ECON 4848 Term Paper

Andrew Mikhaylovskiy, Michael Patel

November 2020

Contents

1 Abstract

Infant birth weight can be influenced by a multitude of factors, from how the mother behaves during pregnancy to the father's age. Studying these factors allows for doctors to make more informed decisions in patient care for future individuals. We specifically want to look at what impacts tobacco use, alcohol consumption, prenatal care, and parental age have on birth weight. To examine how these various factors influence birth weight, we will be using the *bwght2* from the Wooldridge package in R, which contains data on the birth weight of 1832 infants as well as other relevant data for our examination. We then ran several regressions using variables present in the dataset in order to determine which ones were statistically significant. The variables that were determined to be significant were cigarettes per day, number of prenatal care visits, time of prenatal care beginning, and paternal age. Our study found that for every cigarette smoked per day, on average, infant birth weight decreased by 11.3 grams, while each prenatal care visits were correlated with an increased in birth weight of 17 grams per visit. While the month of prenatal care beginning only being significant at the $\alpha = 0.1$ level, the month prenatal care began is very closely tied to total number of visits, with some who began their visit late still being able to obtain the adequate number of care visits.

2 Introduction

The weight of newborn infants is one of the most crucial indicators into how a child will grow and mature physically. A child born between 2.5 kg and 4.5 kg is considered to be of healthy weight, with the World health Organization defining a low weight birth, or LBW, as an infant born weighing less than 2.5 kg [5]. Infants who are born with lower birth weights are more prone to complications in early stages of development. This includes, but is not limited to, trouble gaining weight, inability to fight infections, and inconsistent body temperatures, usually trending on the low side [5]. This leads to low birth weight infants requiring a higher amount of care than those born with a normal birth weight, around 8 pounds or 3.6 kg. This paper will only examine LBW infants in single births, removing the possibility of low birth weight due to multiple fetuses during gestation.

The National Institute of Child Health and Human Development Neonatal Intensive Care Network ran a study in 1991 in which infant mortality rates were calculated for low birth weight children in their care. The study defines surviving as being alive for at least the first 59 days after birth, and found an average lifetime of 15 days for LBW infants who did not survive [3]. The study found that infants born under 750 g had a survival rate of 34 percent, 751 g - 1000 g had a 66 percent survival rate, 1001 g - 1250 g had a 87 percent survival rate, and those with birth weights between 1251 g and 1500 g had a survival rate of 93 percent [3]. This study establishes the importance of birth weight on infant survival rates, and shows a clear connection between increased birth weight and increased chance of surviving. With at least 59 days spent in the hospital for surviving LBW infants, the intensive care needed for these low birth weight infants creates an increased strain not only on the medical system, but also on the parents of these low weight infants. This creates the need to study what causes these low weight births, and how the frequency of them be reduced.

Beginning in the early 20th century, the United States has been studying the links between different characteristics of mothers and how they impact newborns. In 1952, Dugald Baird found a link between maternal height and viability of the fetus, where women who were taller and in good physical condition had a significantly higher rate of survival for their child as compared to children with smaller mothers [2]. Specifically, Baird found that for each additional inch increased in height of the mother, if physical fitness levels are held constant, the chances of an infants surviving increase around 5 percent per inch, with women under 5 foot 1 inch having the highest rate of infant mortality [2]. From this study, scientists in Great Britain were able to draw connections between height and social class, and attempted to draw a connection between these two variables and fetal birth weight [4]. These studies laid the ground work for further inquisition into the obstetrics, or the study of child birth, and raised further questions into what affects child birth.

Studies between cigarette smoke and low birth weight infants are fairly well documented, with decreases in smoking during pregnancy being linked to increases in birth weight. In a 1984 clinical trial conducted by Mary Sexton and Richard Hebel, 935 pregnant smokers were separated into two groups [9]. The control was allowed to keep smoking at the rate they were previously, while the experimental group underwent intervention treatment to help them quit smoking [9]. It was found that in the group which quit smoking, there was an average increase of 92 g per infant [9]. This result was found to be significant in that the change in birth weight is almost exclusively associated with the mother's decision to quit smoking, and not the gestational age of the infant. These results are important in that they show a direct connection between smoking during pregnancy and decreased birth weight, however further prenatal care is not specified in the article.

