

# The Negative Impact of Maternal Smoking on Infant Health

James P. Donohue

January 30, 2022

## Abstract

It has been reported by the U.S. Surgeon General that smoking during pregnancy increases the risk for pregnancy complications and adverse effects to the health and wellbeing of infants. Using data provided by the Stat Labs data site we observe a sample of 1236 baby weights, of which it was recorded whether or not the mother was an active smoker. It is important to make note here that the mean and median birthweights for non-smokers is somewhat higher than that of smokers, by a margin of approximately 8 units. It can be concluded that infants of non-smoking mothers are consistently, on average, a noticeably higher birthweight than those of smoking mothers.

# Background and Significance

It has been reported by the U.S. Surgeon General that smoking during pregnancy increases the risk for pregnancy complications and adverse effects to the health and wellbeing of infants. As cited by the CDC, the Surgeon General and U.S. Department of Health and Human Services: “One in every five babies born to mothers who smoke during pregnancy has low birth weight. Mothers who are exposed to secondhand smoke while pregnant are more likely to have lower birth weight babies. Babies born too small or too early are not as healthy.” It is the goal of this report to confirm and support these findings through our observation of the provided data on maternal smoking and infant health.

## Methods

Using data provided by the Stat Labs data site we observe a sample of 1236 baby weights, of which it was recorded whether or not the mother was an active smoker. Initially, the data was separated for smoking and non-smoking mothers into two groups (datasets), this allowed for the creation of two histograms to plot the distribution of both “smoking birthweights” and “non-smoking birthweights.” These histograms were accompanied by descriptive statistics, i.e. mean, median, standard deviation, skewness, kurtosis, and a five-point summary, as visualized by side by side box and whisker plots. To illustrate the difference in distribution between these two groups, both histograms were plotted together in a grouped bar chart. Finally, to confirm that our two datasets came from populations with a common distribution, quantile plots of each group were created.

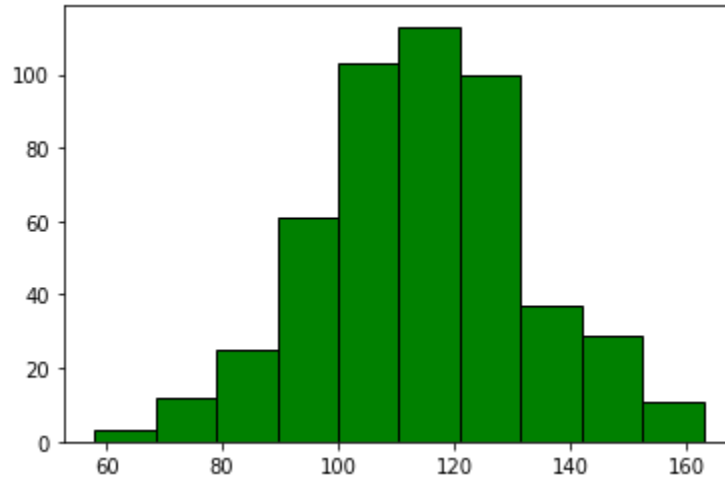
## Results

Descriptive Statistics	Smokers	Non-Smokers
Mean	114.36437246963563	123.04716981132076
Median	115.0	123.0
Standard Deviation	18.241806794356968	17.39868877808028
Skewness	-0.018363532784288962	-0.18736306526595664
Kurtosis	-0.02090922517503202	1.0522126994011294
5-Point Summary	[ 58. 102. 115. 126. 163.]	[ 55. 113. 123. 134. 176.]

It is important to make note here that the mean and median birthweights for non-smokers is somewhat higher than that of smokers, by a margin of approximately 8 units. Although the minimum of the

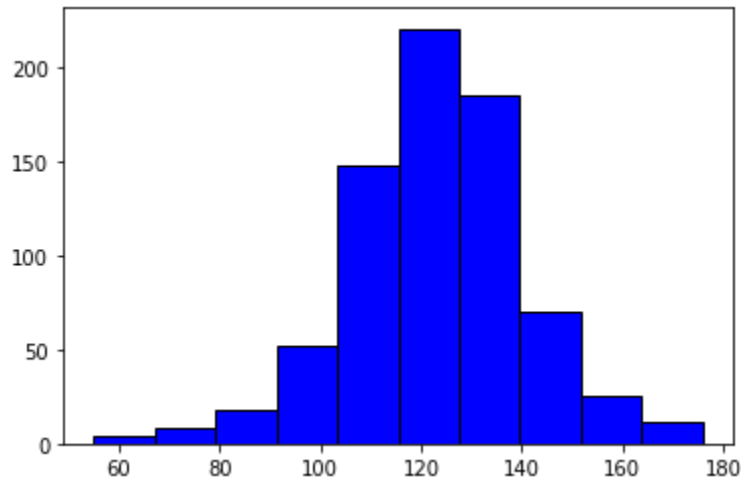
distribution of non-smoker birthweights is lower than that of smokers, which goes against our hypothesis, all other values in the group's five-point summary are noticeably larger than that of the smoker group.

