# Healthcare and Medical Analytics Individual Coursework

**Examining the Prevalence of Low Birth Weight among Infants Born to Black Mothers: A Comprehensive Study**

## Introduction

Low birth weight, defined by the World Health Organization (WHO) as a new born weighing less than 5.5 pounds or 2.5 kilograms at birth, represents a significant global public health concern. Many cases of low birth weight can be attributed to factors such as poor maternal nutrition, inadequate prenatal care, and overall maternal ill health (WHO, n.d.). In the United States, this issue is particularly pressing, affecting 1 in every 12 infants, and underscoring its magnitude (Cleveland Clinic, 2023).

Among the various risk factors associated with low birth weight, race has been identified as a significant consideration. Previous studies have investigated racial disparities in low birth weight and explored the influence of socio-economic and environmental factors on birth outcomes. (Clay, Woodson, and Kersh ,2021) found that being Non-Hispanic Black and receiving government rent assistance increased the likelihood of low birth weight, while factors such as marriage, healthcare access, and living in public housing were associated with a reduced likelihood. Another study by (Catov et al. ,2015) examined data spanning from 1997 to 2011 and revealed a concerning trend of low birth weights, particularly among African American infants. This trend was potentially linked to maternal health risks such as obesity and hypertension. The authors emphasised the need for further research on maternal health status, placenta characteristics, and prenatal care practices to better understand the implications for birth outcomes. (Nkansah-Amankra et al. ,2009) focused on the impact of income inequality and social support on birth outcomes, specifically low birth weight and preterm delivery. Their findings indicated that moderate levels of income inequality in neighbourhoods were associated with low birth weight, while low social support increased the risk for both low birth weight and preterm births. The study highlighted the heightened vulnerability of non-Hispanic Black mothers to adverse birth outcomes, influenced by neighbourhood deprivations associated with low income and limited social support.

Based on these studies, this report aims to investigate the prevalence of low birth weight among infants born to Black mothers, employing a logistic regression analysis to examine the influence of socio-economic and environmental factors. By analysing the intersection of race, income inequality, education, and physical health, this research contributes to the understanding of the complex factors impacting low birth weight and birth outcomes.

## Descriptive Statistics

The dataset used for this report is The National Longitudinal Study of Adolescent to Adult Health (AddHealth) dataset, consisting of five waves of survey responses. This analysis focuses on the fourth wave, which includes questions about general life, pregnancy, and live births. The comprehensive nature of the AddHealth dataset provides valuable insights into socio-demographic characteristics and health behaviours relevant to the investigation.

In this study, the records of the most recent births were included, specifically focusing on infants reported to have been born before the due date. This criterion allowed for the analysis of a total of 705 observations. To classify low birth weight, all the birth weights reported by the mothers as being 5 pounds and below were categorized as low birth weight, assigned a binary value of 1. Birth weights from 6 pounds and above were assigned a value of 0, indicating that the birth weights were not low. Among these observations, 562 (79.7%) were classified as normal birth weight, indicating birth weights above 5 pounds. Additionally, 140 (19.8%) were classified as low birth weight, and 3 (0.42%) had missing records for the infant's birth weight.

In this analysis, the variable of race plays a significant role in determining the birth weight of infants. Therefore, it is crucial to examine the distribution of this variable in relation to the birth weight outcomes. Among the 705 respondents, the distribution of race was as follows: 68.5% identified as White, 23.7% as Black, 1.42% as American Indian, and 2.7% as Asian American Pacific Islander and four dummy variables were created. It is noteworthy that American Indians and Asian American Pacific Islanders were underrepresented in this analysis compared to their respective population proportions.

The bar chart below illustrates the distribution of low birth weight cases among Black and White women based on educational attainment. It is important to note that the data for Asian American Pacific Islanders and American Indians is not included in this graph due to incomplete records for all levels of education. However, a comprehensive graph including all racial groups can be found in the appendix. Among the 59 recorded cases of low birth weight in Black women, the highest percentage (~50%) is observed among those who have completed some form of college or vocational training education. Similarly, for White women of the 75 records of low birth weight recorded, the highest percentage (~48%) of low birth weight cases also comes from those who have completed some form of college or vocational training education. It is worth noting that there is limited representation from individuals who did not

complete high school, accounting for only 7.09% of the total responses. This limited representation may explain why it is not showing that individuals without a high school degree have the highest percentages of low birth weight for both racial groups. Furthermore, the graph suggests that having a bachelor's degree has a significant impact in reducing the likelihood of having a baby with low birth weight, as this educational category shows the second lowest percentage of low birth weight cases among both Black and White women. This highlights the influential role of higher education in promoting healthier birth outcomes.

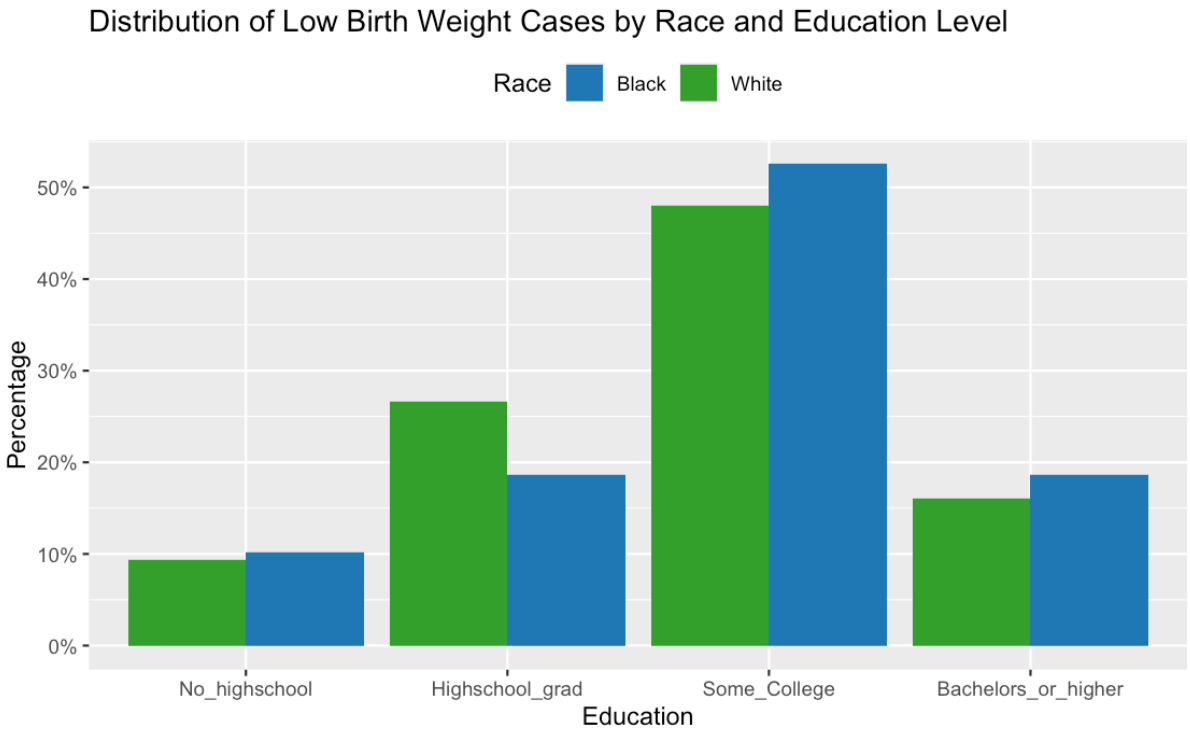



*Figure 1 Distribution of Low Birth Weight cases By Race and Education Level*

Dummy variables were used to represent educational levels, including "no high school degree," "graduated from high school," "completed some college," and "completed a bachelor's degree or higher." Income levels were categorized as "Low income," "Middle income," and "High income." The timing of births was indicated by a binary variable called "Actual_Preterm." Prenatal care access was denoted by a binary variable. "Multiparity status" was represented as a binary variable. Body mass index (BMI) was categorized into "Underweight," "Normal," "Overweight," "Obese I," "Obese II," and "Obese III." Smoking habits during pregnancy were captured by four variables. The gender of the newborn was indicated by a binary variable. Refer to Appendix A and B for a comprehensive description of variables and useful visualisations.

