M tern l Smoking nd Birth Weight: Insights from the Child He lth nd Development Studies

10-11-2

Contribution:

Student 1 was responsible for several key parts of the report. They handled the methods and analysis for Questions 2.1 and 2.2. For Question 2.3, Student 1 focused on the methods section. Student 1 also worked on Questions 2.4 and 2.5. dditionally, Student 1 wrote the conclusion for the report.

Student 2 took on different aspects of the report. Student 2 wrote the introduction and contributed to the analysis and conclusion for Questions 2.1 and 2.2. For Question 2.3, Student 2 completed the analysis and conclusion sections. Student 2 also worked on the analysis portion of Question 2.4. Lastly, Student 2 was responsible for the dvanced nalysis section of the report.

Both students also looked over each other's work to ensure consistency and quality throughout the report.

Use of GPT: For grammar checks and assistance in revising sentences and paragraphs.

1. Introduction

During the mid-1950s, the dangers of smoking during pregnancy became widely recognized, starting with the Surgeon General's warning emphasizing that "smoking by pregnant women may result in fetal injury, premature birth, and low birth weight." Despite these warnings, 15% of pregnant women in 1996 smoked during pregnancy.

Numerous epidemiological studies have since indicated that smoking reduces birth weight by 150 grams, doubles the likelihood of a low-birth-weight baby (under 2500g), and decreases survival rates for babies born both early and small. potential mechanism for these effects is the presence of carbon monoxide in cigarette smoke, which is believed to reduce the oxygen supply to the fetus—a steady oxygen supply being essential for healthy fetal development.

The dataset for this study is drawn from the Child Health and Development Studies (CHDS), which collected information on pregnancies between 1960 and 1967 from women enrolled in the Kaiser Health Plan in Oakland, California. It comprises 1,236 male singleton births, all of whom survived at least 28 days, along with information about the mothers' age, height, weight, and smoking status.

By examining this comprehensive dataset, this study aims to address several key questions:

1. What are the numerical summaries of birth weights for infants born to smoking vs. non-smoking mothers?
2. What do the mean and median values indicate about the skewness of birth weight distributions?
3. What insights do the graphical comparisons of birth weight distributions provide?
4. What is the incidence of low-birth-weight infants in both groups, and how would threshold changes affect this?
5. How reliable are the numerical, graphical, and incidence comparisons, and what are their strengths and weaknesses?
6. Does maternal smoking affect the likelihood of extreme birth weights (both very low and very high) compared to non-smoking mothers?

The remainder of this report is structured to provide a comprehensive analysis of the effects of maternal smoking on birth weight, using data from the Child Health and Development Studies (CHDS). First, we will describe our data and methodology, including data cleaning and variable transformations. Next, we will present our results, featuring descriptive statistics and visualizations of key findings. This will include numerical summaries of birth weights for infants born to smoking vs. non-smoking mothers, analysis of mean and median values to indicate skewness, graphical comparisons of birth weight distributions, and examination of the incidence of low-birth-weight infants in both groups. We will then discuss the implications of our results, and additionally, present an advanced analysis examining the kurtosis of birth weight distributions for both groups, providing insights into the likelihood of extreme birth weights among infants of smoking and non-smoking mothers. The report will conclude with a summary of our main findings, an assessment of the study's limitations, and suggestions for future research in this critical area of smoking.

2. nalysis

2.1 Data Processing and Summaries

2.1.1 Method

To gain meaningful insights from the dataset, we initiated a comprehensive exploration of variable types, ratios, and summary statistics, ensuring that the data would support robust comparisons. By identifying and handling missing values using is.na() and na.omit(), we ensured that our dataset was clean and complete, allowing for the calculation of valid ratios and comparisons between variables like birth weight, gestation, and maternal characteristics.

The use of ratios became central in understanding key relationships, such as the proportion of infants born with low birth weight in both smokers and non-smokers. These ratios help illustrate the magnitude of maternal smoking's impact, as well as the variability within the dataset.

