*Ariel Gilgeours*

*Birth weight Analysis*

Table of Contents

Introduction\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2

Infant Birth Weight and Smoking\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2

Smoking, Education and Prenatal Visits\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_5

Infant Birth Weight, Mother’s Weight Gain and Smoking\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_13

Appendix\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20

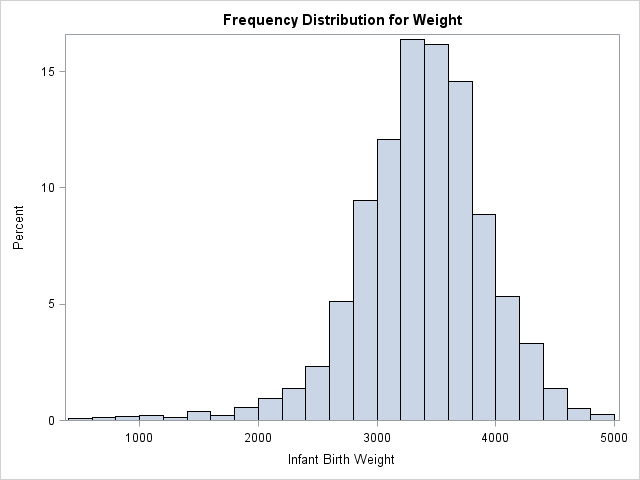
The SAS birth weight data set of 50,000 observations of birth weight data from 1997 from the National Center of Health Statistics. The data records live, singleton births to black or white mothers, age 18 to 45 from the United States. The variables included in this data set are weight, black, married, boy, visit, ed, smoke, cigsper, momage and momwtgain. For the analyses presented in this paper a sample size of 5000 was randomly sampled from the population of observations.

The smoking status of the mother, whether they smoke or do not smoke is the overarching theme of this analysis. The basis of this research analysis is to determine whether smoking has statistically significant effect on the mother and therefore effecting the birth weight of the infant. The reason behind looking at smoking as the main factor in these analyses is given the fact that smoking has determinantal effects to individuals through primary or secondary contact and is one of the leading causes of preventable deaths in the United States. Smoking has been found to harm nearly all organs in the body in some shape or form.

Infant Birth Weight and Smoking

In this analysis we looked at the birth weight of babies born to mothers who smoked compared to the birth weight of babies born to mothers that did not smoke. Infant birth weight which is the weight of a baby taken at birth is given in grams. The variable smoke is indicated as 0 = No and 1 = Yes. Infants born with low birth weight may be healthy, but they can also have very serious health conditions. Low birth weight can affect organ function, and babies they might have trouble eating or staying warm.

The research question that we are going to look at in this analysis; Is there a significant difference between the birth weight of babies born to mothers who smoke compared to mothers that do not smoke?



When looking at the data with the use of a histogram of the infant birth weights. We can see that the data creates a bell-shaped curve and looks similar to a normal distribution. In this case it is slightly left skewed. The curve peaks around 3500 grams and has weights at high as 5000 grams and some lower than 1000 grams.

In order to perform this analysis, we will be using nonparametric statistics given the fact that it is a simple and clear learning tool, that allows us to learn about our sample before performing other analyses. It also empirical and allows us to run a model when the population is unknown. The use of npar1way procedure using wilcoxon score (rank sum) for the variable weight was performed and classified by the variable smoking. We look at the approximation of the sum of ranks, which is the sum of ranks subtracted from its expected value and divide it by the standard deviation.

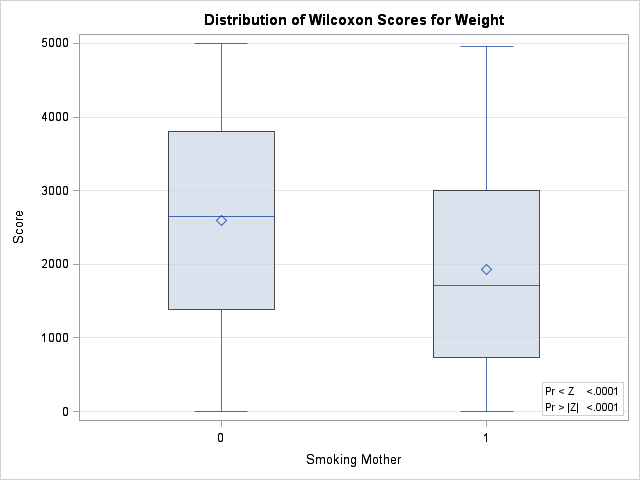
|  |
| --- |
| Frequency Distribution for Weight |

