Caitlyn Bryant (fourth year undergrad, Applied Mathematics major)
Kieran Mann (fourth year undergrad, Math and Economics major)
Matthew Kucirek (fourth year undergrad, Applied Mathematics major)
Edward Chao (fourth year undergrad, Probability and Statistics major)

Case Study 1
Professor Bradic
Math 189
"babies23"

Maternal Smoking and Infant Death

Our objective is to evaluate the Surgeon General warning: "Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight" by comparing the birth weights of babies born to smokers and babies born to non-smokers.

When looking at the factors that attribute to low-birthweight babies, we take into account: (1) whether the birth was preterm labor because that can lead to a low-birthweight baby. (The typical gestation period for a baby is 40 weeks. Preterm labor occurs when the baby is born at any time before 37 weeks.), (2) whether the baby was first born because babies that are not first born are likely to weigh less, (3) the mother's age because if the mom is less than 17 years old or over 35 years old, the baby is likely to weigh less, (4) the mother's height and weight (we combined these two variables into "body mass index") because those mothers that have a low body mass index are more likely to have a low-birthweight baby[1], (5) the father's body mass index because those fathers that have a low body mass index are more likely to have a low-birthweight baby.[2]

The primary data that we studied was part of the Child Health and Development Studies (CHDS) and was collected of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Health Plan of the San Francisco region, comprised of 1236 single-birth (no twins) male babies that all lived at least 28 days.

Other knowledge to take into account: (1) A low-birthweight baby is considered to be any baby that weighs less than 89 ounces; (2) Through epidemiological studies, it has been indicated that smoking is responsible for 150g of reduction in birth weight, and smoking mothers are twice as likely than non-smoking mothers to have a low-birthweight baby[3]; (3) The data collected for our study is an enlarged portion of the CHDS data; (4) The women involved in our study were enrolled in Kaiser Health Plan, obtained parental care in the San Francisco area,

---

[1] "Low Birthweight." March of Dimes., Oct. 2014. Web. 25 Jan. 2017. <http://www.marchofdimes.org/complications/low-birthweight.aspx>.

[2] Rice, Frances, and Anita Thapar. "Estimating the Relative Contributions of Maternal Genetic, Paternal Genetic and Intrauterine Factors to Offspring Birth Weight and Head Circumference." Early Human Development. Elsevier, July 2010. Web. 25 Jan. 2017. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2954294/>.

[3] "The Influence of Maternal Smoking and Exposure to Residential ETS on Pregnancy Outcomes: A Retrospective National Study." Maternal and Child Health Journal. U.S. National Library of Medicine, n.d. Web. 25 Jan. 2017. <https://www.ncbi.nlm.nih.gov/pubmed/23090285>.

and delivered their babies in any of the Kaiser hospitals in Northern California; (5) At the time of the baby's birth, the baby's length and weight are recorded.

Up until the 1940s, no one thought that a pregnant mother could carry a disease that could hurt her baby. In 1941, Dr. Norman Gregg observed an "unusual" number of infants with congenital cataracts and had realized that all of the mothers had contracted rubella (also known as German measles) in their first or second month of pregnancy.[4]

It is commonly believed that the carbon monoxide found in cigarette smoke reduces the oxygen supply to the fetus. Even though the physiological effects of decreased oxygen supply on the development of a fetus are not entirely understood, it is known that a steady supply of oxygen is vital for a developing baby. It is believed that the placenta increases in surface area and the number of blood vessels in order to compensate for the reduced oxygen supply, which then hypothetically leads to "abruptio placentae", or placental abruption. Placental abruption occurs when the placenta detaches from the uterine wall, resulting in preterm labor and fetal death.[5]

The "big questions" that we will answer by our data analyses are: (1) What is the difference in weight between babies born to mothers who smoked during pregnancy and those who did not?; (2) Is this difference important to the health and development of the baby? (In our study, we define a healthy baby to be a baby that weighs greater than 88 ounces and is born between 37 and 42 weeks.)

In the remainder of our paper, we will summarize numerically the differences between the two distributions of babies' birth weights from mothers who smoked while pregnant and mothers who did not. We will then use graphical methods to compare these two distributions, and compare the frequencies of low-birthweight babies for the two groups. Furthermore, we will draw conclusions based on our research in order to answer the questions stated above.

Numerical Analysis:

In this study, there are five different smoking categories: 0 - never smoked, 1- smoked during pregnancy, 2 - smoked up until pregnancy, 3 - smoked at one point but not during pregnancy and 9 - unknown. For the purpose of measuring how smoke affects the health of the baby, we disregard the 9 category as it provides no insight into the question. To see how the smoking affects the weight of the baby, we separated the data based on smoking category and found the respective mean, standard deviation, skewness and kurtosis of each smoking category. There are 544 babies in group 0, 484 in group 1, 95 in group 2 and 103 in group 3. The results are summarized in the following table (see Table 1.1 at top of next page):

---

[4] Dunn, P. M. "Perinatal Lessons from the Past: Sir Norman Gregg, ChM, MC, of Sydney (1892–1966) and Rubella Embryopathy." Archives of Disease in Childhood. Fetal and Neonatal Edition. BMJ Group, Nov. 2007. Web. 25 Jan. 2017. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675410/>.

