

Discovery on Birth Weight

Team 8:

Kevin Castillo

Sushrutha Gujjalwar

Marcelo Tournier

Wei Quan

Stephanie Ziritt

Andy Chen

Abstract

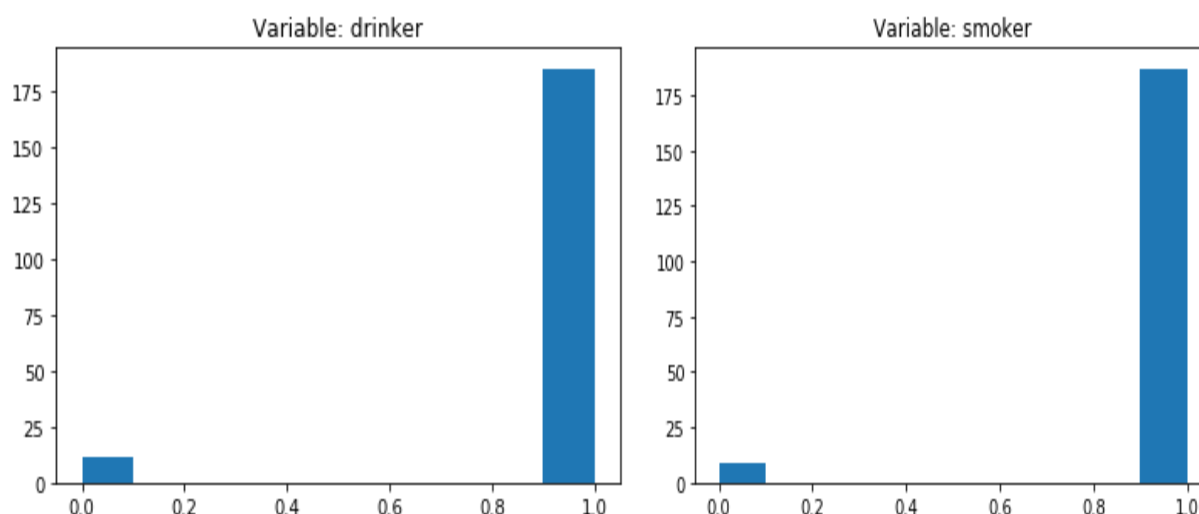
This report has the objective of present and discuss an analysis of risk factors during pregnancy, related to baby weights. To identify insights and predict preventable issues, we studied data of 18 variables in 196 babies, including basic information of the parents such as age, ethnicity, education, drinking, smoking behaviors and gender of baby. The finding indicates the influence of mother's age larger than 54, father's age, father's education, consumption of cigarettes and alcohol. A predictive model with a score of 0.703 (explaining 70.3% of the variability in risk factors influencing babies' weights) emphasizes that mothers aging over 54 years old should be aware of the risk in child birth weight and be cautious with drinking and smoking behavior. At the end, we make a recommendation to put efforts in data collection and feature gathering, to improve preemptive actions based on analytics. A classification model is also suggested to work better with predicting abnormality of birthweight.

“A baby is an inestimable blessing.” – Mark Twain. Babies are blessings for their parents and if you were to ask different parents how much of a blessing their babies were, it wouldn’t take long and much analysis that it’s not an easy task to quantify the amount of joy their baby brings to their lives. For this reason and many others, it’s important to make sure that parents are well informed of what are important features of giving birth to a healthy baby and consider what features can cause a baby to be more likely to be born at an unhealthy weight. Studies have shown that “Babies born prematurely or with low birth weight have an increased risk of early death, and those who survive are more likely to have chronic ill health or disability.”

Considering this, we analyzed different features of pregnancy from both the mother and the father to help us predict the baby’s birth weight. This analysis is very important because we will be able to provide insights and our recommendations regarding what are important things to consider in order to increase a baby’s chance of being born healthy and also increase our predictions.

The dataset we analyzed contained 18 variables and 196 observations, including basic information of the parents such as age, ethnicity, education, drinking, smoking behaviors and gender of baby. Before analyzing the dataset, we identified what would be healthy birth weight ranges and according to Madhu Desiraju, MD from KidsHealth.org, a baby is considered in the low birth range if they are born with less than 2500 grams (2.5kg) and high birth weight if they are born 4000 grams (4kg). This range was important to identify because we wanted to create a model where we could predict the baby’s birth weight and see if it is within the healthy range. We also removed two variables regarding Apgar Score from the analysis and model building,

since they are scores gathered after a baby is born and would not be helpful to predict the baby's birth weight.



