MACS 30200 Problem Set 1 - Part 1
Dataset Assignment

Shuting Chen

April 16, 2018

1. Data accessibility

The 2016 U.S. Natality Data used for this analysis can be accessed through National Bureau of Economic Research (NBER)'s website (http://nber.org/data/vital-statistics-natality-data.html). It contains newborn's birth information, health conditions, and parents' demographic and medical data of all states in the United States. The data is part of the National Vital Statistics System and is curated by National Center for Health Statistics (NCHS). Under the management of the Vital Statistics Cooperative Program, the vital registration offices of all states, the District of Columbia, New York City, Puerto Rico, Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands provided the data to the NCHS. Therefore, it is originally stored on the NCHS website of Centers for Disease Control and Prevention (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm).

2. Cite key papers using this data

The data are widely used as primary source of birth data in several fields:

- Unintended pregnancy: Finer and Henshaw (2006), Finer and Zolna (2011)
- Diseases spread from mothers to children: Lanzieri et al. (2013), Bowen et al. (2015)
- Family planning policy: Dickert-Conlin and Chandra (1999)

3. Data collection

As mentioned in section 1, the 2016 Natality Data were provided by the vital registration offices of all states, the District of Columbia, New York City, Puerto Rico, Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands through the Vital Statistics Cooperative Program of NCHS. Nevertheless, prior to 1972, data are based on a 50-percent sample of birth certificates from all States. Beginning in 1972, data are based on a 100-percent sample of birth certificates from some states and on a 50-percent sample from the remaining States. In 1985, all states and the District of Columbia used 100-percent records of birth certificates.

4. Table of descriptive statistics

Table 1 presents descriptive statistics for selected variables of parents' demographics and newborn's health condition. Based on the interest of exploring potential relationships between parents' demographics and newborn's health condition, I use mother's age (mager), mother's education (meduc), father's age (fagecomb) and father's education (feduc) as primary demographic variables. According to the User Guide to the 2016 Natality Public Use File, mother's and father's education have been classified into eight levels, ranging from 8th grade or less (value of 1) to doctorate or professional degree (value of 8). Newborn's health condition is represented by period of gestation (combgest), birth weight (dbwt), 5-min APGAR score (apgar5) and 10-min APGAR score (apgar10).

Table 1: Descriptive statistics for parents' demographics and newborn's health condition

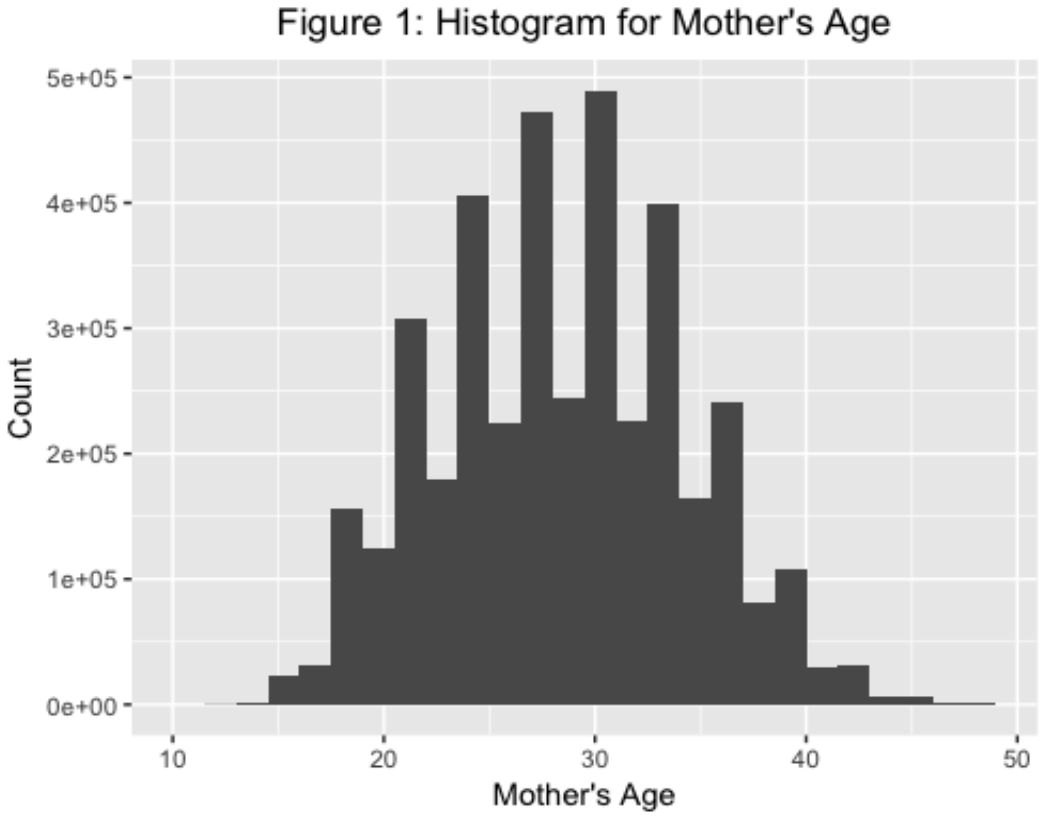
Variables selected for parents' demographics				
	Mother's Age	Mother's Education	Father's Age	Father's Education
Min	12	1	11	1
Max	50	8	98	8
Median	29	4	31	4
Mean	28.709	4.306	31.563	4.225
Std.dev	5.820	1.738	6.805	1.758

