Case Study 1: Maternal Smoking and Infant Death

Dukki Hong (A98058412)

Karl Rummler (A12405136)

Nick Roberts (A11705541)

John Diez (A14751991)

Yiwen Li (A13959913)

I. Introduction

New born children are the beginnings of a new generation of workers, philosophers, entrepreneurs, and our legacy. With these important factors in mind, a growing body of literature in science and media focuses on different factors in what helps a baby develop normally, but also what becomes a risk factor for negative outcomes. What we'll be focusing in this paper are how external factors, such as the ones that come from the environment or mother themselves, could influence an infant's typical development. An infant's typical development is usually defined as its general health, birth weight, and future outcomes. It has been found that having abnormal statistics for these variables, especially birth weight, greatly increases the risk of Sudden infant death [1]. We will analyze the potentially harmful risks of smoking in relation a baby's health.

Terms defined:

- *Sudden Infant Death* (SID): The Sudden Death of an infant that is less than a year old.
- *Low Birth Weight (LBW):* Low Birth Weight is defined as infant birth weight of less than 88.1849oz including 88.1496oz (2.5kg up to and including 2.499kg).
- *Gestation:* Gestation is the process of carrying or being carried in the womb between conception and birth. Can measured by days/weeks.
- *Utero:* Utero is a Latin word for "in the womb" or "in the uterus" in Biology.
- *Smoking*: The rate of smoking by the mother, passive smoking from family members.
- *Odds Ratio (OR)*: Odds Ratio is a measure of association between an exposure and an outcome. The OR is the odds that an outcome will happen under a particular exposure as opposed to an outcome happening in the absence of the exposure. If OR=1, then the exposure

does not affect the odds of outcome. If OR<1, then the exposure is associated with lower odds of outcome. If OR>1, then exposure is associated with higher odds of outcome.

- *Early Preterm Delivery*: Early Preterm Delivery is premature birth that occurs before the 35th week of pregnancy (Birth < 35th week).
- *Preterm Delivery*: Preterm Delivery is premature birth that occurs before the 37th week of pregnancy (Birth < 37th week).

II. Background

Previous research have found that having a lower birth weight increases the probability of a child experiencing sudden infant death (SID) [1-6]. In addition, previous research has found that birth weight of infants is positively correlated with maternal smoking. Timur et al. state that the risk of a birth weight under 2500g is 3.8 higher for maternal smokers compared with non-smokers [2-5]. Also very interestingly, women who smoke are 3.9 times more likely to start feeding their baby with supplementary infant foods at 4 months or earlier than non-smoking mothers. Such phenomenon implies the bad nutrition situation of new born children. Additionally, smoking induces effects not only low weight when they birth, but also exhibits more serious long term impacts. Data were collected from healthcare centers located in a city that has 24 primary healthcare centres. The research sample comprises of 700 women randomly selected from the healthcare centers. In this study, some special cases are excluded.

According to a research on mother's education and birth weight, low birth weight is partially a result of decisions made by the mother before and during pregnancy. Furthermore, the British data shows that maternal education is found to be positively correlated with birth weight [13]. This illustrates that having a more educated mother increases the likelihood for a baby to have a normal birth weight compared to children who have a lower value. Another interesting factor that's related to birthweight was the mother's age [6, 14]. A study conducted in New Jersey newborn infants found that younger mothers (15) and mothers older than 40 were at a higher risk for lower birth weight babies compared to 25-36 years olds [14].

Based on a research on 299 Italian mothers from nine different cities who smoked during pregnancy in 1999 and 2000, maternal smoking during pregnancy is strongly associated with preterm delivery [1, 4, 7, 10]. We can define Infant Gestational periods into two categories: Preterm delivery and early preterm delivery. A preterm delivery is when an infant is delivered before the 37th week of pregnancy. Early preterm delivery is when an infant is delivered before the 35th week of pregnancy [4]. Research showed that mothers who smoked during pregnancy were 1.53 times as likely as mothers who did not smoke during pregnancy to have a preterm delivery [12]. Moreover, mothers who smoked during pregnancy were 2.00 times as likely as mothers who did not smoke during pregnancy to have an early preterm delivery.

III. Data

The primary data comes from data collected by the Child Health and Development Studies (CHDS). Data from the CHDS is collected from all pregnancies in Kaiser Health Plan of the San Francisco region between 1960 and 1967. Women included in this dataset were all enrolled in Kaiser Health Plan. Moreover, this dataset drew a line between women who has obtained prenatal care in San Francisco area and delivered in any of the Kaiser hospitals in Northern California other women from other backgrounds. Data is comprised into babies.txt and babies23.txt files. 1236 data points on babies are included. All studies are boys and single births (no twins) and all new born children lived at least 28 days. Two essential variables in this file are birth weight and maternal smoking factor. Babies weight in ounces are numerical, discrete data while smoking is considered as categorical. Level 0 refers to non-smoking; level 1 refers to yes; level 9 means unknown.

IV. Investigations

We mostly use the dataset babies.txt to analyze the relation between the newborn children and the smoking mothers and non-smoking mothers.

Methods

For the section of numerical summary and analysis, we looked into mean, median, maximum, minimum of the four variables in the babies.txt file. In doing so, we have a better idea on how is the

data distributed in big picture and where is the center of the data comparing between smoking and non-smoking group. In addition, the standard deviation and quartile shows how the data is spread.

For the section of graphic analysis, the histogram provides a direct visualization of the shape and the distribution of the data and compares the differences between two variables and its frequency. Boxplot identifies the outliers and quartiles of the two distributions and shows what is the difference between the range explicitly. Moreover, the Q-Q plot compares the normal distribution with our sample distribution, both with non-smoking and smoking mothers.

