Math 189 - Case Study #1

Effect of Maternal Smoking on Baby Health

February 2nd, 2019

Name	PID
Erin Werner	A12612584
Emma Choi	A12635909
Talal Alqadi	A13816618
Ella Lucas	A13557332
Samantha De La Torre	A13300273

I. Introduction

Epidemiological studies have indicated that smoking mothers are twice as likely as non-smoking mothers to have a low birth weight baby (under 2500g) and babies born smaller have lower survival rates. The purpose of this study is to determine if there is a difference between the birth weight of babies born to smoking mothers and the birth weight of babies born to non-smoking mothers. Additionally, the study aims to identify whether or not a difference in birth weight will affect the health of the baby. The Child Health Development Studies (CHID) obtained this data from women enrolled in the Kaiser Health Plan in Oakland, California from 1960 to 1967.

II. Data

This study utilizes a subset of 1,236 babies out of all the pregnancies from the CHID data. All of the babies in this subset were singular children (not twins), male, and lived for at least 28 days.

Table 2.1. Dataframe variables description.

Variable	Unit	Denotation	Description
Birth Weight	Ounces	bwt	Infant's weight at birth.
Gestation	Days	gestation	Number of days the fetus is in the mother's womb.
Parity	NA	parity	0 - First born child to the mother.
Age	Years	age	Mother's age.
Height	Inches	height	Mother's height.
Weight	Pounds	weight	Mother's pre-pregnancy weight.
Smoke	NA	smoke	0 - The mother does not smoke.

Table 2.2. Numerical Summary Statistics of Smokers

Smokers	bwt	gestation	parity	age	height	weight	smoke
Min.	58	223	0	15	53	87	1
1st Qu.	102	271	0	22	63	112	1
Median	115	279	0	26	64	125	1
Mean	114.11	277.99	0.25	26.73	64.09	126.87	1
3rd Qu.	126	286	0.25	30	66	135.25	1
Max	163	330	1	43	72	215	1

Table 2.3. Numerical Summary Statistics of Non-smokers

Non-smokers	bwt	gestation	parity	age	height	weight	smoke
Min.	55	148	0	17	56	89	0
1st Qu.	113	273	0	23	62	115	0
Median	123	281	0	27	64	127	0
Mean	123.05	280.17	0.26	27.56	64.02	129.60	0
3rd Qu.	134	289	1	31	66	140	0
Max	176	353	1	45	71	250	0

Any unknown value for a variable is represented by either "9", "99", or "999". Unknown smoking behavior was removed from the dataframe, as the purpose of our project is to determine the impact of smoking on a baby's birth-weight. For columns representing gestation, age, height, and weight, the unknown values were replaced with the mean value for the respective column.

III. Background

Fetal Development

Fetal development is influenced by the mother's behavioral characteristics, such as smoking during gestation. To better contextualize the data summarized in the section above, supporting information will be provided on general fetal development. Upon conception, the placenta begins to form and will serve to nourish the baby. An embryo is the fertilized egg and sperm cell in the morula stage, prior to attaching to the mothers uterine wall. A fetus is the next progression of development where extra embryonic structures have emerged. At 4 weeks, the fetus is actively developing structures that will form its face and neck. Internally the heart, blood vessels, lungs, stomach, and liver develop. By 24 weeks, the fetus' organs are fully formed. Finally, the newborn infant (or neonate) is born.

The typical gestation period is 40 weeks long, however some remain in utero up to 42 weeks. Conversely, preterm delivery is categorized when a baby is born before 37 weeks. A 28 week old fetus weighs about 4 to 5 lbs (64/80 oz) and is about 40 cm long. At 32 weeks, a fetus weighs about 5 to 5.5 lbs (80/88 oz) and is about 45 cm long. In the final weeks, the baby gains about 0.2 lbs (3.2 oz) per week. The average newborn ranges from 45 to 55 cm in length and from 5.5 to 8.8 lbs (88/140.8 oz). Babies born at term that weigh less than 5.5 lbs (88 oz) are considered small for their gestational age.

Effects of Smoking Cigarettes on Fetal Development

Within the context of this study it is important to empirically question how maternal smoking affects another variable from a high-level standpoint, considering available literature and alternative studies to inform our analysis. Generally speaking, health studies have presented a relationship between smoking while pregnant and fetal developmental issues including injury, premature birth, and low birth weight.