The NPAR1WAY Procedure

| **Wilcoxon Scores (Rank Sums) for Variable Weight Classified by Variable MomSmoke** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | **N** | **Sum of Scores** | **Expected Under H0** | **Std Dev Under H0** | **Mean Score** |
| **0** | 4316 | 11180367.5 | 10792158.0 | 35073.5332 | 2590.44659 |
| **1** | 684 | 1322132.5 | 1710342.0 | 35073.5332 | 1932.94225 |
| **Average scores were used for ties.** | | | | | |

| **Wilcoxon Two-Sample Test** | |
| --- | --- |
| **Statistic** | 1322132.5000 |
|  |  |
| **Normal Approximation** |  |
| **Z** | -11.0684 |
| **One-Sided Pr < Z** | <.0001 |
| **Two-Sided Pr > |Z|** | <.0001 |
|  |  |
| **t Approximation** |  |
| **One-Sided Pr < Z** | <.0001 |
| **Two-Sided Pr > |Z|** | <.0001 |
| **Z includes a continuity correction of 0.5.** | |

| **Kruskal-Wallis Test** | |
| --- | --- |
| **Chi-Square** | 122.5105 |
| **DF** | 1 |
| **Pr > Chi-Square** | <.0001 |



From this analysis we are able to determine that the number of individuals from the sample that were non-smokers was 4316 and the number of individuals from the sample that were smokers was 664. We obtain a sum of scores value of 1322132.5 which is the sum of the ranks and the z score approximation of -11.0684. Given the fact for a normal distribution at the alpha level of .05 the z score is -1.645 . Since the Z score obtained from this model is less than -1.645, the difference between birth weight in babies of mothers who do smoke is statically significant than babies of mothers who do not smoke. This is also evident in the box-plot above, which shows that the means are different but the standard deviations are similar given the width of the box-plot.

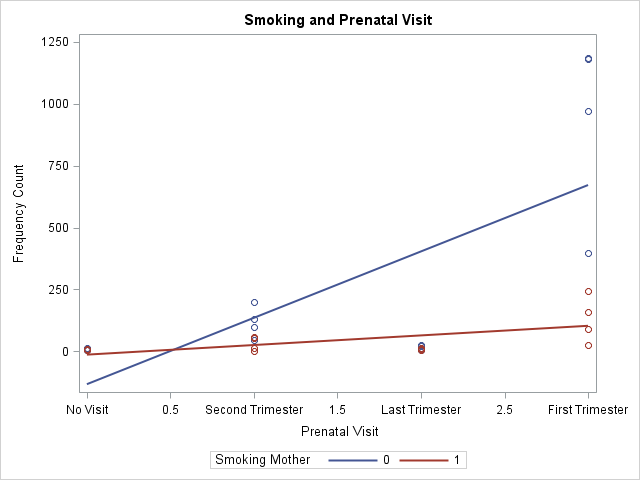
Smoking, Education and Prenatal Visits

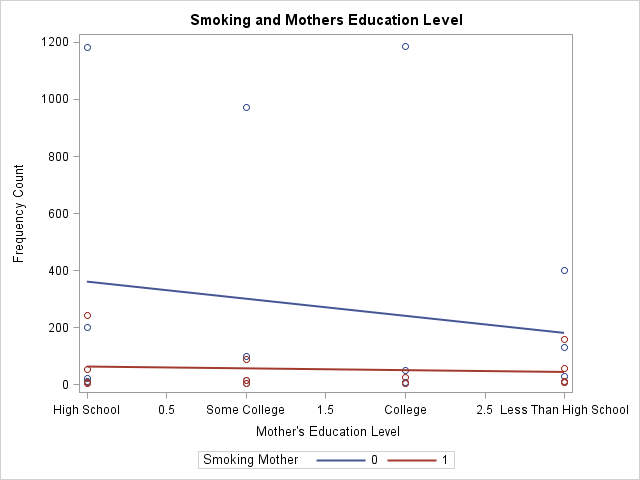
Education is essential for the preventative measures of any health condition. Those who are less educated are more likely to not acquire knowledge about smoking and its health effects compared to those with a higher education. Which might lead to a higher prevalence of smoking in those groups. Not only is education essential for the knowledge of health but the ability to have access to health care that is needed for individuals, especially those at risk such as pregnant women. For this study we are looking for patterns of association between smoking, education level and prenatal visits.

