Shawn Koohy
1/30/22
Dr. Davis
MTH332

# Does Smoking During Pregnancy Effect a Baby's Birthweight

## Abstract

Over the past couple of decades awareness towards smoking has become a very important issue to be brought up by Doctors and Health Professionals to their patients. Smoking is known to cause cancer, heart disease, diabetes, and various other life-threating health issues. Smoking can be especially dangerous when done by pregnant women. The data provided for this study is based on the birthweights of the newborn babies. Through various analytical and statistical measurements, we have decided that smoking during a pregnancy will cause a baby to have a lower birthweight than those who don’t smoke. Furthermore, we will be discussing where this conclusion came from, how it was measurement and how it could be possible to make the conducted study more efficient based off data choices.

## Introduction & Background

This analysis report has been devoted to furthering the awareness of women smoking during their pregnancy. A baby with a higher or lower birthweight can find it harder to fight off infections and regulate blood sugar levels. A baby's weight is a common factor in determining their health status not only at birth but can also play a role later as an adult. The nicotine found in cigarettes or other smoking-based products has shown to reduce maternal blow flow, known as "uteroplacental circulation", which is the way a baby receives nutrients and oxygen. This is where effects such as lower birthweight comes from. We will see soon how our statistical measurements match up with the science and be able to visually see the difference.

## Methods

The data used in this experiment was collected and conducted by the Child Health and Development Studies between 1960 and 1967, where over 1,000 families participated. To plot and analyze the data provided I used Mathematica. From Mathematica I split the data set points into two groups, one for smoking mothers (named "Smoke") and non-smoking mothers (named "NoSmoke"). After separating the data, I graphed a histogram for both groups. The histogram is an easy visual to understand and identify symmetry and deviations. Along with those benefits a histogram allows me to find many important descriptive statistical measurements such as minimum, maximum, mean, median, standard deviation, skewness, kurtosis, first quartile and third quartile. To better understand what these measurements, plots, and data points are telling me, I created a smooth histogram for each group and graphed them together to easily compare their differences. Box and Whisker plots were also created to better visual the five-point summary of the data. Lastly Quantile Plots for "Smoke" and "NoSmoke" were created with the intent to compare the data distribution to the approximate normal distribution.

## Results

From the data collected we can draw the conclusion that smoking does influence the birthweight of a baby. This result was determined by the various statistical measurements, histograms, smooth histograms, box and whisker plots that were used when evaluating the given data by the Child Health and Development Studies.

In Table 1 we can see some major differences in the median and maximum birthweight. More specifically the smoking mothers have a lower median and maximum than that of non-smoking-mothers. It is also important to note that Q1 And Q3 are lower for the smoking mothers versus non-smoking mothers and not a significant difference in terms of minimum weight but not small enough to consider negligible.

| Five Point Summary | | | | | |
|---|---|---|---|---|---|
| Group \ Statistic | Minimum (oz) | First Quartile (oz) | Median (oz) | Third Quartile (oz) | Maximum (oz) |
| Non-smoking Mothers | 55 | 113 | 123 | 134 | 176 |
| Smoking Mothers | 58 | 102 | 115 | 126 | 163 |

Table 1. Five Point Summary

When analyzing Table 2 there is a major difference between the mean of the two groups, smoking mothers being on the low side, and the standard deviations are similar. A moderately or skewed distribution exist between -1 and -0.5 or 0.5 and 1. A highly skewed distribution exists when skewness is less than -1 or greater than 1. From this we can say that both groups are very lightly skewed. A Kurtosis of +/-2 from the normal 3 is also considered good/acceptable.

| Other Relevant Descriptive Statistics | | | | |
|---|---|---|---|---|
| Group \ Statistic | Mean (oz) | Standard Deviation (oz) | Skewness | Kurtosis |
| Non-smoking Mothers | 123.047 | 17.3987 | -0.186984 | 4.03706 |
| Smoking Mothers | 114.11 | 18.0989 | -0.033595 | 2.98803 |

