Normality Test and Distributiom

Hugo Toledo-Alvarado

To check the assumptions of the model.

Read the data

```
filedir<-"https://raw.githubusercontent.com/Hugo-Toledo/Applied-Statistics-R-UNIPD/master/data/cows.txt"
cows<-read.table(file = filedir,stringsAsFactors = FALSE,header = TRUE, sep = "\t")
head(cows) # Show the data
```

```
##             MATR breed    herd parity OP milk dim CLDIM  fat protein cellu
## 1 IT014TN003B016 BRUNA 1805639      7  3 15.6 266     6 4.08    3.84  1006
## 2 IT014TN003C020 BRUNA 1805639      6  3 26.4 151     4 3.57    3.58   510
## 3 IT022000032815 BRUNA 1805639      4  3 36.4  51     2 4.84    3.21   186
## 4 IT022000039390 BRUNA 1805639      3  3 32.9 185     5 4.40    3.53   714
## 5 IT022000039394 BRUNA 1805639      4  3 25.2 176     4 4.77    3.99   374
## 6 IT022000039417 BRUNA 1805639      4  3 34.9  35     1 3.42    3.21    44
##   casein    r   sh   ph
## 1   2.78 17.4 3.10 6.76
## 2   2.72 15.0 2.58 6.80
## 3   2.75 10.2 3.26 6.67
## 4   2.71 16.2 3.05 6.78
## 5   2.98 20.4 3.39 6.72
## 6   2.69 12.0 3.16 6.74
```

Descriptive Statistics and Normality Test

```
milk<-cows$milk      # Select the variable
summary(milk)        # Basic statistics
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.30   24.88   29.05   30.09   35.30   50.90
```

```
sd(milk)             # Standard Deviation function
```

```
## [1] 7.614805
```

```
range(milk)          # Range function
```

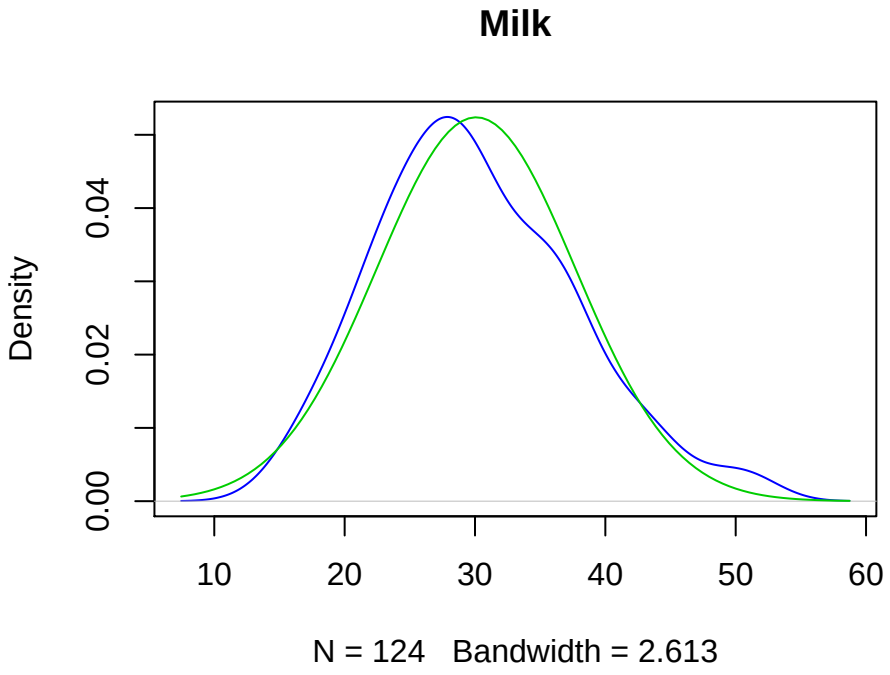
```
## [1] 15.3 50.9
```

```
shapiro.test(milk)   # Shapiro - Wilk normality test
```

