Lab 4: Does Prenatal Care Improve Infant Health?

w203: Statistics for Data Science

April 10, 2017

Introduction

This is a group lab. You may work in teams of 2 or 3.

The file bwght_w203.RData contains data from the National Center for Health Statistics and from birth certificates. Your team has been hired by a health advocacy group to study this data and help them understand whether prenatal care improves health outcomes for newborn infants.

The file includes a birthweight variable. Additonally, the one- and five-minute APGAR scores are included. These are measures of the well being of infants just after birth.

Variable descriptions are provided as follows.

```
load("bwght_w203.RData")
desc
```

```
##    variable                            label
## 1     mage               mother's age, years
## 2    meduc              mother's educ, years
## 3   monpre         month prenatal care began
## 4    npvis  total number of prenatal visits
## 5     fage               father's age, years
## 6    feduc              father's educ, years
## 7    bwght               birth weight, grams
## 8    omaps           one minute apgar score
## 9    fmaps          five minute apgar score
## 10    cigs           avg cigarettes per day
## 11   drink              avg drinks per week
## 12     lbw              =1 if bwght <= 2000
## 13    vlbw              =1 if bwght <= 1500
## 14    male                  =1 if baby male
## 15   mwhte               =1 if mother white
## 16   mblck               =1 if mother black
## 17    moth           =1 if mother is other
## 18   fwhte              =1 if father white
## 19   fblck              =1 if father black
## 20    foth           =1 if father is other
## 21  lbwght                       log(bwght)
## 22  magesq                           mage^2
## 23 npvissq                          npvis^2
```

Assignment

Prepare a report addressing the question of whether prenatal care improves newborn health outcomes.

A successful submission will include

1. A brief introduction

2. A model building process, supported by exploratory analysis. Your EDA should be interspersed with, and support, your modeling decisions. In particular, you should use exploratory techniques to address

- What transformations to apply to variables and what new variables should be created.
- What variables should be included in each model
- Whether model assumptions are met

3. A minimum of three model specifications. In particular, you should include

- One model with only the explanatory variables of key interest.
- One model that includes only covariates that you believe increase the accuracy of your results without introducing bias.
- One model that includes the previous covariates, but also covariates that may be problematic for one reason or another.

4. For your first model, a detailed assessment of the 6 CLM assumptions. For additional models, you should check all assumptions, but only highlight major differences from your first model in your report.
5. A well-formatted regression table summarizing your model results. Make sure that standard errors presented in this table are valid. Also be sure to comment on both statistical and practical significance.
6. A discussion of whether your results can be interpretted causally. In particular, include a discussion of what variables are not included in your analysis and the likely direction of omitted variable bias. Also include a discussion of which included variables may bias your results by absorbing some of the causal effect of prenatal care.
7. A brief conclusion with a few high-level takeaways.

Please limit all submissions to 30 pages. Be sure to turn in both your pdf report and also your source code.