The research question for this analysis; Is smoking among mothers from different education level statistical effect their prenatal visits?

In this study we looked at 5000 births. The three variables examined in this analysis were smoking, prenatal visits and mother’s education level. Within the data each variable was scaled in order to perform this analysis. Smoke status was scaled as 0 = Non-Smoking, 1 = Smoking. Prenatal visits were scaled as 0 = no visit, 1 = visit in second trimester, 2 = visit in last trimester, 3 = visit in first trimester. Mother’s education level was scaled as 0 = high school, 1 = some college, 2 = college, 3 = less than high school.

For the first graph with count vs prenatal visit showed that non-smoking mothers had a much higher frequency of having their first prenatal visit during the first trimester compared to mothers that did smoke. You can also see there are more mothers that smoked who had no visit at all or had a prenatal visit in the last trimester of their pregnancy. For the second graph of count vs. mother’s education level. In this case, the mother’s education for smoking mothers was pretty evenly distributed between the four classes. While more mothers that did not smoke had a higher frequency of high school, some college and college education compared to the mothers that smoked.





Given the fact that we want to determine the relationship between each of the factors and which factors fit the data best. A loglinear model would be best suited for this analysis of smoking, education level and prenatal visits. With loglinear analysis we want to look at the variables and their relationship to one another. By looking at this relationship we test whether the expected cell counts, fits to the observed cell count. These are computed by the likelihood ratio and when the observed values are significant, we are able to see a large chi square. We start with a saturated model in order to determine whether all of the terms are significant with regard to the outcome.

| **Maximum Likelihood Analysis of Variance** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **MomSmoke** | 1 | 28.17 | <.0001 |
| **Visit** | 3 | 628.42 | <.0001 |
| **MomSmoke\*Visit** | 3 | 11.07 | 0.0114 |
| **MomEdLevel** | 3 | 17.42 | 0.0006 |
| **MomSmoke\*MomEdLevel** | 3 | 12.73 | 0.0053 |
| **Visit\*MomEdLevel** | 9 | 79.70 | <.0001 |
| **MomSmoke\*Visit\*MomEdLeve** | 6\* | 6.01 | 0.4224 |
| **Likelihood Ratio** | 0 | . | . |

|  |  |
| --- | --- |
| Note: | Effects marked with '\*' contain one or more  redundant or restricted parameters. |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** |
| **MomSmoke** | **0** | 0.6945 | 0.1308 | 28.17 | <.0001 |
| **Visit** | **No Visit** | -1.5111 | 0.3005 | 25.28 | <.0001 |
|  | **Second Trimester** | 0.4707 | 0.1228 | 14.70 | 0.0001 |
|  | **Last Trimester** | -1.2873 | 0.1354 | 90.39 | <.0001 |
| **MomSmoke\*Visit** | **0 No Visit** | -0.5434 | 0.2843 | 3.65 | 0.0559 |
|  | **0 Second Trimester** | 0.1518 | 0.1187 | 1.64 | 0.2009 |
|  | **0 Last Trimester** | -0.00690 | 0.1121 | 0.00 | 0.9510 |
| **MomEdLevel** | **High School** | 0.5323 | 0.1465 | 13.20 | 0.0003 |
|  | **Some College** | -0.0215 | 0.2260 | 0.01 | 0.9241 |
|  | **College** | -0.9433 | 0.2521 | 14.00 | 0.0002 |
| **MomSmoke\*MomEdLevel** | **0 High School** | -0.0649 | 0.1445 | 0.20 | 0.6533 |
|  | **0 Some College** | -0.0983 | 0.2238 | 0.19 | 0.6606 |
|  | **0 College** | 0.5400 | 0.2275 | 5.63 | 0.0176 |
| **Visit\*MomEdLevel** | **No Visit High School** | -0.4209 | 0.3157 | 1.78 | 0.1824 |
|  | **No Visit Some College** | 0.1456 | 0.5447 | 0.07 | 0.7892 |
|  | **No Visit College** | 0.3347 | 0.3804 | 0.77 | 0.3788 |
|  | **Second Trimester High School** | 0.3028 | 0.1425 | 4.51 | 0.0336 |
|  | **Second Trimester Some College** | -0.1379 | 0.2344 | 0.35 | 0.5562 |
|  | **Second Trimester College** | -0.3778 | 0.1472 | 6.59 | 0.0102 |
|  | **Last Trimester High School** | 0.0272 | 0.1608 | 0.03 | 0.8656 |
|  | **Last Trimester Some College** | -0.0508 | 0.2032 | 0.06 | 0.8027 |
|  | **Last Trimester College** | -0.3863 | 0.2453 | 2.48 | 0.1153 |
| **MomSmoke\*Visit\*MomEdLeve** | **0 No Visit High School** | 0.4631 | 0.2724 | 2.89 | 0.0891 |
|  | **0 No Visit Some College** | -0.3916 | 0.4406 | 0.79 | 0.3742 |
|  | **0 No Visit College** | -0.2972 | 0.2407 | 1.53 | 0.2168 |
|  | **0 Second Trimester High School** | -0.1293 | 0.1381 | 0.88 | 0.3493 |
|  | **0 Second Trimester Some College** | 0.1904 | 0.2329 | 0.67 | 0.4136 |
|  | **0 Second Trimester College** | . | . | . | . |
|  | **0 Last Trimester High School** | -0.0947 | 0.1121 | 0.71 | 0.3984 |
|  | **0 Last Trimester Some College** | . | . | . | . |
|  | **0 Last Trimester College** | . | . | . | . |