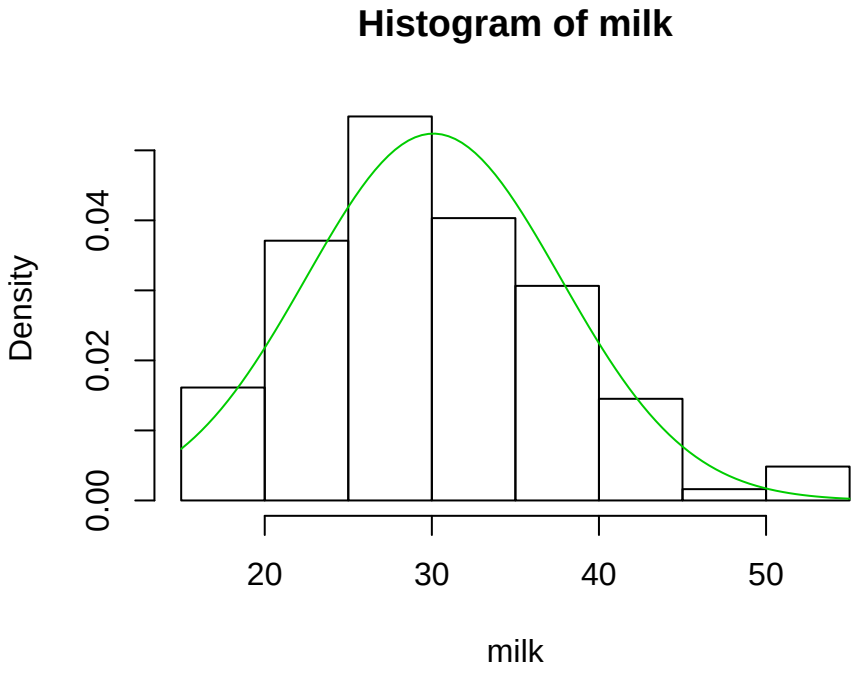
```
##
##  Shapiro-Wilk normality test
##
## data:  milk
## W = 0.97876, p-value = 0.04779
```

Density plot with normal distribution curve

```
plot(density(milk), col=4, main="Milk") # Create a density plot
curve(dnorm(x,mean=mean(milk),sd = sd(milk)),add = T,col=3) # Add normal dist.
```

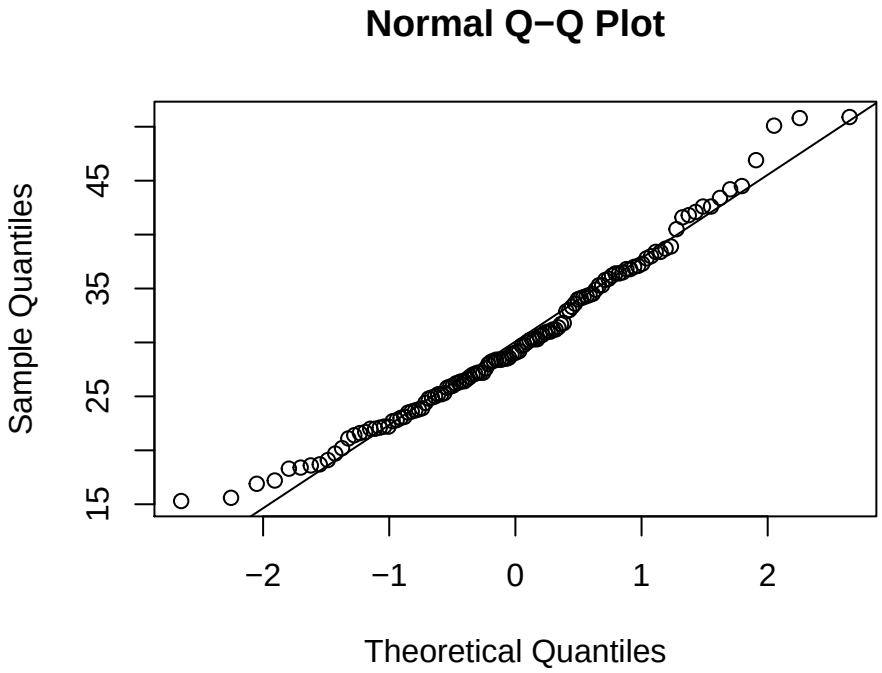



```
hist(milk, breaks = 10, freq = FALSE) # Create histogram
curve(dnorm(x,mean=mean(milk),sd = sd(milk)),add = T,col=3) # Add normal dist.
```



Q-Q Plot

```
qqnorm(cows$milk) #  Create a qq-plot
qqline(cows$milk) # Add line
```




```
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
#NOTE: the library psych has a function
#to describe data including  test for kurtosis and skew

#install.packages("psych")
#library(psych)
#describe(milk,skew = TRUE,ranges = TRUE,quant = c(0.1,0.99))
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```