Report

Contents

I. INTRODUCTION

We conducted analyses on a study to examine the relationship between parental smoking during pregnancy and pregnancy outcomes. There were a total of 680 subjects in the data set. Data were collected through interviews with the mothers early in their pregnancy while data on pregnancy outcomes were gathered at birth. Our analysis will focus on the relationship between maternal cigarette smoking and the birth weight of the baby. Information on the baby, mother, and father was collected and stored in three separate datasets. We combined the data set into one data set to use for our analyses. The primary outcome in this study was the birth weight of the baby and the z-score that was calculated for the birth weight of the baby based on the lower 10% cut-off value of -1.28 was used for analysis.

II. METHODS

All analyses were performed using the SAS statistical package and all statistical tests were performed using a significance level of 0.05. Variables of interest included birth weight of the baby in grams, maternal smoking status (number of cigarettes per day), paternal smoking status (number of cigarettes per day), gestational age (weeks), maternal pre-pregnancy weight (pounds, grams), maternal height (inches), maternal age (years), maternal BMI (wt_kg/ht_m^2), paternal age (years), paternal education(years), paternal height(inches). Two variables, mean birth weight in grams (MEAN_BWTG) and standard deviation (SD_BWTG) were created using PROC MEANS for the entire data set to get statistics for mean and standard deviation of birth weight in grams from our study. A Z-score variable was created based on the mean and standard deviation for birth weight for each child to compare Z-score of birth weight of grams in child to the cutoff value of lower 10% in order to assign children to the 0 or 1 categories for a new variable, SMALL_BWT (1=small, 0=not small). The maternal smoking category was created based on the number of cigarettes smoked in a day to create three categories a non-smoking mother, mother smokes a pack or less a day (1-20 cigarettes), a mother more than a pack a day (21 or more cigarettes). The paternal smoking category was created based on the number of cigarettes smoked in a day to create three categories as non-smoking father, father smokes a pack or less a day (1-20 cigarettes), father more than a pack a day (21 or more cigarettes). We also created two continuous variables modeling the number of cigarettes smoked by the mother: From 0-18 cigarettes; > 18 cigarettes to be used in a piecewise model.

We used PROC FREQ to summarize categorical variables and PROC UNIVARIATE to summarize continuous variables. Tests of linear association were performed using

linear regression analysis (PROC REG) or using (PROC GLM). A one-way analysis of variance was performed to compare the maternal smoking groups on Z-score for birth weight using PROC GLM. The 'lsmeans' statement was used in the GLM procedure with Tukey's adjustment to make pairwise comparisons. A two-factor analysis of variance was performed to compare the maternal and paternal smoking groups on Z-score for birth weight using PROC GLM. An interaction term was also included between maternal and paternal smoking categories. A linear regression was performed to compare the maternal smoking groups and maternal BMI on Z-score for birth weight using PROC REG and PROC GLM was used for pairwise comparisons. Scatter plot were formed using PROC SGPLOT to check the linearity. A multiple linear regression model was performed to test the association of various variables with Z-score for birth weight. A piecewise linear model created using two continuous variables for maternal smoking to predict Z-score for birth weight.
We conducted various regression models and carried out hypothesis tests keeping different predictors in the model to check the relation of parental variables to the low birth weight of child.

III. RESULTS

A. DESCRIPTIVE STATISTICS (QUESTION 2)

Descriptive statistics were produced for the entire sample of 680 participants. Categorical variables are summarized in Table 1 which provides frequencies and percentages. The sample is predominantly not having small birth weight with 596 of the 680 total sample or 87.65% representing the babies that were not having small birth weight. In mothers, half of the sample has never smoked (56.25%), about one fifth of the population smoke less than 20 cigarettes (24.40%), and the remainder were smoking more than 21 cigarettes a day (19.35%). In fathers, one third of the sample has never smoked (32.22%), about one fifth of the population smoke less than 20 cigarettes (24.96%), and the remainder were smoking more than 21 cigarettes a day (42.81%).

Table 1: Descriptive Statistics for Patient Characteristics

Variable	(N=680)
Small birth weight	
Not small	596 (87.65%)
Small	84 (12.5%)
Maternal Smoking Category	
No smoker	378 (56.25)
1-20	164 (24.40)
>21	130 (19.35%)
Paternal Smoking Category	
No smoker	213 (32.22%)
1-20	165 (24.96%)
>21	283 (42.81%)

Continuous data are summarized in Table 2 which displays the sample size, mean, median, standard deviation, and minimum and maximum values. The mean birth weight of the baby in pounds of the sample is 7.52 and it ranged from 3.30 to 11.40. Nutritional information varies widely as seen in Table 2. For each variable, we can examine the minimum, maximum and standard deviation giving us an idea of the range and variability of each variable. We found the mean and median to be somewhat different for certain variables (alcohol consumption, beta-carotene, plasma beta-carotene). For the alcohol consumption variable, it appears that there is a likely outlier of 29 alcoholic drinks/day that may be skewing the data. Another important characteristic to note is that compared to the absolute measure, the natural log for the plasma beta-carotene measure has less variability and is more normally distributed. Note that there was no missing data for any of the patient characteristics. However, a few variables had missing data so the total number is not 680 for all variables.

Table 2: Descriptive Statistics for Patient Characteristics

Descriptive Statistics for Exposure variables						
	N	Mean	Median	Std. Dev.	min	max
Birth weight of baby in pound	672	7.52	7.60	1.10	3.30	11.40
Head circumference	656	13.22	13.00	0.63	11.00	15.00
Length of baby	672	20.27	20.00	0.98	17.00	23.00
Gestational age	673	39.76	40.00	1.88	29.00	48.00
Maternal age	680	25.86	25.00	5.46	15.00	42.00
Maternal BMI	680	21.47	21.13	2.63	15.55	39.71
Paternal age	680	28.80	28.00	6.13	18.00	52.00
Paternal Education	680	13.38	14.00	2.20	6.00	16.00
Paternal Height	680	70.62	71.00	2.64	62.00	79.00
Descriptive Statistics for Outcome Variables						
Z score for birth weight in grams	672	0.00	0.07	1.00	-3.86	3.54

B. Analysis

Part 1 –Model for Z-score for birthweight of the child and maternal smoking.

A global one-factor ANOVA was performed to test for an association between the z-score for the birthweight of the child and the maternal smoking category to determine if maternal smoking category is related to Z-score for birth weight. Maternal

smoking category had 3 factors-

i) non-smoking mother.
ii) mother smokes a pack or less a day (1-20 cigarettes);
iii) Mother smokes more than a pack a day (21 or more cigarettes)

Table 4

Global model	Degrees of Freedom	F Value	P value	R square
	2,662	17.57	<.0001	0.050410

Table 5: This table shows the parameter estimates for each category

	Parameter Estimate	Standard Error	T-value	P value	95% Confidence Limits	
Intercept	0.1961	0.05049	3.88	0.0001	0.09696	0.2952
MATSMKCAT 2	-0.47649	0.0920	-5.18	<.0001	-0.6571	-0.2958
MATSMKCAT 3	-0.4221	0.1005	-4.20	<.0001	-0.6194	-0.2248

<u>The results of table 4 Global model</u>

F statistic = 17.57,
$df = 2{,}662$,
p-value<0.001,
$R^2 = 0.050410$.

