Project 2

STA 9797

Erik Carrion

Fall 2021

Introduction

The Birth Weight dataset consists of 50,000 observations across 10 variables, of which the target of our analysis is Infant Birth Weight. The variables recorded in the dataset include factors such as the mother's race and marital status as well as her level of education and whether she smokes or not. Of concern in this analysis is the effect of Education, if any, on infant birth weight.

It is well established in scientific literature that smoking during pregnancy is associated with a host of negative outcomes, including low birth weight. We must therefore account for this variable in any model we consider.

Before performing any analysis, we first had to modify the factor levels for Mother's Education and the Visit variable to accord with reason. For example, originally, Level 3 of Mother's Education corresponded to "Less Than High School" while Level 0 represented "High School". The Levels were re-set so that 0= "Less Than High School", 1 = "High School", 2 = "Some College", and 3 = "College". A similar re-structuring was done for the Visit Variable, although ultimately it was not considered.

In order to facilitate further analysis, we created a new, categorical target variable named "Above Average" which is an indicator variable for whether or not the given infant is above or below the average Birth Weight.

We begin with an Exploratory Analysis of the entire dataset to better understand the distribution of our target variable with respect to our factors of interest. We then perform the same analysis on a stratified random sample of the data with the levels of Mother's Education used as strata.

The stratified sample is then used to investigate the effect of Education, if any, on Infant Birth. We first investigate a 1-Way ANOVA Model with Education as our only factor of interest and compare this to a 2-Way Model which considers both Education and whether the mother smokes or not. We then account for a potential nuisance variable in the form of Weight Gain as we expect a mother who gains more weight will have a heavier baby.

We proceed to investigate the relationship between Education and whether a mother smokes by making use of the Cochran-Armitage Trend Test and Fisher's Exact Test.

Finally, we conduct a logistic regression where we regress our categorical target variable, "Above Average", against Education, Smoking, and Weight Gain to determine if Education can be used as part of a linear classifier to predict whether a woman's baby will be above or below average birth weight.

Exploratory Data Analysis

Looking at the entire dataset by the Mother's Level of Education and their smoking status, we have the following summary statistics:

Education=0

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
0	0	5864	5864	284.0000000	5330.00	3345.00	3319.66	549.8297506
	1	2109	2109	312.0000000	5245.00	3146.00	3091.75	581.6122909

Education=1

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
1	0	14484	14484	240.0000000	5415.00	3402.00	3369.32	576.2264006
	1	2965	2965	457.0000000	5131.00	3204.00	3178.54	570.3198430