Starting with the saturated model which has all the main effect, three-way interactive and two-way interactive terms, we can see how this model fits to the data. Looking at the maximum likelihood analysis of variance, from the output you can see that the variables and two-way interactive term are all statistically significant at the .05 level. While the three-way interactive term MomSmoke\*Visit\*Education level has a chi-square value of 6.01 and a probability of .4224. Since the probability is above .05, the interactive term is not statistically significant. In regard to the likelihood ratio there are no degrees of freedom, so no chi square is computed, since all of the cells were accounted for. Give the fact that the three-way interactive term is not statistically significant the model must be reduced further. The next step is to run a reduced model without the interactive term which is performed below.

|  |
| --- |
| reduced model w/o 3 way interaction |

| **Maximum Likelihood Analysis of Variance** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **MomSmoke** | 1 | 216.73 | <.0001 |
| **Visit** | 3 | 1827.92 | <.0001 |
| **MomSmoke\*Visit** | 3 | 11.03 | 0.0115 |
| **MomEdLevel** | 3 | 119.10 | <.0001 |
| **MomSmoke\*MomEdLevel** | 3 | 243.47 | <.0001 |
| **Visit\*MomEdLevel** | 9 | 205.83 | <.0001 |
| **Likelihood Ratio** | 6 | 5.53 | 0.4779 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** |
| **MomSmoke** | **0** | 0.9040 | 0.0614 | 216.73 | <.0001 |
| **Visit** | **No Visit** | -1.8570 | 0.1638 | 128.52 | <.0001 |
|  | **Second Trimester** | 0.5706 | 0.0786 | 52.65 | <.0001 |
|  | **Last Trimester** | -1.2536 | 0.1247 | 101.02 | <.0001 |
| **MomSmoke\*Visit** | **0 No Visit** | -0.2002 | 0.1382 | 2.10 | 0.1474 |
|  | **0 Second Trimester** | 0.0508 | 0.0666 | 0.58 | 0.4455 |
|  | **0 Last Trimester** | -0.0320 | 0.1033 | 0.10 | 0.7566 |
| **MomEdLevel** | **High School** | 0.7438 | 0.0853 | 76.00 | <.0001 |
|  | **Some College** | -0.1741 | 0.1162 | 2.25 | 0.1340 |
|  | **College** | -1.2166 | 0.1419 | 73.52 | <.0001 |
| **MomSmoke\*MomEdLevel** | **0 High School** | -0.2968 | 0.0372 | 63.80 | <.0001 |
|  | **0 Some College** | 0.0850 | 0.0465 | 3.33 | 0.0680 |
|  | **0 College** | 0.8139 | 0.0741 | 120.50 | <.0001 |
| **Visit\*MomEdLevel** | **No Visit High School** | -0.0209 | 0.1979 | 0.01 | 0.9160 |
|  | **No Visit Some College** | -0.2811 | 0.2836 | 0.98 | 0.3216 |
|  | **No Visit College** | 0.0325 | 0.2889 | 0.01 | 0.9104 |
|  | **Second Trimester High School** | 0.1905 | 0.0928 | 4.21 | 0.0401 |
|  | **Second Trimester Some College** | 0.0405 | 0.1237 | 0.11 | 0.7434 |
|  | **Second Trimester College** | -0.3524 | 0.1455 | 5.86 | 0.0155 |
|  | **Last Trimester High School** | -0.0393 | 0.1507 | 0.07 | 0.7943 |
|  | **Last Trimester Some College** | 0.0335 | 0.1866 | 0.03 | 0.8574 |
|  | **Last Trimester College** | -0.4026 | 0.2446 | 2.71 | 0.0998 |