When we conducted the global test, we found that there is statistical significance at the level of α=0.05, as p<0.0001, this shows there is a difference in z-score for birth weight among the maternal smoking categories.

The overall model shows that the maternal smoking category account for 5% variability in predicting the z-score for birth weight.

The equation for the linear model of the Z-score can be formed as follows.

Z-score= 0.1961 + -0.4765(Smoke cat 2) + -0.4221(Smoke cat 3).

If the mother smokes less than 20 cigarettes a day, then Z-score will be -0.2804.

If the mother smokes more than 20 cigarettes a day, then Z-score will be -0.2260.

These results show that maternal smoking has a negative effect on the birth weight of the child and can increase the probability of the child having a Z-score for birth weight in the lower 10% of Z-distribution. The results seem a bit bizarre because smoking more

than 20 cigarettes a day does not cause the maternal smoking to be more dangerous than smoking less than 20 cigarettes a day.

Next, we proceed to look at the pairwise comparisons between group 1 and 3, group 2 and 3 and group 1 and 2 to see which groups have significant differences in the mean Z-score for birth weight.
The table 6 shows the adjusted means and standard error for the different groups.

Table 6

Maternal smoking category	Adjusted means	Standard error
1(nonsmoker)	0.19609372	0.05048868
2(smoke <20 cigarettes)	-0.28039184	0.07691844
3(smoke >20 cigarettes)	-0.22604039	0.08687326

Table 7: Pairwise comparision results

	t-value	p-value
Comparison between 1 and 3	-5.17872	<.0001
Comparison between 2 and 3	-4.20121	<.0001
Comparison between 1 and 2	-0.46842	0.8861

We test the hypothesis of a statistically significant difference in the z- score for birth weight between nonsmokers and heavy smokers adjusting for light smokers.

For the first pairwise comparison, we compare non-smokers and heavy smokers adjusting for light smokers and observe the z-score for birth weight for non-smokers to be 0.1961 and observe the z-score for birth weight for heavy smokers to be -0.2260. The p-value for the pairwise comparison group <0.0001, hence the difference between the two groups seems to be significant. The mean difference between z- Score non-smokers and heavy smokers adjusting for light smokers (t-value) is -5.1787.

We test the hypothesis of a statistically significant difference in the z- score for birth weight between light smokers and heavy smokers adjusting for no smokers.

For this pairwise comparison, we compare light smokers and heavy smokers adjusting for no smokers, and observe the z-score for birth weight for light smokers to be -0.2804 and observe the z-score for birth weight for heavy smokers to be -0.2260. The p-value for the pairwise comparison group <0.0001, hence the difference between the two groups seems to be significant. The mean difference between z- Score of light smokers and heavy smokers adjusting for no smokers (t-value) is -4.2012.

We test the hypothesis of a statistically significant difference in the z- score for birth weight between nonsmokers and light smokers adjusting for heavy smokers.

For this pairwise comparison, we compare nonsmokers and light smokers adjusting for heavy smokers and observe the z-score for birth weight for nonsmokers to be 0.1961 and observe the z-score for birth weight for light smokers to be -0.2804. The p-value for the pairwise comparison group 0.8861, hence the difference between the two groups does not seem to be significant. The mean difference between z- Score nonsmokers and light smokers adjusting for heavy smokers (t-value) is -0.46842.

From these results, we can conclude that the mean z-score changes significantly in light smokers and heavy smokers and among nonsmokers and heavy smokers but there is not much difference between no smokers and light smokers.

Part 2- Model to test association between maternal smoking categories and paternal smoking categories and Z-score for birth weight.

A two-factor ANOVA with interaction was performed to test for an association between the maternal smoking categories and paternal smoking categories and the dependent variable is the z-score for birth weight. The maternal smoking category has three levels nonsmoker is level 1, a mother who smokes less than 20 cigarettes a day is level 2 and those who smoke more than 21 cigarettes a day is level 3. Similarly, the paternal smoking category has three levels nonsmoker is level 1, a father who smokes less than 20 cigarettes a day is level 2 and those who smoke more than 21 cigarettes a day is level 3.

Table 8

	F Value	P value	R-square
Overall model	5.05	<.0001	0.059433

When we conducted the global test we found that there is statistical significance at level of α=0.05, as p value<.0001 this shows that *at least* one level of the maternal smoking category or paternal smoking category differs from other with respect to z-score. The maternal and paternal smoking category together explain almost 6% variability in z-score.

By adding parental smoking to the earlier model we can see that the R^2 increased from 5.04% to 5.94%. Hence, the model improved.

Table 9

Source	DF	Type III Sums of Squares	F Value	Pr > F
MATSMK_CAT	2	25.38350840	13.16	<.0001
PATSMK_CAT	2	2.52382765	1.31	0.2709

Source	DF	Type III Sums of Squares	F Value	Pr > F
MATSMK_CA*PATSMK_CAT	4	4.05092462	1.05	0.380

The interaction for maternal smoking category and paternal smoking category is not significant at the p-value of 0.380, at the significance level of α=0.05, hence we removed the interaction term and rerun the model to look at the individual effects.

Table 10

Source	DF	Type III Sums of Squares	F Value	Pr > F
MATSMK_CAT	2	30.31088538	15.71	<.0001
PATSMK_CAT	2	0.74557273	0.39	0.6796

From the results we can see that
The Type 3 Results for maternal smoking category show that f-value=15.71 df=2,4 p<.0001

The p-value for maternal smoking category is <.0001 after adjusting for paternal smoking category, hence we can conclude that there is a linear association between z-score and maternal smoking category after adjusting for paternal smoking categories.

The Type 3 Results for paternal smoking category show that f-value=0.39 df=2,4 p=0.6796

The p-value for paternal smoking category is 0.6796 after adjusting for maternal smoking category, hence we can conclude that there is no linear association between z-score and paternal smoking category after adjusting for maternal smoking categories.

There is unequal allocation in each category of maternal smoking category which leads to an unbalanced design.
In order to determine which level is showing significant differences we conduct post-hoc test using tukey's test.

Table 11

MATSMK_CAT	Z_BWTG LSMEAN	Standard Error	p-value
1	-0.2847	0.0794	0.0004
2	-0.2306	0.0917	0.0122
3	0.1888	0.0523	0.0003

Table to show pairwise comparison between maternal categories.

Table 11

	t-value	p-value
Comparison between 1 and 3	-4.9419	<.0001
Comparison between 2and 3	-3.9850	0.0002
Comparison between 1 and 2	-0.4523	0.8934

For the first pairwise comparison, we compare maternal smoking category 1 with maternal smoking category 2 and observe the adjusted mean z-score for category 1 to be 0.1888 and for the category 2 to be -0.2848. The p-value for the pairwise comparison group is 0.8934, there does not seem to be significant differences between these two groups. The t value for the difference between the two means is -0.4524.

For the second pairwise comparison, we compare maternal smoking category 1 with the maternal smoking category 3 and observe the adjusted mean z-score for category 1 to be 0.1888 and for the category 3 to be -0.2306. The p-value for the pairwise comparison group is <0.0001, there seem to be significant differences between these two groups. The t-value for the difference between the two means is -4.9419.

