Statistical Analysis of Maternal Smoking and Infant Birth Weight

Math 189/289C
Homework 1 Report
Jan. 24th, 2017

By:

Emma Roth, 4th Year Computer Science/Bioinformatics Major

- Introduction, T-test, Histograms, and Histogram Analysis

Ileena Mitra, 1st year Bioinformatics PhD

- Box-plot, QQ plot, Conclusion, Formatting/ Editing

Megan Lee, 3rd Year Mathematics/Economics Major

- Introduction, Numerical Summary

Jing Gu, 2nd year Chemistry MS

- Incidence Comparison

Keven Nguyen, 4th Year Mathematics/Economics Major

- Numerical Summary

Adithya Bharadwaj Balaji, 1st year Graduate student in Electrical and Computer Engineering

- Numerical Summary, Incidence Comparison

1. Introduction

Much research has been conducted to investigate the relationship between smoking during pregnancy and birth weight. One such study was done by the Ohio Department of Health in 1989, which analyzed differing birth weights amongst White and African American babies born to both mothers who smoked during pregnancy and mothers who did not. The study found that a low birth weight was over twice as likely amongst babies born to mothers who smoked than those that did not. It was also concluded that the more cigarettes smoked by the mother, the higher the possibility of low birth weight (Morbidity and Mortality Weekly Report, 1990).

The adverse effects of maternal smoking impinges beyond infant birth weight. Research has shown that babies with lower birth weight have a higher risk of infant death. A 2001 study published by the American Academy of Pediatrics determined that the chance of infants living until discharge from a hospital steadily increased with an increase in birth weight, while mortality rates increased with a decrease in birth weight.

While the chemistry and physiology of how cigarettes interact with a developing fetus are not well known and difficult to test due to ethics, studies on Rhesus Monkeys suggest that nicotine is the main actor and can cross the placenta to react with highly sensitive nicotinic receptors in the developing lungs, altering fetal lung development and the amount of oxygen the fetus receives. A study by the University of Patras Medical School tracked the development of children born to smoking mothers up to two years and found that by the age of two, those children displayed considerable impediment in weight, length, and head circumference.
Studies published in the British Medical Journal in 1989 and 1991, respectively, found that low birth weight is associated with falling FEV_1 (Forced Expiratory Volume in One Second, an indicator of lung capacity and health) measurements taken among both children at age 7 and men aged 59-67.

Another study led by the University of Vermont examined how the frequency and timing of maternal smoking affected birth weights and found that mothers who smoked during the third trimester of pregnancy had the strongest likelihood of birthing low weight babies. The study concluded that regardless of cigarette consumption prior to pregnancy, a reduction during pregnancy would lead to an increased birth weight. From this prior research, we can see that studying birth weight is a worthy indicator of health of the baby throughout his or her life.

The objective of our report is to determine if there is a difference in birth weight between babies born to smoking and nonsmoking mothers, and to analyze the impact that difference has on the health of the baby. Our dataset, babies.txt, is taken from the Child Health and Development Studies which examined data from pregnant women living in San Francisco, under the Kaiser Health Plan, between 1960 and 1967. The purpose of the study was to analyze the effect of maternal smoking on birth weight. The data we analyzed included the following variables: baby weight (measured in ounces), gestation, parity, mother's age, height, weight, and smoking status. Gestation refers to how long the baby was in utero, and parity refers to the number of pregnancies reaching viable gestational age (including live births and stillbirths). Twin pregnancy carried to viable gestational age is counted as 1. We used numerical, graphical, and incidence comparisons to analyze the difference in birth weight between the two data sets, including calculating mean, median, mode, kurtosis and skew coeff of both data sets, generating histograms, box-plots, and quantile-quantile plots. Using these methods we found that both data sets are approximately normally distributed and that babies with mothers who smoke have lower birth weight.

2. Body

2.1 Numerical Summary

Table 1: The 7 variables in the dataset.

Variable Name	Description	Variable Type	Numerical Type
bwt	Body weight of the babies at birth in ounces	Numerical	Continuous
gestation	Gestation period of the baby	Numerical	Continuous
parity	First born baby	Categorical	N/A
age	Age of the mother	Numerical	Continuous
height	Height of the mother	Numerical	Continuous
weight	Pre-pregnancy weight of the mother	Numerical	Continuous
smoke	Smoking status of the mother	Categorical	N/A

Table 2: The brief summary of the various sample statistics.

Data Summary	Babies Born to Smoking Mothers	Babies born to Non-Smoking Mothers
n	459	715
Maximum	163	176
Minimum	58	55
Mean	113.819	123.085
Median	115	123
Mode	115	129
Variance	334.707	303.585
Standard Deviation	18.295	17.423
Skewness	-0.021	-0.163
Kurtosis	2.947	4.016
Third Quartle	126	134

The dataset under consideration here is obtained from the file “babies.txt”. The dataset contains 1236 observations. Entries that had incomplete data containing ‘999’ or ‘99’ were removed from all analysis. After removal of the entries with missing information in the dataset, the final data set used contained 1174 entries. Out of the seven variables in the data, we focus on the categorical variable “smoke” and its impact on the birth weight of the babies (Table 1). This helps us analyze whether a mother’s smoking habit has an impact on the birthweight of the babies. On stratifying the data based on the variable “smoke”, the number of non-smokers comes out to be 715 and the number of smokers is 459 (Table 2).