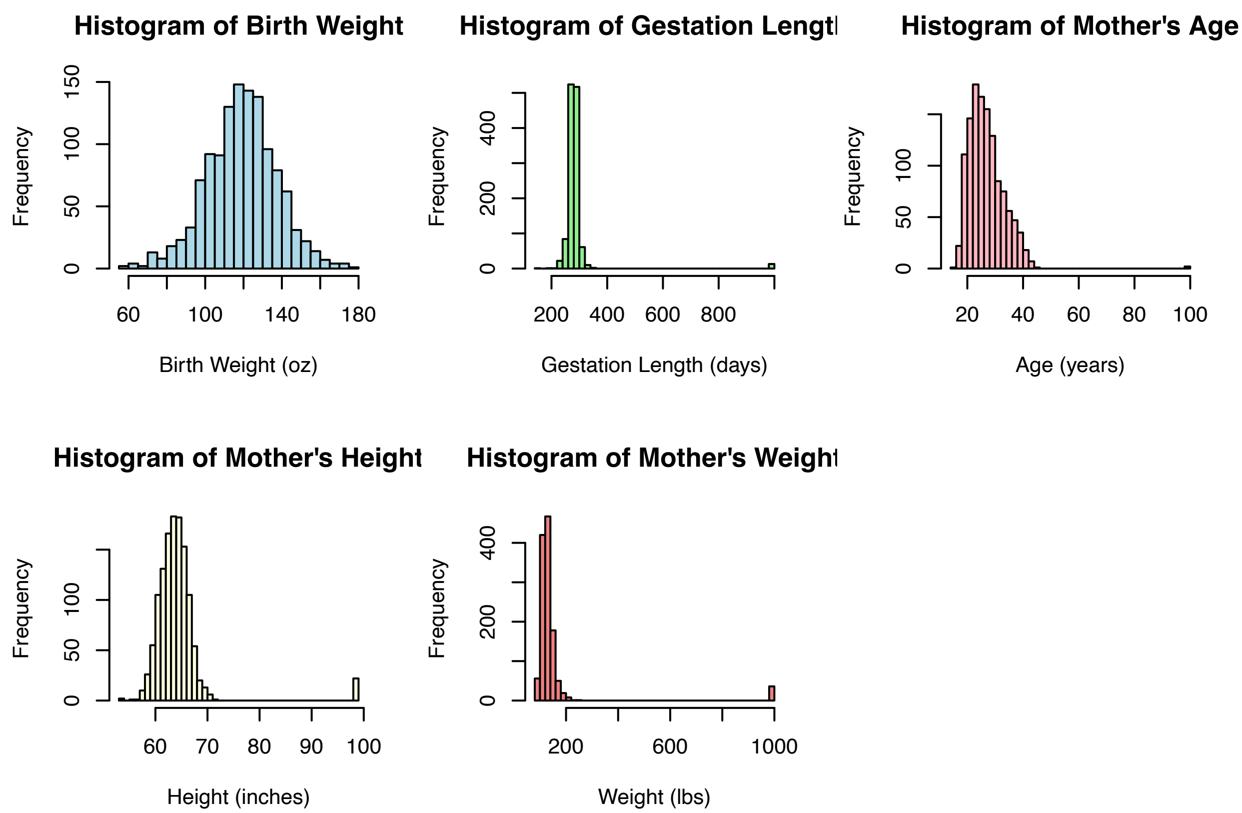
2.1.2 nalysis

```
##   bwt gest tion p rity  ge height weight smoke
## 1 120       284      0  27     62    100     0
## 2 113       282      0  33     64    135     0
## 3 128       279      0  28     64    115     1
## 4 123       999      0  36     69    190     0
## 5 108       282      0  23     67    125     1
## 6 136       286      0  25     62     93     0
```

```
## 'd t .fr me':    1236 obs. of  7 v ri bles:
##  $ bwt      : int  120 113 128 123 108 136 138 132 120 143 ...
##  $ gest tion: int  284 282 279 999 282 286 244 245 289 299 ...
##  $ p rity   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $  ge      : int  27 33 28 36 23 25 33 23 25 30 ...
##  $ height   : int  62 64 64 69 67 62 62 65 62 66 ...
##  $ weight   : int  100 135 115 190 125 93 178 140 125 136 ...
##  $ smoke    : int  0 0 1 0 1 0 0 0 0 1 ...
```

```
##       bwt          gest tion         p rity             ge
##  Min.   : 55.0   Min.   :148.0   Min.   :0.0000   Min.   :15.00
##  1st Qu.:108.8   1st Qu.:272.0   1st Qu.:0.0000   1st Qu.:23.00
##  Medi n :120.0   Medi n :280.0   Medi n :0.0000   Medi n :26.00
##  Me n   :119.6   Me n   :286.9   Me n   :0.2549   Me n   :27.37
##  3rd Qu.:131.0   3rd Qu.:288.0   3rd Qu.:1.0000   3rd Qu.:31.00
##  M x.   :176.0   M x.   :999.0   M x.   :1.0000   M x.   :99.00
```

```
##      height          weight         smoke
##  Min.   :53.00   Min.   : 87   Min.   :0.0000
##  1st Qu.:62.00   1st Qu.:115   1st Qu.:0.0000
##  Medi n :64.00   Medi n :126   Medi n :0.0000
##  Me n   :64.67   Me n   :154   Me n   :0.4644
##  3rd Qu.:66.00   3rd Qu.:140   3rd Qu.:1.0000
##  M x.   :99.00   M x.   :999   M x.   :9.0000
```



