Analysis of Low Birth Weight Dataset

## Abstract

Low birth weight (LBW), a critical indicator of neonatal health, remains a persistent global health challenge. Identifying factors contributing to LBW is essential for targeted interventions and improving maternal and child health outcomes. In this study, we employ logistic regression analysis using the well-documented [Hosmer-Lemeshow dataset](https://search.r-project.org/CRAN/refmans/COUNT/html/lbw.html) to predict the likelihood of LBW, focusing on maternal characteristics and behaviors such as smoking status and age. The findings provide insights into the key predictors of LBW, demonstrating the utility of predictive models in public health contexts.

## Introduction

Low birth weight, defined as a birth weight of less than 2500 grams, is associated with increased risks of infant morbidity and mortality, developmental issues, and chronic diseases later in life. Numerous studies have explored the determinants of LBW, highlighting maternal factors such as age, socioeconomic status, prenatal care, and smoking behavior.

The Hosmer-Lemeshow dataset, which has been extensively used to demonstrate logistic regression techniques, provides a valuable resource for exploring LBW. This dataset contains records of 189 infants and includes information on maternal characteristics like age, smoking status, and parity, allowing for the development of predictive models. Logistic regression, a powerful tool for binary outcome prediction, is applied here to determine the factors most significantly associated with LBW.

## Methods

The dataset consists of 189 observations with the following variables:

* **Low (outcome)**: Binary indicator of low birth weight (1 if LBW, 0 otherwise).
* **Age**: Age of the mother in years.
* **Lwt**: Weight of the mother at the last menstrual period.
* **Race**: Categorical variable indicating race (White, Black, Other).
* **Smoke**: Binary variable indicating smoking status during pregnancy.
* **Ptl**: Number of premature labors.
* **Ht**: History of hypertension (1 if present, 0 otherwise).
* **Ui**: Presence of uterine irritability (1 if present, 0 otherwise).
* **Ftv**: Number of physician visits during the first trimester.

## Results

* The mean birth weight of a child is 2944.5873016.
* The rounded birth weight of a child is 2945.

Figure 1 illustrates the boxplot of mother’s weight in relation to low birth weight.

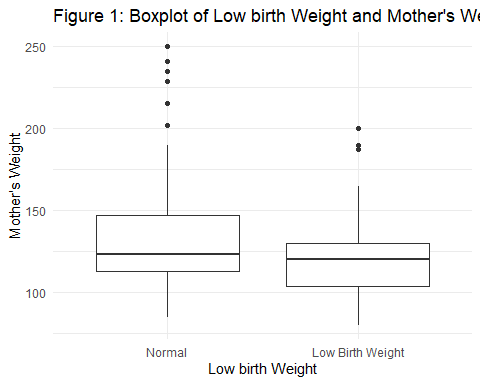
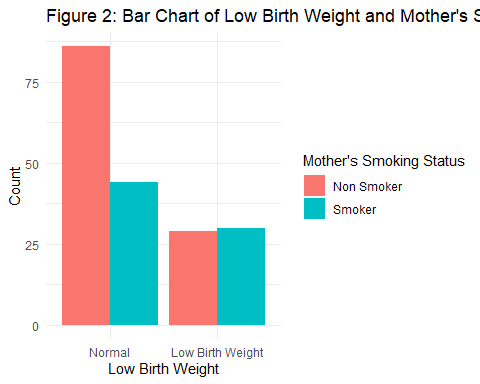


Figure 2 presents a bar chart depicting the counts of mothers categorized by smoking status in relation to low birth weight.



### Linear Regression

model <- lm(bwt ~ lwt + smoke , data = df)  
  
model

Call:  
lm(formula = bwt ~ lwt + smoke, data = df)  
  
Coefficients:  
(Intercept) lwt smokeSmoker   
 2501.125 4.237 -272.081

report(model)

We fitted a linear model (estimated using OLS) to predict bwt with lwt and  
smoke (formula: bwt ~ lwt + smoke). The model explains a statistically  
significant and weak proportion of variance (R2 = 0.07, F(2, 186) = 6.76, p =  
0.001, adj. R2 = 0.06). The model's intercept, corresponding to lwt = 0 and  
smoke = Non Smoker, is at 2501.13 (95% CI [2045.73, 2956.52], t(186) = 10.84, p  
< .001). Within this model:  
  
 - The effect of lwt is statistically significant and positive (beta = 4.24, 95%  
CI [0.90, 7.57], t(186) = 2.51, p = 0.013; Std. beta = 0.18, 95% CI [0.04,  
0.32])  
 - The effect of smoke [Smoker] is statistically significant and negative (beta  
= -272.08, 95% CI [-480.39, -63.77], t(186) = -2.58, p = 0.011; Std. beta =  
-0.37, 95% CI [-0.66, -0.09])  
  
Standardized parameters were obtained by fitting the model on a standardized  
version of the dataset. 95% Confidence Intervals (CIs) and p-values were  
computed using a Wald t-distribution approximation.

## Discussion

The strong association between smoking and LBW corroborates findings from numerous epidemiological studies. Smoking during pregnancy is known to reduce the oxygen supply to the fetus, impairing growth and leading to lower birth weights[@kataoka2018]. The inverse relationship between maternal weight and LBW is also well-documented, with underweight mothers often having inadequate nutritional stores to support fetal growth [@rumrich2020].

This study underscores the importance of targeting maternal smoking cessation programs to reduce LBW prevalence. Additionally, weight gain monitoring during pregnancy can provide early signals of potential risks, enabling healthcare providers to intervene appropriately.

One limitation of the study is the dataset’s relatively small size, which may reduce the power to detect smaller associations. Future research could benefit from larger datasets, more detailed socio-economic variables, and longitudinal data to track pregnancy outcomes over time.

**Conclusion:** Logistic regression analysis of the Hosmer-Lemeshow dataset reveals that maternal smoking and weight are key predictors of LBW. Public health initiatives focused on reducing smoking during pregnancy and promoting maternal health through nutrition and prenatal care are essential for reducing LBW and improving neonatal outcomes.