Bootstrap with R

```
> # boot1.R    Working on the bootstrap
> # Run with    R --vanilla < boot1.R > boot1.out &
> # grades.dat has 4 columns: ID, Verbal SAT, Math SAT and 1st year GPA
>
> marks <- read.table("grades.dat")
> n <- length(marks$verbal)
> n
[1] 200
> marks[1:10,]
   verbal math gpa
1     623  509 2.6
2     454  471 2.3
3     643  700 2.4
4     585  719 3.0
5     719  710 3.1
6     693  643 2.9
7     571  665 3.1
8     646  719 3.3
9     613  693 2.3
10    655  701 3.3
> obscorr <- cor(marks)
> obscorr
          verbal      math       gpa
verbal 1.0000000 0.2746341 0.3224477
math   0.2746341 1.0000000 0.1942431
gpa    0.3224477 0.1942431 1.0000000
> # Question: Is the correlation between Verbal SAT and GPA the same as
> # the correlation between math SAT and GPA?
> # What is the sampling distribution of the difference between correlation
> # coefficients?
> #
> obsdiff <- obscorr[3,1]-obscorr[3,2] # Verbal minus math
> obsdiff
[1] 0.1282046
> # The strategy will be to obtain a 95% bootstrap confidence interval for
> # the difference between verbal correlation and math correlation. This
> # confidence interval will be approximately centered around the observed
> # difference obsdiff = .128.  If the confidence interval does not include
> # zero, we will conclude that the observed difference is significantly
> # different from zero.
>
> BOOT <- 1000  ;  bsdiff <- NULL ; set.seed(9999)
> # Accumulate bootstrap values in bsdiff
> # For clarity, do operations in several separate steps inside the loop
> for(i in 1:BOOT)
+     {
+     bootmarks <- marks[sample(1:n,replace=TRUE),] # sample rows with
+                                                   # replacement
+     bcorr <- cor(bootmarks)  # Correlation matrix of bootstrap sample
+     bdiffer <- bcorr[3,1]-bcorr[3,2] # Differencce between correlation
+                                      # coefficients
+     bsdiff <- c(bsdiff,bdiffer) # Add bdiffer to the end of bsdiff
+     } # Next bootstrap sample
> bsdiff <- sort(bsdiff)
> # Now get lower and upper limits of 95% CI
```

```
> low <- bsdiff[.025*BOOT] ; up <- bsdiff[.975*BOOT + 1]
> low ; up
[1] -0.03643594
[1] 0.3032818
> (low+up)/2
[1] 0.1334230
> obsdiff
[1] 0.1282046
> write(bsdiff,"bsdiff.dat") # Maybe for later analysis
```

```
# boot1.R    Working on the bootstrap
# Run with    R --vanilla < boot1.R > boot1.out &
# grades.dat has 4 columns: ID, Verbal SAT, Math SAT and 1st year GPA

marks <- read.table("grades.dat")
n <- length(marks$verbal)
n
marks[1:10,]
obscorr <- cor(marks)
obscorr
# Question: Is the correlation between Verbal SAT and GPA the same as
# the correlation between math SAT and GPA?
# What is the sampling distribution of the difference between correlation
# coefficients?
#
obsdiff <- obscorr[3,1]-obscorr[3,2] # Verbal minus math
obsdiff
# The strategy will be to obtain a 95% bootstrap confidence interval for
# the difference between verbal correlation and math correlation. This
# confidence interval will be approximately centered around the observed
# difference obsdiff = .128.  If the confidence interval does not include
# zero, we will conclude that the observed difference is significantly
# different from zero.

BOOT <- 100  ;  bsdiff <- NULL ; set.seed(9999)
# Accumulate bootstrap values in bsdiff
# For clarity, do operations in several separate steps inside the loop
for(i in 1:BOOT)
    {
    bootmarks <- marks[sample(1:n,replace=TRUE),] # sample rows with
                                                  # replacement
    bcorr <- cor(bootmarks)  # Correlation matrix of bootstrap sample
    bdiffer <- bcorr[3,1]-bcorr[3,2] # Differencce between correlation
                                     # coefficients
    bsdiff <- c(bsdiff,bdiffer) # Add bdiffer to the end of bsdiff
    } # Next bootstrap sample
bsdiff <- sort(bsdiff)
# Now get lower and upper limits of 95% CI
low <- bsdiff[.025*BOOT] ; up <- bsdiff[.975*BOOT + 1]
low ; up
(low+up)/2
obsdiff
write(bsdiff,"bsdiff.dat") # Maybe for later analysis
pdf("bsdiff.pdf") # Send graphics output to pdf file
hist(bsdiff)
```