For the third pairwise comparison, we compare maternal smoking category 2 with the maternal smoking category 3 and observe the adjusted mean z-score for category 2 to be -0.2848 and for the category 3 to be -0.2306.The p-value for the pairwise comparison group is 0.0002, there seem to be significant differences between these two groups. The t value for the difference between the two means is -3.9850

This result shows that there is significant difference in z-score for birth weight in between light smoker and heavy smoker and no smoker and heavy smoker categories, while there is not a significant difference between low smoker and no smoker categories.

We do not conduct post-hoc test on paternal smoking categories as it is not significant. The adjusted means for the paternal smoking categories are as shown in the table 12.

Table 12

PATSMK_CAT	Z_BWTG LSMEAN	Standard Error
2	-0.15931929	0.07955071
3	-0.09946216	0.06000217
1	-0.06777062	0.07611284

From our results we can conclude that the maternal smoking category is significantly associated with Z-score for birth weight while the paternal smoking category and interaction for paternal and maternal smoking category are not.

Part 3- Model to test association of Z- score for birth weight and maternal smoking category and maternal BMI

We ran a multiple regression model using maternal smoking category and maternal BMI to predict Z- score for birth weight. The overall model is significant with a p-value <0.0001 suggesting that there is an association between Z-score for birth weight and maternal smoking category and maternal BMI. $(\beta_{matsmokecat} \neq \beta_{matBmi} \neq 0)$.

Level of significance: α=0.05

Estimates of interest:
F = 17.74, df = (2,662)
R^2 = 0.0509
p<0.0001
Hence, we can conclude that there is significant evidence of an association between Z-score and at least one of the predictors (maternal BMI and maternal smoking category.) The variables maternal BMI and maternal smoking category together explain 5% of the variability in z-score.

Now we check the interaction between the maternal BMI and maternal smoking category.

The results of the global model show that

Global model
F-statistic= 8.46
df= 5,659
p-value <.0001

R^2 = 0.060285

There is significant evidence of an association between maternal BMI and maternal smoking category and the interaction between maternal BMI and maternal smoking category at α=0.05 level with p value <0.001. Now we move forward to look at the interaction.

Interaction of maternal BMI and Maternal Smoking category

F-statistic= 0.04
df= 659
p-value= 0.9598

R^2 = 0.060285

Conclusion- The interaction is not statistically significant with a p-value of 0.9598 at the level of α=0.05, hence we can conclude that there is no interaction between *maternal BMI and maternal smoking category*.

Since the interaction term is non-significant as a whole, we would not look further at the individual interactions between the separate categories of maternal smoking.

From the results of this model, we see that the p-value is not significant for the interaction of Maternal BMI and maternal smoking category. Hence, we can say that Maternal BMI is not an effect measure modifier.

Maternal BMI is a covariate since the p-value of 0.0090 is significant we can say that it is a covariate.

To check if Maternal BMI is a confounder or not, we use the 10% rule of confounding (adjusted-crude/crude). Taking the crude and adjusted values of parameter estimates from the linear regression models above we find that maternal smoking categories are not confounders for Z-score for birth weight.

-0.4386-0.4765/-0.4765= -0.0793

7.9% is lesser than 10% hence we can conclude that there is no presence of joint confounding on maternal smoking category 2 to Z-score for birth weight by maternal BMI.

-0.4062+0.4221/-0.4221= 0.0378

3.7% is lesser than 10% hence we can conclude that there is no presence of joint confounding on maternal smoking category 3 to Z-score for birth weight by maternal BMI.

A single linear term for BMI is not good enough to use in modeling the relationship of BMI and Z- score for birth weight.

Part 4- Model for z-score of birth weight and the maternal BMI.

We run a simple linear regression model with z-score of birth weight as outcome and the maternal BMI as predictor.

Table 13

	Degree of Freedom	F Value	P value	R-Square
Model	1, 670	11.04	0.0009	0.0162

F-value= 11.04
DF= 1,670
P-value= 0.0009
R^2 = 0.0162

From the results of the model we see that the overall model is significant at the α=0.05 level with a p-value of 0.0009. From this we can conclude that there is significant evidence of a linear association between z-score of birth weight and the maternal BMI.

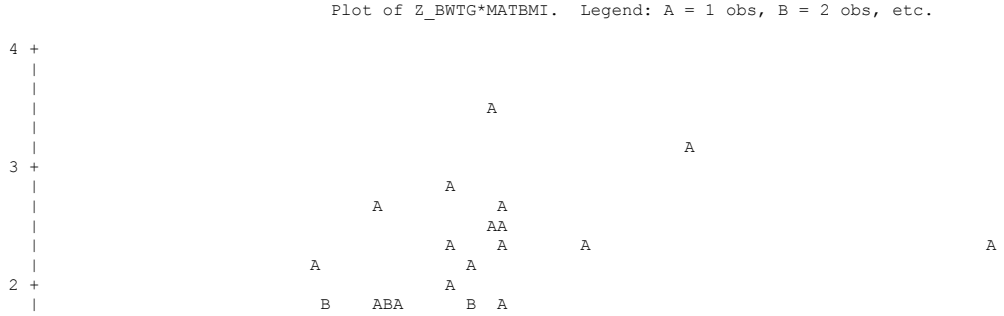
The R^2 of this model is 1.62% hence the maternal BMI is not a good predictor of Z-score of birth weight.

