Investigating the Relationship between Neonatal mortality rate and Mother’s characteristics

M. Abdollahian*, S. Ahmad#, S. Huda+, S. Nuryani, D. Anggraini

*School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia
#Department of Management, College of Business Administration Al Yammamah University, Riyadh Saudi Arabia
+School of Information Technology and Mathematical Sciences, University of Ballarat, Australia
Ulin Hospital (RSUD Ulin) Banjarmasin Indonesia
Department of Mathematics Lambung Mangkurat University, Banjararu, Indonesia

Abstract

Neonatal mortality rate (NMR) is an increasingly important public health issues in many developing countries. Neonatal death now accounts for about two-thirds of the eight million infant deaths that occur globally each year. It is well-documented that low birth weight (LBW) is the most significant factor influencing NMR. This paper deploys regression analysis to explore the relationship between weight of low birth weight babies and various characteristics of mother. The results indicate that there is a significant relationship between weight of low birth weight babies and mother’s weight, age, gestation age and hemoglobin level.

Keywords: Normality test, multivariate normal distribution, simulation and multi-regression

1. Introduction

Newborn size is an important indicator of infant survival and childhood mortality. Simple and accurate method of estimating newborn weight that can be easily applied to all pregnancies is an important means of reducing mortality rate. Several investigators have shown that low birth weight is associated with high prenatal mortality and morbidity [6]. Birth weight may also predict both short- and long term adverse outcomes. For example, higher birth weight among term infants is associated with birth complications [2] as well as a reduced risk of cardiovascular disease and hypertension in later life, but an increased risk of obesity [7-10].

The ultrasound has been used for the examination and evaluation of high-risk pregnancies and for the diagnosis of congenital malformations. During the last two decades ultrasound techniques have been improved and are implemented in most gynecology and obstetric clinics worldwide [11]. The biophysical profile was used to assess fetal well-being and to confirm fetal gestational age by measuring the biparietal diameter and crown rump length [12]. Other investigators have predicted intrauterine fetal weight using ultrasonographic measurement of the fetal abdominal circumference [3]. More recent reports have emphasized the usefulness of this measurement in monitoring normal fetal growth and in detecting intrauterine growth retardation [4]. All these studies were conducted in Western countries where perinatal medical care and ultrasonographic measurements are advanced.

This study was undertaken among pregnant women attending the Banjarmasin Clinic in Indonesia during 2010-2011. A total of 198 pregnant women between the ages of 16 and 42, who attended the clinic for antenatal care or routine follow up were included in the study. For each patient; age, baby weight, patient weight at the time of delivery (Kg), gestation age at the time of delivery, hemoglobin level before (in the third semester of pregnancy) and after (towards the end after consumption of Vitamin C and Sulfas Ferroles) were measured. Out of the 198 deliveries 10 had a low birth weight. Low birth weight is defined as a birth weight of less than 2500 gram and is a well-documented risk factor for neonatal mortality [1, 5]. Information on these characteristics for individual babies was obtained from the records. In this study we have investigated the distribution of all these characteristics for the low weight babies. It was observed that all characteristics follow normal distribution. Multivariate normal distribution based on the observed means and standard deviations is used to obtain 1000 simulated data. Multi-regression analysis is deployed to find the relationship between weight for low weight birth babies and the above characteristics.

This paper is organized in the following manner. Time series and distribution analysis are discussed in section 2. A review
of the multi-regression analysis and multi-normal stimulations are presented in section 3. Discussion based on Simulated and application example with real clinical data is presented in section 4. This followed by conclusion in section 5.

2. Time series and Distribution analysis

2.1 Time series plots are used to evaluate patterns in data over time. It is also used to investigate whether the patterns on different characteristics have uniform or corresponding pattern. For our experimental data, we first decided to compare the time series pattern of the detailed characteristics by including all 198 patients. The plot for the first 30 patients is presented in Figure 1.

The result clearly shows that the trend in baby’s weight follows the trend in mother age, digestion age, hemoglobin level and mother weight. However, the variability in baby’s weight almost consistently matches the trends present in the mother’s age and hemoglobin level; i.e., when the hemoglobin level is down the baby weight is also down.

