Problem Set #1
MACS 30200, Dr. Evans
Bobae Kang

1 Write a data section

The current national birth data is part of the National Vital Statistics System (NVSS), an inter-governmental data sharing program by which the National Center for Health Statistics (NCHS) collects and provides access to the official vital statistics. To collect data, NCHS works in partnership with the vital registration systems on the state and local levels. Anyone can access the data as well as the User's Guide documentation on the NCHS web page.[1] The same data is also accessible via the National Bureau of Economic Research website.[2]

The comprehensiveness of the birth data has allowed many previous studies of various disciplines to generate meaningful empirical results. For example, Bateni et al. (2016) used the birth data to analyze the cesarean delivery rate among twin pregnancies for a specific time period. Another example is Pesko and Currie (2016), which used the birth data to examine the effect of e-cigarette minimum legal sale laws on birth outcomes among pregnant teenagers.

The birth data for the year of 2015 includes total 3,988,733 observations and 240 variables. These variables include time and place of the birth, various demographic features of the infants' parents, mothers' conditions before and during pregnancy (use of cigarette, height and weight, and risk factors), diseases, characteristics of labor and delivery, and more. The NVSS web page provides a link to the official documentation of the data with a full description of all 240 variables.[3]

Table 1 provides a descriptive summary of selected variables, including infants' sex and weight, birth facility, paying method for delivery, and parents' demographic characteristics. Some of these variables are made into binaries (0-1) for simpler presentation in the tabular format. Omitted values of the given variable are not taken into account when calculating summary statistics. In addition, for non-binary variables, values that are merely placeholders for "Unknown" values are excluded in getting summary statistics. For Mother's age, the value of 50 denotes the age of 50 and older in the original data. However, this definition is ignored in calculating the summary statistics for the mother's age since there are only 769 such instances, or less than 0.02% of all observations.

[1] https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm
[2] http://nber.org/data/vital-statistics-natality-data.html
[3] ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2015.pdf

Table 1: Summary statistics of selected variables

	Type	Mean	Std.Dev.	Median	Max	Min	N
Infant's sex (female)	0-1	0.49	0.50	0	-	-	3,988,733
Infant's weight (grams)	Cont.	3270	591.80	3315	8165	227	3,985,085[1]
Birth type (single)	0-1	0.97	0.18	1	-	-	3,988,733
Birth facility (hospital)	0-1	0.98	0.12	1	-	-	3,988,733
Payment source (Medicaid)	0-1	0.43	0.49	0	-	-	3,919,668[2]
Mother's marital status	0-1	0.60	0.49	1	-	-	3,988,733
Paternity acknowledged	0-1	0.68	0.47	0	-	-	1,647,583[3]
Mother's age (years)	Count	28.51	5.86	29	12	50	3,988,733
Mother's race (white)	0-1	0.76	0.43	1	-	-	3,988,733
Mother's education (BA degree or higher)	0-1	0.32	0.47	1	-	-	3,919,668[2]
Mother's nativity	0-1	0.77	0.42	1	-	-	3,988,733
Father's age (years)	Count	31.36	6.83	31	98	13	3,513,343[4]
Father's race (white)	0-1	0.63	0.48	1	-	-	3,988,733
Father's education (BA degree or higher)	0-1	0.4	0.49	0	-	-	3,919,668[2]

[1] 3,648 observations with unknown weight values (coded as 9999 in the original data) are removed
[2] 69,065 observations with omitted (i.e., "NA") values are removed
[3] 2,341,150 observations with non-applicable (''X") values are removed
[4] 475,390 observations with unknown age values (coded as 99 in the original data) are removed

Some of the most notable characteristics are as follows: In the year of 2015, the overwhelming majority of babies were delivered in hospitals. In addition, over two-fifth of all births were paid for by Medicaid, and about 23 percent of babies were born to mothers who were not born in the United States. Approximately 60 percent of newborns were born to mothers who were married. Among those babies likely born out of the wedlock, about 68 percent had fathers whose paternity was legally acknowledged. The average age of known fathers were slightly higher than that of mothers. Also, a greater percentage of fathers held Bachelor's degree than mothers.

Table 2 illustrates the demographic characteristics of infants' parents in greater detail. In the Panel B. the marital status of fathers is imputed from the value of paternity acknowledgement variable. The table shows various interesting aspects of the data. Among white parents who received higher education (college or more), the overwhelming majority were married: 90.5% of mothers and 93.0% of fathers. The same pattern does not hold for non-white parents, especially for fathers. In fact, approximately two-thirds of non-white fathers with advanced degrees were unmarried.

