

## Case Study 1: Birth Weights and Maternal Smoking During Pregnancy

Yu-Chun Chen, A13356506

Yanyu Tao, A13961185

Bolin Yang, A92111272

Shuibenyang Yuan, A14031016

Haoming Zhang, A14012520

Mingxuan Zhang, A13796895

### Introduction:

At the end of 2016, the United States National Institute of Health published an article, claiming that smoking during pregnancy would affect infants' immune system. In fact, debates over the effect of maternal smoking during pregnancy on infants' health have been fierce. Some researchers have proven the causal relationship between maternal smoking and infants' lower birth weight and higher perinatal mortality. In 1972, Butler, Goldstein and Ross conclude that cigarette smoking during pregnancy would increase neonatal mortality by 28% and lower birth weight by 170g, based on surveys of smoking habits conducted on mothers delivering babies within a three-month span and data of births and infant deaths collected in Britain (Butler, Goldstein, Ross). In the same year, John Hopkins Medical Institutions also published its research result of intrauterine growth retardation shown among infants of smoking mothers; babies born to smoking mothers were 250g lighter than those born to non-smoking mothers (Hardy and Mellits). However, the causal relationship, or even the association, of maternal smoking and lower birth weight was also questioned in some later researches. In 1978, research published by University of Oulu, Finland shows that the low birth weights of babies born to smoking mothers have no significant difference from, or even have a higher mean birth weight and lower perinatal mortality than, those born to non-smokers. Yet, babies of smokers do experience higher morbidity up to the age of 5 than their controls (Rantakallio).

The purpose of this research paper, on the one hand, is to investigate the relationship between maternal smoking and infants' birth weight by applying statistical methods on the datasets of the information of approximately 15,000 families' new births collected in Oakland, California from 1960 to 1967. Our hypothesis is that **there is a significant association between lower birth weights and maternal smoking during pregnancy**. In other words, we expect birth weights of babies born to mothers with different smoking habits, either smoke or avoid smoking, during pregnancy has a significant difference. In this paper, the numerical analysis including sample means, variance and quantile and graphical analysis using histograms, box plots, Q-Q plots and statistical moments are elaborated to compare the group of smoking pregnant mothers and its control. Moreover, testing methods including two-sample t-test, chi-square test, and proportion test will be applied. Finally, based on these analyses and tests, the association of maternal smoking during pregnancy and babies' low birth weight will be determined. On the other hand, we also want to use our research results to call for all future or currently pregnant mothers to take care of their own health during pregnancy, especially to avoid smoking, in order to give their babies better health conditions.

## Data:

The data used and analyzed for this project is based on babies23.txt, referencing from the Child Health and Development Studies (CHDS) database. The CHDS data presents pregnancies occurring between 1960 and 1967 of the Kaiser Foundation Health plan in the San Francisco Bay area. The data contains information of approximately 15,000 families' new births.

The data subset includes 1206 women, who had baby boys living at least 28 days. CHDS investigates the relationships between the baby's birth weight and other factors among parents' personal, social, and environmental characteristics. The most helpful factor for finding correlations in this dataset is mother's smoking habit. It also records baby's birth measurements including baby's weight, birth order, and other features.

The raw data contains different variables, and we only need the baby birth weight in this case, so we just extract the baby birth weight column ('wt') from the raw data with correlated categories: non-smoking(0), smoking during pregnancy(1), other smoking (2-3), and unknown (9). Since unknown data may affect our research, we will drop out the unknown smoking category. Information for each variable is detailed below the *Table 1*.

Variable name	Code	Description	Type of data
1. id		identification number	nominal
2. plurality	5	single fetus	categorical
3. outcome	1	live birth that survived at least 28 days,	categorical
4. date		birth date where 1096 = January 1, 1961	categorical
5. gestation		length of gestation in days	numerical
6. sex		infant's sex	categorical
	1	male	
	2	female	
	9	unknown	
7. wt		birth weight in ounces	numerical

	999	Unknown	
8. parity		total number of previous pregnancies including fetal deaths and stillbirths	numerical
	99	unknown	
9. race		mother's race	categorical
	0-5	White	
	6	Mexican	
	7	Black	
	8	Asian	
	9	mixed	
	99	unknown	
10. age		mother's age in years at termination of pregnancy	categorical
	99	unknown	
11. ed		mother's education	categorical
	0	less than 8th grade	
	1	8 <sup>th</sup> – 12 <sup>th</sup> grade, did not graduate	
	2	high school graduate and no other schooling	
	3	high school graduate + trade	
	4	high school graduate + some college	
	5	college graduate	

