Maternal Smoking and Infant Death

Austin Clark, Timothy Lee, Andy Schleicher, Tom Tang, Aiden Yoon, Benson Wu

February 2, 2019

Abstract

Method

This primary statistical analysis was conducted using a sample of 1236 infants collected by the Child Health and Development Studies from 1961 to 1962. Information about the infant's parents were also recorded. A supplementary data set from the Center for Disease Control(CDC) and Prevention with observations of infant mortality in 2012 was used for our analysis. We hypothesize that maternal smoking will lead to a higher infant mortality, and that low infant birth weights correlate with a higher mortality rate. We performed our statistical analysis with the following processes/tests:and plots, resampling techniques, t-test and two-proportion Z-test with confidence level to be 0.05. Data was analyzed and visualized with R software.

Results

There was significant statistical evidence that smoking during pregnancy and the incidence of low birth weight are correlated. Our testing shows that both, the average birthweight of children from mothers who smoke is lower than that of children from mothers who do not smoke, and that smoking increases the chance of having what is classified as a "low".

Conclusion

Maternal smoking during pregnancy negatively affects birth weight. Specifically, it results in lower birth rates on average. Mothers should avoid smoking during their pregnancy due to the increased chance of low baby birth weight. Moreover, low baby birth weight can increase the chance of mortality. Thus, smoking will increase the odds of low baby birth weight which will escalate the likelihood of mortality.

Background

For the last half century, occurrences of low baby birth weights have been associated with smoking during pregnancy. This relationship has been studied in depth with a plethora of literature and information available. One significant outcome that resulted from these studies is The Comprehensive Smoking Education Act of 1984 (Public Law 98-474) enacted by Congress. The main consequence of this study reads: "Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight"[6]. Thus, there has been a significant amount of data to force the CDC and the government to warn the public of this relationship and its consequences.

While studying this relationship, we need to keep in mind that there are other factors that contribute to low baby birth weights. In order to "measure" the difference between mothers who smoke while pregnant and mothers who do not smoke while pregnant is to focus on the mean of the babies birth weights[1], the difference between means, the amount of time spent in the gestation period.

Data Set and Processing

Our primary data set consists of 1,236 observations. Observations that had extreme values that were outside the stated range for each variable were assigned to be missing values, as probably intended by the data collectors. Because the primary focus of this analysis was to look at birth weights and gestation periods, we decided to remove any observations that contained missing values in these two fields in order to prevent incorrect analysis, thus reducing the data set to 1,213. We proceeded by separating the data between babies' whose mothers have never smoked, and mothers who smoke or have smoked for our analyses. The following table shows the summary statistics of birth weight and gestation by smoking status of the mother.

Table 1: Birth Weight (oz.)

Smoking Status	Count	Mean	Min	Max	Std. Dev	Skewness	Kurtosis
Non-smoking	544	122.9	55.0	176.0	17.03	-0.09	4.28
Smoking	682	117.0	58.0	170.0	18.70	-0.10	2.97

Table 2: Gestation (Days)