Parents without college or higher degrees show somewhat different patterns. White mothers without advanced degrees are almost equally divided between married (51.3%) and not married (48.7%). In case of non-college graduate white fathers, more such fathers were married than not (62.6% versus 37.4%). Somewhat differently, greater portion of less educated non-white mothers were unmarried (69.1%) than married

(30.9%). However, non-white fathers without college degrees were somewhat more evenly divided between married and unmarried: 43.3% and 56.7%.

Table 2: Demographic characteristics of parents

Panel A. Mothers

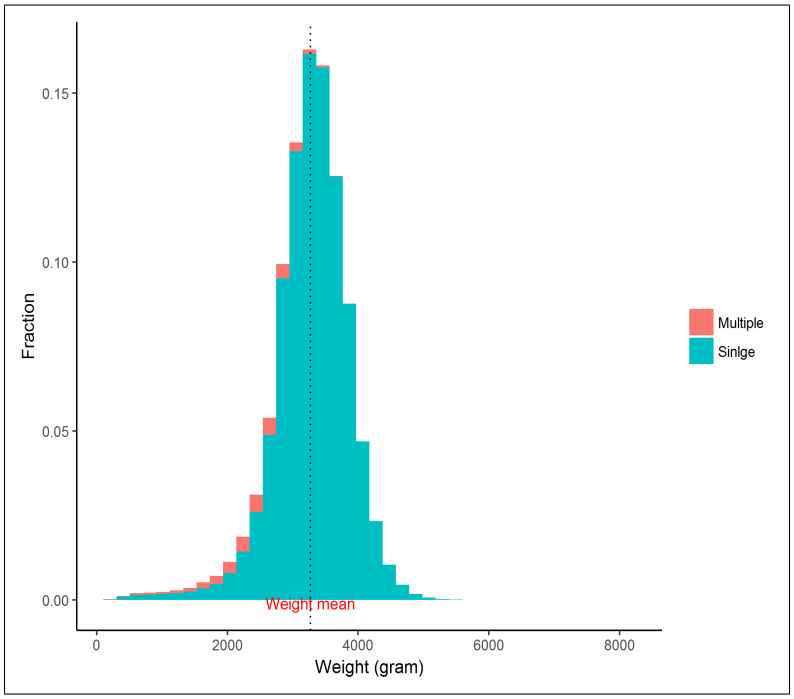
		Married		Unmarried		Total
		U.S. born	Foreign born	U.S. born	Foreign born	
White	Total	1,587,343	354,377	850,208	230,088	3,022,016
		(0.398)	(0.089)	(0.213)	(0.058)	(0.758)
	College	771,258	112,959	72,966	19,590	976,773
		(0.193)	(0.028)	(0.018)	(0.005)	(0.245)
	No college	790,436	233,788	764,844	205,555	1,994,623
		(0.198)	(0.059)	(0.192)	(0.052)	(0.500)
	NA	25,649	7,630	12,398	4,943	50,620
		(0.006)	(0.002)	(0.003)	(0.001)	(0.013)
Non-white	Total	180,603	262,019	452,930	71,165	966,717
		(0.045)	(0.066)	(0.114)	(0.018)	(0.242)
	College	76,629	147,045	35,588	12,825	272,087
		(0.019)	(0.037)	(0.009)	(0.003)	(0.068)
	No college	101,514	107,521	411,000	56,150	676,185
		(0.025)	(0.027)	(0.103)	(0.014)	(0.170)
	NA	2,460	7,453	6,342	2,190	18,445
		(0.001)	(0.002)	(0.002)	(0.001)	(0.005)

Panel B. Fathers

		Married	Unmarried	Total
White	Total	1,781,246	718,997	2,500,243
		(0.447)	(0.180)	(0.627)
	College	744,737	56,136	800,873
		(0.187)	(0.014)	(0.201)
	No college	1,036,509	618,529	1,655,038
		(0.260)	(0.155)	(0.415)
	NA	0	44,332	44,332
		(0)	(0.011)	(0.011)
Non-white	Total	559,904	928,586	1,488,490
		(0.140)	(0.233)	(0.373)
	College	257,701	508,800	766,501
		(0.065)	(0.128)	(0.192)
	No college	302,203	395,053	697,256
		(0.076)	(0.099)	(0.175)
	NA	0	24,733	24,733
		(0)	(0.006)	(0.006)

Note: Figures in parentheses denote proportions, obtained by dividing the number of observations for each category by the size of the entire population, i.e., 3,988,733.