## Regressions

In the first logistic regression model, the association between the mother's race and low birth weight was explored. The results revealed that being a Black mother was significantly associated with approximately 140% (odds ratio = 2.40, $p < 0.001$) higher odds of having a child with low birth weight compared to being a white mother (reference category). This suggests that being a Black mother is a significant predictor of low birth weight in this study. On the other hand, the other race dummy variables (AAPI and American_Indian) did not show significant associations with low birth weight, which may indicate that these variables are not significant predictors of low birth weight in this sample or may require further investigation to fully understand their relationship. The AIC of this model was 690.09.

In the second logistic regression model, the association between low birth weight and more socioeconomic variables was examined. The variables included in the analysis were Race, Income, Education, and access to prenatal care. The results once again indicated that being a Black mother was significantly associated with 114% higher odds (odds ratio = 2.14, $p = 0.00126$) of having a child with low birth weight compared to being a white mother. This suggests that race plays a significant role in determining the likelihood of low birth weight.
Furthermore, being a high school graduate, as opposed to having some college/technical training education, was associated with approximately 109% higher odds (odds ratio = 2.092, $p = 0.00999$) of having a child with low birth weight. This highlights the importance of educational attainment as a contributing factor. However, the middle and upper income categories when compared to the low income category did not show a statistically significant effect on the odds of low birth weight, indicating that income may not be a significant predictor in this particular model. Similarly, although access to prenatal care was associated with a 55% decrease in the odds of low birth weight, this relationship did not reach statistical significance (odds ratio= 0.445, p-value = 0.21824). The AIC of this model was 563.83, indicating better fit than the previous model.

In the third regression model, variables related to perinatal outcomes, physical health and well-being were considered- the mother's BMI, multiparity status, smoking during pregnancy, newborn gender, and preterm birth. Results revealed important associations with low birth weight. Coming from a middle-income household, compared to a lower-income household, was linked to a significant 45% decrease in the odds of low birth weight. Multiparity, indicating multiple live pregnancies, showed a substantial 615% increase in the odds of low birth weight. Notably, being born preterm exhibited an extremely significant association, with an

approximately 1980% increase in the odds of low birth weight. Underweight BMI also played a significant role, with a 366% increase in the odds of low birth weight compared to a normal BMI. Additionally, being born as a male infant, as opposed to a female infant, was associated with a 52% decrease in the odds of low birth weight. The AIC of this model was 434.43 indicating a much better fit than the former two models. These findings highlight the complex interplay of factors contributing to low birth weight, encompassing income, multiparity, gestational age, maternal weight, and infant gender. The results of all the regressions can be seen in Appendix C. For this model, several statistical tests were carried out to measure and rank the goodness of fit. They can be seen in Appendix D.

The last set of regressions involved creating interaction terms between race and education, as well as race and Actual_preterm. These interaction terms were included due to their statistical significance in previous models. However, the regression results did not yield any new insights. None of the interaction variables showed statistical significance, and they had larger standard errors compared to the variables in the third model. The variables from the third model maintained similar statistical significance in this new model. Therefore, the inclusion of these interaction terms did not provide additional explanatory power or uncover any significant interactions between race, education, and Actual_preterm in relation to low birth weight.

The logistic regression models provided valuable insights into factors associated with low birth weight. Being a Black mother consistently emerged as a statistically significant predictor across all models, highlighting the role of race in birth outcomes. Educational attainment, specifically high school graduation, was also found to contribute to low birth weight. However, income and access to prenatal care did not show significant associations. Physical health and perinatal factors, such as BMI, multiparity, preterm birth, and infant gender, were significant predictors of low birth weight. These findings emphasize the complex interplay of socioeconomic and health-related factors in predicting low birth weight. Overall, the models demonstrate the importance of considering race, education, and physical health when understanding and addressing low birth weight.

## Limitations

One major limitation of this report is the reliance on self-reported variables, which introduces potential bias. Self-reported data may be subject to recall bias, impacting the validity of the

results. Another limitation is the exclusion of important variables like maternal age, alcohol consumption, stress, preeclampsia, and gestational diabetes. These variables are known to affect foetal development and should be considered in future analyses. Another limitation of this report is the omission of accounting for weights associated with the variables. Weighting variables help ensure that the sample accurately reflects the characteristics of the target population, accounting for different probabilities of selection and non-response. Failure to consider these weights may result in a potential distortion of the findings and their representativeness of the population under study. Future studies should prioritise the inclusion of appropriate weights to enhance the validity and representativeness of the observed associations between variables. Furthermore, certain racial groups, particularly American Indian and Asian American/Pacific Islander populations, are underrepresented in the survey. This limits the generalisability of findings and introduces potential biases in associations between race and low birth weight.

## Conclusion

The findings of this study highlight the significant association between being a Black mother and an increased likelihood of low birth weight. This underscores the importance of addressing racial disparities in birth outcomes and implementing targeted interventions for Black mothers. Additionally, educational attainment, particularly being just a high school graduate, was found to be associated with higher odds of low birth weight. This emphasizes the need to provide educational opportunities and support for mothers to improve birth outcomes. Although income did not show a significant effect on low birth weight in this study, the association between middle-income status and a decrease in low birth weight odds suggest the potential role of financial stability in reducing the risk. Comprehensive prenatal care and support for healthy behaviours are essential for improving birth outcomes. Policy recommendations include targeted interventions for Black mothers, access to quality prenatal care, educational opportunities, and resources to address maternal health risks. Efforts to reduce income inequality and promote economic stability can also contribute to improved birth outcomes. In conclusion, this study emphasises the need for comprehensive strategies to address racial disparities, educational attainment, maternal health risks, and social support to reduce the prevalence of low birth weight and improve the health outcomes of infants born to Black mothers.

## References

Catov, J.M., Lee, M., Roberts, J.M., Xu, J. and Simhan, H.N. (2015). Race Disparities and Decreasing Birth Weight: Are All Babies Getting Smaller? *American Journal of Epidemiology*, [online] 183(1), pp.15–23. doi:https://doi.org/10.1093/aje/kwv194.

Clay, S.L., Woodson, M.J. and Kersh, R. (2021). An Exploration of Socio-demographic, Economic, and Environmental Factors in Black/White Disparities in Low Birth Weight Outcomes: A Cross-Sectional Study. *Journal of Research in Health Sciences*, [online] 21(2), pp.e00511–e00511. doi:https://doi.org/10.34172/jrhs.2021.49.

Cleveland Clinic (2023). *Low Birth Weight*. [online] Cleveland Clinic. Available at: https://my.clevelandclinic.org/health/diseases/24980-low-birth-weight [Accessed 4 Jul. 2023].

Nkansah-Amankra, S., Dhawain, A., Hussey, J.R. and Luchok, K.J. (2009). Maternal Social Support and Neighborhood Income Inequality as Predictors of Low Birth Weight and Preterm Birth Outcome Disparities: Analysis of South Carolina Pregnancy Risk Assessment and Monitoring System Survey, 2000–2003. *Maternal and Child Health Journal*, [online] 14(5), pp.774–785. doi:https://doi.org/10.1007/s10995-009-0508-8.