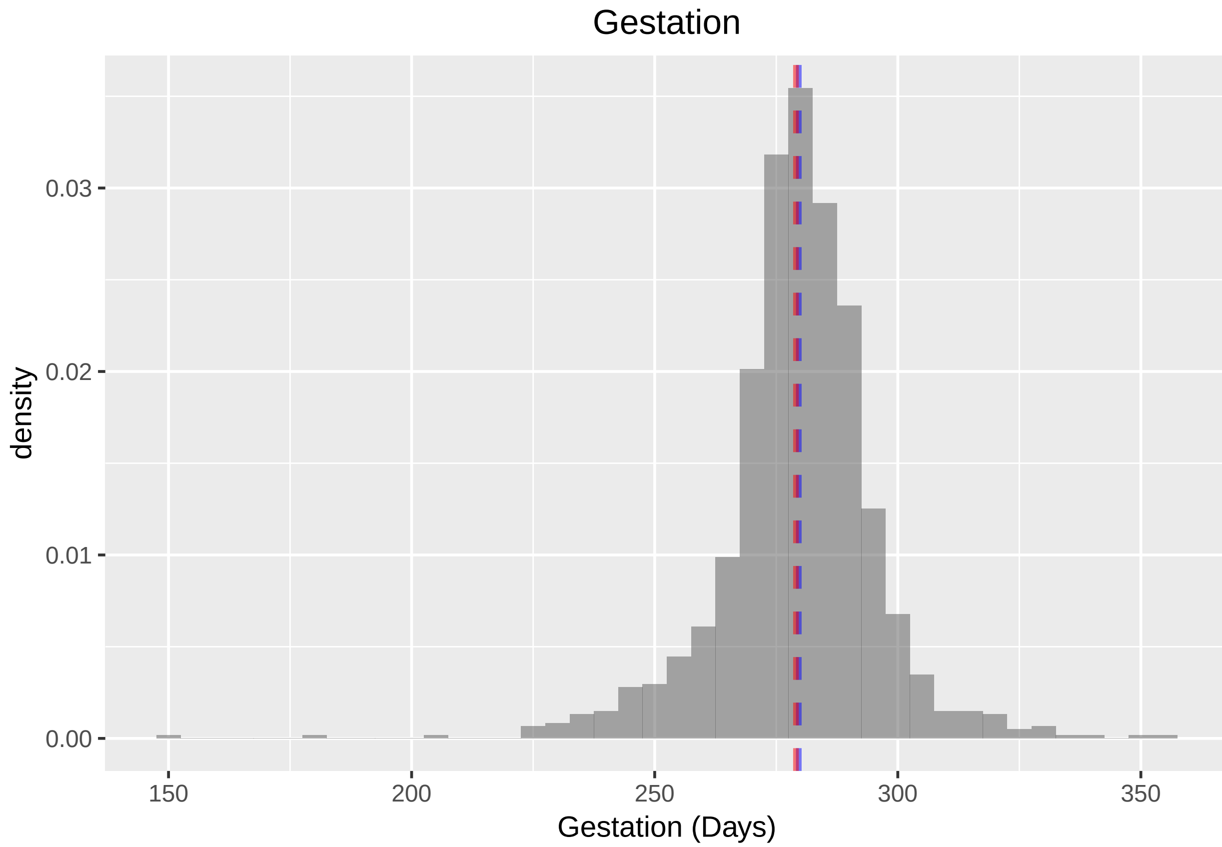
Smoking Status	Count	Mean	Min	Max	Std. Dev	Skewness	Kurtosis
Non-smoking	544	279.6	148.0	353.0	16.60	-1.39	14.02
Smoking	682	279.1	223.0	3515.0	15.63	-0.20	5.15

Another data set that we used consisted of 23401 observations. We followed similar procedures to clean the data set by handling any missing or extreme values; Maternal smoking data was not available, so the only variables used were birth weight and number of days a baby survived. Any row that had a weight category of 14 was removed because these weights were not recorded. Our primary goal of this data set was to use it to see and compare mortality rates based on the weight of babies. Because the weights of the babies were given in grams, we used the basis of 2500g as an "underweight" baby. The weight categories are grouped from 1-13 with the weight ranges indicated on the original data source.

Table 3: Class Counts for Birth Weights (oz.)