To numerically summarize the dataset we calculated several sample statistics (Table 2). The mean among the two groups (smoking and non-smoking mothers) shows the sample average of baby birthweights. The median, a measure that better fights outliers, shows a higher centralized birth weight of 123 ounces for babies born to non-smoking mothers in contrast to 115 ounces for those born to smoking mothers. The mode, or most frequent number, is higher for babies born to non-smoking mothers at 129 ounces, compared to 115 ounces for babies born to smoking mothers. The variance, which measures how far spread the dataset is from the mean, indicates that the spread in birth weight of babies born to smoking mothers is greater than that of babies born to non-smoking mothers. The standard deviation, the square root of the variance, reiterates that there is a greater spread in distribution of data from babies born to smoking mothers than babies born to non-smoking mothers. The skewness measures the asymmetry of the distribution; both birth weights of babies born to smoking and non-smoking mothers skew slightly left, but are approximately symmetric. The kurtosis measures the tails of the distribution. The dataset shows a higher kurtosis in the birthweight of babies born to non-smokers, which indicates more outliers than in the birthweight of babies born to smoking mothers. The third quartile measures the middle number between the median and the maximum of the dataset, and is higher in babies born to non-smoking mothers at 134 ounces compared to the 126 ounces of babies born to smoking mothers.

2.2 Graphical Methods

We used graphical methods to explore the relationship between the birth weight in ounces of babies born to mothers who smoked and who did not smoke during pregnancy. Using the filtered data with all unknown data samples removed (see above explanation), we created the graphs using multiple methods in R. We informally assessed the distribution of the two data sets using a histogram, box-plot, and quantile-quantile plot. The data presented in the below graphs includes baby birth weight in ounces from 715 non-smoking mothers (dataset #1) and from 459 smoking mothers (dataset #2).

Figure 1: Birth weight of babies with mothers who do smoke.

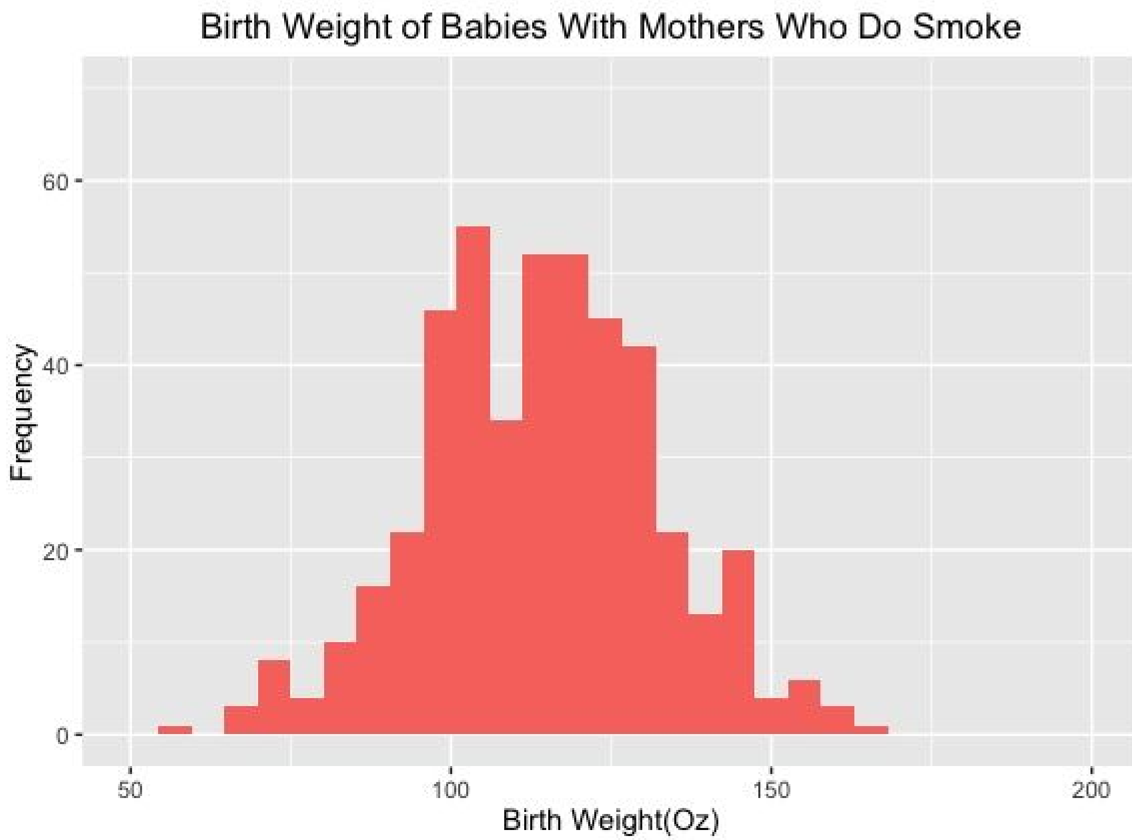


Figure 2: Birth weight of babies with mothers who do not smoke.

