Preliminary Analysis

Bradley Hopkins, Matthew Hoctor

2/6/2021

Contents

Getting Started

Read carefully project outline document to see what you need to do with your project.

Read all support documents to understand what you can do with your project.

Look at the dataset to see what are available (e.g., info/variables)

- Bwt; birth weight
- Gestwks; gestational age in weeks
- Age; mother's age
- Mnocig; number of cigarettes/day smoked by the mother
- Mheight; maternal weight
- Mppweight; pre-partum weight
- New variables to consider
  - Maternal BMI & BMI category
  - Smoking category

Literature Review

Do an intensive online search to find/gain knowledge about some similar/related topics through some studies. Use the resources of literature search and readings as your scientific support for your project, more particularly for your aims.

ACOG Tobacco Cessation Guidelines

> Pregnant women should be advised of the **significant perinatal risks** associated with tobacco use, including orofacial clefts, **fetal growth restriction**, placenta previa, abruptio placentae, **preterm prelabor rupture of membranes**, **low birth weight**, increased perinatal mortality, ectopic pregnancy, and decreased maternal thyroid function. Children born to women who smoke during pregnancy are at an increased risk of respiratory infections, asthma, infantile colic, bone fractures, and childhood obesity. Pregnancy influences many women to stop smoking, and approximately 54% of women who smoke before pregnancy quit smoking directly before or during pregnancy. Smoking cessation at any point in gestation benefits the pregnant woman and her fetus. The greatest benefit is observed with cessation before 15 weeks of gestation. Although cigarettes are the most commonly used tobacco product in pregnancy, alternative forms of tobacco use, such as e-cigarettes or vaping products, hookahs, and cigars, are increasingly common. Clinicians should advise cessation of tobacco products used in any form and provide motivational feedback. Although counseling and pregnancy-specific materials are effective cessation aids for many pregnant women, some women continue to use tobacco products. Clinicians should individualize care by offering psychosocial, behavioral, and pharmacotherapy interventions. Available cessation-aid services and resources, including digital resources, should be discussed and documented regularly at prenatal and postpartum follow-up visits.

https://www.acog.org/en/Clinical/Clinical%20Guidance/Committee%20Opinion/Articles/2020/05/Tobacco%20and%20Nicotine%20Cessation%20During%20Pregnancy

Commentaries on 2014 Int. J. Epidemiol. Republication of original article

https://pubmed.ncbi.nlm.nih.gov/25301860/

Commentary: Competing Explanations of the "Birthweight Paradox"

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4190521/

Conditioning on an intermediate " If we consider the low birthweight infants whose mothers do not smoke, then we know maternal smoking is ruled out as a cause for low birthweight, so that there must have been some other cause, possibly something such as malnutrition or a birth defect, theconsequences of which for infant mortality are much worse.2 By not controlling for the common causes (U) of low birthweight and infant mortality, we are essentially setting up an unfair comparison between the smoking and non-smoking mothers."

Heterogeneity of low birthweight phenotype (can be considered a special case of conditioning on an intermediate)

Jacob Yerushalmy on self-selection and the pitfalls of causal inference https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4200064/

526: Unraveling the racial disparity in preterm birth: experiences of discrimination are associated with preterm birth and low birth weight in black women

https://www.ajog.org/article/S0002-9378(10)01804-1/fulltext (The study in the sample paper uses data that involves 10,000 White women and 3,000 Black women. However, Black women have been noted to have worse birth outcomes than White women even without consideration for variables such as smoking status and there is likely a large cluster of different factors contributing to this.)

Project Aims and Hypothesis

Shape your project aims and hypotheses based on reading literature, your knowledge and skills. Make sure your aims tightly connected to the project goal (i.e., association model building, and prediction model building.)

Test hypotheses explaining 'Birthweight paradox' See notes on supporting literature below Run some descriptive analysis to study distribution of the data for each variable; some exploratory analysis, see the project outline for details. Describe each variable Are data transformations necessary? (gladder function) Scatter plot matrix Check normality of response data Check independence assumption Consider categorization of variables, e.g. smoking Supporting literature

Descriptive and Exploratory Analysis

Run some descriptive analysis to study distribution of the data for each variable; some exploratory analysis, see the project outline for details.