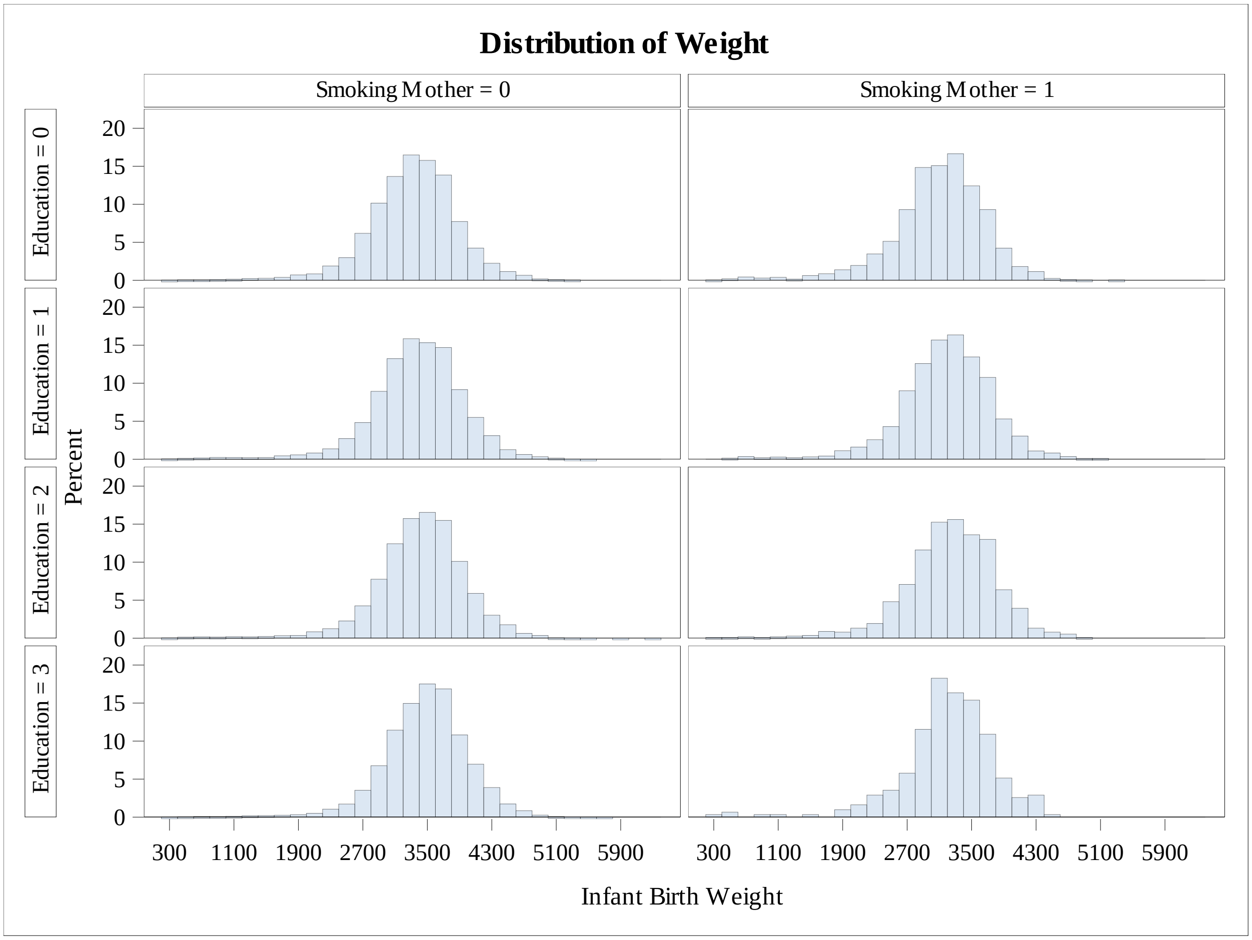
Education=2

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
2	0	10982	10982	330.0000000	6350.00	3430.00	3411.26	560.9301042
	1	1147	1147	369.0000000	4961.00	3255.00	3231.34	569.2570306

Education=3

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
3	0	12137	12137	340.0000000	5642.00	3487.00	3473.50	527.7873055
	1	312	312	322.0000000	4423.00	3260.00	3200.66	587.1402700

We notice immediately that regardless of education level, children born to Mothers who smoke tend to have lower birth weights than children whose mothers do not smoke. The differences between levels of education, regardless of smoking or not smoking, however, seem relatively the same. To investigate, we look at the histograms.



The distributions look to align with what we saw above in that the distribution of Weight across levels of education seems to be the same with the primary difference being driven by whether the mother smokes or not.

After taking a stratified random sample we get the following summary statistics and histograms.

Education=0

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
0	0	102	102	1077.00	4630.00	3359.50	3315.69	576.3027426
	1	48	48	1276.00	4366.00	3330.50	3246.54	514.6009798

Education=1

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
1	0	118	118	936.0000000	4706.00	3384.50	3319.58	621.9472499
	1	32	32	2182.00	4621.00	3073.50	3117.72	549.2452437

Education=2

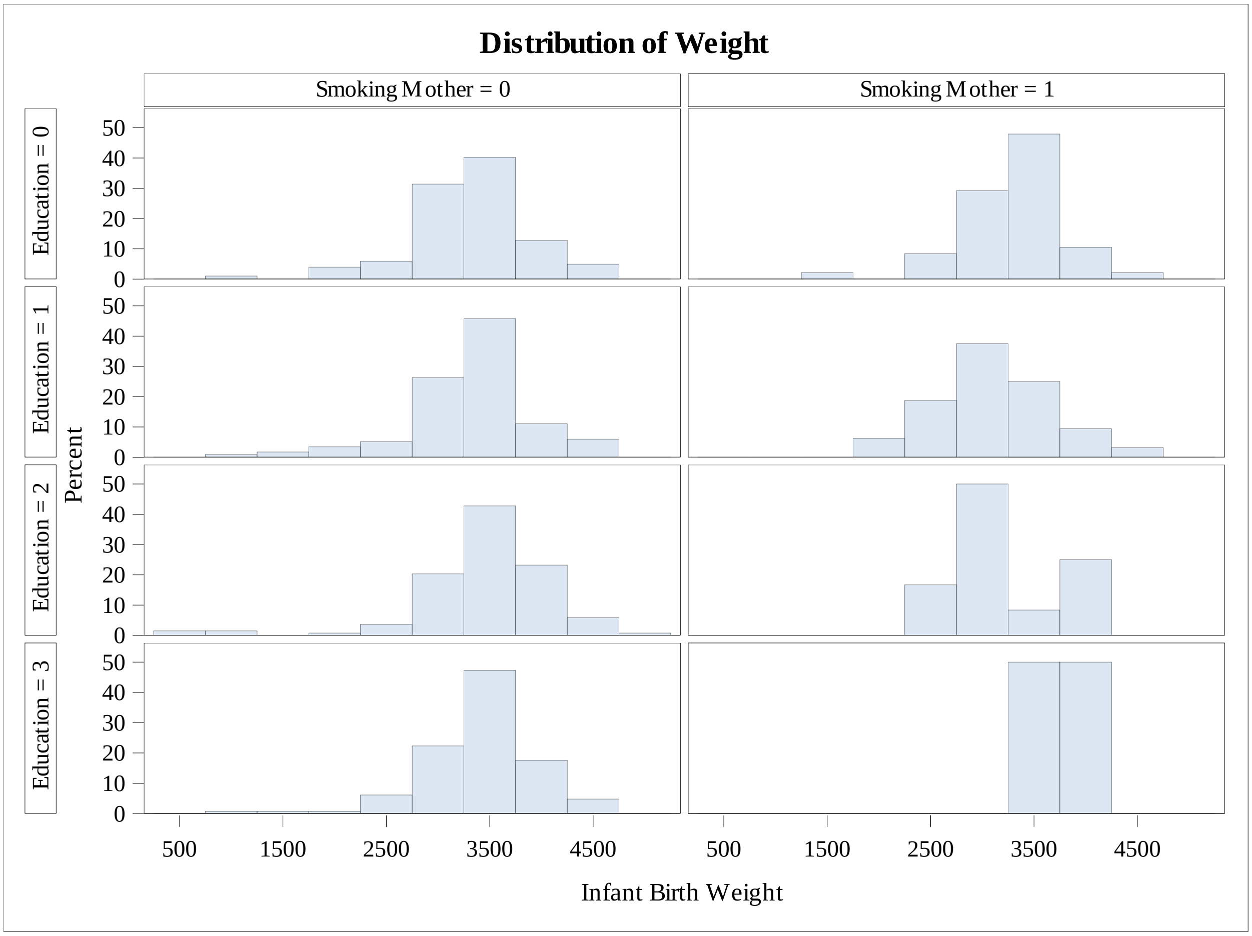
Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
2	0	138	138	369.0000000	4848.00	3547.00	3448.88	661.6207423
	1	12	12	2637.00	4082.00	3143.50	3294.17	494.2187284

