An introductory example to the bootstrap.

Description: Results of an experiment to test whether directed reading activities in the classroom help elementary school students improve aspects of their reading ability. A treatment class of 21 third-grade students participated in these activities for eight weeks, and a control class of 23 third-graders followed the same curriculum without the activities. After the eight-week period, students in both classes took a Degree of Reading Power (DRP) test which measures the aspects of reading ability that the treatment is designed to improve.

Number of cases: 44 Variable Names:

1.Treatment: Whether student participated in activities (treated) or not (control) 2.Response: Score on Degree of Reading Power test

```
> scores
   Treatment Response
 1   Treated       24
 2   Treated       43
 3   Treated       58
 4   Treated       71
 5   Treated       43
 6   Treated       49
 7   Treated       61
 8   Treated       44
 9   Treated       67
10   Treated       49
11   Treated       53
12   Treated       56
13   Treated       59
14   Treated       52
15   Treated       62
16   Treated       54
17   Treated       57
18   Treated       33
19   Treated       46
20   Treated       43
21   Treated       57
22   Control       42
23   Control       43
24   Control       55
25   Control       26
26   Control       62
27   Control       37
28   Control       33
29   Control       41
30   Control       19
31   Control       54
32   Control       20
33   Control       85
```

```
34   Control       46
35   Control       10
36   Control       17
37   Control       60
38   Control       53
39   Control       42
40   Control       37
41   Control       42
42   Control       55
43   Control       28
44   Control       48
```

```
> treated <- scores(1:21)
> control <- scores(22:44)

> n1 <- length(control)
> n2 <- length(treated)
```

A few statistics

```
> # means

> # 1st sample
> mean(control)
[1] 41.52174
> # 2nd sample
> mean(treated)
[1] 51.47619

> # Difference
> mean(treated) - mean(control)
[1] 9.954451


> # Standard deviations

> # 1st sample
> sqrt(var(control)/n1)
[1] 3.575758
> # 2nd  sample
> sqrt(var(treated)/n2)
[1] 2.402002

> # standard deviation of the difference
> sqrt(var(control)/n1 + var(treated)/n2)
[1] 4.307628
```

Bootstrap standard estimates for the mean

```
> B<- 100

> my.boot1.thetastar <- rep(NA,B)
> for (b in 1:B)
{
bt.index <- sample(1:n1, size = n1, replace = T)
my.boot1.thetastar[b] <- mean(control[bt.index])
}

> my.boot2.thetastar <- rep(NA,B)
> for (b in 1:B)
{
bt.index <- sample(1:n2, size = n2, replace = T)
my.boot1.thetastar[b] <- mean(treated[bt.index])
}

> # 1st sample
> sqrt(var(my.boot1.thetastar))
[1] 3.677239
> # 2nd  sample
> sqrt(var(my.boot2.thetastar))
[1] 2.216425
```

Handy function

```
> # appendix: we may use the bootstrap function
> bootstrap<-
function(x, nboot, theta, ..., func = NULL)
...

e.g.

> my.boot1<-bootstrap(control,nboot=100,mean)
> my.boot2<-bootstrap(treated,nboot=100,mean)

> my.boot1<-bootstrap(control,nboot=100,median)
> my.boot2<-bootstrap(treated,nboot=100,median)


> # 1st sample
> sqrt(var(my.boot1$thetastar))
```

```
[1] 3.677239
> # 2nd  sample
> sqrt(var(my.boot2$thetastar))
[1] 2.216425
```

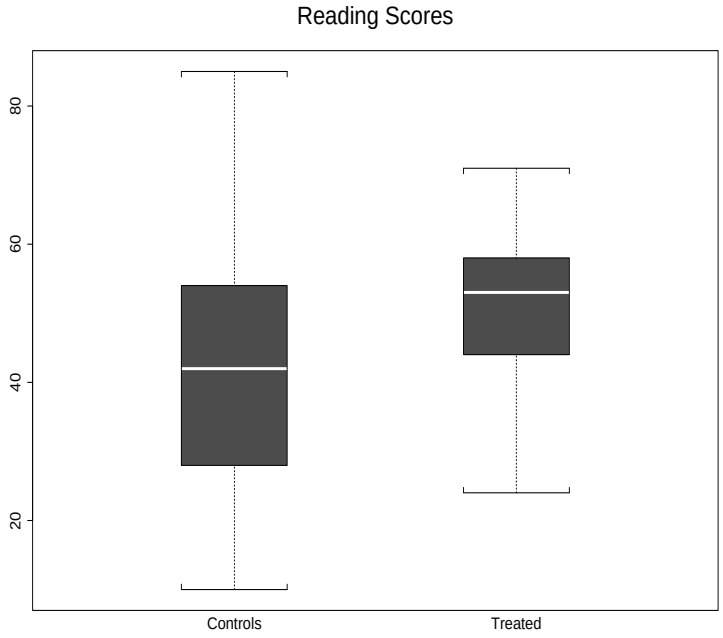

Figure 1: Boxplot

Medians

```
> median(control)
[1] 42
> median(treated)
[1] 53

> # Difference
> median(treated) - median(control)
[1] 11
```