Import the dataset

Import the maternal smoking dataset:

```
CHDS <- read.csv("CHDS.csv")
```

Descriptions of individual variables

Birth Weight

Gestational age

Maternal Age

Cigarettes smoked

Maternal height

Maternal pre-partum weight

Generation of new variables of interest

BMI

The following R Code creates a new variable for body mass index from the maternal weight and pre-partum height data:

```
CHDS$BMI <- 703 * CHDS$mppwt / ((CHDS$mheight)^2)
```

BMI category

The following R code creates a new variable for BMI category:

```
CHDS$BMI_cat <-
  ifelse(CHDS$BMI < 18.5, 0,
  ifelse(CHDS$BMI < 25, 1,
  ifelse(CHDS$BMI < 30, 2,
  ifelse(CHDS$BMI < 35, 3,
  ifelse(CHDS$BMI < 40, 4,
  5)))))
```

Note that 0 corresponds to underweight, 1 to normal weight, 2 to overweight, 3 to class I obese, 4 to class II obese, and 5 to class III obese.

Smoking category

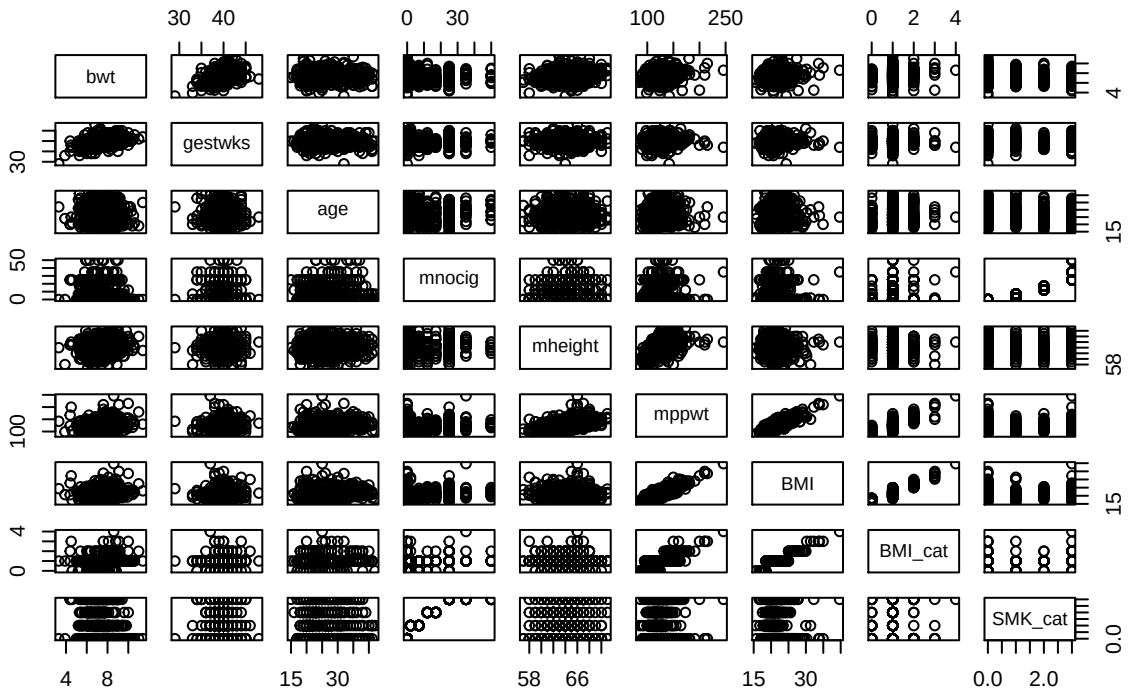
The following R code creates a new variable for smoking category. Patients are categorized as 0, non-smokers; 1, light smokers 1-9 cigarettes/day; 2, moderate smokers, 10-19 cigarettes/day; or 3, heavy smokers 20+ cigarettes/day:

```
CHDS$SMK_cat <-
  ifelse(CHDS$mnocig == 0, 0,
  ifelse(CHDS$mnocig < 10, 1,
  ifelse(CHDS$mnocig < 20, 2,
  3)))
```

Scatter plot matrix

The following R code creates a scatter plot matrix:

```
pairs(~bwt+gestwks+age+mnocig+mheight+mppwt+BMI+BMI_cat+SMK_cat, data = CHDS, main = "Scatterplot Matri
```