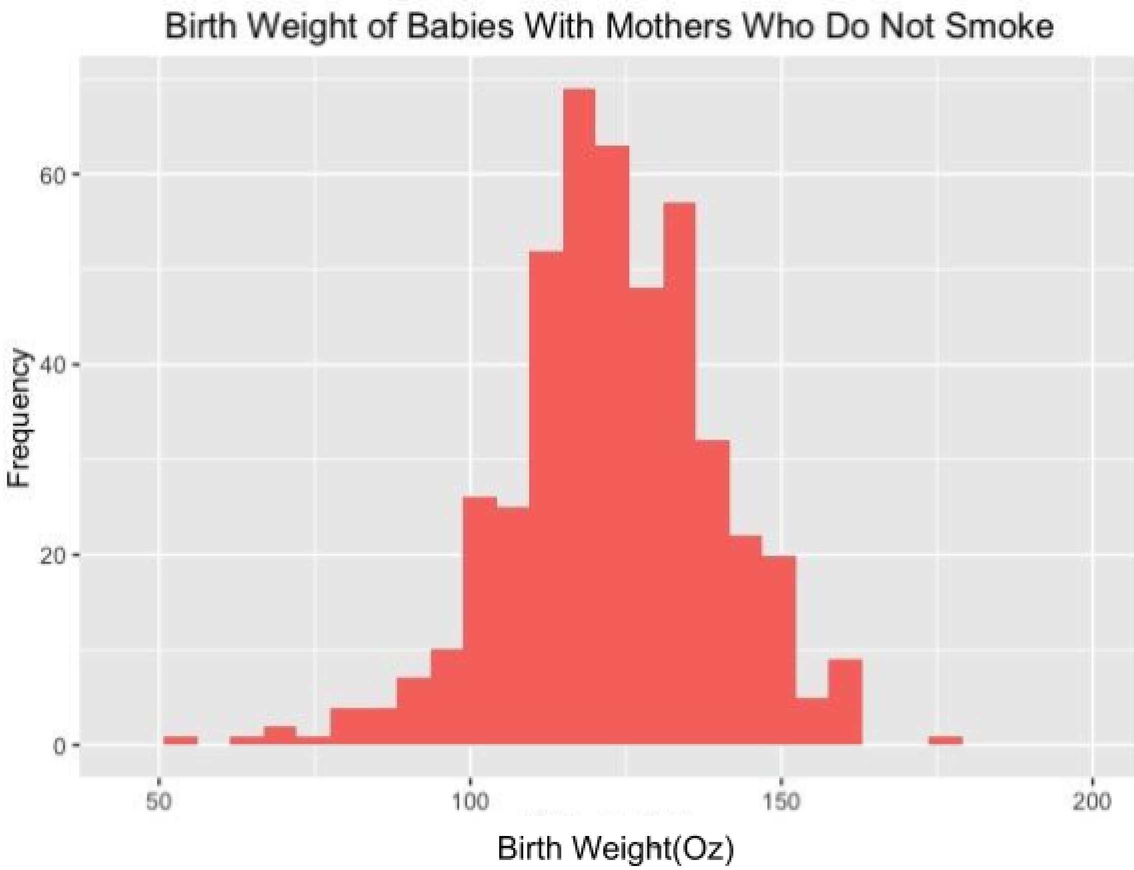
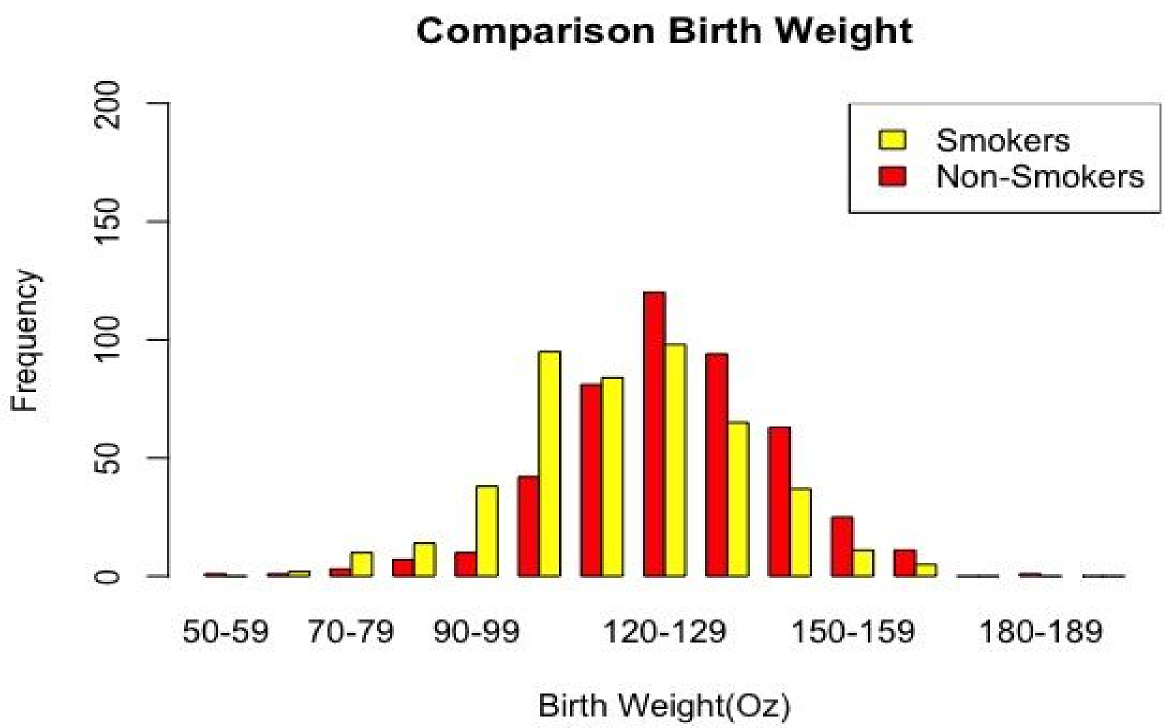


Figure 3: Comparison of birthweight for babies born to non-smoking and smoking mothers.