It is important to note that when this data was collected, 1988, the proportion of that adult population that smoked at some point was measured at 51 percent and those that currently smoked measured at 28.1 percent by the CDC [7]. This varies from the 14 percent we see reported by the CDC today, a drop of almost half just in current smokers [11]. The difference in how smoking was perceived in the 1980's as opposed to today allowed for more studies where pregnant women were allowed and in some cases encouraged to smoke. This leads to a large amount of data from before the 1990's since the adverse effects of smoking on pregnancy were still a relatively new concept.

The consumption of alcohol during pregnancy has also been correlated with a decreased birth weight. consuming one ounce of absolute alcohol in early pregnancy leads to a decrease in birth weight of 91 grams, while consuming one ounce in late pregnancy lead to an even greater decrease of 160 grams [6]. The study also corrected for the association between drinking and heavy tobacco use by using a sample where alcohol and tobacco use were not significantly correlated [6]. This allows for information to be drawn relating specifically to alcohol consumption without tobacco use influencing the infant birth weight.

There are also factors which can positively influence birth weight, as opposed to lowering it as tobacco and alcohol do. Prenatal care is an extremely important part of modern obstetrics, with both number of visits and how early care was administered being crucial to child development [1]. In a meta-analysis conducted by Greg Alexander and Carol Korenbart on studies relating to prenatal care, they found that, on average, women had around 10 14 prenatal visits, usually beginning in the first trimester [1]. They also point out that prenatal care centers typically emphasize quitting smoking as the most important thing during pregnancy, with most centers having this at the top of their list for prenatal health [1].

The evidence that prenatal care visits are important to fetal development is seen in more than just this study. In an examination of births in two California counties in 1978 by Jonathan Showstack, there was as strong association between number of prenatal visits and infant birth weight [10]. The study follows the Institute of Medicine's definition of adequate prenatal care which is set out as care beginning the in first trimester and having at least nine care visits before birth, and inadequate being no care until the third trimester or under four care visits [10]. Showstack found that on average, those infants who received adequate care weighed over 200 grams more than those who received inadequate care [10].

Variables that are outside of medical control also play a part in the weight of new born babies, such as the age of the father. in a study conducted by Nancy Reichman in the American Journal of Public Health, Reichman found that fathers over the age of 35 were 90 percent more likely to have a low birth child than someone ages 18-34 years old [8]. This study identified father's age as an independent risk factor for infant birth weights, placing paternal age as factor doctors need to take into consideration when caring for pregnant mothers.

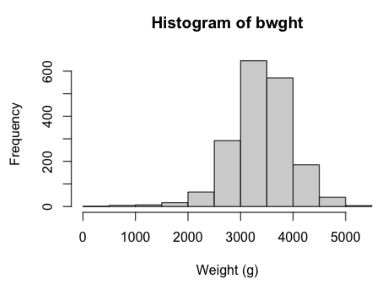
Our study is distinct in that we aim to not only observe how smoking, alcohol, prenatal care, and father's age independently affect the birth weight of infants, but also observe how all these factors together influence infant birth weight. From initial analysis of the data, we will be able to determine if any of the variables are statistically insignificant, and remove them from our analysis. This will give us a clearer picture as to how multiple variables relating to pregnancy affect an infant's birth weight.