Check normality of response data

Birth Weight

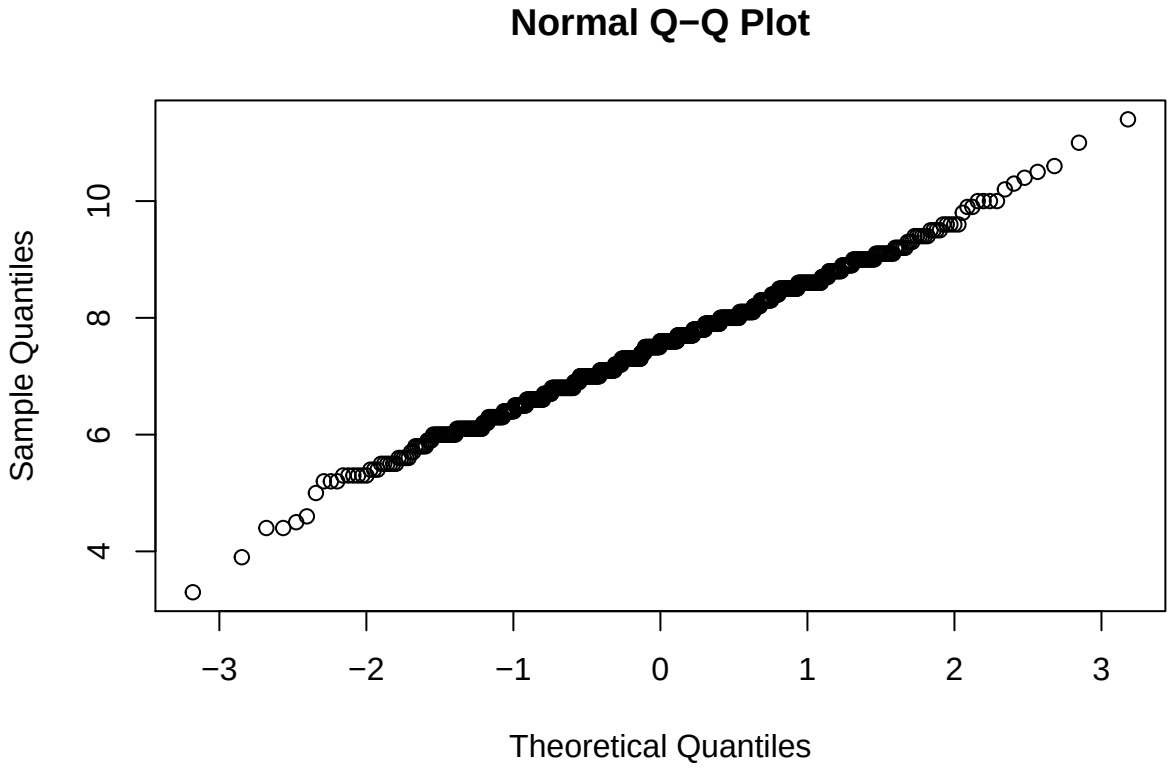
We can perform the Shapiro-Wilk test to to test for normality:

```
shapiro.test(CHDS$bwt)
```

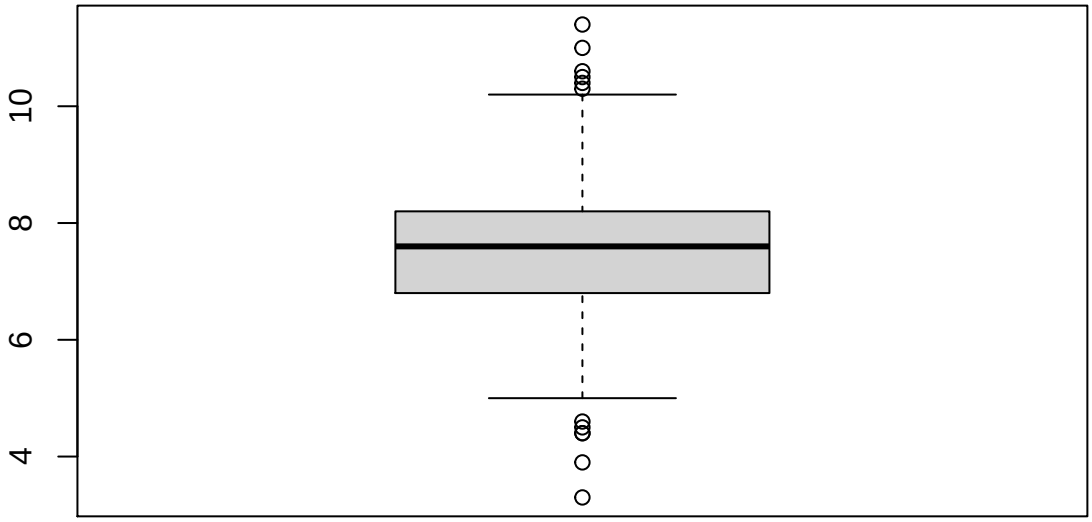
```
##
##  Shapiro-Wilk normality test
##
## data:  CHDS$bwt
## W = 0.99645, p-value = 0.133
```

This borderline p-value suggests that more investigation may be needed. A Q-Q plot and boxplot may help:

```
qqnorm(CHDS$bwt)
```




```
boxplot(CHDS$bwt)
```



These results suggest that birth weight is normally distributed.