WHO (n.d.). *Low birth weight*. [online] www.who.int. Available at: https://www.who.int/data/nutrition/nlis/info/low-birth-weight#:~:text=Low%20birth%20weight%20has%20been [Accessed 4 Jul. 2023].

# Appendices

## Appendix A: Variable Description

| Variable | Description | Source |
|---|---|---|
| AAPI | Dummy variable based on the interviewee's description of respondent's race.<br>= 1, if respondent is Asian American and Pacific Islander Race (AAPI)<br>= 0, if respondent is not AAPI | Wave 4- General Use Dataset |
| American_Indian | Dummy variable based on the interviewee's description of respondent's race.<br>= 1, if respondent is American Indian<br>= 0, if respondent is not American Indian | Wave 4- General Use Dataset |
| Black | Dummy variable based on the interviewee's description of respondent's race.<br>= 1, if respondent is Black<br>= 0, if respondent is not Black | Wave 4- General Use Dataset |
| White | Dummy variable which is also reference category for Race and is based on the interviewee's description of respondent's race.<br>= 1, if respondent is White<br>= 0, if respondent is not White | Wave 4- General Use Dataset |
| low_income | A dummy variable which also represents the reference income category for the model and indicates respondent's household income.<br>= 1, if respondent indicated their household income was below $49,999 i.e. values 1-8<br>= 0, if respondent's income level did not fall in this category | Wave 4- General Use Dataset |
| middle_income | A dummy variable which indicates respondent's household income<br>= 1, if respondent indicated their household income was greater than $50,000 and less than $150,000 i.e. values 9-11<br>= 0, if respondent's income level did not fall in this category | Wave 4- General Use Dataset |

| | | |
|---|---|---|
| upper_income | A dummy variable which indicates respondent's household income<br>= 1, if respondent indicated their household income was greater than $150,000 i.e., value 12<br>= 0, if respondent's income level did not fall in this category | Wave 4- General Use Dataset |
| No_highschool | A dummy variable which indicates respondent's highest educational level<br>= 1, if respondent indicated that they were either in the 8th grade or less or that they had finished some high school i.e., values 1 and 2<br>= 0, if respondent's education level did not fall in this category | Wave 4- General Use Dataset |
| Highschool_grad | A dummy variable which indicates respondent's highest educational level<br>= 1, if respondent indicated they had graduated from high school i.e. value 3<br>= 0, if respondent's education level did not fall in this category | Wave 4- General Use Dataset |
| Some_College | A dummy variable which also served as reference category that indicates respondents highest educational level<br>= 1, if respondent indicated had finished from some college or completed vocational training i.e. values 4,5,6<br>= 0, if respondent's education level did not fall in this category | Wave 4- General Use Dataset |
| Bachelors_or_higher | A dummy variable which indicates respondent's highest educational level<br>= 1, if respondent indicated had completed a bachelor's degree and other post graduate education i.e. values 9-12<br>= 0, if respondent's education level did not fall in this category | Wave 4- General Use Dataset |
| No_cigs | A dummy variable which also serves as a reference category to indicate respondents smoking frequency during pregnancy.<br>= 1, if respondent indicated that they did not smoke while pregnant i.e. value 0 | Wave 4- Pregnancy Dataset |

| | | |
|---|---|---|
| | = 0, if respondent did not fall in this category with regards to smoking during pregnancy | |
| Low_cigs | A dummy variable which indicates respondents smoking frequency during pregnancy.<br>= 1, if respondent indicated that they smoked a few cigarettes, but not every week i.e. value 1<br>= 0, if respondent did not fall in this category with regards to smoking during pregnancy | Wave 4- Pregnancy Dataset |
| Medium_cigs | A dummy variable which indicates respondents smoking frequency during pregnancy.<br>= 1, if respondent indicated that they smoked a few cigarettes a week, but not every day i.e. value 2<br>= 0, if respondent did not fall in this category with regards to smoking during pregnancy | Wave 4- Pregnancy Dataset |
| High_cigs | A dummy variable which indicates respondents smoking frequency during pregnancy.<br>= 1, if respondent indicated that they smoked 10 to greater than 31 cigarettes in a day while pregnant i.e. values 3-6<br>= 0, if respondent did not fall in this category with regards to smoking during pregnancy | Wave 4- Pregnancy Dataset |
| Prenatal_care | Binary variable used to indicate if mother had access to prenatal care during pregnancy:<br>= 1, if mother indicated that she had access to prenatal care<br>= 0, if mother indicated that she did not have access to prenatal care | Wave 4- Pregnancy Dataset |
| Underweight | A dummy variable used to indicate respondent's BMI classification.<br>= 1, if respondent was categorised as underweight <18.5 i.e. value 1<br>=0, if respondent was placed in a different category | Wave 4- General Use Dataset |
| Normal | A dummy variable which served as reference category that indicate respondent's BMI classification.<br>= 1, if respondent was categorised as normal 18.5- <25 i.e. value 2 | Wave 4- General Use Dataset |

| | | |
|---|---|---|
| | =0, if respondent was placed in a different category | |
| Overweight | A dummy variable used to indicate respondent's BMI classification.<br>= 1, if respondent was categorised as overweight 25-<30 i.e. value 3<br>=0, if respondent was placed in a different category | Wave 4- General Use Dataset |
| Obese_i | A dummy variable used to indicate respondent's BMI classification.<br>= 1, if respondent was categorised as obese I 30-<35 i.e. value 4<br>=0, if respondent was placed in a different category | Wave 4- General Use Dataset |
| Obese_ii | A dummy variable used to indicate respondent's BMI classification.<br>= 1, if respondent was categorised as obese II 35-<40 i.e. value 5<br>=0, if respondent was placed in a different category | Wave 4- General Use Dataset |
| Obese_iii | A dummy variable used to indicate respondent's BMI classification.<br>= 1, if respondent was categorised as obese III 40+ i.e. value 6<br>=0, if respondent was placed in a different category | Wave 4- General Use Dataset |
| Multiparity_status | A binary variable that indicates whether a mother has had multiple pregnancies or not.<br>= 1 , indicating that the mother has had more than one baby during the specific pregnancy. i.e values 2-5<br>= 0, indicates primiparity, indicating that the mother has had only one baby during the specific pregnancy. i.e. value 1 | Wave 4- Pregnancy Dataset |
| Actual_Preterm | A binary variable that indicates whether a live birth was preterm<br>= 1 , A value of 1 if baby was delivered more than 3 weeks before due date<br>= 0, if baby was delivered less than 3 weeks before due date | Wave 4- Live Births Dataset |
| baby_gender | A binary variable used to indicate gender of baby<br>=1, if baby was a boy<br>=0, if baby was a girl | Wave 4- Live Births Dataset |
| low_birthweight | A binary variable used if the newborn had low birth weight at time of birth | Wave 4- Live Births Dataset |

| | | |
|---|---|---|
| | =1, if baby weighed less than 5 pounds at birth<br>=0, if baby weighed greater than 5 pounds at birth | |

# Appendix B: Descriptive Statistics

Descriptive Statistics for Low Birth Weights Variable

| low_birthweight | Frequency | Percentage |
|---|---|---|
| 0 | 562 | 79.7163121 |
| 1 | 140 | 19.8581560 |
| NA | 3 | 0.4255319 |