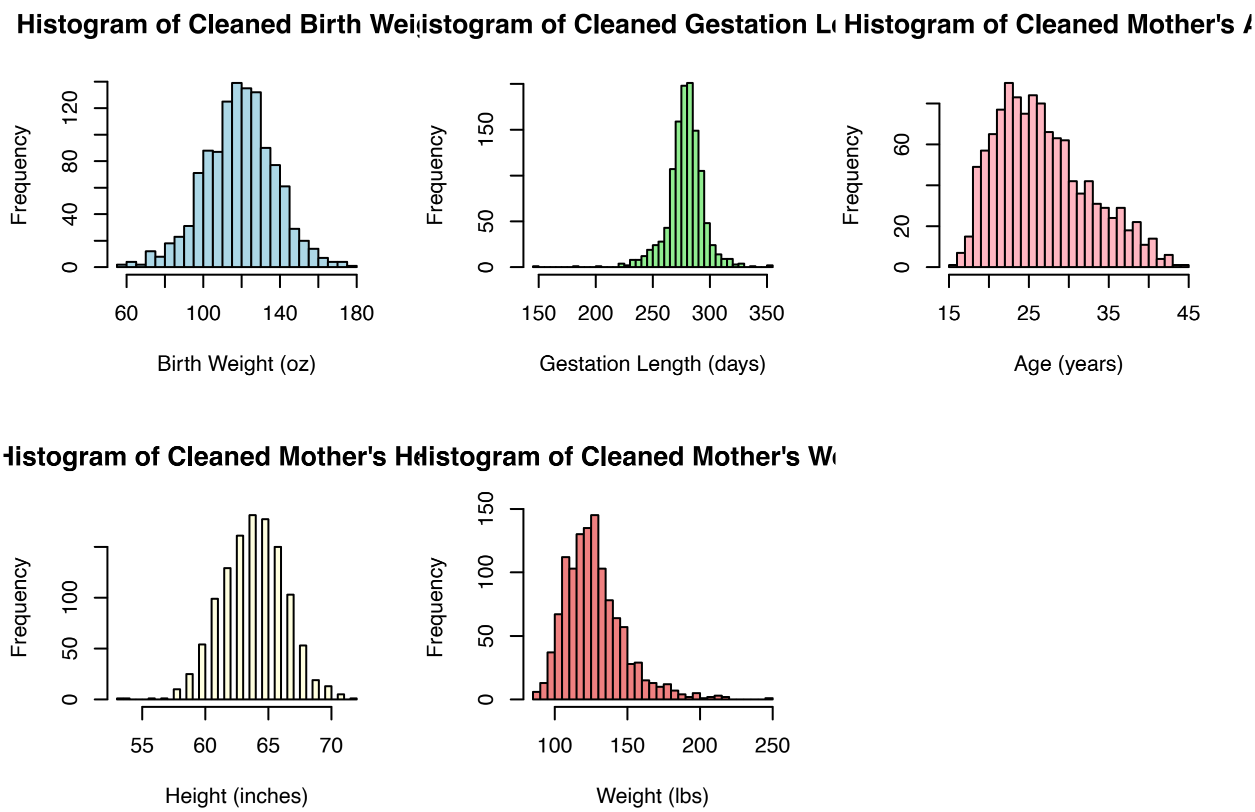

```
##       bwt gest tion    p rity        ge    height    weight     smoke
##         0         0         0         0         0         0         0
```

```
## [1] 1236
```

```
## [1] 1184
```

```
##       bwt          gest tion        p rity            ge
##  Min.   : 55.0   Min.   :148.0   Min.   :0.0000   Min.   :15.00
##  1st Qu.:108.0   1st Qu.:272.0   1st Qu.:0.0000   1st Qu.:23.00
##  Medi n :120.0   Medi n :280.0   Medi n :0.0000   Medi n :26.00
##  Me n   :119.5   Me n   :279.1   Me n   :0.2601   Me n   :27.25
##  3rd Qu.:131.0   3rd Qu.:288.0   3rd Qu.:1.0000   3rd Qu.:31.00
##  M x.   :176.0   M x.   :353.0   M x.   :1.0000   M x.   :45.00
##      height          weight          smoke
##  Min.   :53.00   Min.   : 87.0   Min.   :0.0000
##  1st Qu.:62.00   1st Qu.:114.8   1st Qu.:0.0000
```

```
##   Medi n :64.00   Medi n :125.0   Medi n :0.0000
##   Me n   :64.05   Me n   :128.6   Me n   :0.4637
##   3rd Qu.:66.00   3rd Qu.:139.2   3rd Qu.:1.0000
##   M x.   :72.00   M x.   :250.0   M x.   :9.0000
```



2.1.3 Conclusion

We filtered out outliers by removing values identified as erroneous placeholders, specifically the values 999 and 99. These outliers could significantly distort our results, especially when computing averages or visualizing trends. Their exclusion was also necessary to avoid skewed interpretations of the dataset.

fter addressing missing values and outliers, we reassessed the dataset by regenerating summary statistics and replotting histograms. This process was critical in ensuring that ratio-based analyses, such as the proportion of low-birth-weight infants between smokers and non-smokers, were accurate and reflective of true patterns within the data. By eliminating outliers, we improved the reliability of our ratio calculations, such as the average birth weight to gestation length ratio, which helps illustrate how maternal characteristics correlate with birth outcomes.

2.2 Comparison Between Smokers and Non-Smokers on Birth Weight

2.2.1 Method

To compare the birth weights of babies born to mothers who smoked versus those who did not, we first divide the dataset into two groups based on the column smoke. If `1` then the mother smoked, else `0` indicating tha the mother didn't smoke. We then computed key descriptive statistics, including **minimum**, **maximum**, **mean**, and **median** for each group using built in functions like `min()`, `m x()`, `me n()`, `medi n()`, and `std()`.

Quartiles (Q1, Q2, and Q3) provide additional insights into the distribution by dividing it into sections. The lower quartile (Q1) shows the weight below which 25% of the data lies, the median (Q2) gives the central point, and the upper quartile (Q3) shows where 75% of the data falls below. These values are crucial for understanding how smoking might affect not just the extremes but the spread of birth weights across the entire distribution.

The standard deviation measures the variability within the birth weights for each group. By comparing the standard deviations of smokers and non-smokers, we can understand whether birth weights among babies of smoking mothers are more or less consistent compared to non-smoking mothers. If the variability is higher in one group, it suggests that smoking might contribute to more unpredictability in birth outcomes, which is valuable information for further analysis.