Gestational Age

We can perform the Shapiro-Wilk test to to test for normality:

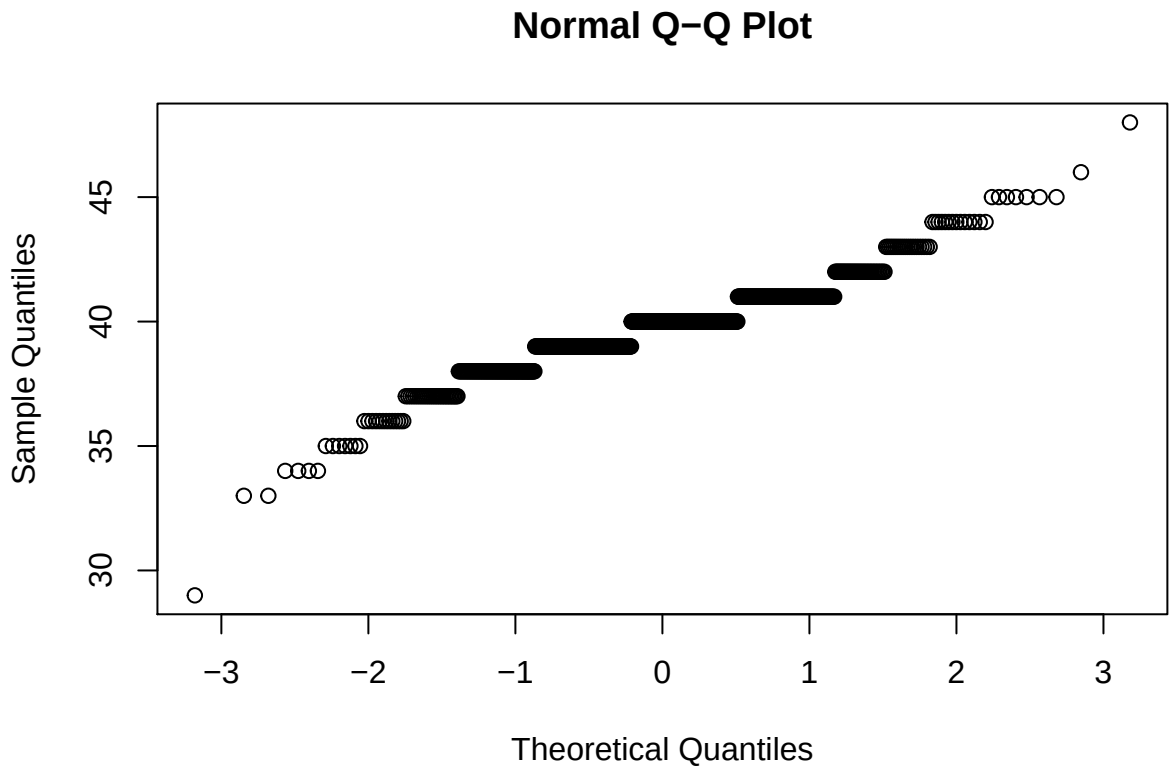
```
shapiro.test(CHDS$gestwks)
```

```
##
##  Shapiro-Wilk normality test
##
```

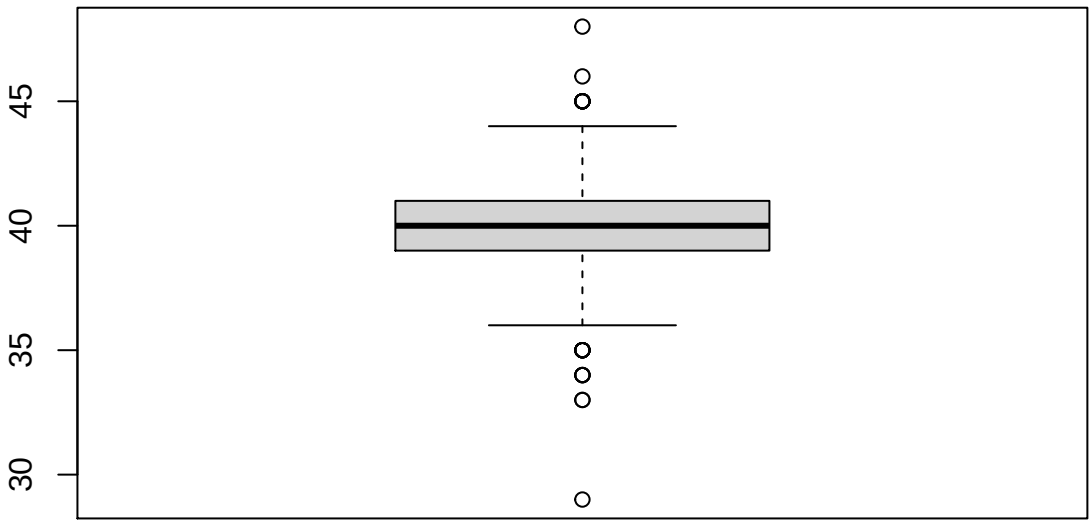
```
## data:  CHDS$gestwks
## W = 0.93902, p-value = 4.551e-16
```