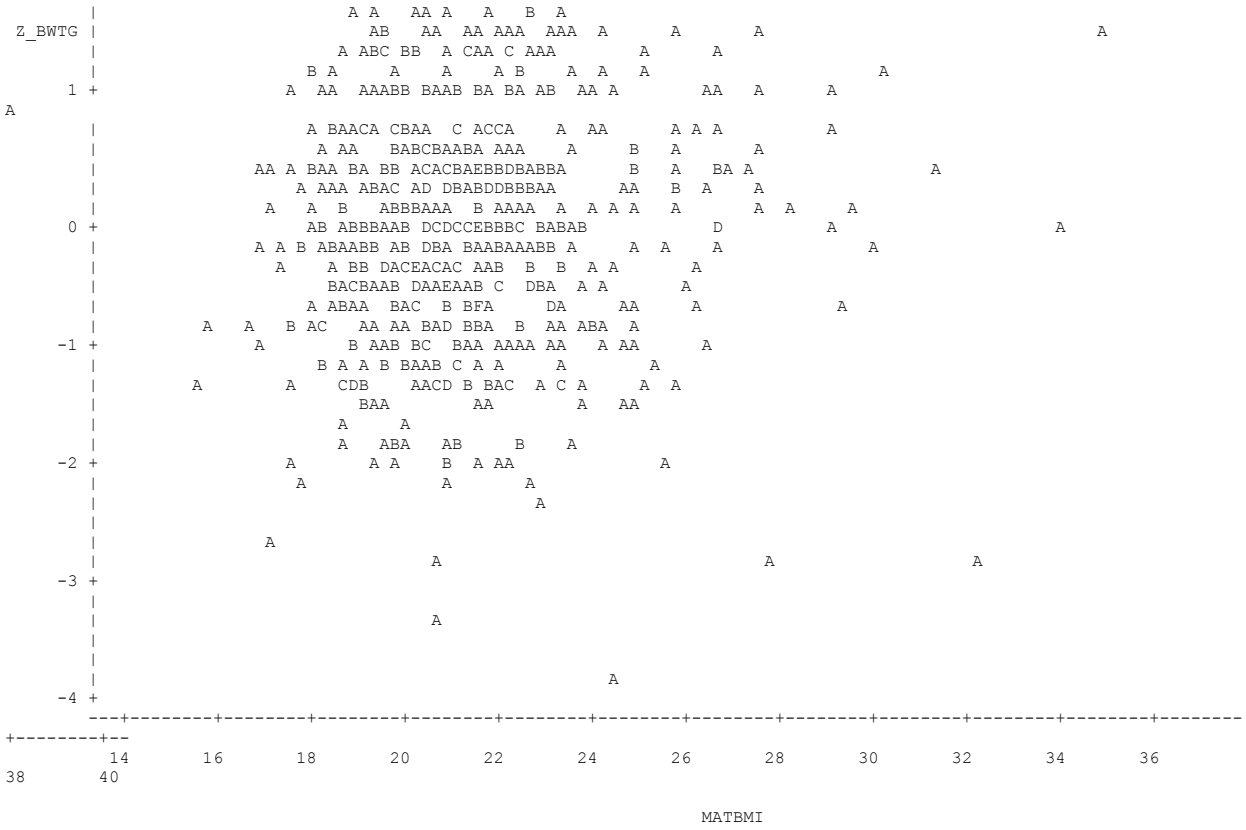
Table 14

	Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	P value	95% confidence interval
Intercept	1	-1.04120	0.31572	-3.30	0.0010	-1.6611, -0.4213
MATBMI	1	0.04850	0.01460	3.32	0.0009	0.01984.0.07717

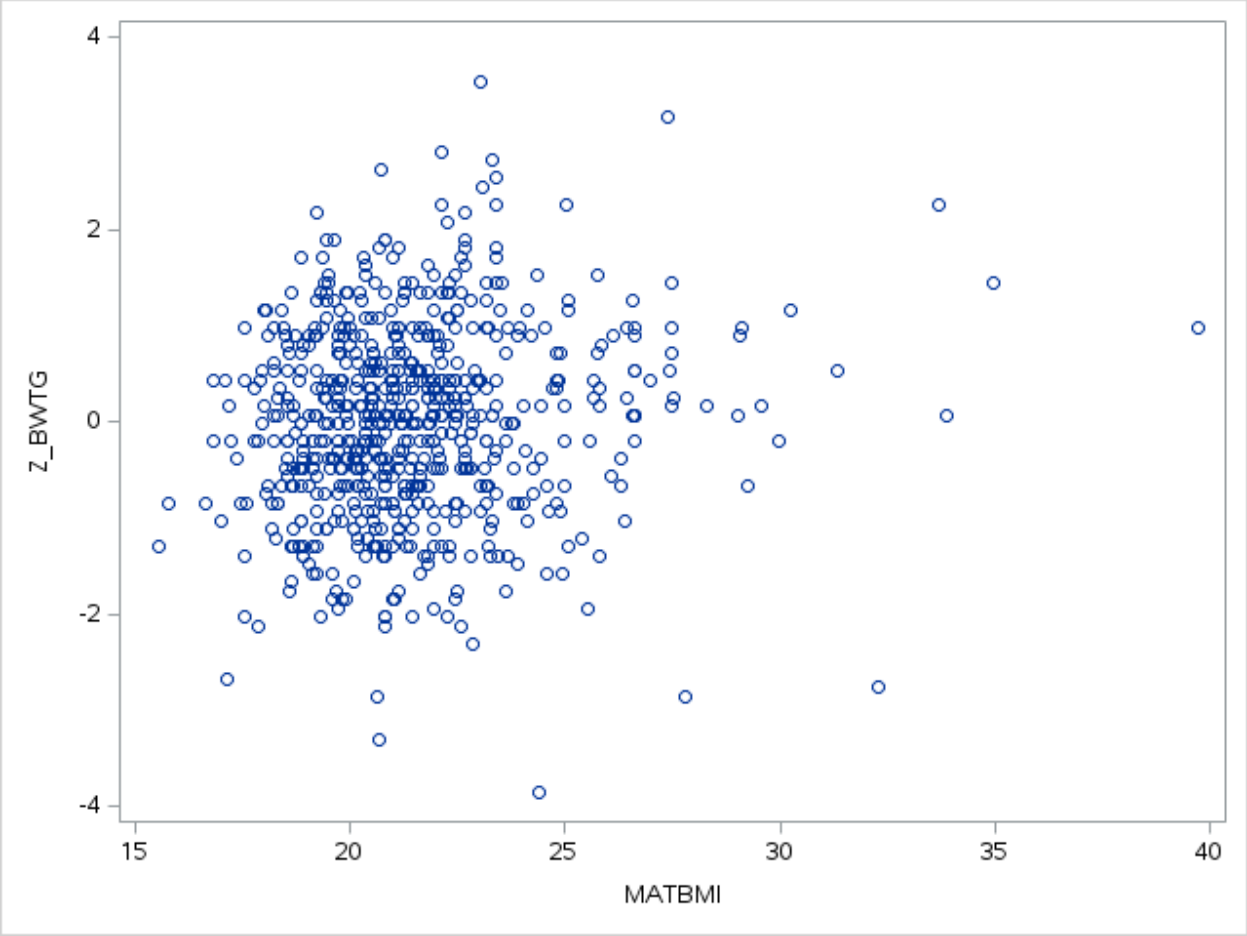
With 1 unit increase in the maternal BMI, the z-score would increase by 0.04850. So, this results show that with increased maternal BMI the z-score would increase and there would be less chances of baby having small birth weight.

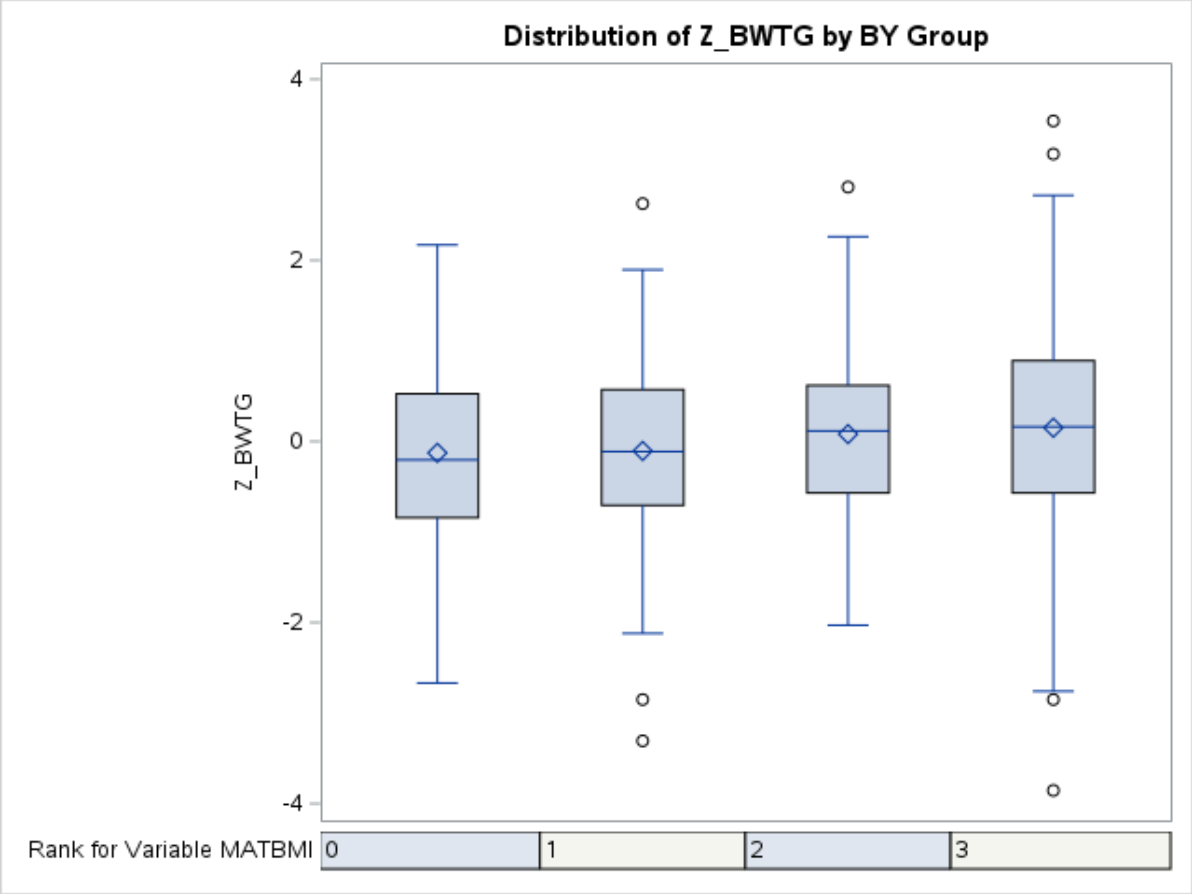
To check the linearity of the maternal BMI variable, we use graphs. The scatter plot and the box plot show that the maternal BMI variable does not follow a linear trend.