Another correlation between weight and infant health came out of a 2003 study claiming that smoking during pregnancy increases the risk of the infant developing asthma. Furthermore, it was found that low birthweight and preterm delivery increases the risk of asthma later in life (at 7 years old) whereas simply being underweight did not (Jaakkola et al.). A 1980 study also designed to tackle the hypothesis that smoking during pregnancy causes legitimate (and measurable) harm to the fetus put forth that the habit led to pre and postnatal growth retardation, placental mortality, and behavioral anomalies (Abel). Placental abruption is another possible outcome in cases where the mother smokes during pregnancy. Carbon monoxide contained within cigarette smoke is thought to reduce oxygen supply to the fetus. While concrete psychological effects of decreased oxygen supply on fetal development aren't completely understood, a steady supply of oxygen is critical for development. It is thought that the placenta attempts to compensate for decreased oxygen supply by increasing its surface area and number of blood vessels. The consequences of depriving the placenta from oxygen include possible placental abruption and placenta previa. Placental abruption occurs when the placenta breaks away from the uterine wall and results in preterm delivery and fatal death.

Research on the effects of smoking during pregnancy in the 20th century led to institutional social awareness, however not immediate eradication of the behavior. Starting in 1995, a Surgeon General warning was circulated to inform the public of the risks associated with smoking while pregnant. Regardless, it was reported that 15% of pregnant women smoked during pregnancy in 1996. Investigation into available literature surrounding smoking during pregnancy showed frequent correlation with the behavior and fetal health complications - which reinforces that a study demonstrating otherwise merits further analysis (see 'Theory' section).

Confounding Variables

In the context of this study, confounders may be defined as possible items that can mask the effect of smoking on the infants death rate. Historically (prior to the 1940s) the public was unaware that a disease carried by a pregnant woman could affect her baby. This began to change in 1941 when Dr. Norman Gregg observed an "unusual" number of infants with congenital cataracts. It was found that those mothers had contracted Rubella within the first or second month of their pregnancies. In this context Rubella exists as a confound for our study, considering that our data is from 1960-1967 and the MMR vaccine was not licensed for use until 1971.

Attributes of the mother that could affect her pregnancy such as age, environment, health record, and socioeconomic status should be considered as possible confounds. One scientific study examined the "most common" causes of stillbirth based on autopsy and placental examination, and serves as a nice generalization of aforementioned confounding factors. The study examined the causes of 132 stillbirths in Italy from 2000-2004 and a reasonable cause was found in 99 out of 124 pregnancies. The major conditions had to do with placental disruptions, abnormalities, and insufficiencies (Bonetti et al.). Since placental abruption has a known association with carbon monoxide from smoking, it is a possible cause of stillbirths when the mother smoked. However, the study on stillbirths demonstrates that this notion cannot be generalized due to the volume of babies who died due to placental complications even in non-smoking mothers.

Is the Difference Important?

The study seeks to determine if different birth weights of babies born to smokers versus nonsmokers could be attributed to differences in the health and development of those babies. Death rates are grouped in four different ways - fetal (death before birth), neonatal (first 28 days after birth), perinatal (combined fetal and neonatal) and infant. Death at different stages of development can indicate a relation to different medical issues. It has been shown that babies born to smokers weigh less than babies born to non-smokers. From previous studies, it has been shown that low birth weights have been correlated with an increase in death rates. Conversely, in our study of interest, babies of smokers had much lower death rates than babies of nonsmokers (rates not adjusted for confounders).

When considering the validity of the surprising results that babies of smokers had lower death rates than babies of non-smokers, it should be noted that the study did not consider possible confounds. To compare, similar studies that did take into account confounding variables can be considered, such as the findings of Wilcox and Russell. When comparing mortality rates of smokers and non-smokers, groupings were handled by sorting babies by their gestational age or relative birth weight (Wilcox et al.). These researchers normalized their data for the cohort that it belonged to, so that comparisons could be fairly drawn. Their conclusion indicated that the mortality rate of smokers was higher.

IV. Investigation

Alternative Explanation

Studies have shown that women who smoke during pregnancy are at an increased risk of fetal intervention, leading to increased rates of cesarean sections (Lurie 2014). Due to the premature removal of the baby via cesarean section, the resultant birth weight is lower than that of a baby born on time. However, it is unclear whether smoking induces low baby weight endangering the baby and requiring a cesarean section, or smoking causes complications leading to the need for a cesarean section early prompting a low birth weight. While the implications of smoking on birth weight are unclear, Lurie's 2014 study showed that smoking increased the odds of an instrumental or cesarean delivery by 24% and it was often due to a non-reassuring fetal heart pattern rather than low birth weight. The evidence suggesting a correlation between smoking, mode of delivery, and birth weight will be explored (see 'Theory: Alternative Hypothesis' section) in an attempt to clarify the relationship between smoking during pregnancy and birth weight.

