Implementing Weibull Distribution to Generate Prediction Function for National Final Examination in Papua Province

Halomoan Edy Manurung  
MSc. Student  
Satyawacana Christian Univ.  
Jl. Diponegoro 52-60 Salatiga

Eko Sediyono  
Lecture, IT Dept. Of SWCU  
Satyawacana Christian Univ.  
Jl. Diponegoro 52-60 Salatiga

Iwan Sembiring  
Lecture, IT Dept. Of SWCU  
Satyawacana Christian Univ.  
Jl. Diponegoro 52-60 Salatiga

ABSTRACT
One of Indonesia government’s program which is being implemented in the last ten years is to increase the education quality nationally. One of the education quality standards is final examination. Final examination is a concrete form of what the school has given before, during, and after the learning process, so the result of school policy in conducting learning process can be known from the national final examination score. Hence, the result of final examination is totally important to be analyzed and predicted based on the classification tested knowledge field.

This thesis examines about the method for estimating the parameters of Weibull distribution with likelihood estimation maximum method by using newton raphson iteration to the final examination score in Papua province. The result of this research is a distribution function which can be used to predict the next year national final examination score based on the previous condition and score that the students achieved. The function of the prediction that has been found before can be used to estimate what kind of condition can be applied to improve the learning quality so that the score of final examination will be better than before.

Keywords
Weibull distribution, Newton Raphson Iteration, Method of Maximum Likelihood Estimation, National Exam Score Papua Province, Parameter Estimation Modeling

1. INTRODUCTION
1.1 Background
The advance of technology creates us to keep preparing our self not to left behind in order to be able to deal with the life that more competitive year by year. school as a formal education institution keep on preparing the learners to be able to compete in the ear of globalization, so the education is one of the need which can’t be separated from daily life. In a process of education, there are evaluation, accreditation, and certification to monitor the progress of education. The use of evaluation is to control the quality of education nationally as an accountability form of education organizer to the interest parties. One of education evaluation form is the final examination for elementary school, junior and senior high school. Final examination is defined as an activity to measure and evaluate students’ competence nationally in primary and secondary education level (the ministerial regulation of national education No. 34 year 2007 about final examination).[3]

In a formal sense, the policy of national final examination based on the ministerial regulation of national education no. 77 year 2008 about the national final examination. The goal of national final examination is to evaluate the graduate competence achievement nationally in certain subjects of science and technology (ministerial regulation no. 77 article 2, year 2008). The result of national final examination is used as one of consideration for; (a) mapping the educational unit and/or education program quality. (b) The basis selection to enter the next education level.(c) deciding whether the students pass from the educational unit or not. (d) Guiding and giving aid to educational unit to improve the quality of education. Referring to the ministerial regulation no. 77 year 2008 article 14 about national final examination, it is stated that the learners can pass the final examination if they fulfill the requirements. The requirements are; the average score is 5,50 for all subject which are tested and they are allowed to have score 4,00 for only two subjects and 4,25 for the rest. For the next action, the score of final examination can be described by using mathematic and statistic model to give description and analyze all data that are researched by finding the parameter grade. Based on the problem above, the writer’s interested to examined this field by taking “Implementing Weibull Distribution to Generate Prediction Function for National Final Examination in Papua Province”.[3]

1.2 Identification Problem
Based on the background problem, the problems of the research are as follows:
1. How to get the parameter estimator obtained distribution.
2. How to apply and interpret the model of distribution that has been gotten in the data of final examination in Papua province.

1.3 Previous Research
Spline regression approach for final examination score modeling in state vocational 3 school in Buduran, Sidoarjo [9] contains the model estimator of multivariable linear spline with one knot point to make the score of national final examination in junior high school, the score of tryout, the score of competence, and the score of school final examination, distance and parents’ salary as the model toward the national final examination score of state vocational school 3 in Buduran, Sidoarjo which the goal is to facilitate the arrangement of decision making strategy, especially the factor decisions which influence final examination score of the vocational school. The result shows that the testing model describes the score of national final examination in junior high school, the score of tryout, the score of competence, and the distance, with the significance level of 5 % variable, have a great influence toward the national final examination score
in state vocational school 3 in Buduran, Sidoarjo. Then the variable of school final examination and parents’ salary do not have any great influence toward the score.