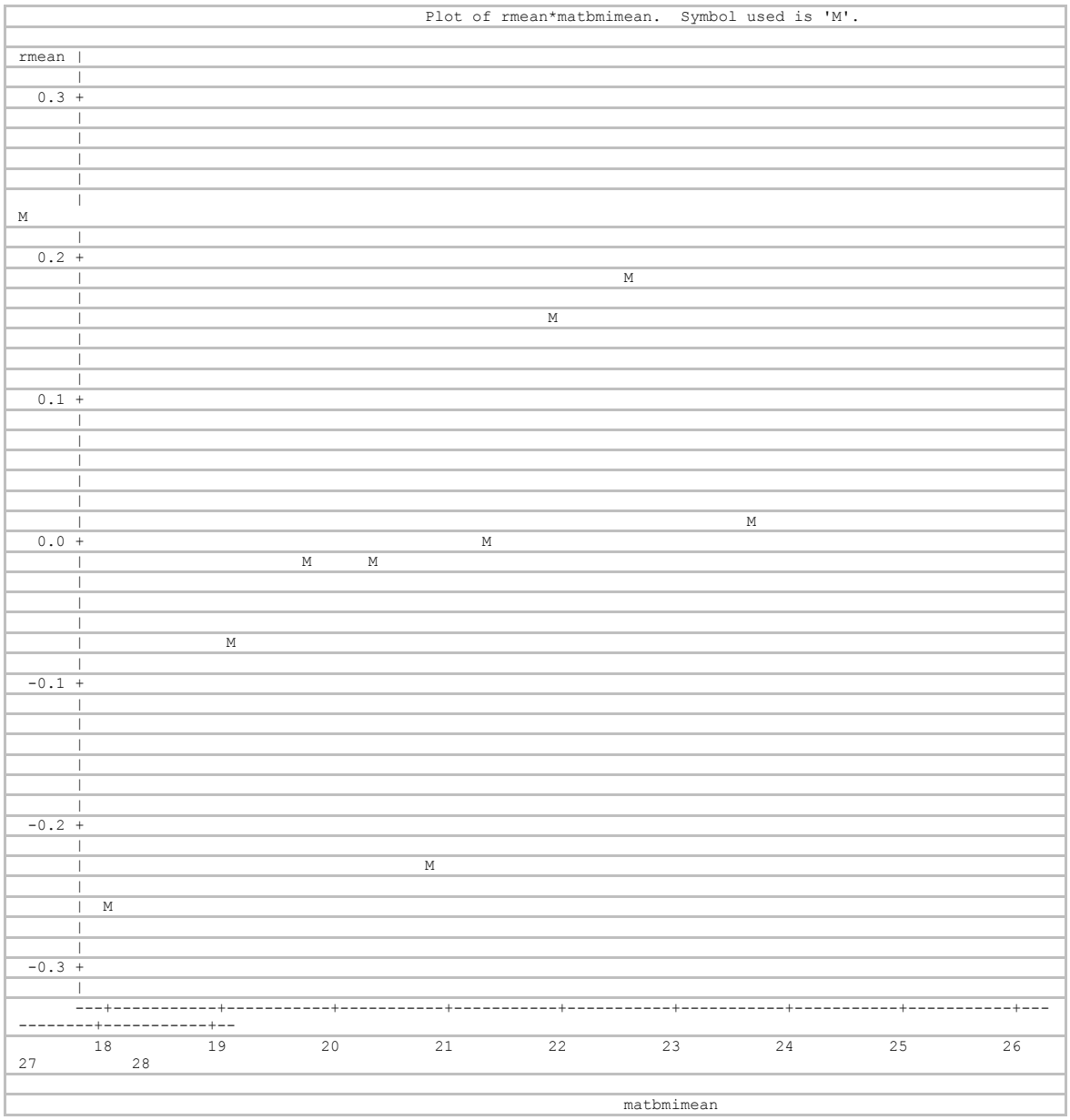


NOTE: 8 obs had missing values.


4
2
0
-2
-4
Z_BWTG
15
20
25
30
35
40
MATBMI



Maternal BMI does not seem to have a linear relationship to z- score for birth weight. In the boxplots, there might be a suggestion of curvature in the relationship between Maternal BMI and Z-score.



Scatter plots indicate that maternal BMI is probably not a good predictor of the outcome z-score as there is lot of variability.
Based on the means plot instead of being linear there is a somewhat curved trend.

Part 5 – Model with gestational age, maternal age, maternal BMI, paternal age, paternal education, paternal height and maternal and paternal smoking categories as predictors of Z-score for birth weight

We run a multiple linear regression model with gestational age, maternal age, maternal BMI, paternal age, paternal education, paternal height and maternal and paternal smoking categories as predictors of Z-score for birth weight.

First we test the global model and then move onto individual effects.

Level of significance: $\alpha=0.05$

Estimates of interest:
$F = 26.23$,
df = (8, 632)
$R^2 = 0.2397$
$p<0.0001$

From this model we conclude that there is significant evidence of an linear association between *z*- score for birth weight and gestational age, maternal age, maternal BMI, paternal age, paternal education, paternal height and maternal and paternal smoking categories. All the variables together explain 23.97% of the variability in *z*-score for birth weight.

Since the global test was significant, we now look at the individual level tests.