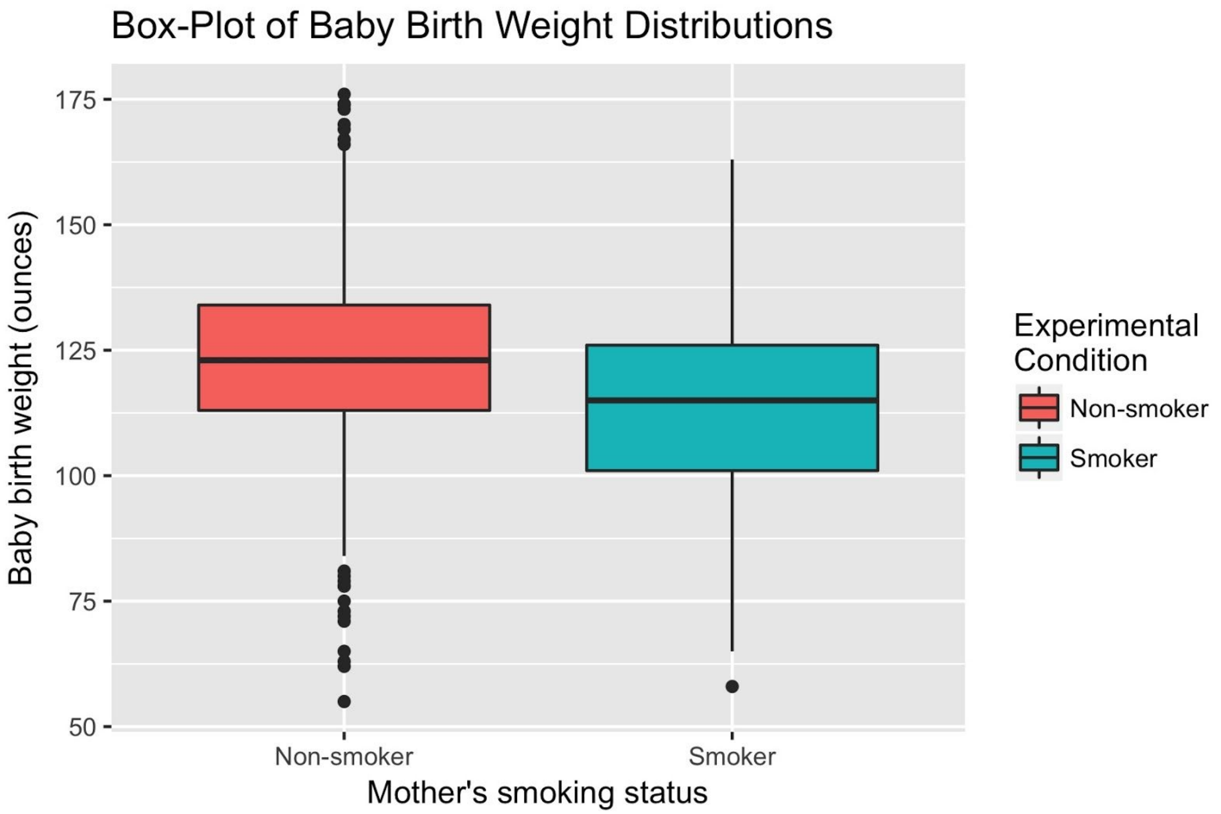
Histograms

The single data set histograms were plotted using the R package ggplot2, while the Comparison Birth Weight graph was plotted using the multhist() function of the plotrix package. The data displayed in Figure 1 is the entire set of birth weights for the smoking data (dataset #1) and Figure 2 contains the 459 random samples from the non-smoking data set, with seed set to 5 for reproducibility. On the single data set graphs, bin size was set to 5. On the comparison chart, bin size was set to 10.

We can discover much about the data by analyzing the shape of its histogram (Figures 1 and 2). Both distributions appear approximately symmetric, however data from mothers who do not smoke (dataset #2) has a bit of a tail and may be slightly left-skewed. Non-smoking birth weight data peaks around 120 ounces. Additionally, the non-smoker distribution indicates several outliers, particularly at 50 ounces and 175 ounces. Both data sets cover roughly the same range of birth weights, between 50 and 180 ounces. The data from mothers who smoke appears to have two peaks, one around 105 ounces and one between 115-120 ounces. The combined comparison histogram, Figure 3, indicates that both data sets have modes in the range 120-129 ounces. In the smoker data set (dataset #2), there are many more babies under 88 ounces, the accepted standard for low birth weight, than the nonsmoker data set (dataset #1). There are more data points to the left of the peak in the smoker group than the nonsmoker group, and more data points from the nonsmoker group to the right of the peak than the smoker group.