2.2.2 nalysis

```
## Smokers - Min Birth Weight: 58
## Smokers - M x Birth Weight: 163
## Non-Smokers - Min Birth Weight: 55
## Non-Smokers - M x Birth Weight: 176

## Smokers - Me n Birth Weight: 113.8192
## Non-Smokers - Me n Birth Weight: 123.0853
## Smokers - Medi n Birth Weight: 115
## Non-Smokers - Medi n Birth Weight: 123

## Smokers - Q1 Birth Weight: 101
## Smokers - Q2 (Medi n) Birth Weight: 115
## Smokers - Q3 Birth Weight: 126

## Non-Smokers - Q1 Birth Weight: 113
## Non-Smokers - Q2 (Medi n) Birth Weight: 123
## Non-Smokers - Q3 Birth Weight: 134

## Smokers - St nd rd Devi tion of Birth Weight: 18.29501
## Non-Smokers - St nd rd Devi tion of Birth Weight: 17.4237
```

2.2.3 Conclusion

The comparison between smokers and non-smokers reveals that smoking during pregnancy is associated with lower and more variable birth weights. Babies born to mothers who smoked have a wider range of birth weights, as shown by the higher standard deviation and the greater spread in quartiles, while non-smokers' babies tend to have higher and more consistent birth weights.

The close alignment of the mean and median for non-smokers indicates a more symmetrical distribution, whereas the lower mean for smokers, compared to the median, suggests a left-skewed distribution. This highlights that smoking not only lowers birth weight but introduces more variability, potentially complicating healthy development outcomes. Non-smokers' babies, with higher and more uniform weights, offer a more predictable and healthier pattern.