Education=3

Analysis Variable : Weight Infant Birth Weight								
Education	**Smoking Mother**	**N Obs**	**N**	**Minimum**	**Maximum**	**Median**	**Mean**	**Std Dev**
3	0	148	148	1077.00	4536.00	3435.00	3427.97	515.1238067
	1	2	2	3346.00	4090.00	3718.00	3718.00	526.0874452

Our sample reflects the larger dataset as a whole where there is a significant difference between smoking and non-smoking mothers while the differences in levels of Education seem insignificant. Of note, however, is the increase in birth weight for a smoking mother when her level of Education is 3. Given there are only 2 observations for this particular level, the effect of chance on this outcome is likely high.

Looking at the histograms, we see that the data still look approximately normally distributed with approximately equal variances. With that said, if we compare the largest variance to the smallest variance, the ratio is less than 2x, allowing us to perform ANOVA with confidence that our results will be valid.



Distribution of Weight
Smoking Mother = 0
Smoking Mother = 1
Education = 0
Education = 1
Education = 2
Education = 3
Percent
0
10
20
30
40
50
500
1500
2500
3500
4500
Infant Birth Weight

1-Way & 2-Way ANOVA

In order to better understand what effect, if any, Education has on Birth Weight we first perform a 1-Way ANOVA with Education as our independent variable. We do so because our response variable, in this case Weight, is continuous and our variable of interest is categorical.

The results of our 1-Way ANOVA are below:

1-Way ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3359224.9	1119741.6	3.27	0.0210
Error	596	204153528.7	342539.5		
Corrected Total	599	207512753.6			

R-Square	Coeff Var	Root MSE	Weight Mean
0.016188	17.42077	585.2687	3359.603

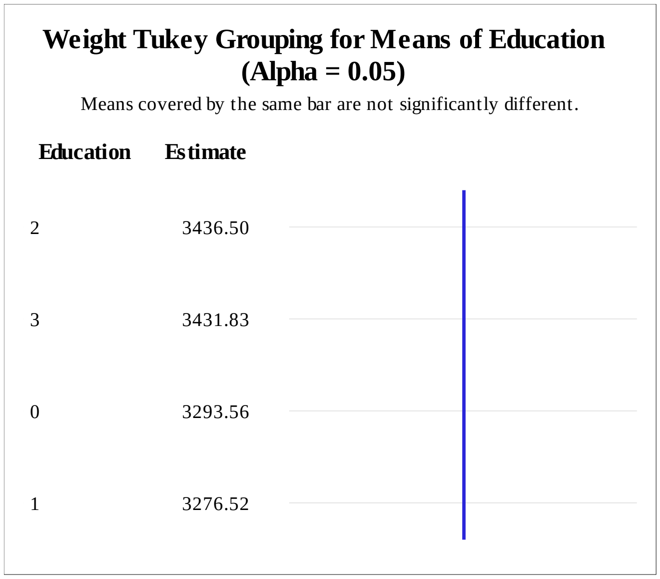
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	3	3359224.860	1119741.620	3.27	0.0210

Levene's Test for Homogeneity of Weight Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Education	3	2.137E12	7.124E11	1.06	0.3646
Error	596	3.997E14	6.706E11		

First, the test for Homogeneity of Variance satisfies model assumptions. Second, the overall model is significant with a p-value of .0210. However, the model only accounts for 1.6188% of the total variation in our response, suggesting that, in our sample, Education does not play a large role in explaining the variability in Infant Birth Weight.

