Maternal Smoking and Infant Death Statistical Analysis

Sophia Vo, Katelyn Zhao

2024-10-13

1. Introduction

This report aims to present the results of a statistical analysis done on the differences in weight between babies born to mothers who smoked during pregnancy and those who did not.

Our analysis aims to answer the question: Is there a significant difference in birth weight between babies born to women who smoked and those who did not? Using numerical, graphical, and incidence comparison approaches to answering this question, we found that [our results].

The dataset used for this analysis contains information from 1236 babies; all male, single births, and lived up to 28 days. Variables in the dataset include birth weight, length of gestation, whether or not this was the mother's first pregnancy, mother's age, mother's height, mother's weight, and whether or not the mother was a smoker.

The report is structured as follows: Section 2 covers the statistical analysis processes and results and Section 3 discusses the conclusions and implications of the results.

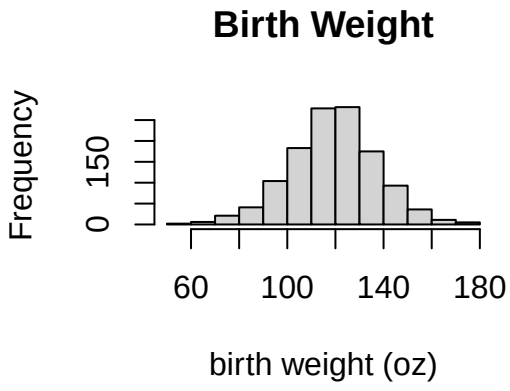
2. Analysis

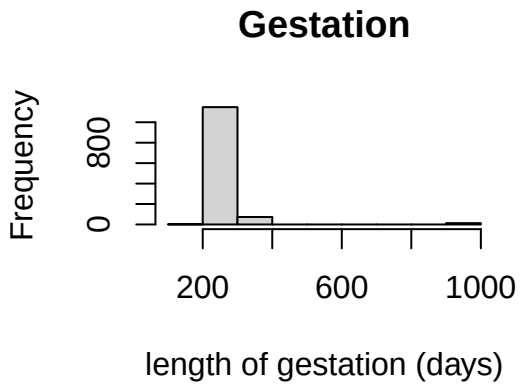
2.1. Data Processing

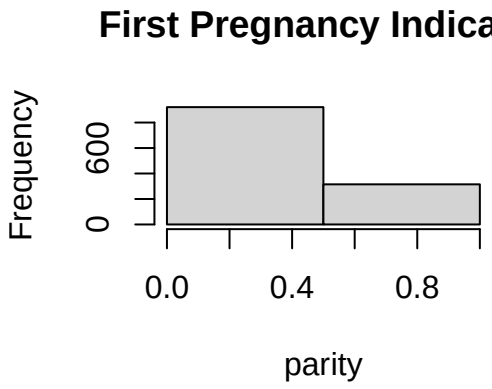
2.1.1. Methods

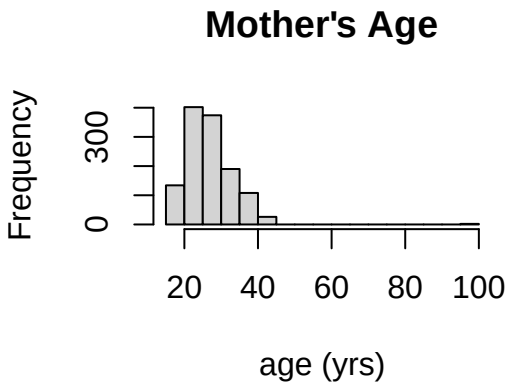
The first step of our analysis was to examine the dataset and understand the characteristics and distributions of each variable. We reviewed all variables and determined their type (numerical or categorical) and identified any outliers or irregularities. To do this, we looked at the distributions of each variable graphically using a histogram.