Both Figure 1 and 2 are approximately symmetric, suggesting that both distributions are approximately normal. Additionally, Figure 3 indicates that smoker data is more normally distributed than data from non-smokers. The imbalance of data seen on the comparison graph suggests that the set of data belonging to non-smoking mothers contains more birth weights above 120 ounces, while the data from the smokers contains more birth weights below 120 ounces. This supports our above calculated means of smoker’s birth weights and non-smokers birth weights of about 113 and about 123, respectively.

Figure 4: Box-plot comparison of birth weight for babies born to non-smoking and smoking mothers.



Box-Plot

The data presented in this graph includes baby birth weight in ounces from 715 non-smoking mothers (dataset #1) and from 459 smoking mothers (dataset #2).The box plot graph (Figure 4) was created in R using the ggplot2 package. The box-plot displays several important statistical values. Each section of the box-plot is composed of 25% of the data, therefore the thick middle line is the median value of the dataset and the length of the box is the Interquartile range (IQR). The shape of each of the box plot graphs (Figure 4) shows us that distribution is generally symmetric for both dataset, because the whiskers are about the same length for each box-plot. However, the range of distribution of the birth weight of non-smoking mothers (dataset #1) is wider compared to smoking mothers (dataset #2). This suggests there is likely a large amount of variability in the birth weight in dataset #1. We can also see from the box plot that the birth weight of non-smoking mothers data (dataset #1) contains a number of outlier values for low and high birth weight. Yet, the median birth weight of babies from mother's who do not smoke (dataset #1) is higher compared to the mothers who do smoke (dataset #2).

Figure 5: Quantile-quantile plot comparison of birth weight distribution compared to a Normal distribution.

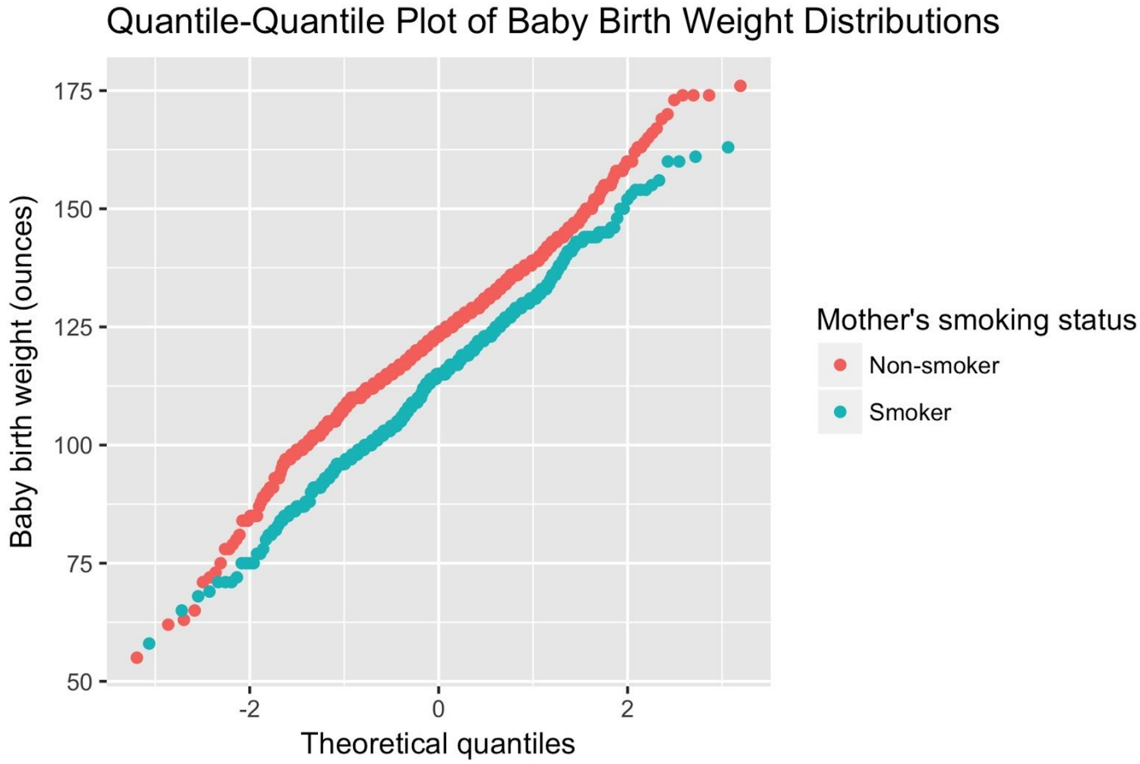
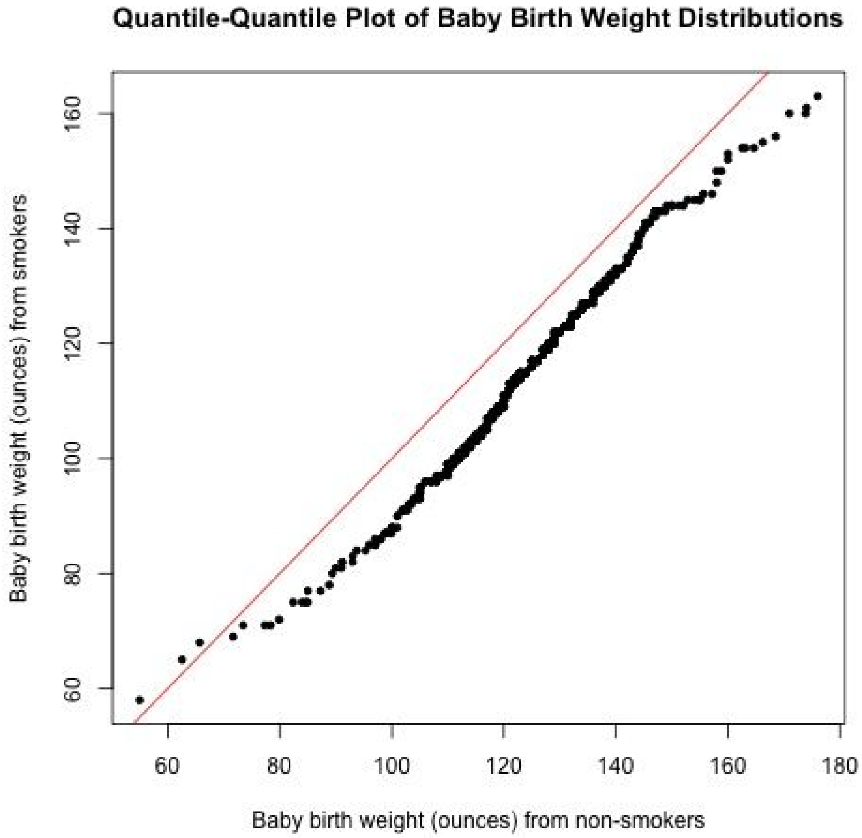