Table 15

	Parameter Estimate	Standard Error	T Value	P value	95% Confidence Limits	
Intercept	-13.87273784	1.24490327	-11.14	<.0001	-16.31739998	-11.42807570
Gage	0.21456821	0.01851694	11.59	<.0001	0.17820582	0.25093060
Matage	0.00010617	0.01123417	0.01	0.9925	-0.02195479	0.02216712
MATBMI	0.03549345	0.01347945	2.63	0.0087	0.00902337	0.06196353
Patage	0.00305428	0.00997798	0.31	0.7596	-0.01653984	0.02264840
Pated	-0.00099242	0.01689556	-0.06	0.9532	-0.03417084	0.03218601
Patht	0.06599874	0.01366061	4.83	<.0001	0.03917290	0.09282459
MATSMK_CAT 2	-0.35859398	0.08751908	-4.10	<.0001	-0.53045840	-0.18672957
MATSMK_CAT 3	-0.38284714	0.09475350	-4.04	<.0001	-0.56891805	-0.19677622

	Parameter Estimate	Standard Error	T Value	P value	95% Confidence Limits	
MATSMK_CAT 1	0.00000000	.	.	.	.	.
PATSMK_CAT 2	0.00129171	0.09612396	0.01	0.9893	-0.18747043	0.19005385
PATSMK_CAT 3	-0.01530868	0.08604915	-0.18	0.8589	-0.18428654	0.15366918
PATSMK_CAT 1	0.00000000	.	.	.	.	.

1) Results for gestational age:

$\hat{\beta}_{gage=}$ 0.21456821
t = 11.59, df = 630
p<.0001
SE= 0.01851694
For every one unit increase in gestational age was associated with an increase in z-score of 0.2145 (p<0.0001), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and gestational age, adjusting for all the other variables in the model.

2) Results for maternal age:

$\hat{\beta}_{mage=}$ 0.00010617
t = 2.63, df = 630
p=0.9925
SE= 0.01123417
For every one unit increase in maternal age was associated with an increase in z-score of 0.0001 (p=0.9925), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and maternal age, adjusting for all the other variables in the model.

3) Results for maternal BMI:

$\hat{\beta}_{matbmi=}$ 0.03549345
t = 0.01, df = 630
p=0.0087
SE= 0.01347945

For every one unit increase in maternal age was associated with an increase in z-score of 0.0355 (p<0.0001), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and maternal BMI, adjusting for all the other variables in the model.

4) Results for Paternal age:

$\hat{\beta}_{patage=}$ 0.00305428
t = 0.31, df = 630
p=0.7596

SE= 0.00997798

For every one unit increase in paternal age was associated with an increase in z-score of 0.003 (p=0.7596), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and paternal age, adjusting for all the other variables in the model.

5) Results for Paternal education:

$\hat{\beta}_{pated=}$ -0.0009924
t = -0.06, df = 630
p=0.9532

SE= 0.00997798

For every one unit increase in paternal age was associated with an increase in z-score of -0.0009 (p=0.9532), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is not significant evidence of an association between z- score for birth weight and paternal education, adjusting for all the other variables in the model.

6) Results for Paternal height:

$\hat{\beta}_{patht=}$ 0.06599874

t = 4.83, df = 630
p<.0001

SE= 0.01366061

For every one unit increase in paternal height was associated with an increase in z-score of 0.0659 (p<0.0001), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and paternal height, adjusting for all the other variables in the model.

7) Results for Maternal smoking category 2:

$\hat{\beta}_{matsmk\_cat2}$ = -0.35859398

t = -4.10, df = 630
p<.0001

SE= 0.08751908

For every one unit increase in maternal smoking category 2 (less than 20 cigarettes a day) was associated with an increase in z-score of -0.3585 (p<0.0001) as compared to maternal smoking category 1 (non smoker), assuming all other variables in the regression model remain constant

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and maternal smoking category, adjusting for all the other variables in the model.

8) Results for Maternal smoking category 3:

$\hat{\beta}_{matsmk\_cat3}$ = -0.38284714

t = -4.04, df = 630
p<.0001

SE= 0.09475350

For every one unit increase in maternal smoking category 3 (more than 20 cigarettes a day) was associated with an increase in z-score of -0.3585 (p<0.0001) as compared to maternal smoking category 1 (non smoker), assuming all other variables in the regression model remain constant.

9) Results for Paternal smoking category 2:

$\hat{\beta}_{patsmk\_cat2}$ = 0.00129171

t = 0.01, df = 630

p-0.9893

SE= 0.09612396

For every one unit increase in paternal smoking category 2 (less than 20 cigarettes a day) was associated with an increase in z-score of 0.001 (p=0.9893) as compared to paternal smoking category 1 (non smoker), assuming all other variables in the regression model remain constant.

10)Results for Paternal smoking category 3:

$\hat{\beta}_{patsmk\_cat3}$= -0.01530868

t = -0.18, df = 630

p-0.8589

SE= 0.08604915

For every one unit increase in paternal smoking category 3 (more than 20 cigarettes a day) was associated with an increase in z-score of -0.0153 (p=0.8589) as compared to smoking category 1 (non smoker), assuming all other variables in the regression model remain constant.

Hence, we can conclude that there is significant evidence of an association between z-score for birth weight and maternal smoking category, adjusting for all the other variables in the model.

The significant predictors in this model for predicting the Z-score for birth weight are gestational age, maternal BMI, paternal height and maternal smoking categories.
To check the level of significance and strength of linear association we get the table of standardized estimates as given below.

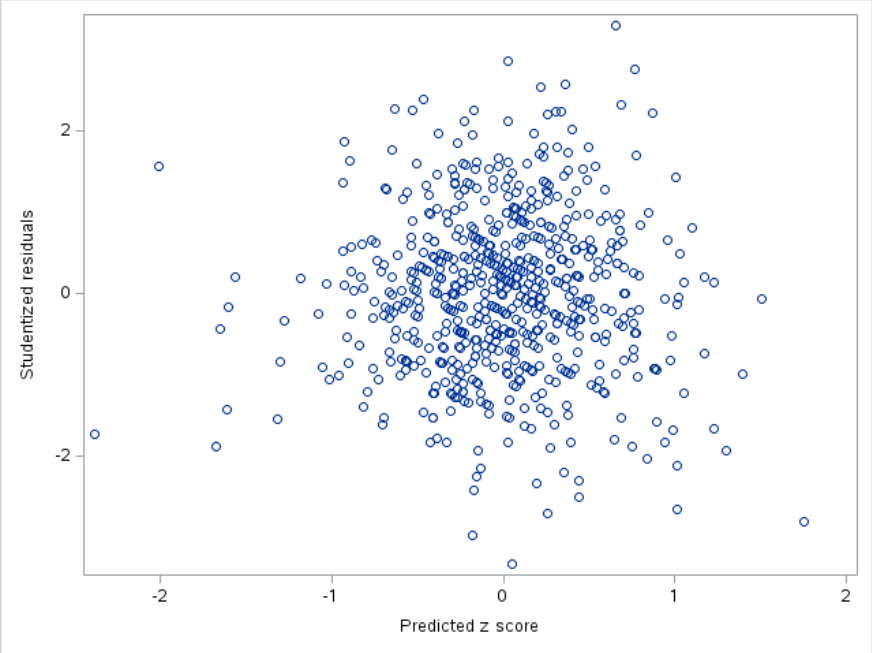
Table 16

Parameter	Estimate	Standardized Estimate
Intercept	-13.872738	0
Gage	0.214568	0.402142
Matage	0.000106	0.000576
MATBMI	0.035493	0.093514

Parameter	Estimate	Standardized Estimate
Patage	0.003054	0.018722
Pated	-0.000992	-0.002178
Patht	0.065999	0.170487
MATSMK_CAT 2	-0.358594	-0.151929
MATSMK_CAT 3	-0.382847	-0.149505
MATSMK_CAT 1	0	0
PATSMK_CAT 2	0.001292	0.000554
PATSMK_CAT 3	-0.015309	-0.007507
PATSMK_CAT 1	0	0

We get this table of standardized beta estimates by using proc glmselect procedure. According to this table, we can rank the variables according to their strength of their linear association with z-score for birth weight.
Gestational age is a variable that has the strongest linear association with z-score for birth. The order of strength of variables is as follows
Gestational age> Paternal height > paternal age>Mother smoking less than 20 cigarettes> Mothers smoking more than 20 cigarettes> Maternal BMI > Fathers smoking more than 20 cigarettes> Paternal education> Maternal age > Father smoking less than 20 cigarettes.