As we all know, drinking and smoking during pregnancy should have negative effect on baby's health which is why they were the first features that we looked at. It is shown in the data that drinking and smoking behaviors are the most significant variables that will negatively affect the birth weight. However, more than 90% of the mothers have reported drinking or smoking during their pregnancy, while the research from the Center for Disease Control and Prevention shows that 15% of the mothers have reported smoking and 10% drinking. This makes it seem that the dataset less representative of the whole population.

The next variable we considered, mother's age, with an average of 40 years in our dataset, is one of the most significant variables affecting birth weight. It is more likely to have a low-weight baby as the mother gets older. Father's age, averaged 39, shares similar characteristics with mother's age, while it has a smaller impact on the birth weight.

Interestingly, many of the other variables did not have a very significant relationship with the birthweight directly. For example, we believed that the ethnicity of the parents along with the

prenatal visits or possibly starting prenatal care too late would show some relationship with the birth weight but unfortunately, in our analysis it did not. However, with education, the father's was interesting because after noticing the distribution, we decided to experiment by transforming the variable into its log form. This essentially meant that there was a non-linear relationship but by transforming it, we could discover how it could affect birth weight. What we discovered is that the average of the baby's birth weight would go up as the education of the father goes up. These features: the drinks per week, cigarettes per day, the mother's age, more specifically if the mother was over the age of 54, the father's age and the father's education were the features that we implemented in our prediction model. With the prediction model developed, we are able to achieve a R^2 score of 0.703 in new (not analyzed) data - which means that we are able to explain 70.3% of the factors involved with the variation of baby weight. Moreover, we can predict the child weight at birth with an average error of 277 grams. With testing different features and trying to create new variables, we believed this was the best model because of the combination with its predictive power and avoiding overfitting the data.

To summarize, what we gathered from our model is that the number of cigarettes, the amount of drinks, and the mother's age impacts a baby's birth weight the most. The more that the mother either smokes or drinks, the more likely the baby is to be born in an unhealthy low birth weight. The same relationships exist with the mother's age and we identified in our analysis that after the age of 54, is where we saw the strongest relationship of helping predict the baby's birth weight. In conclusion, after our analysis we have thought of several things to consider improving our prediction power for birth weight. First, as always, we should gather more data and observations. As the current dataset includes only about 9% of overweight and 8% underweight babies, there is

not enough observations to conclude contributing factors to the abnormality. In the overweight subset, independent variables, such as cigarettes, drinks, the mother's age and the father's age, all indicate healthier prenatal behavior and more suitable fertile age of both parents. In order to get profound insights on the factors to abnormal birth weight, we need to increase sample size of overweight and underweight newborns.

In addition to increasing sample size, it is also recommended to collect more information related to parents. For example, another feature that could be collected and help improve our predicting power would be the mother's weight before and during pregnancy. We discovered that a study from (AOGS), according to Wiley-Blackwell, reveals that both pre-pregnant weight (body mass index, BMI) and weight gain in pregnancy are important predictors of babies' birth weight.

Lastly, we would like to mention another possibility of improving our predicting power. We were tasked with being able to predict the baby's birth weight and since we were trying to predict an exact number, we saw that Linear Regression was the best model for this. However, there are complications with Linear Regression in this specific case because in most cases, Linear Regression is helpful to identify if there's a positive or negative relationship between two different variables. For example, in most cases, you would see that the more a student studies, the higher their mark would be. In the case with babies birth weight, although we have identified that the more a mother smokes, drinks or is over the age of 54, the more likely she is to give birth to a low weight baby, we have to consider that a low weight baby is not the only risky complication. The most common cause of high weight babies (4kg) is Diabetes during

pregnancy. This is a high risk condition, not only for the child (there is a high risk of bone and brain lesions for a big baby during labor), but for the mother, because Pregnancy Diabetes increases the risk of the woman to be diabetic in the future. We faced a problem when trying to identify factors of high weight at birth, because we didn't have any data related to diabetes risk or follow-up, such as the mother's weight, or abdominal circumference (cheap and easy measurements to collect), or the pregnant's blood sugar tests. What makes a baby's weight too low and a baby's weight too high. If the major problem is trying to identify whether the baby's weight will be normal or in a risky weight (either too high or too low), though limited by the small sample size of risky weight data within current dataset, we can consider predicting using a classification model in the future.

Since our research analysis and the model outcomes point to the terrible effects of alcohol and cigarettes during pregnancy, the algorithm can be used to influence women who drink and/or smoke, and are planning to have children, to change their behavior, showing them the potential risks for the baby, based on actual risk data from the woman. Nowadays, with the help of in vitro fertilization, births by women in their 50s have increased by more than 165 percent from 2000 to 2013 (Willets). The algorithm can also be used to help potential mothers understand the risk of getting pregnant at a much older age such as in their mid 50's and higher.