[5] Wickström, R. "Effects of Nicotine During Pregnancy: Human and Experimental Evidence." Current Neuropharmacology. Bentham Science Publishers Ltd., Sept. 2007. Web. 25 Jan. 2017. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656811/>.

	Smoke - 0	Smoke - 1	Smoke - 2	Smoke - 3
Mean	122.78	114.11	123.08	124.44
Std Deviation	17.11	18.10	17.80	18.61
Skew	-0.12	-0.033	-0.38	-0.34
Kurtosis	4.31	2.99	3.56	3.36

TABLE 1.1: *Summary Statistics by Smoking Category*

Before working with the data, we want to validate the distribution and test for normality. To test for normality we run the D'Agostino test of skewness and the Anscombe-Glynn test of kurtosis. For the D'Agostino test, the null hypothesis is that the skew is zero and it is a two tailed test. For the Anscombe-Glynn test, the null hypothesis is that the kurtosis is 3 and it is also a two tailed test. The following table displays the p-values for the tests:

	Smoke - 0	Smoke - 1	Smoke - 2	Smoke - 3
p-Skewness	.24	.76	.12	.14
p-Kurtosis	3.053e-05	.92	.19	.30

TABLE 1.2: *p-Values of hypothesis tests for skewness and kurtosis*

This tells us that categories 1,2 and 3 are normal. In fact, the distribution for smokers performed very well with a 2.99 kurtosis and a -.033 skew. From this, we expect this to follow the rules of the normal distribution very well. The distribution for category 0 has a large kurtosis of 4.31, making it more likely to realize extreme values. Examining the data further, we found that there are a total of 63 underweight babies. Of these underweight babies, 17 came from nonsmokers, 40 came from smokers and 3 each came from categories 2 and 3. Despite the nonsmokers' distribution having a large kurtosis *and* the smokers' distribution normal behavior, most of the underweight babies are coming from the smokers' distribution.

In the mean section of the table, we see that babies born to mothers in category 1 have the smallest average weight. To test if this difference is significant we run a hypothesis test. Let $\mu 0$ and $\mu 1$ be the means for category 0 and 1, respectively. We test the null hypothesis $\mu 0 = \mu 1$ against the alternative hypothesis that $\mu 1 < \mu 0$. Performing Welch Two Sample t-test in R, we get a t-statistic of 7.86 with 996 degrees of freedom and a p-value of 4.816e-15. From this result, we reject the null hypothesis and cannot conclude that $\mu 0 = \mu 1$, hence there is a statistical significance in the difference of the means of the babies' weight. Replacing category 0 with either category 2 or 3 does not change the result of the two sample t-test. This is expected, as the distributions for 0, 2 and 3 are numerically similar.

Looking at each distribution, we see that smoking has little to no effect on the standard deviation. Despite how the mother smoked, the babies' weights have a similar spread of data. We have already verified normality for smokers. For nonsmokers, the skew is that of a normal distribution— however the kurtosis cannot be verified through the hypothesis test. Despite this, we assume that the distribution for nonsmokers is also normal. Let X represent a baby's weight from nonsmoking mothers (category 0) and let Y represent a baby's weight from smoking mothers (category 1). Then X-Y is a random variable that is also normally distributed with mean 8.67 and standard deviation 24.91. The probability that X-Y is greater than zero is 0.636. In other words, there is a 63.6% chance that a baby from a nonsmoking mother outweighs a baby from a smoking mother.

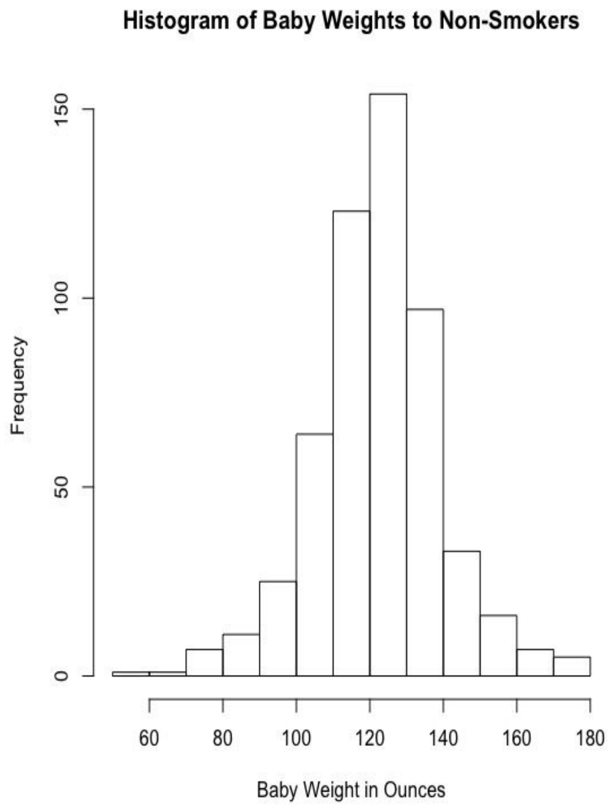
To assess the differences in the distributions of baby weight based on mother's smoking status, we looked at the quantiles and extremes of each distribution. The information in presented in the following table:

	Smoke - 0	**Smoke - 1**	**Smoke - 2**	**Smoke - 3**
Min	55	58	62	65
1st Quartile	113	102	112	112
Median	124	115	122	124
3rd Quartile	132.2	126	136.5	138.0
Max	176	163	163	170

TABLE 1.3: *Quartiles and extremes of baby weight by mother's smoking category*

Again, we see that the distributions from categories 0, 2 and 3 are similar. However, distributions 0 and 1 vary. The first quartile for nonsmokers is near the median for smokers. Similarly the median for nonsmokers is the third quartile for smokers. This supports that the distributions are spread out similarly, yet the distribution for smokers is shifted to the left.

Graphical Analysis:



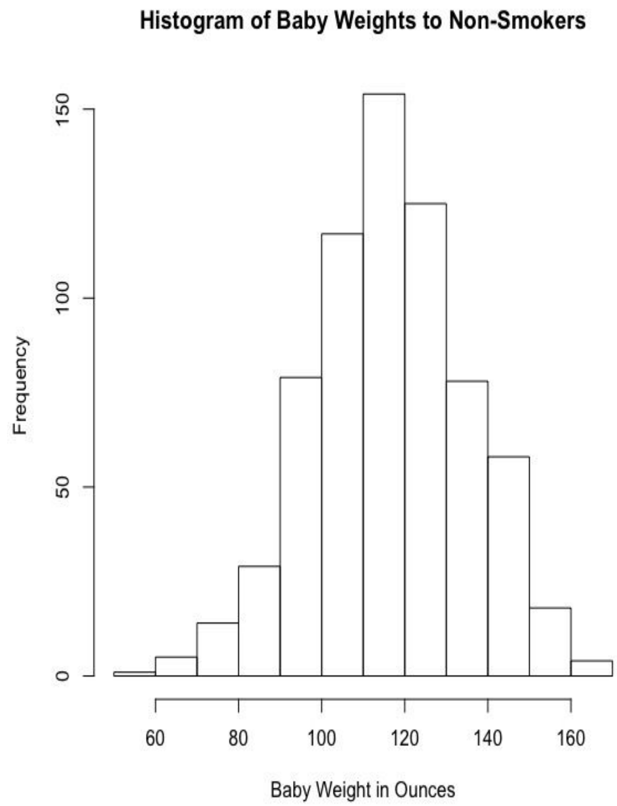


FIGURE 1(a): *Histogram of the baby birth weights to mothers who did not smoke*	**FIGURE 1(b):** *Histogram of the baby birth weights to mothers who smoked*

Both histograms are unimodal, but there should be special attention given to the location of the modals. Figure 1(a) shows a mode at around 125 oz, which is healthy weight for a baby. However, in comparison, Figure 1(b) has a mode to the left, located at around 115 oz, which is slightly, but noticeably less than the weight of babies born to mothers who did not smoke. Approximately 30% of the baby weight in ounces is located to the left of 110 oz in Figure 1(a), while an estimate of 40% is located to the left of 120 oz in Figure 1(b). This result furthers the notion that smoking has an effect on an infant's birth weight, as the distribution is skewed to the right in Figure 1(b), and vice versa in Figure 1(a).

In contrast to the birthweight distribution of mother who smoked versus mothers who did not smoke, we also looked at the how often the mothers smoked.

Histogram of Baby Weights to Frequency Smoked

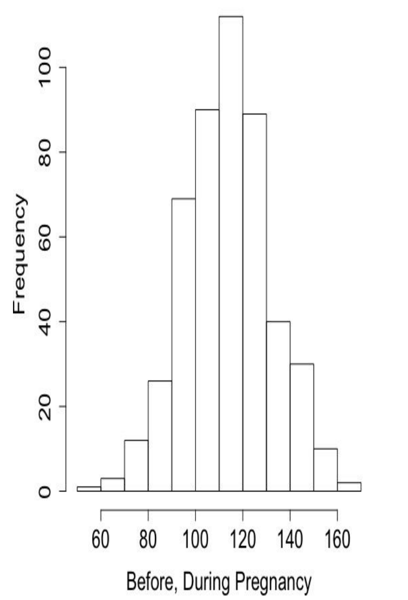


FIGURE 2(a): *birth weights of babies born to mothers who still smoked during pregnancy*

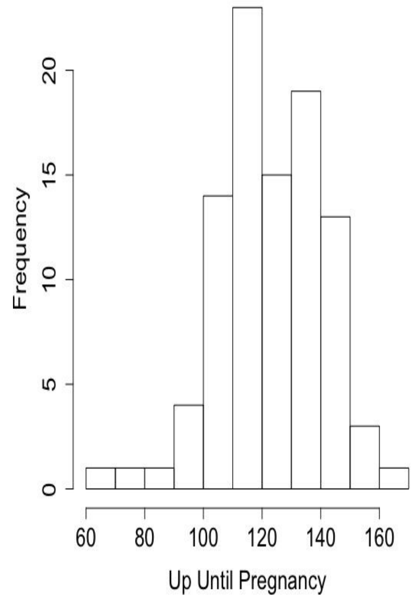


FIGURE 2(b): *birth weights of babies born to mothers who smoked up until pregnancy*