For identifying outliers, we check the studentized residuals. There are 2 points that look like outliers in the residual plot that are over the cut-off value of 3.



The problem points are the observations having studentized residual greater than 3, we find that 2 points could be problematic for our model.

Table 17: For identifying outliers

Obs	ID	r_zbwtg	p_zbwtg
1	442	3.29966	0.65606
17	523	-3.33697	0.05373

For identifying influence points, we find the cook's distance. Those points having value of greater than *4/n=0.006* were identified as influence points. These points are summarized in the table below.

Table 18: for influence points having Cook's D> 0.006

Obs	ID	Studentized residuals	Cook's distance
1	17	-1.72753	0.022767
2	50	1.72144	0.006171
3	123	1.35658	0.007158
4	143	-2.81785	0.049095

Obs	ID	Studentized residuals	Cook’s distance
5	223	-1.57985	0.007294
6	436	1.53521	0.030073
7	442	3.29966	0.017395
8	488	1.53125	0.006500
9	492	-1.88214	0.029848
10	523	-3.33697	0.027041
11	550	-2.30428	0.007517
12	551	-2.71760	0.006742
13	566	-1.68732	0.007378
14	569	-2.33568	0.010073
15	584	2.38333	0.006240
16	690	-2.03193	0.008422
17	720	-2.65554	0.017226
18	759	-1.53636	0.008709
19	771	-2.11997	0.006747
20	866	2.21565	0.006868
21	892	2.76144	0.015102
22	928	-2.51483	0.011441
23	971	2.58245	0.006928
24	976	-1.93165	0.007679
25	1031	2.86794	0.009648

Obs	ID	Studentized residuals	Cook's distance
26	1055	2.23705	0.023924
27	1071	2.54186	0.006834
28	1144	2.19733	0.009896
29	1161	-1.66730	0.015702
30	1186	2.27781	0.011358
31	1188	1.19990	0.007640
32	1308	1.63382	0.006418
33	1313	-1.88481	0.013549
34	1324	-2.98622	0.038845
35	1428	-2.25149	0.010985
36	1532	-1.55429	0.006103
37	1641	-1.42747	0.006533
38	1657	1.55788	0.008095
39	1709	-1.78046	0.007784
40	1712	-2.20449	0.008596

The Variation Inflation Factor values for any variable do not exceed 10 and there is no clear indication of a collinearity problem related to these variables.

Table 19: Variation inflation factor

Variable	Tolerance	Variance Inflation Factor
Gage	0.99170	1.00837
Matage	0.31857	3.13900

Variable	Tolerance	Variance Inflation Factor
MATBMI	0.95616	1.04585
Patage	0.31723	3.15230
Pated	0.86453	1.15669
Patht	0.95449	1.04768
MATSMK_CAT	0.92433	1.08187
PATSMK_CAT	0.88666	1.12783

Table 20: Collinearity Diagnostics

Collinearity Diagnostics (intercept adjusted)										
Number	Eigenvalue	Condition Index	Proportion of Variation							
			Gage	Matage	MATBMI	Patage	Pated	Patht	MATSMK_CAT	PATSMK_CAT
1	1.98828	1.00000	0.00159	0.06703	0.01198	0.06677	0.04235	0.00385	0.00249	0.01083
2	1.30517	1.23426	0.02718	0.00848	0.02418	0.00816	0.02021	0.00090903	0.28600	0.29030
3	1.16038	1.30900	0.00105	0.00018766	0.17317	0.00185	0.28652	0.28177	0.03580	0.00698
4	1.01365	1.40054	0.45167	0.00005438	0.20090	0.00006675	0.01502	0.22413	0.00237	0.05837
5	0.96497	1.43543	0.49804	0.00011671	0.29322	0.00027996	0.00003085	0.21526	2.190798E-7	0.00054155
6	0.77119	1.60568	0.01579	0.01085	0.19513	0.00725	0.08072	0.12384	0.55970	0.16947

Collinearity Diagnostics (intercept adjusted)										
Number	Eigenvalue	Condition Index	Proportion of Variation							
			Gage	Matage	MATBMI	Patage	Pated	Patht	MATSMK_CAT	PATSMK_CAT
7	0.62087	1.78953	0.00096724	0.01171	0.10096	0.01137	0.55362	0.13899	0.11356	0.46342
8	0.17548	3.36604	0.00372	0.90158	0.00045808	0.90425	0.00155	0.01125	0.00008443	0.00007217

The condition index for principal component 8 is approximately 3.3 and over 90% of the variance of paternal age and maternal age is explained by it.
There are no collinearities in this regression as none of the condition index is above 10.

Part 6 – Piecewise Model to test association of z-score for birth weight and smoking groups 1 and 2.

First we run a global model to see if the z-score for birth weight is same across both the smoking groups.

Table 21

	DF	F Value	Pr > F	R square
Global Model	1,663	8.01	0.0048	0.011940

Global model
F-statistic=8.01
df= 1,663
p-value = 0.0048

R^2 = 0.011940

When we conducted the global test we found that there is statistical significance at level of α=0.05, as p=0.0048, this shows there is difference in z-score for birth weight among both the smoking groups.

Now we run piecewise linear model using *smoking category 1 and smoking category 2* to predict z-score for birth weight.

We test the association between predict z-score for birth weight and smoking groups $(\beta_{smk1} \neq \beta_{smk2} \neq 0)$.

Level of significance: α=0.05

Estimates of interest:
F = 21.81,
df = (2,662)
R^2 = 0.0618
P<0.0001

We conclude that there is a significant evidence of a linear association predicting z-score for birth weight and the smoking groups. The smoke groups explain 6.18% of the variability in predicting the z-score for birth weight.

Table 22

DF	F Value	Pr > F	R square
2,662	21.81	<.0001	0.0618

Now we look at individual effects of the smoke groups.