We employed Tukey's Range Test to make pairwise comparisons while constraining our alpha level to .05. The results show no significant difference in birthweight between the levels.

Alpha	0.05
Error Degrees of Freedom	596
Error Mean Square	342539.5
Critical Value of Studentized Range	3.64334
Minimum Significant Difference	174.1



We now consider the 2-Way model which includes whether a mother is a smoker or a non-smoker. We first consider the fully saturated model. While our sample is balanced with respect to Educational Level, they are not balanced with respect to Smoking and so we employ proc glm in our analysis.

2-Way ANOVA – Fully Saturated Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	4971327.2	710189.6	2.08	0.0443
Error	592	202541426.4	342130.8		
Corrected Total	599	207512753.6			

R-Square	Coeff Var	Root MSE	Weight Mean
0.023957	17.41037	584.9195	3359.603

Source	DF	Type II SS	Mean Square	F Value	Pr > F
Education	3	2163260.125	721086.708	2.11	0.0981
MomSmoke	1	1013909.634	1013909.634	2.96	0.0857
Education*MomSmoke	3	598192.697	199397.566	0.58	0.6265

The fully saturated model suggests that the interaction is not significant so we can consider a main effects model only. The results of which are given below.

2-Way ANOVA – Main Effects Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	4373134.5	1093283.6	3.20	0.0129
Error	595	203139619.1	341411.1		
Corrected Total	599	207512753.6			

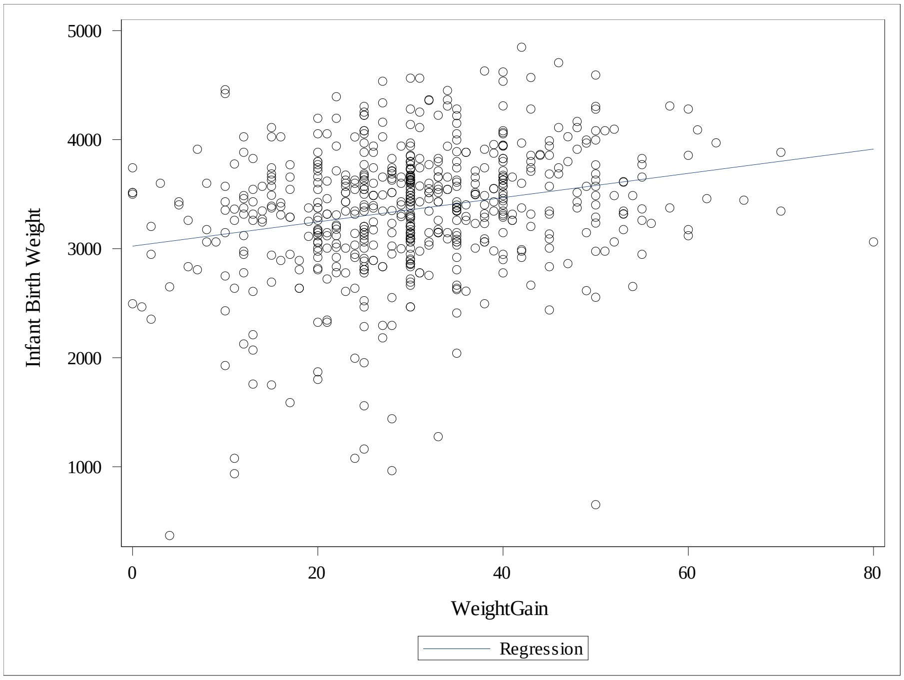
R-Square	Coeff Var	Root MSE	Weight Mean
0.021074	17.39205	584.3040	3359.603

Source	DF	Type II SS	Mean Square	F Value	Pr > F
Education	3	2163260.125	721086.708	2.11	0.0975
MomSmoke	1	1013909.634	1013909.634	2.97	0.0854

While the overall model is significant, the individual variables are not, suggesting there is a missing covariate which could help account for the variation in Infant Birth Weight. Of the variables recorded, we hypothesize that Weight Gain is associated with the response and should be included in the model.