227-499	500-749	1000-1249	1250-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	4500-4999	5000-8165
5493	4115	1383	734	652	1492	1976	2653	2917	1417	347	68

Gestation



We first examine gestation due to the minute difference in means between the and smoking and non-smoking groups. As we can see from the density plots, we can see that the distributions of the two groups are very similar

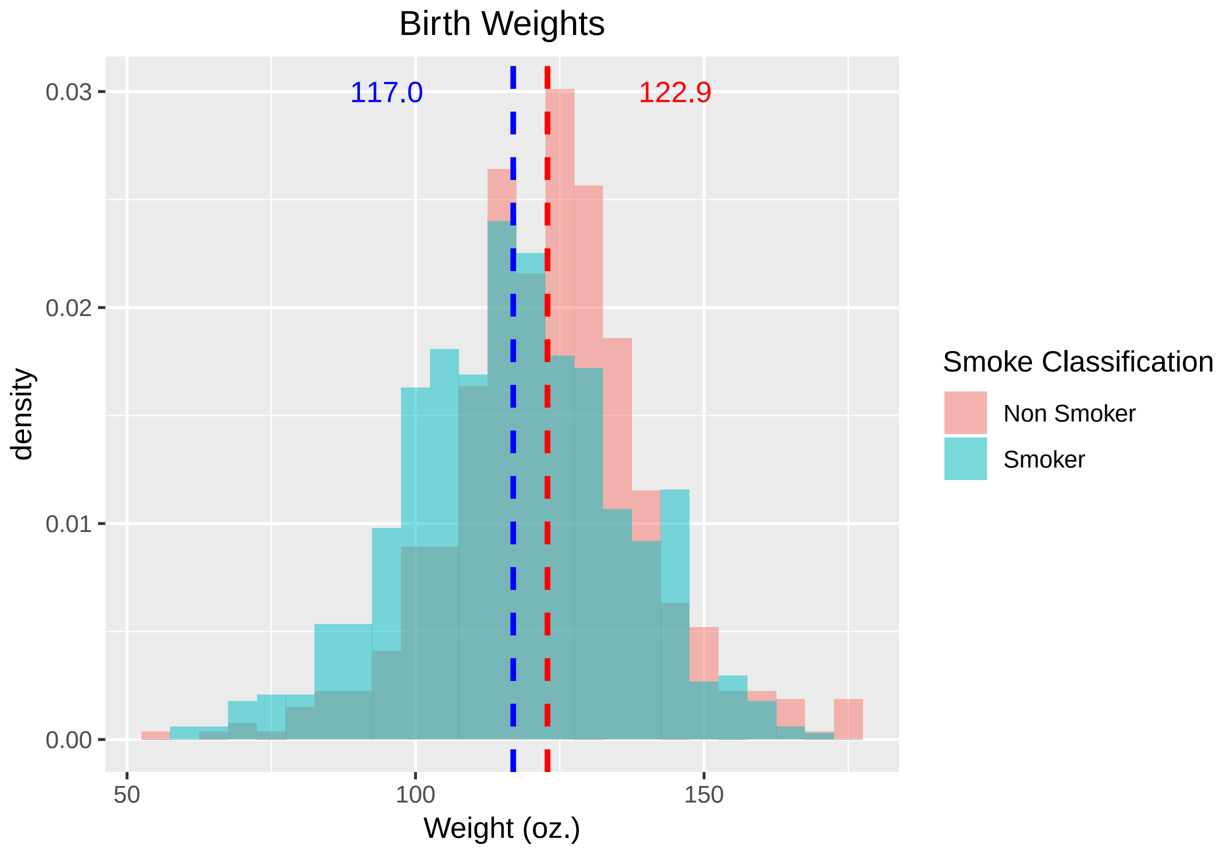
Birth Weights

Looking at the means of the birth weights, it's apparent at first glance that the babies who are affected by maternal smoking versus babies who are not have a lower average birth weight. The birth weights of the non-smoking group also possesses a lower standard deviation, implying that the weights are more centered around the mean.