	6&7	trade school and high school unclear	
	9	unknown	
12. ht		mother's height in inches to the last completed inch	numerical
	99	unknown	
13. wt		mother pre-pregnancy weight in pounds	numerical
	999	unknown	
14. drace		father's race; coding same as mother's race	categorical
15. dage		father's age; coding same as mother's age	categorical
16. ded		father's education; coding same as mother's education	numerical
17. dht		father's height; coding same as for mother's height	numerical
18. dwt		father's weight; coding same as for mother's weight	numerical
19. marital		does mother married?	categorical
	1	married	
	2	legally separated	
	3	divorced	
	4	widowed	
	5	never married	
20. inc		family yearly income in \$2500 increments	categorical

		code separates income by every \$2500	
21. smoke		does mother smoke?	categorical
	0	never	
	1	smokes now	
	2	until current pregnancy	
	3	once did, not now	
	9	unknown	
22. time		If mother quit, how long ago?	categorical
	0	never smoked	
	1	still smokes	
	2	during current preg	
	3	within 1 year	
	4	1 to 2 years ago	
	5	2 to 3 years ago	
	6	3 to 4 years ago	
	7	5 to 9 years ago	
	8	10+ years ago	
	9	quit and don't know	
	98	unknown	
	99	not asked	

23. number		number of cigs smoked per day for past and current smokers	categorical
	0	never	
	1	1-4	
	2	5-9	
	3-8	10-60+	
	9	smoke but don't know	
	98	unknown	
	99	not asked	

*Table 1: Detail Information of Data*

## Background:

Parents' smoking during pregnancy is always a major public health concern. Especially about the influence of nicotine on the developing fetus and the harm associated with it. Before we look at the smoking influence on baby, we have to understand how the fetal grow in a different period in order to fully understand the effect of smoking.

According to the slides in the lecture, the average gestation period for a baby is around 40 weeks. However, there could be some variation to it. Baby born before 37 weeks is called preterm delivery. There are also some cases that babies can stay in the womb up to 42 weeks. After the delivery, the length for newborns ranges from 45 to 55cm and the weight ranges from 5.5 to 8.8 pounds. If a baby with weight lower than 5.5 pounds, it is considered as small. During the previous studies, babies who born too early or with much lower weight tend to be more vulnerable to certain diseases and thus have a higher mortality rate compared to other normal babies. They could potentially suffer from neurodevelopmental disabilities, infection, asphyxia and much more. One of the important things about premature birth is that it could cause lung and breathing problems, which leads to Asthma or Bronchopulmonary dysplasia (also named BPD). Even though some premature birth could survive from the neonatal period, they tend to have delays in physical development, learning, communicating with others, getting along with others and taking care of himself. In the long term, some long-term disabilities could be caused by premature birth, such as behavior problems, attention deficit, and neurological disorders like cerebral palsy that could affect the brain and even the whole body.

There are several factors that can be influential to the fetal growth, including genetic DNAs, nutrition mother that provide to the fetal and also several environmental factors like oxygen and

blood flow. Previous studies show that if fetal are influenced by abnormal factors, newborns tend to have lower weights and heights with higher mortality rate within the first month.

The main influence from cigarette smoking is that the cigarette consists of more than 4,000 chemicals, including cyanide, lead and a lot of compounds that could have potential in resulting in cancers. Among all those harmful chemicals, the two most toxic ones are nicotine and carbon monoxide. When parents smoke during the pregnancy, all those chemicals will be inhaled together with oxygen into mother's body traveling around through blood cells. Thus, there is a potential that those chemicals will be carried towards the umbilical cord, and as a result, be absorbed by the fetal. Therefore, we need to make investigation on up to what degree the cigarettes smoking would affect newborn or even the fetal before born and is it significant that there is a correlation between them. Moreover, if it is significant, what aspects that smoking mothers would influence their babies.

Based on the background and previous studies, we observe that there are controversial conclusions regarding the birth weight and the smoking status of the babies' mothers. In our study, we want to do analysis on the data from our dataset, and see if there is a relationship between birth weights and smoking habits of the babies' mothers, and check if there is another factor from smoking that could affect babies' health and growth.

## Investigation:

To investigate the difference between birth weights of babies born to mothers who smoke during pregnancy and those of babies born to non-smoking pregnant mothers, data is divided into two groups: those with 'smoke' data entries equal to 1, which represents data of babies with mothers who smoked during pregnancy, and those with 'smoke' data entries equal to 0, 2, or 3, which represent data of babies with mothers who did not smoke during pregnancy. The notation or label of "non-smokers" below refers to mothers who avoid smoking during pregnancy and "smokers" only refers to those who smoke during pregnancy instead of general smokers.