The Ordinal Logistics Regression analysis in a case study of vocational school accreditation in east java [10]. The aim of the research is to examine the form of ordinal regression parameter estimation and decide the model of ordinal logistic regression which connect the status of accreditation with the human development index with the school profile which cover the status of the school whether it is state or private, how long the school has existed, the number of the students, the number of the teachers, and the number of alumnus who are accepted in business and industrial field in this last one year when the school proposed the accreditation, the status of the land or building and also the number of the students’ average score with the ordinal logistic regression. The finding of the research shows the significance variable which has a great influence toward the status of accreditation of vocational school in east java is how long the school has existed, the number of the teachers and also the sum of average of the national final examination score when the school proposed for the accreditation.

Estimated time-cycle for semiconductor final testing process with Weibull - distributed waiting time [7]. This research explains that the accurate time cycle is the important basis of an arrangement which is needed by many production applications, especially for the due date, the analysis of matrix performance, the arrangement capacity and schedule. The final testing process is the final step of the complicated manufacture semiconductor process. To increase the response ability and reach the better shipping time at the factories in the final testing, so it is important to develop the accurate cycle of time estimation method. This research uses statistic approach to calculate the time cycle in the final testing semiconductor multi-layer that involves the multiple sum of weibull-distributed waiting time. Beside the percentage from the achieved waiting time is useful for industrial practitioner in the case of the targeted shipping deadline to be on time. the implementation of the time cycle estimation model which is proposed to the semi-conductor final test in industrial park in Hsinchu. The research shows that the calculation of time cycle estimation method is very satisfying.

The implication of weibull distribution to determine the reliability, availability, maintainability and predictive maintenance in boiler system generator at Ipomoni Dalton company [11]. This research explains the determination of the appropriate reliability, availability, and maintainability in generator system steam are done to find out the index of reliability, availability, and maintainability of a system by using weibull distribution. The aim of the maintenance predictive based on the condition is to evaluate the condition of a component by monitoring periodically or continually. From the analysis, it shows the index of reliability, availability, and maintainability for steam generator system for 9500 hours is 0.4051 hour, 0.7775 hour and 0.9077 hour. After doing the predictive maintenance, the result of reliability grade evaluation is different T day from each component with new reliability grade, Rm(t) is 0.88 and 0.55.

1.4 Aims and Contributions
The aims of the research is as follows:

1. For the researcher, to broaden the concept of statistic and apply the model, especially weibull distribution by using likelihood maximum estimation method with raphson newton iteration.
2. For information system scientific, to enrich the scientific concept, especially related to the program modeling using weibull distribution and likelihood estimation maximum method, raphson newton and the application.
3. For education, to give problem solving alternative as evaluation material for education to interpret and take decision for the sake of the education. It is by describing and analyzing, also observing the national final examination progress that’s evaluated step by step.

2. THEORY, MODEL AND ANALYSIS

2.1 The studying result assessment by government
The government of Indonesia considers that national final examination is important to evaluate the students passing competence. It needs national standard in every region in Indonesia since the education in Indonesia is various. Each student must enroll national final examination once if they fulfill the requirements that have been decided by BSNP (Badan Standard Nasional Pendidikan; literacy means the department of national education standard). The national final examination participants will get an identification paper from national exam committee of education unit. The result of the national final examination is for:

1. Mapping the educational unit and/or education program quality.
2. The basis selection to enter the next education level.
3. Deciding whether the students pass from the educational unit or not.
4. Development Guiding and giving aid to educational unit to improve the quality of education.[5][3]

2.2 Two-parameter Weibull distribution
Weibull distribution is one of the first continuous distribution and it is straightforward to determine because the scale parameter can be easily estimated by the program that use the reliability method to estimate the reliability function. The Weibull distribution density function can be expressed as:
\[ f(t, \alpha, \beta) = \frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right] \]

Weibull cumulative distribution function is:
\[ F(t, \alpha, \beta) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right] \]

So that the survival function is:
\[ S(t, \alpha, \beta) = 1 - \left[1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right]\right] \]

This is random variable, \( \alpha \) is the scale parameter and \( \beta \) is a shape parameter. Parameters \( \alpha \) and \( \beta \) are parameters to be estimated.[4][6]