Table 4: Incidence of Low Baby Birth Weights

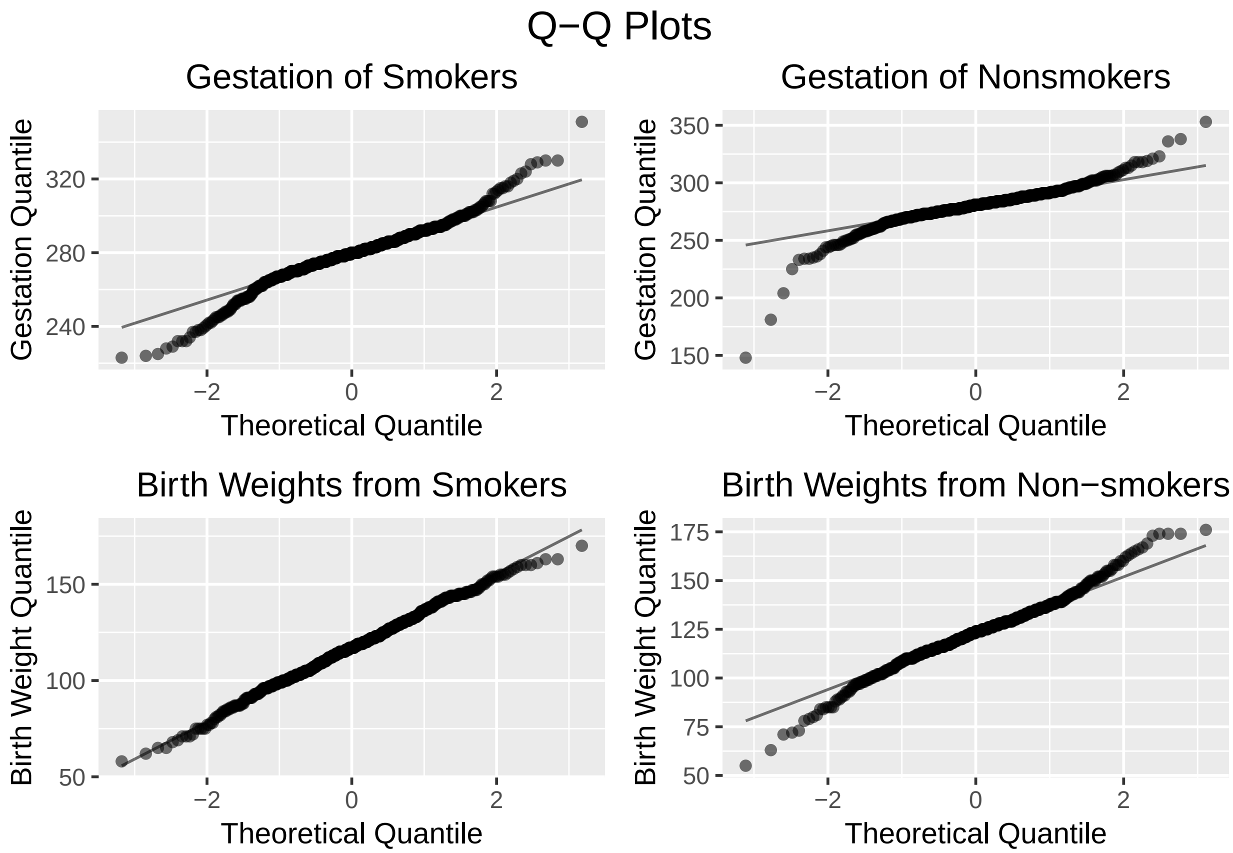
Smoking Status	Incidence	Added	Subtracted
Non-Smoking	0.0297	0.0390	0.0204
Smoking	.0681	.0756	.0607

We also looked at incidence rates for the birth weights of the two groups. For the non-smoking group, the sample proportion of underweight babies was .0297, while the sample proportion of underweight babies for the smoking group was .0681. In order to see if these numbers were reliable, we classified and removed an additional 5 babies for each group; the respective proportions can be seen in the table below. The lack of change in the altered incidences shows that because our sample size for each group is large enough that it will not affect the incidence significantly, showing that our estimates are reliable.