Now we look at the reduced model with the main effects and two-way interactions without the three-way interaction. We notice here that the chi squares for the main effects and two-way effects are all statistically significant at the .05 level. In this model we now have a likelihood ratio. For the likelihood ratio we get a chi square of 5.53 and a probability of .4779 that is associated with the chi square from the three-way interactive term from the saturated model.

The best model for this data would be the reduced model without the three-way interactive term since it fits more appropriately than any other model. This model does the best job in describing what factors affect the outcome in regard to responses in each of the cells. Which means that main effects and two-way interactions are statistically significant.

Mother’s Weight Gain Related to Infant Weight

An infant’s birth weight it determined at birth and is often influenced by a multitude of factors. As mothers are the source of nutrients for a growing fetus, a mother’s weight gain during pregnancy may be a contributing factor to the weight gain of the fetus. While a mother’s smoking status, whether they smoke or not can be harmful to their own health and weight and therefore may be harmful to the infants. Since nicotine, which is found in cigarettes, has been found to increase metabolism that may be the contributing factor that could affect weight gain during pregnancy.

Mother’s Weight Gain was centered at its median in the original data and formatted by adding 30 in order to obtain the “actual mom weight gain.” The variable smoke was set at 0 = Non-Smoking Mother and 1 = Smoking Mother. The infant birth weight is weighted in grams.

The research question for this study is; Does a mother’s weight gain influence birthweight and smoking status?

Before performing an ANCOA, an exploratory analysis on the data needs to be performed in order to see if this data is a good fit for the model.

|  |
| --- |
| Mothers Weight Gain as a Confounding Factor Related to Infant Weight |

The ANOVA Procedure

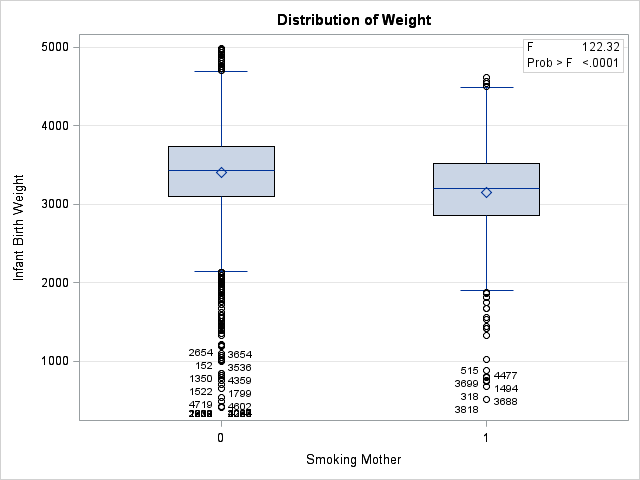
Dependent Variable: Weight Infant Birth Weight

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 1 | 37443182 | 37443182 | 122.32 | <.0001 |
| **Error** | 4998 | 1529926626 | 306108 |  |  |
| **Corrected Total** | 4999 | 1567369808 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Weight Mean** |
| --- | --- | --- | --- |
| 0.023889 | 16.41218 | 553.2701 | 3371.094 |

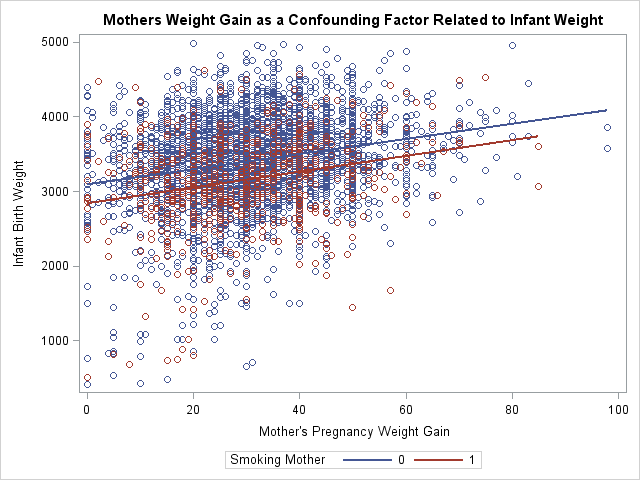
| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | 1 | 37443182.41 | 37443182.41 | 122.32 | <.0001 |

| **Level of MomSmoke** | **N** | **Weight** | |
| --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **0** | **4316** | 3405.54356 | 546.805383 |
| **1** | **684** | 3153.71637 | 592.483837 |