ANCOVA

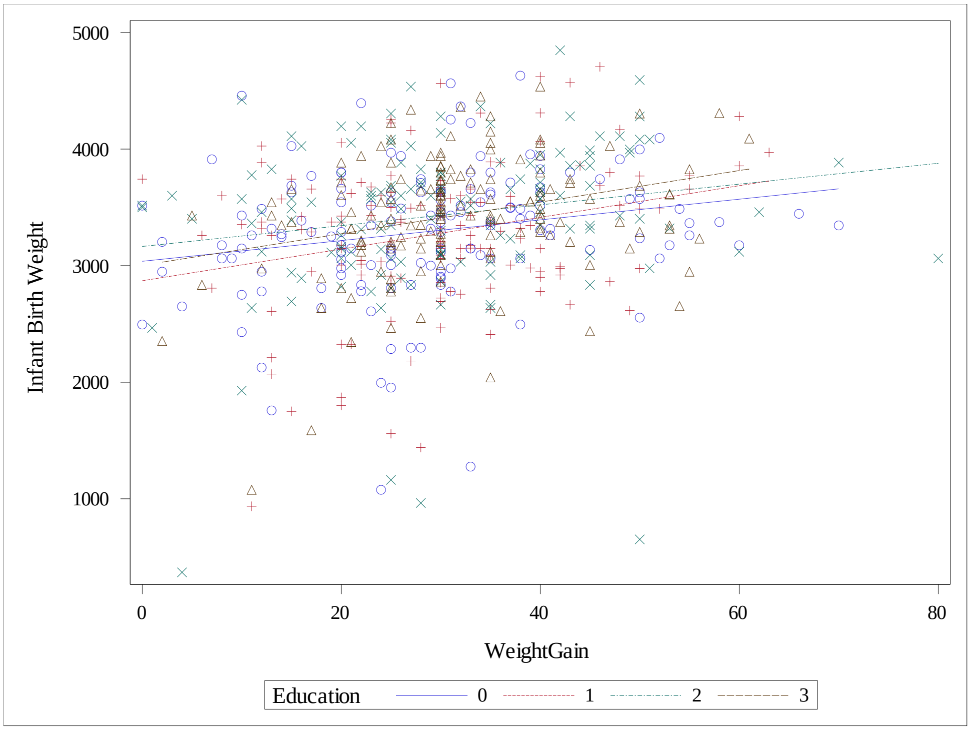
In order to determine if Weight Gain should be included in the model, we first assess its relationship to Infant Birth Weight. We see that there is a positive relationship between Weight Gain and Infant Birth Weight. To determine its significance, we regress Infant Birth Weight on Weight Gain, resulting in a model that is overall significant and whose slope coefficient is significantly different from zero.

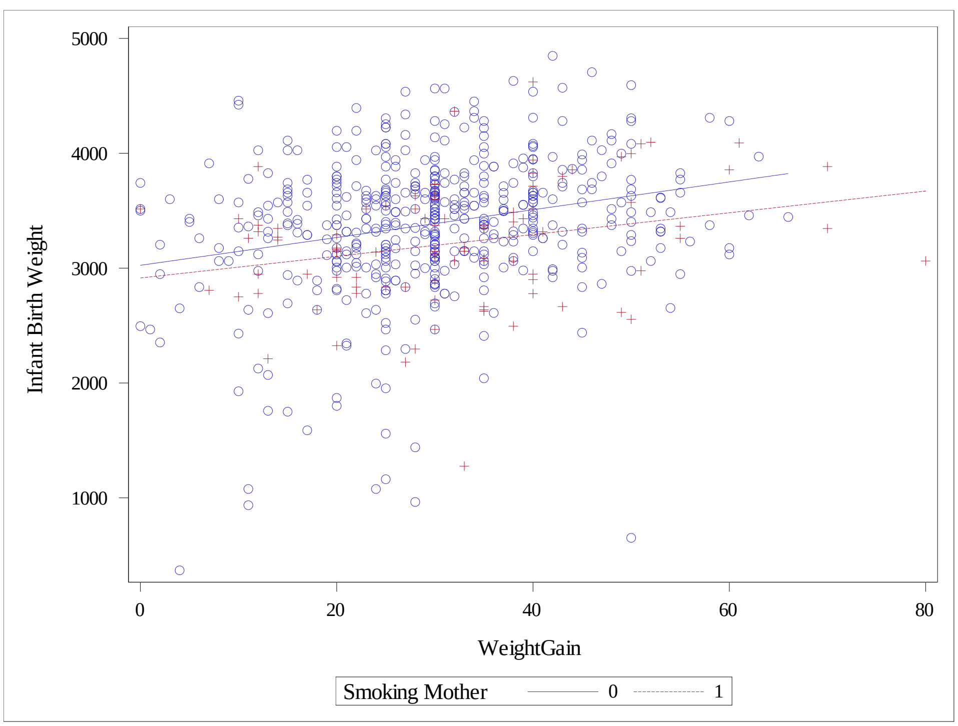


Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	11478319	11478319	35.01	<.0001
Error	598	196034435	327817		
Corrected Total	599	207512754			

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	3023.43644	61.43155	49.22	<.0001
WeightGain		1	11.12706	1.88043	5.92	<.0001

Our next step is to understand how Education and Smoking Status relates to Weight with respect to Weight Gain. As we see in the plots below, there are interactions among the levels of Education. For Smoking, it's a different story and we don't immediately disqualify it from analysis. To determine whether we will proceed, we must first test for equal slopes.