Methodologies Comparison

This study used many statistical methods to compare the birth weights of babies from non-smoking mothers and smoking mothers to determine whether or not there was a difference. The numerical statistics provide a simple and basic numerical representation of the center and spread of the data. Whereas, the graphical summaries, including the boxplots and histograms reflect the whole distribution of data, rather than just summary statistics. In addition to seeing the center and the spread, a graphical summary depicts the symmetry, modality, and tails, which is informative of where the majority of the data points lie. The Kurtosis and Quantile-Quantile plots both serve to compare the data to normalized data sets. However the Kurtosis method plots each histogram separately and compares the skewness of both the non-smoking and smoking mothers to a normalized skewness. Whereas, the Quantile-Quantile plots for the smoker and non-smoker data is plotted linearly on the same graph along with the normalized line. Each statistical method elicits a different benefit. Therefore it is advantageous to utilize all of them to increase the credibility of the study.

V. Theory

Numerical Statistics

Table 5.1. Babies birth-weight distribution without unknown values (i.e. "Smoke" == 9).

babies$bwt	1st Qu.	Median	Mean	3rd Qu.	Std. Dev.
Smokers	102	115	114.11	126	18.10
Non-smokers	113	123	123.05	134	17.40

From Table 5.1, it is shown that the infants of nonsmokers tend to weigh more than those of smokers. This is because both the mean and median birth-weights are greater on average. Yet, the standard deviation and interquartile ranges are similar, demonstrating that both smoking and nonsmoking infants have a similar spreads and distributions. Because the mean and median values are almost equivalent for each group, the data will be symmetric.

Histograms

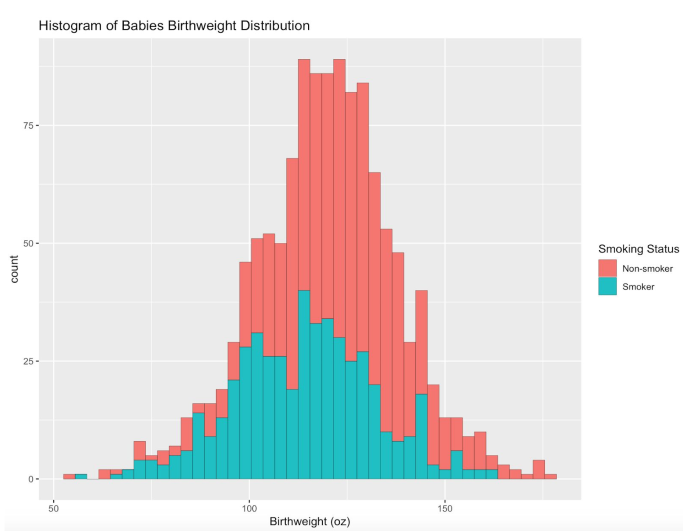

Figure 5.1. Overall distribution of babies birth weight (oz), as distinguished by mothers that were classified as smokers versus non-smokers.

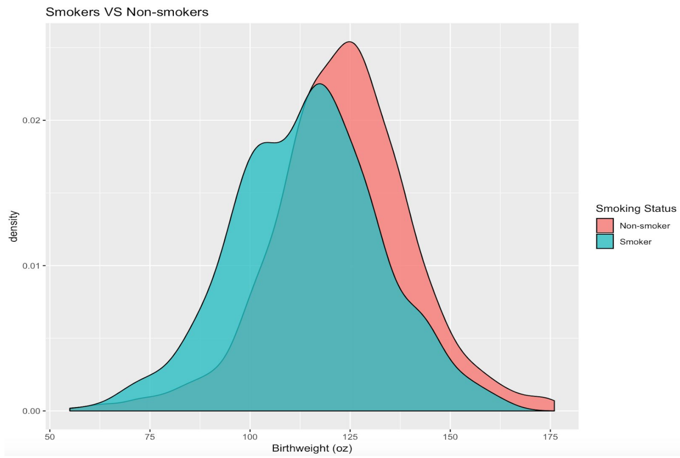

Figure 5.2. Comparative densities of babies birth weights (oz), as distinguished by mothers that were classified as smokers versus non-smokers.