Table 2. Other Relevant Descriptive Statistics

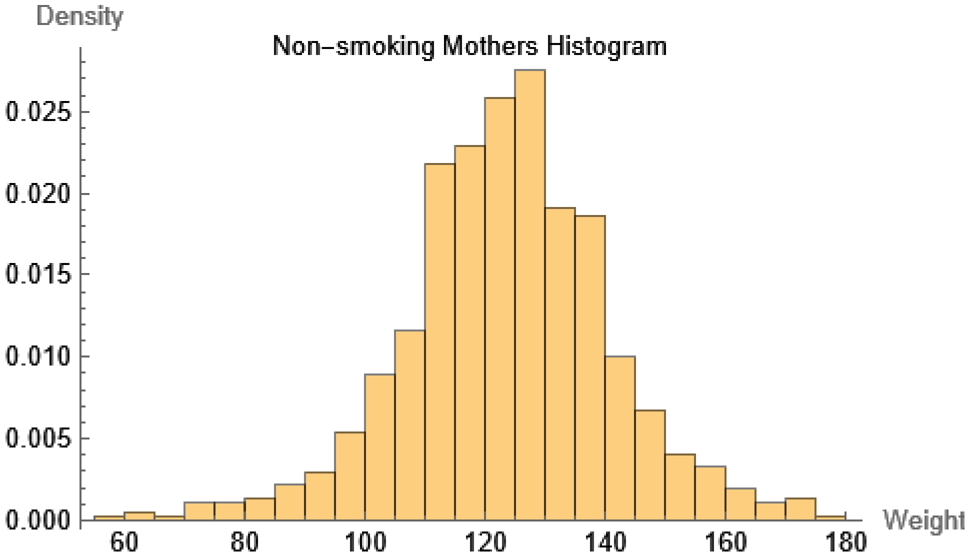


Figure 1. Non-smoking Mothers Histogram

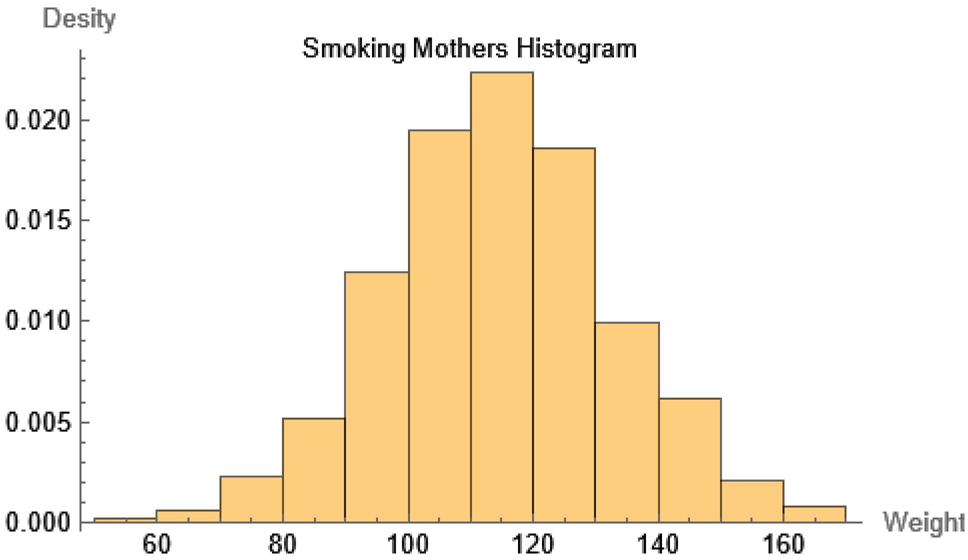


Figure 2. Smoking Mothers Histogram

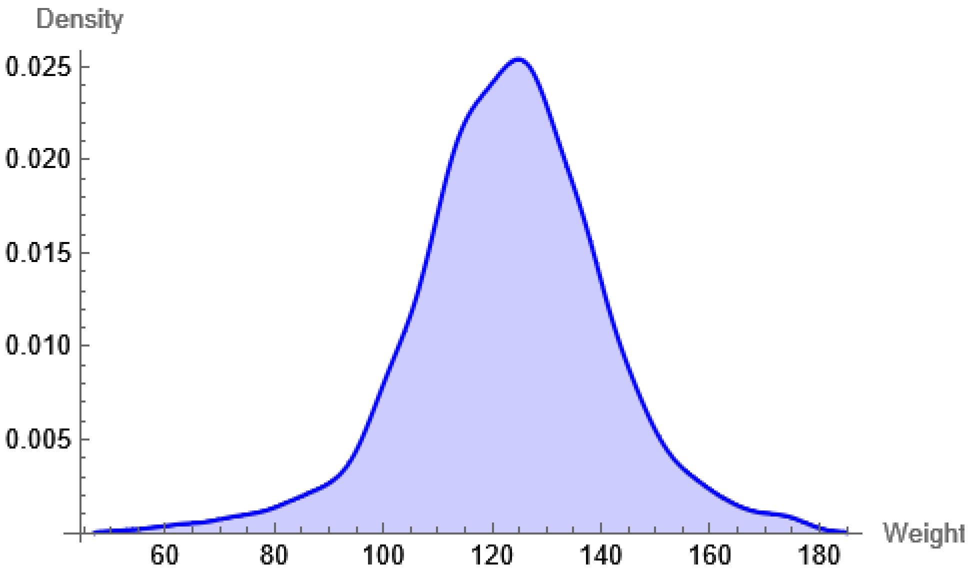


Figure 3. Non-smoking Mothers Smooth Histogram

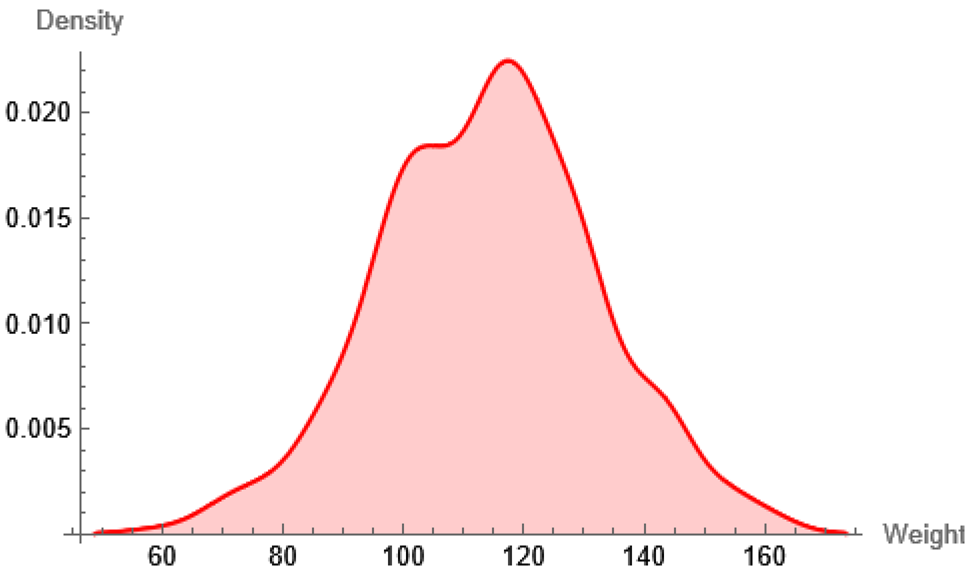


Figure 4. Smoking Mothers Smooth Histogram

The comparison of the smooth histograms (Figure 5.) allows for an easy visualization of exactly what the data and statistical measurements are telling us. We can see a very clear shift to the right (less weight) in smoking mothers compared to non-smoking mothers. Along with that we can see in that the smooth histogram for non-smoking mothers' data is more evenly distributed. We can look back at Table 2. to see how our measurements match the graph. As an example, non-smoking mothers' kurtosis was ~4 which we can see as it is thinner and has a higher peak.