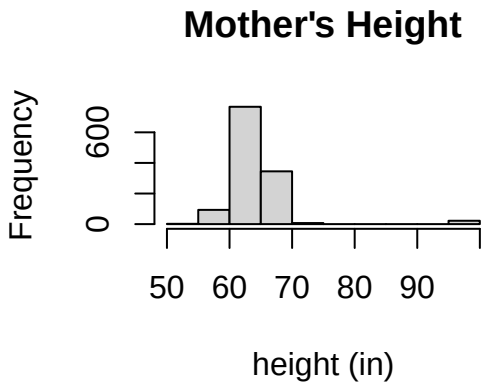
2.1.2. Analysis

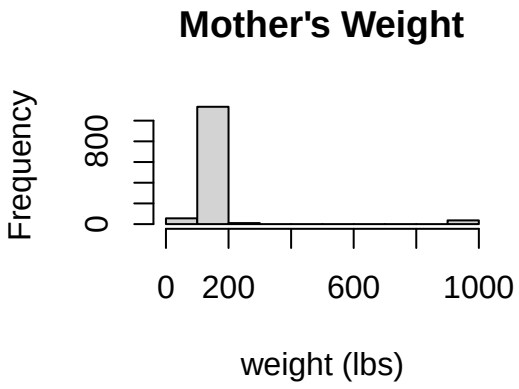


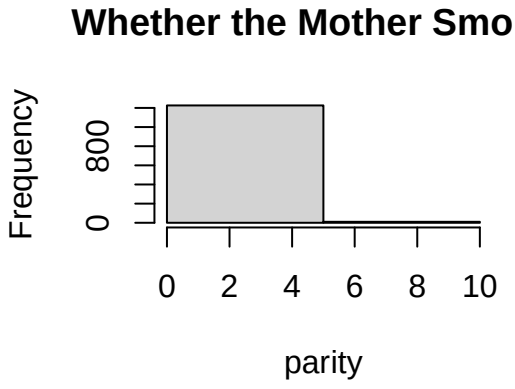






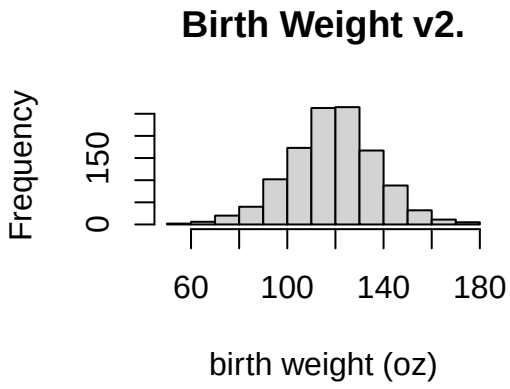


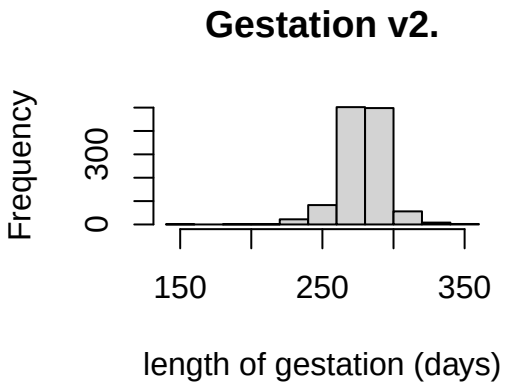


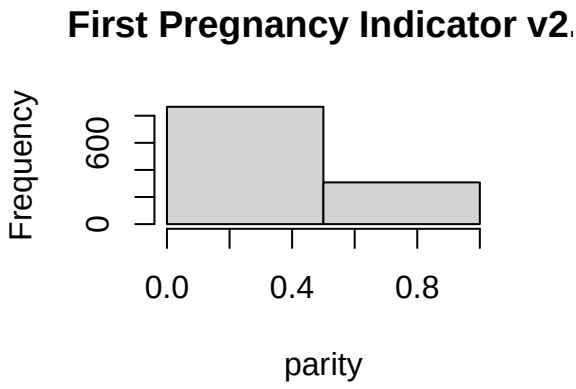


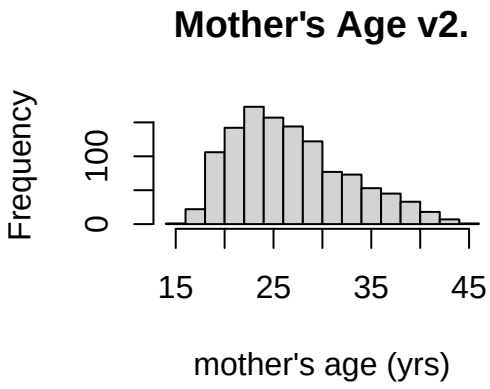
While looking at the histograms, we noticed that some of the data in the dataset were outside of a reasonable range for those variables. Some women had listed gestation periods of up to 1000 days while others were described as over 8 feet or up to 1000 pounds. So, while cleaning the dataset, we removed these outliers to complete the analysis without them.

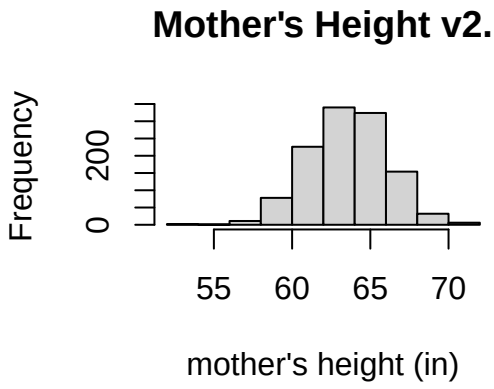
After cleaning the data, we plotted the same histograms again with the cleaned dataset.

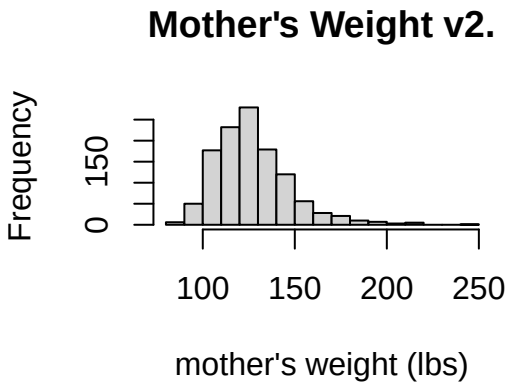

Birth Weight v2.
Frequency
0
150
60
100
140
180
birth weight (oz)

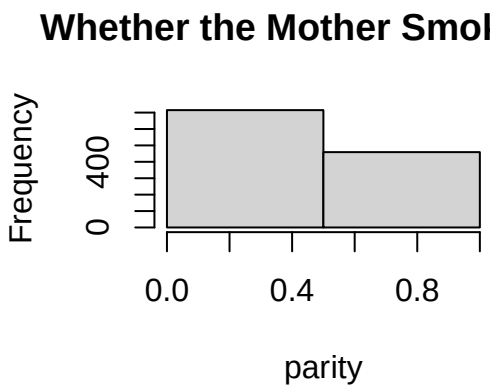

Gestation v2.
Frequency
0
300
150
250
350
length of gestation (days)


First Pregnancy Indicator v2.
Frequency
0
600
0.0
0.4
0.8
parity


Mother's Age v2.
Frequency
0
100
15
25
35
45
mother's age (yrs)


Mother's Height v2.
Frequency
0
200
55
60
65
70
mother's height (in)


Mother's Weight v2.
Frequency
0
150
100
150
200
250
mother's weight (lbs)



2.1.3. Conclusion

We concluded that the dataset does not qualify as a simple random sample because of the method in which it was collected. In order to qualify as a simple random sample, every pregnant women would need to have had an equal chance of being selected. However, in this dataset the pregnant women selected were limited to a certain time frame (1960-1967) and were limited to a specific health plan in a single region. Therefore, this dataset does not qualify as a simple random sample. This means that the dataset is not representative of the entire population and the results of the analysis cannot be generalized.