We explore the model of infant birth weight by whether a mother smoke or not. From the ANOVA procedure you can see that the probability of this model is less than .0001% which means that a mother’s smoking status has a significant effect on the outcome of this model. From the model above you can see that mothers who do not smoke delivers a baby on average 252 grams heavier than a mother that does smoke. From the boxplots you can see that there is difference in mean and quartile values. Therefore, there is clearly an effect of a mother’s smoking status on the outcome measure, infant birth weight.

Next, we want to take a look on mother’s pregnancy weight gain as a confounding effect on infant birth weight. In order to run an ANCOVA model, the data must be best suited for that type of analysis. In order for an ANCOVA to work , the covariate should cause the variables of interest, smoke status to create parallel or roughly parallel lines. In other words, the differences between the smoking and non-smoking mothers is consistent no matter what the mother’s pregnancy weight gain would be. Which is shown below for this data.



Next, a simple regression analysis of mother’s pregnancy weight gain on infant birth weight is performed. Given the model you can see that mother’s weight gain is statistically significant on the outcome variable infant weight. When looking at the R value is roughly 5% of the outcome measure infant weight gain. With a beta value of 10.46126 which shows that as mother’s weight gain increases, infant’s birth weight increases as well.

|  |
| --- |
| Regression Analysis Infant Weight by Mom Weight Gain |

The REG Procedure

Model: MODEL1

Dependent Variable: Weight Infant Birth Weight

|  |  |
| --- | --- |
| **Number of Observations Read** | 5000 |
| **Number of Observations Used** | 5000 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 87813137 | 87813137 | 296.64 | <.0001 |
| **Error** | 4998 | 1479556671 | 296030 |  |  |
| **Corrected Total** | 4999 | 1567369808 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 544.08616 | **R-Square** | 0.0560 |
| **Dependent Mean** | 3371.09360 | **Adj R-Sq** | 0.0558 |
| **Coeff Var** | 16.13975 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | 3049.58144 | 20.19109 | 151.04 | <.0001 |
| **MomWtGain** | Mother's Pregnancy Weight Gain | **1** | 10.46126 | 0.60740 | 17.22 | <.0001 |

Now we look at the ANCOVA since we see there is parallelism within the data above. Looking at the saturated model with main effects and interactive effect. While the model overall looks to be statistically significant with a probability of less than .0001. Given the fact that this is not a balanced data set, this model fits best with the Type III model. Mother’s smoke status and mother’s weight gain turns out to be statistically significant. Noticing that the R squared value has increased from what it previously was in the regression model. However, you can see that the interactive effect in this model MomWtGain\*MomSmoke is not statistically significant with a probability of .7945 and it should be eliminated from the model. Therefore, the analysis of covariance will only depend on mother’s weight gain and mother smoke status. However, we do need to establish that the interactive term is not significant to the model.

|  |
| --- |
| Regression Analysis Infant Weight by Mom Weight Gain |

The GLM Procedure

Dependent Variable: Weight Infant Birth Weight

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3 | 121837701 | 40612567 | 140.36 | <.0001 |
| **Error** | 4996 | 1445532108 | 289338 |  |  |
| **Corrected Total** | 4999 | 1567369808 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Weight Mean** |
| --- | --- | --- | --- |
| 0.077734 | 15.95629 | 537.9014 | 3371.094 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | 1 | 37443182.41 | 37443182.41 | 129.41 | <.0001 |
| **MomWtGain** | 1 | 84374883.04 | 84374883.04 | 291.61 | <.0001 |
| **MomWtGain\*MomSmoke** | 1 | 19635.15 | 19635.15 | 0.07 | 0.7945 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | 1 | 7899440.06 | 7899440.06 | 27.30 | <.0001 |
| **MomWtGain** | 1 | 49850986.20 | 49850986.20 | 172.29 | <.0001 |
| **MomWtGain\*MomSmoke** | 1 | 19635.15 | 19635.15 | 0.07 | 0.7945 |

