1 An Analysis of Smoking During Pregnancy and Birthweight

Nathan Liittschwager[1,+], Arielle Thibeault[1], Katherine Yamamoto[1], Alec Guthrie[1], Alex Leibscher[1], and Mike Ona[1]

[1]University of California, Mathematics, San Diego, USA
[+]all authors contributed equally to this work

ABSTRACT

Many previous studies have suggested that maternal smoking has harmful effects on fetal development and subsequently on neonatal health. This study aims to identify a causal relationship between maternal smoking and infant mortality. Moreover, the analysis examines the hypothesis that maternal smoking influences a lower birth rate for infants and utilizes literature to explore the correlation between lower birth weight and infant mortality. From our analysis, the data suggests that there is a statistically significant difference between the birth weight of infants born to smoking mothers and infants born to non-smoking mothers. This implies that maternal smoking during pregnancy can lead to low birth weight and subsequently a less healthy infant. This result is consistent with similar studies suggesting maternal smoking during pregnancy has a direct relationship to the weight and therefore health of an infant.

Introduction

The Comprehensive Smoking Education Act of 1984 required that all cigarette packages and advertisements be labeled with, among other warnings, "SURGEON GENERAL'S WARNING: Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight." For decades now, the literature has documented the harmful effects of maternal smoking on the health of infants (Abrevaya, 2006; Woolbright, 1984). Utilizing results of reputable studies stating that low birth weight (LBW) infants are subject to more health problems, we investigate the relationship between the smoking status of the mother and the subsequent birth weight of the infant. Is there significant evidence that maternal smoking causes low birth weight and therefore increases the chance of infant mortality?

With a publicly available dataset compiled by the Child Health and Development Studies (CHDS), we aim to answer this question via statistical analysis on both maternal smoking and non-smoking subsets. We calculate the mean and standard deviations for birth weights from each group and evaluate them using a simulated test distribution for comparison. To ensure appropriate analytical methods, we examine sample distributions for normality and perform error analysis on the sampled and simulated data to provide consistency and robustness.

Background

Weight at birth is of particular interest to many researchers because collectively LBW infants are more susceptible to subnormal growth, illnesses, and neurodevelopmental issues (cerebral palsy, blindness, etc.), despite most LBW infants having normal outcomes (Hack 1995). The World Health Organization defines LBW as less than 2,500 grams (88.185 ounces). LBW is mainly caused by either a short gestation period or poor intrauterine growth rate, meaning the fetus is not growing at a normal rate inside the womb. According to the U.S. National Library of Medicine, a normal gestation period lasts between 37 and 42 weeks, or about 259 to 294 days. At around week 35 to 37 the fetus weighs about 2,500 grams, and although it may continue to gain some weight, it is unlikely to get much longer. Usually, the child will be born near the 40th week of pregnancy.

Although birth weight, gestation period, and maternal smoking status are controlled, previous studies have found more than 40 potential determinants of LBW. Racial/ethnic origin, maternal and paternal height, caloric intake, episodic illness, malaria, alcohol consumption, and tobacco chewing among others, have been shown to have direct causal influence on intrauterine growth (Kramer 1987). These potential determinants may possibly act as confounding variables in our analysis and camouflage

the possible effect of smoking on infant birth weight and mortality. To determine if mothers who smoke during pregnancy give birth to less healthy infants than mothers who do not smoke, we test the data for an implication that maternal smoking causes LBW in infants under the assumption that there is a direct connection between LBW infants and infant mortality. If there is indeed a statistical difference between the weight of infants of smoking and non-smoking mothers, to what extent does this difference ultimately contribute to the infant mortality rate?

Description of the Data

The data for this report are a small subset from a study done by Child Health and Development Studies in 1995. That study included pregnancy data from 15,000 families collected between 1960 and 1967 by the Kaiser Foundation Health Plan in Oakland, California, all of whom voluntarily participated in interviews about their health and any lifestyle choices that may affect the mother's pregnancy and the health of the child. Women participating in the study had been under prenatal care in the San Francisco area and delivered at Kaiser hospitals in Northern California. The subset of data used in this report is a sample of baby measurements from 1,236 of those births. All babies included in this report are single birth males who survived at least 28 days after the time of birth. The measurements of interest are birth weight, whether the mother smoked during pregnancy, and the gestational period. Samples with unknown smoking status and gestation period were removed, reducing the total number of samples to 1,213.

Table 1. The key measurement variables are birth weight, mother's smoking status during pregnancy, and baby's gestation period.

Variable	Description	Type
bwt	The infant's body weight in ounces(oz) at birth	Numerical Continuous
smoke	An indicator of whether or not the mother smoked during pregnancy	Categorical indicator
gestation	The baby's gestational period, in days	Numerical Discrete

Table 2. A numerical summary of baby birth weight with respect to mother's smoking status during pregnancy. Babies born to mothers who smoked had a lower mean birth weight compared to mothers who did not smoke during pregnancy, but the statistical significance of this lower mean is yet to be explored.

Bodyweight(oz)	Pooled	Smoker	Non-smoker
Mean	119.6	114.2	123.3
Standard Deviation	18.2	18.1	17.2
Variance	331.7	329.8	300.7
Minimum	55.0	58.0	55.0
1st Quantile	109.0	101.8	113.0
Median	120.0	115.0	123.0
3rd Quantile	131.0	126.0	134.0
Maximum	176.0	163.0	176.0

The data used in this report offer unique insights into a small but representative sample of the Californian population. Most notable is the approximately equal representation of different socioeconomic and education backgrounds (Paul 1992). For example, the data shows family incomes between $2,500 and $22,500 are non-modal. The sample population also includes roughly half of the mothers with and without undergraduate college degrees. On the other hand, some subsets of the population were not evenly represented. Children in this report's data were born to women of a young childbearing age with a median of 26 years old. Furthermore, about 70% of all children were white and 20% black, while races of mixed, Mexican, and Asian children were barely represented. It should also be noted that since the sample includes only mothers at Kaiser hospitals in Northern California that the data likely suffers from convenience bias. These limitations of the sample population should be considered when interpreting this report and making any inferences.