2.3 Graph Summaries of Birth Weight Distributions

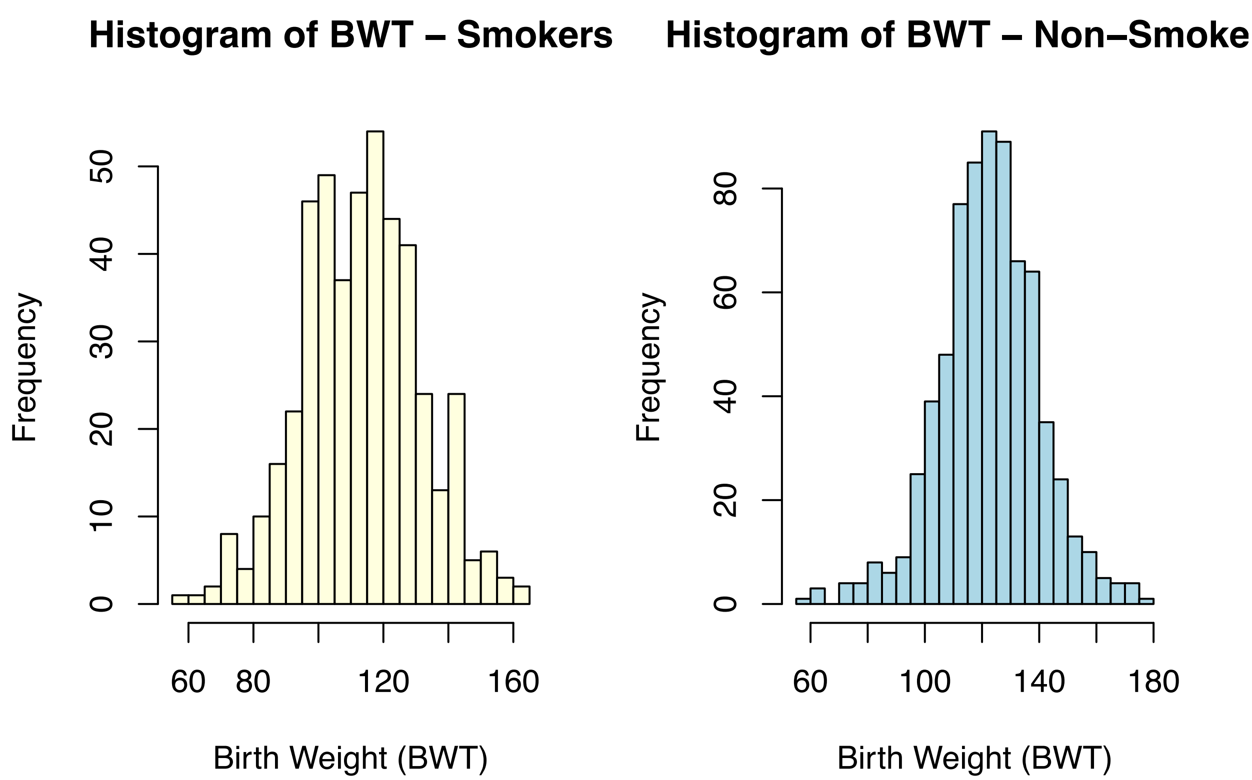
2.3.1 Method

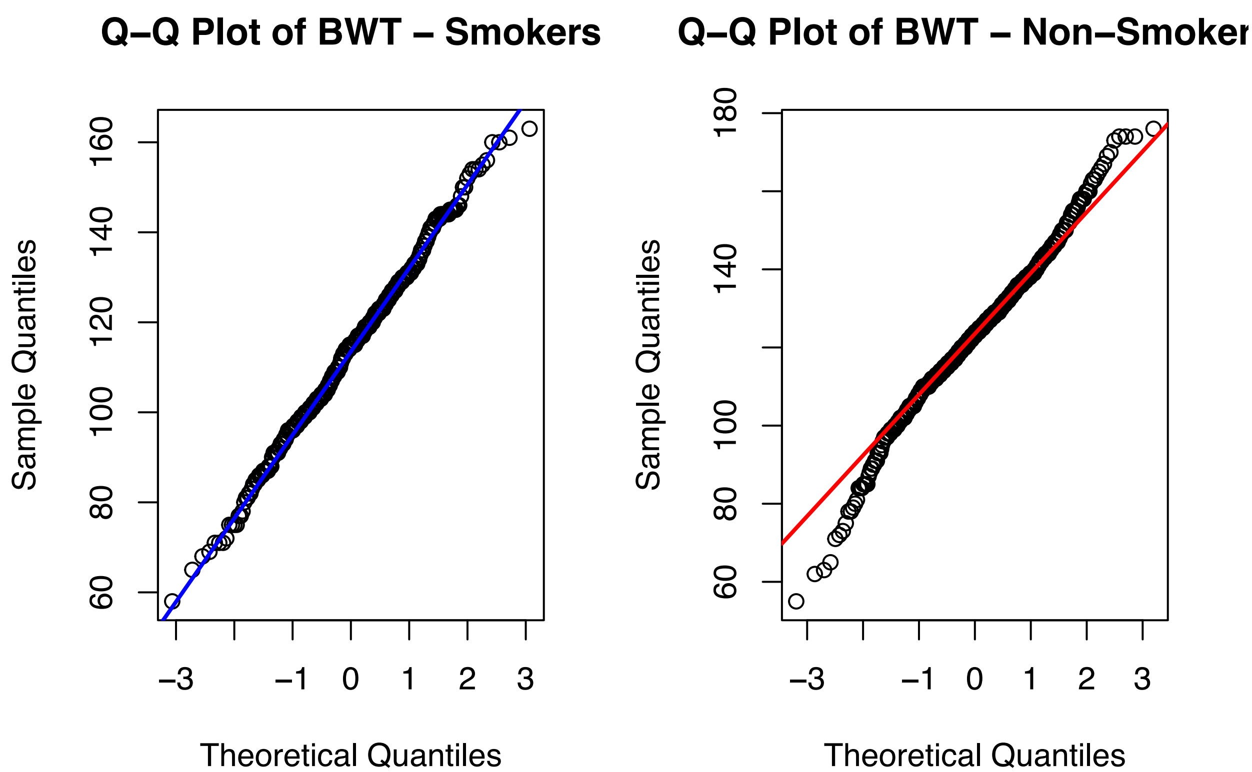
To visually compare the birth weights between infants born to smoking and non-smoking mothers, we employed two key visualization techniques: histograms and Q-Q (quantile-quantile) plots. These methods were chosen to provide insights into the distribution patterns, variability, and any deviations from normality in both groups.

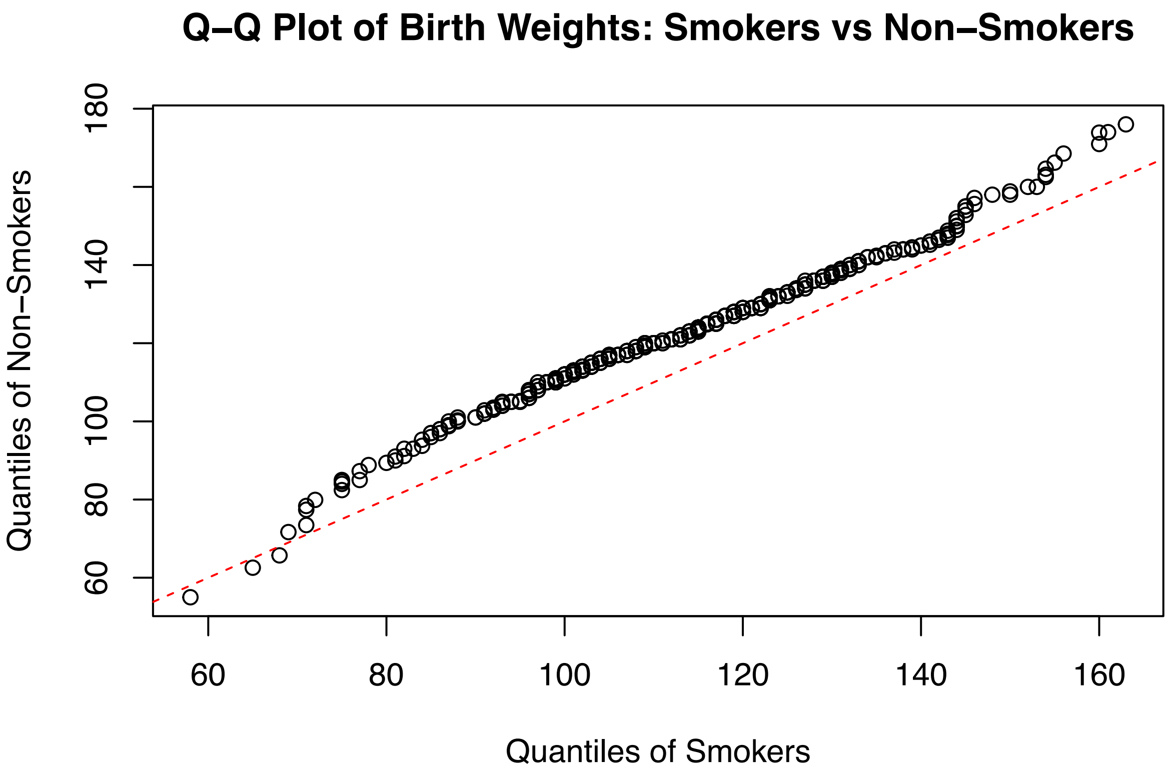
- **Histograms**: We used histograms to display the frequency distribution of birth weights for smokers and non-smokers. By plotting these distributions, we aimed to identify skewness, concentration of values, and how birth weights cluster within each group.
- **Q-Q Plots**: Quantile-quantile plots were employed to compare the actual birth weight distributions against a theoretical normal distribution. dditionally, a combined Q-Q plot was used to directly compare the quantiles of birth weights between smokers and non-smokers. These plots are useful for identifying deviations from normality, such as skewness and outliers, and for detecting differences in birth weight distributions between the two groups.