### 1. Numerical Comparison

From *Table 2* below, the sample is observed to have included more babies born to non-smoking mothers during pregnancy than babies that are born to mothers smoking during pregnancy. Generally, babies born to smoking pregnant mothers have lower sample mean and max birth weights and also varies a little more within the sample compared to its control. However, the sample of babies born to smoking mothers during pregnancy has higher sample min birth weight. Still, the sample median, 1st quantile, and 3rd quartile of birth weights of babies born to mothers smoking during pregnancy are lower than those of babies born to non-smoking pregnant mothers.

By observing skewness and kurtosis of both distributions of sample of smoking and non-smoking mothers during pregnancy, it can be concluded that both distributions are left-skewed, and that sample of birth weights of babies born to mothers smoking during pregnancy is approximately normal with a kurtosis value of 2.99, while its control is not too close to normally distributed with a kurtosis value of 4.04.

	Babies' Birth Weights born to Non-smoking pregnant mothers (oz)	Babies' Birth Weights born to pregnant smoking mothers (oz)
count	742	484
mean	123.047170	114.109504
std	17.398689	18.098946
min	55.000000	58.000000
25%	113.000000	102.000000
50% (median)	123.000000	115.000000
75%	134.000000	126.000000
max	176.000000	163.000000
skewness	-0.187636	-0.033700
kurtosis	4.037060	2.988032

*Table 2: Summary of the Babies' Birth Weight Born to Non-smoking Mothers and Born to Pregnant Smoking Mothers*

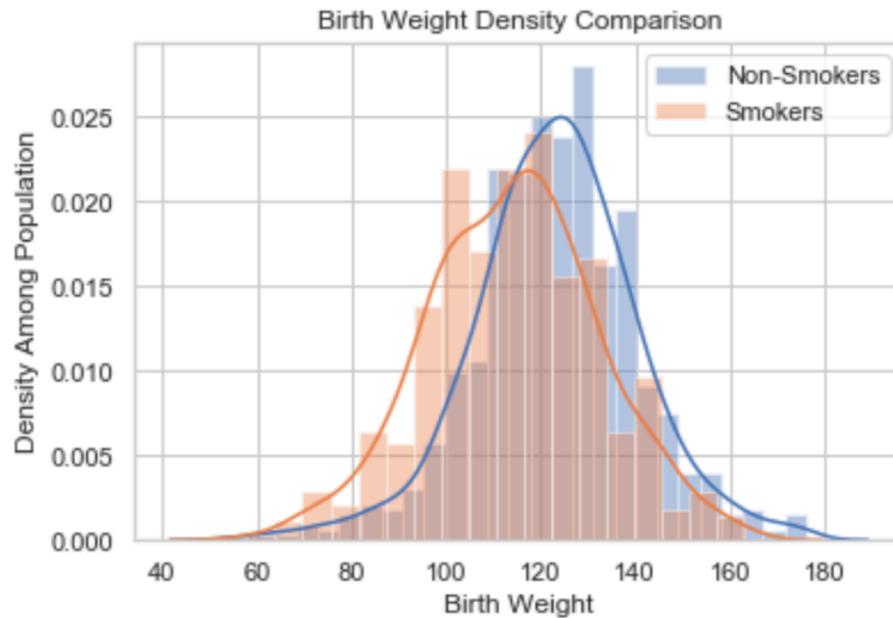
## 2. Graphical Comparison

### Histograms:

As *figure 1* shown below, the orange histogram describes the distribution of birth weights of babies born to mothers who smoke during pregnancy while the blue one describes the distribution of birth weights born to pregnant non-smoking mothers. The orange distribution is roughly bell-shaped, symmetric and bimodal, while the blue distribution is symmetric and unimodal. The center of distribution birth weights of maternal smoking mothers (around 110) is on the left of the one of those of non-smokers (around 130). Moreover, the density value at the center of the orange distribution (around 0.022) is lower than that of the blue distribution (around



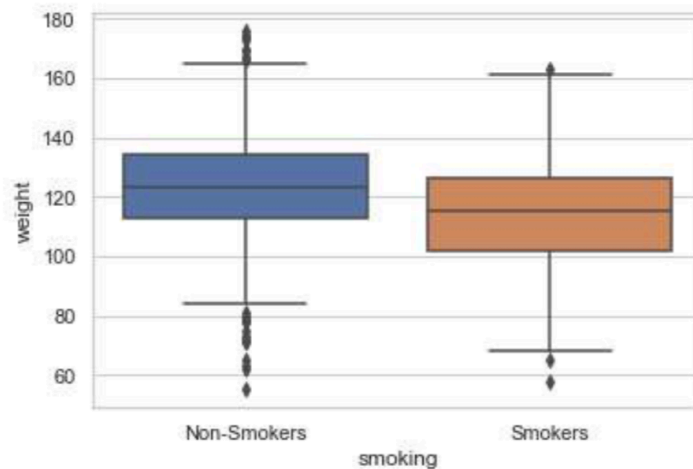
0.025). There are no obvious outliers for both histograms. The spread of smokers is on the left of the spread of non-smokers with similar widths.