Statistical Investigations

In this section we tabulate our results and descriptive statistics. For inquiry on method and underlying theory, refer to "Theory and Methods".

To summarize the data, there are 1226 observations utilized in this analysis: 484 records of mothers who smoked during pregnancy and 742 mothers who did not smoke during pregnancy. The mean birth weights of both groups are 114.1 ounces for

the smoking subset and 123.0 for the non-smoking subsets with standard deviations 18.1 and 17.4 respectively.

In order to assess the distributions of the data concerning bodyweight and quantify our analysis a bit more objectively, the Kurtosis and Skewness of the pooled, smoking, and non-smoking groups was taken (Table 3). We found the pooled and smoking distributions to be relatively close to the ideal Normal Distribution (Kurtosis = 3, Skewness = 0), while the non-smoking group had both more skew and kurtosis, indicating a higher degree of asymmetry.

A quantile-quantile plot for the data (Figure 2) was also given, so that the distributions of the data may be compared against that of a normal distribution. To interpret the graph, the data points should approximate closely to the line – large deviations are an indicator of non-normality.

Analysis of the Distribution of Birth Weights from Pooled, Smoking, and Non-smoking Mothers

Purpose: We investigate the approximate distributions of birth weights from the Pooled, Smoking and Non-smoking groups. Since the distribution of the data are paramount to understanding the differences in the data, and provide useful information on what sort of analyses may be conducted, we investigate the question of whether the various data may be well approximated by a normal model. While the investigation is framed as a statistical test, it should be noted that the procedures here are largely descriptive and are therefore *approximate* in nature. To inspect the quantile-quantile plots of smoking and non-smoking groups, as well as the kernal density estimates, see **Figure 2.**.

Procedure:

For a simple collection of kurtosis and skewness values of each of the groups, see **Table 4**

The skewness and kurtosis values for the total observed birth weights are calculated as -0.1406 and 3.443, respectively. To further the analysis, the Monte Carlo simulation is generated for the same sample size and then reiterated a thousand times to yield a simulated mean skewness value of 0.0016 with quantiles (-0.1306, 0.1333) and a mean kurtosis value of 2.999 with quantiles (2.757, 3.298).

Separating the data into maternal-smoking and maternal-non-smoking subsets under the criterion that the smoking group contains only women who had smoked during pregnancy, the skewness and kurtosis values for the maternal smoking group were calculated as -0.0336 and 2.988, respectively. From inspection, it appears the data set is roughly normally distributed. The Monte Carlo simulation yields a mean skewness value 0.0010 with quantiles (-0.2307, 0.2146) and a mean kurtosis value of 2.994 with quantiles (2.633, 3.526). These values vary slightly from the prior test due to the difference in simulated sample size. This is consistent with the Central Limit Theorem.

Similarly for the maternal-non-smoking group, the calculated skewness and kurtosis values are 0.1870 and 4.037, respectively. From inspection, these values appear to indicate that the data is not normally distributed. The Monte Carlo simulation yields a mean skewness value -0.0004 with quantiles (-0.1766, 0.1674) and a mean kurtosis value of 2.992 with quantiles (2.688, 3.364).

Result: Since the sample skewness and kurtosis values lie outside 95% of the simulated data, the distribution of the **pooled groups** observed birth weights may not follow a normal distribution.

Likewise, comparing the observed and simulated values for the smoking group that the sampled data *is normally distributed* for the **smoking group**.

Since the sample skewness and kurtosis values lie outside 95% of the simulated data, we and conclude that the distribution of birth weights of infants born to **non-smoking group** may not be well approximated by a normal model.

Group(bwt)	Kurtosis	Skewness
Pooled	3.443	-0.1406
Smoking	2.965	-0.0163
Non-smoking	3.847	-0.1041

Table 3. A comparison of the kurtosis and skewness of the distributions concerning bodyweight, smoking and non-smoking mothers. A normal distribution has a kurtosis of 3 and a skew of 0. Note the skews are relatively slight, indicating symmetry, but the kurtosis of the non-smoking group is somewhat high. The negative sign in the skewness values indicates the data is skewed to the left.

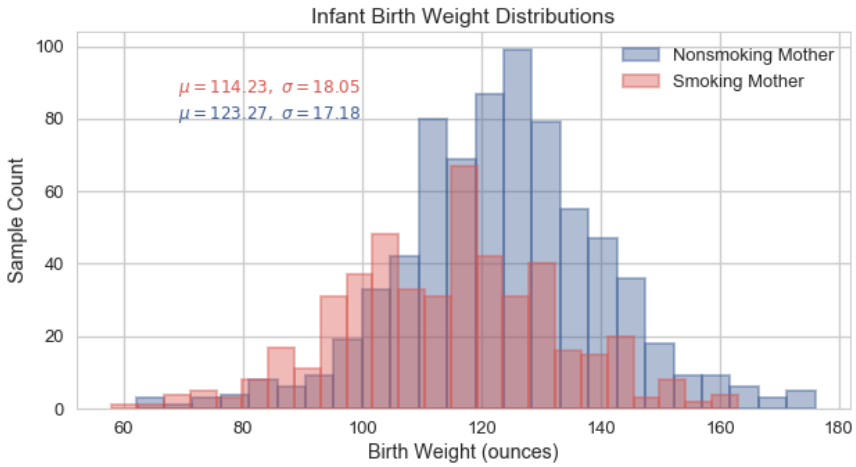
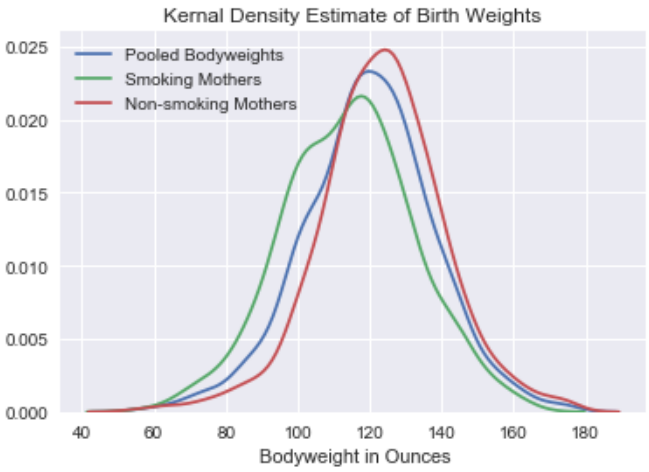