Trends in the data become more visible through histograms. These histograms demonstrate the distribution and density of the babies' birth weights in relation to whether the mother was classified as a smoker versus a nonsmoker. From Figure 5.1, it is clear that the overall distribution of birth weights is unimodal and symmetric as it has no major skew. We can also see that there are more babies that fall under the nonsmoking category than smoking. Figure 5.2 compares both densities directly, revealing that each group is centered around a different mean. Yet, the smoker subset is more spread out whereas the nonsmoker distribution has a more pronounced peak, which would indicate that the kurtosis is larger. As the nonsmoker subset is centered around a higher birth weight, our hypothesis that smoking leads to a decrease in a babies' birth weight is reinforced.

Box-and-Whisker Plot

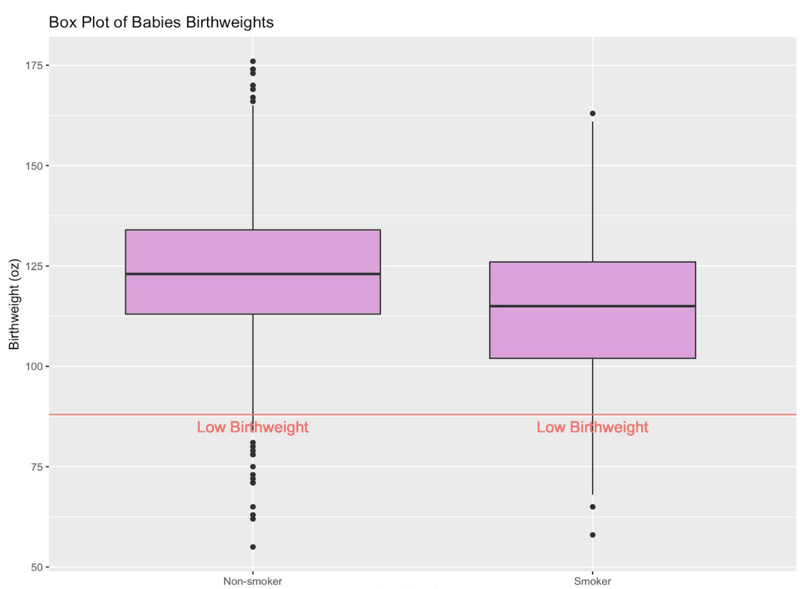

Figure 5.3. Box plot of babies birth weights (oz), grouped by the mothers' smoking classification.

Box plots demonstrate the spread of values within a group in relation to its distribution quartiles. The data is separated into groups based on the mothers smoking classification. The red line indicates a standard "low birth weight" measure for reference. As shown in Figure 5.3, the lower quartile of babies' birth weights from the smokers group is lower than that of the non-smokers group, thus indicating that babies born to mothers that smoke are more likely to have a lower birth weight. Additionally, there are less babies below the "Low Birth Weight" line from the Non-smoker group than from the Smoker group. These results provide more evidence that smoking causes babies to be born with a lower birth weight.

Simple Monte Carlo Procedures (Kurtosis/Skewness)

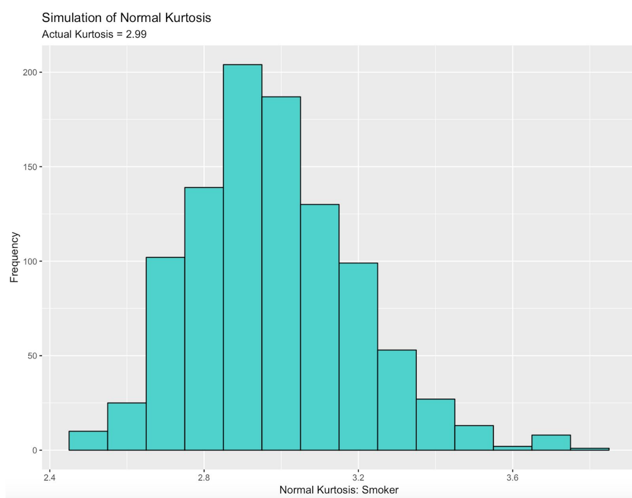

Figure 5.4. Histogram for the simulation of normal kurtosis on the group classified as "Smokers".

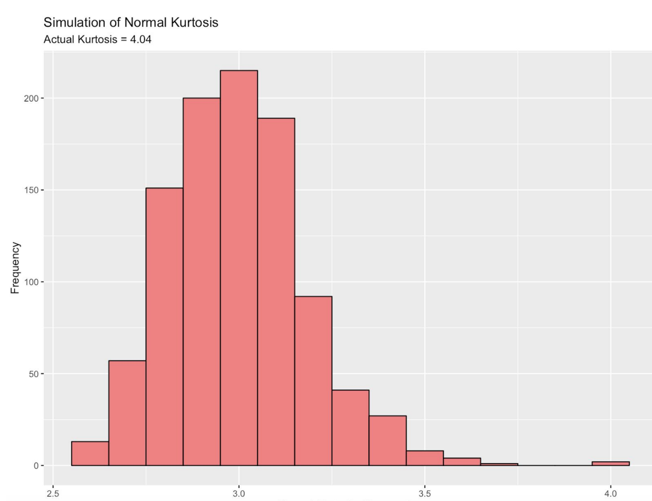

Figure 5.5. Histogram for the simulation of normal kurtosis on the group classified as "Non-smokers".