Figure 6: Quantile-quantile plot comparison of distribution for birth weight from babies born to non-smoking and smoking mothers.



Quantile-Quantile Plot

The quantile-quantile plot is useful for informally interpreting a number of important statistical properties about the data. The quantile-quantile plot (Q-Q plot) (Figure 5) comparing the two datasets to a theoretical Normal distribution was created in R using the ggplot2 package, the Q-Q plot comparing the distributions of the two datasets (Figure 6) was made using qqplot in R. First, looking at (Figure 5), we can infer that both datasets approximately follow a normal distribution. We can make this inference because the points roughly follow a diagonal straight line, indicating the observed values (y-axis) approximately match the theoretical value (x-axis). Next, the Q-Q plot in (Figure 6) is plotting the sample quantiles of baby birth weight from non-smoking mothers (x-axis) versus smoking mothers (y-axis). From this plot we can view the difference in distribution between the two datasets. Due to the fact that the points follow a general diagonal straight line, we can infer the shape of distribution for the two datasets is roughly the same. However, the points fall below the red line of slope = 1 and y-intercept = 0, which indicates the mean of the birth weight from non-smokers (dataset #1) is greater than the mean of the birth weight from smokers (dataset #2).

It is not a precise method of analysis to compare data by histogram, box-plot, and Q-Q plot graphs. While graphical analysis allows us to make rough estimates about things such as the normality of the distribution and the summary statistics, we need further numerical statistical analysis to prove our visual conclusions true. Mathematical analysis is provided in the alternate sections.

2.3 Incidence Comparison

The minimum birth weight for a healthy newborn baby is 5.5 pounds or 88 ounces. So babies born with birth weight lower than 5.5 pounds are considered to be low-birth-weight babies. Considering the 715 non-smoker mothers (dataset #1) and the 459 smoker mothers (dataset #2). Out of the babies born to smokers, 7.843% of the babies are underweight and out of the babies born to non-smokers 2.937% of the babies are underweight. We further used simulation to randomly sample 300 data from birth weights of babies born from non-smoking mothers and computed the proportions of low birth weights (under 88 ounces) for 1000 times. The histogram of proportions in figure 7 of low birth weights shows that the mean of proportions is still around 2.9%. When a few more or less birth weights classified as low birth weights should not significantly change the fraction cutoff which is 2.937% as low birth weight for babies born to mothers who did not smoke during pregnancy. Same for the group of smokers.

Figure 7: Distribution of proportions of low birth weights from 300 subset of birth weights of babies born from mothers who did not smoke during pregnancy.

Simple Monte Carlo simulation of kurtosis coefficients were used to check for normality of sampled data. Kurtosis coefficient shows how pronounced the peak of the distribution is, which is a more formal and quantitative way to decide whether a departure from normal distribution if big or small comparing to the graphic method. It was found that the Kurtosis coefficients of birth weight babies born from women who smoked during their pregnancy and those from women who did not smoke was 2.95 and 4.02 separately. Since sample sizes for the two groups are different (459 for smokers and 715 for nonsmokers), random variables from normal distribution were sampled twice with different total number of observations. Figures 8 and 9 are histograms of kurtosis values for 459 and 715 pseudo-random observations sampled from a normal distributions over 1000 repetitive computations.

Figure 8: Distribution of kurtosis coefficients from 459 normally distributed samples.

