Baby Weights of Smoking Mothers vs Baby Weights of Non-Smoking Mothers

McCord Murray

January 27, 2022

Abstract

Over time, many studies have shown health risks attributed to smoking tobacco. Data from an investigation of all pregnancies among women in the Kaiser Foundation Health Plan in the San Francisco–East Bay area that occurred between 1960 and 1967 was collected to observe if smoking during pregnancy affects the newborn's body weight. Upon examining the distributions of the babyweights of smoking mothers versus non-smoking mothers, not much could be immediately determined by the histograms alone. When overlaying the two histograms on top of each other, it becomes much more clear with this visualization that more babies born from smoking mothers were born underweight. The descriptive statistics of the distributions showed that babies born from mothers who smoked weighed, on average, about half a pound lighter than babies born from mothers who did not smoke.

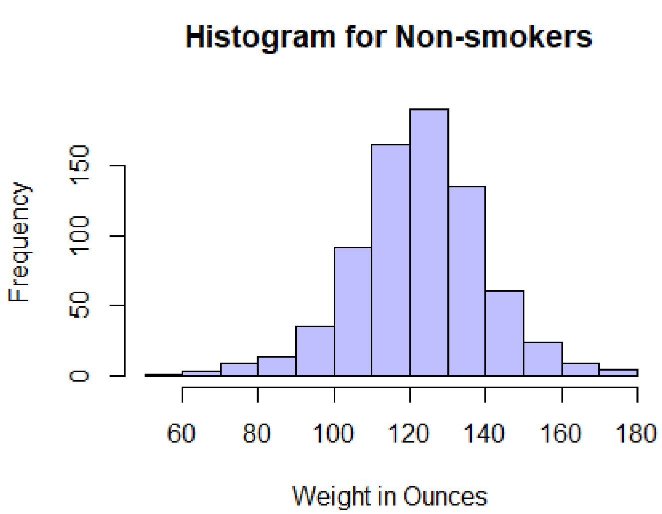
Background and Significance

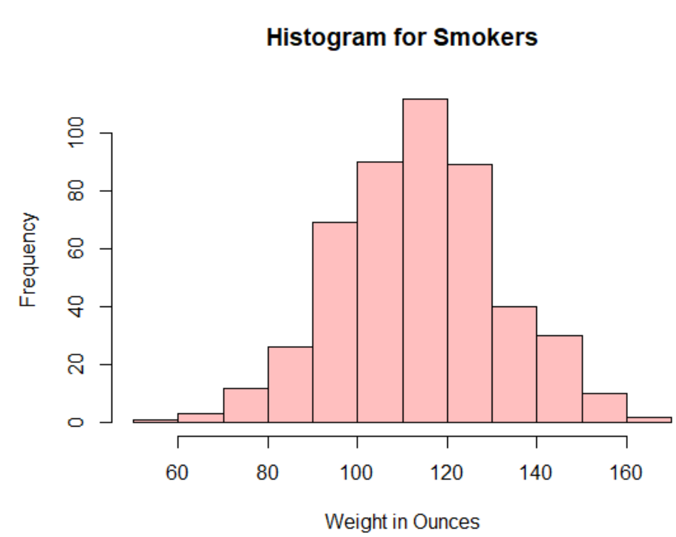
Over time, many studies have shown health risks attributed to smoking tobacco. As stated by the CDC (2020): “Smoking leads to disease and disability and harms nearly every organ of the body,” which is why smoking during pregnancy is widely considered to be irresponsible since the baby relies on the functionality of the mother’s organs to survive. However, despite this fact many women still continue to smoke during pregnancy. A health warning placed on the side of cigarette packages says: “Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight,” (Nolan & Speed, 2000, p.2), The goal of this study is to confirm whether or not babies who were born by mothers who smoked during pregnancy have a higher chance of being underweight and/or born prematurely. The hypothesis is that babies born by smoking mothers do have a higher chance of being underweight and/or born prematurely.