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 2870.197494 | B | 43.64116419 | 65.77 | <.0001 |
| **MomSmoke Non-Smoking Mother** | 251.218902 | B | 48.07915383 | 5.23 | <.0001 |
| **MomSmoke Smoking Mother** | 0.000000 | B | . | . | . |
| **MomWtGain** | 10.599995 | B | 1.43906616 | 7.37 | <.0001 |
| **MomWtGain\*MomSmoke Non-Smoking Mother** | -0.412554 | B | 1.58367737 | -0.26 | 0.7945 |
| **MomWtGain\*MomSmoke Smoking Mother** | 0.000000 | B | . | . | . |
|  |  |  |  |  |  |

When looking at the main effects only model, when looking at the overall model you can see that it is statistically significant given the fact that its probability is less than .0001 at the .05 level. When looking at the sum of squares of the main effect model compared to the sum of squares for the saturated model it can be seen that there is barely a difference indication that the interactive term added little significance to the model. When looking at the main effects Type III, given the fact that this is an unbalanced data set. We can see that mother’s smoke status is statistically significant to an infant’s weight gain. And mother’s weight gain is also statistically significant to an infant’s weight gain.

|  |
| --- |
| Regression Analysis Infant Weight by Mom Weight Gain |

The GLM Procedure

Dependent Variable: Weight Infant Birth Weight

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 121818065 | 60909033 | 210.55 | <.0001 |
| **Error** | 4997 | 1445551743 | 289284 |  |  |
| **Corrected Total** | 4999 | 1567369808 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Weight Mean** |
| --- | --- | --- | --- |
| 0.077721 | 15.95480 | 537.8512 | 3371.094 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | 1 | 37443182.41 | 37443182.41 | 129.43 | <.0001 |
| **MomWtGain** | 1 | 84374883.04 | 84374883.04 | 291.67 | <.0001 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MomSmoke** | 1 | 34004928.36 | 34004928.36 | 117.55 | <.0001 |
| **MomWtGain** | 1 | 84374883.04 | 84374883.04 | 291.67 | <.0001 |

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 2879.308903 | B | 26.09785508 | 110.33 | <.0001 |
| **MomSmoke Non-Smoking Mother** | 240.102104 | B | 22.14558616 | 10.84 | <.0001 |
| **MomSmoke Smoking Mother** | 0.000000 | B | . | . | . |
| **MomWtGain** | 10.259345 |  | 0.60072426 | 17.08 | <.0001 |

From the results of each of the research questions presented in this analysis. We can say that smoking has an effect on the birth weight of babies. It was found that babies born to mother that smoke had a smaller birth weight compared to mothers that didn’t smoke. It was also found that mothers that smoked gained less weight during pregnancy compared to mothers that did not smoke. It showed that not only smoking, but mothers weight gain had a statistically significant effect on the infant’s weight gain. It was also determined that mother’s education, and prenatal visits were statistically significant when compared to whether a mother was a smoker or non-smoker. While the exact detrimental effects of smoking during pregnancy might vary case by case, it is shown to affect the infant.

**Appendix**

The data used in order to run the analyses below was randomly selected from SAS 1997 Birth Weight Data. For more information can be found at the link below.

<https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.4&docsetId=statug&docsetTarget=statug_sashelp_sect008.htm&locale=en>

Infant Birth Weight and Smoking

**proc** **surveyselect** data = sashelp.BWeight method = SRS rep = **1**

sampsize = **5000** seed = **767** out = birthwtsamp;

id \_all\_; **run**;

**proc** **npar1way** data=birthwtsamp wilcoxon

plots=(wilcoxonboxplot);

class MomSmoke;

var Weight;

**run**;

**proc** **sgplot** data = birthwtsamp;

title 'Frequency Distribution for Weight';

histogram Weight;

**run**;

Smoking, Education and Prenatal Visits

**proc** **surveyselect** data = sashelp.BWeight method = SRS rep = **1**

sampsize = **5000** seed = **767** out = birthwtsamp;

id \_all\_; **run**;

**PROC** **FREQ** data=birthwtsamp;

tables MomSmoke\*Visit\*MomEdLevel/out = birthwtfreq nocol nopercent relrisk;

**RUN**;

**proc** **format**;

value vfmt **0** = 'No Visit' **1** = 'Second Trimester'

**2** = 'Last Trimester' **3** = 'First Trimester';

value efmt **0** = 'High School' **1** = 'Some College'