Numerical Summary of Data

In the table below, we include the minimum, maximum, mean, median, first quartile, third quantile of the babies weights, gestation period, age of mothers, parity of Non-smoking mothers and the babies' birth weights with smoking mothers. First quartile represents the the cutoff point where 25% of data lies below it, and third quartile represents the cutoff point where 75% of data lies below it. In the numerical analysis, we take out the whole row of the data if any column entry has a "999", which means that specific entry of data is unknown.

Table 1 Birth Weight. Analysis *The mean and median of smoking mothers have their babies' birth weight lower than non-smoking mothers. The 1st Quartile and 3rd Quartile also follow the same pattern. Therefore, the numerical summary in data reveals that smoking mothers tend to have lower birth weight infants, but we don't know if it's a significant difference.*

	Min	Max	Mean	Median	Count	1st Quantile	3rd Quantile	Std.
Smoking Mother	58	163	114	115	479	101.2	126.0	18.180
Non-Smoking	55	176	123	123	732	113.0	134.0	17.353

Table 1: Smoking vs Non Smoking Birth Weight (oz) Data

Additional Analysis of Confounding Factors:

Table 2 Gestational Period. Analysis *The mean, median and quartiles of the gestation period of smoking and nonsmoking mothers are similar, so we conclude that there could be no strong relation between gestation period and smoking or nonsmoking mothers.*

	Min	Max	Mean	Median	Count	1st Quantile	3rd Quantile	Std.
Smoking Mother	223	330	277	279	479	271	286	15.087
Non-Smoking	148	353	280	281	732	273	289	16.638

Table 2: Smoking vs Non Smoking Gestation Period (days) Data

Table 3 Age. Analysis *The mean, median and quartiles of the age of smoking and nonsmoking mothers are similar, so we conclude that there could be no relation between the age and smoking or nonsmoking mothers.*

	Min	Max	Mean	Median	Count	1st Quantile	3rd Quantile	Std
Smoking Mother	15	43	26	26	479	22	30	5.654
Non-Smoking	17	45	27	26	732	23	31	5.836

Table 3: Smoking vs Non Smoking Age (years) Data

Table 4 Parity. Analysis*. Parity represents if the mother has carried their baby for >20 weeks. The mean, median and quartiles of the parity of smoking and nonsmoking mothers are similar, then we consider there is no strong relation between parity and smoking or nonsmoking mothers.*

	Min	Max	Mean	Median	Count	1st Quantile	3rd Quantile	Std.
Smoking Mother	0	1	0.253	0	479	0	1	0.435
Non-Smoking	0	1	0.261	0	732	0	1	0.439

Table 4: Smoking vs Non Smoking Parity Data

Additional Statistical Graphic Analysis

Question: Does the Mother's age and Smoking Status influence a baby's Birth Weight?

To replicate the statistic that mother age matters for birth weight, as well as whether it could be due to smoking, we created figure 1 below.The Scatter plot below (Figure 1) shows the relationship between mother's age and infant birth weight for mothers that smoked during pregnancy and those that did not. Infants born from mothers that smoked are shown in red dots and infants born from mothers that did not smoke are shown in blue dots. It is evident from Figure 1 that the majority of infants born from mothers that smoked had a lower weight compared to infants born from mothers that did not smoke. It is also important to note that there is a positive correlation between smoker mother's age and infant's weight. Cluster of red dots can be found near the lower left corner of the plot. This means that infants born from mothers that smoked and had a lower age also had a lower birth weight. On the other hand, more blue dots appear on the upper portion of the plot. This means that despite the age of the mothers that did not smoke, more infants had higher birth weights.

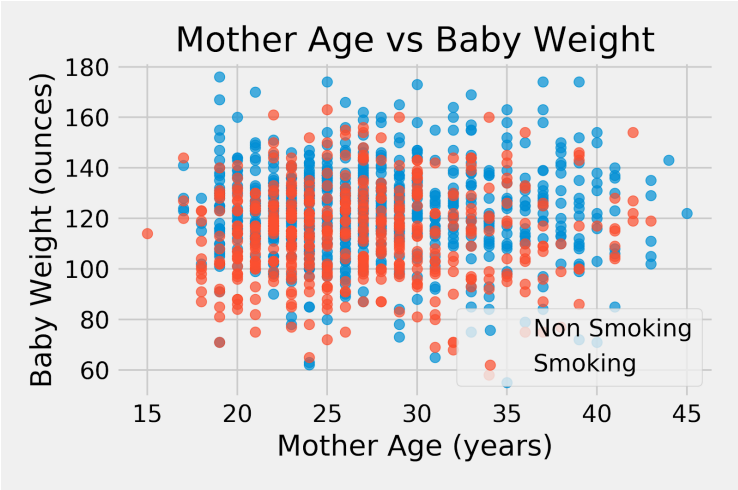

Figure 1: Scatter Plot of mother's age (in years) and Baby's weight (in oz)

Histogram: Comparison between Smokers and Nonsmokers

Question: Does Maternal Smoking affect an Infant's Birthweight?