*Figure 1: Histograms and Density Curve Comparison between the Babies' Birth Weight Born to Non-smoking Mothers and Born to Pregnant Smoking Mothers*

#### Box plot:

Figure 2 below is a boxplot showing the two distributions of baby birth weight with different mother's smoking habits during pregnancy, either smoking or non-smoking. The weights range from 80 oz to 160 oz for non-smoking mothers during pregnancy, with a relatively long tail at the end. The weights range from 60 oz to 160 oz for smoking mothers during pregnancy, with relatively short tails on both ends. The inter-quartile ranges for both groups lie between 100 oz and 140 oz, with a slightly lower median for the non-smoking group. By comparison, the smoking group has a higher percentage of babies' weights lying below 120 oz, and the non-smoking group has a higher percentage of babies' weights lying above the 120 oz. In general, the whole shape of the smoking group is lower than the non-smoking group.



*Figure 2: Boxplot of Two Distributions*

### QQ-plot:

QQ-plot is a graphical method for comparing two probability distribution by plotting their quantiles against each other. Since several statistical analyses require the assumption of being normally distributed, QQ-plots are used to visualize if the distributions generated from data are close to the normal distribution. The following two curves are the QQ-plots for birth weights of babies for both mothers who smoke during the pregnancy group and mothers who do not smoke during the pregnancy group. The dashed line represents the theoretical normal distribution. For the non-smoking group, except the middle part, two tails deviate from the dashed line a lot, indicating that the data for this group is not normally distributed. In the smoking group, the quartiles lie very close to the real normal distribution with a linear line, indicating that the data for this group is normally distributed. However, the tails for both graphs are quite sparse comparing to the real normal distribution, mainly because there could be not so many outliers in the weights of babies. One of the two groups are not close to the standard normal distribution.

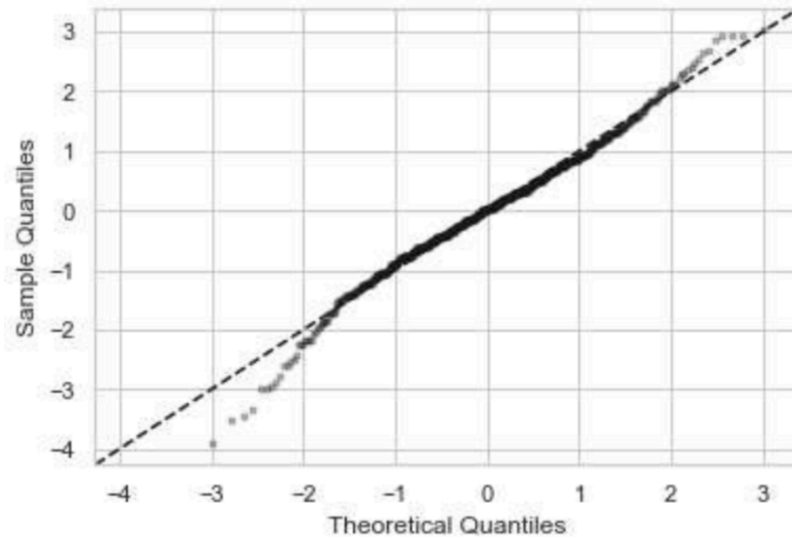


Figure 3: Quantile-Quantile Plot for Non-smoking distribution

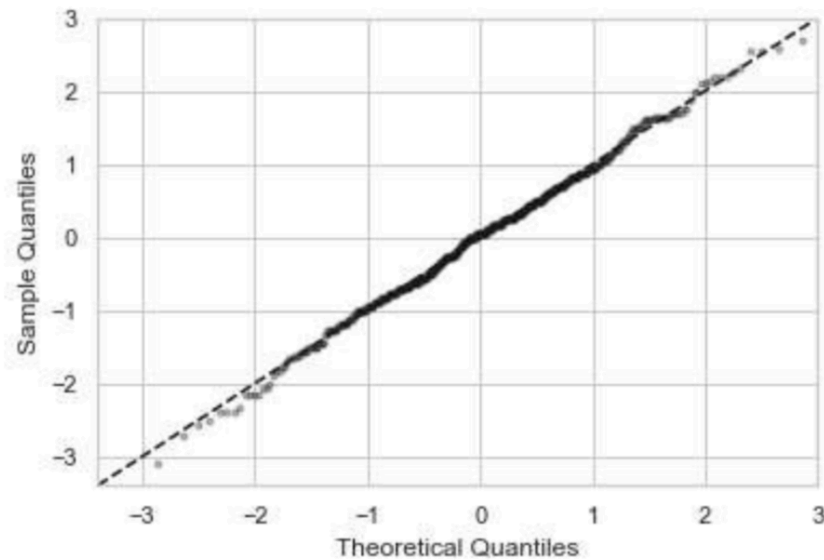


Figure 4: Quantile-Quantile Plot for Smoking distribution

Validating that the distributions are not both normal:

