

Birth weight of Smoke and Nosmoke Mothers

Edward Huynh

Feb 2, 2020

## 1. Abstract

I investigate the differences in weight between babies born to mothers who smoked during pregnancy and those who did not. Assess the importance to the health of babies of the differences in three types of comparisons (numerical, graphical, incidence) by using Rstudio program method. I find that the weight of babies who born from non-smoked mother is higher and more healthier than the ones from smoked mothers.

## 2. Introduction & Background

In this statistical report, i'm going to analyze the data "babiesI.data" which was provided from the Child Health and Development Studies (CHDS)—a comprehensive investigation of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco–East Bay area (Yerushalmy [Yer71]). I analyze this study by using Rstudio program to find the differences in birthweights between smoke and no-smoke mothers to answer whether or not smoking during pregnancy will affect a baby's birth weight and whether or not a baby's birth weight has an effect on its health. Typically, smaller babies and babies born early have lower survival rates than larger babies who are born at term.

## 3. Methods and Results

- ❖ *Data collection:* The data available for this report, "babiesI.data", is a subset of a much larger study — the Child Health and Development Studies. The entire CHDS database includes all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan in Oakland, California. This information collected for 1236 babies — those baby boys born during one year of

the study who lived at least 28 days and who were single births (i.e., not one of a twin or triplet). The information available for each baby is birth weight and whether or not the mother smoked during her pregnancy.

❖ Analysis: I use Rstudio to import dataset “babiesI.data”, then analyze steps by steps to collect results and plots for this study as below:

- A. Summarize numerically the two distributions of birth weight for babies born to women who smoked during their pregnancy and for babies born to women who did not smoke during their pregnancy.

Import data into Rstudio and make data frame for data set.

```
> data<-read.table("C:/Users/Sang/Documents/babiesI.data", header=TRUE,sep="")
> str(data)
'data.frame': 1236 obs. of 2 variables:
 $ bwt : int 120 113 128 123 108 136 138 132 120 143 ...
 $ smoke: int 0 0 1 0 1 0 0 0 0 1 ...
```

Extract and data frame the list of birth weights of babies of non-smoking mothers from the data frame. Summary of five number( the Median, the Upper Quartile, the Lower Quartile, the Maximum value, and Minimum value) of birth weight for babies born to women who didn't smoke during pregnancy.

```
> nosmoke<-subset(data,smoke==0)
> str(nosmoke)
'data.frame': 742 obs. of 2 variables:
 $ bwt : int 120 113 123 136 138 132 120 140 114 115 ...
 $ smoke: int 0 0 0 0 0 0 0 0 0 0 ...
> summary(nosmoke)
      bwt      smoke
Min.   : 55   Min.   :0
1st Qu.:113   1st Qu.:0
Median :123   Median :0
Mean   :123   Mean   :0
3rd Qu.:134   3rd Qu.:0
Max.   :176   Max.   :0
```

Extract the list of birth weights of babies of smoking mothers from the data frame. Summary of five number( the Median, the Upper Quartile, the Lower Quartile, the Maximum value, and Minimum value) of birth weight for babies born to women who smoked during pregnancy.