To get a better picture of each of the datasets, it is important to check for normality through kurtosis and skewness. In Figure 5.4 and 5.5, we run a Monte Carlo simulation of the normal kurtosis.

The kurtosis of the smoker group is 2.99, indicating that it is nearing a normal distribution. The kurtosis of non-smoker group is 4.04. This suggests two things: that it is far from the normal distribution, and may contain extreme values. The slight difference in kurtosis values between the two groups suggests that they are not similar data sets in terms of normality. The increase in kurtosis value for the non-smoker group may be due to a higher central peak of birth weight along with extreme values at the tails. The extreme values could be due to either medical confounds unrelated to smoking which lead to low birth weights, or invalid data about smoking presented by the mothers.

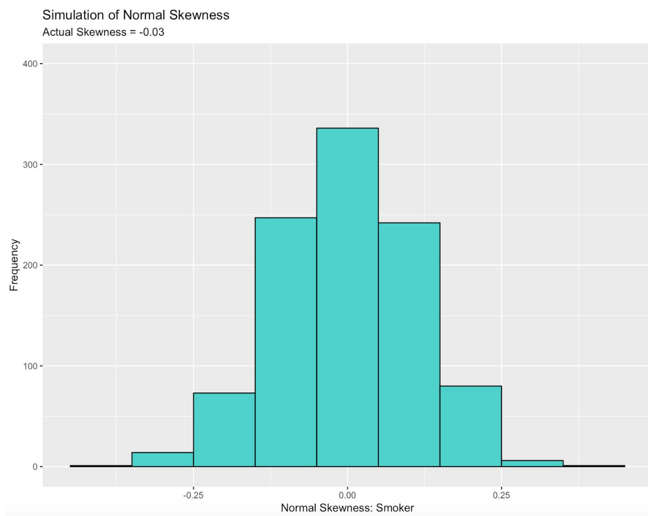

Figure 5.6. Histogram for the simulation of normal skewness on the group classified as “Smokers”.

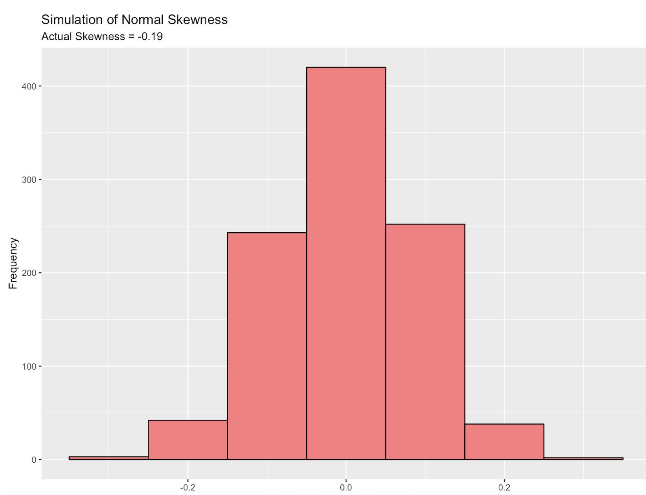

Figure 5.7. Histogram for the simulation of normal skewness on the group classified as “Non-smokers”.

Furthermore, we continue this test of normality with the skewness test in a similar Monte Carlo simulation. Figure 5.6 and 5.7 are both histograms that suggest both data sets are nearly normally distributed, as their skewness values approach 0. This provides evidence for data symmetry in both the non-smoking and smoking group. However, the non-smoker data has a slightly higher skewness value, meaning that it is a little less distributed than the smoker group.