Test for Slope Equality

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MomSmoke	1	160883.747	160883.747	0.50	0.4811
WeightGain	1	7363670.416	7363670.416	22.75	<.0001
WeightGain*MomSmoke	1	110113.855	110113.855	0.34	0.5600

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Education	3	954686.86	318228.95	0.98	0.4033
WeightGain	1	11210307.35	11210307.35	34.40	<.0001
WeightGain*Education	3	486570.02	162190.01	0.50	0.6840

We note that the interaction is not significant for either Education or Smoking and thus we will be able to proceed with our analysis. First, we consider a fully saturated model with Education, MomSmoke, and WeightGain which shows that the interactions are not significant allowing us to focus on a main effects model.

The results of our main effects analysis show that Education is not significant and can be excluded from the model.

ANCOVA – Fully Saturated Model Output

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Education	3	247172.216	82390.739	0.25	0.8595
MomSmoke	1	13451.931	13451.931	0.04	0.8391
Education*MomSmoke	3	66600.105	22200.035	0.07	0.9769
WeightGain	1	4801589.636	4801589.636	14.72	0.0001
WeightGain*Education	3	153603.332	51201.111	0.16	0.9252
WeightGain*MomSmoke	1	21609.383	21609.383	0.07	0.7970
Weight*Educat*MomSmo	3	72388.116	24129.372	0.07	0.9740

ANCOVA – Main Effects Model Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	15919366.8	3183873.4	9.87	<.0001
Error	594	191593386.8	322547.8		
Corrected Total	599	207512753.6			

R-Square	Coeff Var	Root MSE	Weight Mean
0.076715	16.90476	567.9329	3359.603

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Education	3	1467128.28	489042.76	1.52	0.2092
MomSmoke	1	1808018.57	1808018.57	5.61	0.0182
WeightGain	1	11546232.28	11546232.28	35.80	<.0001

Fisher’s Exact Test & Cochran-Armitage Test

Given Education is not a significant predictor, we ask if there is a relationship between Education and Smoking and if so, what is the nature of that relationship, and does it affect Education’s significance? In order to investigate this line of questioning, we employed Fisher’s Exact Test & the Cochran Armitage Trend Test.

From the output below, we see that the Chi-Square Test for Row Independence has a p-value of <.0001. As the level of mother’s Education increase, there is a significant difference in the percentages of mothers who do not smoke in the sample. If we look to the Cochran-Armitage test, we note that the statistic is 7.93 with a p-value of <.0001 suggesting a positive trend in the row percentages so that as the level of education increases, the percentage of mothers who do not smoke also increases.

These results are confirmed by Fisher’s Exact test (in SAS this is actually the Freeman-Halton test for general R X C tables) which concludes that there is a significant association between the two variables.

As the level of Education increases, our proportion of smokers decreases, however, the mere presence of smoking within our model is strong enough to absorb any residual variation that would have led to Education being significant. Thus, the relationship between Education and Smoking is such that it allows us to eliminate Education entirely.

Statistic	DF	Value	Prob
Chi-Square	3	63.9307	<.0001
Likelihood Ratio Chi-Square	3	72.4843	<.0001
Mantel-Haenszel Chi-Square	1	62.8771	<.0001
Phi Coefficient		0.3264	
Contingency Coefficient		0.3103	
Cramer's V		0.3264	

Cochran-Armitage Trend Test	
Statistic (Z)	7.9361
One-sided Pr > Z	<.0001
Two-sided Pr > $\lvert Z \rvert$	<.0001

Table of Education by MomSmoke			
Education	MomSmoke(Smoking Mother)		
Frequency Percent Row Pct Col Pct	0	1	Total
0	102 17.00 68.00 20.16	48 8.00 32.00 51.06	150 25.00
1	118 19.67 78.67 23.32	32 5.33 21.33 34.04	150 25.00
2	138 23.00 92.00 27.27	12 2.00 8.00 12.77	150 25.00
3	148 24.67 98.67 29.25	2 0.33 1.33 2.13	150 25.00
Total	506 84.33	94 15.67	600 100.00

Fisher's Exact Test	
Table Probability (P)	<.0001
Pr <= P	<.0001

Logistic Regression

Given that Education was deemed insignificant in our prior analysis, we then asked if it could instead be used as part of a classifier to determine whether a baby would be above or below the average Infant Birth Weight. We use "Above Average" as our target variable.