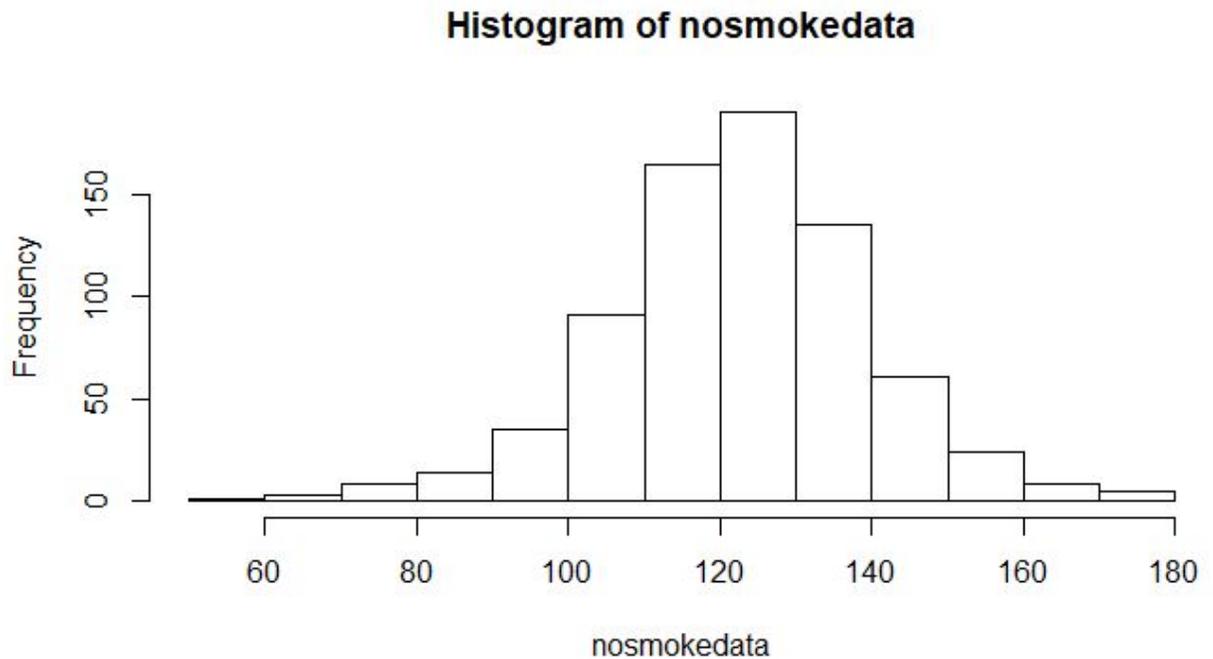
```
> smoke<-subset(data,smoke==1)
> str(smoke)
'data.frame':  484 obs. of  2 variables:
 $ bwt  : int  128 108 143 144 141 110 92 115 119 115 ...
 $ smoke: int   1  1  1  1  1  1  1  1  1  1 ...
> summary(smoke)
      bwt      smoke
Min.   : 58.0   Min.   :1
1st Qu.:102.0   1st Qu.:1
Median :115.0   Median :1
Mean   :114.1   Mean   :1
3rd Qu.:126.0   3rd Qu.:1
Max.   :163.0   Max.   :1
```

In numerical summary, my results show babies born to mothers who are smokers have an average birth weight 114.1 ounces and babies born to mothers who are non-smoke have an average birth weight 123 ounces. The first quartile in the summary tables also indicates that 25% of the babies born to smoking mothers have birth weights lower than 102 ounces and 25% of the babies born to non-smoked mothers have birth weights lower than 113 ounces. The minimum weight of the babies born to smoking and non-smoking mothers are 58 and 55 ounces respectively. The maximum weight of babies born to smoking and non-smoking mothers are 163 and 176 ounces respectively.

➤ B. Use graphical methods to compare the two distributions of birth weight.

Creating and data framing a vector of numeric values of the birth weights of the babies of non-smoking mothers and histogram of this set of data.