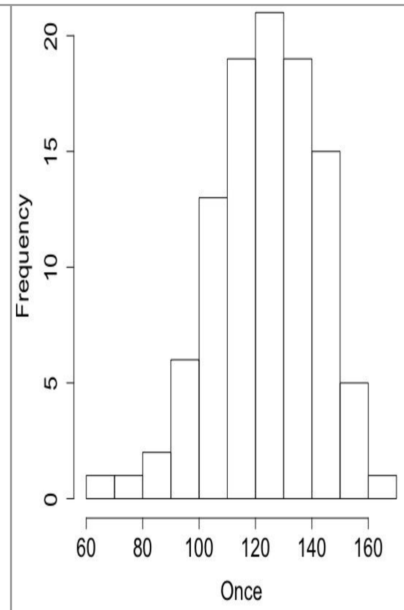


FIGURE 2(c): *birth weights of babies born to mothers only smoked once before pregnancy*

We explored three different cases: mothers who smoked while pregnant, mothers who smoked up until pregnancy, and mothers who only smoked once before pregnancy. Figure 2(a) shows the distribution of mothers continued smoking, showing a distribution similar to Figure 1(b), with one mode at around 115 oz. Figure 2(c) shows a distribution similar to Figure 1(a), but showing more variance. The histogram that carries the most interest is Figure 2(b), which shows two modes, one at around 115 oz, and the second mode at around 135 oz. The second mode on the right is noticeably smaller than the mode on the left, but also suggests that the distribution is skewed to the right. Similarities in distribution across the graphs in Figures 1 and 2 exhibit a relation between smoking and infant birth weight, while even implying that the frequency plays a role.

To delve into the greater detail on how much the mothers smoked, we used a box plot to ascertain the relation between the baby's birthweight and the number of cigarettes smoked by the mother.

