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 both graphical and numerical 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.
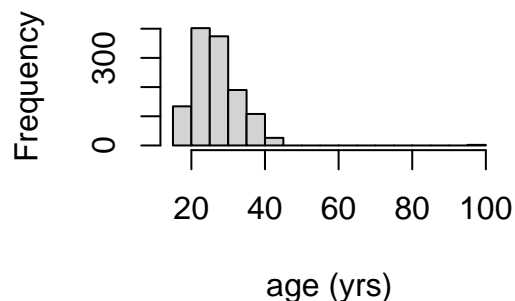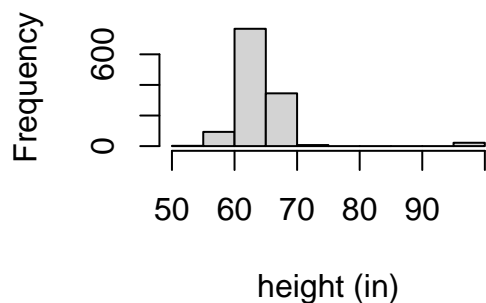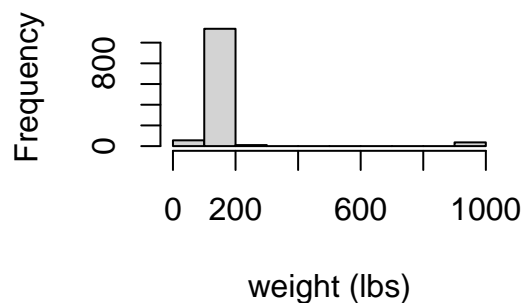
2.1.2. Analysis
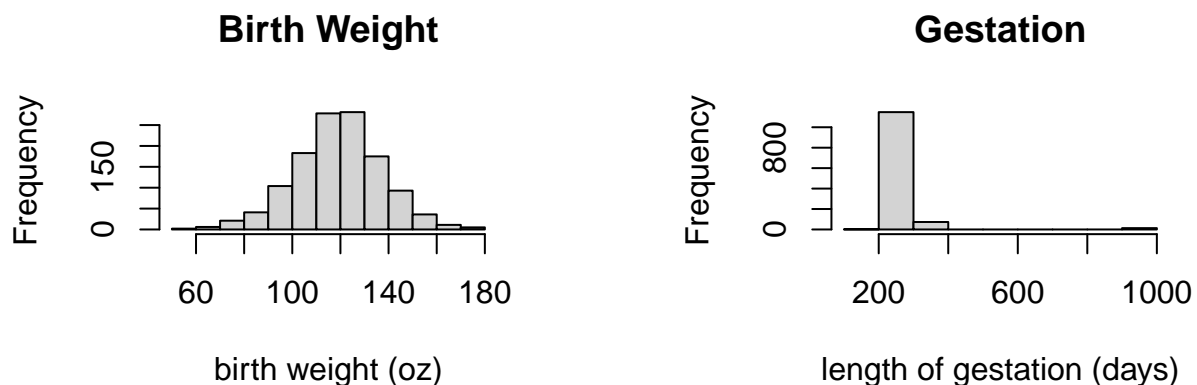
First Pregnancy Indicator
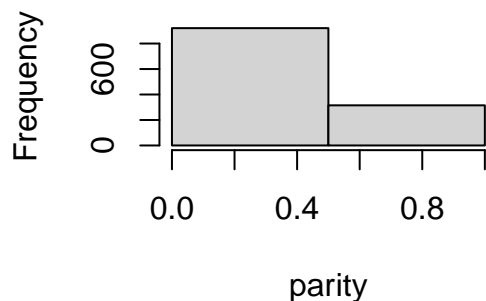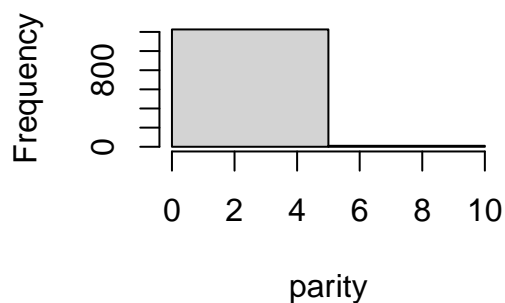