The Histogram below (Figure 2) shows the standardized birth weight of infants between mothers that smoked during pregnancy and mothers that did not. According to the figure, the mean birth weight of infants born from mothers that smoke (blue dotted line) is greater than the mean birth weight of infants born from mothers that does not smoke (red dotted line). This shows how infants born from mothers that did not smoke generally have higher weights. Comparing the maximum frequency of the

weight of the two groups, infants born from mothers that smoke (colored in red) and infants born from mothers that do not smoke (colored in blue), it is apparent that infants of mothers that does not smoke had a greater maximum (specific numeric values in Table 2). The histogram for the infants born from mothers that did not smoke is slightly left skewed compared to the red histogram and it is also evident that the converse is also true. This shows, at face value, that infants born from mothers that do not smoke will likely to have a higher weight than those born from mothers that smoke.

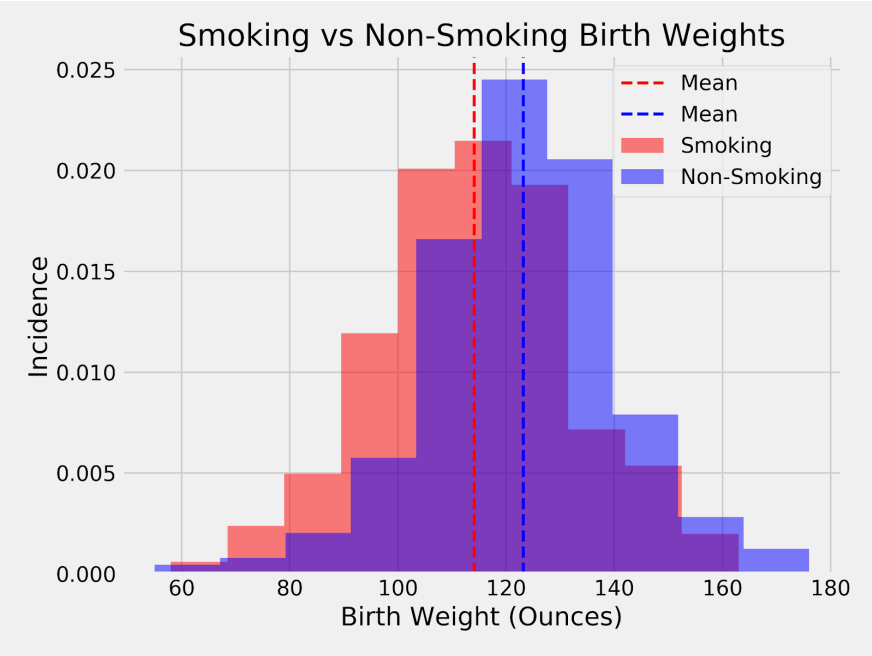

Figure 2: Histogram of the frequency of standardized Birth Weights between Smokers and Non-smokers. Lines represent means for the respective color distribution.

Low-Birth-Weight Incidence Analysis

From the data we see that non-smoking mothers have a low-birth-weight baby at 3.01%, whereas a smoking mother has a low-birth-weight baby at 8.35%. This is a difference of 5.34%, which is significant. However, this value is based on a baby being classified as low-birth-weight if its under 2500g, which is about 88.18oz. Figure 5 below shows that different thresholds for birth weights (+/- 5oz and 10oz) still results with low birth weight incidence being higher with smoking mothers than non-smoking mothers. The higher the birth weight, the higher the difference of smoking rate and non-smoking rate. This suggests that our estimates are reliable.

Birthweight (oz)	78.18	83.18	88.18	93.18	98.18
Smoking Rate - Non-smoking Rate (%)	1.765	2.608	5.345	7.600	12.577

Table 5: Low-birth-weight differences between smoking and non-smoking mothers across multiple thresholds

The Importance of the Differences:

After investigating the relationship between low birth weights and their mother's smoking habits using statistical methods, it's clear that a mother who smokes during pregnancy negatively affects their child's birth weight. We viewed the data three ways: numerical, graphical, and incidence. What we found numerically was that the means of the two groups are different from one another, where non-smoking mothers had higher average birth weights compared to smoking mothers. Graphically, we found were able to assess the spread of the birth weights, which allows us to understand whether the value could be possibly skewed by outliers and whether or not we can conduct further statistical analysis. Incidence rates helps us specifically see how the difference in percentages of when low birth weights occur. We were able to see that smoking mothers had a 5% increase in incidence of infants who have low birth weight. Other studies have found smoking to be a risk factor for low birthweight and sudden infant death [2-5, 11]. The odds ratio for some of those studies tend to be about 2.35, and we notice that the 5.34% increase from 3.01% is actually almost 275% increase or similarly, an odds ratio of 2.75 [3]. This could mean that smoking could really be a major factor in SID as it seems to be related with birth weight strongly. Further comparisons is located in the conclusion.

Boxplot

Question: Does Maternal Smoking affect an Infant's Birthweight?

The boxplot splits the data set into quartiles and helps us visualize the data in a better way to assess the data's spread and pattern. In Figure 3, the median (represented by the Red line) of the babies' birth weight corresponding to smoking mothers is smaller than that corresponding to non-smoking mothers. According to the figure, the box length, known as the distance between third quartile and

first quartile, is larger in smoking mothers' data set. This shows that babies' weight of smoking mothers ranges in lower weight usually and comparatively more spread out than those of non-smoking mothers. Moreover, the two horizontal lines, called whiskers, extend from the box shows the outliers of the non-smoke data. Outliers in the box plot is represented by the dots past each whisker end, which has been logically determined to be any value Q1 - 1.5x(Q3-Q1) and Q3 + 1.5 x (Q3-Q1) . Reading from Figure 3, the upper outlier for the birth weight of babies from non-smoking mothers is slightly greater than that of the birth weight of babies from smoking mothers. On the other hand, the lower outlier of the birth weight of babies from non smoking mothers is a lot higher than that of the birth weight of babies from non-smoking mothers. This shows how the lower and upper bounds of the birth weight of babies born from non-smoking mothers are higher than those of the birth weight of babies from smoking mothers.