3 Model

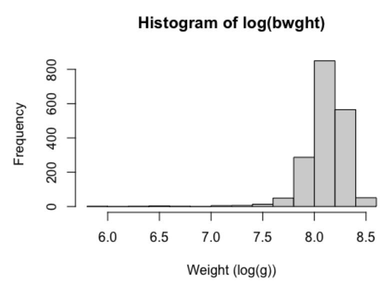
We are using the data set, *bwght2* from the Wooldridge package. This dataset was obtained from Dr. Zhehui Luo, from Michigan State University. She got them from state files linking birth and death certificates, and from the National Center for Health Statistics Natality and Mortality Data. The data includes 1832 observations on 23 variables. After our initial research, we decided that we wanted to look at factors that impact birth weight. The data set includes several variables of interest, such as: mother's age, number of prenatal visits, smoking habits, drinking habits - among others. These were the factors that stood out to us, intuitively, as they would seem to be the most viable in terms of affecting birth weight. Our initial regression ended up including four independent variables.*cigs* represents the average number of cigarettes smoked per day while pregnant, *drink* represents the average number of drinks per week while pregnant, *npvis* is the total number of prenatal visits, and lastly, monpre is the month in which prenatal care began. Note that *monpre* is not the specific month (i.e., December), but rather the amount of months that had passed since becoming pregnant, before beginning prenatal care. Additionally, fage and *mage* represent mother's age and father's age, respectively. We generated the following initial model below:

$$bwght = \hat{\beta_0} + \hat{\beta_1} cigs + \hat{\beta_2} drink + \hat{\beta_3} npvis + \hat{\beta_4} monpre + \hat{\beta_5} fage + \hat{\beta_6} mage \quad (1)$$

We will be running a level-level model. When deciding the type of model to use, we needed to see if our birth weight data was skewed. In order to compare if the log-level model was better fitting, we tested both level-level and log-level. As shown below, the distribution of birth weight is more normally distributed under a Level-Level model, and therefore we confirmed that our model is correct.



