Resampling Methods.

Exercises.

Anatoli Iambartsev
IME-USP

Bootstrap. The use of the term bootstrap derives from the phrase to pull oneself up by ones bootstraps, widely thought to be based on one of the eighteenth century "The Surprising Adventures of Baron Munchausen" by Rudolph Erich Raspe: *The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.*

Bootstrap. [CL, p.5].

Let $T(\cdot)$ be a functional of interest, for example estimator of a parameter. We are interested in estimation of $T(F)$, where F is population distribution. Let F_n be an empirical distribution based on sample $x = (x_1, \ldots, x_n)$. Bootstrap:

1. generate a sample $x^* = (x_1^*, \ldots, x_n^*)$ with replacement from the empirical distribution F_n for the data (boostrap sample);

2. compute $T(F_n^*)$ the bootstrap estimate of $T(F)$. This is a replacement of the original sample x with a bootstrap sample x^* and the bootstrap estimate of $T(F)$ in place of the sample estimate of $T(F)$;

3. M times repeat steps 1 and 2 where M is large, say 100000.

Bootstrap. [CL, p.5].

Now a very important thing to remember is that with the Monte Carlo approximation to the bootstrap, there are two sources of error:

1. the Monte Carlo approximation to the bootstrap distribution, which can be made as small as you like by making M large;
2. the approximation of the bootstrap distribution F_n^* to the population distribution F.

If $T(F_n^*)$ converges to $T(F)$ as $n \to \infty$, then bootstrapping works.

Bootstrap. [CL, p.5].

"If $T(F_n^*)$ converges to $T(F)$ as $n \to \infty$, then bootstrapping works. It is nice that this works out often, but it is not guaranteed. We know by a theorem called the Glivenko-Cantelli theorem that F_n converges to F uniformly. Often, we know that the sample estimate is consistent (as is the case for the sample mean). So, (1) $T(F_n)$ converges to $T(F)$ as $n \to \infty$. But this is dependent on smoothness conditions on the functional T. So we also need (2) $T(F_n^*) - T(F_n)$ to tend to 0 as $n \to \infty$. In proving that bootstrapping works (i.e., the bootstrap estimate is consistent for the population parameter), probability theorists needed to verify (1) and (2). One approach that is commonly used is by verifying that smoothness conditions are satisfied for expansions like the Edgeworth and Cornish-Fisher expansions. Then, these expansions are used to prove the limit theorems."

Bootstrap. [CL, p.24].

“One function in the basic R packages that lies at the heart of resampling is the `sample()` function, whose syntax is

```
sample(x, size, replace = FALSE, prob = NULL)
```

The first argument `x` is the vector of data, that is, the original sample. `size` is the size of the resample desired. `replace` is `TRUE` if resampling is with replacement, and `FALSE` if not (the default). `prob` is a vector of probability weights if the equalweight default is not used. Any arguments omitted will assume the default. If `size` is omitted, it will default to the length of `x`.”

Bootstrap. [CL, p.24-25].

“For our purposes, it will usually be easiest to resample the indices of the data from a sample of size n, rather than the data itself. For example, if we have five data in our set, say

```
> x=c(-0.3, 0.5, 2.6, 1.0, -0.9)

> x

[1] -0.3 0.5 2.6 1.0 -0.9
```

then

```
> i = sample(1:5, 5, replace=TRUE)

> i

[1] 3 2 3 2 2

> x[i]

[1] 2.6 0.5 2.6 0.5 0.5
```

is the resample of the original data.”

Bootstrap standard error.