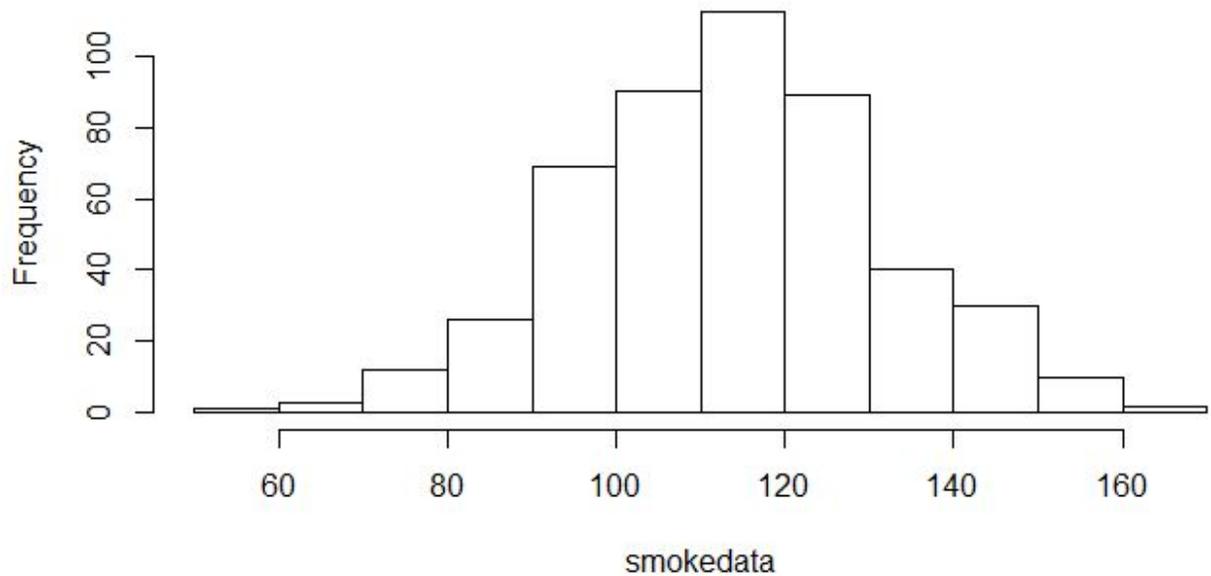
```
> nosmokedata<-nosmoke[['bwt']]
> str(nosmokedata)
 int [1:742] 120 113 123 136 138 132 120 140 114 115 ...
> hist(nosmokedata)
```



Creating and data framing a vector of numeric values of the birth weights of the babies of smoked mothers and histogram of this set of data.

```
> smokedata<-smoke[['bwt']]
> str(smokedata)
 int [1:484] 128 108 143 144 141 110 92 115 119 115 ...
> hist(smokedata)
```