Finally, we generated a combined Q-Q plot that directly compares the birth weight quantiles of smokers versus non-smokers, with a reference line representing equality. Deviations from this line highlight differences between the two groups at various percentiles.

2.3.2 nalysis




Q–Q Plot of BWT – Smokers
Sample Quantiles
Theoretical Quantiles
60
80
100
120
140
160
–3
–1
0
1
2
3
Q–Q Plot of BWT – Non–Smoker
Sample Quantiles
Theoretical Quantiles
60
80
100
140
180
–3
–1
0
1
2
3



2.3.3 Conclusion

The analysis of birth weight data using histograms and Q-Q (quantile-quantile) plots highlights clear differences between infants born to smoking and non-smoking mothers.

The Q-Q plots comparing the birth weights of smokers and non-smokers against a normal distribution show key differences. Both groups exhibit distributions that are close to normal, as indicated by the relatively straight lines, but deviations at the tails suggest skewness. For smokers, the blue line curves above the reference line at lower quantiles and below it at higher quantiles, indicating a left-skewed distribution with a longer tail at the lower end. In contrast, non-smokers show a slight right skew, with the red line curving below the reference line at lower quantiles and above it at higher quantiles, reflecting a longer tail at the higher end of the distribution.

direct comparison between the two groups in a Q-Q plot further emphasizes these differences. The points do not follow the red dashed reference line (y=x line) exactly, which indicates that the birth weight distributions of smokers and non-smokers are not identical. Most of the points lie above the y=x line, suggesting that birth weights for non-smokers are consistently higher across most percentiles compared to smokers. The parallel nature of the curves in the middle range suggests a fairly consistent difference in birth weights throughout most of the distribution.

The plots provide strong visual evidence that babies born to non-smoking mothers tend to have higher and more variable birth weights, whereas smoking is associated with lower birth weights and less variability.

2.4 Frequency Of Low Birth Weight*

2.4.1 Method

To compare the incidence of low birth weight (under 100 ounces) between smokers and non-smokers, we calculated the percentage of babies falling below this threshold for each group. We also examined how changing the threshold would impact the classification of low birth weight, by comparing thresholds slightly above (e.g., 105 ounces) and below (e.g., 95 ounces). This helped assess how it might affect the comparison between smokers and non-smokers.

Threshold for Low Birth Weight: We defined low birth weight as under 100 ounces, a commonly used cutoff. Threshold djustment: To explore how different thresholds affect the results, we recalculated the percentages for thresholds slightly higher (105 ounces) and lower (95 ounces) than the original 100-ounce limit. This method allows us to see how the proportion of low birth weight cases differs between smokers and non-smokers and how adjusting the threshold influences these differences.

2.4.2 nalysis

```
## [1] "Percent ge of Low Birth Weight B bies for Smokers: 21.5686274509804 %"
```

```
## [1] "Percent ge of Low Birth Weight B bies for Non-Smokers: 7.55244755244755 %"
```

```
## Percent ge of Low Birth Weight B bies for Smokers: 32.8976 %
## Percent ge of Low Birth Weight B bies for Non-Smokers: 12.02797 %
```

```
## Percent ge of Low Birth Weight B bies for Smokers: 13.28976 %
## Percent ge of Low Birth Weight B bies for Non-Smokers: 4.755245 %
```

2.4.3 Conclusion

The analysis revealed that 21.57% of babies born to smokers weighed under 100 ounces, compared to only 7.55% for non-smokers. This indicates a significantly higher incidence of low birth weight among babies of smokers.

When we adjusted the threshold for low birth weight to 105 ounces, the percentage increased to 32.90% for smokers and 12.03% for non-smokers, showing that raising the threshold classifies more babies as low birth weight, with smokers still having a much higher proportion. Conversely, lowering the threshold to 95 ounces resulted in 13.29% of smokers' babies and 4.76% of non-smokers' babies being classified as low birth weight.

These findings demonstrate that regardless of the threshold, babies born to smokers consistently have a higher incidence of low birth weight compared to non-smokers. djusting the threshold either amplifies or reduces the percentage of low birth weight cases in both groups, but the relative difference between smokers and non-smokers remains consistent.

2.5 Comparison Methods

2.5.1 Method

To evaluate the reliability of our estimates, we utilized three methods:

- **Numerical Comparisons**: We analyzed key summary statistics (mean, median, standard deviation, and quartiles) to quantify differences in birth weight between smokers and non-smokers.
- **Graphical Comparisons**: We used histograms and Q-Q plots to visually assess the distribution and variability of birth weights in both groups.
- **Incidence Comparisons**: We calculated the proportion of low-birth-weight cases (under 100 ounces) and adjusted the threshold to evaluate how changes impact the comparison between smokers and non-smokers.