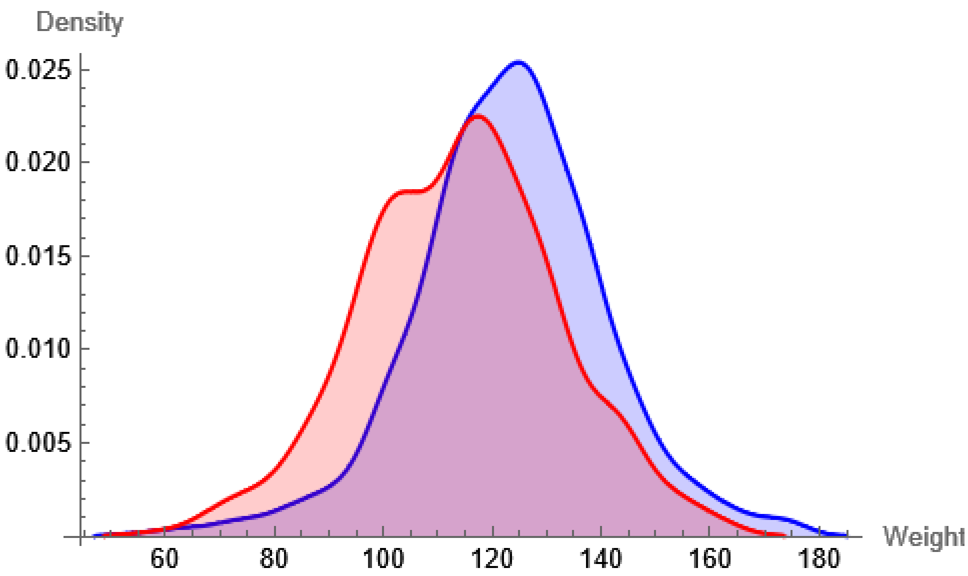


Figure 5. Comparison of Smoking (Red) vs. Non-smoking Mothers (Blue) Smooth Histogram

We can see in Figure 6. and Figure 7. (Next page) that there is a significant difference between the linearity of each Q-Q plot. We see a very normally distributed and linear combination of points in non-smoking mothers were as in smoking mothers there is a clear non-uniform distribution. The Q-Q plot for smoking mothers tells us that there is a noticeable deviation from what we expect to be the normal distribution.