*Figure 2 Descriptive Statistics for low birth weight variable*

Descriptive Statistics for Race Variable

| Race | Frequency | Percentage |
|---|---|---|
| AAPI | 19 | 2.695035 |
| American_Indian | 10 | 1.418440 |
| Black | 193 | 27.375887 |
| White | 483 | 68.510638 |
| NA | 0 | 0.000000 |

*Figure 3 Descriptive statistics for Race Variable*

Descriptive Statistics for Income Variable

| Income | Frequency | Percentage |
|---|---|---|
| low_income | 346 | 49.078014 |
| middle_income | 305 | 43.262411 |
| upper_income | 18 | 2.553192 |
| NA | 36 | 5.106383 |

*Figure 4 Descriptive Statistics for Income Variable*

Descriptive Statistics for Education Variable

| Education | Frequency | Percentage |
|---|---|---|
| No_highschool | 50 | 7.092199 |
| Highschool_grad | 103 | 14.609929 |
| Some_College | 372 | 52.765957 |
| Bachelors_or_higher | 179 | 25.390071 |
| NA | 1 | 0.141844 |

*Figure 5 Descriptive Statistics for Education Variable*

Descriptive Statistics for Prenatal Care Variable

| Prenatal_care | Frequency | Percentage |
|---|---|---|
| No Prenatal Care | 13 | 1.843972 |
| Prenatal Care | 616 | 87.375886 |
| NA | 76 | 10.780142 |

*Figure 6 Descriptive statistics for Prenatal Care Variable*

Descriptive Statistics for Baby Gender Variable

| Gender | Frequency | Percentage |
|---|---|---|
| Girl | 344 | 48.79433 |
| Boy | 361 | 51.20567 |
| NA | 0 | 0.00000 |

*Figure 7 Descriptive Statistics for Baby Gender Variable*

Descriptive Statistics for Actual Preterm Variable

| Status | Frequency | Percentage |
|---|---|---|
| Term | 576 | 81.702128 |
| Preterm | 121 | 17.163121 |
| NA | 8 | 1.134752 |

*Figure 8 Descriptive Statistics for Actual Preterm Variable*

Descriptive Statistics for BMI Variable

| BMI | Frequency | Percentage |
|---|---|---|
| Underweight | 14 | 1.9858156 |
| Normal | 198 | 28.0851064 |
| Overweight | 189 | 26.8085106 |
| Obese_i | 143 | 20.2836879 |
| Obese_ii | 70 | 9.9290780 |
| Obese_iii | 85 | 12.0567376 |
| NA | 6 | 0.8510638 |

*Figure 9 Descriptive Statistics for BMI Variable*

Descriptive Statistics for Smoking_Preg Variable

| Smoking_Preg | Frequency | Percentage |
|---|---|---|
| No_cigs | 509 | 72.198582 |
| Low_cigs | 28 | 3.971631 |
| Medium_cigs | 38 | 5.390071 |
| High_cigs | 54 | 7.659574 |
| NA | 76 | 10.780142 |

*Figure 10 Descriptive Statistics for Smoking_Preg Variable*

Descriptive Statistics for Multiparity Status Variable

| Multiparity_status | Frequency | Percentage |
|---|---|---|
| Uniparous | 617 | 87.517730 |
| Multiparous | 13 | 1.843972 |
| NA | 75 | 10.638298 |

*Figure 11 Descriptive Statistics for Multiparity Status Variable*

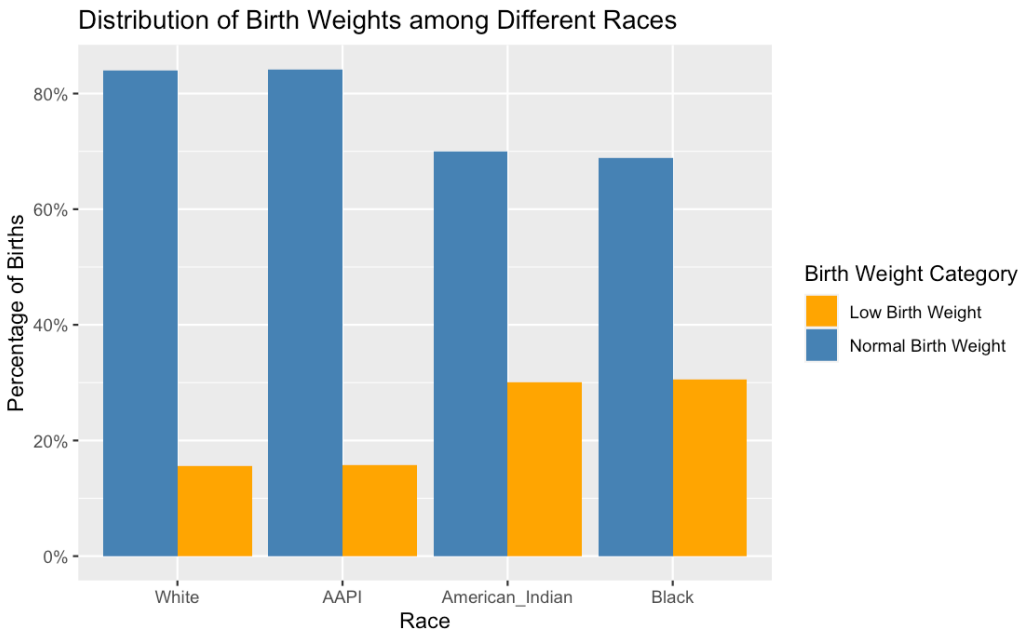



*Figure 12 Graph Showing Distribution of Birth Weights Among Different Races*

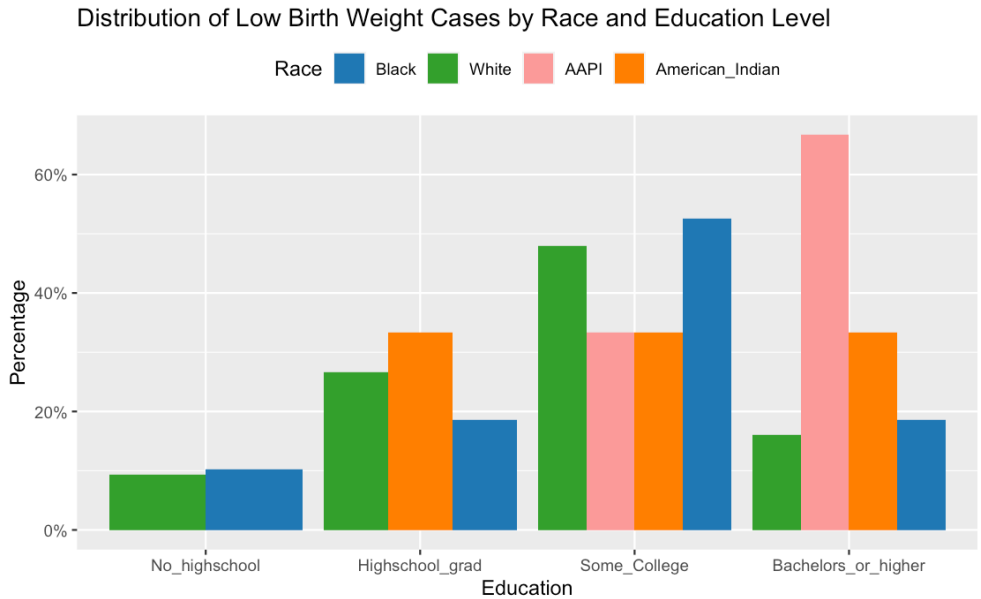



*Figure 13 Graph Showing Distribution of Low Birth Weight Cases by Race and Education Level*