This is very consistent with normal distribution. We will do a Q-Q plot and boxplot for completeness:

```
qqnorm(CHDS$gestwks)
```




```
boxplot(CHDS$gestwks)
```

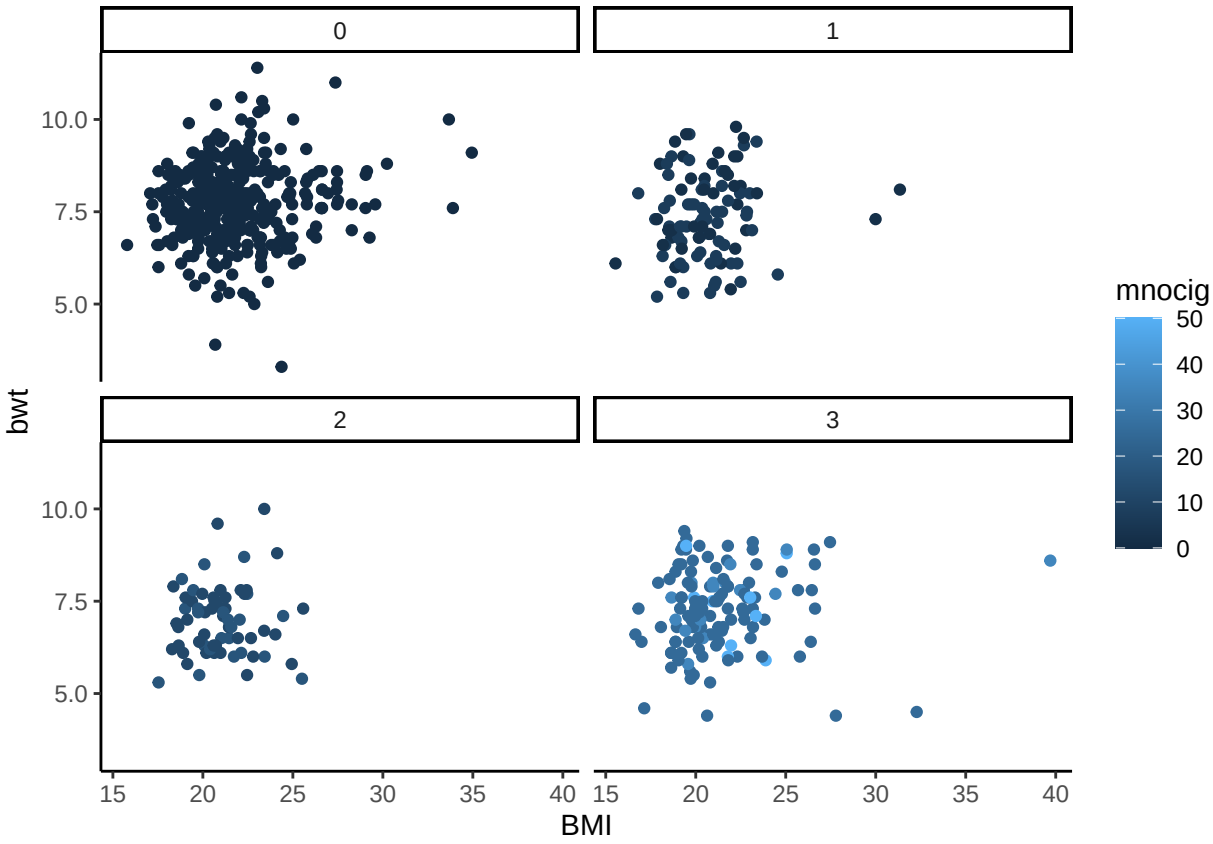


These results suggest that gestational age is normally distributed.

Exploratory Analysis

A preliminary exploratory analysis of the data is shown below, to provide a concept for further evaluation later on.

```
ggplot(CHDS) +
  geom_point(aes(x = BMI,
                 y = bwt,
                 color = mnocig)) +
  theme_classic() +
  facet_wrap( ~ SMK_cat)
```



References: