Maternal Smoking and Infant deaths

Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight.

- Is this warning to be taken seriously ?
- Dataset: CHDS - Berkeley, California.
- Taken entirely from Chapter 1 of the book
  Stat Labs: Mathematical Statistics Through Applications by Deborah Nolan and Terry P. Speed
- Local Copy: Chapter 1 File
- Check website: https://www.stat.berkeley.edu/users/statlabs/

Maternal Smoking and Infant deaths: Data set

- At birth, measurements on the baby were recorded. They included the baby's length, weight, and head circumference.
- `Babies1.data` is a subset of this information collected. Contains data on:
  - weight of 1236 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).
  - if the mother smoked during her pregnancy.

The data set is available in the shared dropbox folder.

Maternal Smoking and Infant deaths: Data set

- Epidemiological Studies indicate that birth weight is a measure of the baby's maturity (health).
- Reading:- From Chapter 1, please read sections on *Fetal Development*, *Rubella* and *Physical Model*.

Maternal Smoking and Infant deaths: Question

- Is there a difference between birth weights of babies born to smokers and those born to non-smokers ?
- Is the difference important to the health and development of the baby ?

Maternal Smoking and Infant deaths

[Yer71] Yerushalmy:

- The babies were grouped according to their birth weight;
- Within each group, the numbers of babies that died in the first 28 days after birth for smokers and nonsmokers were compared.
- To accommodate the different numbers of babies in the groups, rates instead of counts are used in making the comparisons.

Maternal Smoking and Infant deaths

He calculated Neonatal mortality rates per 1000 births by birth weight (grams) for live-born infants of white mothers, according to smoking status. They are given in table

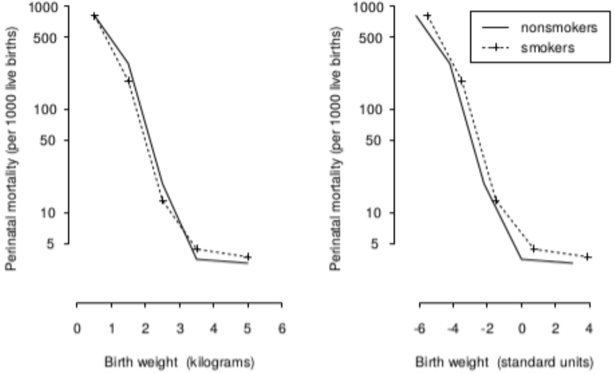
Weight in Grams	Nonsmoker	Smoker
$\leq$ 1500	792	562
1500-2000	406	346
2000-2500	78	27
2500-3000	11.6	6.1
3000-3500	2.2	4.5
3500+	3.8	3.6

Maternal Smoking and Infant deaths: Conclusion 1

[Yer71] Yerushalmy found that although low birth weight is associated with an increase in the number of babies who die shortly after birth, the babies of smokers tended to have much lower death rates than the babies of nonsmokers.

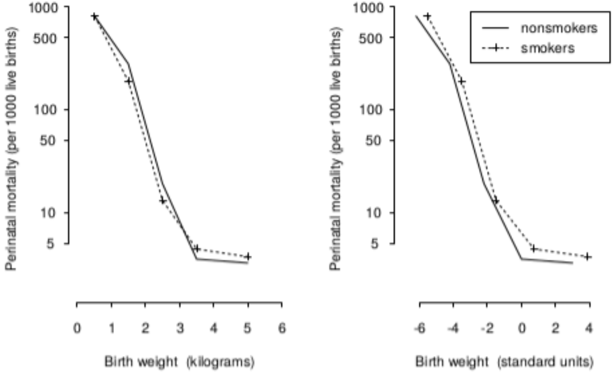
Maternal Smoking and Infant deaths

Wilcox and Russell [WR86] advocate grouping babies according to their relative birth weights. Plotted Mortality rate for perinatal stage.



Maternal Smoking and Infant deaths: Conclusion 2

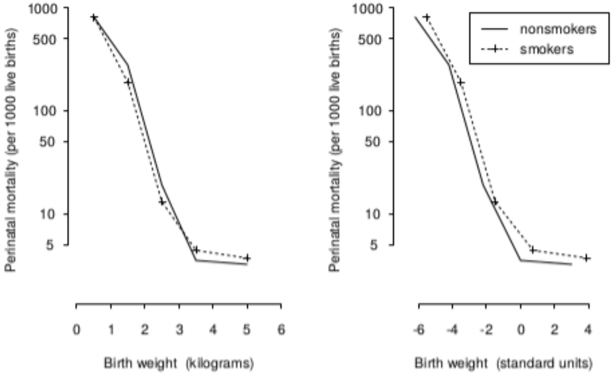
[Wr86] They found that for babies born at term, smokers have higher rates of perinatal mortality in every standard unit.



Maternal Smoking and Infant deaths: Two conclusions?

- [Y71]
  - not adjusted for the mother's age.
  - young smoker versus old non-smoker. Medical factors.
- [MKLS88] Adjusted for above:
  - found that the risk of neonatal death for babies who were born at 32+ weeks gestation is roughly the same for smokers and nonsmokers.
  - It was also found that the smokers had a higher rate of very premature deliveries (20-32 weeks gestation), and so a higher rate of early fetal death.

Maternal Smoking and Infant deaths



[Wr86]

- Babies born to smokers tend to be smaller, the mortality curve is shifted to the right relative to the nonsmokers curve. verify?
- If the babies born to smokers are smaller but otherwise as healthy as babies born to nonsmokers, then the two curves in standard units should roughly coincide.

Normal Distribution

```
> help(Normal)
```

will tell you what functions of Normal distribution that are provided in R.

Normal Distribution: PDF

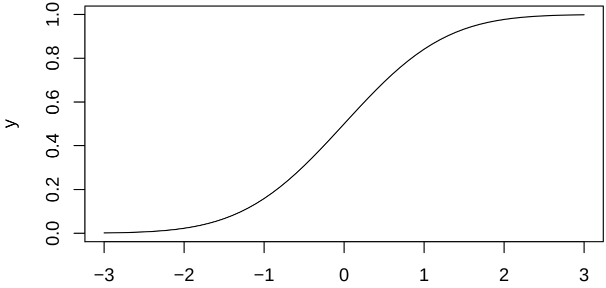
You can calculate the values of the normal density function using the the `dnorm` command.

```
> dnorm(0)
[1] 0.3989423
> dnorm(1)
[1] 0.2419707
> dnorm(0, mean=4, sd=3)
[1] 0.05467002
```

Normal Distribution: CDF

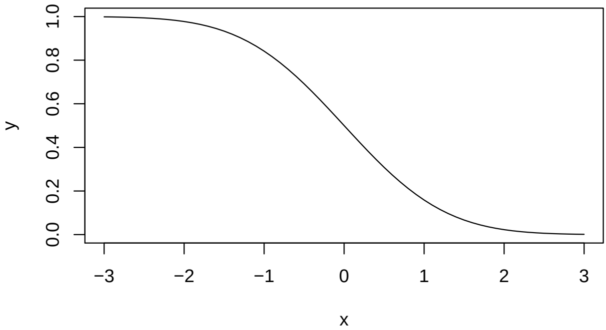
You can calculate the values of the cummulative distribution function of the normal using the the `pnorm` command.

```
> pnorm(0)
> pnorm(1)
> x = seq(-3,3, by=0.1); y = pnorm(x) ;plot(x,y, type="l")
```



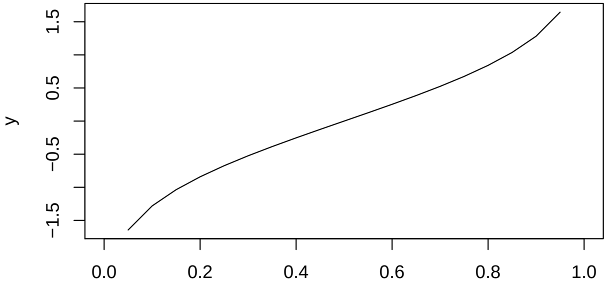
Normal Distribution: Tail Probabilities

```
> pnorm(0, lower.tail=FALSE)
> pnorm(1, lower.tail=FALSE)
> x = seq(-3,3, by=0.1); y = pnorm(x, lower.tail=FALSE)
> plot(x,y, type="l")
```



Normal Distribution: quantiles

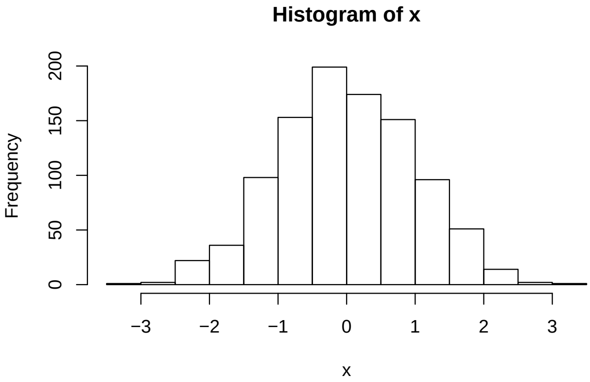
```
> qnorm(0.68); qnorm(0.95);qnorm(0.997)
[1] 0.4676988
[1] 1.644854
[1] 2.747781
> x = seq(0,1, by=0.05); y = qnorm(x);plot(x,y, type="l")
```



Normal Distribution: samples

```
> x=rnorm(1000)

> hist(x)
```

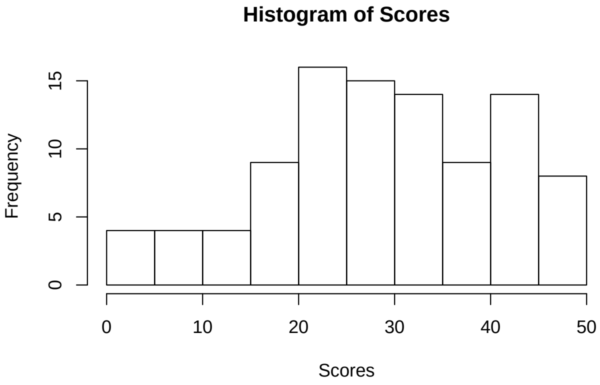


Normal Distribution: Key Probabilities

```
> pnorm(1) - pnorm(-1) # within one standard deviation
[1] 0.6826895
> pnorm(2) - pnorm(-2) # within two standard deviation
[1] 0.9544997
> pnorm(3) - pnorm(-3) # within three standard deviation
[1] 0.9973002
```

Are the Scores normal ?

```
> Scores = scan("Scores")
> hist(Scores)
```



Are the Scores normal ?

```
> summary(Scores)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   22.00   30.00   29.28   40.00   50.00
> sd(Scores)
[1] 12.06816
> cs = (Scores-mean(Scores))/(sd(Scores))
```

Are the Scores normal ?

```
> mean(cs)
[1] 7.922276e-17
> sd(cs)
[1] 1
> onesdcs = cs[cs >-1& cs <1] # within one sd
> twosdcs = cs[cs >-2& cs <2] # within two sd
> threesdcs = cs[cs >-3& cs <3]# within three sd
```

Are the Scores normal ?

```
> length(onesdcs)/length(cs)
[1] 0.628866
> length(twosdcs)/length(cs)
[1] 0.9587629
> length(threesdcs)/length(cs)
[1] 1
```

This is called the 68-95-99.7 first check to see if data is normal or not.

Kurtosis and Skewness

Let $\{x_i : 1 \leq i \leq n\}$ be given data set. Then the `mean` and Standard Deviation, `sd`, are given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \sigma_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Another check for normal are *Kurtosis* and *Skewness* coefficients.