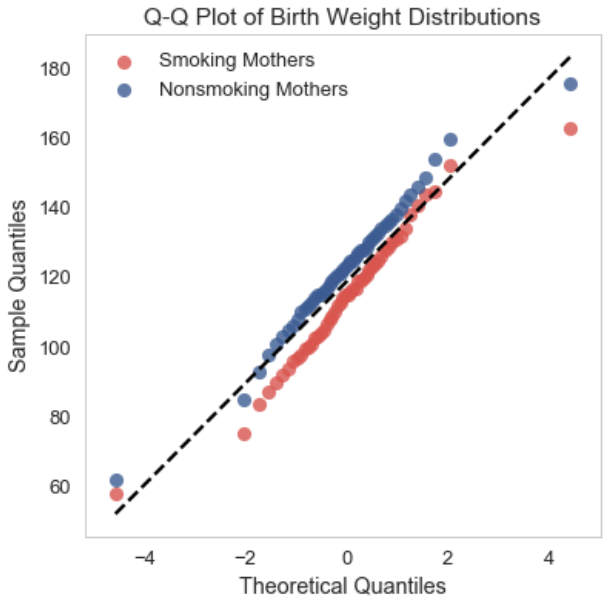
Figure 1. A comparison of the bodyweight of infants from mothers who smoked during pregnancy and mothers who did not. Note that the histogram of data from smoking mothers (red) is shifted relatively to the left, suggesting a difference in distribution. Moreover, note the peaks of the non-smoking data almost appear to be bimodal, and the relative thickness of the tails in comparison to the smoking group.

The *Kernal Density Estimate*(KDE) is the value on the y-axis on the left-hand graph given in **Figure 3**. It is a non-parametric method that is used to estimate probability distributions. Distributions are more or less visualized histograms such that their bin width is taking infinitesimally small. The (KDE) is a function that attempts to smooth the histograms so that the reader is more easily able to see their approximate distribution. The normal distribution is visualized as a perfectly symmetrical "bell shape."

The *Quantile-Quantile Plot* is another distribution estimate tool. The normal distribution and data are transformed into a straight line, so that they may be compared. A set of data which is approximately normal will hold close to the center line.



(a) Kernal Density Estimates of Groups



(b) Quantile-Quantile Plot against a normal distribution

Figure 2. The close approximation of the data points in the quantile-quantile plot is a strong indicator of normality, despite the slight deviations found in the Kurtosis and Skew. Likewise, the kernal density estimates provide a graphical representation of the data. While the smoking mothers has two peaks, it is actually better approximated by the normal distribution than the non-smoking mothers. The distributions of the data follow an approximately normal distribution. Moreover, the t-test is fairly robust to departures from normality (Casella and Berger, ("Statistical Inference").

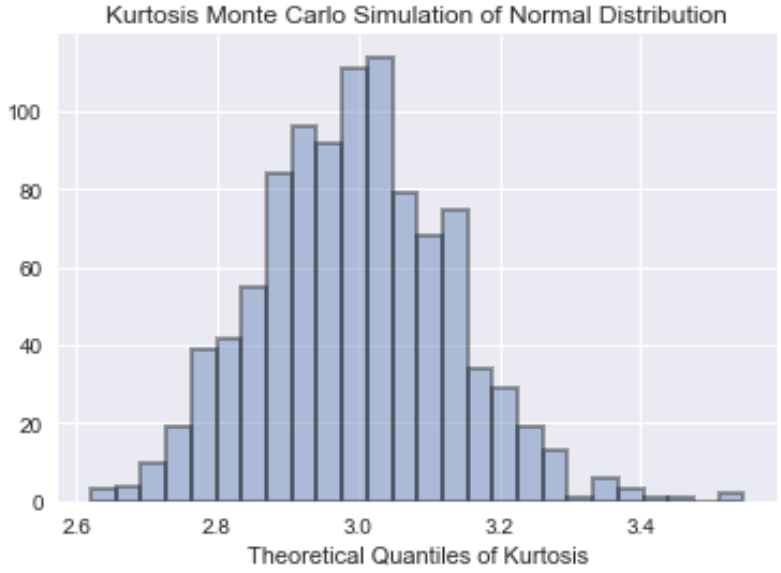


Figure 3. Monte Carlo Simulation of the Kurtosis. In order to conduct our Kurtosis testing, we simulated random samples from a normal distribution, with sample size equal to the length of our data. Then a kurtosis value of that data was taken. This process was repeated 1000 times, with the values collected into this histogram. Note that the majority of the data is centered over the kurtosis of 3, which is that of a perfectly normal distribution.

Testing Difference in Observed Means in Birth Weight

Null Hypothesis ($\alpha = 0.01$): There is no difference in birth weight means between the maternal smoking and maternal nonsmoking subsets.