Variables selected for newborn's health condition				
	Period of Gestation	Birth Weight (g)	5-min APGAR Score	10-min APGAR Score
Min	17	227	0	0
Max	47	8165	10	10
Median	39	3310	9	7
Mean	38.634	3266.867	8.789	5.933
Std.dev	2.483	592.189	0.821	2.749

As shown in Table 1, the median age of mother and father are quite similar, which are 29 years old and 31 years old respectively. However, the distribution of father's age is more spread out with larger range (87) and larger standard deviation. Mother's and father's education basically share the similar summary statistics and the median level of education for both mother and father is some college credit, but not a degree (value of 4). For newborn's health condition, the average birth weight is 3266.867 grams and the standard deviation of birth weight is 592.189. The APGAR score is a measure of the need for resuscitation and a predictor of the infant's chances of surviving the first year of life. The higher APGAR score means that the infant is in a better health condition. By comparing last two columns, the distribution of 10-min APGAR score is more spread out with higher standard deviation of 2.749.

5. Visualization

Figure 1 visualizes the distribution of mother's age for all newborns in the United States. The most common ages for delivering a baby are around 28 and 31. This not only follows physiological law but also coincides with the fact that these ages are probably the transformative points for individuals' life and people are more likely to have a baby at these ages. Besides, Figure 1 shows that the occurrence of delivering a baby decreases dramatically when mother's age beyond 38. Presumably, this might be attributed to the higher risk of having a baby at advanced maternal age.

