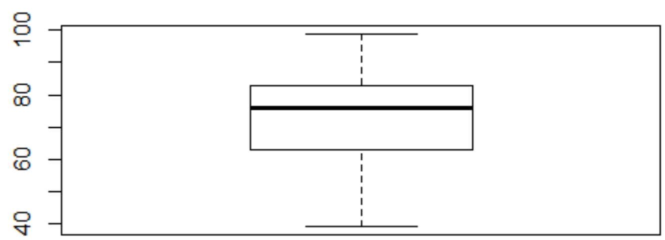
STA2023 R Labs

Lab 2

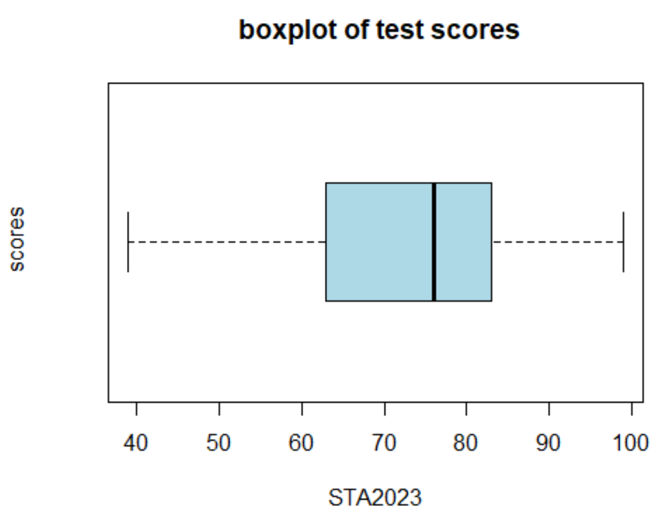
Topics: Measures of Central Tendency. Calculate and interpret the mean, median, and mode of a set of numbers. Calculate and interpret the weighted mean. Determine the inter-relationships between the mean, median, and mode for skewed and symmetrical distributions. Calculate and interpret the mean for a frequency distribution. Measures of Variation. Calculate and interpret the range, variance, and standard deviation of a set of numbers. Calculate and interpret the variance, and standard deviation for grouped data. Measures of Relative Standing. Calculate and interpret quartiles and percentiles. Calculate and interpret z-scores for a normal distribution

```
> x<-c(10,15,20,30)
> mean(x)
[1] 18.75
> sd(x)
[1] 8.539126
> var(x)
[1] 72.91667
> median(x)
[1] 17.5
> y<-c(2,3,1,4,5,7,1,3,4,9,0,4)# consider variable y
> sort(y) # it allows to identify 4 as the mode.
 [1] 0 1 1 2 3 3 4 4 4 5 7 9
> range(y)
[1] 0 9
> #therefore, the range is 9-0 = 9
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.750   3.500   3.583   4.250   9.000
> quantile(y)
  0%  25%  50%  75% 100%
0.00 1.75 3.50 4.25 9.00
> quantile(y, prob = c(0.15, 0.20, 0.35))
 15%  20%  35%
1.00 1.20 2.85


> x<-c(10,15,20,30) # same variable we have on top. Do not input again.
> frq<-c(2,3,4,5)# let's consider a weight for each data point in x
> data.frame(x, frq) # to get a picture of it.
   x frq
1 10   2
2 15   3
3 20   4
4 30   5
> install.packages("Weighted.Desc.Stat") # using package Weighted.Desc.Stat:
weighted descriptive statistics
> require(Weighted.Desc.Stat)
> w.mean(x,frq)
[1] 21.07143
> sigma<-w.sd(x,frq) # w.sd retrieve the population sd
> sigma
[1] 7.362689
> n<-sum(frq)
> s<-sigma*sqrt(n/(n-1)) # in order to obtain the sample standard deviation:
> s
[1] 7.640623
```