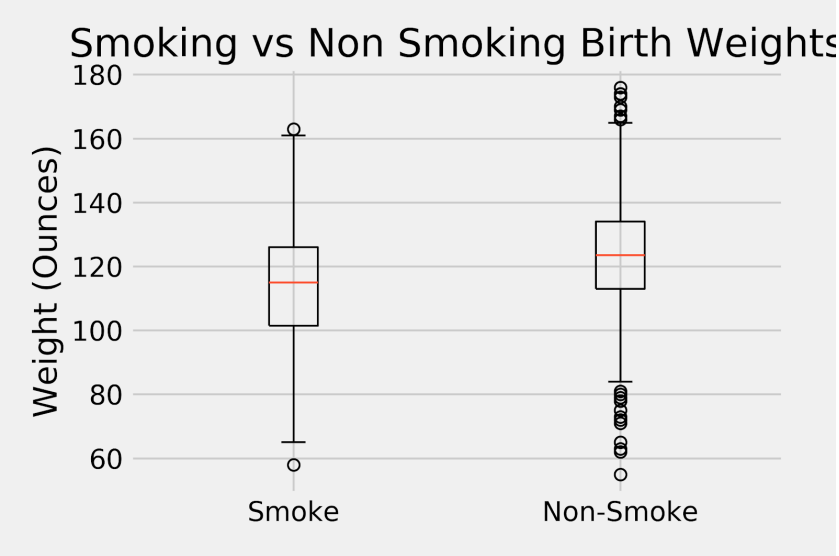

Figure 3: Box plot of Birth Weights of Infants from Smokers and Non-smokers (in oz)

Quartile-Quartile Plot

To conduct our analysis, we need to know if the data is Normally distributed (explained deeper in theory), and a QQ plot helps determine how normal it is. The figure below (Figure 4) shows the QQ-Plot for the birth weight of infants of smoking mothers. It is apparent in Figure 4 that the shape of the distribution of the birth weight of infants of smoking mothers is tending towards a linear line across the graph. This means that the birth weight of infants of smoking mothers is almost normally

distributed, i.e. the Kurtosis (a numeric value that calculates how normal a data is) of Figure 4 is close to a 3 (Kurt ≈ 2.936). Figure 5 shows the QQ-Plot for the birth weight of infants of smoking mothers. Compared to the data from Figure 4, the shape of the distribution in Figure 5 is different. This means that there is more data on the left side of the standard normal distribution, which is equivalent to saying that the distribution is skewed to the left. The Kurtosis value for Figure 5 is far from 3, which was approximately 4.032, evidently proving that the distribution is not normal. To confirm normality, we will be conduct further analyses in the next set of figures.

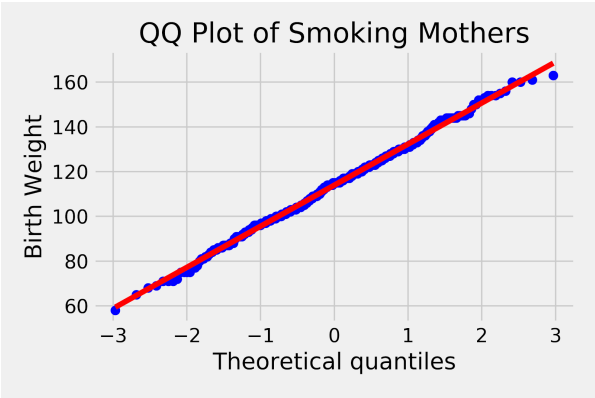

Figure 4: Quantile-Quantile Plot for birth weight of infants of Smoking mothers

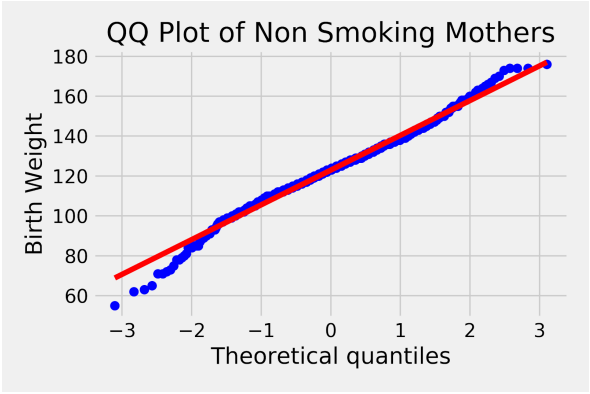

Figure 5: Quantile-Quantile Plot for Non-smoking mothers

Hypothesis Testing:

According to the Central Limit Theorem, given a sufficiently large sample size, the sample mean will approximately equal to the mean of the population. To confirm that the population of smoking and non-smoking mother birth weights are normal, we've conducted a bootstrap to allow us to simulate a population of sample means. To do this, we randomly sampled 2000 samples with replacement from the data set, both non-smoking and smoking mothers, for 1000 times (further details is explained in the *Theory Section*). Read from the bootstrapped histogram below, the average of the birth weights is approximately normally distributed. Doing another analysis of Kurtosis (normal analysis), we found that the non-smoking value is 2.74 (rounded to the hundredth digit) which is close to the value of 3.