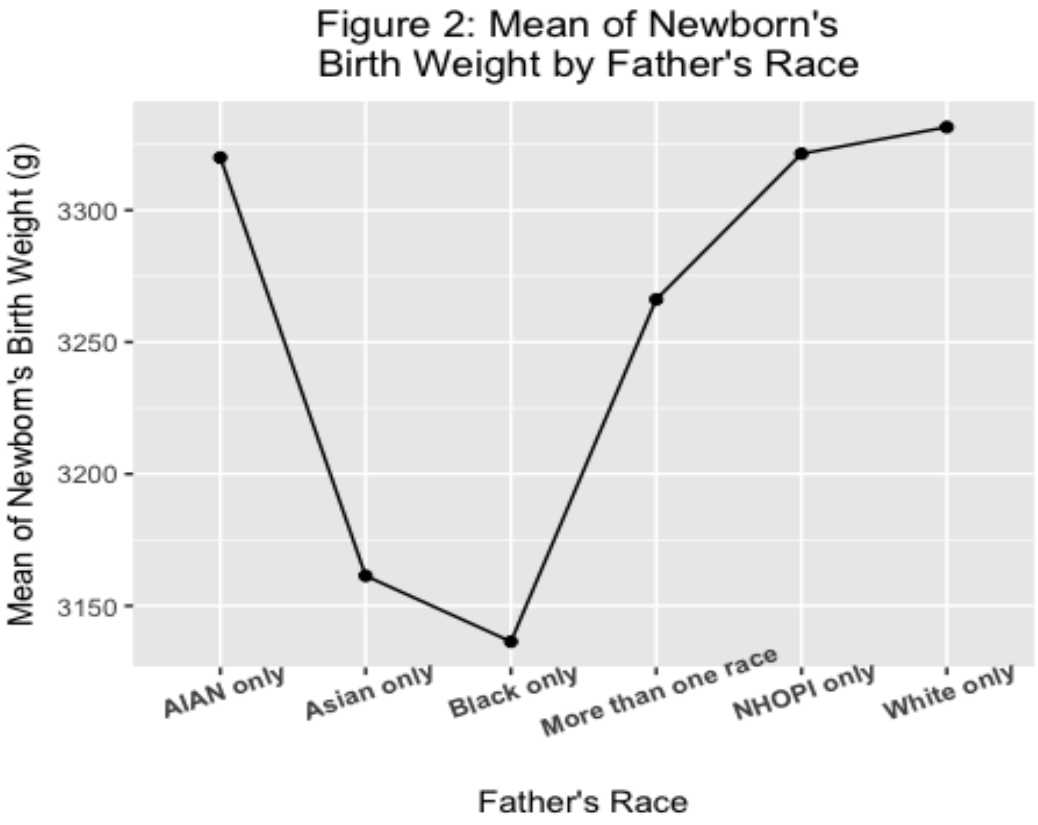


6. Conditional description of the data

Table 2 and Figure 2 display the differences in newborn's average birth weight by father's race. It can be easily recognized that the average newborn's birth weight remains at the similar level for white (only), AIAN (only) and NHOPI (only) fathers, which is bigger than 3300 grams. However, the average newborn's birth weight is lower than 3175 grams if the infants' fathers are Asian (only) and Black (only). And if the infant's father has more than one race, the average birth weight for the infant stands between that for other infants with fathers having only one race. This probably implies that gene controlling for newborn's birth weight is different from race to race but we need to conduct more analysis to rule out confounding factors.

Table 2: Mean of newborn's birth weight by father's race

Race	White only	Black only	AIAN only	Asian only	NHOPI only	More than one race
Mean of Newborn's Birth Weight (g)	3331.485	3136.437	3319.952	3161.397	3321.406	3266.173



References

Bowen, Virginia, John Su, Elizabeth Torrone, Sarah Kidd, and Hillard Weinstock. "Increase in incidence of congenital syphilis-United States, 2012-2014." *MMWR Morb Mortal Wkly Rep* 64, no. 44 (2015): 1241-1245.

Dickert-Conlin, Stacy, and Amitabh Chandra. "Taxes and the Timing of Births." *Journal of political Economy* 107, no. 1 (1999): 161-177.

Finer, Lawrence B., and Stanley K. Henshaw. "Disparities in rates of unintended pregnancy in the United States, 1994 and 2001." *Perspectives on sexual and reproductive health* 38, no. 2 (2006): 90-96.

Finer, Lawrence B., and Mia R. Zolna. "Unintended pregnancy in the United States: incidence and disparities, 2006." *Contraception* 84, no. 5 (2011): 478-485.

Lanzieri, Tatiana M., Sheila C. Dollard, Cassandra D. Josephson, D. Scott Schmid, and Stephanie R. Bialek. "Breast milk–acquired cytomegalovirus infection and disease in VLBW and premature infants." *Pediatrics* 131, no. 6 (2013): e1937-e1945.