Ashaen Damunupola - 23883137

STAT2402 Assignment 2

Executive Summary

This report examines the effect of maternal smoking and other factors on low birthweight in infants. A total of 187 data records were analyzed, including variables such as maternal age, weight, race, smoking status, hypertension, uterine irritability, and birthweight. After exploring several models, the logistic regression model focusing on key predictors like smoking, maternal weight, race, hypertension, and uterine irritability (model2) was selected as the best fit.

model2 was preferred over other more complex models with interaction terms, mainly because it provided a superior fit, while also being considerably simpler. The model showed that smoking as well as hypertension were strong predictors of low birthweight, with smokers being 2.71 times more likely to have a low-birthweight baby, and maternal hypertension increasing the risks of a low birthweight baby by 6.31 times.

Introduction

Low birthweight, which in this case is described as less than 2.5kg or 2500g, can be a critical health issue that could eventually result in long term stunted development, and in the worst-case scenario: increased infant mortality. Identifying factors of causation is crucial for early intervention and prevention of said issues.

One well-researched maternal factor that leads to low birthweight in newborns is smoking during pregnancy, which has been found to negatively affect cognitive function and increase risk for chronic diseases and cardiovascular disorders later in life [1]. Not only does smoking reduce birthweight it may also have epigenetic effects.

Danileviciute et al. (2012) used multiple linear regression models to estimate the relationship between maternal smoking, genetic polymorphisms and birthweight [3]. Their best model for predicting birthweight reduction due to smoking and genetic factors was:

Birthweight Reduction = $\beta_{\square} + \beta_{\square}$(Smoking) + $\beta_{\square}$(GSTT1-null) + $\beta_{\square}$(GSTM1-null) + ε

- where **GSTT1-null** and **GSTM1-null** are genetic polymorphisms that modify the effect of smoking on birthweight.

The results indicated significant reductions in birthweight in infants born to mothers carrying the GSTT1-null and GSTM1-null genotypes. However, this model suffers from a small sample size for certain genotypes as well as limited generalization in the chosen population. This is because the population included primarily Lithuanian women, which may not represent smoking behaviours in other parts of the world.

The Japan Environment and Children’s Study by Suzuki et al. confirms the detrimental effects of smoking on birthweight; however, their findings must be interpreted with caution as the study relied on self-reported smoking data, which introduces bias if the population group under-reported [2].

The dataset analyzed in my study includes 187 observations from mothers and babies. It contains variables such as maternal age, weight, race, smoking status, medical history, and of course the baby's birth weight. There appears to be no missing data. Our model focuses on the direct impact of smoking and other factors on low birthweight, aiming to extend the understanding of this relationship through statistical modelling. We shall hypothesize that maternal smoking is one of the primary factors leading to low birthweight in infants.

Methodology

To examine the relationship between smoking, as well as other variables, and birthweight, explanatory data analysis (EDA) will be performed on the birth dataset. The EDA will consist of histograms to visualize the distribution of birthweight and bar plots to observe the proportions of low birthweight across the various categorical variables in the dataset. Missing values will be checked for as well.

Following the EDA, logistic regression will be used to model the probability of low birthweight as a function of maternal factors. Interaction terms will be considered. The model will then be reduced to significant terms only, using a stepwise selection method, as a simple model is preferred for ease of analysis.

All statistical analysis will be conducted in the R statistical environment [4]. Statistical significance will be taken at $\alpha = 0.05$ (5%).

Results

A logistic regression model: model1 was initially fitted to the data. This model included all available predictors. Then, the p-values were analyzed to find out which predictors were the most significant. Several interaction terms relating to the variable 'smoke' were added to the model to explore if they would result in a superior fit (models 3, 4, 5 & 6 where the interaction terms were smoke * race, smoke * ht, smoke * ui, smoke * lwt), however they all proved to be statistically worse than the original. This is because the p-values of the interaction in every case revealed to be quite large.