2.2 Normality test

To carry out statistical analysis one often requires normal data. The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed. The authors have used statistical package Minitab to fit the commonly used normal distribution function to individual characteristic of the low birth weight data. For all the characteristics in the data set the p-value of the fit was less than 0.01 indicating that all the six characteristics follow normal distribution. The graph is presented in Figure 2.

We have used the multi-normal distribution based on the means and standard deviations of these fitted individual normal distributions to generate one thousand samples of size one with six characteristics in each sample. The Normal probability plots for the simulated data together with the corresponding p-value for the individual characteristics are given in Figure 3.

Figure 1: Time series plot of all the six characteristics
Figure 2: Normal probability plot of the actual data, where the p-value for each characteristic is listed.

Figure 3: Normal probability plot of the 1000 simulated data with the corresponding p-value for individual characteristic.
Figure 3 shows that the p-value corresponding to all characteristics is greater than 0.01. Therefore we can conclude that the 1000 simulated data follow normal distribution.

3. Multi-Regression Model and Multinormal Simulation

Multiple-regression is an appropriate approach used to accurately model the relationship between a scalar variable \( y \) and one or more explanatory variables denoted by \( X \). It is assumed in multiple regressions that the residuals (predicted minus observed values) are distributed normally. Even though most tests (specifically the \( F \)-test) are quite robust with regard to violations of this assumption, it is always a good idea, before drawing final conclusions, to review the distributions of the major variables of interest.

Given a data set \( \{y_i, x_{i1}, x_{i2}, \ldots, x_{ip}\} \) of \( n \) statistical units, a linear regression model assumes that the relationship between the dependent variable \( y_i \) and the \( p \)-vector of regressors \( x_i \) is linear. This relationship is modelled through a so-called “disturbance term” \( \epsilon_i \), an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

\[
y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad i = 1, \ldots, n.
\]

In this study the authors are interested to find the relationship between; the baby weight for low birth weight babies (dependent variable) and mother age, mother weight, gestation age at the time of delivery and hemoglobin level (before and after). The independent effect of a number of variables on the baby weight was calculated by using multiple linear regressions. Using 1000 multi-normal simulated data based on the means and standard deviations of the normal distribution fitted to individual observed characteristics, we obtained the following regression model:

\[
\text{Baby } W = 1.46 - 0.00126 \text{ age} + 0.00299 \text{ mother W} + 0.00483 \text{ digestion W} + 0.263 \text{ Hem before} - 0.200 \text{ Hem After}
\]

The correlation coefficient \( r = 0.8 \) for the regression equation given in (3.2).

To assess the efficacy of the proposed model, we have deployed the model to predict the observed weight of the actual data. The predicted value, together with the standard error of the prediction, observed and predicted 95% confidence interval, error of prediction and the actual observed weight are given in table 1.

### Table 1: Predicted Values for the actual baby weight based on the proposed multi regression model.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95% CI</th>
<th>95% PI</th>
<th>Error of prediction</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.35634</td>
<td>0.00717</td>
<td>(2.34228, 2.37040)</td>
<td>(2.19305, 2.51962)</td>
<td>0.04366</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>2.17726</td>
<td>0.00688</td>
<td>(2.16375, 2.19077)</td>
<td>(2.01402, 2.34050)</td>
<td>0.02274</td>
<td>2.2</td>
</tr>
<tr>
<td>3</td>
<td>2.24332</td>
<td>0.00716</td>
<td>(2.22927, 2.25736)</td>
<td>(2.08003, 2.40661)</td>
<td>0.05668</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>2.23267</td>
<td>0.00513</td>
<td>(2.22621, 2.24273)</td>
<td>(2.06968, 2.39566)</td>
<td>0.06733</td>
<td>2.3</td>
</tr>
<tr>
<td>5</td>
<td>2.22985</td>
<td>0.00580</td>
<td>(2.21848, 2.24122)</td>
<td>(2.06677, 2.39293)</td>
<td>-0.12985</td>
<td>2.1</td>
</tr>
<tr>
<td>6</td>
<td>2.44998</td>
<td>0.00607</td>
<td>(2.43806, 2.46190)</td>
<td>(2.28686, 2.61310)</td>
<td>-0.04998</td>
<td>2.4</td>
</tr>
<tr>
<td>7</td>
<td>2.27899</td>
<td>0.00548</td>
<td>(2.26824, 2.28974)</td>
<td>(2.11595, 2.44202)</td>
<td>0.12101</td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>2.44922</td>
<td>0.00557</td>
<td>(2.43830, 2.46015)</td>
<td>(2.28617, 2.61277)</td>
<td>0.05078</td>
<td>2.5</td>
</tr>
<tr>
<td>9</td>
<td>2.45621</td>
<td>0.00509</td>
<td>(2.44622, 2.46620)</td>
<td>(2.29322, 2.61920)</td>
<td>-0.05621</td>
<td>2.4</td>
</tr>
<tr>
<td>10</td>
<td>2.19371</td>
<td>0.00636</td>
<td>(2.18123, 2.20681)</td>
<td>(2.03055, 2.35687)</td>
<td>-0.09371</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Table 1 presents the predicted weight in column 2 under “Fit”, corresponding 95% confidence interval under “95% CI”, Actual observed weight which is listed in the last column under “Actual value” and error of the prediction based on the proposed model under “Error of prediction”. It can be seen that the maximum forecasting error for the proposed model corresponds to observation 7 which is 121 grams.