**Birthweights of Smokers' Infants**



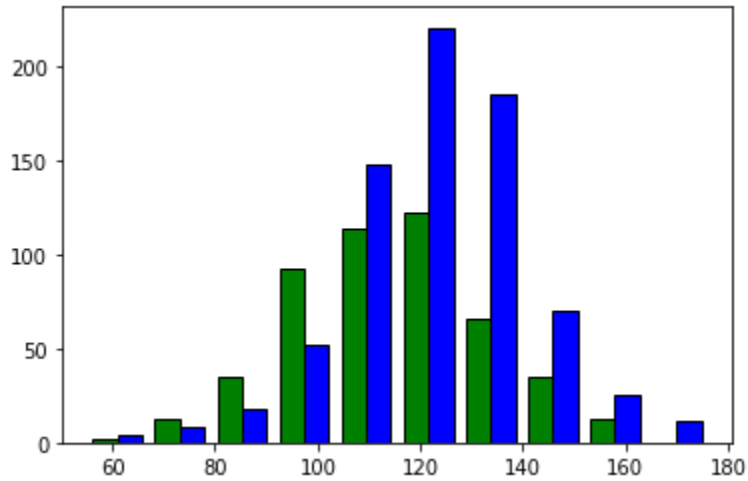
Note above the histogram of birthweights of smokers' infants. This distribution reflects the results observed by our descriptive statistics.

**Birthweights of Non-Smokers' Infants**



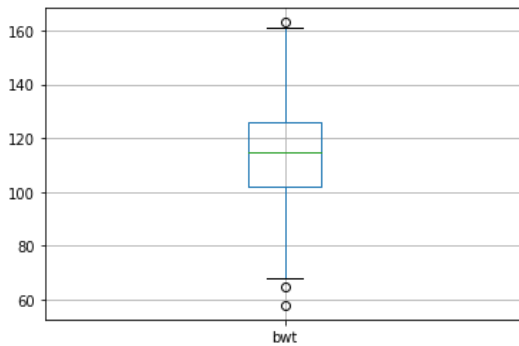
Note above the histogram of birthweights of non-smokers' infants. This distribution reflects the results observed by our descriptive statistics.

### Birthweights of Smokers' Infants vs. Non-Smokers' Infants

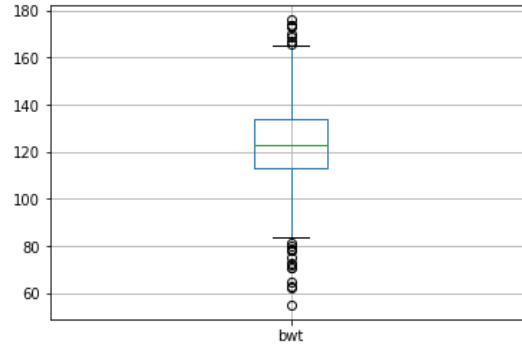


Note above the grouped bar chart of birthweights of smokers' infants against that of non-smokers' infants. This comparison makes the greater skewness and kurtosis of non-smokers' infant birthweights much more apparent.

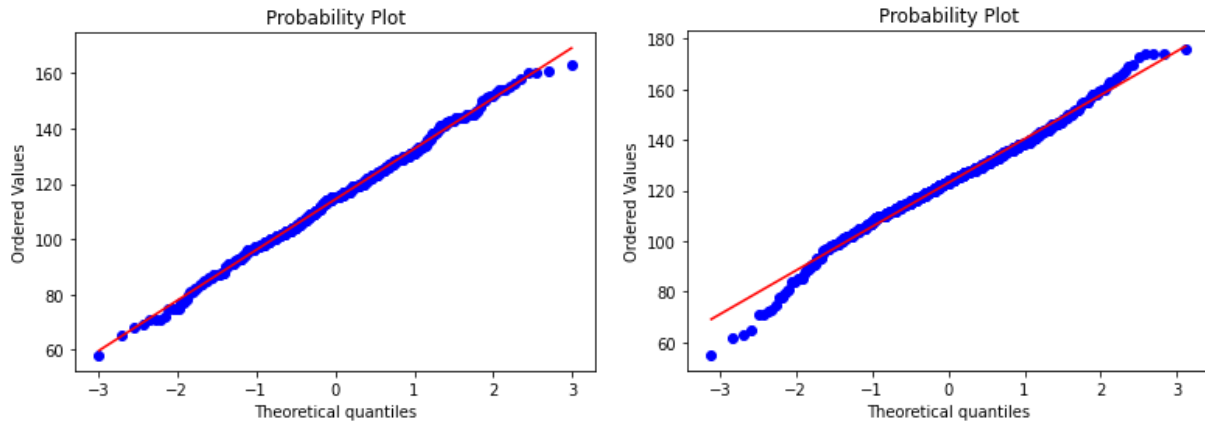
#### Distribution of Smokers' Infants



#### Distribution of Non-Smokers' Infants



Note above the boxplots of the distributions of both smokers' infants and non-smokers' infants birthweights. These plots reflect the above table of descriptive statistics as well as displaying a noticeable amount of outliers in the distribution of non-smokers' infant birthweights. This group also has a noticeably smaller interquartile range.