Table 23

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.35572	0.18766	-1.90	0.0584
smk1	1	-0.04301	0.00675	-6.37	<.0001
smk2	1	0.02971	0.01036	2.87	0.0043

Smoke group 1:
We test the hypothesis of a linear association between z-score for birth weight and smoke group 1 after adjusting for smoke group 2($\beta_{smoke1} \neq 0$| smk2)
Level of significance: α=0.05

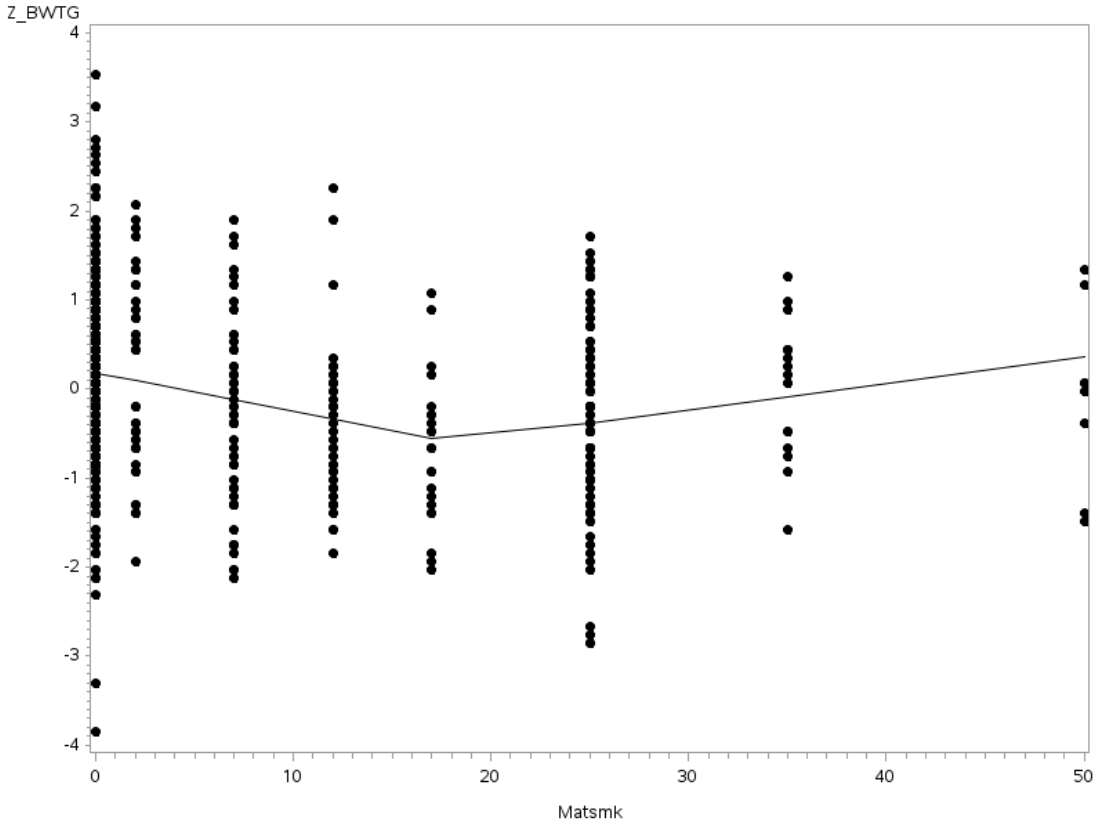
Results:
$\hat{\beta}_{smk1}$= -0.04301
t = -6.37 df =662
p<0.0001
SE= 0.00675
For every one unit change in smoke 1 the z-score for birth weight will decrease by -0.04301 (p<0.0001), assuming all other variables in the regression model remain constant.

Conclude that there is significant evidence of an association between z-score for birth weight and smoke group 1 after adjusting for smoke group 2
Individual estimated slope for smoke 1 = -0.35572+smoke1*-0.04301.

Smoke group 2:
We test the hypothesis of a linear association between z-score for birth weight and smoke group 2 after adjusting for smoke group 1($\beta_{smoke2} \neq 0$| smk1)
Level of significance: α=0.05

Results:
$\hat{\beta}_{smk2=}$ 0.02971
t = 2.87 df =662
p=0.0043
SE= 0.01036
For every one unit change in smoke 1 the z-score for birth weight will increase by 0.02971 (p<0.0001), assuming all other variables in the regression model remain constant.

Conclude that there is significant evidence of an association between z-score for birth weight and smoke group 2 after adjusting for smoke group 1
Individual estimated slope for smoke 2 = -0.35572+ smoke2*0.02971.



There is a change point at the 18 cigarettes and we can see there is no significant increase in the two groups in z-score for birth weight.
Based on the plot, it appears that the relationship between Z-score for birth weight and the number of cigarettes smoked is non- linear.
From this plot, we can conclude that there is a slight increase in probability of z-score for birth weight being less than 10% cut-off for mothers who smoke less than 18 cigarettes a day as the number of cigarettes increase but the

Now we compare the slopes of smoke group 1 and smoke group 2
Table 24

	DF	F Value	Pr > F
Test 1 results	1,662	21.59	<.0001

Results:
F value = 21.59
df for numerator and denominator = 1,662
$p<0.0001$

We conclude from this test that slope of smoke group 1 is not equal to slope of smoke group 2 as the p-value is statistically significant at level of $\alpha=0.05$.

Part 7- Conclusion

A number of analyses were completed on the 680 participants of the study. From our analyses, we see that there is an influence of parental smoking on the birth weight of the child.
From the first model we concluded that there is significant association of maternal smoking categories with Z-score for birth weight. Based on the parameter estimates of the individual smoking categories we saw that there is not much change in effect based on the number of cigarettes smoked per day but smoking can increase the probability of a child being low birthweight. The results were a bit strange as we see that mothers who smoke less than 20 cigarettes a day were having greater chances of having a child who was low birth weight as compared to mothers who smoked more than 20 cigarettes a day. In order to see the relation among the maternal smoking cateogries on Z-score for birth weight we carried out pairwise comparisions, the results of these showed that there is significant difference between high smokers and non smoker in terms of child with low birth weight. There was also significant difference between low smokers and high smokers for low birth weight child but not among low smokers and non smokers.
Next, we added paternal smoking categories to the model. The model was still significant showing that parental smoking has a significant effect on low birth weight of the child. However, the interaction between maternal and paternal smoking categories was not significant. The paternal smoking category individually was not significant for causing low birth weight in a child.

In the third model, we tested the association of maternal BMI and maternal smoking categories to Z-score for birth weight and concluded that the overall model was significant. The interaction between maternal BMI and maternal smoking categories was not significant. From our analysis, we found out that maternal BMI is not an effect modifier or confounder but it is a covariate.
In the fourth model, we test the association of maternal BMI and Z-score for birth weight. The results of this model show that there is a significant association of maternal BMI to Z-score for birth weight. However, this association is not linear based on the scatter plot and the means plot. Since there is lot of variability in the outcome the means plot is the best method to check the linearity. The means plot indicates that there is a somewhat curved trend in the plot.
In the model 5, we run a multilinear regression with gestational age, maternal age, maternal BMI, paternal age, paternal education, paternal height and maternal and paternal smoking categories as predictors of Z-score for birth weight. Gestational age should be included in the model as it contributes to the effect of Z-score of birth weight. The results of this model show that adding all these variables to our model increases our Rsquare to 25%. The significant predictors in this model were gestational age, maternal BMI, paternal height, and maternal smoking categories. We also found a few influence point in the model based on the Cook's distance. The Variation Inflation Factor does not exceed 1 hence there is no collinearity problem relating to these variables. Finally, we created a piecewise model to test association of Z-Score with the two smoking groups created as continuous variables, the results show that the model is statistically significant and the slopes for both the groups are not equal.

Overall it looks like gestational age, paternal height and maternal smoking and maternal BMI has an influence on the birth weight of the child.