2.2. Birth Weight Distributions

2.2.1. Methods

Now that we had examined the dataset and each variable as a whole, we looked at the two distributions of birth weights for babies born to women who smoked during their pregnancy and for babies born to women who did not smoke during their pregnancy. Summarizing these distributions numerically, we calculated the minimum, maximum, mean, median, standard deviation, and quartile values.

2.2.2. Analysis

We found that the numerical summaries were as follows:

Table 1: Numerical Differences in Distributions of Smoking and Non-Smoking Mothers

	Smokers	Non_Smokers
Minimum	58.0000	55.0000
Maximum	163.0000	176.0000
Mean	113.8192	123.0853
Median	115.0000	123.0000
Standard Deviation	18.0989	17.3987

Q1	101.0000	113.0000
Q3	126.0000	134.0000

Most of these numerical summaries show that the birth weights of babies of mothers who are non-smokers are larger than the birth weights of babies of mothers who are smokers. To determine whether or not those differences are significant would require further analysis.

2.2.3. Conclusion

We found that the mean and median birth weights of the smokers were 114.1095 and 115 respectively while the mean and median birth weights of the non-smokers were 123.0472 and 123 respectively. Since the means and medians of the two distributions were very similar, we can infer that the shapes of the two distributions are almost symmetrical.

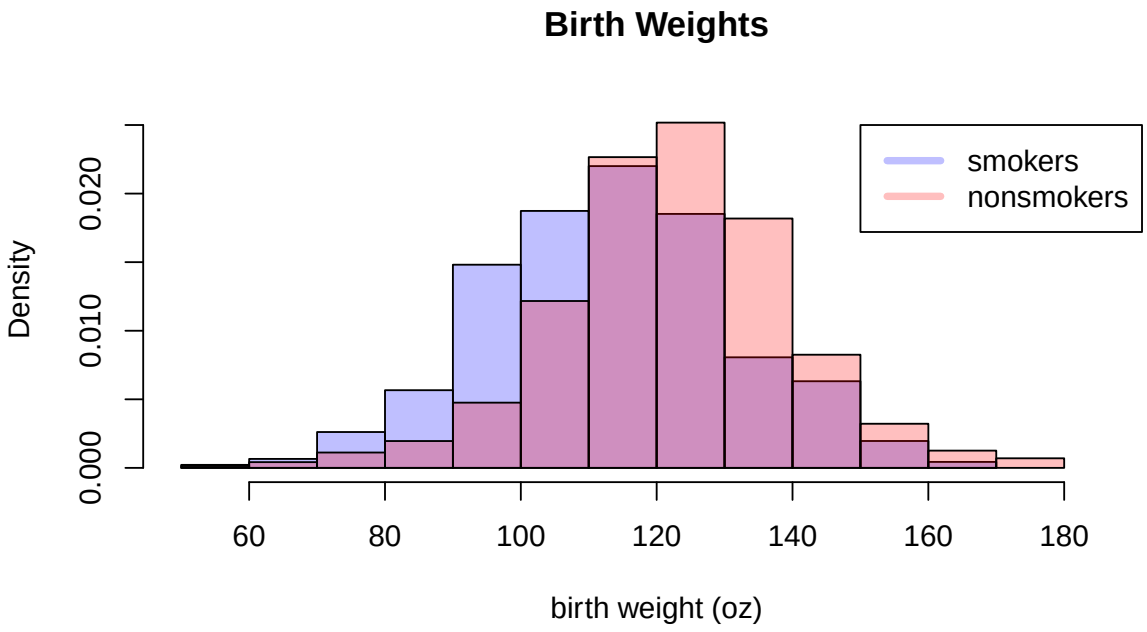
2.3. Graphical Analysis

2.3.1. Methods

Now that we've summarized the two distributions numerically, we also want to look at them graphically.

