

MACS 302

Problem Set 1

Reid McIlroy-Young

April 19, 2017

The data used for this analysis are from the *2015 Natality Public Use File* provided by the US Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS). The specific data file used is provided by the National Bureau of Economic Research (NBER) and has been converted to CSV. The data are available from both sources, binned by year from 1968 to 2015, but this analysis is only done with the 2015 data. Different years have greatly varying levels of coverage in both sample population coverage and questions, with 2015 having coverage of all US resident's births within the US and 240 reported variables per birth. The original data can be found [here](#) and the NBER compilation [here](#).

US hospitals are mandated to report all births of live infants to the CDC, although the exact data reported varies between states and by year, this is then combined and most of it made public with some columns removed for privacy such as much of the fine geographic data being removed for the last few years. Note that the definition of 'within the US' is defined for 2015 as "*births registered in the 50 States, the District of Columbia, and New York City*" [1] and thus is excluding some number of births of Americans on US soil (e.g. overseas embassies, unregistered births) and possibly including some number not on US soil (e.g. foreign embassies). Notably, US territories are also included for 2015 although some years only have them as a supplement.

These data have been used in a variety of different domains. The CDC has an annual health survey of vital statistics that they produce most years, the latest being released in 2015 [2] this used the natality data set. The data are also used to answer more specific questions such as those examining diseases spread from mothers to children [3] [4] or as a starting point for longer studies such as those examining early childhood interventions [5]. There also have been multi-year longitudinal analysis such as those looking at health outcomes for pregnant women [6], the effects of birth control access [7] or teen birth patterns [8].

Since the available data has 3 988 734 observations with 240 values looking at the complete set is in piratical and beyond the scope of this report. Although we estimate only 100 or so of these variables contain non-redundant information as most columns have an accompanying ‘Recode’ column and the different column groups are divided by ‘Filler’ columns.

Here we will look at those relevant to our analysis: The mother’s smoking habits which is given by four estimates of her number of cigarettes per day. A few measures of mother’s health: where she gave birth, her weight at birth, her height and her weight gain after birth. Along with the month of birth, figure 3 shows the summary statistics. Note that born in hospital is a categorical variable with 1 being born in hospital and 0 being other location. Also note that most of the fields have a no answer value and responses with it have been dropped.

One interesting result of examining the data on number cigarettes smoked daily by the mother by trimester, figure 1 shows the histograms of cigarette usage by trimester for smokers. Two things are notable, first that some people are likely over reporting their per day usage. Secondly that as the pregnancy progresses the mothers tend to reduce their smoking. The cause of this will be explored later.

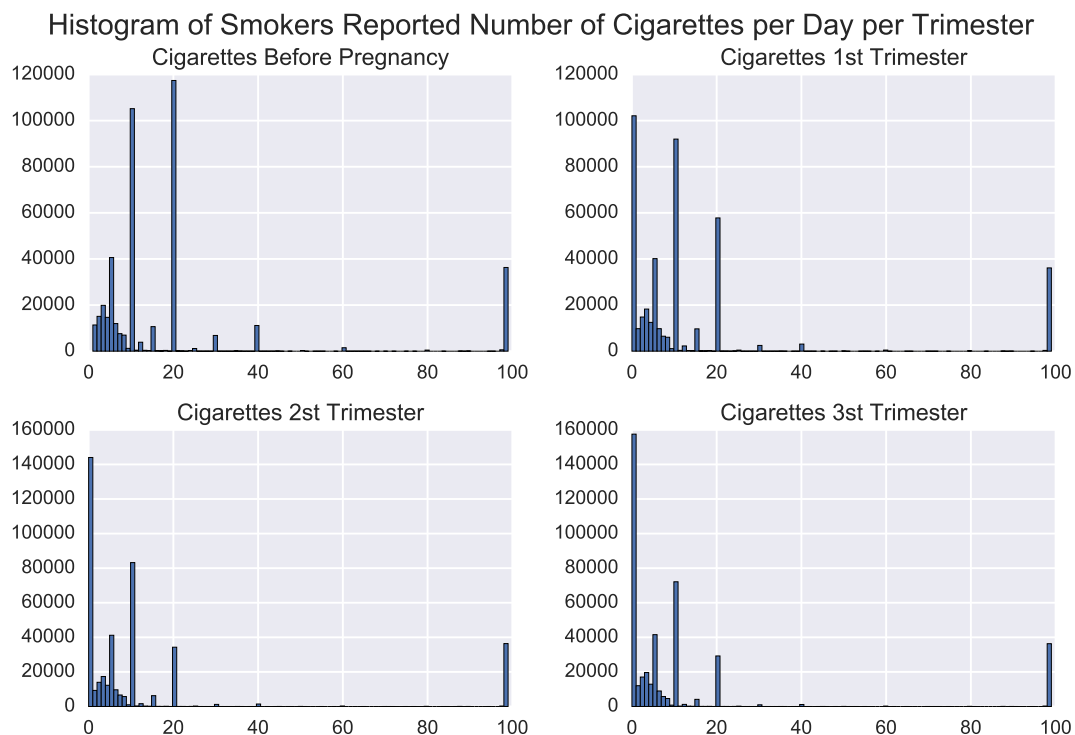


Figure 1: *Number cigarettes smoked daily by the mother by trimester*

We also see an interesting seasonal component to most of the metrics, figure

2 shows the mean of the cigarette usage, hospital birth and number of month of prenatal care as compared to the birth month. These all show some season variance, which we will show is statistical significant. In particular the amount people smoke tends to increase during the winter months, even if they are pregnant. This plot also shows how much of a drop occurs between trimesters in the mothers smoking.