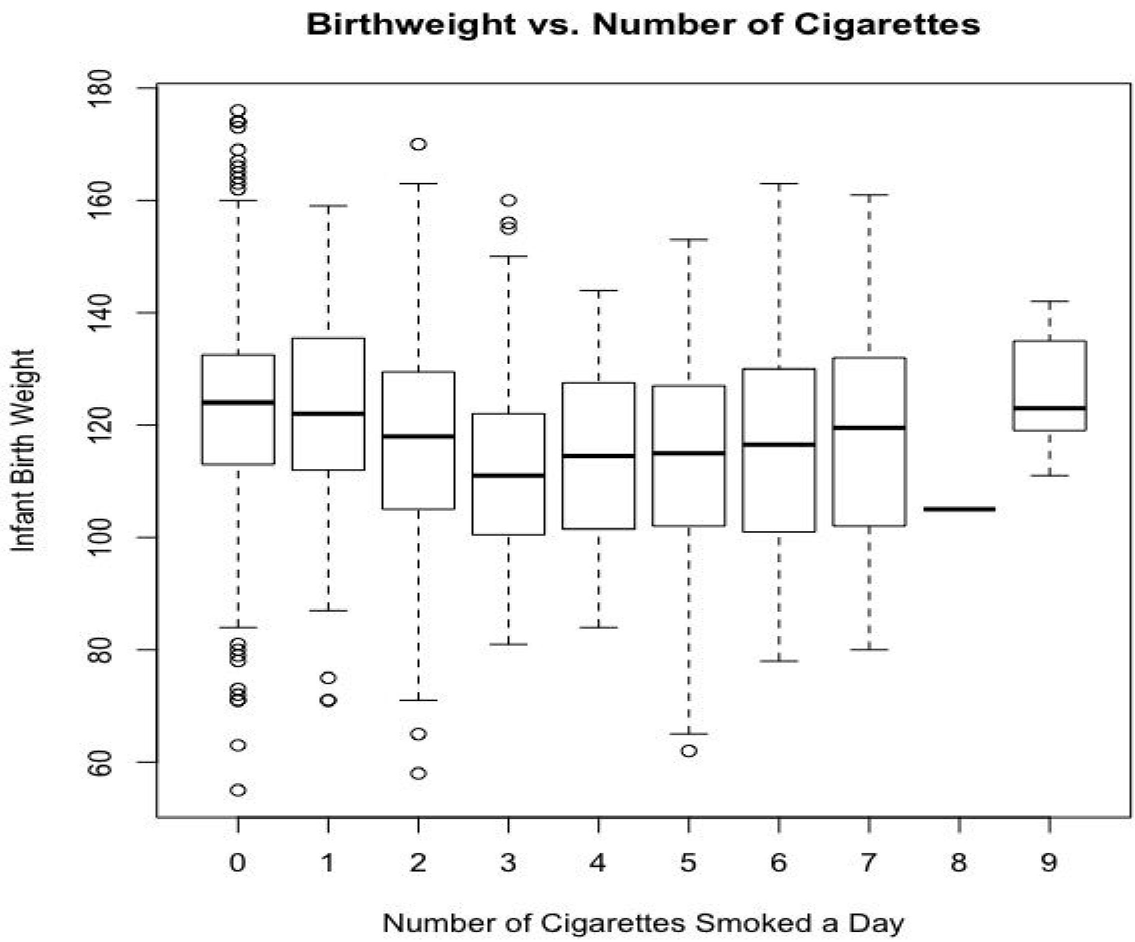


FIGURE 3: *Box Plot of birth weights vs the number of cigarettes a day the mother smoked.*
0 = never smoked; 1 = 1-4; 2 = 5-9; 3 = 10-14; 4 = 15-19; 5 = 20-29; 6 = 30-39; 7 = 40-60; 8 = 60+, 9 = unknown amount

All data points outside the quantiles are outliers, and we see the most of these when x = 0, mothers who never smoked. However, when x=0, the box is relatively small, suggesting smaller variance in birth weights, and a median around 130 oz, in line with the distribution in Figure 1(a). For x = 3-7, mothers who smoked from 10-60 cigarettes a day, the medians range from lower to slightly lower (from as low as 110 oz to around 118 oz, with the median increasing as more cigarettes are smoked. And as the median increases, the boxes grow in size, implicating a wider variance as the more cigarettes are smoked. The plots for x = 1-2 are more or less the same than the x = 3-7, but in the reverse order, although both still show a lower birth weight median than for x=0. There was only one case where an individual smoked 60+ a day, and that median is significantly lower than all other medians. For x=9, the two quantiles of the box are of an uneven size, giving way to the idea of more variability in the upper middle quantile, but less so in the lower middle quantile. Cigarette smoking was a common variable in x = 1...8, and because those medians fell beneath the median of mothers who did not smoke (x=0), it is suggested that an infant's health is affected by the frequency of smoking as well.

Outside of the frequency and amount smoked, we also paid attention to the Body Mass Index (BMI) of the mothers in relation to the infant birth weight.

