`Homework 7 Group Project`

Birth Weight Exercise Questions

<mark>Teams: Please duplicate this exercise file: "File - Make a Copy"</mark>

Due: 12/08/2020 5pm

Created on: 11/23/2020

In this data analysis project, you will examine various factors that may be associated with baby's birth weight, starting with descriptive analyses then followed by a set of regression analyses. Please use the [google sheet](#) to submit your work. (NYU NetID required to login)

Part One:

1. Conduct a descriptive analysis of this dataset follows [Sample Demographic Characteristic Table 1](#)
    a. Variables:
        i. Race: race of mother (1 = white, 2 = black, 3 = other)
        ii. GestWeek: gestational weeks (?)
        iii. Age35: whether mother is older than 35 at the time of child birth
        iv. Smoke: whether mother smokes
        v. Ed: mother's level of education
        vi. Hyperten: whether mother has hypertension
        vii. SecondBirth: whether the baby is first birth (secondbirth=0 if first birth)
        viii. GestDiabetes: whether mother has gestational diabetes
        ix. InfantSex: gender of the baby

x.   BMI: mother's BMI before pregnancy

    xi.   WeightGain: mother's weight gain during pregnancy

    xii.   FatherAge35: whether father is older than 35 years old

    xiii.   FatherBMI: fatherBMI

b.   Using the three categories of variable "Race" as the column heading for Table 1, and the rest of the variables in the above list as row headings. The last row is the outcome variable.

c.   If the row variable is categorical, report n and %; if the row variable is quantitative, report mean and SD. You may want to split the table into two blocks, the first block include categorical variables; and the bottom block include quantitative variables.

Example of Descriptive Table 1

Variables	Black	White	Other	All
Categorical	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Mother age > 35				
Smoke				
...				
Quantitative	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Father's BMI				
...				
Birth Weight (grams				

d.   Write a short paragraph summarizing the table. Which ones of these variables do you think may be related to baby birth weight (how much do you know about determinants of baby's birth weight? A quick wiki check can be helpful!)

2. Include a figure that illustrates the trend of birth weight over years and by race.
    a. You can make a line plot:
        i. Y: mean birth weight
        ii. X: year
        iii. Three lines each represents a racial group
    b. Alternatively, you can make a set of boxplots to display the distribution of birth weight for every race at each year in a cluster, then for each year across the x axis.
        i. Use code like `boxplot(Y ~ time + race)`
        ii. Note: you may need to adjust the example code to get to the desired pattern
3. Researchers are particularly interested in the racial differences in birth weights.
    a. Fit a regression using race as predictors to compare birth weights across different races.
    b. How does racial gap change over time? What regression model would you use to answer this question?
4. Create a subset of data for 2011. Conduct a more detailed analysis on the associations of various factors with birth weight.
    a. Start with the model that only includes race.
    b. Include the rest of the variables in the model (the "big" model).
        i. Do we need to include "year" to analyze this subset of data? Why or why not?
        ii. Pay attention to variable types!
    c. Create a table reporting the regression results (follow the style of Sample Regression Table 3 on APA style guide
    d. Compare the race only model with the "big" model, what differences do you notice?
        i. What happens to racial gap in birth weight after we control for other variables?

ii. Which other variables are significantly associated with Birth weight? Anything surprising?

iii. What's the overall model fit (R squares) of the "big" model? Is it much greater than the "small" model? Which variables contribute most to this change?

d. Do you think the "big" model is the best model that characterizes the relationship between birth weight and other factors? Why and why not? What are other ways that you would like to explore to improve the model fit and model interpretation?