In the density plots of the babies' birth weights for both smoking and non-smoking groups. We note that both plots visually appear to be normally distributed for both smokers and non-smokers. Notice that the mean of the baby weight for smokers is lower than that of non-smokers. This is what we expected and is concurrent with our hypothesis. This conclusion leads to an investigation on whether the distribution of gestation periods and birth weights among the two groups are normally distributed.

Testing for Normality



From the Q-Q plots of the gestation periods, we notice that the tails of the points behave similarly in both the smoker and non-smoker groups such that they both follow trends of normally distributed Q-Q plot. This points to the possibility of there not being as much difference in distribution between the two groups for gestation.

We can also infer that for both smokers and non-smokers, a majority of the data is clumped around the center line and only the extreme values toward the end of the tails seem to trail off. If we look at the Q-Q plot for the smoking group, we can see that the data for smokers is very close to the reference line indicating that it is likely to be a normally distributed dataset. This is strengthened by the fact that skewness -0.10 and kurtosis is 2.97, which is very close to the skewness and kurtosis of a perfect normal distribution. Looking at the non-smoker group, it is harder to conclude that the dataset is normally distributed because the tails deviate away from the reference line, while the center remains on the reference line. The kurtosis turns out to be 4.28, indicating that the distribution has heavy to heavy tails. We believe that this is due to other confounding variables that affect the baby's birth weight that we didn't take account for.

Bootstrap Resampling

We used bootstrapping resampling again to find the bootstrap confidence interval of gestation length birth weights and for the two groups. Bootstrap is a resampling technique where we randomly sampled our data with replacement. We checked that our sample mean was within the bootstrapped confidence interval to ensure that our observed data set can be assumed to be normally distributed.

Table 5: Incidence of Low Baby Birth Weights

Smoking Status	Mean Birth Weight	90% CI
Non-Smoking	0.0297	0.0390
Smoking	.0681	.0756

The mean of the smokers group's birth weight, 116.93 ounces, sits within the 90% bootstrap confidence interval of [115.72, 118.12]. This gives more evidence that the smoker group's birth weights are normally distributed. The same is shown by non-smoker's baby birth weight with a sample mean of 122.87 ounces and a 90% bootstrap confidence interval of [121.6643, 124.0821]. This is consistent with what we expected to result from a normal distribution. We will proceed with the assumption that our data of smoker and non-smoker weights come from a normal distribution. This allows us to use perform t-tests and Z-tests with the assumption of normal data.

Statistical analysis

We first looked at the effect of smoking on gestation periods to find out if smoking results in shorter gestation periods which can be a cause for low birth weights. We used a t-test to test for a difference in means for gestation periods.

> H_0: Mean gestation period from smoking mothers is equal to the mean gestation period from non-smoking mothers
>
> H_1: Mean gestation period from smoking mothers is less than the mean gestation period from non-smoking mothers

Our data did not find significant statistical difference in the mean gestation period of smokers and non-smokers, as the resulting p-value was 0.5355. For the rest of our analysis, we continued by looking primarily at birth weights.

We continued by examining the birth weights. In order to see if there was statistical significance with babies' weights and maternal smoking, we wanted to examine the proportions of underweight babies from both the smoking and non-smoking groups. Under the assumption that both our samples come from a normal distribution, we proceed to use a t-test to determine if the difference in means is statistically significant.

> H_0: Mean birth way from smoking mothers is equal to the mean birth weight from non-smoking mothers
>
> H_1: Mean birth weights from smoking mothers is less than the mean birth weight from non-smoking mothers

The test results in an extremely low p-value of 4 x 10-9 so we reject the null hypothesis in favor of the alternative. We conclude that the difference in means of the two populations is in fact statistically significant and that the mean birth weights from smoking mothers is in fact less than the mean birth weights from non-smoking mothers.