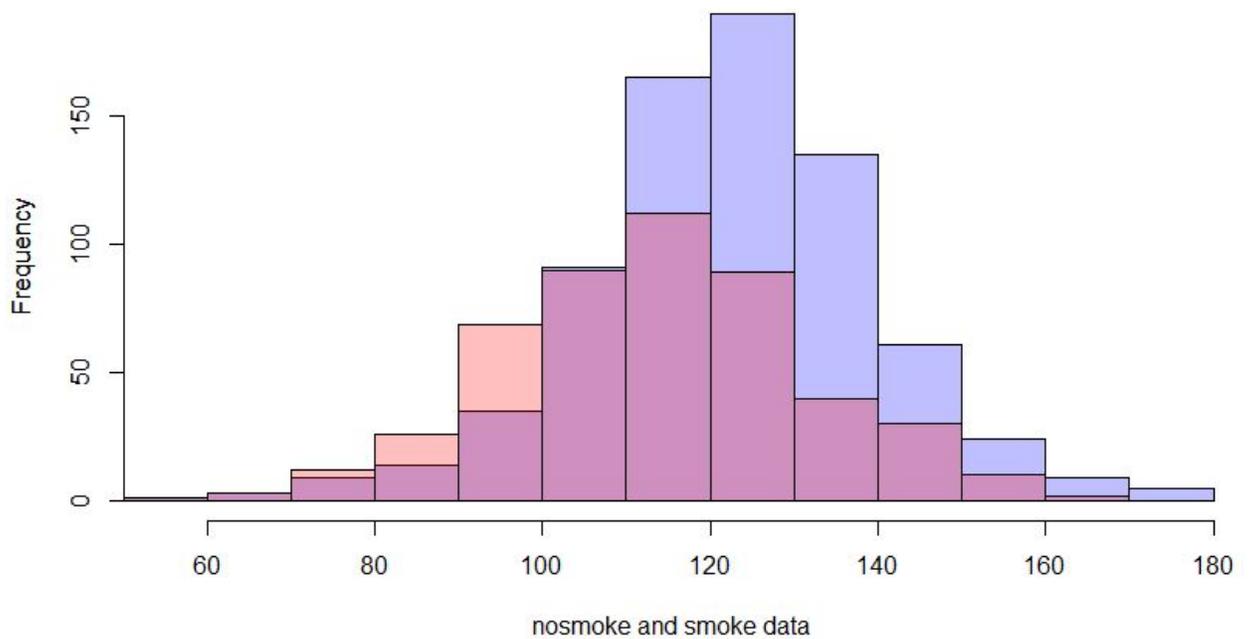
## Histogram of smokedata



In order to see the difference of two set of data birthweight of smoke and non-smoke mothers, i use the overlapping histogram to compare. I set the name for both data, then set the min and max values, set the color of histograms in different colors ( red = Smokers , blue and Non-smokers) as commands below.

```
> p1<-hist(nosmokedata)
> p2<-hist(smokedata)
> a<-min(min(nosmokedata),min(smokedata))
> b<-max(max(nosmokedata),max(smokedata))
> plot(p1,col=rgb(0,0,1,1/4),main="",sub="",xlab="",xlim=c(a,b))#first histogram
> plot(p2,col=rgb(1,0,0,1/4),main="",sub="",xlab="",xlim=c(a,b),add=T)#second histogram
> title(main="Histograms of nosmoke and smoke data",xlab="nosmoke and smoke data",ylab="Frequency")
```

**Histograms of nosmoke and smoke data**



In graphical method summary, The overlapping histogram plot shows that the distribution of babies from smoking mothers are slightly to the left of the distribution of babies from non-smoke mothers. This result indicates that babies from non-smoke mothers tend to have a higher birth weight in general. Moreover, the babies with non-smoke mothers have higher frequency of high birth weight (110 - 140 ounces) while babies with smoking mothers have a low frequency of high birth weight.

- C. Compare the frequency, or incidence, of low-birth-weight babies for the two groups.  
How reliable do you think your estimates are?

In order to compare the values of frequency or incidence, I have to download the packed of moment in library as command below

```

> install.packages("moments")
WARNING: Rtools is required to build R packages but is not currently installed
Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Sang/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/moments_0.14.zip'
Content type 'application/zip' length 56559 bytes (55 KB)
downloaded 55 KB

package 'moments' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Sang\AppData\Local\Temp\Rtmp4A84uU\downloaded_packages
> library(moments)

```

Compare the sd, skewness, kurtosis values as commands belows

```

> skewness(nosmokedata)
[1] -0.1869841
> skewness(smokedata)
[1] -0.03359498