Quantile-Quantile Plot: Comparing Distributions (theoretical/normal vs. actual)

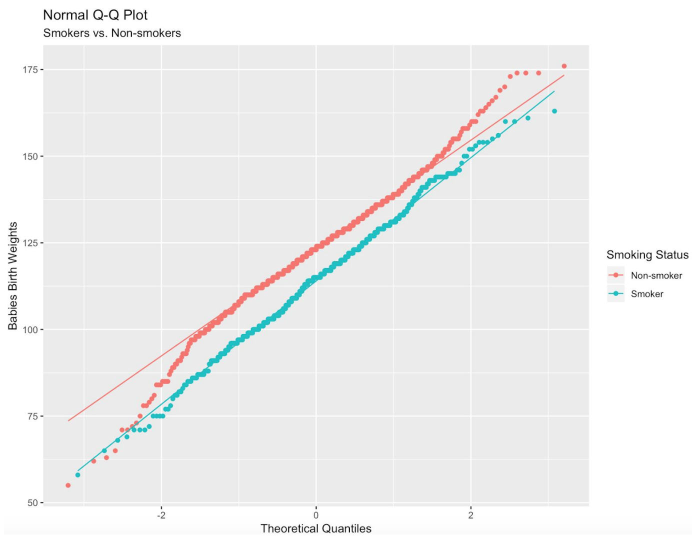

Figure 5.8. Normal quantile-quantile plot of babies' birth weights for both groups classified as smoking and non-smoking.

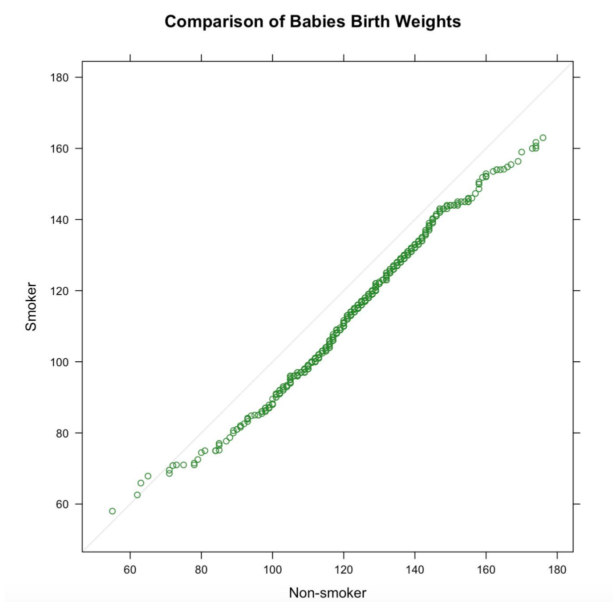

Figure 5.9. Quantile-quantile plot comparison of babies' birth weights.

To test whether the data sequences of birth weights to both smokers and non smokers are normally distributed, we use the Quantile-quantile plot. Figure 5.8 compares the birth weight distribution to its theoretical normal distribution, while also differentiating between smokers and non smokers, as blue and red respectively. The birth weight of the smokers all fall onto the line, supporting the idea that they are approximately normal distributed. The data of the non smokers looks mostly symmetric with the addition of heavy tails on both sides, indicating the presence of extreme values. The data is mostly linear and intercepts equal to zero, with the non-smoker data starting off as irregular and becoming bimodal.

In Figure 5.9, we graph the actual data of the birth weights by comparing smoker to non-smoker. The data points deviate from the quantile-quantile line toward the non-smoker axis, and show an additive shift, as they are parallel to the normal line. The distributions are not identical as they don't fall on the ***x*=*y*** line, with intercept not equal to 0. This visualization shows that the mean of birthweight is greater in non-smokers than it is to smokers, reinforcing our hypothesis that smoking during pregnancy can lead to a decrease in the babies' birth weight.

Central Limit Theorem and further Normality

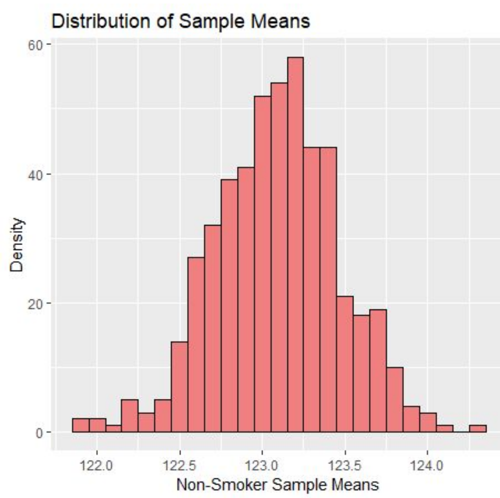

Figure 5.10. Histogram of bootstrap "Non-smoking" sample means.