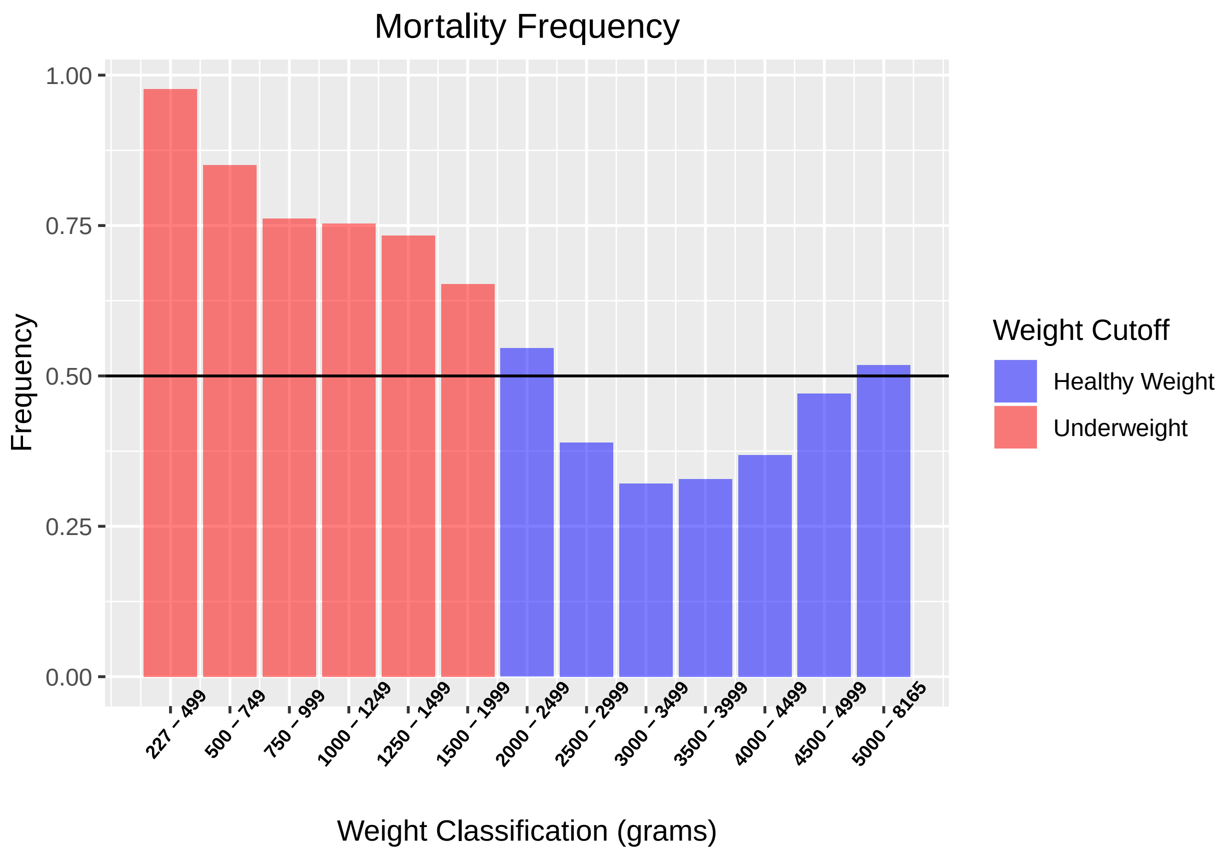
Furthermore, we wanted to determine if smoking leads to a higher incidence of what is considered a "low" (less than 88 oz.)5 birth weight. We proceeded with a two proportion Z-test to see if the proportion of babies born from smokers with a "low" birth weight (0.0681) is higher than that of the non-smokers (0.0297). Our hypothesis was the following:

> H_0: Proportion of "low" birth weight is equal between the smoking mothers and non-smoking mothers
>
> H_1: Proportion of "low" birth weights is higher for smoking mothers than non-smoking mothers