**2** = 'College' **3** = 'Less Than High School';

**run**;

**proc** **sgplot** data=birthwtfreq;

format Visit vfmt.;

title 'Smoking and Prenatal Visit';

reg x= Visit y=Count / Group = MomSmoke;

**run**;

**proc** **sgplot** data=birthwtfreq;

format MomEdLevel efmt.;

title 'Smoking and Mothers Education Level';

reg x= MomEdLevel y=Count / Group = MomSmoke;

**run**;

**proc** **surveyselect** data = sashelp.BWeight method = SRS rep = **1**

sampsize = **5000** seed = **767** out = birthwtsamp;

id \_all\_; **run**;

**PROC** **FREQ** data=birthwtsamp;

tables MomSmoke\*Visit\*MomEdLevel/out = birthwtfreq nocol nopercent relrisk;

**RUN**;

**proc** **format**;

value vfmt **0** = 'No Visit' **1** = 'Second Trimester'

**2** = 'Last Trimester' **3** = 'First Trimester';

value efmt **0** = 'High School' **1** = 'Some College'

**2** = 'College' **3** = 'Less Than High School';

**run**;

**proc** **catmod** data = birthwtfreq;

format Visit vfmt. MomEdLevel efmt.;

title 'proc catmod for 2X4X4 tables';

weight Count;

model MomSmoke\*Visit\*MomEdLevel =\_response\_ / noresponse noiter;

loglin MomSmoke|Visit|MomEdLevel;

**run**;

**proc** **catmod** data = birthwtfreq;

format Visit vfmt. MomEdLevel efmt.;

title 'reduced model w/o 3 way interaction';

weight Count;

model MomSmoke\*Visit\*MomEdLevel =\_response\_ / noresponse noiter;

loglin MomSmoke|Visit MomSmoke|MomEdLevel Visit\*MomEdLevel;

**run**;

**proc** **catmod** data = birthwtfreq;

format Visit vfmt. MomEdLevel efmt.;

title '2 way interaction w/o Visit|MomEdLevel';

weight Count;

model MomSmoke\*Visit\*MomEdLevel =\_response\_ / noresponse noiter;

loglin MomSmoke|Visit MomSmoke|MomEdLevel;

**run**;

**proc** **catmod** data = birthwtfreq;

format Visit vfmt. MomEdLevel efmt.;

title '2 way interaction w/o MomSmoke|MomEdLevel';

weight Count;

model MomSmoke\*Visit\*MomEdLevel =\_response\_ / noresponse noiter;

loglin MomSmoke|Visit Visit|MomEdLevel ;

**run**;

**proc** **catmod** data = birthwtfreq;

format Visit vfmt. MomEdLevel efmt.;

title '2 way interaction w/o MomSmoke|Visit';

weight Count;

model MomSmoke\*Visit\*MomEdLevel =\_response\_ / noresponse noiter;

loglin MomSmoke|MomEdLevel Visit|MomEdLevel ;

**run**;

Infant Birth Weight, Mother’s Weight Gain and Smoking

**proc** **surveyselect** data = sashelp.BWeight method = SRS rep = **1**

sampsize = **5000** seed = **767** out = birthwtsamp;

id \_all\_; **run**;

**data** birthwtsamp;

set birthwtsamp;

MomWtGain = MomWtGain + **30**;

**proc** **format**;

value MSfmt **0** = 'Non-Smoking Mother' **1** = 'Smoking Mother';

**run**;

**proc** **sgplot** data=birthwtsamp;

title 'Mothers Weight Gain as a Confounding Factor Related to Infant Weight';

reg x= MomWtGain y=Weight / Group = MomSmoke;

**run**;

**proc** **anova** data=birthwtsamp;

class MomSmoke;

model Weight = MomSmoke;

means MomSmoke;

**run**;

**proc** **reg** data=birthwtsamp;

title 'Regression Analysis Infant Weight by Mom Weight Gain';

model Weight = MomWtGain;

plot Weight\*MomWtGain;

**run**;

**proc** **sort**;

by MomSmoke;

**proc** **glm** data=birthwtsamp;

format MomSmoke MSfmt.;

class MomSmoke;

model Weight = MomSmoke MomWtGain MomSmoke\*MomWtGain/ solution;

**run**;

**proc** **glm** data=birthwtsamp;

format MomSmoke MSfmt.;

class MomSmoke;

model Weight = MomSmoke MomWtGain / solution;

lsmeans MomSmoke / stderr pdiff cov out=adjmeans;

**run**;

**proc** **print** data = adjmeans;

**run**;