

Newton Raphson's algorithm is a looping procedure used to solve non-linear equations. This algorithm utilizes first order derivative vectors and second order derived matrices of maximized functions. Fisher Scoring algorithm is similar to Newton Raphson's algorithm. The difference is that fisher scoring uses the expected value of the second order derivative matrix to the parameters in the model (Ehlers, 2002).

Expectation Maximization algorithm is a common algorithm used to calculate the maximum likelihood estimates used for the circumstances that include missing observations. The first step is to divide the data into two parts,

---

Corresponding Author:- Switamy Angnitha Purba.
Address:- Faculty of Math and Science Universitas Sumatera Utara.

namely the missing and nonmissing part, then estimate the value of data lost through linear regression so that the data becomes complete. The initial regression used was taken from the observed data only. In the next iteration process the estimated missing data is obtained from the complete linear data regression equation in the previous data.

Research on the estimation of parameter values using the Maximum Likelihood method has indeed gained a great deal of attention for researchers. The focus of this study is to estimate the parameter values in Maximum Likelihood model based on Newton Raphson and Fisher Scoring and Expectation Maximization algorithm. The purpose of comparing the three algorithms is to find out the estimation value on the three algorithms have the same result or have big result difference.

Method:-
Maximum Likelihood:-
Maximum likelihood is an important estimation method in statistical analysis. The main objective of maximum likelihood estimation is to find parameters that maximize joint likelihood of a data. Let $X_1, X_2, ..., X_n$ has joint density $f_\theta(x_1, x_2, ..., x_N) = f(x_1, x_2, ..., x_n|\theta)$ in this case the observed values are $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ so that the likelihood of θ is:

$$L(\theta) = \prod_{i=1}^{N} f(x_i|\theta)$$

Then we used logarithm to maximize likelihood function or better known as loglikelihood function. The loglikelihood function is:

$$l(\theta|x_1, x_2, ... , x_N) = \sum_{n=1}^{N} \ln f(x_N|\theta)$$

Newton Raphson Algorithm:-
One of the most commonly used methods of statistical optimization is the newton raphson method. The method is based on estimation of a function that wants to be quadratically optimized. The optimum of the estimation will generally be easier if the initial predicted value of the function is good. However, if the initial predicted value of the function is not nearly near the optimum value so the new estimation will be recalculated and the process will be repeated continuously. Millar (2011) reveals that the newton raphson algorithm derives from the quadratic estimation of the objective loglikelihood function through derivatives of linear estimation. Where if $\theta \epsilon \mathbb{R}^s$ so its first derivative are $l'(\theta) = \frac{\partial l(\theta;y)}{\partial \theta}$ and the parameter value $\theta^{(k)}$ on the iteration k. The newton raphson algorithm predicts $l'(\theta)$ using the extension of taylor series by $\theta^{(k)}$. Its produce the Newton Raphson Algorithm as this:

$$\theta^{(k+1)} = \theta^{(k)} - H\left(\theta^{(k)}\right)^{-1} l'\left(\theta^{(k)}\right)$$

In this case $H\left(\theta^{(k)}\right)$ is the Hessian matrix $s \times s$ the second derivative of $l(\theta)$ evaluated when $\theta^{(k)}$.

Fisher Scoring Algorithm:-
Scoring algorithm or commonly known as fisher scoring is a form of newton method used in statistics to solve the maximum likelihood equation. Fisher scoring algorithm found by Ronald Fisher. Fisher scoring algorithm is similar to newton raphson algorithm, the diff erence is fisher scoring using information matrix. The information matrix is the negative of the expected value of the derived matrix of both functions to be maximized while the newton raphson algorithm uses the second derived matrix of the observed value (Ehlers, 2002).
Fisher Scoring's iteration formula is:

$$\beta^{(r+1)} = \beta^{(r)} + I\left(\beta^{(r)}\right)^{-1} l'(\beta^{(r)})$$

In this case $I^{(r)}$ the rth estimate of the observed information matrix. The matrix of information in this paper is the negative of the expected value of the derived two matrix loglikelihood functions. The matrix of information is $I = -E\left(\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_k}\right)$.

Expectation Maximization Algorithm (EM Algorithm):-
The EM algorithm is a commonly used algorithm for calculating the maximum likelihood estimates used for circumstances that include missing or incomplete observations. The EM algorithm was first researched by Dempster, Laird, and Rubin (1977). The EM algorithm is a two-step process for estimating the parameters of an incomplete data model.