The above quantile plots illustrate the close to normal distributions of the birthweights of both smokers' and non-smokers' infants. The slightly greater skewness of the non-smokers distribution can be seen on the lower end of the group's quantile plot. The above plots confirm the close-to-normal distributions of both datasets as well.

## Conclusion

It can be concluded that infants of non-smoking mothers are consistently, on average, a noticeably higher birthweight than those of smoking mothers. As shown in the descriptive statistics and illustrated in the above grouped bar chart and box and whisker plots, the mean and median birthweights of non-smokers' infants is noticeably higher in both value and population than that of smokers.

## Bibliography

Centers for Disease Control and Prevention. (2020, April 28). Smoking during pregnancy. Centers for Disease Control and Prevention. Retrieved January 31, 2022, from [https://www.cdc.gov/tobacco/basic\\_information/health\\_effects/pregnancy/index.htm](https://www.cdc.gov/tobacco/basic_information/health_effects/pregnancy/index.htm)

Highlights: Overview of findings regarding Reproductive Health. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved January 31, 2022, from [https://www.cdc.gov/tobacco/data\\_statistics/sgr/2010/highlight\\_sheets/pdfs/overview\\_reproductive.pdf](https://www.cdc.gov/tobacco/data_statistics/sgr/2010/highlight_sheets/pdfs/overview_reproductive.pdf)

Stat Labs Data Page. (n.d.). Retrieved January 31, 2022, from <https://www.stat.berkeley.edu/~statlabs/labs.html#babiesI>  
Cited for providing the given dataset

# Appendix

## Problem 19:

$$\begin{aligned}\sum(x_i + c)^2 \\ \sum(x_i + c)(x_i + c) \\ \sum(x_i^2 + c^2 - 2x_i c) \\ \sum(2x_i + 0 - 2c) = 0 & \quad \text{take derivative} \\ \sum(2x_i - 2c) = 0 \\ 2\sum(x_i - c) = 0 \\ \sum(x_i - c) = 0 \\ \sum(x_i) - \sum(c) = 0 \\ \sum(x_i) - nc = 0 & \quad \text{by law of summation} \\ \sum(x_i) = nc \\ c = \sum(x_i)/n \\ c = \bar{x}\end{aligned}$$

## Code:

```
# -- coding: utf-8 --
"""
Created on Thu Jan 20 11:05:55 2022

@author: James P. Donohue
"""

import pandas as pd
from matplotlib import pyplot as plt
import numpy as np
import statsmodels.api as sm
import pylab
import scipy.stats as stats

# reading csv files
data = pd.read_csv(r'/content/babies.csv', sep=',')
#print(data)

dfSmokers = data.loc[(data["smoke"]>0)]
#print(dfSmokers)
#descriptive statistics for smoking mothers
```

```

smokeMean = dfSmokers['bwt'].mean()
print(smokeMean)
smokeMed = dfSmokers['bwt'].median()
print(smokeMed)
smokeStd = dfSmokers['bwt'].std()
print(smokeStd)
smokeSkew = dfSmokers['bwt'].skew()
print(smokeSkew)
smokeKurt = dfSmokers['bwt'].kurtosis()
print(smokeKurt)
#five-point summary (min, first quartile, median, third quartile, max)
smokeFivePoint = np.percentile(dfSmokers['bwt'], [0, 25, 50, 75, 100],
interpolation='midpoint')
print(smokeFivePoint)
dfNotSmokers = data.loc[(data["smoke"]<1)]
#print(dfNotSmokers)
#descriptive statistics for smoking mothers
nonSmokeMean = dfNotSmokers['bwt'].mean()
print(nonSmokeMean)
nonSmokeMed = dfNotSmokers['bwt'].median()
print(nonSmokeMed)
nonSmokeStd = dfNotSmokers['bwt'].std()
print(nonSmokeStd)
nonSmokeSkew = dfNotSmokers['bwt'].skew()
print(nonSmokeSkew)
nonSmokeKurt = dfNotSmokers['bwt'].kurtosis()
print(nonSmokeKurt)
#five-point summary
nonSmokeFivePoint = np.percentile(dfNotSmokers['bwt'], [0, 25, 50, 75,
100], interpolation='midpoint')
print(nonSmokeFivePoint)

list_one = dfSmokers.bwt.tolist()
list_two = dfNotSmokers.bwt.tolist()

# histograms of baby weight
plt.hist(list_one, ec='black', color='green') #smokers
plt.show()

plt.hist(list_two, ec='black', color='blue') #non smokers
plt.show()

plt.hist([list_one, list_two], ec='black', color=['green', 'blue'])

```

```
#smokers vs non-smokers
plt.show()

# box and whisker plots

dfSmokers.boxplot(column='bwt')
plt.show()

dfNotSmokers.boxplot(column='bwt')
plt.show()

# quantile plots

#smoker
stats.probplot(list_one, dist = 'norm', plot=pylab)
pylab.show()
#non-smoker
stats.probplot(list_two, dist = 'norm', plot=pylab)
pylab.show()
```