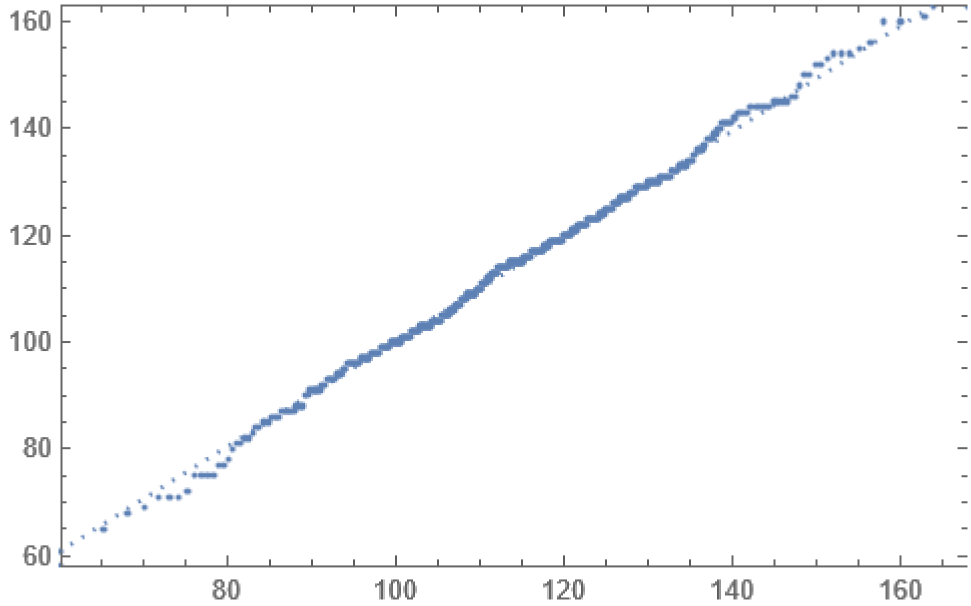


Figure 6. Non-smoking mothers Q-Q Plot

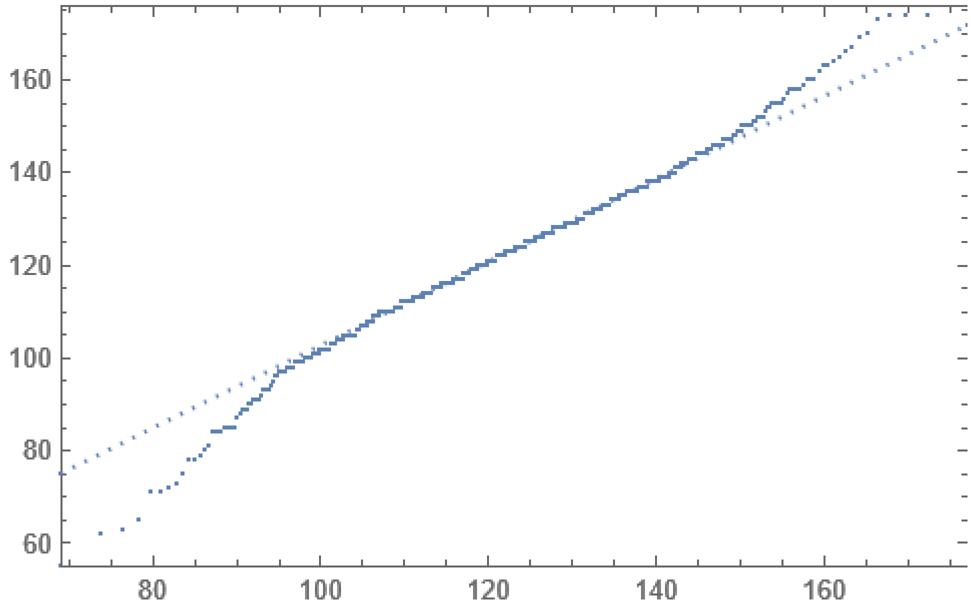


Figure 7. Smoking Mothers Q-Q Plot

The box and whisker plot allows us to better visualize the five-point summary that was presented in Table 1. We can see the visually smaller median , maximum, Q1, and Q3, as discussed before.

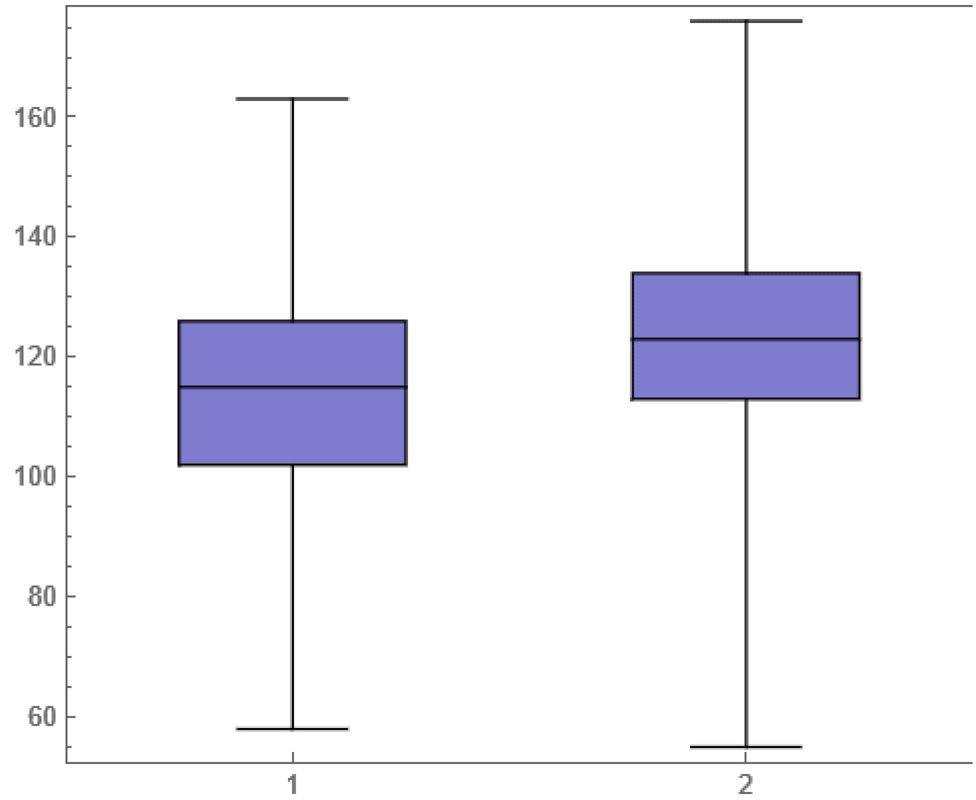


Figure 8. Box and Whisker Plot of Smoking Mothers (1) and Non-smoking Mothers (2)