Therefore, model1 proceeded to be refined through a stepwise selection process (AIC). This resulted in model2, which excluded the non-significant predictors: age and physician visits. The AIC value of model2 was 218.4 which is an improvement over model1, which had an AIC value of 221.8.

The next step was to perform a drop in deviance test to make sure that model2 is a better fit. The test showed no significant difference between the full model and the nested, simpler model. The large p-value (0.7345) indicated that removing these variables had no significant effect. Therefore, the simpler model with less predictors was indeed better.

The final step was to compare diagnostic plots for model1 and model2 (see Figure 1). The results of the graphs showed that model2 handled the data better as there was less pronounced non-linearity in residuals vs fitted plot, residuals were closer to normal distribution in the Q-Q plot, reduced heteroscedasticity in the scale-location plot and even though high leverage points existed in the residuals vs leverage plot, they were less impactful than in model1.

As a result, the final model (model2) is:

$$\log(\text{Low Birthweight}) = -0.575 - 0.01396 \times \text{lwt} + 0.45498 \times \text{race} + 1.8426 \times \text{ht} + 0.99579 \times \text{smoke} + 0.82035 \times \text{ui}$$

Odds ratios were calculated to interpret the impact of predictors on a binary outcome such as low birthweight. For smoking, the odds ratio is an indicator of how much mor likely a smoker is to have a low birthweight baby than a non-smoker. Here, it was revealed that mothers who smoke are 2.71 times more likely to have a low-birthweight baby.

model2 was then visualized to show how well it fits (see Figure 2). The graph shows that most predicted probabilities for low birthweight fall between 0.3 and 0.6, indicating that while the model can estimate the likelihood of low birthweight, it lacks absolute certainty. However, this may be due to unmeasured factors such as genetics.

Discussion

All retained variables in the final model, including smoking, maternal weight (lwt), race, hypertension (ht), and uterine irritability (ui) appear to be significant predictors of low birthweight.

- Smoking has a strong positive effect on the odds of low birthweight babies, with an odds ratio of 2.71, indicating that smoking nearly triples the odds of low birthweight.
- For every unit increase in maternal weight, the odds of low birthweight decrease by a factor of 0.99, meaning higher maternal weight reduces the likelihood of low birthweight slightly.
- “other” racial groups have an increased likelihood of low birthweight, with the odds increasing by 1.58 times for certain racial categories.
- Hypertension was one of the strongest predictors, with mothers who had hypertension being 6.31 times more likely to give birth to a low-birthweight baby.
- Mothers with uterine irritability are 2.27 times more likely to have a low-birthweight baby.

A drop in deviance test between the full and refined models showed no significant loss in model fit when removing non-significant variables. The visualizations from the exploratory data analysis supported these findings, showing higher proportions of low birthweight among smokers and hypertensive mothers.

These findings support the hypothesis that maternal smoking is indeed a major factor for the birth of low birthweight babies. My study also agrees with the study by Ardissino et al., which demonstrated that maternal hypertension is a significant risk factor for low birthweight [5]. Their Mendelian randomization study revealed a direct causal relationship between elevated systolic blood pressure and reduced birthweight