As it is observed above that one distribution, or maybe even both distributions, are not normal, Chi-Square Goodness of Fit test is performed to check the normality of both distributions. The null hypothesis is that the distribution is normal, and the significance level is 0.05 for both tests. P-values and conclusions from both tests are shown below in *Table 3*. The results also verify

our rough analysis according to kurtosis and skewness values and also QQ-plots shown above in *Table 2* and *Figure 3, 4*.

	P-value for Chi-Square GOF Test	Conclusion
Distribution of Birth Weights of Babies Born to Non-Smoking Mothers During Pregnancy	2.50963e-05	The null hypothesis is rejected and that this distribution is not normally distributed.
Distribution of Birth Weights of Babies Born to Smoking Mothers During Pregnancy	0.948874	The null hypothesis cannot be rejected and that we cannot say that this distribution is not normally distributed.

*Table 3: Chi-Square Goodness of Fit Test Results for Both Distributions*

### 3. Incidences:

#### Frequency of Low-Weight Babies:

To find the relationship between the babies' low birth weight and the mothers' smoking habits, the low birth weight standard should be defined. From the research on website, Children's Hospital of Philadelphia suggests that the birth weights less than 2,500 grams (88.1849 ounces) are diagnosed as low birth weights, which is also the low weight cutoff used in *Figure 5* below. Babies weighing less than 1,500 grams (52.91094 ounces) at birth are considered very low birth weight. The following graph shows the frequency of low baby birth weights across the two populations of non-smoking and smoking at pregnancy mothers (Children's Hospital of Philadelphia). We can see that there is a greater proportion of smoker group that has birth weights less the threshold.

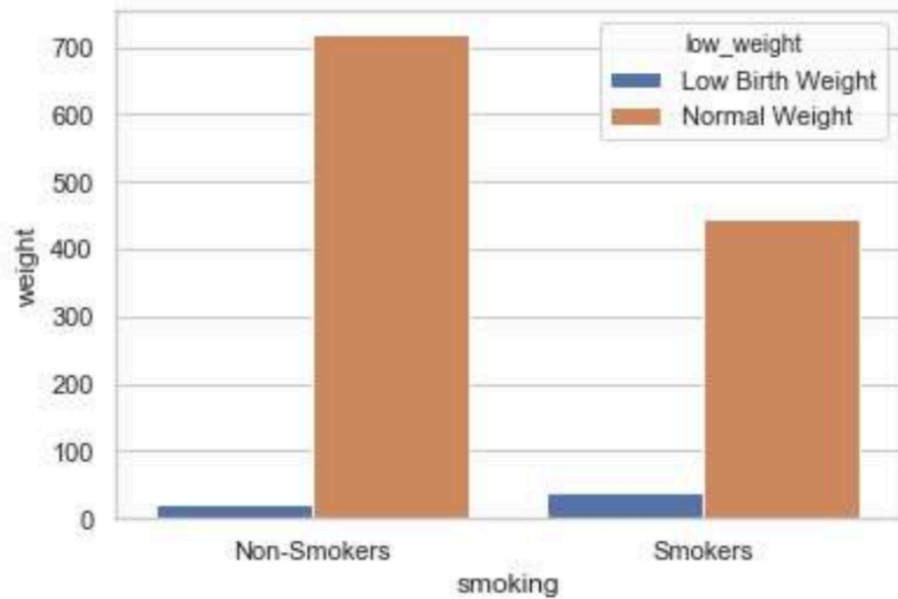


Figure 5: Frequency of babies' having low birth weight in each population

#### Chi-square test for Low Weight Cut-off:

In order to check that 88.2 ounces is a fair standard of babies' low weight cutoff, chi-square test of independence is performed to test the relationship between babies' birth weight and smoking habit. A list of birth weight cut-off points from 75 ounces to 99 ounces for every 5 ounces as threshold is created below. *Table 4* indicates that p-values are smaller than 0.025 between 85 ounces and 99 ounces, which means that the difference between both groups regarding the proportions of low-weight babies is significant with the significance level of 0.05, suggesting that the low-weight standard is within this range and 88.2 ounces a fair cutoff.

Moreover, the p-value difference for the cut-off between 85 and 90 ounces is very small so that choosing other low-weight cutoff from 85 to 90 ounces will not alternate the conclusion. Therefore, 2500g (88.2 ounces) is a suitable low weight standard for this study to investigate the association between mothers' smoking habits and the babies' birth weights given the threshold for low birth weight is between 85 to 99.