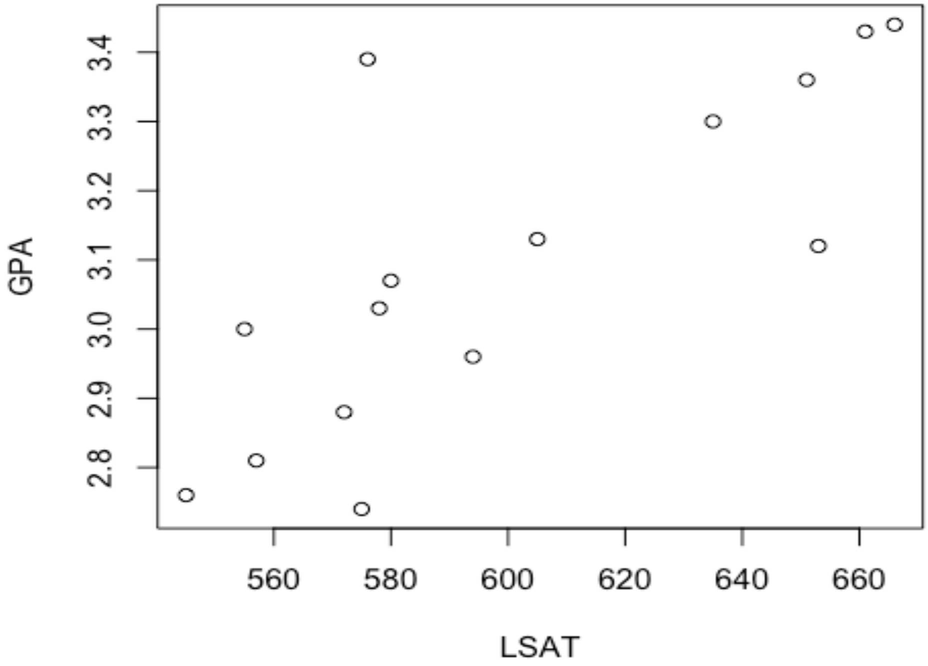
From bootstrap sampling we can estimate any aspect of the distribution of $\widehat{\theta} = s(y)$ (which is any quantity computed from the data $y = (y_1, \ldots, y_n)$, for example its standard error is

$$s.e.b.(\widehat{\theta}) = \Big(\frac{1}{B-1}\sum_{b=1}^{B}\big(\widehat{\theta}^*(b) - \widehat{\theta}^*(\cdot)\big)^2\Big)^{1/2}$$

where $\widehat{\theta}^*(b)$ is the bootstrap replication of $s(y)$ and

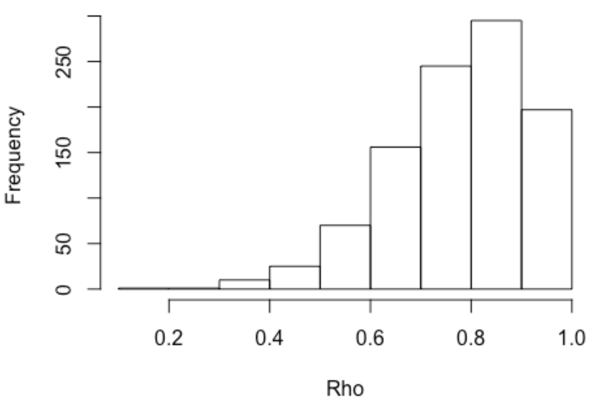
$$\widehat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^*(b).$$

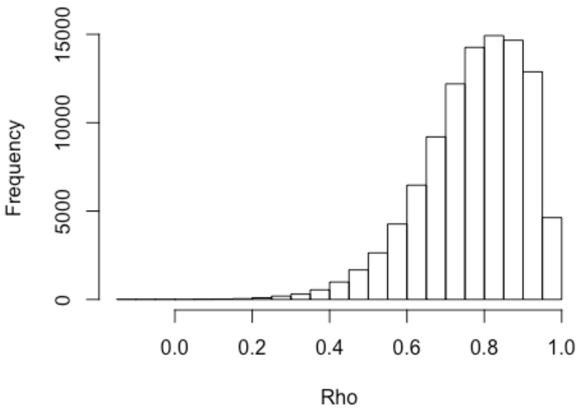
Example [EG]. The 15 points represent various entering classes at American law schools in 1973. On x-axis the average average LSAT score of entering students at school i, on y-axis undergraduate GPA score of entering students at school i.



Example [EG].

We want to attach a nonparametric (bootstrap) estimate of standard error to observed Pearson coefficient for these $n = 15$ pairs, which is $\widehat{\rho} = 0.777$. Let $B_1 = 1000(B_2 = 100000)$, the number of bootstrap replications.