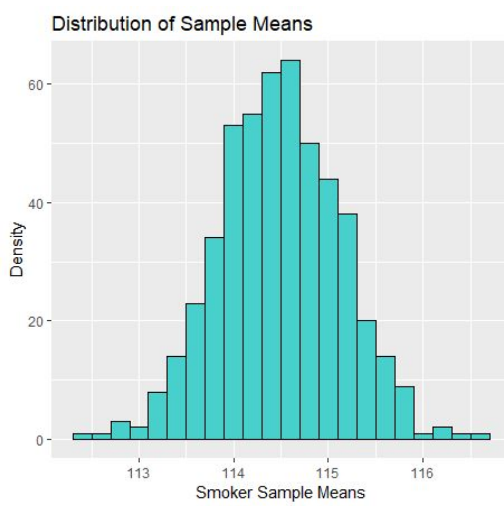

Figure 5.11. Histogram of bootstrap "Smoking" sample means.

To further understand the distributions of the smoker and non-smoker data sets, and whether they are normally distributed or not, the Central Limit Theorem allows us to generate bootstrap samples, and to receive a distribution of the sample means, and then be able to study the distribution formed. For the non-smoking data, 500 samples with 742 values are generated. The mean of each of those samples are graphed in Figure 5.10. It is easy to see the right skewness, agreeing with our analysis of the non-smoker data in Figure 5.5. Nevertheless, the distribution still approaches that of a normal distribution with some irregularities. Similarly, for the smoker data, 500 samples with 484 values are generated. Figure 5.11 shows that the smoking sample means are a closer approximation to the normal distribution than the non-smoking sample means. This further proves the discrepancy in normality when comparing the two samples. In addition, just observing the mean of both histograms, it's easy to see that smokers have significantly lower birth weight means.

Nonsmoker:

```
One-sample Kolmogorov-Smirnov test
data:  (nonsmoke.mean - mean(nonsmoke.mean))/sd(nonsmoke.mean)
D = 0.030098, p-value = 0.7556
alternative hypothesis: two-sided
```

Smoker:

```
One-sample Kolmogorov-Smirnov test
data:  (smoke.mean - mean(smoke.mean))/sd(smoke.mean)
D = 0.022929, p-value = 0.9552
alternative hypothesis: two-sided
```

Furthermore, both sample means show that they follow normal distributions with a 95% confidence interval using the Kolmogorov-Smirnov test, executed by ks.test().

Comparison of Incidence

Now that we have calculated and illustrated different statistical analyses, it is important to define whether these findings are truly reliable or due to chance occurrences. That is, we want to find out if there exists a correlation between smoking and birth weights by differing the amount of babies classified as low birth weight. An effective way to show this is through a Chi-Squared simulation, which will report a statistic based on the probability that two different variables have a dependent relationship. A Chi-Squared test is used to see if the distribution of our data is due entirely to chance or if there does exist a correlation; however, it is crucial to note that this correlation does not prove causation, but indicates how high of a likelihood our data has of existing the way it does only due to chance. Therefore, for the purpose of using this method, the lower the p-value score we obtain the more certain we can claim that our data is not an extreme case and contains a relationship. A Chi-Squared test takes our categorical data, smoking vs. non-smoking, and calculates a “goodness of fit” probability that our observed data fits with a distribution that is to be expected. In Table 5.2, it is shown as we increase the cutoff of low birth weight, our p-score tends closer to 0, meaning our results are more reliable as more babies are distinguished as low birth weight.

Table 5.2. Quantile of low-weight babies and corresponding p-values

Quantile of low bwt	.025	.050	.075	.100
P-value	.11006	.0004254	.000000549	.000000003392

The decline we observe in p-value indicates that as we increase the amount of babies declared as low birth weight, the more reliable our estimates become and we can say confidently that our results are not due to chance. Also, we can say that as fewer babies are classified as low birth weight the higher our chances are of obtaining this data due to chance rather than the variables depending on one another. In summary, this reiterates our previous claims that there does exist a correlation between smoking and low birth weight.

Alternative Hypothesis

While we have illustrated the relationship between smoking and birth weight in babies, we would like to explore other ways smoking impacts a pregnancy. We decided to look into delivery method of the mother to see whether smoking has an effect on which method is used and if there is a correlation between baby weights and these factors. We retrieved data from the Centers for Disease Control and Prevention where we specify grouping by birth weight, delivery method, and tobacco use, and only look into pregnancies that are single birth males, like our previous data set. We expect to see that smokers more often get cesarean sections due to complications during the pregnancy brought along by smoking, and this advanced birth causes the baby to have a low birth weight. On the other hand, we expect to see non-smokers have cesarean sections less than the vaginal delivery method and the birth weights should be higher than those who were born to smokers and had cesarean sections.