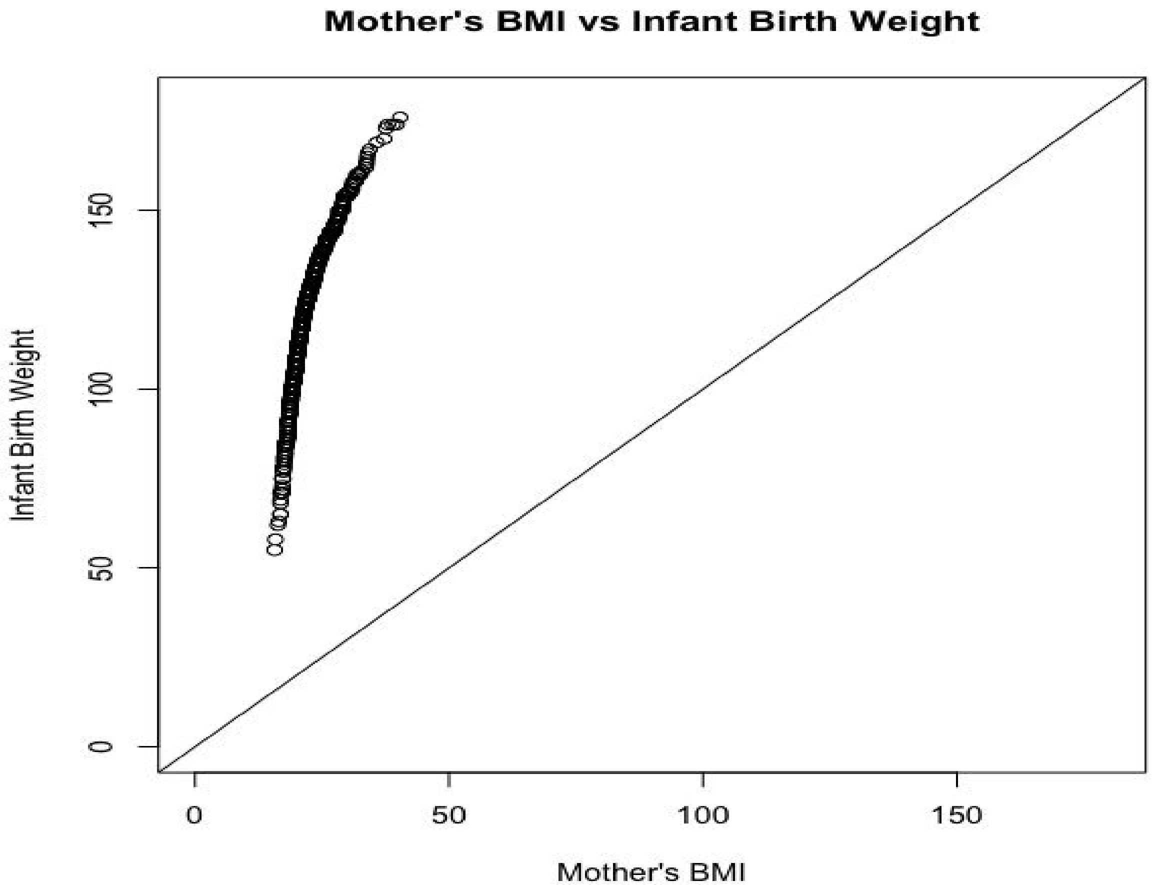


FIGURE 4: *Quantile plot of the mother's BMI and the infant birth weight*

Given the mother's height and weight, we calculated the BMI, and compared it to the weight of the babies at birth. On the plot, the data points are shaped in an upward curve, suggesting a right-skewed data. The skewness of the quantile plot relates to a departure of normality in the data points, in where the two distributions of the samples differ widely. Data points exist entirely above the reference line (y=x). The curve in the middle also implies a long right tail. Overall, the data distribution of the two samples are far from normality, as they exhibit little relation to one another. So from here, we can rule out the mother's BMI as a heavy factor in the weight of the baby, and even its health.

Frequency Analysis:

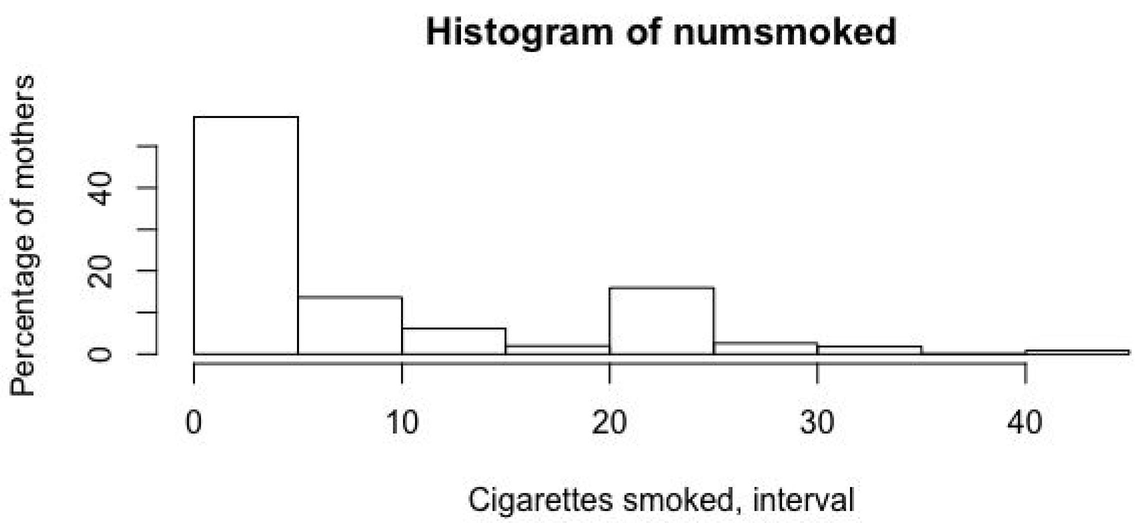


FIGURE 5: *Frequency of number of cigarettes smoked by smoking mothers. Notably, there are two main focal points for the frequency: at about half a pack a day, and one at a pack a day.*

Studying the frequency of number of cigarettes smoked, it was notable that mothers seemed to adhere to two basic patterns of smoking: choosing with greater frequency to smoke either half a pack or a pack of cigarettes a day. This clustering in frequency can be seen in the above histogram by the high percentage of the area contained in the bars corresponding to 5-10 and 20-25 cigarettes per day. Analyzing the difference in frequency of low birthweight babies for each group didn't seem to produce anything significant. So perhaps this behaviour is an interesting quirk to the data which says more about the effect of portion size and marketing on the behaviour of addicts than about the health of the babies from what we have here.

Frequency of birth weights by category, All Mothers

Range (g)	[55,66)	[66,77)	[77,88)	[88,99)	[99,110)	[110,121)	[121,132)	[132,143)	[143,154)	[154,165)	[165,176)
Number of Mothers	6	15	37	83	188	306	306	172	82	31	9

Non-smoking Mothers

Range (g)	[55,66)	[66,77)	[77,88)	[88,99)	[99,110)	[110,121)	[121,132)	[132,143)	[143,154)	[154,165)	[165,176)
Number of Mothers	4	5	13	28	85	187	208	134	56	22	9