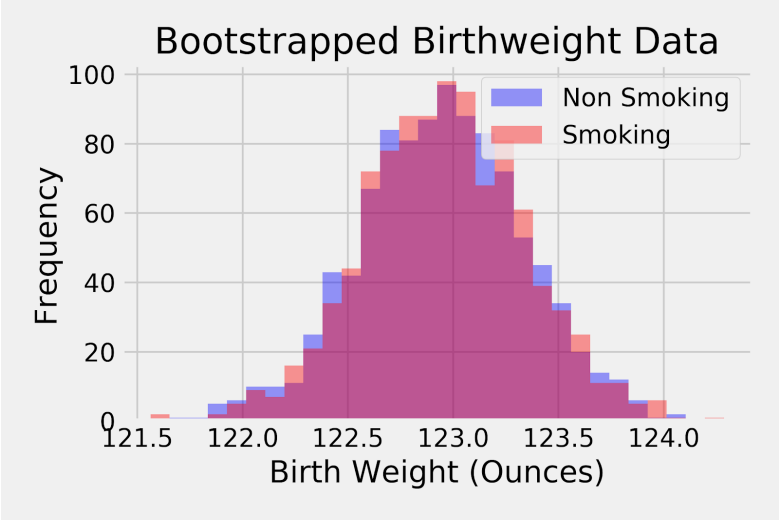

Figure 6: Bootstrapped Birthweight Plot for Non-smoking and smoking mothers

T-Test Analysis

Question: Is the Difference of Birthweight between Smoking and Non-Smoking Mothers statistically significant?

To compare the means of two distributions, we would conduct something called an independent samples t-test. This methodology helps compares two independent sample distribution means and tests whether the difference is statistically significant or have a low probability of actually occurring. To conduct this test, we need to answer two assumptions. The first assumption is normality, which we have confirmed to be true for the population distribution of sample means by utilizing the bootstrap method. Also that the dependent variable is normally distributed. Normality of the distribution of the dependent variable. A kurtosis test shows that this distribution is 3.43 which is close to the 3 value that would mean normality (Figure 10).

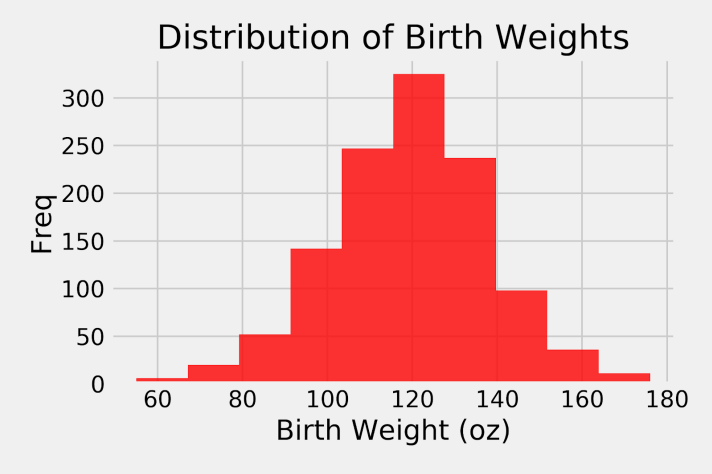

Figure 7: Distribution of Birth Weight (oz)

The second assumption we need to check is the homogeneity of variances or whether the variance or spread between both distributions are the same or not. To test this, we use Levene's test for the homogeneity of variance, which created a resulted in a test statistic of 2.84 which has the p value of .09 which is not statistically significant. This shows that the variances are not that statistically different and we can conduct the independent sample t-test.

Hypothesis testing:

- Null hypothesis - There is no difference between the birth weights between smoking and non-smoking mothers.
- Alternative hypothesis - There is a difference between the birth weights between smoking and non-smoking mothers.

When we conducted the test (without the no-response numeric values), we found a test statistic of $t(732) = -8.65$, $p < .001$. This means that the probability of finding a difference of means this large between the two groups is less than .1%. We have sufficient evidence to reject the null hypothesis that there is no difference between the weights of smoking and non-smoking mothers. Smoking mothers tends to have smaller birth weights than non-smoking mothers.

Additional Question: Does education level affect the birth weight of a baby?

The instigation for this question comes from the potential correlation of a well-educated mother making healthier choices while pregnant and the child's birth weight.

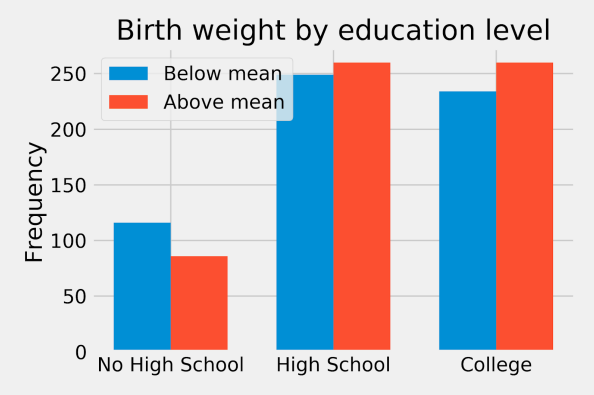

Figure 8: Birth Weight by Mother's Educational Level

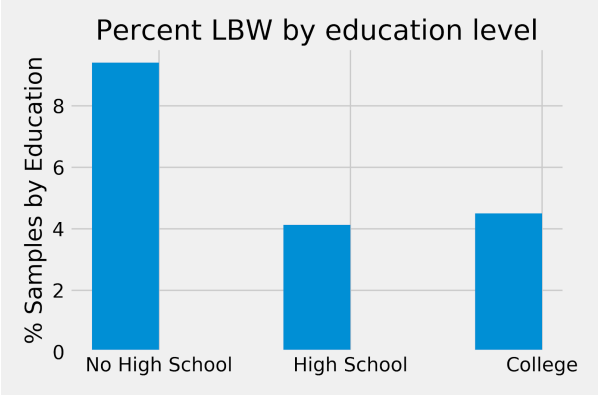

Figure 9: Percent Low Birth Weight vs Mother's Education Level

For figure 8 we are describing how many participants in three different educational levels lie below the mean birth weight. Looking at figure 8, we can see a direct pattern between educational level and the birth weight of the child. As the mother has a higher education, the weight of the child is more likely to be above the mean birth weight than not. This difference increases as the maximum educational level increases as well.

