Advanced Data Programming with R - Homework 1

Vanessa Carvalho - 25267189

June 3, 2025

1 Introduction

For this homework I have chosen a dataset of **Risk Factors Associated with Low Infant Birth Weight**. The data were retrospective collected at Baystate Medical Center, Springfield, Massachusetts during 1986. It investigates the risk factors for low birth weight in newborns, with the goal to explore whether variables such as maternal age, weight, smoking status, and race are associated with this outcome.

It includes 189 observations of live births and contains 10 variables to be considered

- `low`: indicator of birth weight less than 2.5 kg.
- `age`: mother's age in years.
- `lwt`: mother's weight in pounds at last menstrual period.
- `race`: mother's race (1 = white, 2 = black, 3 = other).
- `smoke`: smoking status during pregnancy.
- `ptl`: number of previous premature labours.
- `ht`: history of hypertension.
- `ui`: presence of uterine irritability.
- `ftv`: number of physician visits during the first trimester.
- `bwt`: birth weight in grams.

2 Statistical Analysis

The aim was to develop a logistic regression model for predicting low birth weight, using .

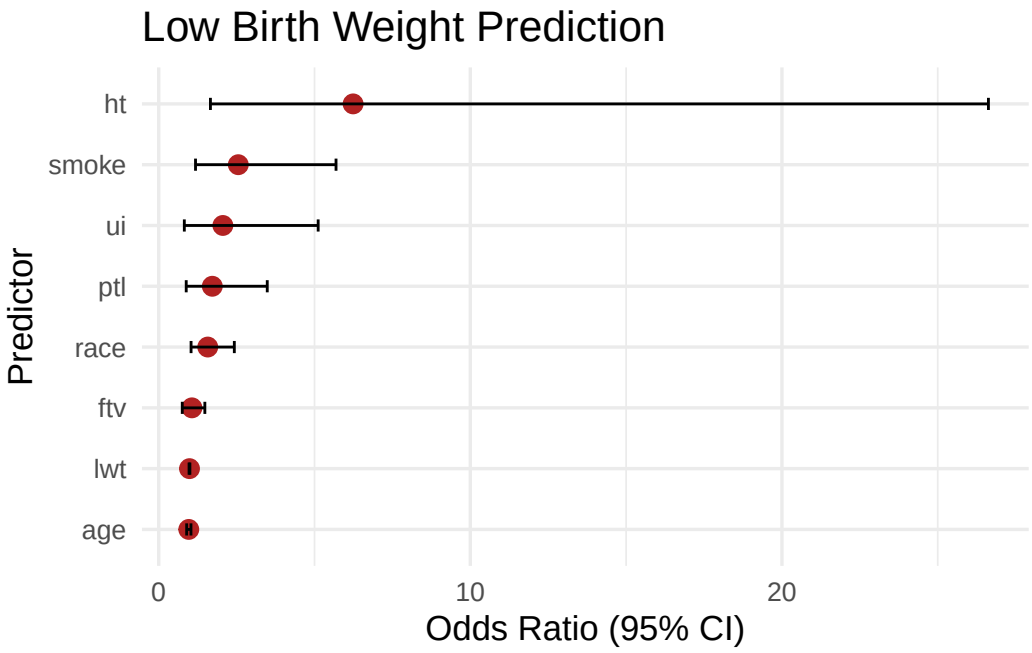

Figure 1: Odds Ratios for Predicting Low Birth Weight using Logistic Regression

The logistic regression analysis was conducted to identify maternal and lifestyle factors associated with an increased risk of low birth weight. The outcome variable was binary (`low`: Low vs Normal birth weight), and multiple predictors including age, weight, smoking status, hypertension, and race were considered.

The results show that **smoking** during pregnancy had a strong and significant association with low birth weight, with an odds ratio well above 1. Similarly, **maternal hypertension (`ht`)** and certain race categories appeared to increase the odds of low birth weight. These findings are consistent with known clinical risk factors.

Conversely, variables such as **mother's age** and **number of prior pregnancies** (`ptl`, `lwt`, etc.) did not show a significant effect in this dataset. Some of the confidence intervals crossed 1, indicating statistical non-significance.