MACSC30200: Problem Set 1
Dataset: 2015 U.S. Natality Data

Joanna Tung
April 19, 2017

Introduction

2015 Natality data for all births occurring within the United States were obtained from the National Bureau of Economic Research (NBER).[1] The data are collected by the National Center for Health Statistics (NCHS) and is made available for public use through the National Vital Statistics System.[2] Data are compiled from a 100% sample of birth certificates in all U.S. states, including the District of Columbia. Table A on page 39 of the codebook confirms that the dataset contains only those births occurring within any of the fifty United States, excluding U.S. Territories of Puerto Rico, the Virgin Islands, Guam, America Samoa, and the Northern Marianas.[3] Variables provided include both maternal and paternal demographic characteristics (ex. race, age, and marriage status), as well as physical and medical characteristics (ex. body mass index (BMI), height, birth weight, and APGAR score range) and other population data (ex. State, county, population size).

The NCHS Natality database is widely used across a multitude of disciplines and for a diversity of applications. Though the NCHS Natality data is commonly used for trending purposes,[4] it has also been employed for a variety of other analyses, including epidemiological studies of crack cocaine and infant mortality clinical design studies to address early preterm birth. [5,6] Researchers have even utilized NCHS Natality data to assess the impact of various recent historical event and

[1] Csv and other formats of natality data can be obtained for U.S. states, years 1968-2015 and U.S. Territories, years 1994-2014, from the NBER website (https://nber.org/data/vital-statistics-natality-data.html).
[2] The National Center for Health Statistics is a division of the U.S. Centers for Disease Control and Prevention, see their website (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Births) to access the natality dataset in its original format. U.S. state and federal laws mandate the completion, collection and publication of birth certificates for all births. The Division of Vital Statistics, a subsidiary of the NCHS, collects the data and publishes it for public use through the Federal National Vital Statistics System. Per the NCHS' request, the following disclaimer is provided here: "any analyses, interpretations, or conclusions reached" are attributed to the author (Joanna Tung) "and not to NCHS, which is responsible only for the initial data." Disclaimer text quoted from NBER website (https://nber.org/natality/cite.html)
[3] The row count of the dataset matches the "occurrence" column of the table. This confirms that the provided natality excludes births occurring in the territories. See National Center for Health Statistics. User Guide to the 2015 Natality Public Use File. Hyattsville, Maryland: National Center for Health Statistics. Annual product 2016, page 39.
[4] See Kost K and Maddow-Zimet I, *U.S. Teenage Pregnancies, Births and Abortions, 2011: National Trends by Age, Race and Ethnicity*, New York: Guttmacher Institute, 2016, http://www.guttmacher.org/report/us-teen-pregnancy-trends-2011. Also Branum AM; Parker JD; Schoendorf KC; "Trends in US sex ratio by plurality, gestational age and race/ethnicity." *Hum Reprod* 2009; 24 (11): 2936-2944.
[5] Crack Cocaine study: Fryer RG; Heaton PS; Levitt SD; Murphy KM. "Measuring Crack Cocaine and Its Impact." *Economic Inquiry* 2013;51 (3) :1651-1681. Infant Mortality study: Li F; Wu T; Lei X; Zhang H; Mao M; et al. "The Apgar Score and Infant Mortality." *PLOS ONE* 8(7): e69072.
[6] Early preterm birth study: Zhang, C; Garrard, L; Carlson, S; Gajewski, B; "Subgroup identification of early preterm birth (ePTB): informing a future prospective enrichment clinical trial design," *BioMed Central* 2017 (17)18.

policy impacts on birth rates.[7] The longitudinal reach and completeness of this dataset makes it an especially powerful tool for many different purposes. For this paper, I will examine the following ten-variable-subset reported in the table below. I theorize that the chosen eight demographic and pre-birth maternal medical characteristics might correspond to increasing natal health, as expressed by birth weight and the APGAR score.[8]

Figure 1: Variables of Interest

Demographic Characteristics	Pre-birth Maternal Medical Characteristics	Natal Characteristics
Married	Prior Pregnancy Terminations	5-Minute APGAR Score Range
Maternal Age Range	Month Prenatal Care Began	Birth Weight (grams)
Maternal Education Level	Maternal BMI	
Paternal Education Level	Gestation Period in Weeks	

Data Cleanup

All "Unknown", "N/A" and null values were excluded from the descriptive statistic calculations below. NCHS initially classified educational levels on a numerical scale from 1-8 (9 being "Unknown"). This was unnecessarily granular for our level of analysis; consequently, the data was reclassified on a 5-point scale instead. Additional detail about the classification categories can be found in the footnotes for Figures 2 and 3 below.