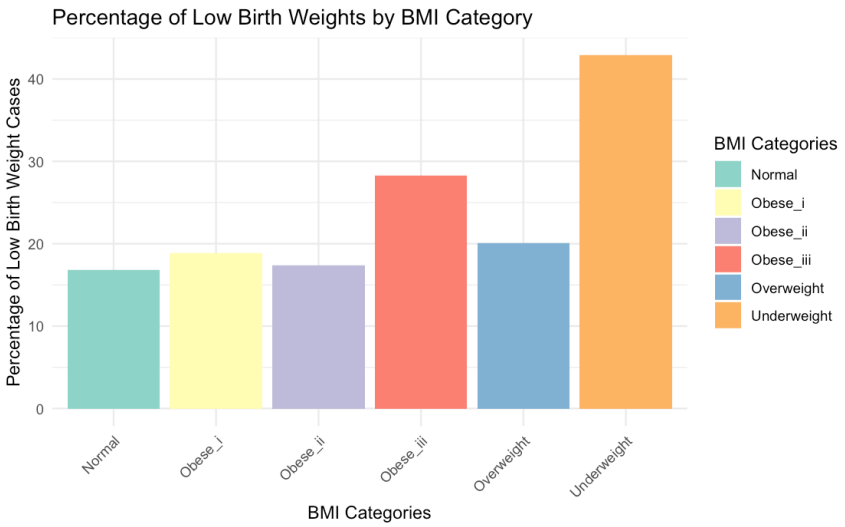



*Figure 14 Graph Showing Percentage of Low Birth Weight Cases by BMI Category*

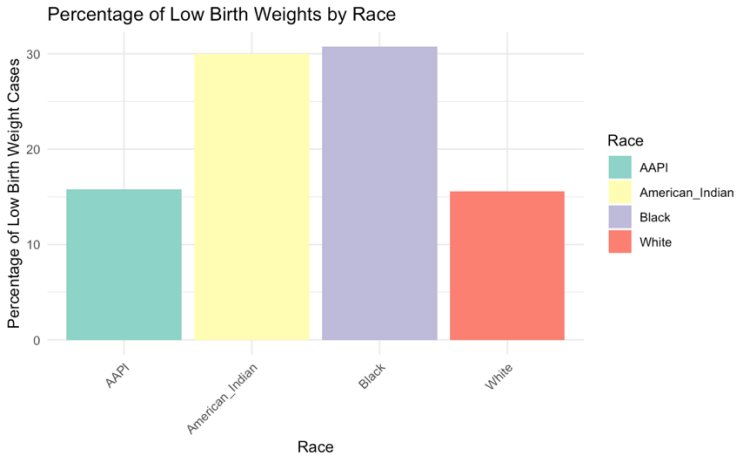



*Figure 15 Graph Showing Percentage of Low Birth Weight Cases by Race*

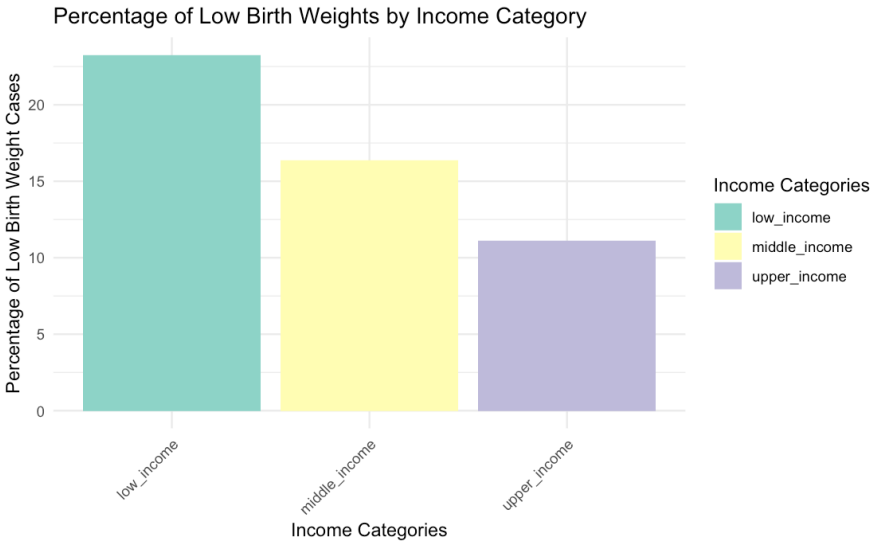



*Figure 16 Graph Showing Percentage of Low Birth Weight Cases by Income*

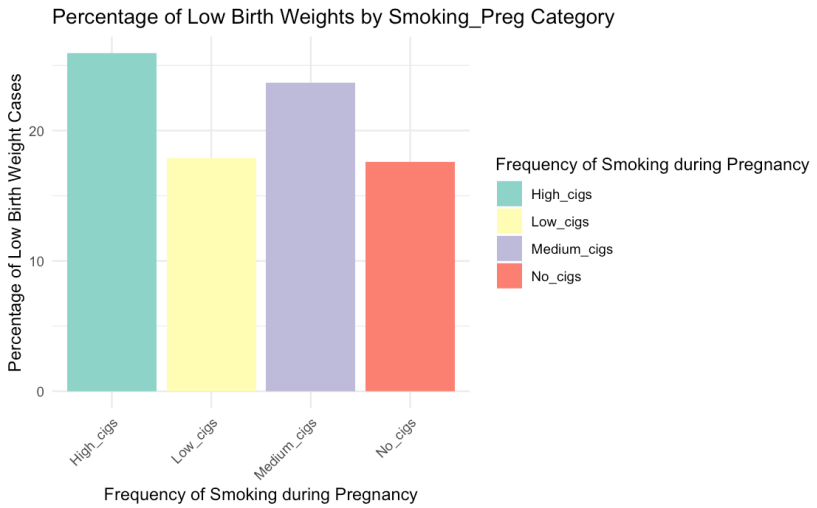



*Figure 17 Graph Showing Percentage of Low Birth Weight Cases by Frequency of Smoking During Pregnancy*

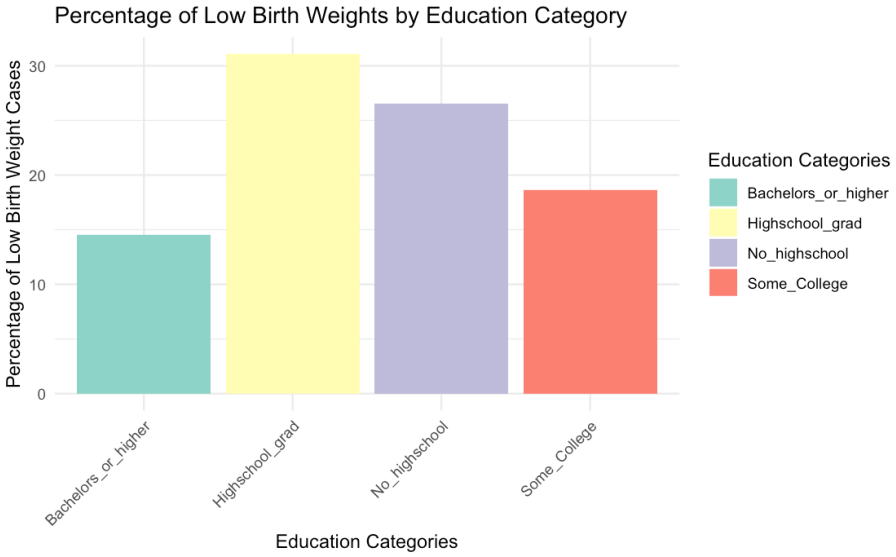



*Figure 18 Graph Showing Percentage of Low Birth Weight Cases by Education Level*

