

Project Work Statement

Sponsor

National Institutes of Health

Maternal Smoking and Infant Health

Participants

Yu Du, ydu10@jhu.edu

Potential Participants

Francis Collins, Francis.Collins@od.nih.gov

Sue Wu, wsue@od.nih.gov

Bill Smith, billsmith@od.nih.gov

Date: October 15, 2012

Any apparent association of this work to NIH is fictional one, and the sole purpose of this work is a class exercise

1 Background

The National Institutes of Health (NIH), part of the U.S. Department of Health and Human Services, is the nation's medical research agency—making important discoveries that improve health and save lives. Thanks in large part to NIH-funded medical research, Americans today are living longer and healthier. Life expectancy in the United States has jumped from 47 years in 1900 to 78 years as reported in 2009, and disability in people over age 65 has dropped dramatically in the past 3 decades. In recent years, nationwide rates of new diagnoses and deaths from all cancers combined have fallen significantly.

2 Problem Statement

One of the U.S. Surgeon Generals health warnings placed on the side panel of cigarette packages reads: “Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight.” Epidemiological studies [1] also indicate that smoking is responsible for a 150 to 250 gram reduction in birth weight and that smoking mothers are about twice as likely as nonsmoking mothers to have a low-birth-weight baby (under 2500 grams).

Since NIH is concerned about this fact, it thus sponsors the team to carry out this statistics experiment to determine the relationship between maternal smoking status and birth weight. Typically, smaller babies have lower survival rates than larger babies who are born at the same term. Therefore, birth weight is also a measure of baby health and maternal smoking status has a potential to influence the baby health. The objective NIH has in mind is to assess whether there is an impact of maternal smoking on birth weight and how if there is any impact. Furthermore, since birth weight is such an important variable, NIH is also interested in knowing what other variables influence the baby birth weight other than maternal smoking status that might include maternal height, weight, age, etc.

In this context, the variables that concern us are birth weight, maternal smoking status, maternal weight, height, age. These variables are what we are going to consider endogenous in this project because we are going to measure the relationship among those important variables. We come up with these variables by research and also with intuition to see whether these chosen predictors have a significant impact on the response variable birth weight NIH is concerned about. Variables other than those might also have an impact on the baby birth weight, like the environment of the hospital where mothers labored and so on. But we are not going to consider those variables in this context because we mainly want to know if variables related to mothers have an impact on the baby birth weight. Those other variables are not related to mothers directly therefore those variables

can be labeled as exogenous.

3 Approach

Data was collected from Child Health and Development Studies (CHDS)— a comprehensive investigation of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San FranciscoEast Bayarea [2].

Data Structure is as follows:

Variable	Description
bwt	Birth weight in ounces (999 unknown)

age	mother's age in years
height	mother's height in inches (99 unknown)
weight	Mother's prepregnancy weight in pounds (999 unknown)
smoke	Smoking status of mother 0=not now, 1=yes now, 9=unknown

Figure 1: Data Structure

Data Cleaning:

- Delete the data with missing values
- Delete the detected outliers by Boxplot

First thing to do in this project is to compare the birth weights of babies born to smokers and nonsmokers and see if there is a significant difference. For this task we simply focus on the variable birth weight and the variable Smoke Status. Therefore split the birth weight into two groups according to the maternal smoking status and then two distributions of birth weight for babies born to women who smoked during their pregnancy and for babies born to women who did not smoke during their pregnancy will be summarized numerically and graphically. Moreover, a two sample hypothesis test will be conducted to test if there is a significant difference in birth weights for these two groups as well as if non-smoking mothers' babies tend to have a larger weight.

Second, focus on the whole dataset that contains the variable birth weight, maternal smoking status, maternal height, weight and age. Do multiple covariates regression analysis for birth weight (response) against other variables including maternal smoking status, maternal height, weight, age as predictors to measure the relationship among those variables.

Original Model:

$$\begin{aligned} \text{Birth Weight} = & \beta_0 + \beta_1(\text{Maternal Smoking Status}) + \beta_2(\text{Maternal Height}) \\ & + \beta_3(\text{Maternal Weight}) + \beta_4(\text{Maternal Age}) + \epsilon(\text{error term}) \end{aligned}$$

Generate residual plots to assess whether the residual assumptions are violated. Also use VIF score to identify the collinearity if any. Then use step function in `r` to backward select the best model with the least AIC score, an evaluation score for the model that balances the bias and variances. When we have the final model, based on this final model, the prediction of the birth weight given the values of other variables can be made. Also the way those predictors influence the birth weight will also be known.

4 Milestones

Major Deadlines:

- Work Statement due date, Sep 28, 2012,
- Midterm Presentation due date, Oct 12, 2012,
- Progress Report due date, Oct 26, 2012,
- Final Presentation due date, Nov 6, 2012,
- Final Report due date, Nov 30, 2012.

5 Deliverable

5.1 From Team to Sponsor

The following outputs are expected from this project:

- A report regarding whether maternal smoking status has an impact on baby birth weight
- A software that produces the prediction interval of baby birth weight given the values of the predictors as the input

5.2 From Sponsor to Team

In order for our project to be of successful one, we will need:

- Access to the datasets of Child Health and Development Studies where the data regarding birth weight, maternal smoking status, maternal height, weight and age are provided,
- Computing resources,
- Timely responses to inquiries,
- Symposium attendance travel expenses.

References

- [1] I. Merkatz and J. Thompson. *New Perspectives on Prenatal Care*. Elsevier, New York, 1990.
- [2] J. Yerushalmy. The relationship of parents cigarette smoking to outcome of pregnancy implications as to the problem of inferring causation from observed associations. *Am. J. Epidemiol*, 93:443–456, 1971.