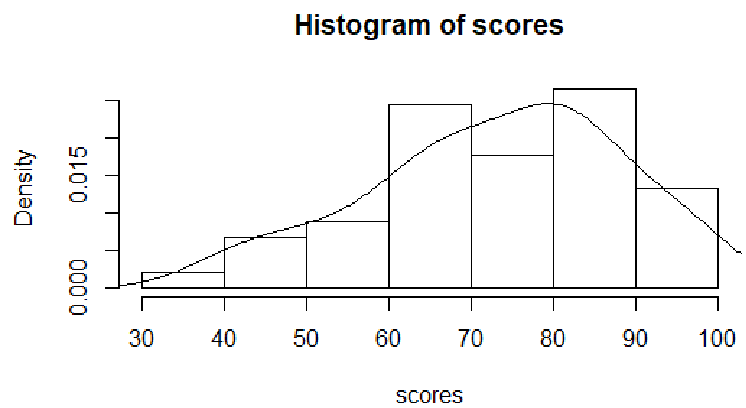
```
> scores<-c(85,77,86, 81, 42, 53, 87, 69, 92, 75, 46, 46, 51,76, 61, 99,55, 8
3, 90, 95, 71, 96,67, 39, 92,63, 81, 97, 76, 59, 81, 65, 81, 71, 83, 62, 69,
79,61, 82, 67, 68, 79,81, 69) # test scores

> summary(scores)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  39.00   63.00   76.00   73.07   83.00   99.00 

> quantile(scores, prob=c(0.1,0.20,0.70,0.9))
 10%  20%  70%  90% 
51.8 61.0 81.0 92.0 

> boxplot(scores)# "plain" boxplot.
```



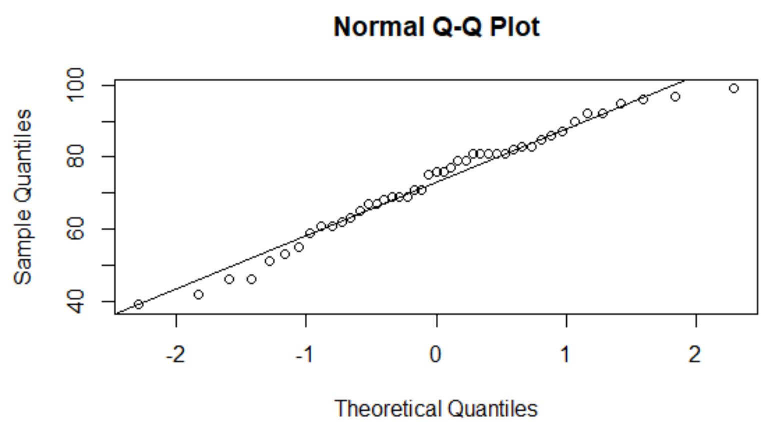

```
>boxplot(scores, main="boxplot of test scores",xlab="STA2023", ylab="scores",
col="lightblue", horizontal = T) # same boxplot, adding labels, horizontal.
```




```
> hist(scores, prob=T)# histogram.
> lines(density(scores)) # adding the density curve to the histogram.
```




```
> qqnorm(scores);qqline(scores)#two functions in one prompt, separated by ;
```




```
> # finding zcores.
> m<-mean(scores)
> sd1<-sd(scores)
> zcores<-(scores-m)/sd1 # the z score formula
> Ztable1<-cbind(scores, zcores)
> head(Ztable1, 4)
     scores    zcores
[1,]     85 0.7837382
[2,]     77 0.2583271
[3,]     86 0.8494145
[4,]     81 0.5210326

> pnorm(1)-pnorm(-1);pnorm(2)-pnorm(-2);pnorm(3)-pnorm(-3)#The Empirical rule
[1] 0.6826895
[1] 0.9544997
[1] 0.9973002
> # for a population whose mean is 10, and sigma (standard deviation) is 0.7,
assuming the variable is normal distributed, then:
> pnorm(10.7, 10, 0.7)-pnorm(9.3, 10, 0.7) # prob between 1 stand dev above
the mean (10+0.7) and 1 std dev below the mean (10-0.7)
[1] 0.6826895
> pnorm(11.4, 10, 0.7)-pnorm(8.6, 10, 0.7) # prob between 2 stand dev above t
he mean (10+2*0.7) and 2 std dev below the mean (10-2*0.7)
[1] 0.9544997
>pnorm(12.1, 10, 0.7)-pnorm(7.9, 10, 0.7) # prob between 3 stand dev above
the mean (10+3*0.7) and 3 std dev below the mean (10-3*0.7)
[1] 0.9973002
```