How significant is the difference? Bootstrap estimates.

```
> # New bootstrap sample
> sqrt(var(my.boot1$thetastar))
[1] 3.460334
> sqrt(var(my.boot2$thetastar))
[1] 3.167517
```

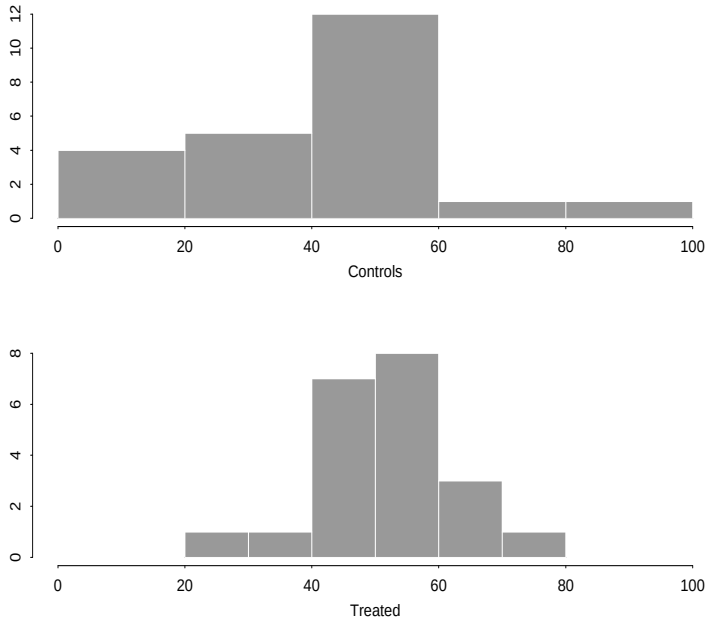

Figure 2: Comparison of the two samples

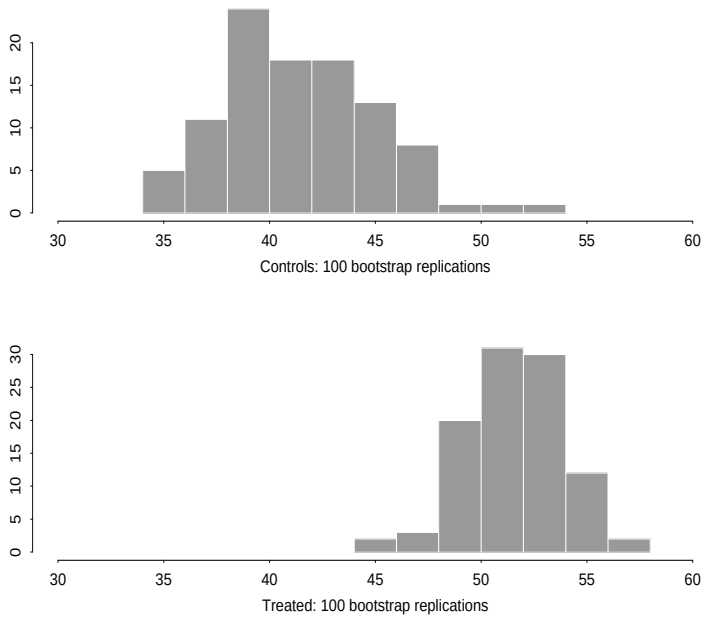

Figure 3: Bootstrap histogram

```
> sqrt(var(my.boot1$thetastar)+var(my.boot2$thetastar))
[1] 4.69117
```

Or

```
> my.boot.diff.thetastar <- (my.boot1$thetastar - my.boot2$thetastar)
> sqrt(var(my.boot.diff.thetastar))
```

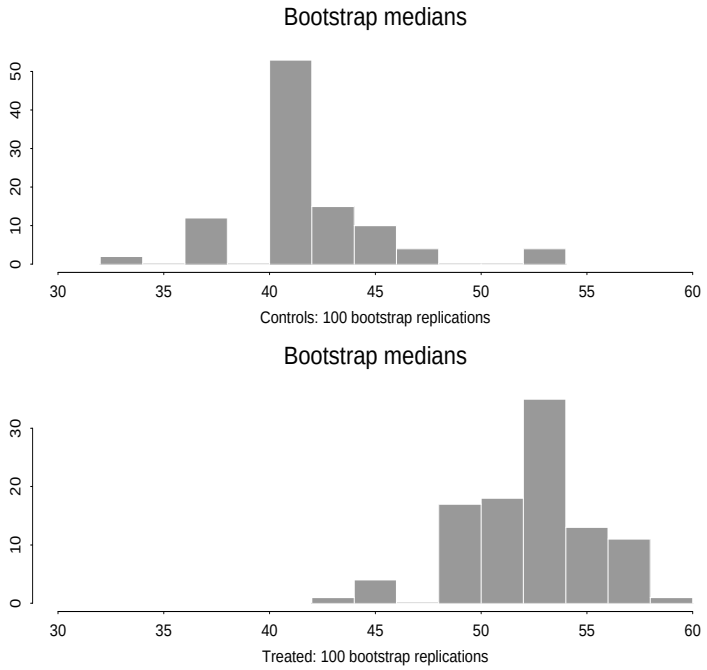

Figure 4: Boxplot