Mothers, Smoking, and Their Children

A statistical analysis of the impacts of smoking mothers on the birth weights of their babies

By: Katelyn Banalewicz
Spring 2021
MTH 332-01

Abstract

A look into the dangers of smoking led to the warnings on cigarette packaging seen today. Special attention was paid to the impacts of smoking on the fetus of a woman who was pregnant. Increasingly, the US Surgeon General highlighted the risk for "fetal injury, premature birth, and low birth weight" on cigarette packaging to warn mothers against smoking during pregnancy. The data presented in this study took data from over 1000 babies and separated this data into two categories: babies born to mothers who smoked while pregnant and those born to whose mothers did not. Analysis of this data observed trends in birth weights, which tend to be in accordance with the Surgeon General's warning against smoking while pregnant, leading to lower birth weights on average. However, rates of data groups may be impacted by not accounting for factors outside of whether or not the women smoked. The data presented below explores deeper the weights of babies born to mothers who smoke to see how numerically valid the Surgeon General's warning truly is.

Introduction and Background

The US Surgeon General has placed health warnings on cigarette packaging for decades, warning the public about the potential dangerous risks associated with smoking. One such warning reads highlights that pregnant women who smoke risk "fetal injury, premature birth, and low birth weight." Despite this, many women, including 15 percent of women who bore children in 1996, smoke while pregnant. This activity can lead to detrimental impacts on the baby's health, as smoking can lead to a reduction in birth weight, one of the metrics for fetal maturity. Usually, babies born at lower birth weights tend to be born earlier and are less likely to survive than those who are bigger and are carried to term.

Fetal development is judged under the assumption of a "normal" 40 week gestational period. Therefore, any baby born before completion of 37 weeks is considered premature or preterm. In 1941, Australian doctor, Dr. Norman Gregg, challenged the widely accepted belief that babies were completely protected within the womb and were not susceptible to any disease or chemical that may impact the mother. Studies of babies born to mothers who had contracted diseases had also been impacted by the ailments their mothers dealt with. Because of this, the chemical agents in cigarette smoke inhaled by the mother will also impact the fetus. Carbon monoxide and other chemicals cause problems such as the restriction and decrease of oxygen supply to the fetus, whose parietal pressure in their blood is already less than that of an adult’s. A decrease in oxygen supply can lead to complications in fetal development that make preterm delivery more likely, leading to low birth weights and increased possibility of fatality.

Methods

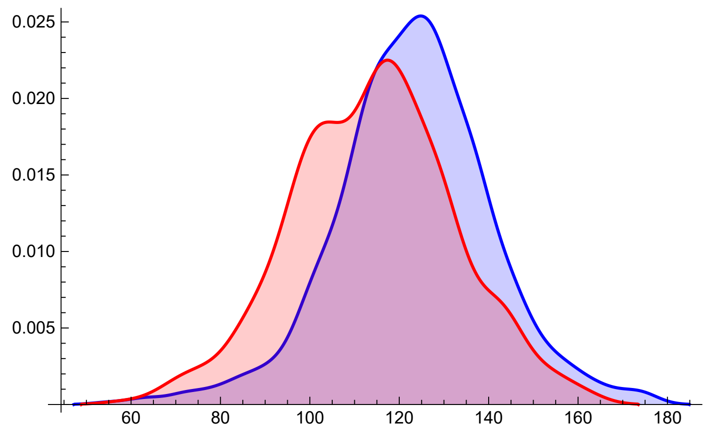
The data in this study was collected to analyze the impacts of smoking on birth weight and infant fatality. The data analyzed in this report is a subset of the larger Child Health and Development Studies (CHDS), conducted to study all pregnancies among the women in the Kaiser Foundation Health Plan in Oakland, CA from the period between 1960-67. This subset is of 1236 baby boys who lived for at least a period of four weeks (28 days) after birth and were not twins, triplets, etc.

Measurements of length, weight and head circumference were taken and recorded at the time of the baby’s birth. Variables for the study were created and labeled as *Birth Weight* and *Smoking Status.* Birth Weight was defined as the weight of the baby in ounces; Smoking Status is defined to note (1) a mother who smoked during her pregnancy, or (0) one who did not.

Analysis of birth weight relative to smoking status was not the only thing taken into account in the study. Comparison of overall mortality rate of babies of smoking mothers against those of mothers who did not smoke was also conducted.

Results

Figure 1:



The figure above is the comparison of all provided birth weights (in ounces) of babies born to smoking (red) and non-smoking (blue) mothers.

