Statistics with R – Beginner Level

Section 5

Checking Assumptions

Lesson 27 - Checking the Normality Assumption - Numerical Method

```
demo <- read.csv("demographics.csv")

View(demo)

##########
### how to perform the Shapiro-Wilk normality test
##########

### we will check whether the variable income is normally
distributed

shapiro.test(demo$income)

### the shapiro.test function is included in the stats
package,
### which is loaded by default when you start R
```

Lesson 28 - Checking the Normality Assumption - Graphical Methods

```
demo <- read.csv("demographics.csv")

View(demo)

##########
### how to check for normality
```

```
### with the help of the histogram
##########

### we will check whether the variable income is normally
distrbuted
### by visual inspection of the histogram
### we will also add the normal distribution curve to the
chart

### load the graphics package

require(ggplot2)

### compute the mean and standard deviation of the income

m <- mean(demo$income)

std <- sd(demo$income)

### build the histogram chart with the normal distribution
curve

ggplot()+
  geom_histogram(data=demo, aes(x=income, y=..density..),
fill="red")+
  stat_function(fun=dnorm, args = list(mean = m, sd = std),
aes(x=demo$income))

### dnorm is a function that creates a normal distribution
### it comes with the stats package
```

Lesson 29 - Detecting the Outliers

```
demo <- read.csv("demographics.csv")

View(demo)

##########
### how to detect the outliers in a data series
### with the help of the standardized values
##########

### we will check whether the variable income presents
outliers
```

```
### compute the standardized values of the variable

zinco <- scale(demo$income, scale=TRUE)

### sort the vector, decreasing

sort(zinco, decreasing=TRUE)
```