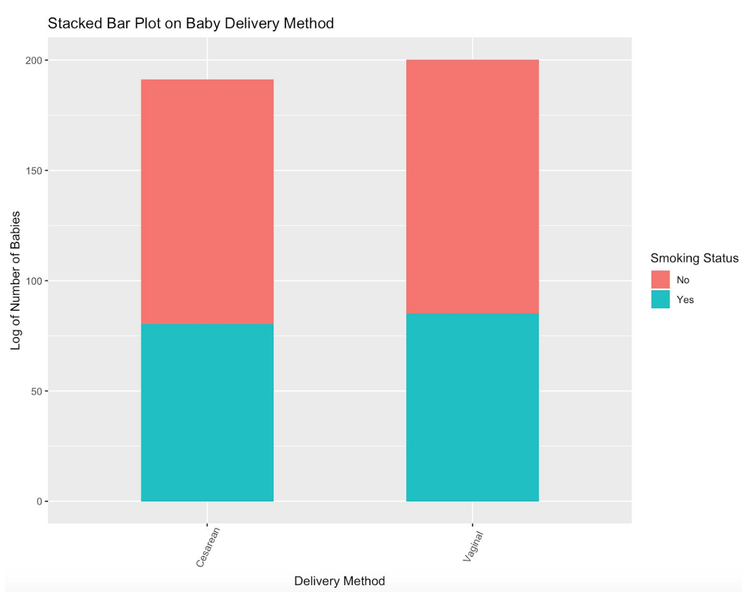

Figure 5.12. Stacked bar plot of delivery method comparing “Smokers” vs “Non-Smokers”.

Our results show that delivery method is not necessarily affected by whether the mother smokes or not. The proportion of babies born by cesarean sections to smokers is less than the proportion of babies born by vaginal method to smokers. This is not what we expected since smoking is known to cause health complications that lead to premature removal of the baby, thus a low birth weight (Lurie 2014). Since the amount of babies born by vaginal method is highest in smokers and non-smokers, we are led to believe that delivery method is not impacted by tobacco use. A limitation of this data set is the amount of cesarean births compared to the amount of vaginal births; there are a lot more vaginal births recorded in this study that could create bias toward delivery method which could be the reason why we see smokers choose vaginal method more often than cesarean. Future studies could maintain the same amount of births for each delivery method to ensure smoking status is not skewed toward a particular delivery method potentially because of its large sample size.

VI. Conclusion

Through various numerical statistics and forms of graphical presentations, we are able to showcase how smoking does share a negative correlation with babies' birth weights. Just by observing the mean baby birth weight between smokers and non-smokers in Table 5.1, we see that the weight is about 10 ounces less than in smokers than in non-smokers. Also, each categories' spreads are almost equal and Figure 5.1 illustrates their distributions being similar. Our quantile-quantile plots, box and whisker plots, and histograms all displayed the same result: there exists a relationship between smoking and birth weight. However, it is significant to acknowledge the limitations of our study and discuss some biases that could arise within our observations. We examine confounding variables that could be affecting our data to show its certain outcome; yet, our results could be due to these variables that we are unaware of. In addition, our results are not generalizable, meaning they are not meant to be assumed for the entire population. We can only give exact conclusions on the group we sampled from, which was the Kaiser Health Plan in Oakland, California. Our results could be further explored by future research conducted over a wider sample size so we can make larger generalizations about the population and with variables like age or socioeconomic status held constant to control for confounding so we can definitively state it is the smoking that causes low birth weight in babies.

Works Cited

Abel, Ernest L. “Smoking During Pregnancy: A Review of Effects on Growth and Development of Offspring.” Human Biology, vol. 52, no. 4, 1980, pp. 593–625. JSTOR, www.jstor.org/stable/41464571.

Bonetti, L.R., Ferrari, P., Trani, N. et al. Arch Gynecol Obstet (2011) 283: 231. https://doi.org/10.1007/s00404-009-1317-4.

Jouni J. K. Jaakkola, Mika Gissler, “Maternal Smoking in Pregnancy, Fetal Development, and Childhood Asthma”, American Journal of Public Health 94, no. 1 (January 1, 2004): pp. 136-140.

Lurie Samuel, Ribenzaft Shay, Boaz Mona, Golan, Abraham & Sadan, Oscar (2014). The effect of cigarette smoking during pregnancy on mode of delivery in uncomplicated term singleton pregnancies, The Journal of Maternal-Fetal & Neonatal Medicine, 27:8, 812-815, DOI: 10.3109/14767058.2013.842551

Wilcox AJ, Skjaerven R. Birth weight and perinatal mortality: the effect of gestational age. American Journal of Public Health, 1992;82(3):378-382.