Procedure:

Upon controlling for gestation so that both smoking and non-smoking groups consist of only full term infants (between 259 and 294) days, we took the arithmetic means of birth weight(oz) for both groups and found that that $\bar{x} = 115.054$ (95% CI $[113.41, 116.70]$) for the smoking group and that $\bar{y} = 123.97$ (95% CI $[122.68, 125.25]$) for the non-smoking group. Since the data sample sizes are large and lack heavy skew, the arithmetic means are approximately normal under the central limit theorem. We performed a bootstrapping two-sample Welch's t-test to test the Null Hypothesis. Under the central limit theorem ($n_x = 392, n_y = 590$), both $\bar{x}$ and $\bar{y}$ and their accompanying sample standard deviations are approximately normally distributed. Moreover, we assume that the samples are independent, as there is no reason to assume the choice of smoking in one group has influence on the other. Despite the slight departures in normality in the samples, Welch's 2-sample t-test is robust under non-normal conditions provided a large enough sample size and lack of skew in the data, which are both present. To account for the difference in sample sizes (which may cause the two sample t-test to fail), we performed a bootstrap algorithm by sub-sampling $n = 300$ from each sample and passed the samples through Welch's t-test statistic. We repeated this process 1,000,000 times and summed over the p-values > 0.01, divided by the total number of simulations, with error taken into consideration.

Result: An approximate p-value of $p = 0.00009$ was obtained. The significantly small p-value implies that there is a statistically significant difference in birth weight means between smoking and non-smoking groups. Thus, the null hypothesis is rejected and we cannot attribute the observed differences to chance alone.

Testing Difference in Observed Gestation Period

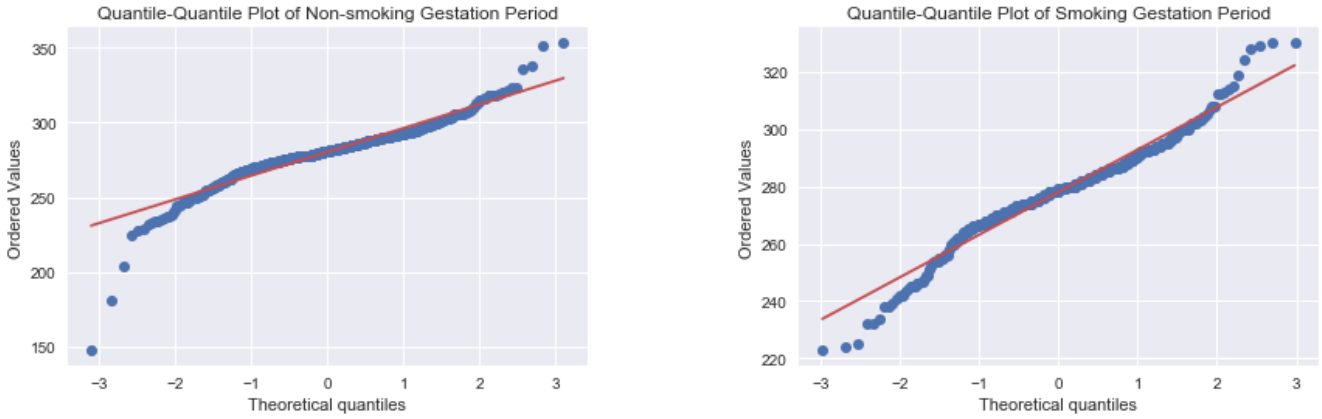
Null Hypothesis ($\alpha = 0.05$): There is no difference in gestation periods between maternal smoking and maternal nonsmoking subsets.

Procedure: The pooled sample mean in gestation period was $\bar{z} = 279.31$, (95% CI [278.39, 280.21]), and the sample means in gestation between the smoking and non smoking groups (in days) to be $\bar{x} = 277.98$ (95% [276.64, 279.33]) and $\bar{y} = 280.19$(95%$[278.98, 281.38]$, respectively, thus giving an observed difference of $\bar{y} - \bar{x} = 2.21$ days.

Because the distribution of the gestation periods does not appear to follow a normal distribution, in that they show heavy

skew and lack of symmetry, (See figure 3 and 4) the prior described t-test algorithm cannot be applied. The following bootstrapping algorithm was used in place. Taking $\bar{y} - \bar{x} = 2.21$ to be our test statistic, we shifted the smoking and non-smoking gestational data so that they both had a sample mean approximate to the pooled mean (to simulate the null hypothesis). We then bootstrapped the shifted samples and calculated the difference in gestational sample means for the $B = 1,000,000$ simulates. We summed the number of differences larger than our observed difference of 2.21 and divided by the total number of simulations performed, adding 1 to the numerator and denominator to account for error and include the initial observation

Result: We produced an approximate p-value of of $p = 0.0066 < 0.05$. Since the p-value is less than the significance level, there is a statistically significant difference in gestation period between groups. Thus, the null hypothesis is rejected and we cannot attribute the observed differences to chance alone.



(a) Non-smoking Gestational Period

(b) Smoking Gestational Period

Figure 4. A comparison of the Gestational against a normal distribution. The tails in both charts deviate heavily from the center and have a slight bend towards the center, indicating the normal distribution is not an adequate fit

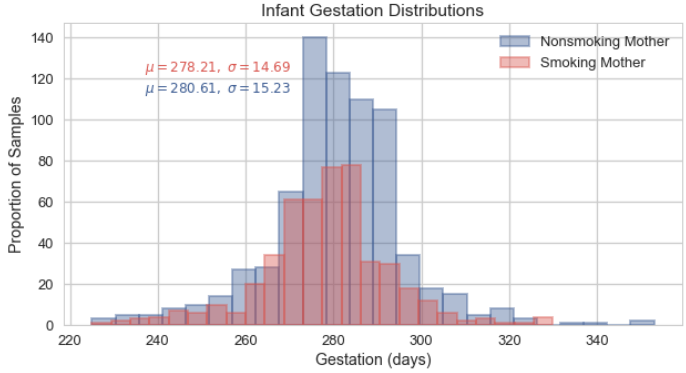


Figure 5. A comparison of the gestation data between smoking and non-smoking mothers. Despite some reported literature, the distributions appear to match closely. A statistical test was performed to see if the distributions really give rise to a comparable mean gestation date.

