IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Understanding The Relationship Between Low Birth Weight and Gestation Period Using Regression Modeling

Jaya Kolluri

The Winsor School, 103 Pilgrim Rd, Boston, MA 02215, USA; jayakolluri100@gmail.com

ABSTRACT: The high prevalence of preterm deliveries (i.e., short gestation periods) and their impact on neonatal mortality rates have been a significant challenge for the medical community. It is also widely known that low birth weight is one of the leading causes of neonatal mortalities. This study focused on understanding the relationship between low birth weight (LWB) and short gestation periods using a data set consisting of patient records of over 900 pregnant women in King County. Preliminary data exploration identified a high correlation between low birth weight and preterm pregnancies. A simple regression model was developed using the R statistical package to validate the linear relationship between LBWs and gestation periods. We built a simple regression model that worked functionally and set the first steps to creating models that can predict birth weights on real-time data. However, we also noticed that the R2 value indicated that only 27% of the variance in the birth weights could be explained by the gestation periods. The remaining 73% can be attributed to other variables not included in the model. The low R2 suggests that there is room for improving the model performance by adding external related variables such as patient demographics or fitting another model type to the data.

KEYWORDS: Bioinformatics; Low Birth Weight; Neonatal Mortality; Linear Regression; Neonatal Period.

Introduction

The first twenty-eight days - the neonatal period - is the most vulnerable survival time for an infant. Children face the highest risk of dying at an estimated average of seventeen deaths per one-thousand live births in their first month.[1] There are approximately 6,700 neonatal deaths daily, with almost three-quarters occurring within the first week.[1] Although the medical community has made a conscious effort and has accelerated its research, the dangers these infants face within the first few weeks of life are still severe.

Many extensive research projects have been conducted to discover the root cause. Globally infectious diseases, including pneumonia, diarrhea, malaria, preterm birth, birth asphyxia, and congenital anomalies, were the leading causes of death for children under one.[2] Almost 45% of deaths within these age brackets stemmed from nutrition-related factors such as low birth weight and malnutrition. Olack *et al*. explain how factors such as low birth weight control the rising neonatal mortality rates.[2] Infants under the weight of 2500g have been reported to be the most common group who fail to make it past their neonatal period.2 ENAP (The Every Newborn Action Plan) is a roadmap to reduce neonatal deliveries and spread awareness of a problem like this.[2]

Babies born small (low birth weight) are at increased risk for several medical and developmental complications. While these medical success stories highlight the power of medical technology, serious questions remain about how these infants will develop and whether they will have normal, productive lives. Children with low birth weight can be born at term or before their term and have a wide range of social and medical risks.[3] Herbst et al. show that low birth weight (LBW) is the most critical risk factor affecting infant mortality and also show how it leads to a higher risk of adverse developmental outcomes.[4] Legislators have tried to address this through public programs like First Steps, which was implemented in the early 1990s to reduce the number of low birth weight infants.[5] The First Steps program was designed to help low-income pregnant individuals get the services they may need. Follow-up research, like that of Arima *et al.*, shows that First Steps significantly reduced LBW infant births, even when adjusting for confounding factors like maternal age and ethnicity.[5]

Our work focuses on directly understanding predictive factors of LBW births. We use a dataset from King County, WA, which consists of 2500 samples of birth certificates from children born in this county in 2001. Some members of this dataset had mothers participating in the First Steps program, while the majority did not. The data are restricted to singleton births (i.e., not twins, triplets, or multiple). Our exploratory data analysis showed a strong correlation between gestation period and birth weight. We then developed a regression model to validate this correlation. Finally, we suggest the steps and future opportunities continue this line of research.

Methods

The Kings County data set consists of 2500 records of newborn babies along with 17 variables as follows: gender of baby, number of babies, mother's age, race, parity (number of previous live infants the mother delivered), mother's marital status, birth weight in grams, number of cigarettes mother smoked per day during pregnancy, number of drinks mother had per day during pregnancy, if mother participated in First Steps program, welfare amounts, if mother was a smoker, if mother was a drinker, mothers weight prior pregnancy, mothers weight gain during pregnancy, the highest level of education mother

DOI: 10.36838/v5i2.11

finished, and gestation (the number of weeks the pregnancy lasted).

This study followed a three-step process. To begin with, an exploratory data analysis was carried out across all variables to understand the relationships between various factors. A more detailed bivariate analysis followed it to investigate further the relationship between the variables Gestation and Birth Weight. After confirming a correlation between these two variables, a simple linear regression model was developed to see how the gestation time period could be used to predict the birth weight of premature infants. The fitted regression model was:

BirthWeight =124(Gestation) - 1411.124 with $p < 0.05$*, and with an* R^2 *of 0.27.*

The model estimated the birth weight to increase by 124g every week of gestational age. Now that we have a working model to predict the increased birth weight rate, we checked to see how accurate our predicted values were compared to the actual data given to us. To further validate the model performance, the dataset was resampled 1000 times to generate multiple models.

```
for (x in 1:10) {

  #re sampling data from original
  sample_data = data[sample(1:nrow(data), nrow(data), replace = TRUE), ]

  model <- lm(bwt ~ gestation, data=sample_data)

  # save coefficients
  sample_coef_intercept <- c(sample_coef_intercept, model$coefficients[1])

  sample_coef_slope <- c(sample_coef_slope, model$coefficients[2])
}
```

Figure 6: Data Sampling and Modeling Code

Results and Discussion

As described above, we picked Birth Weight and Gestational Period as our two variables for the exploratory data analysis.

Figure 1: Histogram of Birth Weight

As expected, the values formed a bell curve with the spike in weight around 3000g and 4000g, the average birth weight for newborn babies. After analyzing the data for the variable BWT (Figure 1), an analysis was conducted on the variable Gestation (Figure 2).