Methods

In order to test our hypothesis, data from an investigation of all pregnancies among women in the Kaiser Foundation Health Plan in the San Francisco–East Bay area that occurred between 1960 and 1967 was collected to observe if smoking during pregnancy affects the newborn's body weight. This data contains the weight of the baby at birth in ounces and whether or not the mother smoked during pregnancy. The first step was to separate the women who smoked during pregnancy from the women who did not smoke during pregnancy into two different sets using R programming. Next, histograms showing the distribution of the babies’ weights for smokers and nonsmokers were created and compared with each other. For each set, we found the mean, median, standard deviation, skewness, kurtosis, and 5-point summary to get as much information about the distributions for each histogram as possible. After this, box and whisker plots were constructed for each set and displayed side by side to compare the 5-point summaries of both histograms. Finally, Q-Q plots were constructed for both histograms to compare their distributions with a normal distribution. If both distributions are close to normal, we can use a t-test to determine the probability the distributions are the way they are.

Results

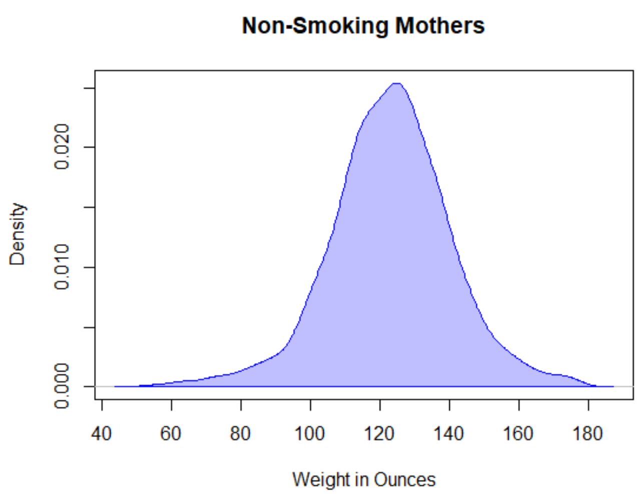


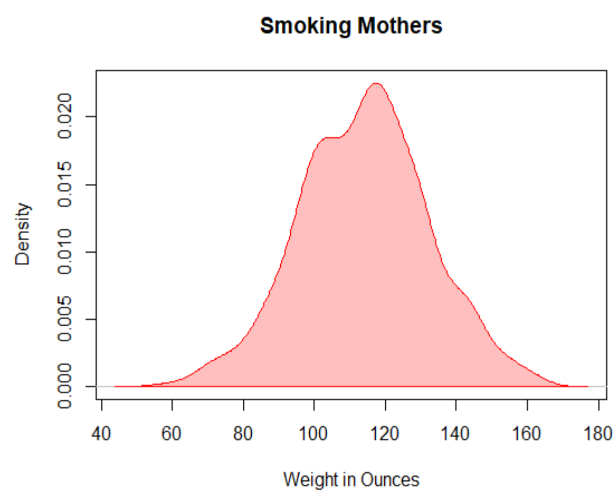


Upon examining the distributions of the babyweights of smoking mothers versus non-smoking mothers, not much could be immediately determined by the histograms alone. In order to get a better idea of what we were looking at, finding the descriptive statistics of these distributions was necessary.

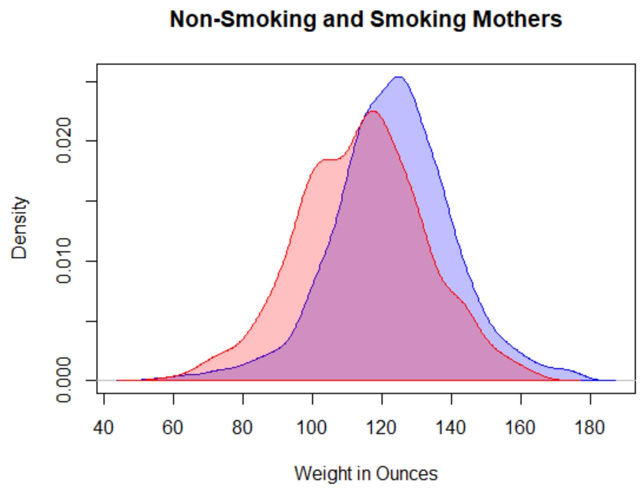
Descriptive Statistics	Non-Smokers	Smokers
Mean	123.0472	114.1094
Median	123	115
Standard Deviation	17.39869	18.09895
Skewness	-0.186941	-0.03359498
Kurtosis	4.03706	2.98032
5-Number Summary	55 113 123 134 176	58 102 115 126 163