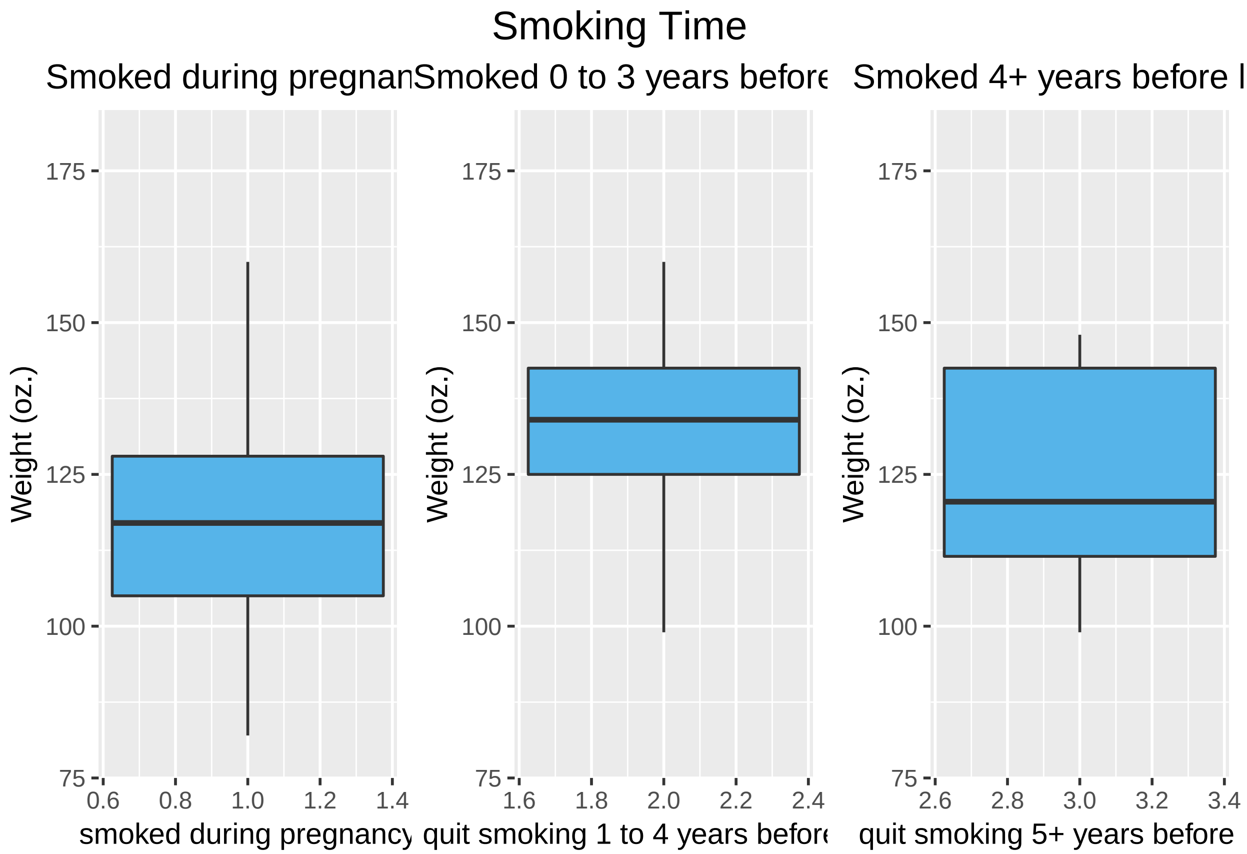
The two proportion Z test yielded a p value of 0.0019, thus indicating that the proportion of underweight babies for smoking mothers is statistically greater than underweight babies for non-smoking mothers.