Given that Weight Gain was significant in prior analyses, we include it in our Logistic Regression Models. First, we consider the fully saturated model and then move on to the main effects model. For the fully saturated model including Education, MomSmoke, and WeightGain, we have the following results:

Logistic Regression – Fully Saturated Model

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-1.1463	0.5090	5.0716	0.0243
Education	1		1	0.5856	0.7320	0.6400	0.4237
Education	2		1	1.2688	0.6770	3.5124	0.0609
Education	3		1	0.7331	0.7378	0.9874	0.3204
MomSmoke	1		1	0.2653	0.8929	0.0883	0.7664
Education*MomSmoke	1	1	1	-1.2188	1.5114	0.6502	0.4200
Education*MomSmoke	2	1	1	-3.2889	2.1236	2.3985	0.1215
Education*MomSmoke	3	1	1	-20.9378	985.6	0.0005	0.9831
WeightGain			1	0.0414	0.0169	5.9767	0.0145
WeightGain*Education	1		1	-0.0191	0.0237	0.6524	0.4192
WeightGain*Education	2		1	-0.0273	0.0221	1.5268	0.2166
WeightGain*Education	3		1	-0.0160	0.0235	0.4598	0.4977
WeightGain*MomSmoke	1		1	-0.0213	0.0273	0.6085	0.4354
Weight*Educat*MomSmo	1	1	1	0.0228	0.0460	0.2458	0.6200
Weight*Educat*MomSmo	2	1	1	0.0572	0.0494	1.3439	0.2464
Weight*Educat*MomSmo	3	1	1	0.5581	22.2665	0.0006	0.9800

The results show that the interactions are not significant and so we consider a main effects model.

Logistic Regression – Main Effects Model

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.5793	0.2654	4.7658	0.0290
Education	**1**	1	-0.1052	0.2381	0.1954	0.6585
Education	**2**	1	0.3125	0.2450	1.6276	0.2020
Education	**3**	1	0.1732	0.2492	0.4831	0.4870
MomSmoke	**1**	1	-0.7353	0.2498	8.6640	0.0032
WeightGain		1	0.0253	0.00699	13.0901	0.0003

No level of Education is deemed significant and so we conclude that Education can not be used as part of a linear classifier.

Summary

Our aim in this analysis was to investigate what effect, if any, Education has on Infant Birth Weight.

To arrive at a determination, we first asked if there was a significant relationship between Education and Weight and used a 1-Way ANOVA to discover that it does indeed have a relationship with our response variable, albeit a minor one.

Because smoking is known to cause adverse outcomes for fetal development, we asked how the relationship between Education and Infant Birth Weight was affected by the presence of a mother smoking or not. The 2-Way ANOVA produced a marginally better model and suggested we were missing part of the picture which led us to ask which of our recorded Variables could be acting as a nuisance.

We found that Weight Gain was positively associated with Infant Birth Weight and so we proceeded with an Analysis of Covariance which led us to a model which concluded that Education was not a significant variable.

We then proceeded to ask whether there was a dependent relationship between Education and Smoking that could explain why Education is not significant. We employed a Chi-Square, Fisher's Exact, and Cochran-Armitage test to determine that the two variables were significantly associated with each other. When this is the case, the variables tend to have an overlap in information, rendering one of them redundant. This is dependent on the relative amount of information each has with respect to the variation in our response. In this case, Smoking has more information than Education and thus in its presence, Education is rendered insignificant.

Finally, despite its prior insignificance, we asked if Education could be used as part of a linear classifier to predict whether a baby would be above or below average weight. As in our previous analysis, in the presence of Smoking and Weight Gain, Education was rendered insignificant.

Conclusion

During our initial investigation of the data, we hypothesized that a woman with greater Education would give birth to a heavier baby. A more educated woman will, on average, earn more than her less educated counterparts through better paying jobs and through these better paying jobs, would have access to better quality healthcare. We believed it reasonable that a more educated, higher paid woman, with better access to health care, would be more likely to engage in healthier living practices which would then lead to larger babies.

Our analysis shows that Education, in the presence of Smoking and Weight Gain, plays nearly no role in explaining the variation in Infant Birth Weight. The effect of Smoking and Weight Gain absorbs any residual variation that would have been taken by Education to render it significant.

With that said, our models are still lacking. Our best model in our ANCOVA analysis, the Main Effects model with just MomSmoke and WeightGain as variables, has an R^2 of 0.069645. There are variables which are positively related to our response but not accounted for in the dataset. These variables may include things such as annual income, annual health care expenditures, percentage of diet that is organic, the amount of pre-natal vitamins and supplements taken, and even genetic markers. Without access to these data, all we can say for certain is that a woman who doesn’t smoke and gains more weight throughout the pregnancy increases their likelihood of having an above average weight child.