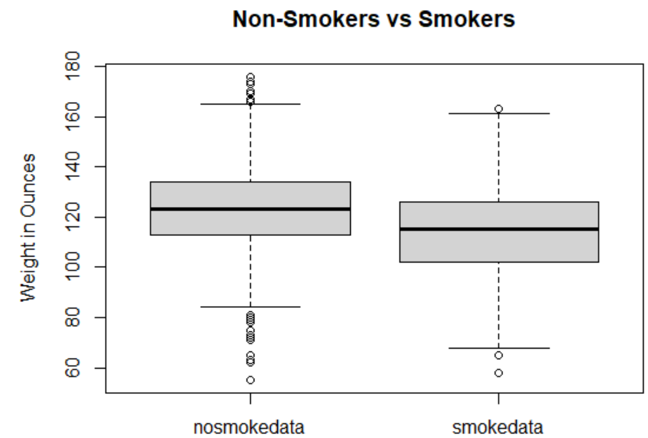
The descriptive statistics showed that babies born from mothers that smoke weigh less on average. Additionally, the kurtosis for the Non-Smokers distribution is high. When smoothing the histograms, it can be observed that the peak of the Non-Smokers histogram is higher than the Smokers histogram. This means more babies born from Non-Smoking mothers weighed close to the median.



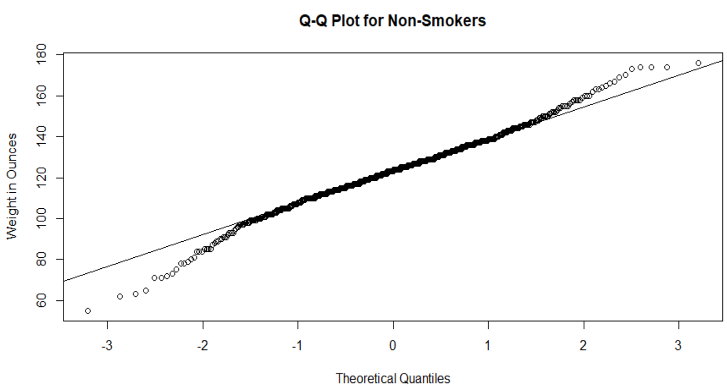


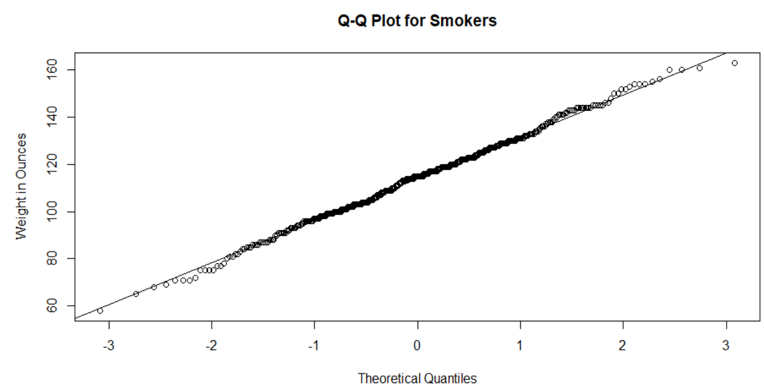
When overlaying the two histograms on top of each other, it becomes much more clear with this visualization that more babies born from smoking mothers were born underweight. The box and whisker plot also shows that there were lots of outliers in the weights of babies born from Non-Smoking mothers, whereas there were very few outliers for the babies born from smoking mothers. Approximately 69% of all babies born from Non-Smoking mothers were heavier than the median of babies born from smoking mothers.



