Executive summary

Low birth weight is one of the most common indicators of infant health. While it is not this report’s objective to establish causation or identify main factors, we are trying to find main associations which might cause low birth weight. A total of 1236 babies from a US hospital were evaluated in applied statistical methods including an unpaired t-test, analysis of variance (ANOVA) and a simple linear regression model.

We are aiming to measure mean baby birth weight of smoking and non-smoking mothers is compared using a t-test in order to investigate a difference effected by smoking habit during pregnancy.

The ANOVA compares the mean baby birth weight of three different ethnicities of mothers to explore differences between the considered ethnic groups 0-5=white 6=mex 7=black 8=asian 9=mixed 99=unknown. The linear regression model intends to firstly describe the relationship of infant weights and multiple variables and further on to create a prediction model that allows to foresee children’s birth weights some significant variables.

This study found that there is a difference in mean baby birth weights between the babies who have smoking mothers and non-smoking mothers. Statistically significant differences in the mean birth weight were found between the ‘White’ and ‘Other’ as well as the ‘White’ and ‘Black’ ethnic groups. This, in fact, suggests that the ethnicity can be a determinant for infant weight. The simple linear regression reveals a weak linear relationship between the mother’s weight and the baby weight. The fit of the regression is insufficient to act as predictive model.

Bootstrapping?

Despite this, future research using additional exploratory factors influencing weights of infants could be used in multivariate regression models.

All these findings are highly limited in its application because of the limited sampled population, as it is very small and collected in 1964. Further research in this topic could deepen our understanding of this connection but would fail to help in preventing low birth weights. Nonetheless, the findings in this report allow to point research into an adequate direction. The contribution of this analysis is to identify mother’s smoking habits and ethnicity as possible determinants of children birth weight along with other attributes that might affect birth weight.

Introduction

Previous studies shown low birth weight “is the single most important factor affecting neonatal mortality and a significant determinant of post-neonatal mortality”. Naturally, there is an elevated interest of parents to avoid and therefore also to explore the determinants of low birth weights. In this study we are aiming to test certain factors and their relationship to birth weights. This report investigates the effects of mothers’ attributes mainly smoking habit and ethnic race on the weight of newborn babies, in order to make possible future suggestions about prevention of low birth weights.

The dataset used for this report consists of 1236 observations of babies collected from a US hospital in the 1960s.

Which variables we have selected?

In this data analysis three main suggestions are that there is a difference on babies’ weight due to races, that there is evidence that smoking has a negative influence on babies’ weight and finally the relationship between mothers’ weight and babies’ weight is possibly not simple linear.

methods

Findings

Explanation Final Model Selection

Using the Adjusted R squared and p values to determine the covariates by backward selection.

The R squared measures the percentage of change in the weight that can be explained by the independent variables. This is done by taking the standard error (distance of data points from the regressed line) and dividing it by the total variance within the weights (distance of weights from the mean weight). Our full model has a higher R square than the updated model. This does not necessarily mean that the variables are all significant determinants of baby’s weight. As you add more and more variables to a model the R squared will increase as each variable helps explain away the squared errors. Thus, we use the adjusted R squared which has a component that helps determine if the variable we added or subtracted was statistically significant in determining the baby’s weight. If we remove significant determinants the percentage of the change in weights will go down and vice versa. Thus, we can conclude that mother’s age did not significantly impact the baby’s weight according to the observations given to us.

Continuing this process, we use p-values to determine which variable to remove until we reach the point where removing or adding variables lowers our adjusted R-squared.