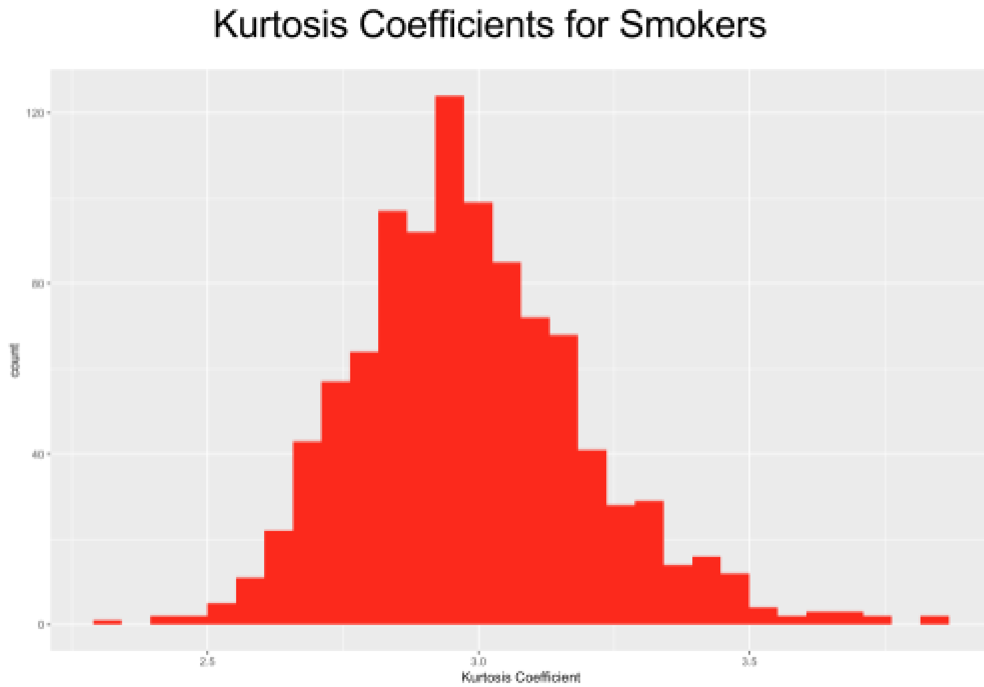
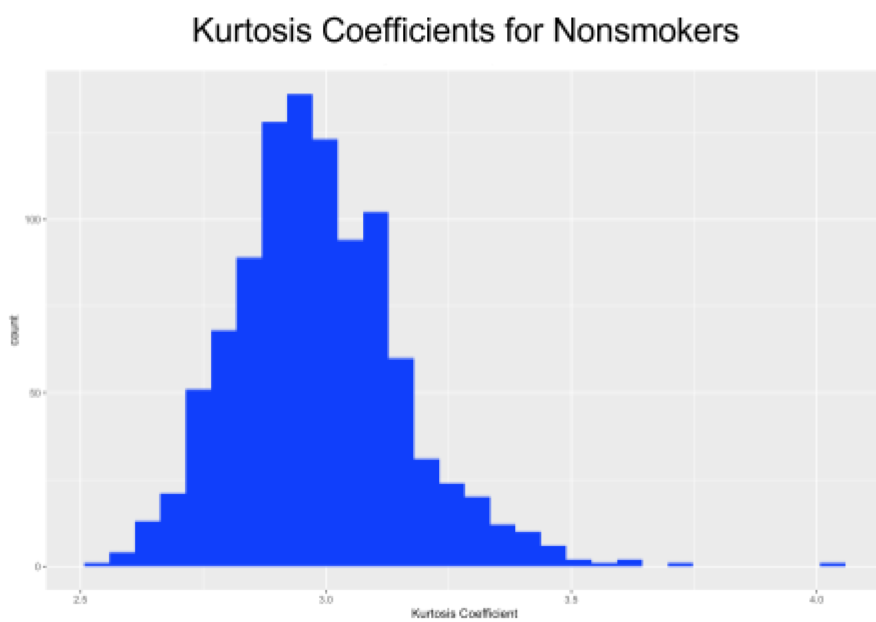


Figure 9: Distribution of kurtosis coefficients from 714 normally distributed samples.



From Figure 9, 2.95 is close to the mean of kurtosis coefficients simulated for smokers case but 4.02 is more than three times standard deviations away from the mean of kurtosis coefficients simulated for baby weights for babies born from non-smokers. This suggests that birth weights of babies born to non-smoking mothers have a much bigger departure from normality than those of babies born to smoking mothers. Therefore, the two groups should have different distributions. The weights between babies born to mothers who smoked during pregnancy and those who did not are indeed significantly different.

3. Conclusion(s)/Discussion

Our numerical, graphical, and incidence analyses show a general concordance. The histogram, Q-Q plot, and kurtosis calculation support that both datasets have an approximately Normal distribution. Our graphical analysis shows a larger spread of birth weight values from non-smoking mothers (dataset #1, range = 121 compared to dataset #2, range = 105), however the numerical analysis concludes the variance of dataset #1 is lower (dataset #1, var = 121 304 vs. dataset #2, range = 335). The variance is potentially lower due to the sample size being much larger (dataset #1, n = 715 vs. dataset #2, n = 459).

We set out to statistically determine if there is a difference in birth weight between babies born to smoking and nonsmoking mothers. Due to the observation that the mean and median birth weight of babies from non-smoking mothers (dataset #1, mean = 123, median = 123) is higher compared to smoking mothers (dataset #2, mean = 114, median = 115), we support the idea that mothers who smoke have babies with lower weight. Based on our background research, our data suggests a relationship between mothers who smoke and developmental problems throughout the life of the low birth weight baby due to the evidence that low birth weight often predicts higher mortality rates, stunted growth, and respiratory problems.