Mother's Age

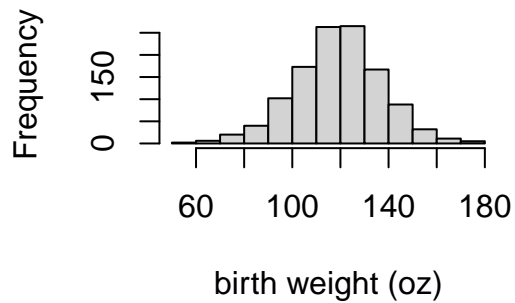## Mother's Height

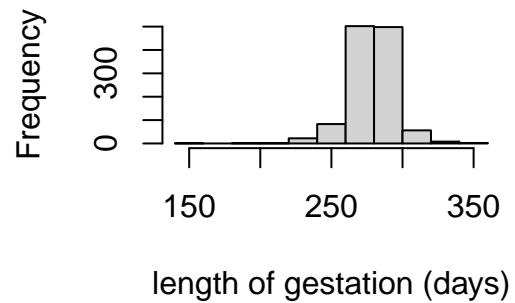## Mother's Weight

Whether the Mother Smokes

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.
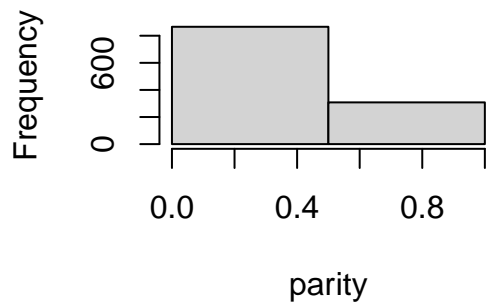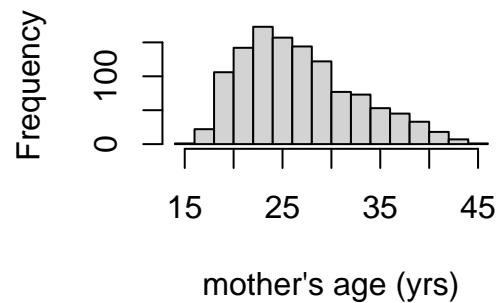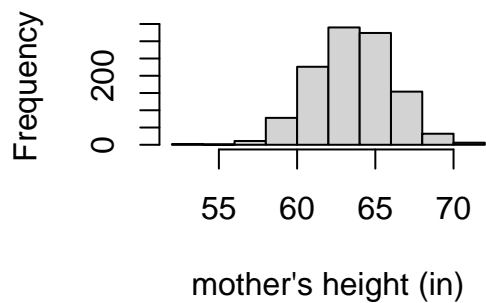
Birth Weight v2.

Gestation v2.

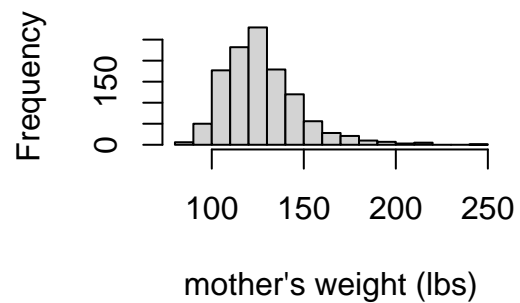## First Pregnancy Indicator v2.

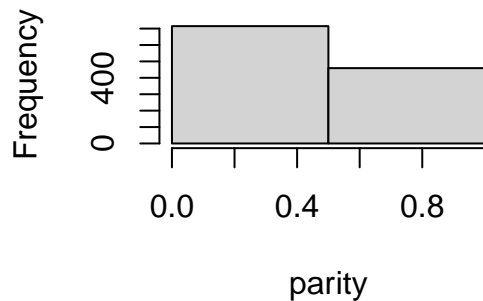## Mother's Age v2.

Mother's Height v2.

Mother's Weight v2.

Whether the Mother Smokes v

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:

Minimum birth weight among smokers: 58 oz Minimum birth weight among non-smokers: 55 oz

Maximum birth weight among smokers: 163 oz Maximum birth weight among non-smokers: 176 oz

Mean birth weight among smokers: 113.8192 oz Mean birth weight among non-smokers: 123.0853 oz

Median birth weight among smokers: 115 oz Median birth weight among non-smokers: 123 oz

Standard deviation among smokers: 18.0989 oz Standard deviation among non-smokers: 17.3987 oz

Q1 birth weight among smokers: 101 oz Q1 birth weight among non-smokers: 113 oz

Q3 birth weight among smokers: 126 oz Q3 birth weight among non-smokers: 134 oz

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.

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.
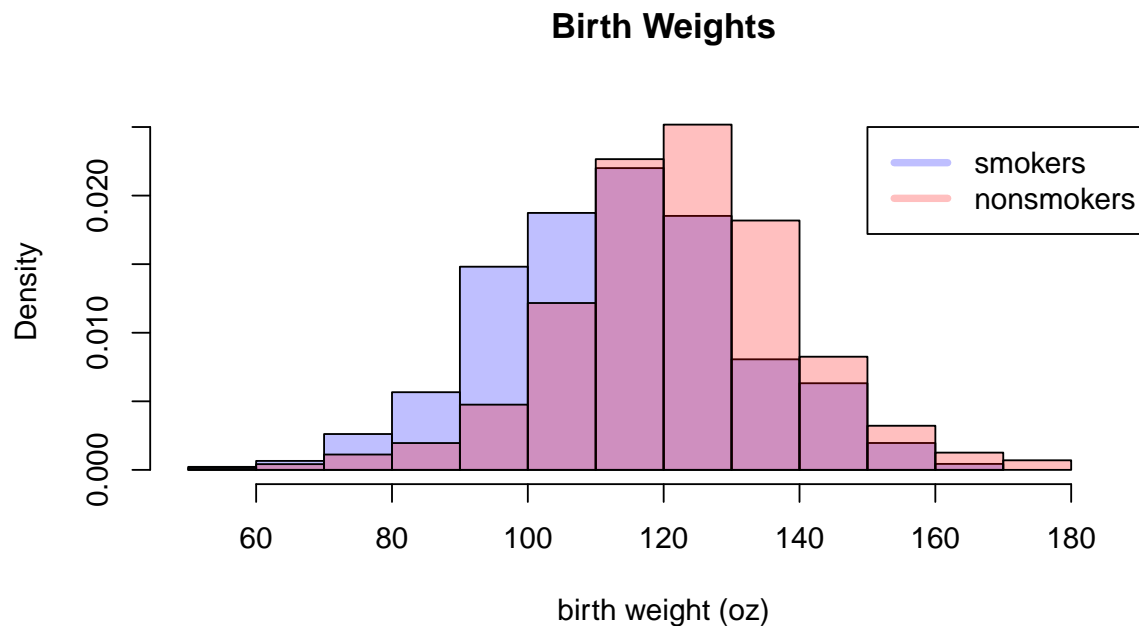
2.3.3. Conclusion

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

2.4.2. Analysis

2.4.3. Conclusion

3. Discussion