In figure 9, we wanted to see whether or not education level showed a higher frequency of % of low birth weight infants. We see in figure 9, that the percentage of those with no high school education have significantly higher rates of low birth weight than those with at least a high school diploma. Face value, it seems that lower education leads to higher probability of lower birth weights, but the difference is not seen for someone with a highschool diploma or college degree.

	No High School Diploma	High School Diploma	College Degree
Low Birth Weight % (Under 2500g)	9.40	4.12	4.50

Table 6:Maximum Mother's Education and Low Birth Weight %

Additional Question #2: Does gestational Age affect an Infant's Birthweight?

Graphical Summary

Previous Research has linked gestational age (how many days it took until conception) to a lower birth weight for an infant. Specifically smoking has an effect on gestational age, but we wanted to compare whether or not gestational age already influences an infant's birthweight. Figure 10 below represents the mean birth weight for each of the three groups. Previous research has determined what gestational weeks/days are considered very preterm, preterm, and normal (represented under the figure description). Face value, we notice that the birth weight between the three groups steadily increases as we go from preterm to normal. The three different groups have different means. The next few figures will discover whether or not these groups are statistically significant.

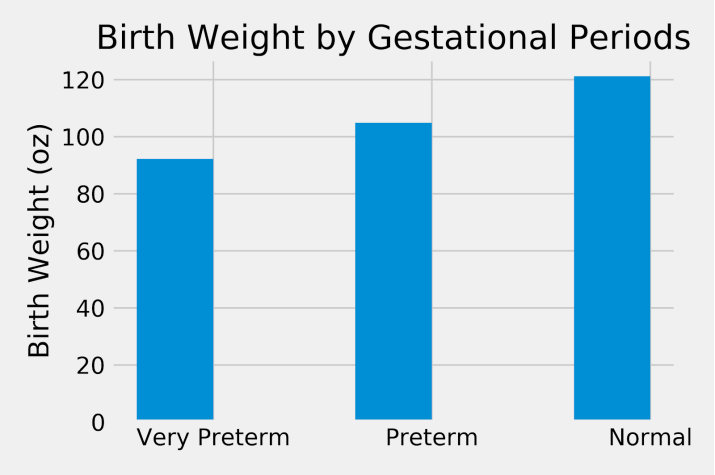

Figure 10: AVG birth weight by gestational period. Very preterm are data points that contain births that occurred <245 days, preterm represents data points between 245-258 days, and normal is >258.

Testing for Statistical Significance:

T-tests usually only compare the means of two groups, but to compare multiple groups with a single Independent variable, then we would need to conduct an ANOVA test. An ANOVA tests helps us understand whether there is a significant difference of the means between two or more groups. To conduct this test, we would need to answer 3 assumptions.

1st assumption:

Normality of the distribution of the dependent variable. A kurtosis test shows that this distribution is 3.43 which is close to the 3 value that would mean normality.

2nd assumption:

Homogeneity of Variance. The variance between the three groups (preterm, very preterm, and normal) have to be equal to do this test. A Levene's test of variance illustrated a difference of p value of .002, which means that the homogeneity of variance assumption is violated. To fix this, we will be conducting a kruskal test.

3rd assumption:

Independence of data. We have to assume that the data was independently collected.

A Kruskall test of the median, a non-parametric test, finds a p-value of <.0001, which means that at least one of the groups is statistically different than the others in terms of mean (Explained more in detail in theory). This means that we can conclude there is a statistically significant difference between the three groups.

Follow up Question: Does Maternal Smoking and Non-Smoking have different Gestational Period?

The research we analyzed talked about how smoking affected gestation period which would should affect birth weight from our previous analysis. However, looking at the graph below we see that their means are pretty close to each other. Face value, by looking at the cutoff points for very-preterm and preterm, there doesn't look to be a large difference between the groups of smoking mothers and gestational period compared to non-smoking test.