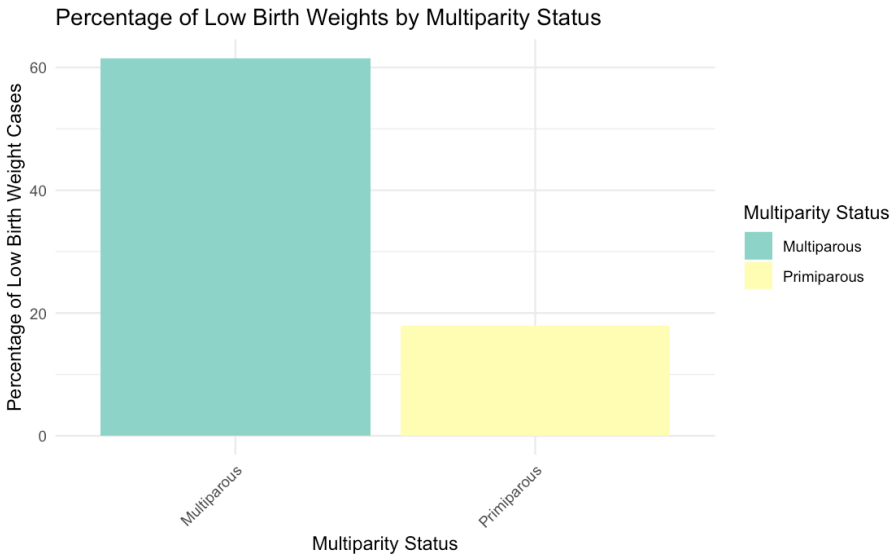



*Figure 19 Graph Showing Percentage of Low Birth Weight Cases by Multiparity Status*

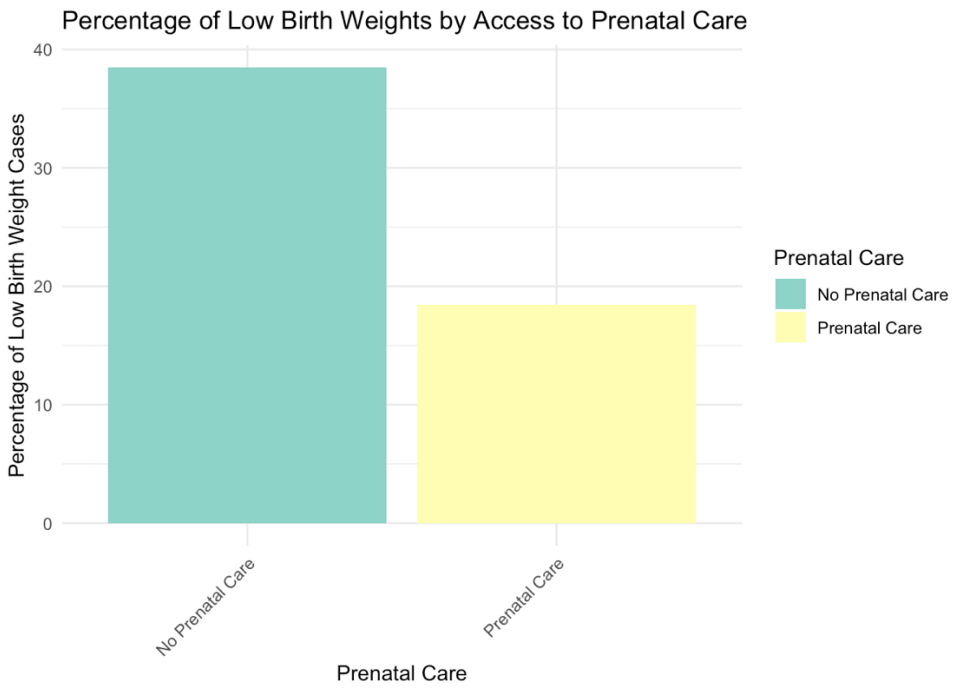



*Figure 20 Graph Showing Percentage of Low Birth Weight Cases by Access to Prenatal Care*