2.3 Methods of Maximum Likelihood Estimation
Methods of Maximum Likelihood Estimation is one method parameter estimation which can be used to estimate the parameters in a model known distribution[1]. To obtain a model parameter estimator with Methods of Maximum Likelihood Estimation is to maximize the likelihood function.
Let $X_1, X_2, ..., X_n$ be a sample of size $n$ from a population whose distribution depends on a set of $p$ parameters, namely $\theta_1, \theta_2, ..., \theta_p$. We can regard this sample as forming a random vector $X = (X_1, X_2, ..., X_n)'$. Suppose that $X$ has density function $f(x, \theta)$, where $x = (x_1, x_2, ..., x_n)'$ and $\theta = (\theta_1, \theta_2, ..., \theta_p)$. This density function is usually referred to as the likelihood function of $X$, we denote it by $L(x, \theta)$. For a given sample, the maximum likelihood estimate of $\theta$, denoted by $\hat{\theta}$, is the value of $\theta$ that maximizes $L(x, \theta)$. If $L(x, \theta)$ has partial derivatives with respect to $\theta_1, \theta_2, ..., \theta_p$, then $\hat{\theta}$ is often obtained by solving the equations:

$$\frac{\partial \log(L(x, \theta))}{\partial \theta_i} = 0, \; i = 1, 2, ..., p \tag{2.3.2}$$

Thus $\hat{\theta}$ satisfies the equation:

$$\frac{\partial \log(L(x, \theta))}{\partial \theta_i} = 0, \; i = 1, 2, ..., p \tag{2.3.2}$$

Equations (2.3.2) are known as the likelihood equations. Suppose that $X_1, X_2, ..., X_n$ for a sample of size $n$ from a normal distribution with an unknown mean $\mu$ and a variance $\sigma^2$. Here, $\theta = (\mu, \sigma^2)$, and the likelihood function is given by:

$$L(x, \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

Let $L'(x, \theta) = \log L(x, \theta)$, then

$$L'(x, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

The likelihood equations in formula (2.3.2) are of the form:

$$\frac{\partial L'}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \tag{2.3.1}$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \tag{2.3.2}$$

Equations (2.3.1) and (2.3.2) can be written as:

$$\sum_{i=1}^{n} (x_i - \mu) = 0 \tag{2.3.3}$$

$$\sum_{i=1}^{n} (x_i - \mu)^2 = n\sigma^2 \tag{2.3.4}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, if $n \geq 2$, then equations (2.3.3) and (2.3.4) have the solutions:

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

These are the maximum likelihood estimates of $\mu$ and $\sigma^2$. Hence, for $\mu = \hat{\mu}$ and $\sigma^2 = \hat{\sigma}^2$, we have

$$\frac{\partial^2 L'}{\partial \mu^2} = \frac{n}{\sigma^2} > 0$$

$$\frac{\partial^2 L'}{\partial \sigma^2} = -\frac{n}{2\sigma^4}$$

Thus $\frac{\partial^2 L'}{\partial \mu^2} < 0$ and $\text{det}(A) = \frac{n^2}{2\hat{\sigma}^4} > 0$. Therefore $(\hat{\mu}, \hat{\sigma}^2)$ is a point of local maximum of $L'$.[1]

### 2.4 Newton-Raphson Method

Let $x_0$ be an initial point in the domain of $f(x)$. By Taylor’s expansion of $f$ in a neighborhood of $x_0$, it is possible to approximate $f(x)$ with the quadratic function $\phi(x)$ given by:

$$\phi(x) = f(x_0) + (x - x_0)\Delta f(x_0) + \frac{1}{2}(x - x_0)^2 H_f(x_0)(x - x_0)$$

Where $H_f(x_0)$ is the Hessian matrix of $f$ evaluated at $x_0$. Oon the basis of formula (2.4.1), we can obtain a reasonable approximation to the minimum of $f(x)$ by using the minimum of $\phi(x)$. If $\phi(x)$ attains a local minimum at $x_0$, then we must necessarily have $\nabla \phi = 0$ that is:

$$\nabla f(x_0) + H_f(x_0)(x_0 - x_0) = 0 \tag{2.4.2}$$

If $H_f(x_0)$ is nonsingular, then from equation (2.4.2) we obtain:

$$x_1 = x_0 - H_f^{-1}(x_0)\nabla f(x_0)$$

If we now approximate $f(x)$ with another quadratic function, by again applying Taylor’s expansion in a neighborhood of $x_1$, and the repeat the same process as before with $x_1$ used instead of $x_0$, we obtain the point:

$$x_2 = x_1 - H_f^{-1}(x_1)\nabla f(x_1)$$

Further repetitions of this process lead to a sequence of points, $x_0, x_1, x_2, ..., x_k, ..., x_n$, such that:

$$x_{i+1} = x_i - H_f^{-1}(x_i)\nabla f(x_i)$$

The Newton-Raphson method requires finding the inverse of the Hessian $H_f$ at each iteration. Newton-Raphson method will be used to get the value iteration until a convergent estimator.[1]