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

Normal Distribution: Kurtosis and Skewness

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

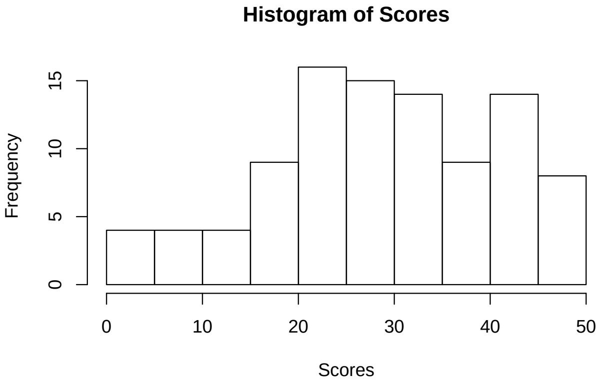
- Skewness 0 indicates that the distribution is symmetric
- Kurtosis is a measure of the peak of the distribution.
- For standard normal Kurtosis is 3.

If Kurtosis and Skewness coefficients are *far* from 3 and 0 we can conclude that the data is not normal.

Are the Scores normal ?

```
> Scores = scan("Scores")

> hist(Scores)
```



Are the Scores normal ?

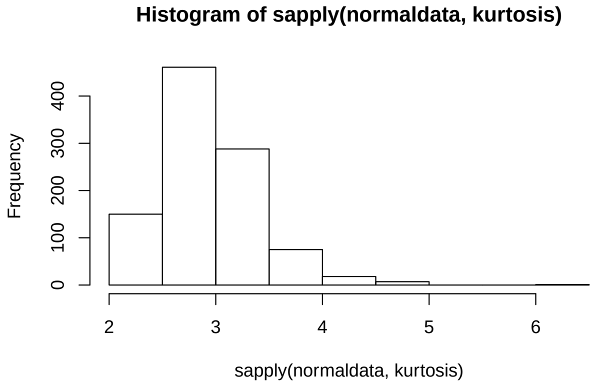
```
> Scores = scan("Scores")
> library(moments)
> Kurtosis = kurtosis(Scores)
> Kurtosis
[1] 2.591014
> Skewness = skewness(Scores)
> Skewness
[1] -0.3548957
```

Are the Scores normal ?

```
> normaldata = replicate(1000, rnorm(97,0,1),
+ simplify=FALSE)
> library(moments)
> mean(sapply(normaldata, kurtosis))
[1] 2.952687
> mean(sapply(normaldata, skewness))
[1] 0.0002724214
```

Are the Scores normal ?

```
> normaldata = replicate(1000, rnorm(97,0,1),
+ simplify=FALSE)
> library(moments)
> hist(sapply(normaldata, kurtosis))
```



Normal Distribution: Quantiles

If Z is standard Normal random variable then the α-th quantile of Z is denoted by z_α where

$$P(Z \leq z_\alpha) = \alpha, \qquad 0 < \alpha < 1.$$

Let $\{x_i : 1 \leq i \leq n\}$ be given data set then we can order them to get

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

we view $x_(k)$ as the $\frac{k}{n+1}$ sample quantile.

Normal Distribution: Quantiles

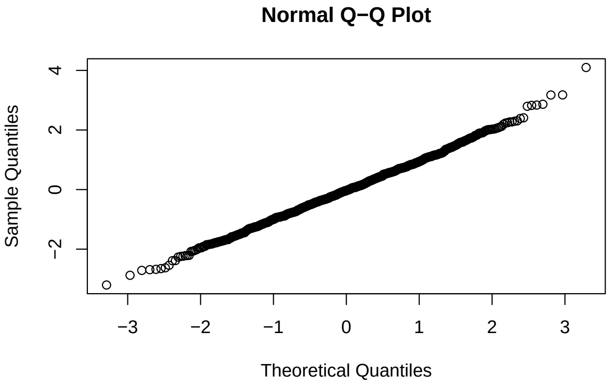
Another check whether data is normal is to plot

$$\left(z_{\frac{k}{n+1}}, x_{(k)}\right)$$

- if the plot is a straight line then it indicates that data is normal

Normal Distribution: Sample Normal-Quantile plot

```
> x=rnorm(1000)
> qqnorm(x)
```



Scores Normal-Quantile-plot

```
> qqnorm(Scores)
```