Data Description

Overall Features

Most births occurred within married, rather than unmarried couples; however, children were born out of wedlock more often than anticipated. Parental education levels appear similar among men and women, though women show a slightly elevated mean compared to men, which could be due to the higher proportion of female students in the graduate school population, though this is not definitively known. Mean and median maternal BMI hewed more closely towards the "overweight" classification than expected. Similarly, the man and median birth weights suggest that the national average baby birth weight exceeds 7 pounds, which was higher than expected from anecdotal reports. The full range of descriptive statistics for the ten variables of interest is provided in Figure 2, below.

[7] A recent Pew Institute study linked declines in birth rate to the recession: Taylor, P; Kochlar, R; Livingston, G; Cohn D; Wang, W; Dockterman, D; "U.S. Birth Rate Decline Linked to Recession," Washington DC: Pew Research Center, 2016. The NIH also examined the impact of various family planning programs on birth rate: Bailey, MJ; "Reexamining the Impact of Family Planning Programs on US Fertility: Evidence from the War on Poverty and the Early Years of Title X," *Am Econ J Appl Econ*. 2012 April; 4(2): 62-97.

[8] See footnote 5 for Figure 2, for more information about the APGAR score.

Figure 2: 2015 NHCS Natality Data, Descriptive Statistics

	Mean (Std. Dev.)	Min	Median	Max	Count
Married[1]	1.40 (0.49)	1	1	2	3988733
Maternal Age Range[2]	4.30 (1.21)	1	4	9	3988733
Maternal Education Level[3]	2.88 (1.21)	1	3	5	3870528
Paternal Education Level[3]	2.83 (1.21)	1	3	5	3354672
Prior Pregnancy Terminations	0.41 (0.87)	0	0	30	3966486
Month Prenatal Care Began	2.94 (1.50)	1	3	10	3728140
Maternal BMI[4]	2.89 (1.19)	1	3	6	3794443
Gestation Period in Weeks	38.65 (2.49)	1	39	47	3985756
5-Minute APGAR Score Range[5]	3.83 (0.45)	1	4	4	3974151
Birth Weight (grams)	3270.14 (591.80)	227	3315	8165	3985085

National Center for Health Statistics (*2015*). Data File Documentations, Natality, *2015*, National Center for Health Statistics, Hyattsville, Maryland.

[1] 1 = married, 2 = unmarried

[2] Maternal Age Ranges are as follows: 1 = Under 15 years, 2 = 15-19 years, 3 = 20-24 years, 4 = 25-29 years, 5 = 30-34 years, 6 = 35-39 years, 7 = 40-44 years, 8 = 45-49 years, 9 = 50-54 years

[3] Education Levels were reclassified from the NBER dataset as follows: 1 ≤ High school degree, 2 = High school degree or equivalent, 3 = Post-High school coursework up to Associate's degree, 4 = Bachelor's degree, 5 = Graduate degree

[4] Maternal BMI levels are classified as follows: 1 = underweight, 2 = normal, 3 = overweight, 4 = obesity I, 5 = obesity II, 6 = extreme obesity

[5] 5-Minute APGAR Scores are a common measure of natal health. The following classifications were applied to the score ranges, based on the 2014 *Neonatal Encephalopathy and Neurologic Outcome*, Second Edition: 1 = nonspecific sign of illness, 2 = moderately abnormal, 3 = reassuring/low-end, 4 = reassuring/high-end

Features of Marriage Status

The dataset was also split into married and unmarried sub-categories for further investigation. On average, married mothers were more likely to have children later in life and report higher maternal and paternal educational levels, compared to unmarried mothers. The higher education levels and married status may help explain the earlier start in prenatal care among married women, though this is only conjecture. Regardless, natal health was similar for children of both married and unmarried mothers. Interestingly, though the maximum number of previous abortions reported for married mothers was 20% larger than that for unmarried numbers, on average, the number of previous abortions were nearly identical between married and unmarried mothers. The full set of descriptive statistics is provided in Figure 3, below.

Figure 3: 2015 NHCS Natality Data, Sorted by Marriage Status at Time of Birth

	Mean (Std. Dev.)		Min		Median		Max		Count	
	Married	Unmarried	Married	Unmarried	Married	Unmarried	Married	Unmarried	Married	Unmarried
Maternal Age Range[1]	4.67 (1.06)	3.75 (1.20)	1	1	5	4	9	9	2384342	1604391
Maternal Education Level[2]	3.31 (1.19)	2.25 (0.94)	1	1	3	2	5	5	2313488	1557040
Paternal Education Level[2]	3.16 (1.20)	2.14 (0.89)	1	1	3	2	5	5	2274622	1080050
Prior Pregnancy Terminations	0.40 (0.86)	0.41 (0.90)	0	0	0	0	30	25	2371152	1595334
Month in Term Prenatal Care Begun	2.77 (1.36)	3.20 (1.66)	1	1	2	3	10	10	2250614	1477526
Maternal BMI[3]	2.83 (1.15)	2.98 (1.24)	1	1	2	3	6	6	2272494	1521949
Gestation Period in Weeks	38.74 (2.31)	38.51 (2.72)	17	17	39	39	47	47	2383114	1602642
5-Minute APGAR Score Range[4]	3.84 (0.43)	3.81 (0.48)	1	1	4	4	4	4	2375771	1598380
Birth Weight (grams)	3321.76 (578.27)	3193.41 (603.25)	227	227	3355	3233	8165	8165	2382154	1602931

National Center for Health Statistics (*2015*). Data File Documentations, Natality, *2015*, National Center for Health Statistics, Hyattsville, Maryland.