### 2.5 The Method of Research Analysis

#### 2.5.1 The Data Sources and Research Variable

The data collection in this research is secondary data, it is the data of national final examination in Papua from 2008-2012, taken from learning and education department in Papua province, Software maple 13 and matlab 2013 is used to process the data. The research variable is responses variable, it’s the average score of national final examination in Papua province, that will be categorized to each region and then will be evaluated theoretically to get the parameter so that it can describe or characterize and analyze the whole data.[2]

#### 2.5.2 The analysis method

Based on the aim of the research, the analysis method which is used in this research as the follow:

1. Determining Determine the likelihood maximum from data distribution as follow below:
   - Form the survival function and derivative it, namely:
     $$S(\text{Survival}) = e^{-\left(\frac{t_{1k}}{\alpha}\right)^\beta}$$
   - The result is wellbull distribution from PDF (probability Density Function), it is:
     $$P(\text{pdfsurvival}) = -\left(\frac{t_{1k}}{\alpha}\right)^\beta e^{-\left(\frac{t_{1k}}{\alpha}\right)^\beta}$$
   - Do log transformation to the likelihood function, it is:
     $$L(\text{likelhoodsurvival}) = \log \prod_{k=1}^{n} P_k(\text{pdfsurvival})$$

2. Evaluate the parameter estimation using raphson newton, as the following step:
   - Find the gradient vector which contains the parameter, it is:
     $$\bar{g} = \begin{bmatrix} \frac{\partial L}{\partial \alpha} \\ \frac{\partial L}{\partial \beta} \end{bmatrix}$$
3. RESULTS AND DISCUSSION

By taking one region in the coastal area, mountainous area and province, it achieved the analysis subject toward the average of national final examination score. The sample to analyze those three regions is biology subject in Jayapura city (figure 9) as the follow:

### Table 3.1 Biologi

<table>
<thead>
<tr>
<th>Year</th>
<th>Shape parameter s</th>
<th>Scalar parameter s</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>The peak value</th>
<th>Survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>8.1834</td>
<td>13.55</td>
<td>7.857</td>
<td>0.7780</td>
<td>0.606</td>
<td>90%</td>
</tr>
<tr>
<td>2009</td>
<td>7.1842</td>
<td>8.332</td>
<td>6.741</td>
<td>1.2950</td>
<td>0.429</td>
<td>44%</td>
</tr>
<tr>
<td>2010</td>
<td>7.3962</td>
<td>7.338</td>
<td>6.916</td>
<td>1.1799</td>
<td>0.368</td>
<td>52%</td>
</tr>
<tr>
<td>2011</td>
<td>8.1364</td>
<td>14.28</td>
<td>7.945</td>
<td>0.4905</td>
<td>0.644</td>
<td>90%</td>
</tr>
<tr>
<td>2012</td>
<td>7.7710</td>
<td>15.19</td>
<td>7.539</td>
<td>0.4960</td>
<td>0.715</td>
<td>80%</td>
</tr>
</tbody>
</table>

It can be concluded that in the last two years has increased very well despite two previous years decreased so for years to come opportunities the average value of seven (7) can be maintained.

### 3.1 Biak County

For Biak region representing coast-line (figure 1 to 4), field of study of science is the most prominent in the following table:

#### Table 3.2 Biak County

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>The peak value</th>
<th>Survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>7.4867</td>
<td>6.84</td>
<td>7.2525</td>
<td>7.4321</td>
</tr>
<tr>
<td>2012</td>
<td>8.0546</td>
<td>1.0730</td>
<td>24.0618</td>
<td>0.466</td>
</tr>
</tbody>
</table>

It can be concluded that in the last two years has increased very well despite two previous years decreased so for years to come opportunities the average value of seven (7) can be maintained.