Gestation(days)	Kurtosis	Skewness
Pooled	6.596	2.053
Smoking	2.053	-0.224
Nonsmoking	8.760	-1.079

Table 4. A collection of Kurtosis and Skewness values from the gestation data. The extremely large kurtosis of the pooled and non-smoking groups, as well as the strong skew indicates the data likely is not normally distributed.

Testing the difference in birth weights from light and heavy smoking mothers

Null Hypothesis ($\alpha = 0.05$): There is no difference in the mean birth weights of infants from women classified as heavy smokers, and women classified as light smokers.

Procedure: Mothers were partitioned into two groups, those who were considered "light" smokers reported smoking less than 10 cigarettes a day during pregnancy. The "heavy" smokers were those who reported smoking more than 10 cigarettes a day during their pregnancy. The sample mean of birth weights found in the light smoke group was $\bar{x} = 117.27$, (95% CI [114.72, 119.81]) and the sample mean of birth weights in the heavy smoker group was found to be $\bar{y} = 113.55$, (95% CI [111.40, 115.71]), with the observed difference $\bar{x} - \bar{y} = 3.72$ While there is a small amount of overlap between the confidence intervals, in order to more accurately test the difference in means, a bootstrapping procedure was utilized. Under the null hypothesis, B=1,000,000 replicates of the difference in means was simulated and then compared to the observed difference of 3.72. Refer to TABLE for a graphical representation of the data.

Results: An approximate p-value of $p = 0.014 < 0.05$ was obtained. Since the p-value is less than the significance level, there is a statistically significant difference in gestation period between groups. Thus, the null hypothesis is rejected and we cannot attribute the observed differences to chance alone.

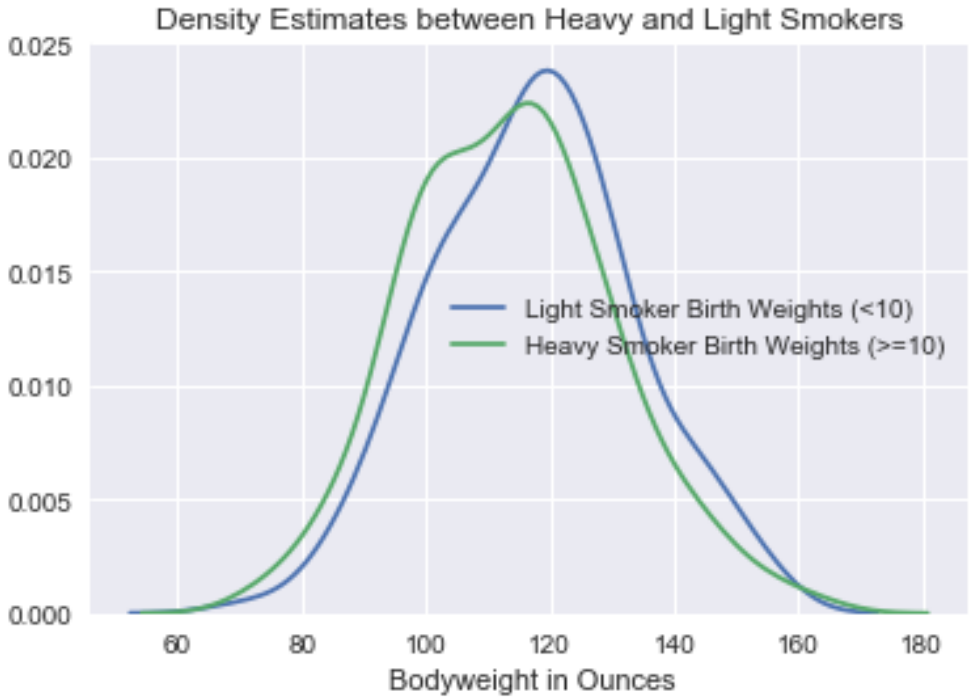


Figure 6. A kernal density estimate of birth weights given by mothers who smoke less than 10 cigarettes a day (light) and more than 10 cigarettes a day (heavy). The slight discrepancy in peaks suggests the data come from a different distribution. Moreover, though slight, the heavy smoking group is shifted to the left (with a slight bimodal peak). In conjunction with the statistical test performed, there is evidence that cigarettes may be dose-dependent.

Discussion

Consistent with previous studies, the analysis indicates a statistically significant difference in the birth weights of infants of non-smoking and smoking mothers while controlling for gestation so that both groups consisted of full-term newborns. Mothers who smoked during pregnancy gave birth to infants who weighed, on average, 250 grams (8.82 ounces) less than those of the mothers who did not smoke. Furthermore, when observing whether there is a correlation between maternal smoking and gestation period, the analysis contained evidence that mothers who smoked had slightly shorter gestation periods than those who didn't. While the difference was minimal, a shorter gestation period has a non-negligible effect on birth weight.

The population from which the data was collected consists of families who were members of the Kaiser Foundation Health Plan in Oakland, California. The mothers, who voluntarily participated in the study, were women who received prenatal care within the Bay Area and delivered in a Northern California Kaiser hospital. These criteria could have confounding effects when studying a correlation between maternal smoking and low birth weight. Because the women were members of the KP Health Plan, it can be assumed that they were able to afford health insurance which indicates a certain socioeconomic standing. Not only that, but the infants were delivered in a particular geographical region, thus the data may not be representative of the entire California population. All of the participants also received previous prenatal care, but having different health insurance plans could result in women receiving different prenatal care which could affect the birth weight of the infant. The most obvious source of sample bias stems from the fact that this data was collected voluntarily.