[1]Maternal Age Ranges are as follows: 1 = Under 15 years, 2 = 15-19 years, 3 = 20-24 years, 4 = 25-29 years, 5 = 30-34 years, 6 = 35-39 years, 7 = 40-44 years, 8 = 45-49 years, 9 = 50-54 years

[2]Education Levels were reclassified from the NBER dataset as follows: 1 ≤ High school degree, 2 = High school degree or equivalent, 3 = Post-High school coursework up to Associate's degree, 4 = Bachelor's degree, 5 = Graduate degree

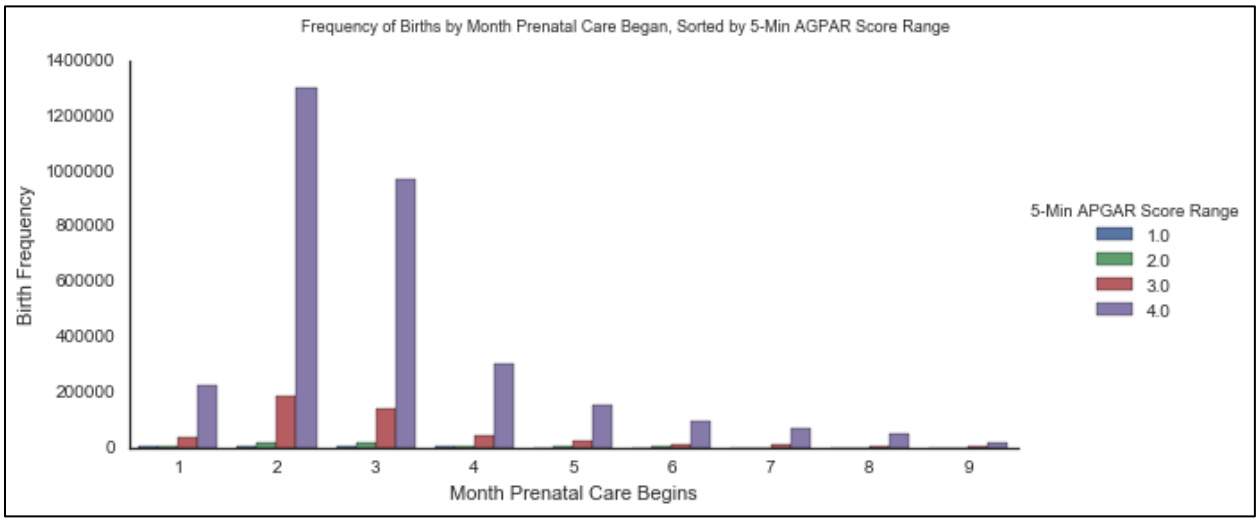
[3]Maternal BMI levels are classified as follows: 1 = underweight, 2 = normal, 3 = overweight, 4 = obesity I, 5 = obesity II, 6 = extreme obesity

[4]5-Minute APGAR Scores are a common measure of natal health. The following classifications were applied to the score ranges, based on the 2014 *Neonatal Encephalopathy and Neurologic Outcome,* Second Edition: 1 = nonspecific sign of illness, 2 = moderately abnormal, 3 = reassuring/low-end, 4 = reassuring/high-end

Natal Health and Prenatal Care

Lastly, I examined birth frequency by natal health using the reported 5-Min APGAR Score Range and the Month in the Pregnancy Term when Prenatal Care began, see Figure 4, below. The frequency histogram illustrates that the overwhelming number of births result in high-scoring APGAR ratings, regardless of when Prenatal Care began. This is a hopeful indicator about the state of natal health in the United States, though may also indicate the limitations of early prenatal care on natal health. Without further investigation, it is impossible to draw any correlations between prenatal care and natal health.

Figure 4: Birth Frequency by Natal Health Score (5-Min APGAR Score Range) and Start of Prenatal Care



National Center for Health Statistics (*2015*). Data File Documentations, Natality, *2015*, National Center for Health Statistics, Hyattsville, Maryland.

*5-Minute APGAR Scores are a common measure of natal health. The following classifications were applied to the score ranges, based on the 2014 *Neonatal Encephalopathy and Neurologic Outcome*, Second Edition: 1 = nonspecific sign of illness, 2 = moderately abnormal, 3 = reassuring/low-end, 4 = reassuring/high-end