### Table 2: Summary statistics for the forecasting error of the proposed model

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>0.0032</td>
<td>0.0255</td>
<td>0.0807</td>
</tr>
</tbody>
</table>

Table 2 shows that the proposed regression model can predict the true value of the new borne weight for LBW babies with the mean error of 0.0032 (3.2 g) and the standard
error of 0.0255. Therefore we can claim that the accuracy of the model is significant and precise.

4. Discussion

The study was conducted among women enrolled in a maternity clinic in Banjarmasin Indonesia. One of the objectives in this research experiment was to investigate the relationship between baby weight (for the low birth weight babies) and mother age, weight, gestation age at the time of delivery, hemoglobin level (before and after consumption of Vitamin C and Sulfas Ferroses) and propose a model to predict weight of LBW babies based on these characteristics. The independent effect of a number of variables on the baby weight was calculated by using multi-regressions. The slope (the beta-coefficient) that shows the amount of change in the dependent (baby weight) for one unit change in an independent variable, such as, mother weight, and other variables together with the Pearson correlation coefficient were used to identify the effect of independent variables on the baby weight. All the analyses were performed using the MINTAB statistical software package.

The slopes and their corresponding p-values in the proposed regression model show that the most effective variables are the gestation age (slope of .005), hemoglobin level before (slope = .263)and after (slope = -.2) and the least effective variables are mother weight (slope = .003) and age (slope = -.00126).

The proposed regression model has a correlation coefficient of 80%. Therefore, one can claim that the relationship is strong enough to predict the newborn weight based on the other four independent variables. The model was then used to predict the observed sample weight data to assess its predicting accuracy. Results presented in tables 1 and 2 shows that the estimated weight using the proposed model is very close to the actual recorded weight for the LBW babies with the mean error prediction of 3.2 grams.

5. Conclusion

Low birth weight is an increasingly prevalent factor in the Maternal Mortality Rate (MMR). Therefore many studies have attempted to identify the sources of variation in the newborn weight. In this study, multi-regression model is used to assess the independent effects of the mother age, weight, gestation age at the time of delivery and hemoglobin level (before and after) on the new born weight for low weight babies.

One thousand Multi-normal simulated data based on the means and standard deviations of the recorded low birth weight newborns were used to estimate the model. The results show that for low birth weight babies there is a statistically significant relationship between the newborn weight and the independent variables; mother age, weight, gestation week and hemoglobin level with a correlation coefficient of 80%.

The proposed model was used to estimate the recorded weights together with their corresponding 95% confidence interval. Analysis of the prediction errors shows that the mean prediction error for the recorded data is 3.2 grams. Therefore one can conclude that the proposed multi-regression model is capable of accurately predicting the weight for the low birth weight babies based on the characteristics of the mother. The model is based on one thousand simulated data using the sample measurements and in future would be tested on a larger set of observed data.

Acknowledgement

The authors would like to thank the medical practitioners and staff in the maternity clinic in Banjarmasin Indonesia for their effort in collecting this data and providing them to us for this research.

References