While we do control for many potential confounders, another source of selection bias exists in the lack of information on smoking status and gestation period. This information was unknown for only 1.9% of the mothers included in the sample. Unfortunately, with this limited dataset we were unable to control certain variables that are proven to be direct determinants of intrauterine growth, such as racial/ethnic origin, paternal height and weight, caloric intake, inherent genetic potential for LBW, or alcohol consumption (Kramer 1987). It is important to recognize this, because the difference that we uncovered, although strongly supported by both our own analysis and the remaining literature, still has the potential of being the result of one of the confounding variables mentioned previously. For future analyses, we would improve this investigation by ensuring stronger control over the noise in the data.

In conclusion, this study confirmed that maternal smoking during pregnancy increases the risk of low birth weight and indicated that gestational age had a modifying effect on the correlation. In order to reduce infant mortality due to low birth weight, it is necessary to provide more educational programs for mothers as well as more intensive prenatal care for mothers who smoke during their pregnancy as they are more susceptible to birthing infants with low birth weight.

Methods and Theory

In this section, we provide statistical and mathematical reasoning to justify the selection of tests used in this analysis. It is assumed that the reader of this section is familiar of basic notation and terminology used in statistics.

A. The Normal Distribution and Central Limit Theorem

One of the primary concerns in data analysis is the distribution function, $f_Y(y)$ from which the random vector $(X_1, X_2, ..., X_n)$ arises, where $X_i, 1 \leq i \leq n$ is a random variable. One of the most widely used and important distributions in statistics is the Normal (or Gaussian) Distribution, which depends two parameters, μ and σ, the **mean** and **standard deviation**, respectively. The normal distribution may be represented by the following probability density function,

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Where $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < x < \infty$

Since large samples of data from random events often follow a normal distribution, which has many satisfying statistical and mathematical properties, it is imperative to test data for normality. The knowledge that our data is normally distributed allows for the application of many statistical tests, including the Welch's t-test for two independent samples, which was conducted to test a difference in means.

One of the affromentioned properties in question is the Central Limit Theorem, whose statement we owe to the book "Mathematical Statistics with Applications" by Larsen and Marx (5th edition)

Central Limit Theorem. Let $X_1, X_2, ...$ be a sequence of independent random variables each having the same distribution, $f_W(w)$. Suppose that the mean μ and the variance σ^2 of $f_W(w)$ are both finite. If $\sum_{i=1}^{n} W_i = S_n$, then for any numbers a and b,

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Proof. A full proof of this theorem is beyond the scope of this paper. For a sketch of a proof, please see "Mathematical Statistics and Applications" by Larsen and Marx.

□

It should be noted that $f_W(w)$ may indeed be any distribution, and it is this fact which allows the construction of confidence intervals as done in the Analysis of this paper. That is, if $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a random variable denoting the sample mean coming from the sampling distribution $f_Y(y)$, μ and σ are the corresponding population means and variances, then by the Central Limit Theorem above, it follows that $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has approximately a standard normal distribution, provided n is large enough.

B. Bias and Consistency

Bias is defined as the difference between the expected value of the estimator and the true value of the parameter. For an estimate to be useful, it must be unbiased. We evaluated our parameters throughout the analysis and concluded that the biases were sufficiently small.

An estimator $\hat{\theta}$ said to be consistent for θ if it converges in probability to θ, that is, if for all $\varepsilon > 0$,

Consistency refers to an estimator's long-term behavior: an estimator is consistent if as more data points are added, it converges to the true value of the parameter. The estimators we utilized for the analysis were chosen for consistency.

C. Confidence Intervals

Throughout this report, we have given the approximate 95% confidence intervals associated the with arithmetic mean as a range of values to which the true mean μ_0 can take on. Here we justify their use and construction using the normal distribution.

If μ_0 is the true mean, σ is the standard deviation, n the sample size, and z^*_α so that $P(Z \geq z^*_\alpha) = \alpha$, then the "95% confidence interval is a range of values with the form

$$\left[\bar{X} - z*_{(\alpha)} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z*_{(\alpha)} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

"Confidence" implies that should we construct enough intervals in this way, 95% of them would contain the true unknown mean, μ_0.

What follows is a short justification of the construction of confidence intervals: by applying the Central Limit Theorem under appropriate conditions, $\frac{\bar{X}-\mu}{\sigma/n}$ follows an approximately normal distribution, and choosing $z*_\alpha = 1.96$, one can find that

$$P\left(-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 95\%$$

implies that

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

Which is the form of the 95% confidence intervals about the sample mean used in this report.

D. Welch's Two Sample t-Test

We state Welch's Two Sample t-Test as a theorem, which we owe to "Mathematical Statistics and Data Analysis" by John Rice. It is really a slightly more general two sample t-test.

Theorem C. Suppose that $X_1,...,X_n$ are independent and normally distributed random variables with mean μ_X and variance σ_X^2 and that $Y_1,...,Y_m$ are independent and normally distributed random variables with mean μ_Y and variance σ_Y^2 and that Y_i are independent of X_i. Then the statistics,

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} \tag{1}$$

follows an approximate t distribution with v degrees of freedom, where v is

$$v = \frac{[s_X^2/n + s_Y^2/m]^2}{\frac{(s_X^2/n)^2}{n} + \frac{(s_Y^2/m)^2}{m}} - 2 \tag{2}$$

rounded to the nearest integer.

Here $\bar{X}$ and $\bar{Y}$ are the arithmetic means of the random vectors $(X_1,...,X_n)$ and $(Y_1,...,Y_m)$ respectively, with their accompanying sample variances, s_X^2 and s_Y^2.