(a) Level-Level Model



(b) Log-Level Model

Figure 1: Distribution of Birth Weight

3.1 Selection and Verification of Regressors

After verifying that a level-level model would best fit our regression, we must now determine the variables that will best fit our model. After running several regressions with various regressors, we were able to narrow our model down.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3011.425    106.677  28.229  < 2e-16 ***
cigs         -11.063      3.331  -3.321 0.000916 ***
drink        -23.056     48.551  -0.475 0.634932    
npvis         17.270      3.884   4.446 9.26e-06 ***
monpre        19.283     11.555   1.669 0.095313 .  
fage           9.236      3.267   2.827 0.004748 ** 
mage          -4.540      3.958  -1.147 0.251482    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 571 on 1825 degrees of freedom
Multiple R-squared:  0.02235,	Adjusted R-squared:  0.01914 
F-statistic: 6.953 on 6 and 1825 DF,  p-value: 2.531e-07
```

Figure 2: Initial Regression Summary

3.1.1 F-Test

After running our initial regression, we wanted to ensure that we could say they have an effect on birth weight. We decided that we would isolate *cigs*, and test whether the other regressors have joint significance.

H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_1: H_0 is incorrect

After obtaining our F-statistic for this hypothesis, we found the F-stat to be equal to **5.864**, which is greater than the critical value of **1.850**. Therefore we will reject H_0, signifying that at least one of *drink*, *npvis*, monpre, *fage*, and *mage* have a partial effect on birth weight. This is not overly important, as this could mean that one and only one of our regressors has a statistically significant effect on birth weight, or it could mean that all five have an effect. Because of this, we will now test each variable to see if they have an effect.

3.1.2 P-Value and Critical Value Approach

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0$

for j = 1, 2, 3, 4, 5, 6

We will test these hypothesis using both the p-value and the critical value approach. Using the p-value approach, we must run our original regression, and compare the p-values attained to our significance level. As can be seen above in Figure 2, both drinks per week and mother's age are not statistically significant at a significance level 5%. While, month prenatal care began is not significant at 5% significance level, it does still have significance at 10% level. Figure 2 also shows us that cigarettes smoked per day, total number of prenatal visits, along with father's age are statistically significant at a significance level of 5%. Since we want to only include variables that are significant, we will use a 5% significance level. This means we will reject both H_1, H_3, and H_5 as they all have p-values less than 0.05. We will fail to reject H_2, H_4, and H_6 as their p-values are greater than significance level 0.05.

Now we will look at the critical value approach to confirm our results from the p-value approach. In order to do this, we will compare the t-value for each regressor to the critical value for 5% significance, which is equal to 1.645. Again, with reference to Figure 2 above, we can see that *drink* and mage are not statistically significant.

At this point, we are considering dropping the drinks per week and mother's age, as neither is statistically significant, nor are they relatively close to being significant. Because of this, we wanted to make sure our model only considered variables that have an impact on birth weight, so we have dropped. This helped us narrow down the variables we felt were most important, and we ended up having our final model:

$$bwght = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 npvis + \hat{\beta}_3 monpre + \hat{\beta}_4 fage + u \tag{2}$$