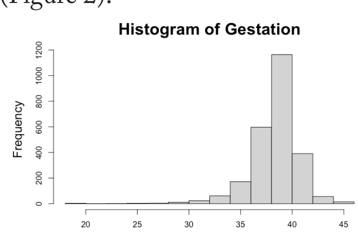

Figure 2: Histogram of Gestation

The average pregnancy lasts around 200 days or 40 weeks, meaning a preterm or premature baby is delivered at around 37 weeks; extreme cases of preterm deliveries range from 23 weeks to 28 weeks. Most women delivered at approximately 40 weeks, as seen in the graph, although many babies were born prematurely.

Since the focus of the current research work is on birth weight, a bivariate analysis was done to explore the factors that cause low birth weight.

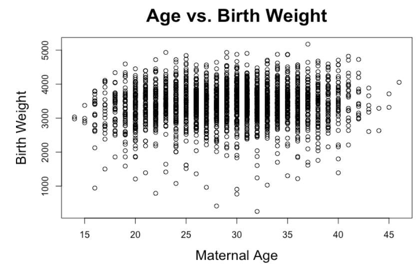

Figure 3: Scatterplot between Age and Birth Weight

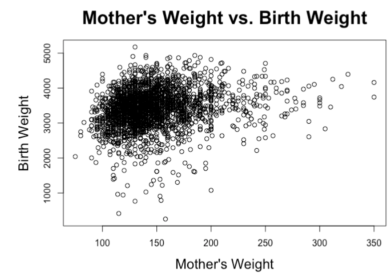

Figure 4: Mother's Weight and Birth Weight of Infant

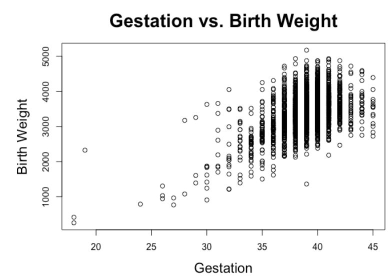

Figure 5: Scatter Plot between Gestation and Birth Weight

Figures 3 - 5 show three variables plotting against Birth Weight to see the correlation between values. While there was a slight correlation between Mother's Weight before pregnancy and the Mother's Age, the variable Gestation showed a significant positive correlation. As gestational time increases, so does the infant's weight. This analysis also showed us that a shorter pregnancy (less gestational time) would result in lower birth weight.

Overall, the predicted values differed from the actual birth weights by an average of 366g, which is relatively minor compared to the actual birth weights. A confidence level of 95% confirmed our hypothesis that the gestation period strongly predicts birth weight. However, the observed R squared value resulted in a low score of 0.27, suggesting that there is room for further improving the model performance by adding external related variables such as patient demographics.

DOI: 10.36838/v5i2.11

Conclusion

In this study, we focused on the relationship between low birth weight (LBW) and the average gestational periods of expecting mothers. Throughout the past decades, preterm deliveries (often resulting in underweight babies) and the increase in neonatal mortality rates have been one of the biggest challenges for the medical community. Previous studies and medical professionals have flagged low birth weight as one of the leading causes of neonatal deaths. As LBW is one of the highest neonatal mortality indicators, we must understand what underlying factors affect both. Using a past data set of 2,500 pregnant mothers' medical records, we developed a linear regression model to predict the weight of a newborn post-delivery. The resulting model was both functional and accurate, with a high confidence level leading us to believe we could use this same procedure on real-time data as our next steps.

However, the low R^2, indicating that only 27% of the variance in birth weights can be explained by gestational periods, suggests that there is room for improvement with the addition of external demographics. Current research is in progress to explore the benefits of including patient demographic data. Future extensions to this study will consider evaluating the role of active learning models that can take real-time patient data using health monitoring devices such as smartwatches. We are also exploring individualized models that can be built based on data specific to a given patient.

Acknowledgments

This study was carried out as the final term project for the Harvard STATStart program for high school students. I want to thank my project mentor, Anuraag Gopaluni, course coordinator Talbot Octavious and the Harvard STATStart program staff for giving me this opportunity to participate in the summer session and for guiding me through this project work.

References

1. Neonatal Mortality. https://data.unicef.org/topic/child-survival/neonatal-mortality/ (accessed 2022-02-12)
2. Olack, B., Santos, N., Inziani, M., Moshi, V., Oyoo, P., Nalwa, G., OumaOtare, L. C., Walker, D., & Otieno, P. A. (2021). Causes of preterm and low birth weight neonatal mortality in a rural community in Kenya: evidence from verbal and social autopsy. *BMC Pregnancy and Childbirth*, 21(1), 536. https://doi.org/10.1186/ s12884-021-04012-z
3. Hack, M., Klein, N. K., & Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *The Future of Children*, 5(1), 176–196. https://doi.org/10.2307/1602514
4. Herbst, M. A., Mercer, B. M., Beazley, D., Meyer, N., & Carr, T. (2003). Relationship of prenatal care and perinatal morbidity in low-birth-weight infants. *American Journal of Obstetrics and Gynecology,* 189(4), 930–933. https://doi.org/10.1067/s0002-9378(03)01055-x
5. Arima, Y., Guthrie, B. L., Rhew, I. C., & De Roos, A. J. (2009). The impact of the First Steps prenatal care program on birth outcomes among women receiving Medicaid in Washington State. *Health Policy (Amsterdam, Netherlands)*, 92(1), 49–54. https://doi.org/10.1016/j.healthpol.2009.02.004

Authors

Jaya Kolluri is currently a sophomore at The Winsor School in Boston, MA. Her interests include but are not limited to, Biology, Chemistry, and Computer Sciences. Jaya plans to further work in this field of computational biology.