The two sample t-test is primarily used to hypothesis test on the null hypothesis $H_0 : \mu_X - \mu_Y = 0$ against the alternative hypothesis $H_A : \mu_X - \mu_Y \neq 0$ in order to compile statistical evidence whether there is an observable difference in treatment and control groups. Note that it is assumed the data come from a normal distribution, however, it can be shown via simulation (See: "An Introduction to Mathematical Statistics and its Applications" by Larsen and Marx, Section 7.5 and 9.2) that the t-test is robust against non-normal data, provided sample size is large enough and skew is minimal, and that this fact may be extended to Welch's t-Test.

E. Bootstrap Algorithm

Nearly all of the information presented here we owe to "An Introduction to the Bootstrap" by Bradley Efron and Robert Tibshirani. A thorough explanation of the bootstrap is beyond the scope of this paper. We provide the necessary definitions and a description of the algorithm used in this paper. For a more in-depth discussion on this topic, please see the aforementioned monograph.

The bootstrap is a simulation method that draws **with** replacement from the sample distribution in order to make inferences about the population. Suppose that $(x_1,...,x_n)$ is a random vector representing data drawn from the distribution F. Define $\hat{F}$ to be the empirical distribution of $(x_1,...,x_n)$ putting probability $1/n$ of drawing the data point $x_i, i = 1,...,n$. We define the **bootstrap sample** to be the vector $\mathbf{x}^* = (x_1^*,...,x_n^*)$, where each of the $x_j^*, j = 1,...,n$ is drawn with replacement from $\hat{F}$, hence x_i^* need not equal x_i.

A description of the algorithm

Let $s(\mathbf{x})$ be a test statistic, in particular, an estimator $\hat{\theta}$. Corresponding to the bootstrap sample $\mathbf{x}^*$ is the bootstrap replicate $\hat{\theta}^* = s(\mathbf{x}^*)$ so that $\hat{\theta}^*$ is the test statistic $s(\cdot)$ applied to the bootstrap sample. This processes is repeated many times, say B times times, and the resulting B bootstrap replicates is the collected. Moreover, the processes is done in such a way that each bootstrap sample is independent of the others - in essence adding an element of variability to the B bootstrap replicates $\hat{\theta}^*$. From there, the standard error of the estimator $\hat{\theta}$ may be approximated using the bootstrap replicates in order to construct confidence intervals and conduct hypothesis tests.

In order to conduct a hypothesis test, one first notes their observed test statistic $s(\mathbf{x})_{obs}$, then manipulates $\hat{F}$ so that it is the empirical distribution *under the null hypothesis* H_0. Then B replicates of the bootstrapped test statistic $s(\mathbf{x}^*)$ are computed and

an approximate p-value is calculated by

$$p \approx \frac{1+\sum_{i=1}^{B} I(s_i(\mathbf{x}^*) \geq s(\mathbf{x})_{obs})}{(B+1)} \tag{3}$$

Where I is the indicator function.

That is, we sum the number of simulated observations as extreme or more extreme than our actual observation, $s(\mathbf{x})_{obs}$, and the $+1$ is added to the numerator and denominator to include the initial observation and avoid numerical errors. While this p-value is approximate, it may be made arbitrarily accurate to the true p-value by taking B arbitrarily large. How large should B be? No precise mathematical model exists. However, the bootstrap is fairly robust if the sample size is fairly large, which in our case it is. Due to limitations in computing power, we choose $B = 999,999$, so that $(1-\alpha)(B+1) = k$ is an integer, as recommended in (Efron, 1993).

A description of the bootstrapped difference in means hypothesis test

Let $\mathbf{x} = (x_1, ..., x_n)$ be the random vector from distribution F, and $\mathbf{y} = (y_1, ..., y_m)$ be the random vector from empirical distribution G. Take $\bar{z}$ to the pooled arithmetic mean of $(x_i, y_j), i = 1, ..., n, j = 1, ..., m$. Also, put $\bar{x}$ and $\bar{y}$ be the arithmetic mean of $\mathbf{x}$ and $\mathbf{y}$ respectively. The test statistic used to check a difference in means between two groups is the difference $s(\mathbf{x}, \mathbf{y}) = \bar{x} - \bar{y}$. Then H_0 is simulated by taking the empirical distributions $\hat{F}$ and $\hat{G}$ to put equal probability on the shifted data points,

$$\tilde{x_i} = x_i - \bar{x} + \bar{z}, \quad i = 1, ..., n$$
$$\tilde{y_j} = y_j - \bar{y} + \bar{z}, \quad j = 1, ..., m$$

Samples are drawn with replacement from $\hat{F}$ and $\hat{G}$ to form the B bootstrap replicates $(\mathbf{x}^{*\mathbf{b}}, \mathbf{y}^{*\mathbf{b}})$, $b = 1, ..., B$, which are then passed through our test statistic $s(\cdot)$. An approximate p-value is calculated using (3).

The Bootstrapped Two Sample Welch's t-test

In regards to the Bootstrapped t-test the same process applies, with some minor differences.If $\mathbf{x}$ consists of n observations and $\mathbf{y}$ consists of m observations, we do not resample with n and m bootstrap samples, instead we *subsample* from F and G while fixing the number of subsamples. In this case, choosing $n_s = m_s = 300$, where n_s, m_s are the number of bootstrap samples from $\hat{F}$ and $\hat{G}$, respectively. Then, replace $s(\cdot)$ with $p(\cdot) = P(t(\mathbf{x}) \geq t_{\alpha/2}) + P(t(\mathbf{x}) \leq -t_{1-\alpha/2})$, where $t(\mathbf{x}, \mathbf{y})$ is given in (1) and where t_α is such that $P(T \leq t_\alpha) = 1 - \alpha$ with the degrees of freedom given in (2). Fixing at the $\alpha > 0$ significance level, our approximate p-value is calculated as being

$$p \approx (1 + \sum_{i=1}^{B} I(p_i > \alpha))/(B+1)$$

where I is the indicator function.