Key Terms:

Mean is defined as the average of numbers when summed and divided by the total number of inputs. **Median** is defined as the middle number when the list of total inputs is sorted into an ascending or descending list. **Standard deviation** is defined to be the statistic measure of the dispersion of data relative to its normal distribution; it is calculated as the square root of the variance. **Skewness** is defined to be the measure of the symmetry of distribution of data points. **Kurtosis** is defined as the measure of the heavy- or light-tailed nature of the data relative to normal distribution; data with light tails tend to have a lack of outliers and therefore a low kurtosis, while data with heavy tails tend to have more outliers and a higher kurtosis.

Statistical Analysis of Birth Weights

Separating the data based on smoking status reveals that the average birth weight of a baby born to a mother who smoked while pregnant was 114.11 ounces, while the median weight if these same babies was 115 ounces. Since the median is slightly greater than the average birth weight, the data is skewed to the left as represented by a negative value, -0.033595. This means that there is a larger number of values on the side below the median and a smaller number of values above that line. The standard deviation of the data for babies born to smoking mothers is 18.0989, meaning that the average difference between data points and the mean is about 18 ounces. The kurtosis of this data is 2.98803, meaning that the data is not far off from the normal distribution, which is supported by the small skewness value.

The average birth weight of a baby born to a non-smoking mother in this study was 123.047 ounces and the median was 123 ounces. The skewness value for this data is -0.186984, and there are slightly more values above the median than below. The standard deviation for these values is 17.3987, meaning that the average difference between data points and the mean is just over 17 ounces. The kurtosis of this data is 4.03706, meaning that there are wider tails to the data and is indicative of having many outliers in the set.

Figure 2 shows the normal distribution compared to the actual distribution for babies born to smoking mothers; Figure 3 shows the normal distribution compared to the actual distribution for those born to non-smoking mothers. The figure on the left shows that the birth weights of babies whose mothers smoked is much more evenly distributed compared to the weights of those whose mothers did not. This analysis is in line with the kurtosis values that reveal there are more outliers in the non-smoking data.

Figure 2

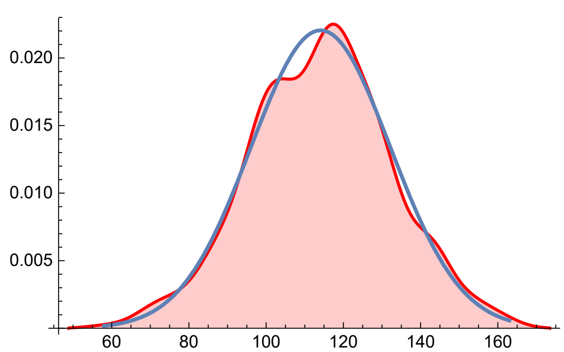
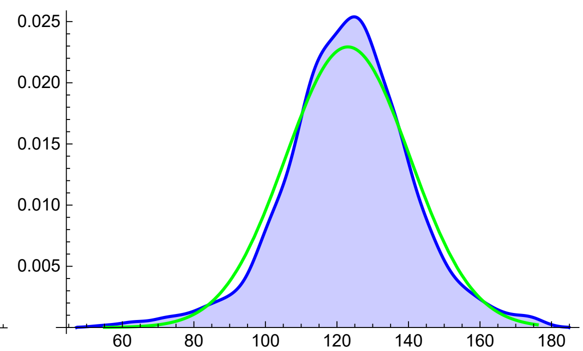


Figure 3



Five-Number Summaries and Quantile Plots

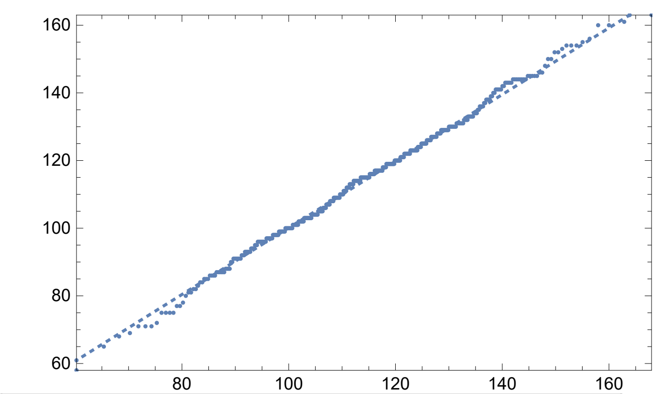
For Smoking Mothers	Median	115
Quartiles	102	126
Extremes	58	163

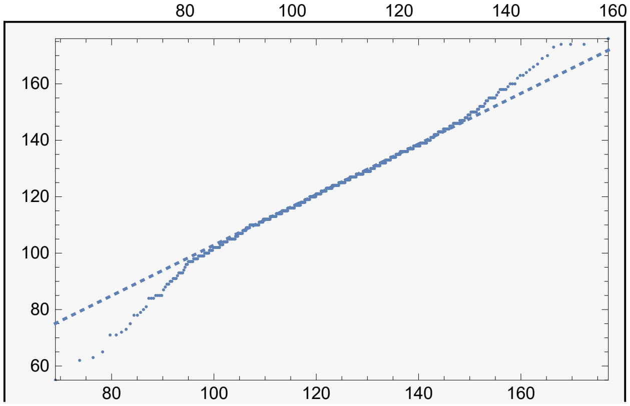
For Non – Smoking Mothers	Median	123
Quartiles	113	134
Extremes	55	176