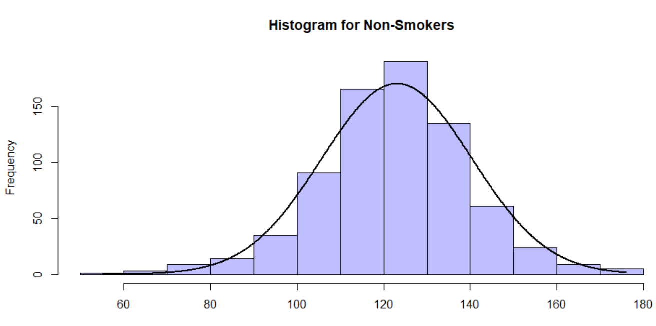


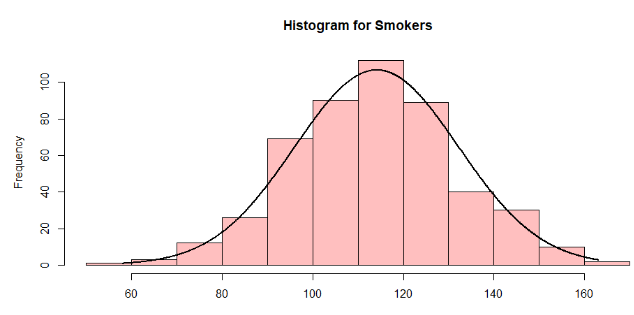
Lastly, when comparing both distributions to a normal distribution with Q-Q Plots, it was observed that the Q-Q Plot for Smokers resembled the normal distribution closely. However, we see the Q-Q Plot for the Non-Smokers deviates from the normal line after 2 quantiles, meaning the distribution for Non-Smokers is not normal.





Drawing a normal curve over the histograms the distribution of the babies' weights from the smokers mothers much more closely resembles the normal curve then the distribution of the babies' weights for smoking mothers. This result is consistent with the Q-Q plots.





Conclusion

The goal of this study was to determine whether or not smoking while pregnant increased the chance of babies being born underweight. The descriptive statistics of the distributions of women who smoked during pregnancy and women who did not smoke during pregnancy showed that babies born from mothers who smoked weighed, on average, about half a pound lighter than babies born from mothers who did not smoke. Around 69% of all babies born from Non-Smoking mothers were heavier than the median of babies born from smoking mothers. This difference was made especially clear when the two histograms were displayed on top of each other. Additionally, the kurtosis of the distribution of baby weights from non-smokers was high. The kurtosis of a distribution is high when it has a tall peak and/or fat tails. In this case, the distribution of baby weights born from non-smokers had a bit of both. The high peak, as seen from the histogram, tells us there is a high chance a baby born from a non-smoker weighs close to the median. This is consistent with the idea that not smoking during pregnancy leads to the baby being born at a healthy weight. However, an explanation for the high number of outliers, as seen from the box and whisker plot, has yet to be determined.

References

Nolan, D., & Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications* (p. 2). Springer.

Health Effects of Smoking and Tobacco Use. Centers for Disease Control and Prevention. (2020). Retrieved 25 January 2022, from https://www.cdc.gov/tobacco/basic_information/health_effects/index.htm.

Appendix

Prove that the mean, $\overline{x}$, of a finite numerical data set $\{x_1, \ldots, x_n\}$ is the value for c that minimizes:

$$\sum_{i=1}^{n} (x_i - c)^2$$

First we take the derivative of the function:

$$\frac{d}{dx}\left(\sum_{i=1}^{n}(x_i - c)^2\right) = \sum_{i=1}^{n} 2(x_i - c)$$

Then we set the derivative equal to 0 to find the absolute minimum:

$$2\sum_{i=1}^{n}(x_i - c) = 0 \Rightarrow \sum_{i=1}^{n}(x_i - c) = 0$$

Using the properties of a summation we can separate the function into a difference of sums:

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} c = 0 \qquad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} c$$

Finally, we conclude that ‘c’ must be equal to the mean to satisfy the equation:

$$\sum_{i=1}^{n} x_i = nc \Rightarrow \boxed{c = \bar{x}}$$