F. Monte Carlo Simulations The Monte Carlo simulation technique exists as a method to estimate probabilities and probability distributions using a randomly-generated simulation sample of a given unknown distribution. The method consists of randomly generating observations for the distribution under inspection on the given domain, performing a computation on the generated data, repeating the process many times, and resulting in a generated distribution that is an unbiased estimator to the unknown distribution. It should be noted that we utilize the Monte Carlo under the assumption that the number of data points simulated is large enough to reasonably eliminate any non-representative values that arise due to randomness.

G. Skewness and Kurtosis Let $(x_1, ..., x_n)$ be a sample of observed data points from the sample distribution $f_X(x)$ with sample mean $\bar{x}$ and sample standard deviation s. The lack of symmetry is measured by skewness, (γ_1), where

$$\gamma_1 = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i - \bar{x}}{s})^3 \tag{4}$$

The amount of distribution in the tails of the distribution or how pronounced the peak of the distribution is can be measured by kurtosis, (κ), where

$$\kappa = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i - \bar{x}}{s})^4 \tag{5}$$

Skewness and kurtosis provide useful information about the shape of the sample distribution which can be used to compare with other known distributions. From the comparison, the distribution of the sample data can ideally be identified. For our analysis, the skewness and kurtosis for the sample distribution are compared against that of a standard normal distribution (0 and 3, respectively) which was generated via Monte Carlo simulation. From the Monte Carlo, the 1000 randomly generated values collected to form a roughly standard normal distribution. For $\alpha = 0.05$, we computed the confidence interval for the approximately standard normal generated distribution:

$$[q_{0.025}, q_{0.975}] \tag{6}$$

where 95% of the generated distribution lies within the interval

By evaluating our sample kurtosis values with the generated values, we can determine the state of the sample distribution and thus reject or fail to reject our null hypotheses.

References

Abrevaya, J. (2006). Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach. Journal of Applied Econometrics, 21(4), 489-519. Retrieved from http://www.jstor.org/stable/25146441

Almond, D., Chay, K., & Lee, D. (2005). The Costs of Low Birth Weight. The Quarterly Journal of Economics, 120(3), 1031-1083. Retrieved from http://www.jstor.org/stable/25098762

Chen, Peter J. "Fetal Development." MedlinePlus, U.S. National Library of Medicine, 26 Aug. 2017, medlineplus.gov/ency/article/002398.h

Gray, R., Bonellie, S., Chalmers, J., Greer, I., Jarvis, S., Kurinczuk, J., & Williams, C. (2009). Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994-2003: Retrospective population based study using hospital maternity records. BMJ: British Medical Journal, 339(7726), 906-908. Retrieved from http://www.jstor.org/stable/25672961

Hack, M., Klein, N., & Taylor, H. (1995). Long-Term Developmental Outcomes of Low Birth Weight Infants. The Future of Children, 5(1), 176-196. doi:10.2307/1602514

Johansson, A., Dickman, P., Kramer, M., & Cnattingius, S. (2009). Maternal Smoking and Infant Mortality: Does Quitting Smoking Reduce the Risk of Infant Death? Epidemiology, 20(4), 590-597. Retrieved from http://www.jstor.org/stable/25662707

Kramer, M. (1987). Determinants of low birth weight: methodological assessment and meta-analysis. Bulletin of the World Health Organization, 65(5), 663-737.

Kramer, M. (1987). Intrauterine Growth and Gestational Duration Determinants. Pediatrics, 80(4), 502-511. Retrieved from pediatrics.aappublications.org/content/pediatrics/80/4/502.full.pdf

Muscati, S., Gray-Donald, K., & Newson, E. (1994). Interaction of Smoking and Maternal Weight Status in Influencing Infant Size. Canadian Journal of Public Health / Revue Canadienne De Sante'e Publique, 85(6), 407-412. Retrieved from

http://www.jstor.org/stable/41991221

Paul B. English, Brenda Eskenazi (1992). Reinterpreting the Effects of Maternal Smoking on Infant Birthweight and Perinatal Mortality. International Journal of Epidemiology. 21(6), 1097-1105. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1

Wilcox AJ, Skjaerven R. (1992). Birth weight and perinatal mortality: the effect of gestational age. American Journal of Public Health. 82(3), 378-382.

Woolbright, L. (1994). The Effects of Maternal Smoking on Infant Health. Population Research and Policy Review, 13(3), 327-339. Retrieved from http://www.jstor.org/stable/40229739

Allen J. Wilcox; Birth Weight and Perinatal Mortality: The Effect of Maternal Smoking, American Journal of Epidemiology, Volume 137, Issue 10, 15 May 1993, Pages 1098–1104, https://doi.org/10.1093/oxfordjournals.aje.a116613 (to discuss preterm babies and low birth weight)

Suzuki, K., Shinohara, R., Sato, M., Otawa, S., Yamagata, Z. (2016). Association Between Maternal Smoking During Pregnancy and Birth Weight: An Appropriately Adjusted Model From the Japan Environment and Children's Study. Journal of Epidemiology, 26(7), 371–377. http://doi.org/10.2188/jea.JE20150185 (japanese study, argues with regression that smoking causes 150-200g reduction in birthweight)

Drake P, Driscoll AK, Mathews TJ. Cigarette smoking during pregnancy: United States, 2016. NCHS Data Brief, no 305. Hyattsville, MD: National Center for Health Statistics. 2018. (7.2% of women still smoke during pregnancy in 2016)

Determinants of low birth weight https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491072/pdf/bullwho00076-0086.pdf

How does smoking during pregnancy harm the baby? (https://www.cdc.gov/reproductivehealth/maternalinfanthealth/tobaccousepregnancy/i

Acknowledgements

We provide thanks to the Teacher Assistants of Math 189, for the help they provided in guiding this project. We especially thank Professor Jelena Bradic, who's suggestions on the bootstrap procedure were invaluable.

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.