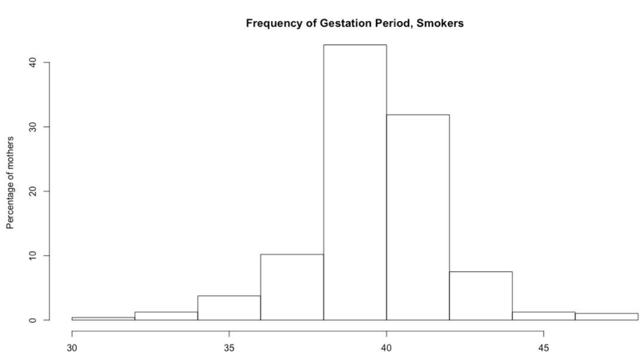
Smoking Mothers

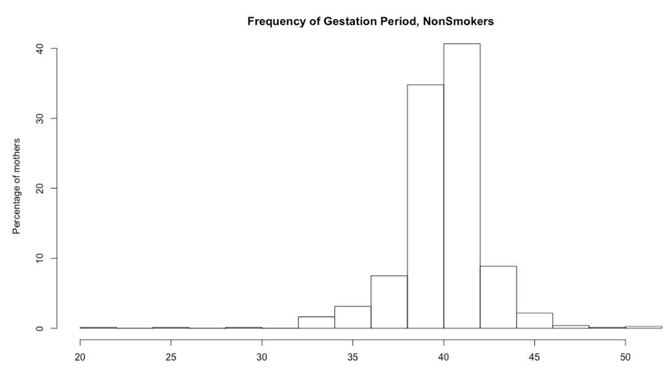
Range (g)	[55,66)	[66,77)	[77,88)	[88,99)	[99,110)	[110,121)	[121,132)	[132,143)	[143,154)	[154,165)	[165,176)
Number of Mothers	2	10	24	55	103	119	98	38	26	9	0

TABLES: 1.2-5: Frequency of birthweight by category, split into 11g intervals, births categorized as underweight are indicated in red

Analyzing the total frequencies of birth weight by category for all mothers, a total of 58 babies were born prematurely in our cleaned dataset. Operating on the hypothesis by that carbon monoxide exposure during the pregnancy itself is influencing the low birth weights, it makes sense to split the mothers into those who responded yes to smoking during pregnancy (smoke value of 1 in dataset “babies23”) and those who did not smoke during pregnancy, whether they once smoked, or quit just before (those who responded 0, 2, or 3). When looking at birthweight frequencies by group, differentiating between smoking and non-smoking mothers as mentioned previously, it becomes clear that a large number of the underweight births are, in fact, occurring in the smoking mothers. For the groups of 11 I used in the tables above, 24 of the 37 total babies (65%) in the “just underweight” [77,88) bracket were from smoking mothers. In the lower two brackets, 10 out of 15 total babies and 2 out of 6 total babies come from the [66,77) and [55,66) brackets respectively. It would seem that babies that are categorized as underweight by this grouping are born more frequently to mothers that smoked during pregnancy than to those that did not, whether they smoked in their lives before, quit at pregnancy or had never smoked.

Further analyzing the tables above, if the standard for underweight births was moved up a bit, perhaps to under 99 grams, then 55 out of the 83 babies in the next bracket, [88,99), or another 66% of the babies that would now be considered underweight by the new standard would have been born to smoking mothers. So, increasing the standard for healthy birth weight would make the effect we’re studying here even more pronounced. Of course, moving standards you initially took as assumptions in ways that happen to better prove one’s point is a slippery slope, but it is also worth considering that we seek to measure the health of the baby, so perhaps the neighborhood of the standard should be considered as well.





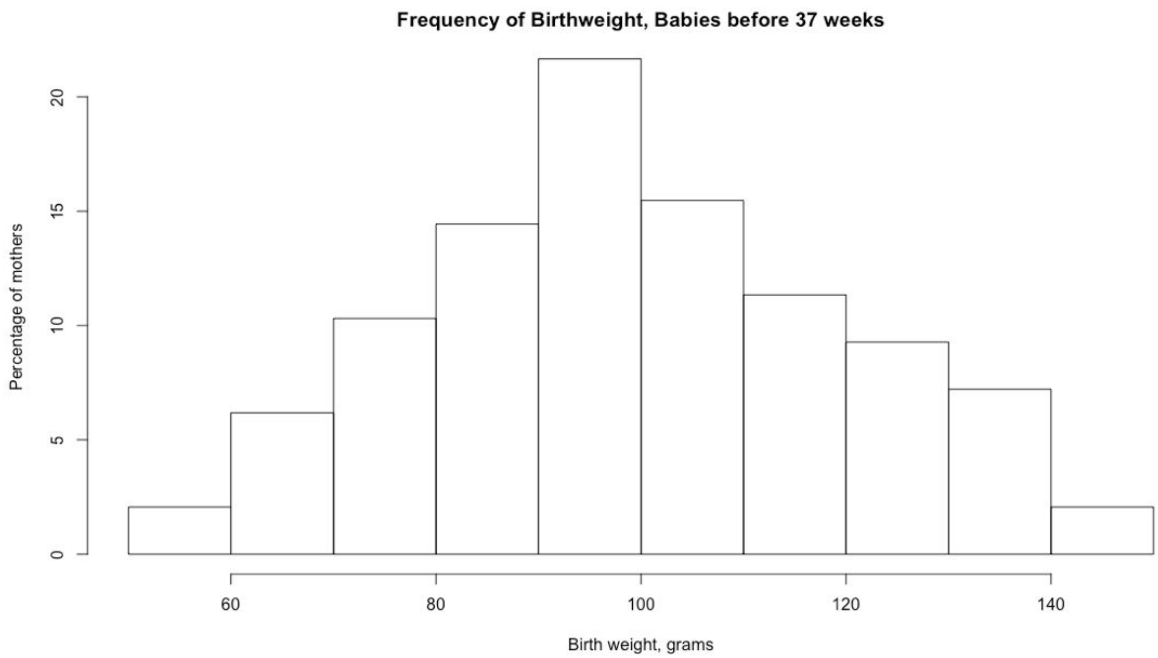


FIGURE 1.3: Histogram of Gestation Period of smokers	**FIGURE 1.4**: Histogram of Gestation Period of non smokers