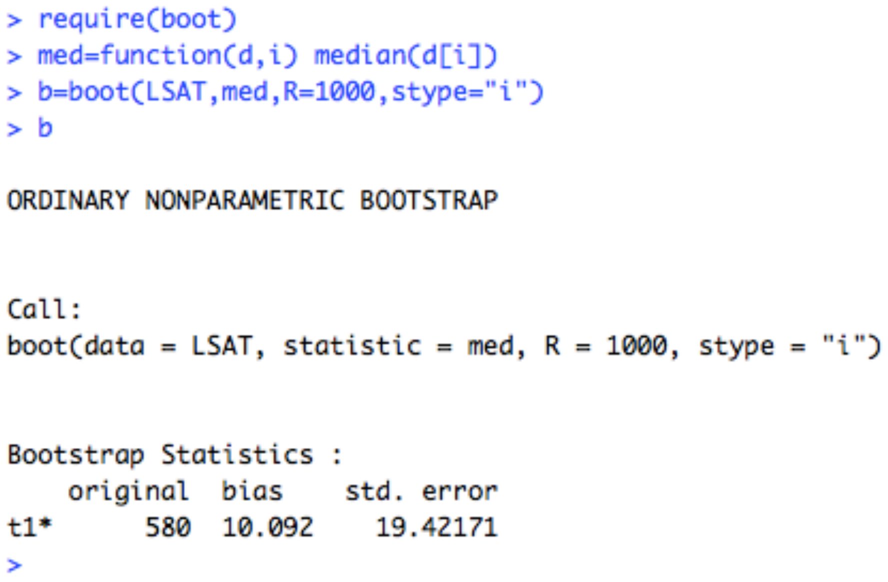


The standard errors are $\widehat{\sigma}_{B_1} = 0.135$ and $\widehat{\sigma}_{B_2} = 0.133$ correspondingly. When $\widehat{\sigma}_{Norm} = \frac{1-\widehat{\rho}^2}{\sqrt{n-3}} = 0.110$.

Example.

[EG]: "One thing is obvious about the bootstrap procedure: it can be applied just as well to any statistic, simple or complicated, as to the correlation coefficient"

Assume we want calculate the standard error for the median of LSAT. Use bootstrap:



```
> require(boot)
> med=function(d,i) median(d[i])
> b=boot(LSAT,med,R=1000,stype="i")
> b

ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = LSAT, statistic = med, R = 1000, stype = "i")


Bootstrap Statistics :
    original  bias    std. error
t1*      580  10.092    19.42171
>
```

Bootstrap bias-reduction.

Let $\widehat{\theta}$ be a consistent estimator, but biased. Target: to reduce the bias of the estimator.

The bias of $\widehat{\theta}$ is the systematic error $bias = \mathbb{E}_F\widehat{\theta} - \theta$. Em general the bias depends on the unknown parameter θ, because why we cannot to have $\widehat{\theta} - bias$.

Consider the following bootstrap bias correction

$$\widehat{\theta}_B = \widehat{\theta} - bi\widehat{a}s.$$

where

$$bi\widehat{a}s = \widehat{\mathbb{E}_F}\widehat{\theta} - \widehat{\theta} = \widehat{\theta}^*_{(\cdot)} - \widehat{\theta},$$

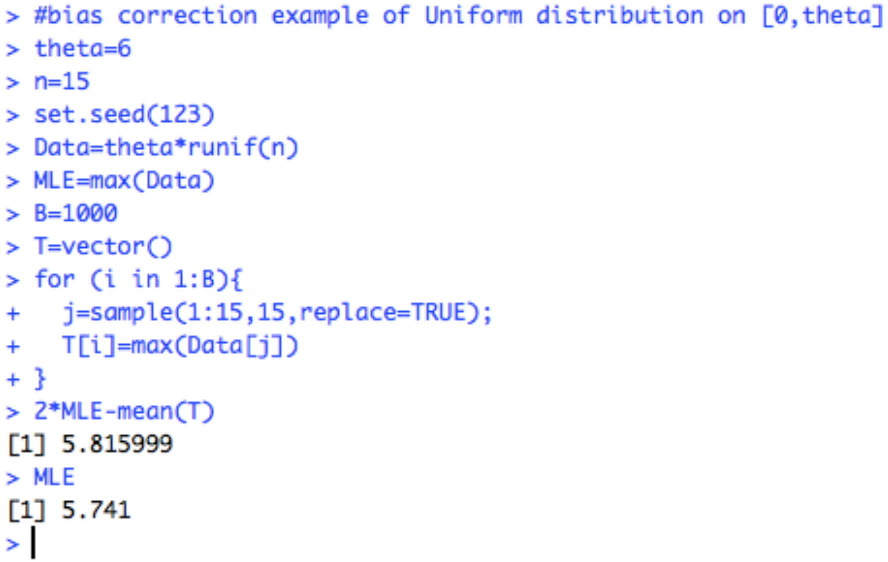
where $\widehat{\theta}^*_{(\cdot)}$ is the average of bootstrap estimators, i.e.