Birth Weight and Mortality

Our next step was to determine whether or not there is a relationship between a baby's weight and its mortality. Because the smoking dataset capped off a baby's survival at 28 days, we kept our data sets consistent by counting any mortality earlier than 28 days as a baby who has died. In order to examine a possible relationship, we decided to look at the proportion of classified mortalities based on the weight of the babies. The results show that babies who are classified as underweight (Classifications 1-7) tend to have a higher frequency of mortalities before 28 days. Babies who had "lower birth rates" had over a 50% mortality rate before 28 days. This shows that there is a correlation between mortality and a baby being underweight.



Confounding Variables



As we can see from the boxplot of birth weights by smoking history, we can see that the birth weights decline as the time that the mother smokes is closer to the pregnancy.

Mother's Weight

Another possible confounding variable is the weight of the mothers during pregnancy. It is commonly known that smoking is an appetite suppressant, and there is a correlation in mothers weight and baby's weight. We tested to see if on average smoking mothers weighed less than non-smoking mothers. We ran a two sample t-test to test for a difference in means. The average weight for non-smoking mothers was 129.76 ounces, and the average weight for smoking mothers was 127.55 ounces. Our hypothesis was the following:

H_0: Mean weight of non-smoking and smoking mothers are equal

H_1: Mean weight of smoking is less than non-smoking mothers

We obtained a p-value of .04479, leading us to reject the null hypothesis to agree that there is statistical evidence that smoking mothers on average are of lower weight than non-smoking mothers. We note that the p-value is close enough to failing to reject that this warrants further study in future experiments. Since the data is so close together in this case, we do not believe the mother's weight was statistically different enough to heavily impact the weight of the babies but do believe that relationship should be investigated further.

Conclusion

We have shown that smoking and incidence of low baby birth weights appear to be correlated. By separating the data by whether or not the mother smokes, we see that the average birth weight for babies born from actively smoking mothers is significantly lower than that of the birth weight of babies from mothers who do not smoke at all. We have attempted to eliminate confounding variables so that the significant difference in baby birth weight can only be explained by the smoking status of the mother. Furthermore, we have shown that smoking mothers have a much higher chance of having what is classified as a "low" baby birth weight. Future experiments should be replicated with datasets that contain a diversity of geographic location in different states in America and even countries to show that these results in this experiment have external validity. Since having a low baby birth rate in itself is a health risk we agree with the Surgeon General's warning that smoking negatively impacts the health of newborn babies and can lead to higher mortality rates.

Sources

Source 1: Woolbright, L. (1994). The Effects of Maternal Smoking on Infant Health. Population Research and Policy Review, 13(3), 327-339. Retrieved from http://www.jstor.org/stable/40229739

Source 2: Simpson, R., & Smith, N. (1986). Maternal Smoking and Low Birthweight: Implications for Antenatal Care. Journal of Epidemiology and Community Health (1979-), 40(3), 223-227. Retrieved from http://www.jstor.org/stable/25566651

Source 3: MacDorman, M., & Mathews, T. (2009). The Challenge of Infant Mortality: Have We Reached a Plateau? Public Health Reports (1974-), 124(5), 670-681. Retrieved from http://www.jstor.org/stable/25682291

Source 4: Rush, D., & Cassano, P. (1983). Relationship of Cigarette Smoking and Social Class to Birth Weight and Perinatal Mortality among All Births in Britain, 5-11 April 1970. Journal of Epidemiology and Community Health (1979-), 37(4), 249-255. Retrieved from http://www.jstor.org/stable/25566423

Source 5: https://www.chop.edu/conditions-diseases/low-birthweight (birth weight reference)

Source 6: Smoking & Tobacco Use. (2015, July 21). Retrieved January 29, 2019, from https://www.cdc.gov/tobacco/data_statistics/sgr/2000/highlights/labels/index.htm

Source 7: Cigarette Smoke Exposure Reprograms the Hypothalamic Neuropeptide Y Axis to Promote Weight Loss(2006, March 6). Retrieved January 29, 2019. https://www.atsjournals.org/doi/abs/10.1164/rccm.200506-977OC