FIGURE 1.5: Histogram of birth weight amongst babies born prematurely

Operating on the hypothesis that carbon monoxide exposure during pregnancy may be related to early birth, thus birth weight, we examined gestation periods' frequency, again differentiating between smokers and non-smokers as described above. It appears that earlier births occur slightly more frequently among smokers, but the effect could hardly be called pronounced. In short, it is hard to draw much conclusive from the relative frequencies here. Additionally, it can be seen from the next figure that the birth weights of babies born prematurely, before 37 weeks, seem roughly normally distributed, just as all birth weights were, further downplaying the significance of a potential gestation period effect, as considered above.

Conclusion/Discussion:

In our research, we utilized three graphs: histograms, box plots, and quantile plots. All three graphs gave more insight to the distribution of an infant's birth weight when born to a mother that smoked or a mother that did not smoke, and if so, the frequency of how much the mother smoked. Using R, we were able to plot our graphs to better study whether or not smoking affected the health of the baby in terms of the baby's birth weight.

Examining the difference in birth weight between babies born to mothers who smoked during pregnancy and those who did not, we were able to conclude from the data that mothers who smoked during pregnancy give birth to babies classifying as underweight with significantly greater frequency than those babies born to mothers who did not smoke during pregnancy. And further than that, the amount of cigarettes smoked played a key role in determining the varying medians of the baby's birth weight. However, we can conclude that a mother's BMI does not seem to be associated with any predictable change in infant's birth weight.

In determining whether this difference is important to the overall health and development of the baby, we compared both the baby's birthweight and the baby's gestation period for babies born to nonsmoking mothers and smoking mothers. Following from our response to our previous research question, we conclude that this difference is important to the health and development of the baby because babies born too small (low-birthweight) or too early (shorter gestation period) can have serious health problems, including but not limited to damages to the baby's developing lungs and brain.[6]

All of the methods of analysis we attempted on the dataset (numerical, graphical and frequency) seemed to imply some level of negative impact on a statistic that could be interpreted as a proxy for health of the baby. Our analysis would seem to justify, then, in the interest of the health of newborns, the Surgeon General's warning mentioned in the introduction. Also, our group's recommendation to any expectant mother is to refrain from smoking as soon as pregnancy is detected in order to improve likelihood of full gestation period and a healthy birth weight.

Appendix/Appendices:

Appendix A:

In the numerical analysis portion, we also calculated conditional probabilities to gain insight on the question. Keeping the data partitioned by mothers smoking status, we look at only the portion of babies that are low-birth.

TABLE A.1: Probability of low birth status conditioned on smoking category

	Smoke - 0	Smoke - 1	Smoke - 2	Smoke - 3
Low birth	0.014	0.032	0.0024	0.002
Not low birth	0.43	0.36	0.074	0.081
Prob(Smoke - x)	0.444	0.392	0.0764	0.083

From the table we can calculate the conditional probabilities. Let P(x) be the probability of an event "x". Then P(low birth given smoke - 0) = P(low birth and smoke - 0)/P(smoke - 0). Doing that for each birth status/smoke combination we obtain a table of conditional probabilities:

[6] Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014

TABLE A.2: Probability of birth weight status given smoking status

	Smoke - 0	Smoke - 1	Smoke - 2	Smoke - 3
P(low birth given smoke x)	.0313	0.0826	0.032	0.029
P(not low birth given smoke - x)	0.9687	0.917	0.968	0.971

Again categories 0,2 and 3 have similar result and smokers stick out. Here we see that if the mother smoked during pregnancy, there is an approximate 5% increase in having a low birth baby. Although these conditional probabilities seem promising, we would have to run a statistical test to see if the difference in probabilities is significant.

Appendix B:

In the graphical analysis, multiple graphs were used to study the distribution of different variables and samples. Histograms helped in giving an insight to the mean, the standard deviation, and the outliers. Within it, we were able to deduce where different percentages of the sample population lie. To compare the multiple variables simultaneously, we used box plots to discover the location of the different quantiles of the distribution and where the median lies.

Quantile plots were used as well to provide more information on two sample distributions by ordering the data in an ascending order, and then separating it quantiles. The quantiles were determined by the equation: $\Phi(zq) = q$, $0 < q < 1$. The data is compared to the normal distribution, referenced by the y=x line, so if data points were to lie above the line, the it is clear that they depart from normality. Curved graphs also show a sign of bimodal.