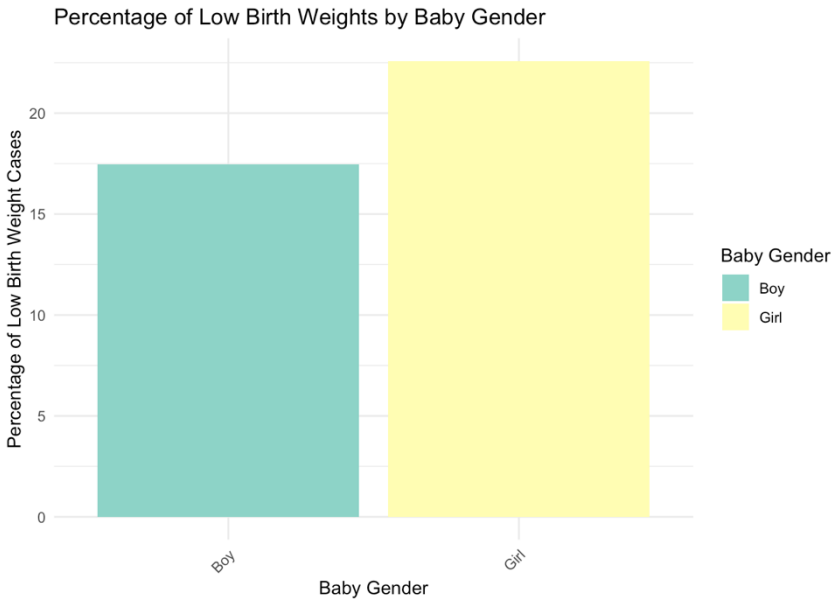



*Figure 21 Graph Showing Percentage of Low Birth Weight Cases by Baby's Gender*

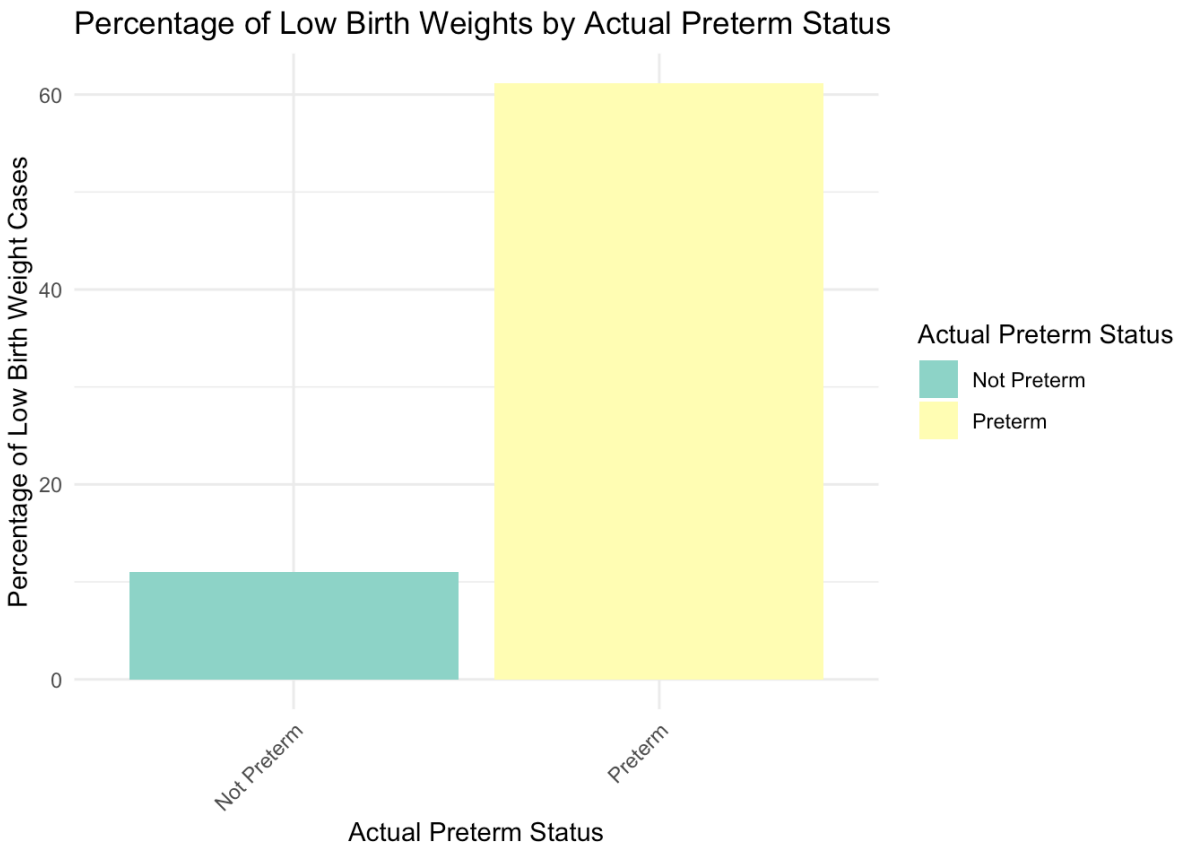



*Figure 22 Graph Showing Preterm Births by Actual Preterm Status*

# Appendix C: Regressions and Results

```
stargazer(logitModel,logitModel3,logitModel4, type="text", title= "logistic regression models for low birth we
t prediction")
```

```
##
## logistic regression models for low birth weight prediction
## ===============================================================
##                                  Dependent variable:
##                              ---------------------------
##                                     low_birthweight
##                                (1)       (2)       (3)
## ---------------------------------------------------------------
## RaceAAPI                      0.015     0.372     1.239
##                              (0.642)   (0.667)   (0.756)
##
## RaceAmerican_Indian           0.842     0.328    -0.322
##                              (0.701)   (0.849)   (1.027)
##
## RaceBlack                    0.876***  0.763*** 0.985***
##                              (0.201)   (0.236)   (0.302)
##
## Incomemiddle_income                    -0.338   -0.592**
##                                       (0.237)   (0.295)
##
## Incomeupper_income                     -1.112   -1.389
##                                       (1.058)   (1.200)
##
## EducationNo_highschool                  0.307     0.125
##                                       (0.427)   (0.568)
##
## EducationHighschool_grad               0.738***  0.670*
##                                       (0.287)   (0.366)
##
## EducationBachelors_or_higher           -0.312   -0.203
##                                       (0.291)   (0.349)
##
## Prenatal_care1                         -0.809   -0.339
##                                       (0.657)   (0.890)
##
## Multiparity_status1                              1.967**
##                                                 (0.811)
##
## Actual_Preterm1                                 3.034***
##                                                 (0.308)
##
## BMIUnderweight                                   1.538**
##                                                 (0.779)
##
## BMIOverweight                                    0.216
##                                                 (0.364)
##
## BMIObese_i                                      -0.247
##                                                 (0.394)
##
## BMIObese_ii                                     -0.161
##                                                 (0.539)
##
## BMIObese_iii                                     0.046
##                                                 (0.463)
##
## Smoking_PregLow_cigs                             0.015
##                                                 (0.640)
##
## Smoking_PregMedium_cigs                          0.119
##                                                 (0.533)
##
## Smoking_PregHigh_cigs                            0.220
##                                                 (0.509)
##
## baby_gender                                     -0.734***
##                                                 (0.275)
##
## Constant                     -1.689***   -0.833   -1.814*
##                               (0.126)   (0.673)   (0.952)
##
## ---------------------------------------------------------------
## Observations                    702       596       581
## Log Likelihood               -341.045  -271.916 -196.217
## Akaike Inf. Crit.            690.090   563.831   434.433
## ===============================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

*Figure 23 Result of the three initial logistic regressions*

```
##
## logistic regression models involving interaction terms for low birth weight prediction
## ================================================================================
##                                                  Dependent variable:
##                                              ---------------------------
##                                                     low_birthweight
##                                                   (1)           (2)
## --------------------------------------------------------------------------------
## RaceAAPI                                          0.545         1.033
##                                                  (1.112)       (0.822)
##
## RaceAmerican_Indian                              -13.100       -13.370
##                                                 (827.307)     (639.593)
##
## RaceBlack                                        1.077***      0.885***
##                                                  (0.401)       (0.341)
##
## EducationNo_highschool                           0.108          0.075
##                                                  (0.698)       (0.575)
##
## EducationHighschool_grad                         0.679         0.660*
##                                                  (0.440)       (0.366)
##
## EducationBachelors_or_higher                     -0.202        -0.227
##                                                  (0.442)       (0.352)
##
## Incomemiddle_income                              -0.548*       -0.625**
##                                                  (0.301)       (0.298)
##
## Incomeupper_income                               -1.345        -1.381
##                                                  (1.208)       (1.190)
##
## BMIUnderweight                                   1.468*        1.533**
##                                                  (0.789)       (0.770)
##
## BMIOverweight                                    0.199          0.160
##                                                  (0.372)       (0.369)
##
## BMIObese_i                                       -0.278        -0.276
##                                                  (0.397)       (0.394)
##
## BMIObese_ii                                      -0.109        -0.216
##                                                  (0.547)       (0.548)
##
## BMIObese_iii                                     0.009          0.021
##                                                  (0.464)       (0.463)
##
## Multiparity_status1                              1.923**       1.970**
##                                                  (0.821)       (0.801)
##
## Prenatal_care1                                   -0.264        -0.047
##                                                  (0.938)       (0.998)
##
## Actual_Preterm1                                  2.995***      2.880***
##                                                  (0.311)       (0.353)
##
## Smoking_PregLow_cigs                             -0.013         0.036
##                                                  (0.639)       (0.633)
##
## Smoking_PregMedium_cigs                          0.107          0.108
##                                                  (0.532)       (0.526)
##
## Smoking_PregHigh_cigs                            0.208          0.201
##                                                  (0.509)       (0.504)
##
## baby_gender                                      -0.726***     -0.750***
##                                                  (0.277)       (0.276)
##
## RaceAAPI:EducationNo_highschool
##
##
## RaceAmerican_Indian:EducationNo_highschool
##
##
## RaceBlack:EducationNo_highschool                  0.0005
##                                                  (1.155)
##
## RaceAAPI:EducationHighschool_grad                -13.374
##                                                 (1,455.398)
##
## RaceAmerican_Indian:EducationHighschool_grad      12.223
##                                                 (827.308)
##
## RaceBlack:EducationHighschool_grad                0.097
##                                                  (0.838)
##
## RaceAAPI:EducationBachelors_or_higher             2.417
##                                                  (1.801)
##
## RaceAmerican_Indian:EducationBachelors_or_higher  28.364
##                                                 (1,674.103)
##
## RaceBlack:EducationBachelors_or_higher            -0.470
##                                                  (0.743)
##
## RaceAAPI:Actual_Preterm1                                        15.078
##                                                                (1,455.398)
##
## RaceAmerican_Indian:Actual_Preterm1                             13.711
##                                                                (639.594)
##
## RaceBlack:Actual_Preterm1                                       0.445
##                                                                (0.736)
##
## Constant                                         -1.878*       -1.984*
##                                                  (0.994)       (1.033)
##
## --------------------------------------------------------------------------------
## Observations                                       581           581
## Log Likelihood                                   -193.638      -195.206
## Akaike Inf. Crit.                                443.276       438.412
## ================================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