Figure 1: Histogram of the distribution of weight of infant at birth

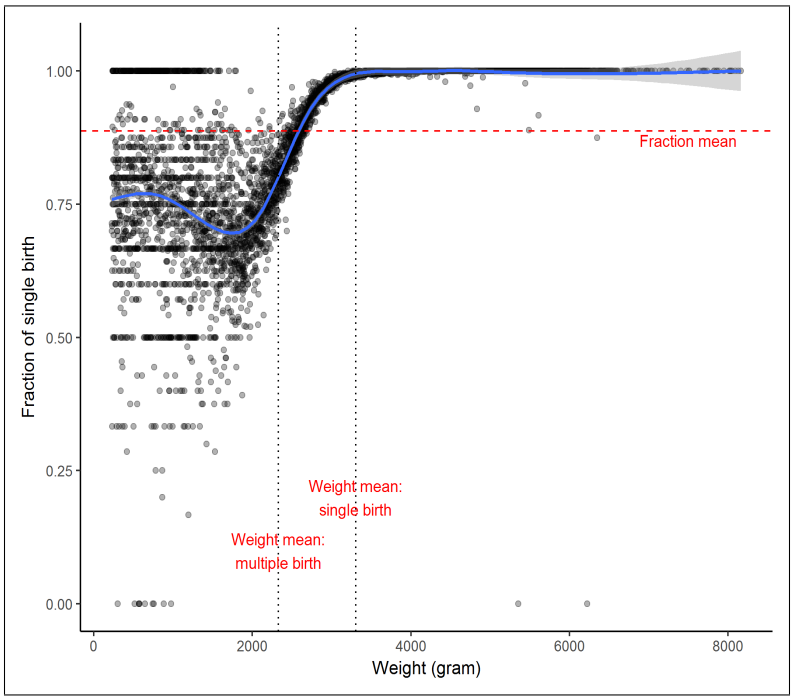


Note: The vertical dotted line marks the average weight of all infants (3270.14 grams). The two colors mark the different types of each birth: single (blue) and multiple (red). 3,648 observations with unknown weight values (coded as 9999 in the original data) are removed (N = 3,985,085).

Figures 1 and 2 take a closer look at the correlation between infants' weight and the type of birth. First, Figure 1 presents a stacked histogram illustrating the distribution of weight of infant. In this plot, different colors are used to show different types of birth: single (only one child at a birth) and multiple (the number of offspring at birth is two or more). The weight distribution looks roughly Gaussian. As shown in Table 1, the majority of all births (approximately 96%) are single births. One notable feature of weight distribution is that infants delivered in multiple births, colored blue in the stacked histogram, tend to be lighter in weight than others.

Figure 2 presents a scatter plot of the fraction of single birth by weight. The scatter plot shows that while there is little notable correlation between weight of infant and the proportion of single birth for the given weight for infants lighter than 2,000 grams (about 4.4 pounds), a strong positive correlation appears for infants whose weight is between 2,000 and 3,000 grams, approximately. That is, in the weight interval of 2,000 and 3,000 grams, the proportion of infants delivered in single births is likely to increase as weight of infant increases. The great majority of infants that are heavier than 3,000 grams were delivered in single births.

Figure 2: Fraction of single birth by weight of infant at birth



Note: The horizontal red dashed line marks the mean of the fraction of single birth at each weight (0.887). The two vertical dotted lines mark the means of the weight of infant for different birth types: single (3303.81 grams) and multiple (2326.11 grams). 3,648 observations with unknown weight values (coded as 9999 in the original data) are removed (N = 3,985,085).

References

Bateni, Zhoobin H., Steven L. Clark, Haleh Sangi-Haghpeykar, Kjersti M. Aagaard, Yair J. Blumenfeld, Susan M. Ramin, Henry C. Lee, Karin A. Fox, Amirhossein Moaddab, Amir A. Shamshirsaz, Bahram Salmanian, Pardis Hosseinzadeh, Diana A. Racusin, Hadi Erfani, Jimmy Espinoza, Gary A. Dildy, Michael A. Belfort, and Alireza A. Shamshirsaz, "Trends in the delivery route of twin pregnancies in the United States, 2006-2013," *European Journal of Obstetrics and Gynecology*, 2016, *205*.

Pesko, Michael F. and Janet M. Currie, "The effect of e-cigarette minimum legal sale age laws on traditional cigarette use and birth outcomes among pregnant teenagers," *NBER Working Paper*, 2016.