Birth Weight Cut-off Point (ounces)	P-value
75	1.483697e-01
80	1.193173e-01

85	3.455917e-02
90	2.999904e-04
95	2.950961e-07
99	2.980443e-11

*Table 4: Chi-square Test results of association between babies' birth weight and mothers' smoking habits during pregnancy under each threshold for low birth weight*

### **Proportion Test:**

In addition to chi-square test of independence, we also perform proportion test on data with low-weight cutoff of 88.18 ounces to confirm that the difference between birth weights of babies born to mothers who avoid smoking during pregnancy and mothers who smoke during pregnancy is significant. The null hypothesis for proportion test is that the proportion of babies with low-weights is not significantly different. Our resulting p-value is 6.23601e-05, which is smaller than 0.025. Thus, we reject the null hypothesis and conclude that the proportions for low-weight babies of both groups are significantly difference with significance level of 0.05.

### **Two-Sample t-Test:**

Since birth weight distribution has sample size over a thousand, and that we have validated above that not both distributions are normal, t-test to test the difference of means is performed instead of z-test. Thus, two-sample t-test is performed to investigate whether there is a statistically significant difference in babies' birth weights across non-smoking mothers during pregnancy and smoking during pregnancy mothers or not. The null hypothesis is that the true difference in means is zero while the alternative hypothesis is that true difference in means is not equal to 0. The outcome of the p-value is 3.524130116-17, which is significantly small and smaller than the significance level 0.01, indicating that the null hypothesis must be rejected. So, the true difference in means is not equal to 0. Therefore, it can be concluded that the birth weights of babies born to mother who smoke during pregnancy is significantly smaller than the babies' birth weights of its control.

### **From Another Angle (additional investigation):**

As it is shown that there is a strong association between low birth weights and mothers' smoking habits during pregnancy, and that mothers who smoke during pregnancy have a higher risk at bearing babies who have lower weights, we become curious about the association between the birth weights of babies and the general smoking habits of mothers. We would like to see if there is a significant difference between birth weights of babies born to mothers who do not smoke at all, and birth weights of babies born to mothers who smoke but avoid smoking during pregnancy. If the difference is significant, then we can conclude that not only smoking during pregnancy has

an association with low birth weights, but smoking regularly, although not during pregnant period, also associate with low birth weights.

To investigate the difference between birth weights of babies born to mothers who smoke but not during pregnancy and those of babies born to non-smoking mothers, data is divided into two groups: those with ‘smoke’ data entries equal to 2 or 3, which represent data of babies with mothers who smoked but avoided smoking during pregnancy, and those with ‘smoke’ data entries equal to 0, which represents data of babies with mothers who did not smoke. The notation or label of “non-smokers” in this part refers to mothers who do not smoke at all and “smokers” only refers to those who smoke, but avoid smoking during pregnancy.

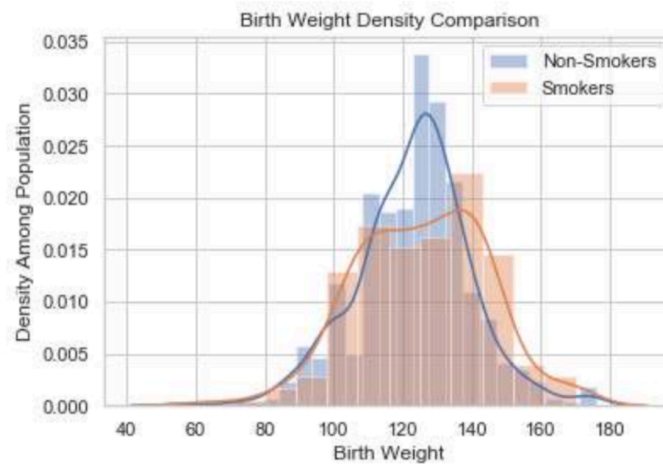
The numerical statistics of both groups are shown in *Table 5* below. From the table, we observe that there are fewer number of mothers in the sample who smoke but avoid smoking during pregnancy than non-smoking mothers. Although the group of birth weights of babies born to non-smoking mothers has lower min and higher max values, and also varies a little less compared to the other group, both groups have similar means, 1<sup>st</sup> quantiles, medians, and 3<sup>rd</sup> quantiles. Both groups have negative values of skewness, indicating that they are both skewed to the left. As the group of mothers who smoke but not during pregnancy have kurtosis value closer to 3, its distribution is closer to a standard normal distribution.