Appendix - Code

```
/* Read In and Modify the Data */
data BirthWeight;
	set sashelp.BWeight;
	/* Re-Factor Visit */
		if Visit = 0 then PreNatal = 0;
		else if Visit = 1 then PreNatal = 2;
		else if Visit = 2 then PreNatal = 3;
		else PreNatal = 1;
	/* Re-Factor Education */
		if MomEdLevel = 0 then Education = 1;
		else if MomEdLevel = 1 then Education = 2;
		else if MomEdLevel = 2 then Education = 3;
		else Education = 0;
	/* Create New Binary Variable for BirthWeight */
		if Weight > 3370.76 then AboveAverage = 1; else AboveAverage = 0;
	/* Adjust Age to fold back in their median */
		Age = MomAge+27;
		WeightGain = MomWtGain + 30;
	/* Drop the Modified Variables */
		drop Visit MomEdLevel MomAge MomWtGain;
run;
/* Use PROC SURVEYSELECT to randomly sample the 50K observations */
/* First Sort by Education as it is our factor of interest */
proc sort data = BirthWeight;
	by Education;
run;
proc surveyselect data = BirthWeight method = srs seed = 1003 n=150
out=BirthWeightSRS;
	strata Education;
run;

/*************************************************/
/*************************************************/

/* EDA */
/* Summary Statistics */
proc means data = BirthWeight std n min max mean median;
	class Education MomSmoke;
	by Education;
	var Weight;
	title "Summary Statistics for Birth Weight by Level of Education and
Smoking - All Data";
run;

proc means data = BirthWeightSRS std n min max mean median;
	class Education MomSmoke;
	by Education
	var Weight;
	title "Summary Statistics for Birth Weight by Level of Education and
Smoking - Sample";
```

```
run;

/* HISTOGRAMS */
proc univariate data = BirthWeight;
      class Education MomSmoke;
      by Education;
      var Weight;
      histogram Weight/nrows = 4;
      ods select histogram;
      title "Distribution of Birth Weight by Level of Education - All Data";
run;
proc univariate data = BirthWeightSRS;
      class Education MomSmoke;
      by Education;
      var Weight;
      histogram Weight/nrows = 4;
      ods select histogram;
      title "Distribution of Birth Weight by Level of Education - Sample";
run;

/*************************************************/
/*************************************************/

/**** 1-WAY ANOVA ****/
proc anova data = BirthWeightSRS;
      class Education;
      model Weight = Education;
      means Education/tukey hovtest=levene;
      title "1-Way ANOVA Weight v Education";
run;

/**** 2-WAY ANOVA ****/
/* Fully Saturated Model */
proc anova data = BirthWeightSRS;
      class Education MomSmoke;
      model Weight = Education|MomSmoke;
      means Education MomSmoke;
      title "2-Way ANOVA Weight v Smoking & Education - Fully Saturated";
run;

/* Main Effects Model */
proc anova data = BirthWeightSRS;
      class Education MomSmoke;
      model Weight = Education MomSmoke;
      means Education MomSmoke;
      title "2-Way ANOVA Weight v Smoking & Education - Main Effects";
run;

/*************************************************/
/*************************************************/

/**** ANCOVA ****/
/* WEIGHT = WEIGHTGAIN - IS WEIGHTGAIN A COVARIATE? */
proc reg data=BirthWeightSRS;
      model Weight = WeightGain;
```

```
run;


/* Interaction Plots */
proc sgplot data=BirthWeightSRS;
      reg x=WeightGain y=Weight/group=Education;
run;
proc sgplot data=BirthWeightSRS;
      reg x=WeightGain y=Weight/group=MomSmoke;
run;


/* TEST FOR EQUAL SLOPES */
proc glm data=BirthWeightSRS;
      class MomSmoke;
      model Weight = MomSmoke|WeightGain;
run;
proc glm data=BirthWeightSRS;
      class Education;
      model Weight = Education|WeightGain;
run;

/* 2-WAY ANCOVA */
proc glm data=BirthWeightSRS;
      class Education MomSmoke;
      model Weight = Education|MomSmoke|WeightGain;
      title "2-WAY ANCOVA - Fully Saturated";
run;
proc glm data=BirthWeightSRS;
      class Education MomSmoke;
      model Weight = Education MomSmoke WeightGain;
      title "2-WAY ANCOVA - Main Effects";
run;
ods rtf close;

/************************************************/
/************************************************/

/**** FISHER'S EXACT, COCHRAN-ARMITAGE, CHI-SQUARE ****/

proc freq data = BirthWeightSRS;
      exact fisher;
      tables Education*MomSmoke/chisq trend;
      title "Proc Freq - Education x Smoking";
run;

/************************************************/
/************************************************/

/**** LOGISTIC REGRESSION ****/
proc logistic data = BirthWeightSRS;
      class Education(ref="0") MomSmoke(ref="0")/param=ref;
      model AboveAverage(event="1") = Education|MomSmoke|WeightGain / ctable
lackfit;
      title "Logistic Regression - Fully Saturated";
```

```
run;
proc logistic data = BirthWeightSRS;
      class Education(ref="0") MomSmoke(ref="0")/param=ref;
      model AboveAverage(event="1") = Education MomSmoke WeightGain / ctable
lackfit;
      title "Logistic Regression - Main Effects";
run;

proc logistic data = BirthWeightSRS;
      class MomSmoke(ref="0")/param=ref;
      model AboveAverage(event="1") = MomSmoke|WeightGain / ctable lackfit;
      title "Logistic Regression - No Education - Saturated";
run;
proc logistic data = BirthWeightSRS;
      class MomSmoke(ref="0")/param=ref;
      model AboveAverage(event="1")= MomSmoke WeightGain / ctable lackfit;
      title "Logistic Regression - No Education - Main Effects";
```