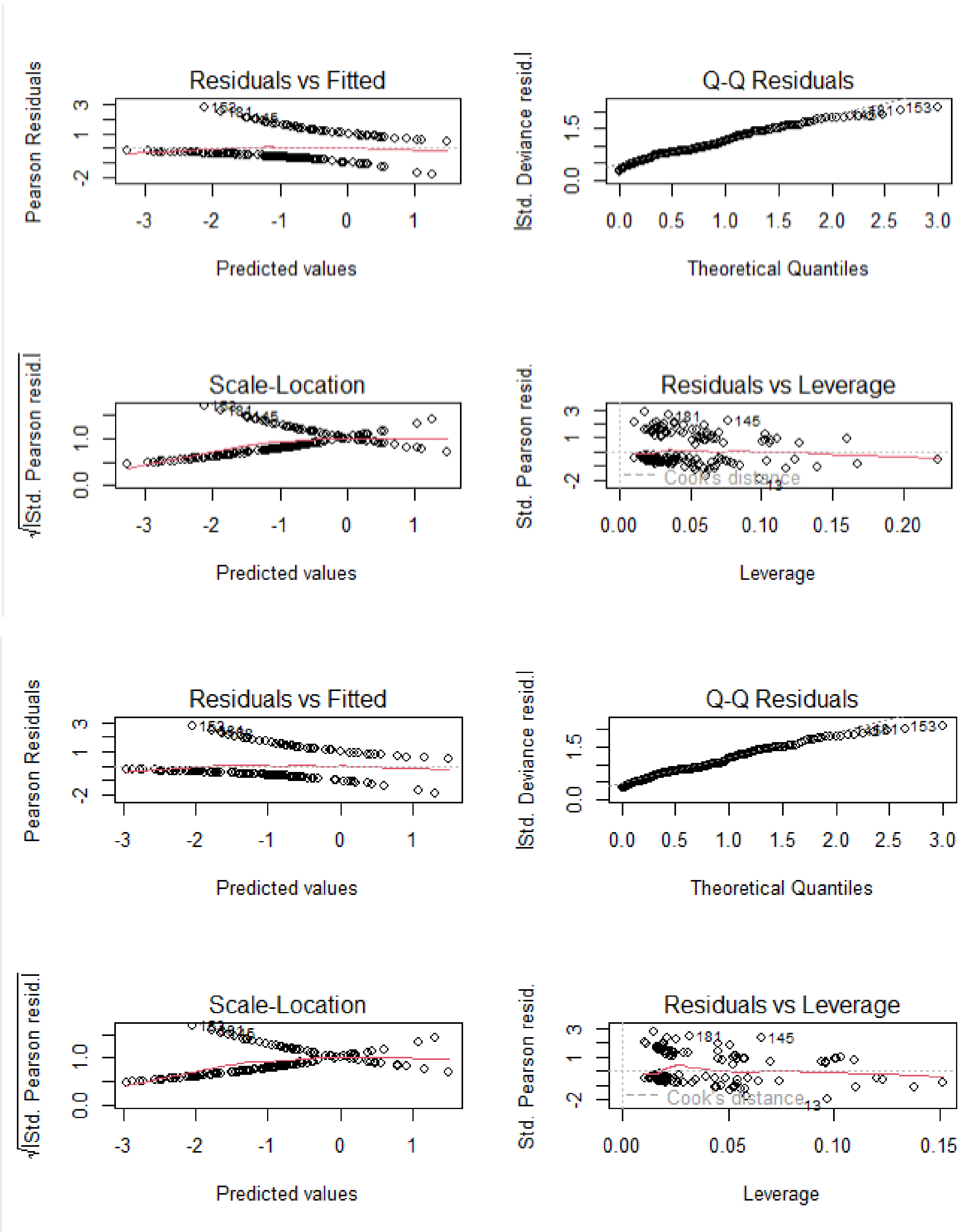

Figure 1: Comparison of diagnostic plots for model1 and model2

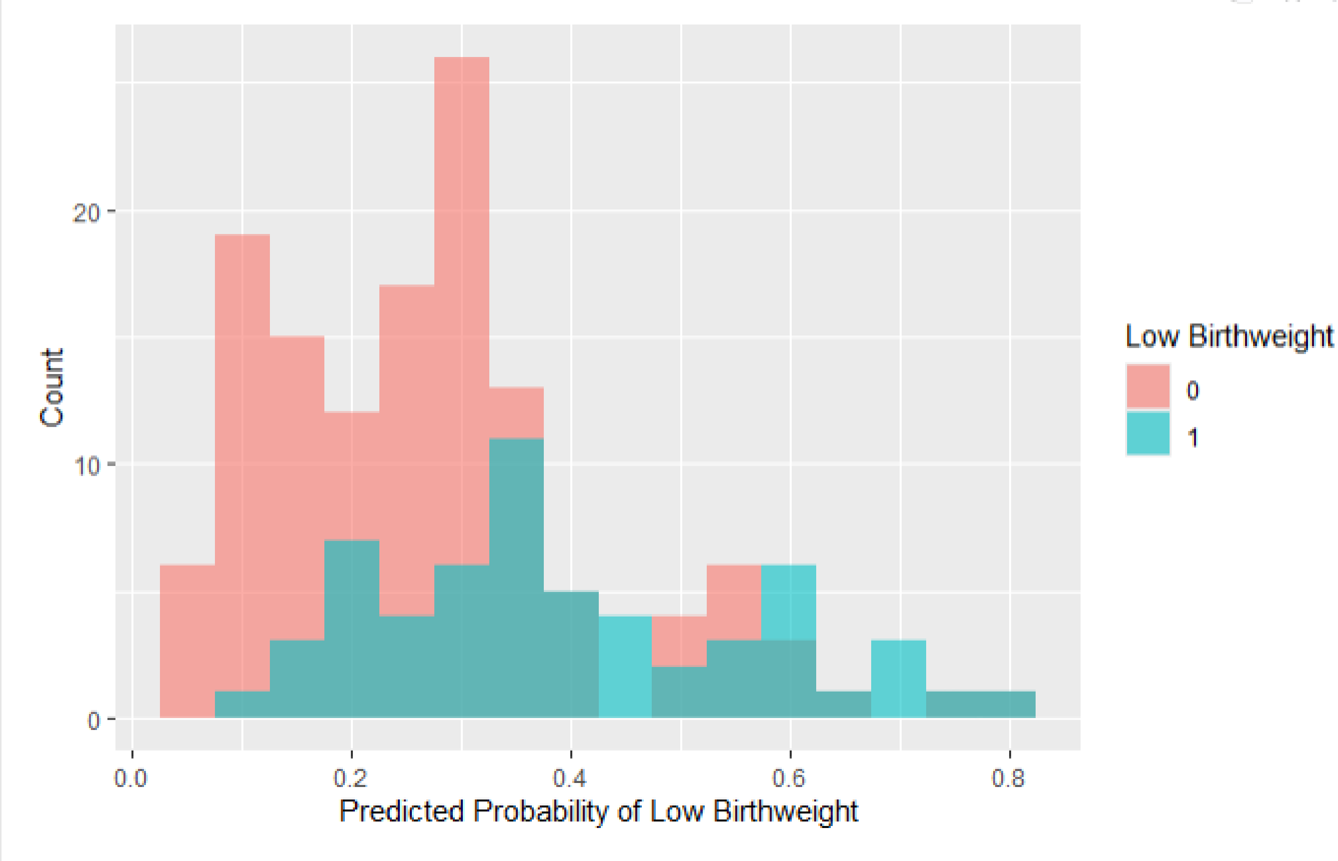

Figure 2: Predicted Probabilities of Low Birthweight (model2)

References:

1. Witt WP, Wisk LE, Cheng ER, Hampton JM, Hagen EW. Preconception mental health predicts pregnancy complications and adverse birth outcomes: a national population-based study. Matern Child Health J. 2018 Apr;22(4):557-68.

2. Suzuki K, Shinohara R, Sato M, et al. Association between maternal smoking during pregnancy and birth weight: an appropriately adjusted model from the Japan Environment and Children’s Study. J Epidemiol. 2016;26(7):371-7.

3. Daniłeviciute A, Grazuleviciene R, Paulauskas A, Nadisauskiene R, Nieuwenhuijsen MJ. Low-level maternal smoking and infant birthweight reduction: genetic contributions of GSTT1 and GSTM1 polymorphisms. BMC Pregnancy Childbirth. 2012 Dec;12(1):161.

4. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2023.

5. Ardissino M, Völlmecke C, Petkova MV, Morris AP, Chappell LC, Lawlor DA, et al. Maternal hypertension increases risk of preeclampsia and low fetal birthweight: genetic evidence from a Mendelian randomization study. *Hypertension*. 2022;79(1):163–72.