```

```

> kurtosis(nosmokedata)
[1] 4.03706
> kurtosis(smokedata)
[1] 2.988032

```

```

> sd(smokedata)
[1] 18.09895
> sd(nosmokedata)
[1] 17.39869

```

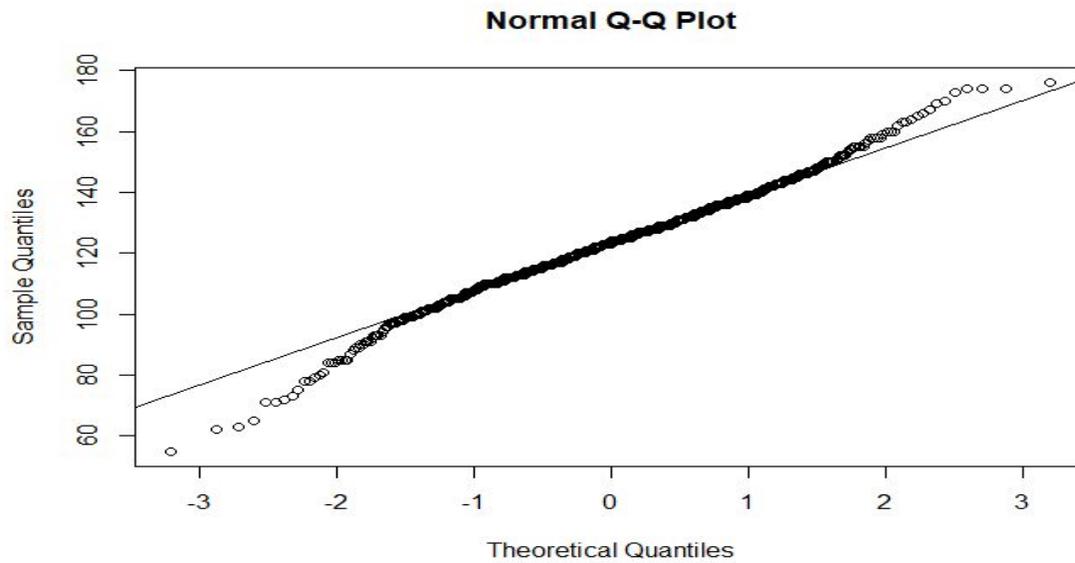
## Quantile (=QQ) plots

To compare a standard normal distribution  $N(0,1)$  with the distribution of non-smoke mothers data set and to fit a line of best fit through the above points, use the command:

```

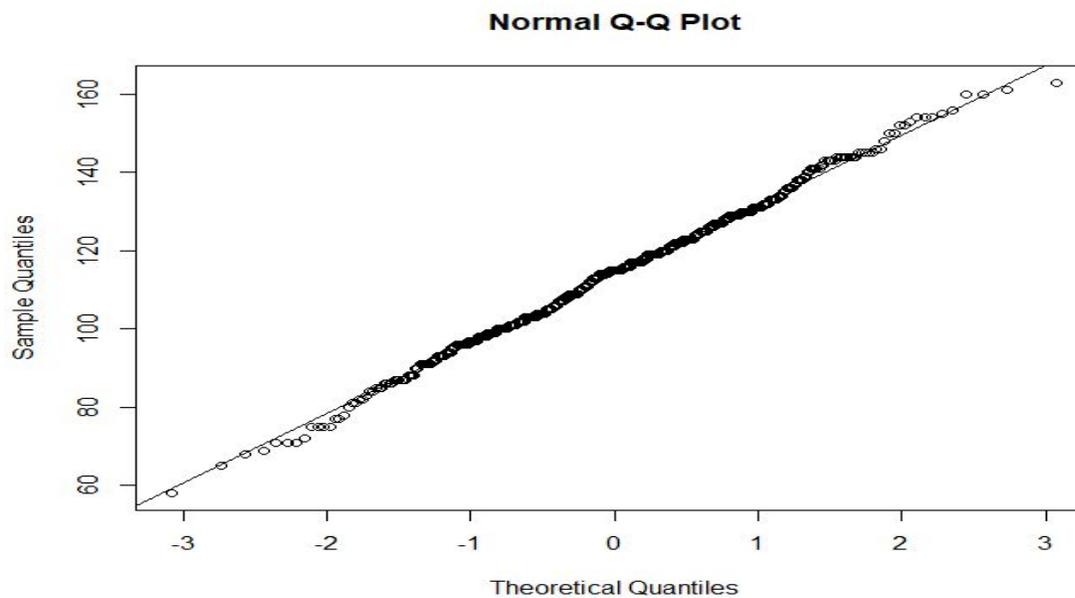
> qqnorm(nosmokedata)
> qqline(nosmokedata)