*Figure 24 Result of logistic regression with interaction terms*

# Appendix D: Statistical Tests

VIF values

| | GVIF | Df | GVIF^(1/(2*Df)) |
|---|---|---|---|
| Race | 1.270843 | 3 | 1.040755 |
| Income | 1.292667 | 2 | 1.066281 |
| Education | 1.375717 | 3 | 1.054601 |
| Prenatal_care | 1.061156 | 1 | 1.030124 |
| Multiparity_status | 1.059315 | 1 | 1.029230 |
| Actual_Preterm | 1.193486 | 1 | 1.092468 |
| BMI | 1.303076 | 5 | 1.026826 |
| Smoking_Preg | 1.243818 | 3 | 1.037034 |
| baby_gender | 1.087454 | 1 | 1.042811 |

*Figure 25 Variance Inflation Factor (VIF) Values for Predictor Variables from Third Regression Model*

```
Variable_Importance= varImp(logitModel4)
library(knitr)

# Print the result_df using kable for a nicely formatted table
kable(Variable_Importance, format = "html", caption = "Variable_Importance") %>%
  kable_styling(bootstrap_options = "striped")
```

Variable_Importance

| | Overall |
|---|---|
| RaceAAPI | 1.6384537 |
| RaceAmerican_Indian | 0.3134998 |
| RaceBlack | 3.2565774 |
| Incomemiddle_income | 2.0022479 |
| Incomeupper_income | 1.1579012 |
| EducationNo_highschool | 0.2193987 |
| EducationHighschool_grad | 1.8321631 |
| EducationBachelors_or_higher | 0.5821006 |
| Prenatal_care1 | 0.3806982 |
| Multiparity_status1 | 2.4253800 |
| Actual_Preterm1 | 9.8585522 |
| BMIUnderweight | 1.9743063 |
| BMIOverweight | 0.5937457 |
| BMIObese_i | 0.6284711 |
| BMIObese_ii | 0.2985013 |
| BMIObese_iii | 0.0989243 |
| Smoking_PregLow_cigs | 0.0234124 |
| Smoking_PregMedium_cigs | 0.2239337 |
| Smoking_PregHigh_cigs | 0.4323943 |
| baby_gender | 2.6726630 |

*Figure 26 Variable Importance for Predictor Variables*

```r
chi_square_result <- anova(logitModel4, test = "Chisq")

# Convert the anova result to a data frame
result_df <- as.data.frame(chi_square_result)

# Load the knitr package
library(knitr)

# Print the result_df using kable for a nicely formatted table
kable(result_df, format = "html", caption = "Chi-Square Test Results") %>%
  kable_styling(bootstrap_options = "striped")
```

Chi-Square Test Results

| | Df | Deviance | Resid. Df | Resid. Dev | Pr(>Chi) |
|---|---|---|---|---|---|
| NULL | NA | NA | 580 | 557.9958 | NA |
| Race | 3 | 9.3042687 | 577 | 548.6915 | 0.0255074 |
| Income | 2 | 7.1232525 | 575 | 541.5683 | 0.0283926 |
| Education | 3 | 8.8940600 | 572 | 532.6742 | 0.0307331 |
| Prenatal_care | 1 | 1.8827387 | 571 | 530.7915 | 0.1700232 |
| Multiparity_status | 1 | 12.2679614 | 570 | 518.5235 | 0.0004608 |
| Actual_Preterm | 1 | 113.0409473 | 569 | 405.4826 | 0.0000000 |
| BMI | 5 | 5.2208647 | 564 | 400.2617 | 0.3895246 |
| Smoking_Preg | 3 | 0.4484882 | 561 | 399.8132 | 0.9300535 |
| baby_gender | 1 | 7.3798363 | 560 | 392.4334 | 0.0065959 |

*Figure 27 Chi-Square Test Results for Predictor Variables*

```r
# Remove rows with missing values
final_columns_complete <- final_columns[complete.cases(final_columns), ]

# Fit the null model
null_model <- glm(low_birthweight ~ 1, data = final_columns_complete, family = binomial())

# Fit the full model
full_model <- glm(low_birthweight ~ Race + Income + Education + Prenatal_care + Multiparity_status + Actual_Prete
rm + BMI + Smoking_Preg + baby_gender, data = final_columns_complete, family = binomial())

# Perform the likelihood ratio test
lr_test <- anova(null_model, full_model, test = "Chisq")

# Print the test results
library(stargazer)
stargazer(lr_test, type="text", title= "Results of Likelihood Ratio test")
```

```
##
## Results of Likelihood Ratio test
## ================================================
## Statistic  N  Mean    St. Dev.   Min      Max
## ------------------------------------------------
## Resid. Df  2 570.000  14.142    560      580
## Resid. Dev 2 475.215 117.070  392.433 557.996
## Df         1 20.000              20       20
## Deviance   1 165.562           165.562 165.562
## Pr(> Chi)  1  0.000               0        0
## ------------------------------------------------
```

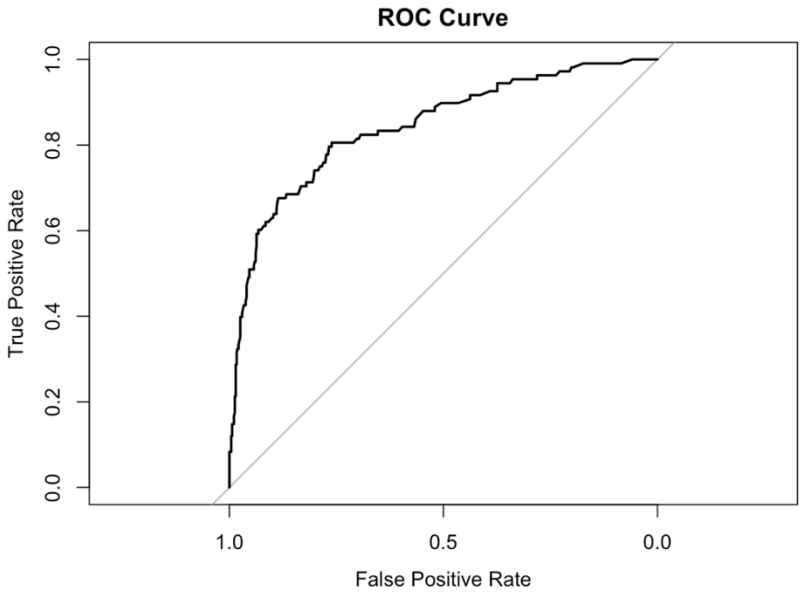
*Figure 28 Results of Likelihood-Ratio test*

```
logitModel4 <- glm(low_birthweight ~ Race + Income + Education + Prenatal_care + Multiparity_status + Actual_Pret
erm + BMI + Smoking_Preg + baby_gender , data = final_columns, family = binomial())

# Use the same dataset to generate predicted probabilities
predicted_probs <- predict(logitModel4, type = "response", newdata = final_columns)

# Create a ROC curve and calculate the AUROC
roc_obj <- roc(final_columns$low_birthweight, predicted_probs)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
# Plot the ROC curve
plot(roc_obj, main = "ROC Curve", xlab = "False Positive Rate", ylab = "True Positive Rate")
```




```
# Calculate the AUROC
auroc <- auc(roc_obj)

# Print the AUROC value
cat("AUROC:", auroc, "\n")
```

```
## AUROC: 0.8433756
```

*Figure 29 ROC curve and AUROC result*