2.3.2. Analysis

We plotted two graphs, one with the birth weight distribution of women who are non-smokers and another with the birth weight distribution of women who are smokers. We then layered the two histograms for direct comparison.



2.3.3. Conclusion

Based on the layering of the two distributions, the distribution of the baby weights of the mothers who are smokers is shifted to the left of the the distribution of the baby weights of the mothers who are non-smokers. This shows that overall, the baby weights of mothers who are smokers is less than the baby weights of mothers who are non-smokers. To determine whether or not this amount is significant, further analysis would need to be done.

2.4. Incidence of Low Birth Weights

2.4.1. Methods

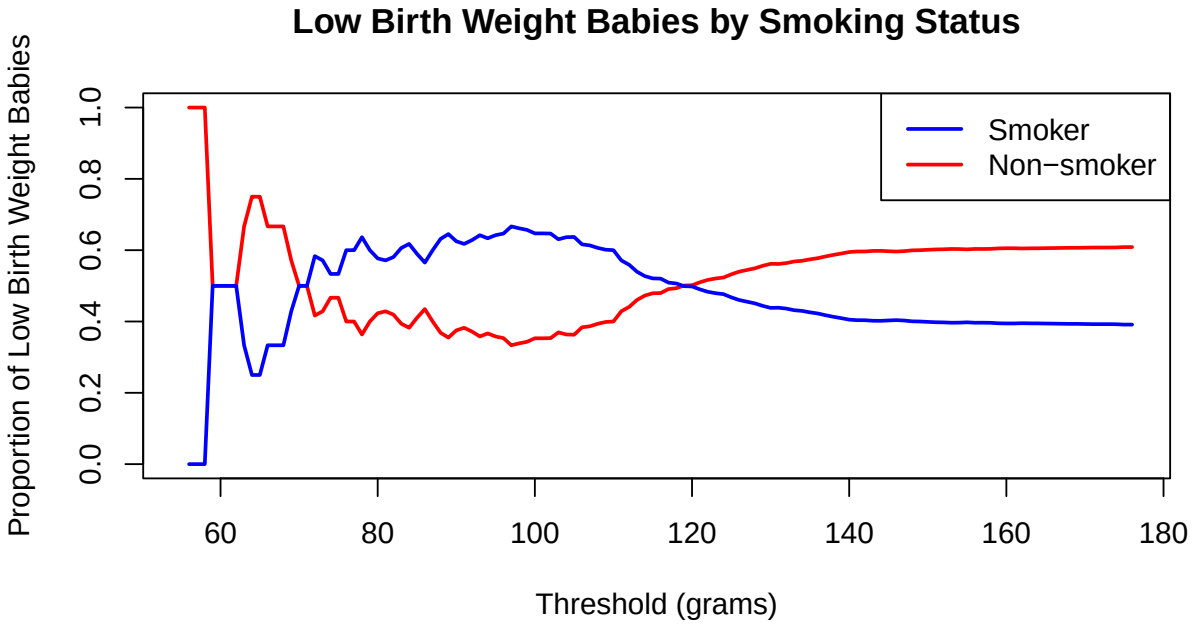
Now that we've done both numerical and graphical analysis, we want to use one final comparison approach: incidence comparison. Using 100 ounces as the threshold of a low-weight baby, we found the percentages of babies that weigh under that threshold for both smoking and non-smoking mothers. We then adjusted the threshold to observe any changes in the incidences of low-weight births.

2.4.2. Analysis

We found that the percentages of low-weight babies are as follows:

Percentage of low-weight babies in smoking mothers: 64.70588%

Percentage of low-weight babies in non-smoking mothers: 35.29412%



2.4.3. Conclusion

3. Discussion