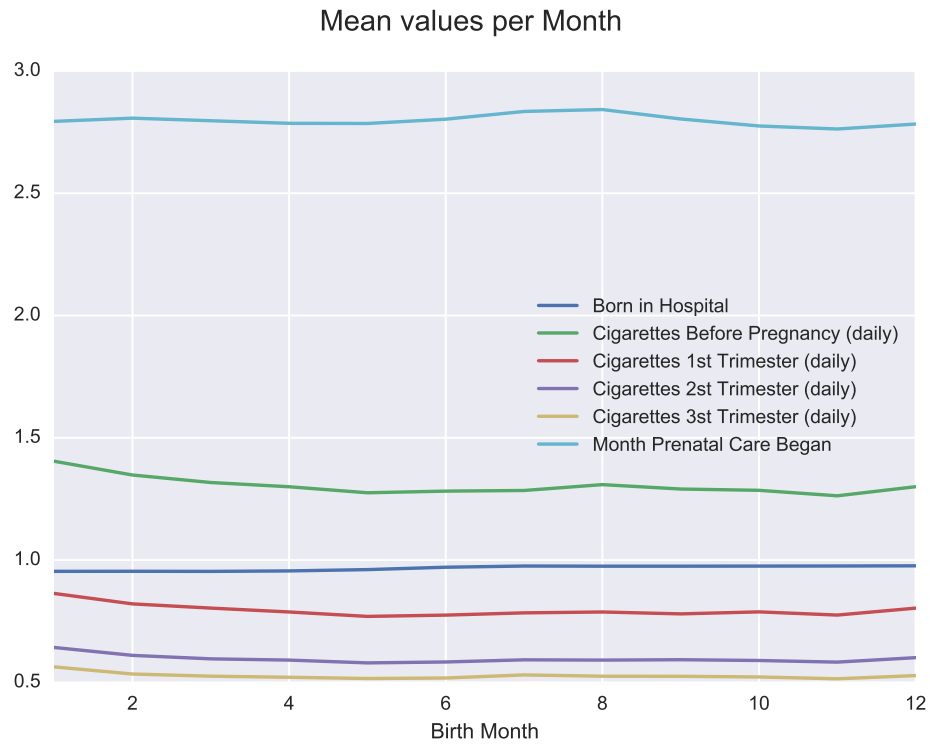


Figure 2: The mean of the cigarette usage, hospital birth and number of month of prenatal care as compared to the birth month

Figure 3: Summary statistics

	Born in Hospital	Cigarettes Daily Before Pregnancy	Cigarettes Daily 1st Trimester	Cigarettes Daily 2nd Trimester	Cigarettes Daily 3rd Trimester
mean	0.9671	1.3040	0.7944	0.5957	0.5267
std	0.1783	5.0204	3.6024	2.9626	2.7643
min	0.0000	0.0000	0.0000	0.0000	0.0000
25%	1.0000	0.0000	0.0000	0.0000	0.0000
50%	1.0000	0.0000	0.0000	0.0000	0.0000
75%	1.0000	0.0000	0.0000	0.0000	0.0000
max	1.0000	98.0000	98.0000	98.0000	98.0000

	Month before birth of Prenatal Care	Mother's Height (Inches)	Mother's Weight at Birth (pounds)	Mother's Weight Gain (pounds)	Birth Month
mean	2.7981	64.1520	186.1590	30.1786	6.5747
std	1.5932	2.8301	40.3255	14.7537	3.4203
min	0.0000	30.0000	100.0000	0.0000	1.0000
25%	2.0000	62.0000	158.0000	21.0000	4.0000
50%	2.0000	64.0000	180.0000	30.0000	7.0000
75%	3.0000	66.0000	207.0000	38.0000	9.0000
max	10.0000	78.0000	400.0000	98.0000	12.0000

References

- [1] National center for health statistics, 2015.
- [2] Michelle JK Osterman, Kenneth D Kochanek, Marian F MacDorman, Donna M Strobino, and Bernard Guyer. Annual summary of vital statistics: 2012–2013. *Pediatrics*, pages peds–2015, 2015.
- [3] Tatiana M Lanzieri, 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(6):e1937–e1945, 2013.
- [4] Virginia Bowen, 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(44):1241–5, 2015.
- [5] Carrie Shapiro-Mendoza, Milton Kotelchuck, Wanda Barfield, Carol A Davin, Hafsatou Diop, Michael Silver, and Susan E Manning. Enrollment in early intervention programs among infants born late preterm, early term, and term. *Pediatrics*, pages peds–2012, 2013.
- [6] William M Callaghan, Susan Y Chu, and Denise J Jamieson. Deaths from seasonal influenza among pregnant women in the united states, 1998–2005. *Obstetrics & Gynecology*, 115(5):919–923, 2010.
- [7] Richard P Dickey. The relative contribution of assisted reproductive technologies and ovulation induction to multiple births in the united states 5 years after the society for assisted reproductive technology/american society for reproductive medicine recommendation to limit the number of embryos transferred. *Fertility and sterility*, 88(6):1554–1561, 2007.
- [8] Centers for Disease Control, Prevention (CDC, et al. Vital signs: Repeat births among teens–united states, 2007–2010. *MMWR. Morbidity and mortality weekly report*, 62(13):249, 2013.