	Babies' Birth Weights born to Non-smoking pregnant mothers (oz)	Babies' Birth Weights born to mothers who smoke but not during pregnancy (oz)
count	544	198
mean	122.777574	123.787879
std	17.109661	18.193139
min	55.000000	62.000000
25%	113.000000	112.000000
50% (median)	124.000000	123.000000
75%	132.250000	137.000000
max	176.000000	170.000000
skewness	-0.121790	-0.353945

kurtosis	4.310783	3.451769
----------	----------	----------

*Table 5: Summary of the Babies' Birth Weight Born to Non-smoking Mothers and Born to Mothers who Smoke but not During Pregnancy*

The histograms and density curves of both groups of birth weights are shown below in *Figure 6*. The smoking group is bimodal while the non-smoking group is unimodal. We observe that although non-smoking group has a higher density at the mean value, both distributions have similar mean birth weight values.



*Figure 6: Histograms and Density Curve Comparison between the Babies' Birth Weight Born to Non-smoking Mothers and Born to Smoking but Avoid Smoking during Pregnancy Mothers*

Two-sample t-test and proportion test are again performed on groups of birth weights of babies born to non-smoking mothers and mothers who smoke but avoid smoking during pregnancy. The resulting p-value for t-test is 0.4972085, which is larger than 0.025, and thus we cannot reject the null hypothesis that the true difference in means is zero with significance level of 0.05. In other words, the difference of birth weights of babies born to non-smoking mothers and mothers who smoke but avoid smoking during pregnancy is not significant enough to reject the two-sample t-test null hypothesis with significance level 0.05. Moreover, the resulting p-value for proportion test is 0.947511, and the null hypothesis that the difference in proportions is zero with significance level 0.05 cannot be rejected. Thus, we conclude that the difference of low-weight babies born to non-smoking mothers and mothers who smoke but avoid smoking during pregnancy is not significant enough to reject the proportion test null hypothesis with significance level 0.05.

From all the comparisons and tests above for birth weights of babies born to non-smoking mothers and mothers who smoke but avoid smoking during pregnancy, it is shown that there is no significant difference between babies' weights whether the mothers smoke regularly or not, given the fact that both kinds of mothers avoid smoking during pregnancy. Therefore, we can conclude that smoking during pregnancy has a stronger association with babies' low weights when compared to smoking habits not during pregnancy.



## Theory:

### Numerical Summaries:

- **Mean:** the “center” of the distributed measurements.
- **Median:** the middle number of the measurements.
- **Quartile:** division the data into groups of data set based on how the values compare to the distribution.
- **Standard Deviation:** the measure that tells how far an individual value may deviate from the center of the distributions.

### Graphical Summaries:

- **Histogram:** A histogram consists of rectangles whose areas represent the percentages of the data lying in the specific interval. It provides an accurate representation of distributions of numerical data and an estimate for the probability distribution of population.
- **Boxplot:** This plot provides an effective means of displaying the range of the data (minimum to maximum), likely range of the data (IQR, from lower quantile to upper quantile) and the typical value (the median). Also, box plot can clearly show the outliers (more than 3IQR away from the box) and suspected outliers (more than 1.5IQR away from the box)
- **Quantile-Quantile Plot:** QQ- plots plot pairs of  $(Z_{k/(n+1)}, X_{(t)})$  and provides a graphical means of comparing two data distributions. When the distributions are identical, the plot should have intercept 0 and slope 1. Here we are actually applying normal quantile plot which compares the data distributions to the normal distribution. If the plotted points point roughly on the line, then it indicates that the data have an approximate normal distribution.
- **3<sup>rd</sup> and 4<sup>th</sup> Moments:** Skewness and kurtosis are both measurements for normality of a distribution. Skewness, the average of the third power of the standardized data, measures the symmetry of a distribution, and will have a value of 0 for a standard normal distribution (which is symmetric), negative if the distribution is left-skewed, and positive if the distribution is right-skewed. Kurtosis, the average of the fourth power of the standardized data on the other hand, measures the heaviness of the tail of the distribution, and will have a value of about 3 for a standard normal distribution.

### Statistical Methods:

- **Chi-Square Test of Independence:** This statistical test helps us to determine if there is a significant relationship between two categorical variables (babies’ birth weight and smoking habit in this study). The test data  $X^2 = \sum [ (O_{r,c} - E_{r,c})^2 / E_{r,c} ]$ , where  $E_{r,c} = (n_r * n_c) / n$  are the expected values, follows the Chi-Square distribution with degrees of freedom  $(r-1)(c-1)$ . Then the p-value from the Chi-Square distribution can tell the relationship between the two variables.
- **Two Sample t-Test:** This test is helpful for questions related to the mean where data is pulled from two random data samples. It is used for comparing the means of two variables or two distinct groups, offering information about whether the means of the two populations differs.
- **Proportion Test:** This test is used to test the difference between two population proportions  $p_1$  and  $p_2$  when a sample is randomly selected from each population. The test statistic is  $\hat{p}_1 - \hat{p}_2$ , and the standardized test statistic is

$$z = \frac{(\hat{p}_{11} - \hat{p}_{12}) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ where } \bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ and } \bar{q} = 1 - \bar{p}.$$

## Discussion & Analysis:

From the numerical statistics, we can observe that the mean, median, and all the quantiles are lower in the group of mothers who smoke during pregnancy than the group of mothers who do not smoke during pregnancy, except for the min value. Thus, the general statistics of the data distribution for birth weights from the smoking group is lower than those from the non-smoking group, which suggests that there is a difference between these two groups.