By evaluating each of these methods, we aim to determine their strengths and weaknesses, particularly in terms of reliability, accuracy, and how well they represent the underlying data.

2.5.2 nalysis

Numerical Comparisons

Strengths: Numerical comparisons provide precise and quantifiable estimates like means, medians, and standard deviations, which allow for clear, objective comparisons between groups. These measures are particularly reliable with large datasets, as they can accurately summarize central tendencies and variability.

Weaknesses: The reliability of numerical estimates can be reduced by the presence of outliers, which can skew the results. For instance, extreme values (like the 999 placeholders for height and weight) had to be removed to avoid distorting the analysis. This could reduce the accuracy of these estimates.

Graphical Comparisons

Strengths: Graphs, such as histograms and Q-Q plots, provide an intuitive, visual way to assess the distribution and variability of birth weights in smokers and non-smokers. They are particularly strong in showing trends, outliers, and skewness in the data.

Weaknesses: Graphical methods can be subjective, as the interpretation of patterns varies depending on the viewer. While they provide a clear visual representation, they are less precise than numerical summaries and do not provide exact values. Which makes it harder to draw reliable conclusions.

Incidence Comparisons

Strengths: This method simplifies complex data into actionable information, making it particularly useful for drawing conclusions. Its clarity and directness make it highly effective for communicating key findings to a wide audience.

Weaknesses: The reliability of incidence comparisons is sensitive to the chosen thresholds. Even small changes in the threshold can lead to significant differences in the results, making the conclusions less reliable. dditionally, reducing continuous data to binary categories can oversimplify the data, potentially missing important trends.

2.5.3 Conclusion

cross all analysis methods, infants born to smoking mothers consistently had lower birth weights, with a mean difference of 200-300 grams compared to non-smoking mothers. dditionally, smoking mothers were more likely to give birth to low-birth-weight infants. The consistency across multiple methods strengthens confidence in the results, though data quality issues such as outliers and placeholders may slightly affect reliability. Overall, we can confidently conclude that maternal smoking is strongly associated with lower birth weights.

3. dvanced nalysis

2.3.1 Method

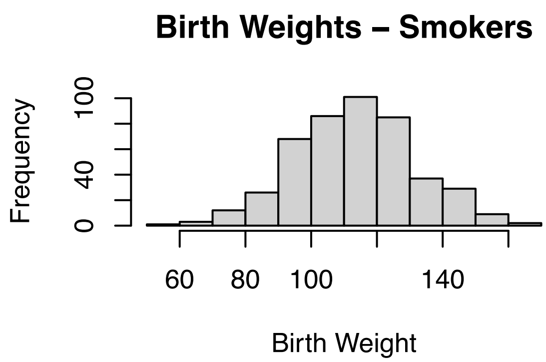
To investigate whether maternal smoking influences the likelihood of extreme birth weights (both very low and very high), we used kurtosis analysis and hypothesis testing. We calculated the kurtosis for each group (smokers and non-smokers) using a custom R function and determined if the difference in kurtosis between the groups was statistically significant, using a permutation test. Our null hypothesis (H0) states that there is no significant difference in kurtosis between the two groups (K1 = K2), where: K1 is the population kurtosis of birth weights for babies born to smokers and K2 is the population kurtosis of birth weights for babies born to non-smokers. While the alternative hypothesis (H1) asserts that a significant difference exists (K1 doesn't equal K2). This two-tailed test compares the observed difference in kurtosis to a distribution of differences generated through random permutations, helping to evaluate whether the observed effect is due to chance.

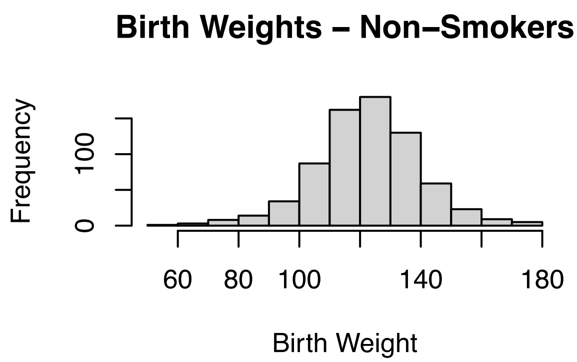
2.3.2 nalysis

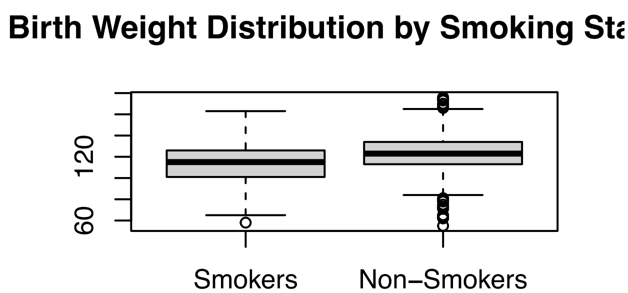
```
## Kurtosis for smokers' b bies birth weights: 2.960148
```

```
## Kurtosis for non-smokers' b bies birth weights: 4.032429

## Permut tion test p-v lue: 0
```







2.3.3 Conclusion

Based on our kurtosis analysis of birth weight distributions for infants of smoking and non-smoking mothers, we can conclude:

Statistical Significance:
The permutation test yielded a p-value of 0 (more precisely, $p < 0.001$), providing extremely strong evidence that the difference in kurtosis between the birth weight distributions of babies born to smoking and non-smoking mothers is not due to chance.