The five-number summary of the data reveals that fifty percent of babies born to smoking mothers weighed between 102 and 126 ounces, and at least one weighed 163. The data is roughly symmetrical, but slightly skewed as the upper quartile is marginally closer to the median (a difference of 11 ounces) compared to the lower quartile (a difference of 13 ounces). Similarly, the five number summary of the data for non-smoking mothers showed that half of these babies weighed between 113 and 134 ounces, though the data appears to have more outliers. This is because while the differences between the quartiles are relatively close to each other, the upper extreme is much closer to the median than the lower extreme.

Similarly, the quantile plots, which show the predicted values based on a normal distribution compared to the actual data, support the finding that the data for non-smoking mothers has more outliers toward the extremes than that of mothers who do smoke.

Figure 4 (quantile plot for babies born to smoking mothers) *Figure 5 (quantile plot for babies born to non-smoking mothers)*





Discussion and Conclusion

The data analysis revealed that babies born to non-smoking mothers tend to have a higher birth weight than those who were born to mothers who smoked while pregnant. This result is in line with the warning of the surgeon general which states the inherent dangers to the fetus that smoking poses. However, the data for smoking mothers is more evenly distributed than that of the data for non-smoking mothers. These deviations can be attributed to a number of factors. One such factor is the size of the groups in the study. When sorting the data, babies were grouped by birth weight, so the data means and medians are impacted based on the number of data points available from the group. Because the groups were differently sized, the rates of their birth weights are counted, rather than the individual birth weights themselves. Another factor in these deviations could be that the data was not adjusted to account for maternal age and other health factors which would contribute to outliers in birth weight. This is significant because older mothers tend to have issues with their pregnancy whether they smoke or not, leading to preterm delivery or low birth weights.

References

Nolan, D. A. & Speed, T. (2001). Maternal Smoking and Infant Health. In Stat labs: Mathematical statistics through applications (pp. 1-25). New York, NY: Springer.

How to find the mean. (n.d.). Retrieved February 06, 2021, from https://www.mathsisfun.com/mean.html#:~:text=The %20 mean %20 is %20 the %20 average,sum %20 divided %20 by %20 the %20 count.

Ganti, A. (2020, December 14). Median definition. Retrieved February 06, 2021, from https://www.investopedia.com/terms/m/median.asp#:~:text=The %20 median %20 is %20 the %20 middle,data %20 set %20 than %20 the %20 average.

Hargrave, M. (2021, February 02). Standard deviation. Retrieved February 06, 2021, from https://www.investopedia.com/terms/s/standarddeviation.asp#:~:text=The %20 standard %20 deviation %20 is %20 a,square %20 root %20 of %20 the %20 variance.&text=If %20 the %20 data %20 points %20 are,the %20 higher %20 the %20 standard %20 deviation.

Skewness - statista definition. (n.d.). Retrieved February 06, 2021, from https://www.statista.com/statistics-glossary/definition/390/skewness/#:~:text=Skewness %20 is %20 a %20 measure %20 of,a %20 distribution %20 is %20 its %20 mode.&text=A %20 distribution %20 is %20 skewed %20 if,the %20 other %3 A %20 it %20 is %20 asymmetrical.

1.3.5.11. measures of skewness and kurtosis. (n.d.). Retrieved February 06, 2021, from https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm#:~:text=Kurtosis %20 is %20 a %20 measure %20 of,would %20 be %20 the %20 extreme %20 case.

Appendix

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

$$\sum_{i=1}^{n} (x_1 - c)^2$$

Let: *min* $\sum_{i=1}^{n} (x_i - c)^2$ be equal to $\sum_{i=1}^{n} (x_i - \overline{x})^2$

The mean of the set is the value that minimizes the sum of the standard deviation squared.

Proof:

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

Now simplify the middle term:

$$\sum_{i=1}^{n} (x_i - \overline{x}) (\overline{x} - c) = \overline{x} \sum_{i=1}^{n} x_i - c \sum_{i=1}^{n} x_i - \overline{x} \sum_{i=1}^{n} \overline{x} + \overline{x} \sum_{i=1}^{n} c$$

$= n\,\overline{x}^2 - cn\overline{x} - nx^2 + n\overline{x}c$

$= 0$

So now, $\sum_{i=1}^{n} (x_1 - \overline{x} + \overline{x} - c)^2$ can be written as:

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

The above expression is minimized when $c = \overline{x}$.