From histograms and density curve comparison between the two groups, we can tell that the distribution for both groups are different, since density curve is quite far away from each other with a 10 oz difference between two means. Moreover, the box-plot show that the mean, median, 1st quartile, and 3rd quartile are significantly lower in the non-smoking group compared to the smoking group. Thus, we hypothesize that mothers smoking during pregnancy period have some influence on the weights of their babies.

From the QQ-plot, we observe that the data for mothers who do not smoke during pregnancy group does not look similar to the normal distribution. Thus, we use Chi-Square Goodness of Fit test to test whether the data from two groups are normally distributed. The p-value is only 2.50963e-05, indicating that the data of the mothers who do not smoke are not normally distributed. Thus, we need to use two sample t-test to check whether the means for two distributions are the same or not. The p-value for t test is 3.524130116-17, which is significantly small and smaller than the significance level 0.01. Therefore, we reject the null hypothesis that the means for the two distribution are not the same. And the mean for the smokers' group are lower than the non-smokers' group.

From the previous study, we acknowledge that the cut-off line for low weights for babies is 2,500 grams (88.1849 ounces), which means if a baby is born with weights lower than 88.2 ounces, they are considered low weights, which could have a higher risk of having diseases and higher mortality rate. In order to validate this, we use Chi-square Test of Independence for low weight cut-off and found that 88.2 ounces could indeed be treated as the threshold for low weights. Histograms are generated for the two groups, and we observe that there are much more babies falling into the low weight category for the group of mothers who smoke during the pregnancy than for its control group.

To check if it is only due to the mother's smoking during pregnancy that lead to the abnormal birth weights, we also study the data of mother who smoked before but stop smoking during the pregnancy. From the two sample t-test, we observe that the p-value is much higher than 0.05, thus we cannot reject the null hypothesis that the mean birth weights are the same for the mothers who do not smoke entirely and the mothers who smoked before but stop smoking during the pregnancy.

Therefore, all results from the numerical, graphical and statistical evidences suggest that there is a strong association between mothers smoking during their pregnancy and low birth weights of their babies, confirming our hypothesis. In other words, if mothers smoke during pregnancy periods, babies born tend to have lower weights. Linking back to previous studies that

babies with lower birth weights tend to have higher mortality rates, further hypotheses regarding maternal smoking habits during pregnancy, lower birth weights, and mortality rates of infants can be formed. Thus, it is better if mothers avoid smoking during pregnancy period for infants' health condition. In conclusion, although causal relationship between maternal smoking during pregnancy and lower birth weights of babies cannot be concluded in this research due to the fact that it is based on observational studies instead of controlled experiments and several confounders are unable to be eliminated, we can conclude that there is a strong association between lower birth weights and maternal smoking habits during pregnancy. Thus, we recommend that all mothers avoid smoking during pregnancy for better health conditions of their babies.

## Work Cited

Butler, N. R., Goldstein, H., & Ross, E. M. (15 april 1972). CigaretteSmoking inPregnancy: ItsInfluenceon Birth Weight and Perinatal Mortality. *British Medical Journal*,127-130. Retrieved January 24, 2019.

Cigarette smoking during pregnancy linked to changes in baby's immune system. (2016, December 23). Retrieved January 24, 2019, from <https://www.nichd.nih.gov/newsroom/releases/122316-smoking-pregnancy>

Children's Hospital of Philadelphia. (2014, August 24). Low Birthweight. Retrieved from <https://www.chop.edu/conditions-diseases/low-birthweight>

Hardy, J., & Mellits, E. (1972). Does Maternal Smoking During Pregnancy Have A Long-Term Effect On The Child? *The Lancet*,300(7791), 1332-1336. doi:10.1016/s0140-6736(72)92777-8

Long-term health effects of premature birth. (2013, October). Retrieved January 28, 2019, from <https://www.marchofdimes.org/complications/long-term-health-effects-of-premature-birth.aspx>

Rantakallio, P. (1978). The effect of maternal smoking on birth weight and the subsequent health of the child. *Early Human Development*,2(4), 371-382. doi:10.1016/0378-3782(78)90064-6