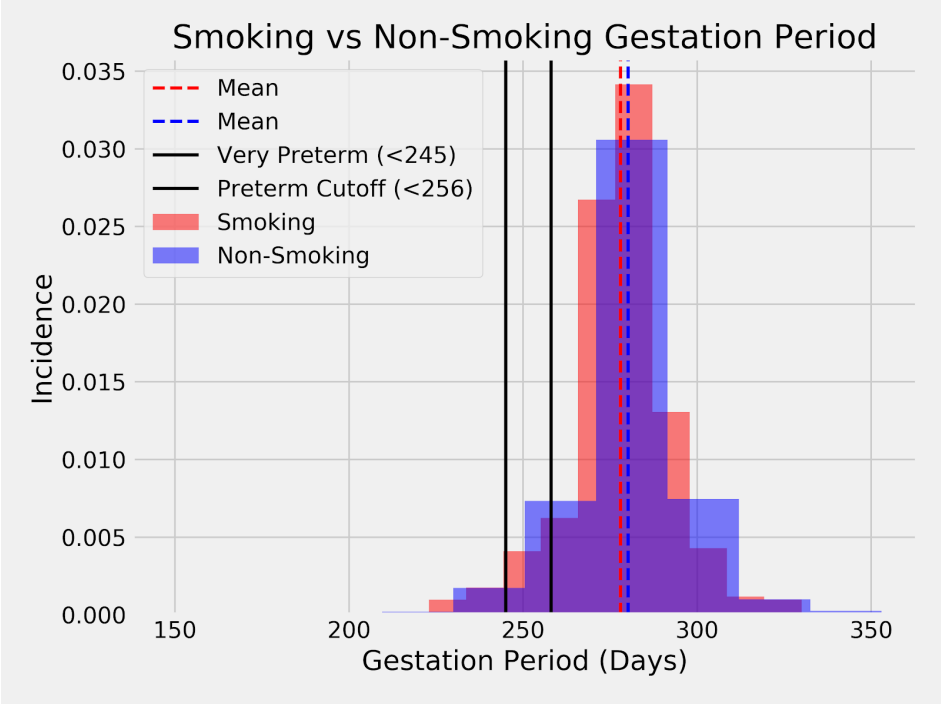

Figure 11: Smoking vs Non-smoking Gestation Period. Also indicated are the data in which either group falls under very preterm or preterm (represented in the black lines). The dotted lines represent the mean gestational period of smoking (red) mothers and non-smoking (blue) mothers.

V. Theory

Numerical

Mean — Mean, also called the arithmetic mean, mathematical expectation or average, is defined as the central value of discrete set of numbers. Consider a set of random variables, $X_1, X_2, \ldots, X_n$, the mean is calculated by the sum of the values divided by the number of variables.

Median — Median is the middle point value, separating the first and second half of a data sample. For example, in the data set {1, 3, 5, 6, 6, 7, 8}, the median is 6. In continuous probability distribution, the median is the value such that a number is equally likely to fall in the dataset above the median and below the median. In statistical analysis, median is commonly used as of measure the skewness of the dataset.

Mode	The mode is the most frequent value in a dataset. It is also the maximum value of probability mass function of a data sample. In the scenario of bootstrapping, mode is the value most likely to be chosen.
Skewness	Skewness is a measure of asymmetry property of the probability distribution by comparing the median with the mean of a data sample. Most of the time, skewness is used to check the normality of distribution. For a standard normal distribution, the skewness coefficient is approximately equal to zero. Skewness is calculated by the third power of the average of the data sample.
Standard Deviation & Variance	Standard deviation is how far away a given value is from the center of the distribution. A standard normal distribution has its mean as zero. Variance is the mean of a square of a random variable minus the square of the mean of a random variable. It is also equal to the square of a standard deviation.
Kurtosis	Kurtosis is a measure of the "tailedness" of a probability distribution of a random variable. Kurtosis is similar to skewness, in terms of describing the shape of the distribution. Kurtosis is calculated by the fourth power of the average of the data sample. Specifically, the kurtosis of any univariate normal distribution is 3. The mathematical definition of Kurtosis is: $$\text{Kurt}[X] = E[(\frac{x-\mu}{\sigma})^4]$$
T-Test	T-test is used to determine the significance in the difference of means between the two groups when there was no difference in population from which the sample was drawn. Null hypothesis denotes that there is no significant difference between the mean of the groups. When the null hypothesis is rejected, the means are statistically different.
Correlation	Correlation is equal to the covariance divided by the product of the standard deviations. It is also denoted as the correlation coefficient, $p_{x,y}$. If two

random variables are independent, the correlation coefficient is equal to zero. The converse is not true, i.e. correlation coefficient of zero does not imply that two random variables are independent. The value of the correlation coefficient ranges from -1 to 1. Correlation coefficient of 1 denotes positive correlation between two variables and correlation coefficient of -1 denotes negative correlation between two variables. The mathematical definition of correlation is:

$$p_{x,y} = corr(X, Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

Graphical

Histogram

Histogram is a direct, accurate way of data visualization and gives an idea on how the data is distributed and the density of the distribution. Usually, a range of values are separated into bins, considered as unit of values. In each interval, the height of the bin describes the frequency density of the range of value.

Boxplot

Boxplot is a method for graphically depicting the groups of numerical data in using quartiles. It has a box lying between the upper and lower quartiles. Outliers are plotted as individual points outside of the box. Boxplot gives an idea on the degree of dispersion, spread and skewness of the data, and also the outliers.

Bootstrap

Bootstrap method is a resampling technique which is used to estimate statistics on a population by sampling a dataset with replacement. Bootstrap is mostly used to estimate mean and standard deviation. Also, it is a very widely used method in machine learning models to make predictions on population when there are not sufficient data provided.

QQ-Plot	Quantile-quantile plot is a probability plot used to compare the similarities and differences between two probability distributions by plotting their quantiles against each other. A Q-Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar between two distribution. A normal Q-Q plot comparing standard normal data with standard normal population, for which the linearity of their relation suggests that the data are normally distributed.
ANOVA	ANOVA stands for analysis of variance. It is a group of statistical models that test the existence of a significant difference between the means. It also tests whether the means of the samples are equal. Null hypothesis means that all groups are random samples from the same population. When a p-value is less than a previously specified value, null hypothesis is rejected, showing a statistically significant result.

Testing Methods

Central Limit Theorem	Central Limit Theorem states that given a large sample size from a finite population variance, the sample mean will approximately equal to the mean of the population. In a more rigorous sense, given n i.i.d. random samples from a distribution with mean μ and standard deviation σ, then the mean of the sum of all the variables will have a normal distribution.
Independence	Two events are independent if the probability of one occurring does not affect the probability of the other occurring. More rigorously, two events are independent if the product of two events is equal to the intersection of the two events.
ANOVA Test	Since T-tests usually only compare the means of two groups, but to compare multiple groups with a single Independent variable, then we would need to conduct an ANOVA test. An ANOVA tests helps us understand whether

there is a significant difference of the means between two or more groups. To conduct this test, we would need to answer 3 assumptions.