Yet, future work is required, such as hypothesis testing, to determine if the difference in birth weight is statistically significant between the two groups of mothers. We cannot definitively say that smoking is a causal factor for low baby birth weight, because we did not perform a rigorous randomized study. Also the low birth weight of the babies could be due to confounders like history of chronic medical illness, history of abortion, history of premature delivery, hard physical work during pregnancy, illness during current pregnancy, haemoglobin level, the practices of taking iron tablets, consuming nutritious food (beans, greens or meat) daily and drinking alcohol during pregnancy, just to name a few.

4. Appendix/Appendices

4.1 Two sample T-test by Emma

A two sample t-test was used to attempt to determine the statistical significance of the difference in mean bwt in the smokers and nonsmokers data sets. The null hypothesis is that the difference in birth weight is due to chance.

$$Standard\ Error\ of\ the\ Mean\ = sqrt\left(\frac{variance(nonsmokers)}{sample\ size(nonsmokers)} + \frac{variance(smokers)}{sample\ size(smokers)}\right)$$

$$sqrt\left(\frac{303.585}{715} + \frac{334.707}{459}\right) = 1.074$$

$$t = \frac{mean(nonsmokers) - mean(smokers)}{SEM}$$

$$t = \frac{123.085 - 113.819}{1.074} = 8.63$$

Our data set is very large and has many degrees of freedom. We estimate the appropriate critical value by comparing our calculated t value to d.f. = 100 and d.f. = ∞. For d.f. = 100, p = 0.01 when t ≥ 2.63. For an infinitely sized data set, p= 0.01 when t ≥ 2.58. Our value is significantly greater than 2.58 and 2.63, so we can reject the null hypothesis and say that the probability of the apparent difference in birth weight arising by chance is $\leq 1\%$.

4.2 Works Cited

Barker, D J P, K M Godfrey, C Fall, C Osmond, P D Winter, S 0 Shaheen. Relation of birth weight and childhood respiratory function to adult lung function and death from chronic obstructive airways disease." BMJ, vol. 303, no. 7, 1991, pp. 671-675. http://www.bmj.com/content/bmj/303/6804/671.full.pdf

Bernstein, Ira M., MD, Joan A. Mongeon, MS, Gary J. Badger, MS, Laura Soloman, PhD, Sarah H. Heil, PhD, and Stephen T. Higgins, PhD. "Maternal Smoking and Its Association With Birth Weight." ResearchGate. N.p., Dec. 2005. Web. 20 Jan. 2017. https://www.ncbi.nlm.nih.gov/pubmed/16260516

"Effects of Maternal Cigarette Smoking on Birth Weight and Preterm Birth -- Ohio, 1989." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 28 Sept. 1990. Web. 20 Jan. 2017. https://www.cdc.gov/mmwr/preview/mmwrhtml/00001782.htm

HARMANJATINDER S. SEKHON , JENNIFER A. KELLER , NEAL L. BENOWITZ , and ELIOT R. SPINDEL. "Prenatal Nicotine Exposure Alters Pulmonary Function in Newbord Rhesus Monkeys." *American Journal of Respiratory and Critical Care Medicine*, vol. 164, no. 6, 2001. Web. Accessed 21 January 2017. http://www.atsjournals.org/doi/full/10.1164/ajrccm.164.6.2011097

Harmanjatinder, Sekhon, Yibing Jia, Renee Raab, Alexander Kuryatov, James F. Pankow, Jeffrey A. Whitsett, Jon Lindstrom, and Eliot R. Spindel. "Prenatal nicotine increases pulmonary α7 nicotinic receptor expression and alters fetal development in monkeys." The Journal of Clinical Investigation, vol. 103, no. 3, 199, pp. 637-647. Web. 21 Jan 2017. https://www.jci.org/articles/view/5232

Lemons, James A. Charles R. Bauer, William Oh, Sheldon B. Korones, Lu-Ann Papile,Barbara J. Stoll, Joel Verter, Marinella Temprosa, Linda L. Wright, Richard A. Ehrenkranz,Avroy A. Fanaroff, Ann Stark, Waldemar Carlo, Jon E. Tyson, Edward F. Donovan, Seetha Shankaran, David K. Stevenson. "ery Low Birth Weight Outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 Through December 1996." *American Academy of Pediatrics Journal*, vol. 107, no. 1, 2001. Web. 21 Jan 2017. http://pediatrics.aappublications.org/content/107/1/e1.short

Karatza, Ageliki A., MD, Anastasia Varvarigou, MD, and Nicholas G. Beratis, MD. "Growth up to 2 Years in Relationship to Maternal Smoking During Pregnancy." Sage Journals. N.p., 1 July 2003. Web. 21 Jan. 2017. http://journals.sagepub.com/doi/abs/10.1177/000992280304200608

Sudesh Raj Sharma, Smith Giri, Utsav Timalsina, Sanjiv Sudarshan Bhandari, Bikash Basyal, and Kusum Wagle "Low Birth Weight at Term and Its Determinants in a Tertiary Hospital of Nepal: A Case-Control Study", *US National Library of Medicine and National Institutes of Science,* vol. 10, no. 4, 2025. Web. Accessed 22 January 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4390309/