Discussion & Conclusions

Now that the results and all data have been present, we can conclusively say that smoking does influence the birthweight of newborn babies, in the mothers of this study. More specifically the birthweight of babies from smoking mothers is less than the birthweight of babies from non-smoking mothers. The major determinants of this conclusion are from the significant deviations between the mean, median, first quartile and third quartile that are all below that of the non-smoking mothers. The non-uniform distributing of the smooth histogram in smoking mothers is one of the clearest visual factors in seeing the significant effects. Throughout nearly all analytical and statistical measurements, various graphs, and plots there is undeniable proof that smoking led to lower-than-normal birthweights in babies. It should be mentioned that though not a major difference there is a slight diversion in the minimum, skewness, and kurtosis in both groups. There is a difference in the minimum by a magnitude of 3, a magnitude of 1.04903 for kurtosis and 0.153389 for skewness. These can be considered statistically insignificant for the purposes of this research. Rebuttal toward this conclusion is not out of the picture though. The data and information provided doesn’t take account for various other issues other than smoking that may cause an abnormal birthweight such as age, weight, or general health of the mother. There is a significant trend in the data of smoking mothers enough to make a valid conclusion that yes, smoking does affect a baby's birthweight.

As this data and study was collected and conducted between 1960-1967, the diversity of mothers in the United States has grown a lot especially in terms of ones, economic, social, and racial characteristic, all of which could have positive or negative effects on the health of a baby. This study contained approximately 66% Caucasian women, 20% African American women and the remaining 14% being Oriental, mixed or other races. It might be better to create a more even distribution of races because it is possible that African American women produce babies with lighter weights than Caucasian women do. This would bias or distort the data to give off the idea that smoking was the cause of the differentiating weights when it is possible it is more due to a person genetics, race, and/or ethnic background.

References

Kataoka, M., Carvalheira, A., Ferrari, A., Malta, M., De Barros Leite Carvalhaes, M., & De Lima Parada, C. (2018, March 12). Smoking during pregnancy and harm reduction in birth weight: A cross-sectional study. Retrieved from https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-018-1694-4#:~:text=Nicotine%20likely%20plays%20an%20important,and%20compromised%20fetal%20neurological%20development

Nolan, D. A., & Speed, T. (2001). *Stat labs: Mathematical statistics through applications*. New York, NY: Springer.

S. (2007, June 05). Born to lose: How birth weight affects adult health and success. Retrieved January 29, 2022, from https://news.umich.edu/born-to-lose-how-birth-weight-affects-adult-health-and-success/

S. (n.d.). Low birthweight. Retrieved January 29, 2022, from https://www.marchofdimes.org/complications/low-birthweight.aspx

Tina Donvito December 16, 2. (2018, December 16). How your weight impacts your baby's weight. Retrieved January 29, 2022, from https://www.parents.com/pregnancy/my-body/pregnancy-health/how-your-pregnancy-weight-directly-affects-your-babys-size/

## Appendix

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

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

*Proof.* Let $f(c) = \sum_{i=1}^{n}(x_i - c)^2$. So,

$$\begin{aligned} f(c) &= (x_1 - c)^2 + (x_2 - c)^2 + ... + (x_n - c)^2 \\ f'(c) &= -2(x_1 - c) - 2(x_2 - c) - ... - 2(x_n - c) \\ &= -2[(x_1 - c) + (x_2 - c) + ... + (x_n - c)] \end{aligned}$$

Let $f'(c) = 0$. So,

$$\begin{aligned} 0 &= -2[(x_1 - c) + (x_2 - c) + ... + (x_n - c)] \\ &= (x_1 - c) + (x_2 - c) + ... + (x_n - c) \\ &= x_1 - c + x_2 - c + ... + x_n - c \\ &= (x_1 + x_2 + ... + x_n) - nc \\ nc &= (x_1 + x_2 + ... + x_n) \\ c &= \frac{x_1 + x_2 + ... + x_n}{n} \\ c &= \overline{x} \end{aligned}$$

$\square$