```

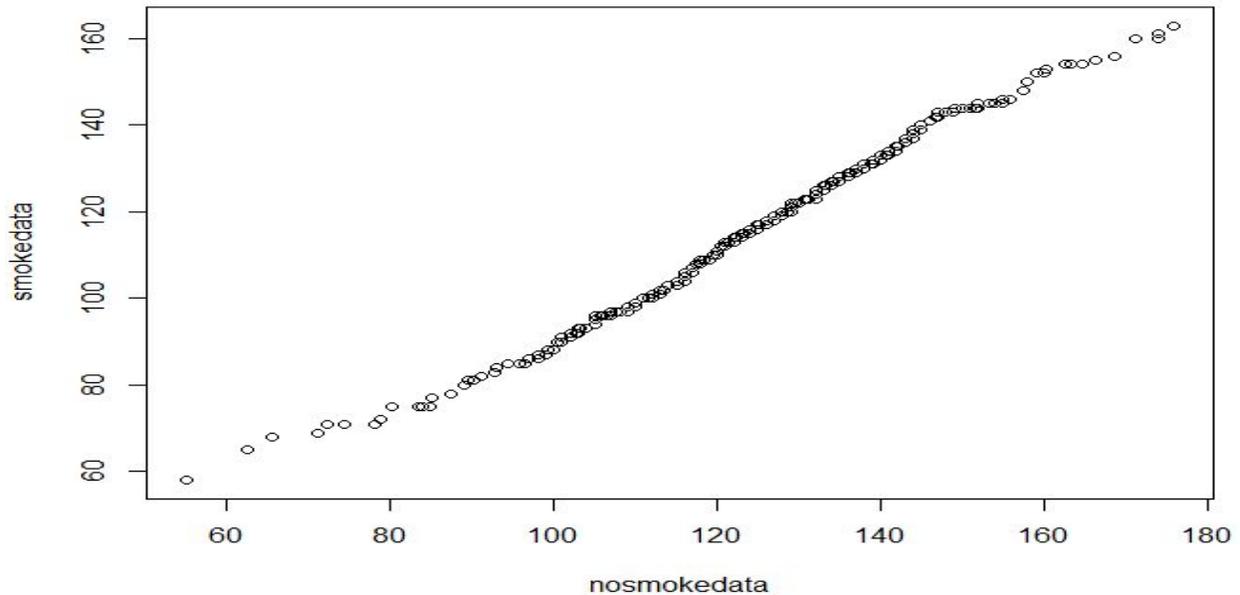


To compare a standard normal distribution  $N(0,1)$  with the distribution of smoke mothers data set and to fit a line of best fit through the above points, use the command:

```
> qqnorm(smokedata)
> qqline(smokedata)
```



Construct a QQ plot for two data set `> qqplot(nosmokedata, smokedata)`



In summary of comparing frequency, standard deviation, the kurtosis, the skewness values would tell me how far away to the mean value of both set of data as the measure of standard score. Plus with the two Q-Q plots show that most of the data points fall onto the reference line. Therefore, I can determine that a normal distribution is a valid statistical model for my data.

#### 4. Discussion & Conclusions

Relate the findings in my numerical and graphical analysis indicate a differences in weight between babies born to mothers who smoked during pregnancy and those who did not. Assess the importance to the health of babies of the differences correlation between the birth weights of the babies and whether or not their respective mothers smoked during pregnancy. Babies of smoking mothers have a lower average birth weight than babies of non-smoking mothers. My graphical analysis also indicates that the variance of birth weight with non-smoking mothers is lower than the birth weight with smoking mothers. The frequency of babies from non-smoking mothers are also more likely to stay in healthier weight from (100 - 140 ounces);

meanwhile, the frequency of babies from smoking mothers are more likely to stay in low-weight birth area. To sum up, the babies born to non-smokers have a higher average birth weight in general which means that the babies would be healthier than the babies born to smoking mothers.