3.2 Gauss-Markov Assumptions

This subsection will verify that our OLS multiple regression model is not biased, and is therefore valid. In order to do this, we must validate the six Gauss-Markov assumptions of multiple regression. We conclude that our OLS multiple regression model satisfies all six Gauss-Markov assumptions and therefore does not exhibit bias.

In order for our regression to be valid, we must first ensure we are not violating any of the six Gauss-Markov assumptions. In this subsection, we will go through each assumption and ensure it satisfies each assumption for multiple regression, which will show us the data is not biased.

3.2.1 MLR 1 - Linear In Parameters

Because our regression model does not include any quadratic terms in the regressors, we have not violated the first Gauss-Markov Assumption.

3.2.2 MLR 2 - Random Sampling

We used birth weight data from Jeffrey Woolridge's, "Introduction to Econometrics". More specifically, we drew from Dr. Zhehui Luo, a recent Michigan State PhD. The data set was collected from both state and national natality and motility data. Considering our data set, the probability that any infant birth from the population being selected as part of the sample is the same as any individual birth in the population, thus we have satisfied the second Gauss-Markov assumption.

3.2.3 MLR 3 - No Perfect Collinearity

We found that there is no perfect collinearity between our regressors. The table below will show the correlation between each variable. Generally, if the R-squared value is greater than or equal to 0.9, we will observe multicollinearity. For our case, we can see in the table below that there is not perfect collinearity between our regressors.

	cigs	*npvis*	*monpre*	*fage*
cigs	1.000	-0.037	0.096	-0.032
npvis	-0.037	1.000	-0.300	0.052
monpre	0.096	-0.300	1.000	-0.100
fage	-0.032	0.052	-0.100	1.000

Table 1: MLR 3 - Correlation Between Regressors

We can also use the Variance Inflation Factor, or VIF, to detect multicollinearity in our regression analysis. The VIF test quantifies how much the variance is inflated, which contributes directly to multicollinearity. A variance inflation factor exists for each of the predictors in a multiple regression model. Generally, if the VIF value is greater than or equal to ten, we will say that we face multicollinearity. In our case, the highest VIF value is 1.116 on the *monpre* variable, and since 10 is greater than 1.116, we have confirmed again that we do not need to worry about multicollinearity.

	cigs	npvis	*monpre*	*fage*
VIF	1.010	1.010	1.116	1.101

Table 2: MLR 3 - Variance Inflation Factor of Regressors

3.2.4 MLR 4 - Zero Conditional Mean

Under MLR 4, the error term, u, has an expected value of zero given any values of the independent, or explanatory variables. Because we have confirmed we are using a level-level model, we can assume this satisfies.

3.2.5 MLR 5 & 6 - Homoskedasticity the Normality Assumption

Under the fifth Gauss-Markov Assumption, the error term, u, has the same variance given any values of the explanatory variables, x_i. Referring to Figure 3, below, we have plotted the residual values against the predicted, or fitted, values.

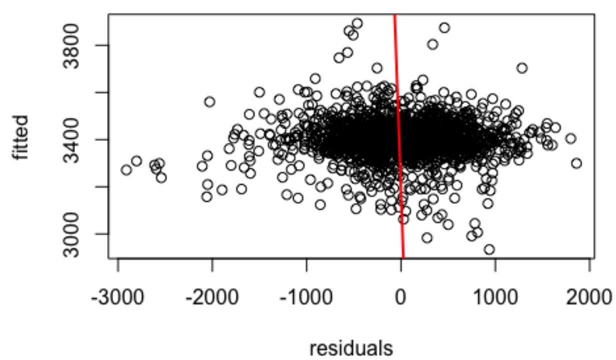


Figure 3: Residuals vs. Fitted Values

As for the Normality Assumption, it states the population error, u, is independent of the explanatory variables $x_1, x_2, ..., x_k$ and is normally distributed with zero mean and variance. This is demonstrated in Figure 3, above, as the residuals have a mean that is very close to zero.

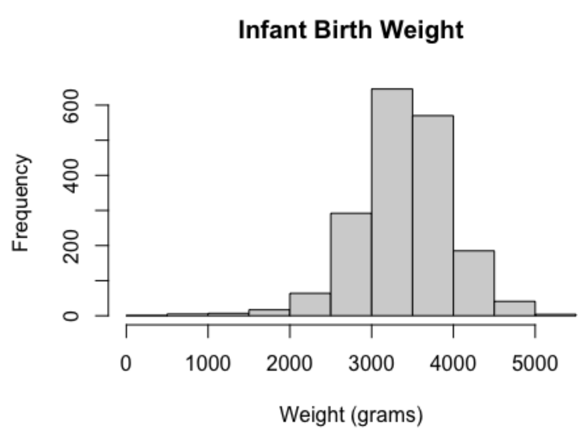


Figure 4: Distribution of Infant Birth Weight

4 Findings

In this section, we will analyze the effects that each independent variable has on infant birth weight.