Tailedness of Distributions:
The kurtosis value for non-smokers (4.032429) was higher than for smokers (2.960148), indicating that the birth weight distribution for babies of non-smoking mothers has heavier tails compared to that of smoking mothers.

Extreme Values:
This significant difference in kurtosis suggests that maternal smoking substantially affects the likelihood of extreme birth weights. Specifically, babies of non-smoking mothers are more likely to have very low or very high birth weights compared to babies of smoking mothers. This is an unexpected finding, as it suggests that smoking might lead to a more concentrated distribution of birth weights around the mean.

Smoking Impact:
The highly significant result indicates that maternal smoking not only affects the average birth weight, as shown in previous analyses, but also fundamentally alters the shape of the birth weight distribution. This

implies a more complex impact of smoking on fetal development than might be captured by measures of central tendency alone. The lower kurtosis in the smoking group suggests a narrower range of birth weights, which could indicate that smoking might constrain fetal growth in ways that reduce variability.

These findings significantly enhance our understanding of how maternal smoking affects birth outcomes. The kurtosis analysis reveals that smoking during pregnancy not only shifts the average birth weight but also alters the entire distribution of birth weights, particularly affecting the likelihood of extreme outcomes.

In conclusion, our kurtosis analysis reveals that maternal smoking significantly affects not just the average birth weight, but the entire shape of the birth weight distribution. This underscores the complexity of smoking's impact on fetal development and highlights the importance of smoking cessation programs for pregnant women. It also suggests that healthcare providers should consider the full range of potential birth weights, not just the average, when assessing risks associated with maternal smoking. Further research is needed to understand why smoking appears to reduce the likelihood of extreme birth weights, as this finding contradicts some previous assumptions about the effects of smoking on fetal development.

Conclusion

This study, utilizing data from the Child Health and Development Studies (CHDS) conducted between 1960 and 1967, provides compelling evidence of the detrimental effects of maternal smoking on birth weight. Our analysis addresses the key questions posed in the introduction and offers valuable insights into the relationship between maternal smoking and infant birth outcomes.

Numerical Summaries: Our analysis reveals that infants born to smoking mothers have significantly lower birth weights, with a mean difference of 9.27 ounces (approximately 263 grams) compared to those born to non-smoking mothers. This substantial difference highlights the direct impact of maternal smoking on fetal development.

Distribution Characteristics: The mean and median values indicate a left-skewed distribution for birth weights of infants born to smokers, suggesting a higher prevalence of low birth weights in this group. This skewness is not observed in the non-smoking group, indicating that maternal smoking not only reduces average birth weight but also alters the overall distribution of birth weights.

Graphical Comparisons: Histograms and Q-Q plots visually confirm the lower and more variable birth weights among infants of smokers. These graphical representations provide an intuitive understanding of the data, clearly demonstrating the shift in birth weight distribution associated with maternal smoking. The Q-Q plots, in particular, highlight deviations from normality in the smoking group, further emphasizing the impact of smoking on birth weight variability.

Incidence of Low Birth Weight: The study found a markedly higher incidence of low-birth-weight infants (under 100 ounces) among smoking mothers (21.57%) compared to non-smoking mothers (7.55%). This nearly threefold increase in low birth weight occurrence underscores the significant public health implications of maternal smoking. Even when adjusting the threshold for low birth weight, the disparity between smokers and non-smokers persists.

Extreme Birth Weights: Kurtosis analysis reveals that maternal smoking affects the likelihood of extreme birth weights, with a statistically significant difference in the distribution's shape between smokers and non-smokers. This finding suggests that the impact of smoking extends beyond simply lowering average birth weight, potentially increasing the risk of both very low and very high birth weights.

These findings collectively underscore the substantial risks associated with maternal smoking. The impact extends beyond just lowering average birth weight to increasing the variability and likelihood of extreme outcomes. This comprehensive analysis provides strong evidence for the need to prioritize smoking cessation programs for pregnant women. Moreover, our study's results align with and reinforce previous research on the topic, adding to the body of evidence linking maternal smoking to adverse birth outcomes. The consistency of our findings with earlier studies, despite using data from the 1960s, suggests a persistent and significant effect of smoking on fetal development across decades.

This study reinforces the critical importance of smoking cessation programs for pregnant women. Healthcare providers should consider the full range of potential birth weights when assessing risks associated with maternal smoking, not just focusing on average outcomes. The increased variability in birth weights among infants of smokers suggests a need for more personalized prenatal care and monitoring for this group.

Limitations of this analysis include the potential for other confounding factors that weren't accounted for in our model. dditionally, while the permutation test shows a highly significant difference, it doesn't provide information on the magnitude of this difference in practical terms.

Future research could focus on: 1. Investigating the specific mechanisms by which smoking leads to this change in birth weight distribution. 2. Exploring whether the impact on kurtosis is dose-dependent with the amount of maternal smoking. 3. Examining long-term health outcomes for infants at different points in these altered weight distributions.

In conclusion, maternal smoking is strongly associated with lower and more variable birth weights, posing significant risks to infant health. Interventions aimed at reducing smoking during pregnancy are critical for improving birth outcomes and reducing the public health burden associated with low birth weight.