### 3.2 Jayawijaya County

For Jayawijaya region representing high-line (figure 5 to 8), field of study of science is the most prominent in the following table:

#### Table 3.3 Jayawijaya County

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>The peak value</th>
<th>Survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>6.8344</td>
<td>7.66</td>
<td>7.5098</td>
<td>7.5750</td>
</tr>
<tr>
<td>2012</td>
<td>6.0579</td>
<td>1.3060</td>
<td>0.5179</td>
<td>0.3734</td>
</tr>
</tbody>
</table>

It can be concluded that in the last two years has increased very well despite two previous years decreased so for years to come opportunities the average value of seven (7) can be maintained.

### 3.3 Jayapura City

For Jayapura region representing province (figure 9 to 12), field of study of science is the most prominent in the following table:
The conclusion, from those three regions, is the most prominent science subject in the first number is chemistry at Jayapura city, the second one is physics at Jayapura city, the third is math at Jayapura city and the last one is biology at Biak regency. The quality, mean, facilitation, and the quality of the teachers in province is much better than in coastal or mountainous area. The mountainous area is lower than coastal area or province, so it needs a lot of things to do to make the success of the national final examination and the quality of the school in that region better than before. A cooperation with government to increase teachers’ mean, facilitation, professionalism, distribution protection and prosperous is a really serious attention to create progress and improvement of the national final examination score to be better year by year and can be maintained in order to all regions in Papua has a very good quality.

4. CONCLUSION

One of the important aspects in inference statistic is to estimate the parameter of a population by analyzing the data collection of the population itself. After getting the parameter estimation, the next step is to take conclusion about the parameter of the population based on the information of the sample data of the population. Weibull distribution uses likelihood estimation maximum method with raphson newton iteration can give a good description and prediction in analyzing and giving interpretation the data of the national final examination average score.

### Table 3.4 Jayapura City

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>The peak value</th>
<th>Survival (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>2008-2012</td>
<td>(7.8175, 8.2925, 8.1310, 8.1555, 8.3776)</td>
<td>(0.9301, 0.8125, 0.5685, 0.3890, 0.4121)</td>
<td>(0.48, 0.579, 0.638, 0.467, 0.505)</td>
<td>(83%, 95, 92%, 83%, 90%)</td>
</tr>
<tr>
<td>Physics</td>
<td>2008-2012</td>
<td>(6.5235, 7.6245, 7.6530, 7.6245, 8.1610)</td>
<td>(1.1394, 1.0070, 0.9548, 0.5873, 0.4851)</td>
<td>(0.423, 0.62, 0.533, 0.657, 0.704)</td>
<td>(36%, 83, 81%, 81%, 94%)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>2008-2012</td>
<td>(7.0035, 7.3090, 8.1950, 7.8627, 7.7805, 0.4680)</td>
<td>(0.8721, 1.6386, 0.5877, 0.7641, 0.4680)</td>
<td>(0.481, 0.296, 0.7, 0.549, 0.661)</td>
<td>(54%, 64%, 95, 87%, 85%)</td>
</tr>
<tr>
<td>Biology</td>
<td>2011-2012</td>
<td>(7.8570, 6.7415, 6.9160, 7.9459, 7.5395)</td>
<td>(0.7780, 1.2950, 1.1799, 0.4905, 0.4960)</td>
<td>(0.606, 0.429, 0.368, 0.644, 0.715)</td>
<td>(90%, 44%, 52%, 90%, 80%)</td>
</tr>
</tbody>
</table>
Fig 5: Plot PDF, CDF, Survival Biology of Jayawijaya County

Fig 6: Plot PDF, CDF, Survival Physics of Jayawijaya County

Fig 7: Plot PDF, CDF, Survival Chemistry of Jayawijaya County

Fig 8: Plot PDF, CDF, Survival Mathematics of Jayawijaya County

Fig 9: Plot PDF, CDF, Survival Biology of Jayapura City

Fig 10: Plot PDF, CDF, Survival Physics of Jayapura City

Fig 11: Plot PDF, CDF, Survival Chemistry of Jayapura City

Fig 12: Plot PDF, CDF, Survival Mathematics of Jayapura City
5. ACKNOWLEDGMENTS
We thank Paulus Harsono and Irfan Wahyudi for their assistance in preparing the data; Felix Reba and Yuly Yolanda Warer for their critique of the findings of this study.

6. REFERENCES