The first step is to divide the data into two parts, namely the missing and nonmissing part, and then estimate the value of data lost through linear regression so that the data becomes complete. The initial regression used was taken from the observed data only. In the next iteration process the estimated missing data is obtained from the complete linear data regression equation in the previous data. The step will continue until the lost data becomes convergent, so maximum parameters are obtained. Chan (2015) revealed that the steps of Expectation Maximization Algorithm (EM Algorithm) are as follows:

Expectation Step (E-Step):-
Given y with the distribution formulation for the complete data $f(x|y, \theta^{(t)})$ is as follows:
$$Q(\theta|\theta^{(t)}) \underline{\text{deff}} \, E_{x|y,\theta^{(t)}}[\log f(X|\theta)]$$

Maximization Step (M-Step):-
Performed by maximizing $Q(\theta|\theta^{(t)})$ to θ as follows:
$$\theta^{(t)} = \underset{\theta}{\arg\max} \, Q(\theta|\theta^{(t)})$$

Result:-
Application on Accident Data:-
This following is the number of accidents that occurred in Indonesia by age group

Age (year)	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39
Total	683	1.307	1.704	4.729	4.175	2.665	2.041	2.017
40-44	45-59	50-54	55-59	60-64	65-70	71-74	75-79	80-84
1.929	1.867	1.828	1.432	1.176	756	534	312	173

In doing parameter estimation using Maximum Likelihood method we need to know the type of distribution from the accident data. In this paper check the type of population distribution of accident data using SPSS. By using SPSS, it is obtained that the data is normally distributed. This is indicated by obtaining a significance value of 0.124 (0.124> 0.05) then it can be said that the data is normally distributed.

This following is a common probability density function (pdf) of normal distribution with parameters μ and σ^2.
$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Likelihood function of normal distribution is:
$$L(\theta|y_1, y_2, \ldots, y_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{N}{2}}\exp\left(-\sum_{i=1}^{N}\frac{(y_i-\mu)^2}{2\sigma^2}\right)$$

Loglkelihood function of normal distribution is:
$$l(\theta|y_1, y_2, \ldots, y_N) = \ln\left[(2\pi\sigma^2)^{-\frac{N}{2}}\exp\left(-\sum_{i=1}^{N}\frac{(y_i-\mu)^2}{2\sigma^2}\right)\right] = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{N}\frac{(y_i-\mu)^2}{2\sigma^2}$$

In this research will be done parameter estimation based on Newton Raphson, Fisher Scoring and Expectation Maximization.

Newton Raphson Algorithm:-
$$\theta^{(k+1)} = \theta^{(k)} - H(\theta^{(k)})^{-1}l'(\theta^{(k)})$$

$$\theta^{(k+1)} = \binom{\mu}{\sigma^2}^k - \begin{bmatrix} -\dfrac{n}{\sigma^2} & -\dfrac{\sum_{i=1}^{N}(x_i - \mu)}{\sigma^4} \\ -\dfrac{\sum_{i=1}^{N}(x_i - \mu)}{\sigma^4} & \dfrac{n}{2\sigma^4} - \dfrac{\sum_{i=1}^{N}(y_i - \mu)^2}{\sigma^6} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \mu) \\ -\dfrac{n}{2\sigma^2} + \sum_{i=1}^{N}\dfrac{(y_i - \mu)^2}{2\sigma^4} \end{bmatrix}$$

Newton raphson algorithm is an iteration algorithm by determining the initial value for each parameter value by any arbitrary and the parameter estimation will be repeated continuously until the convergent value is obtained. If done manually will take a long time therefore in this paper parameter estimation will be done by using software Matlab 2016a. Here iteration estimation result of parameter μ and σ^2 by using Newton Raphson algorithm with initial value $\mu = 234$ and $\sigma^2 = 345$ and with value $\varepsilon = 0.00000001$. Obtained that $\mu = 1725{,}1764706$ dan $\sigma^2 = 1444795{,}0706868$ with the number of iterations is 27.

Fisher Scoring Algorithm:-

$$\theta^{(k+1)} = \theta^{(k)} + I(\theta^{(k)})^{-1}l'(\theta^{(k)})$$

$$\theta^{(k+1)} = \binom{\mu}{\sigma^2}^k + \begin{bmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{n}{2\sigma^4} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \mu) \\ -\dfrac{n}{2\sigma^2} + \sum_{i=1}^{N}\dfrac{(y_i - \mu)^2}{2\sigma^4} \end{bmatrix}$$

By using software Matlab 2016a, Here iteration estimation result of parameter μ and σ^2 by using Newton Raphson algorithm with initial value $\mu = 234$, $\sigma^2 = 345$ and $\varepsilon = 0.00000001$. Obtained that $\mu = 1725{,}1764706$ dan $\sigma^2 = 1444795{,}070686$ with the number of iterations of 2.