```
Residuals:
     Min       1Q   Median       3Q      Max
-2911.20  -319.85    20.64   353.03  1860.65

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2959.901     96.164  30.780  < 2e-16 ***
cigs         -11.234      3.276  -3.430 0.000618 ***
npvis         16.973      3.874   4.381 1.25e-05 ***
monpre        21.136     11.445   1.847 0.064951 .
fage           6.622      2.352   2.815 0.004925 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 570.9 on 1827 degrees of freedom
Multiple R-squared:  0.02153,   Adjusted R-squared:  0.01938
F-statistic: 10.05 on 4 and 1827 DF,  p-value: 4.81e-08
```

Figure 5: Final Regression Summary

As can be seen above, in Figure 4, *cigs* is the only variable that has a negative effect on birth weight. Our model estimated $\hat{\beta}_1$ = -11.234. This result was expected, based on our initial research. Based on the results, for an increase of one in average cigarettes smoked per day, we expect birth weight to decrease by about 11 grams, ceteris peribus. This is quite significant, considering that among those who smoked during pregnancy, they average more than 13 cigarettes per day. This is key when considering advising pregnant women and promoting public health. It has become increasingly known that cigarettes are not advisable for women while pregnant. Additionally, we have observed a fall in overall cigarette consumption in our country, which helps us maintain a healthier public and helps lead to more healthy infant birth weight.

Something in the data that stood out was the rate at which pregnant women with low birth weights choose not to answer certain questions in data collection. In our sample, women with low infant birth weight did not answer how much tobacco and alcohol they consumed at twice the rate of women without low birth weights. We believe this is due to a multitude of psychological factors related to child birth. These factors could be a mother believing that her child's low weight was caused by their use of alcohol and tobacco, and therefore choosing not to report how much of these substances they used, although we cannot be certain.

We can also see that both total number of prenatal visits and the month in which prenatal care began have the largest effects on birth weight. Our model estimates $\hat{\beta_2} = 16.973$, meaning that, all else equal, for each additional prenatal visit, we expect to see an increase in birth weight by almost 17 grams. This is quite significant and shows how vital it is for pregnant women to get prenatal care.

Additionally, we estimate $\hat{\beta_3} = 21.136$, which implies that we expect an increase of 21.136 grams for each additional month before birth prenatal care began, holding all else equal. This is another metric to help us demonstrate how important prenatal care is, and how it can have a rather substantial impact on whether we observe low birth weights or not. It is known that prenatal care is prudent for infants, and these metrics both back that claim up.

Lastly, our model predicts that $\hat{\beta_4} = 6.622$, so for each additional year older the father is, we expect the birth weight to increase by about 6.6 grams, ceteris peribus. Initially, we included paternal age because of our background research indicating that it has been found to have significance. We were surprised upon finding how the paternal age statistically significant, while maternal age was not. We found in our background research that when father's age is greater than 35, there is a 90% chance of having a low birth weight infant[8]. This disagrees with our findings which actually show an increase in birth weight of 6.6 grams for each additional year of age of the father. Going through the data, we also found that there is an average difference of 2.36 years between paternal and maternal age. We did test a new variable, *ageDiff*, to see if it was statistically significant, but did not find that to be the case.

5 Conclusion

Throughout this paper, we have set out explain the variation in infant birth weight. Initially, we looked at cigarettes smoked per day, alcoholic drinks per week, total number of prenatal visits, month in which prenatal care began, and both paternal and maternal age. After running our initial regression, we found drinks and maternal age not to be statistically significant, and so we decided to drop them from our model. Of course, when it comes to infant birth weight, there is a rather large intercept, considering there is a limit as to how much or how little an infant will weight at birth.

Future studies should include other factors relating to the mother's physical health and characteristics. It would be interesting to have data on mother's weight, as that is likely an indicator for infant birth weight. Other factors that are worth studying and may have influence are mother's height, as Baird found in his early work in obstetrics [2]. Examining the mother's diet would be a crucial piece of information into how healthy the child is born, however this information would be difficult to quantify. Most prenatal care centers do emphasize a proper diet, however we believe that having actual information into macro-nutrients, such as quantity of proteins and carbohydrates, would yield a great deal of information into infant birth weights [1].

References

[1] Greg Alexander. "The Role of Prenatal Care in Preventing Low Birth Weight". In: *The Future of Children* (1995), pp. 103–120.

[2] Dugald Baird. "Preventative Medicine in Obstetrics". In: *New England Journal of Medicine* (1952), pp. 679–686.

[3] Maureen Hack. "Very Low Birth Weight Outcomes of the National Institute of Child Health and Human Development Neonatal Network". In: *Pediatrics* 87 (5 1991), pp. 587–597.

[4] John C. Kleinman. "THE EFFECTS OF MATERNAL SMOKING, PHYSICAL STATURE, AND EDUCATIONAL ATTAINMENT ON THE INCIDENCE OF LOW BIRTH WEIGHT". In: *American Journal of Epidemiology* 125 (1985), pp. 843–855.

[5] M.S. Kramer. "Determinants of low birth weight: methodological assessment and meta-analysis". In: *Bulletin of the World Health Organization* (1987), pp. 663–737.

[6] Ruth E. Little. "Moderate Alcohol Use during Pregnancy And Decreased Infant Birth Weight". In: *American Journal of Public Health* 67 (12 1977), pp. 1154–1156.

[7] Morbidity and Mortality Weekly Report Editors. "Cigarette Smoking Among Adults – United States, 1988". In: *Morbidity and Mortality Weekly Report* (1991).

[8] Nancy Reichman. "Paternal Age as a Risk Factor for Low Birthweight". In: *American Journal of Public Health* (2006).

[9] Mary Sexton. "A Clinical Trial of Change in Maternal Smoking and Its Effect on Birth Weight". In: *JAMA* (), pp. 911–915.

[10] Jonathan Showstack. "Factors Associated with Birthweight: An Exploration of the Roles of Prenatal Care and Length of Gestation". In: *American Journal of Public Health* (1984), pp. 1003–1008.

[11] Office of Smoking and Health Editors. *Current Cigarette Smoking Among Adults in the United States*. URL: `https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm`. (accessed: 28/11/2020).