1. The samples are independent.
2. Each sample is from a normally distributed population.
3. The population standard deviations of the groups are all equal. This property.

Kruskal Test	Krukal Test is a non-parametric method for testing whether samples originate from the same distribution. This is used when one of the assumptions of the ANOVA test is not met. ● Null hypothesis: Null hypothesis assumes that the samples (groups) are from identical populations. ● Alternative hypothesis: Alternative hypothesis assumes that at least one of the samples (groups) comes from a different population than the others.

VI. Conclusion

After investigating the relationship between low birth weights and their mother's smoking habits using statistical methods, it's clear that a mother who smokes during pregnancy negatively affects their child's birth weight. We conducted a statistical analysis to compare the differences between smoking mother birth weights and non-smoking mother birth weights and found that smoking-mother birth weights were on average smaller. We recommend informing mothers that smoking could increase the risk for a lower birth-weight which also related to negative outcomes, however future studies should directly measure frequency of smoking compared to simply just whether or not the mother smoked. This could be done by measuring the amount of cigarettes a mother smokes a day before, during, and post pregnancy.

Looking at education level and birth weight, we found that educational level does influence a baby's birth weight. We found that mothers with no high school diploma had a higher frequency of lower birth weight babies compared to mothers with a highschool diploma or college degree. We would suggest providing more resources for mothers without a high school diploma, but we should further see if this is a covariate or could be affected by other variables such as socioeconomic status. There could be a relationship with lower socioeconomic status mothers having a higher probability of not having a high education.

When we looked at age, we did find a slight positive correlation with birth weight and mother's age in both smoking and non-smoking mothers. This could suggest that older mothers tend to have normal weighted babies and that mothers should wait longer until they have a baby. However, we do not know if it's a causal relationship, so we would have to replicate the study.

Finally, we analyzed whether or not birth weight averages were different between different gestational periods. We found that the average birth weights between very preterm, preterm, and normal babies were significantly different, where very preterm had the lowest birth weight and normal babies being the highest birth weight. We found no significant difference when we grouped the participants by smoking. We suggest to create specific precautions for babies that are born below the normal baby age threshold.

VII. References

[1] Mathews, T. J., & MacDorman, M. F. (2007). Infant mortality statistics from the 2004 period linked birth/infant death data set. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 55(14), 1-32.

[2] Chase, H. C. (1967). International comparison of perinatal and infant mortality: the United States and six west European countries. Washington, DC, U.S. Government Printing Office, 1967 (Public Health Service Series 3(6), Publication No. 1000).

[3] Chase, H. C. (1969). Infant mortality and weight at birth: 1960 United States birth cohort. American journal of public health, 59: 1618-1628.

[4] Puffer, P. R. & SERRANO, C. V. (1973). Patterns of mortality in childhood: report of the Inter-American investigation of mortality in childhood. Washington, DC, 1973 (Pan American Health Organization Scientific Publication No. 262).

[5] Saugstad, L. F. (1981). Weight of all births and infant mortality. Journal of epidemiology and community health, 35: 185-191.

[6] Mitchell, E. A., Ford, R. P. K., Stewart, A. W., Taylor, B. J., Becroft, D. M. O., Thompson, J. M. D., ... & Roberts, A. P. (1993). Smoking and the sudden infant death syndrome. Pediatrics, 91(5), 893-896.

[7] Anderson, H. R., & Cook, D. G. (1997). Passive smoking and sudden infant death syndrome: review of the epidemiological evidence. Thorax, 52(11), 1003-1009.

[8] Marufu, T. C., Ahankari, A., Coleman, T., & Lewis, S. (2015). Maternal smoking and the risk of still birth: systematic review and meta-analysis. BMC Public Health, 15(1), 239.

[9] Klonoff-Cohen, H. S., Edelstein, S. L., Lefkowitz, E. S., Srinivasan, I. P., Kaegi, D., Chang, J. C., & Wiley, K. J. (1995). The effect of passive smoking and tobacco exposure through breast milk on sudden infant death syndrome. Jama, 273(10), 795-798.

[10] Timur Taşhan, S., Hotun Sahin, N., & Omaç Sönmez, M. (2017). Maternal smoking and newborn sex, birth weight and breastfeeding: A population-based study. The Journal of Maternal-Fetal & Neonatal Medicine, 30(21), 2545-2550.

[11] Bernstein, I. M., Mongeon, J. A., Badger, G. J., Solomon, L., Heil, S. H., & Higgins, S. T. (2005). Maternal smoking and its association with birth weight. Obstetrics & Gynecology, 106(5), 986-991.

[12] Fantuzzi, G., Aggazzotti, G., Righi, E., Facchinetti, F., Bertucci, E., Kanitz, S., ... & Fabiani, L. (2007). Preterm delivery and exposure to active and passive smoking during pregnancy: a case–control study from Italy. Paediatric and perinatal epidemiology, 21(3), 194-200.

[13] Chevalier, Arnaud and O'Sullivan, Vincent (2007). Mother's Education and Birth Weight. IZA Discussion Paper No. 2640.

[14] Reichman, N. E., & Pagnini, D. L. (1997). Maternal age and birth outcomes: data from New Jersey. Family planning perspectives, 268-295.