Expectation Maximization Algorithm:-
Expectation Step:-

$$Q\left(\mu, \sigma^2 \middle| \mu^{(t)}, \sigma^{2(t)}\right) \underline{\text{deff}} E_{x|y,\mu^{(t)},\sigma^{2(t)}}[\log f(X|\theta)]$$

$$= E_{x|y,\mu^{(t)},\sigma^{2(t)}}\left[-\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{M}\frac{(x_i - \mu)^2}{2\sigma^2} - \sum_{i=M+1}^{N}\frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{M}\frac{(x_i - \mu)^2}{2\sigma^2} - \sum_{i=M+1}^{N}\frac{E_{x|y,\mu^{(t)},\sigma^{2(t)}}(x_i - \mu)^2}{2\sigma^2}$$

$$E_{x|y,\mu^{(t)},\sigma^{2(t)}}(x_i - \mu)^2 = E_{x|y,\mu^{(t)},\sigma^{2(t)}}(x_i^2 - 2x_i\mu + \mu^2) = \sigma^2$$

$$Q\left(\mu, \sigma^2 \middle| \mu^{(t)}, \sigma^{2(t)}\right) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{M}\frac{(x_i - \mu)^2}{2\sigma^2} - \frac{n-m}{2\sigma^2}[\sigma^2]$$

Maximization (M-Step):-

$$\frac{\partial}{\partial\mu}Q(\mu|\mu^{(t)}) = 0 \text{ and } \frac{\partial}{\partial\sigma^2}Q\left(\sigma^2\middle|\sigma^{2(t)}\right) = 0$$

$$\mu^{(t+1)} = \frac{\sum_{i=1}^{m}x_i}{n} \text{ and } \sigma^{2(t+1)} = \frac{\sum_{i=1}^{m}(x_i - \mu)^2}{n}$$

By using software Matlab 2016a, Here iteration estimation result of parameter μ and σ^2 by using Newton Raphson algorithm with initial value $\mu = 234$, $\sigma^2 = 345$ and $\varepsilon = 0.00000001$. Obtained that $\mu = 1725{,}1764706$ and $\sigma^2 = 1444795{,}2041522$ with the number of iterations is 2.

Discussion:-

According to Bain and Engelhardt (1992), the maximum likelihood method uses values in the parameter space Ω corresponding to the maximum possible value of the observed data as an estimate of unknown parameters. In its application L (θ) denotes the density probability function from a random sample. If Ω parameter space is an open interval and L (θ) is a function that can be derived and assumed to be a maximum at Ω.

Mai, et al (2014) exemplifies a random X vector with N observations in which case Y is continuous.

Ehlers (2002) reveals that the newton raphson method is a looping rule for generating β values of β that maximizes the function of g (β).

Lange (2004) reveals that in the gradient statistics of loglikelihood called the second negative score and negative of loglikelihood is the observed information.

One of the uses of the newton raphson algorithm in statistical problems is that loglikelihood in many problems approaches a quadratic function spreading at its maximum. That is what is associated with a normally distributed predicted maximum probability that the logarithm of the normal distribution is a quadratic function. Hence the forecast with loglikelihood is a very good approach and approaches its maximum point (Storvik, 2011).
One advantage of the EM algorithm is its numerical stability where the EM algorithm leads to a steady increase in the probability observed (Lange, 2004).

Conclusion:-

So it can be concluded that the three algorithms have the same estimation results in a normal distributed of accident data. Where in this case estimation using Newton Raphson algorithm has more iteration number than the other two algorithms.

Acknowledgments:-

References:-

1. Bain, J., dan Engelhardt, M. (1992). Introduction to probability and mathematical statistics 2nd edition. Duxbury. USA.
2. Bolstad, B.M. (1998). Comparing some iterative method of parameter estimation for censored gamma data. The University of Waikato.
3. Chan, S. H. (2015). Expectation maximization algorithm. University PURDUE.
4. Ehlers, R. (2002). Maximum likelihood estimation prosedures for categorial data. University of Pretoria, South Africa.
5. Kulter, M. dkk. (2004). Applied linear regression model. Mc Graw Hill/ Irwin Series: Operation and Decision.
6. Lange, K. (2004). Optimization. Springer Texts in Statistics, USA
7. Mai, A.T . dkk. (2014). On optimization algorithms for maximum likelihood estimation. CIRRELT, 64.
8. Miller, R. B. (2011). Maximum likelihood estimation and inference with example in R, SAS and ADMB 1st edition. John Wiley and Sons Ltd. United Kingdom.
9. Storvik, G. (2011). Numerical optimization of likelihoods: additiona literature for STK2120. University of Oslo.
10. Wu,J . C. (1983). On convergence properties of the EM algorithm. University of Wisconsin. The Annals of Statistics. Vol.11, No. 1,95-103.