$$\widehat{\theta}^*_{(\cdot)} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^*_b.$$

Thus

$$\widehat{\theta}_B = \widehat{\theta} - bi\widehat{a}s = 2\widehat{\theta} - \widehat{\theta}^*_{(\cdot)}$$

Bootstrap bias-reduction. Example.



```
> #bias correction example of Uniform distribution on [0,theta]
> theta=6
> n=15
> set.seed(123)
> Data=theta*runif(n)
> MLE=max(Data)
> B=1000
> T=vector()
> for (i in 1:B){
+   j=sample(1:15,15,replace=TRUE);
+   T[i]=max(Data[j])
+ }
> 2*MLE-mean(T)
[1] 5.815999
> MLE
[1] 5.741
> 
```

Jackknife.

In some sense the bootstrap method is a generalization of the method jackknife, in the sense that the resampling is made randomly and not deterministically as in jackknife “leave-one-out”.

Jackknife.

1. We have a sample $y = (y_1, \dots, y_n)$ and estimator $\widehat{\theta} = s(y)$.

2. Target: estimate the bias and standard error of the estimator.

3. The *leave-one-out* observation samples
$$y_{(i)} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n),$$
for $i = 1, \dots, n$ are called *jackknife samples*.

4. Jackknife estimators are $\widehat{\theta}_{(i)} = s(y_{(i)})$.

Jackknife bias-reduction. Quenouille bias.

The bias of $\widehat{\theta} = s(y)$ is defined as

$$bias_J(\widehat{\theta}) = (n-1)\big(\widehat{\theta}_{(\cdot)} - \widehat{\theta}\big),$$

where $\widehat{\theta}_{(\cdot)}$ is the average of Jackknife estimators $\widehat{\theta}_{(i)}$

$$\widehat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\theta}_{(i)}.$$

This leads to a bias-reduced *jackknife estimator* of parameter θ

$$\widehat{\theta}_J = \widehat{\theta} - bias_J(\widehat{\theta}) = n\widehat{\theta} - (n-1)\widehat{\theta}_{(\cdot)}$$

Jackknife bias-reduction. Quenouille bias.

```
> theta=6

> n=15

> set.seed(123)

> Data=theta*runif(n)

> Data

 [1] 1.7254651 4.7298308 2.4538615 5.2981044 5.6428037 0.2733390
3.1686329 5.3545143 3.3086101 2.7396884

[11] 5.7410001 2.7200049 4.0654238 3.4358004 0.6175481
```

The maximal value is 5.7410001 and the second maximal value is 5.6428037.

Jackknife bias-reduction. Quenouille bias.

The maximal value is 5.7410001 and the second maximal value is 5.6428037.

The average of Jackknife estimators $\widehat{\theta}_{(i)}$

$$\widehat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{(i)} = \frac{5.6428037 + 14 \cdot 5.7410001}{15} = 5.734454.$$

The bias-reduced *jackknife estimator* of parameter θ

$$\begin{aligned} \widehat{\theta}_J &= n\widehat{\theta} - (n-1)\widehat{\theta}_{(\cdot)} \\ &= 15 \cdot 5.7410001 - 14 \cdot 5.734454 \ = \ 5.832645. \end{aligned}$$

The bias-reduced *bootstrap estimator* of parameter θ was 5.815999.

Bootstrap hypotheses testing.

- Set the two hypotheses.
- Choose a test statistic T that can discriminate between the two hypotheses. We do not care that our statistic has a known distribution under the null hypothesis.
- Calculate the observed value t_{obs} of the statistic for the sample.
- Generate B samples from the distribution implied by the null hypothesis.
- For each sample calculate the value $t_{(i)}$ of the statistic, $i = 1, \dots, B$.
- Find the proportion of times the sampled values are more extreme than the observed.
- Accept or reject according to the significance level.

Bootstrap hypotheses testing.

Suppose two samples $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_m)$. We wish to test the hypothesis that the mean of two populations are equal, i.e.

$$H: \; \mu_x = \mu_y \quad \text{vs} \quad A: \; \mu_x \neq \mu_y$$

Use as a test statistic $T = \bar{x} - \bar{y}$.

Under the null hypothesis a good estimate of the population distribution is the combined sample $z = (x_1, \dots, x_n, y_1, \dots, y_m)$

For each of the bootstrap sample calculate $T^*_{(i)}$, $i = 1, \dots, B$.

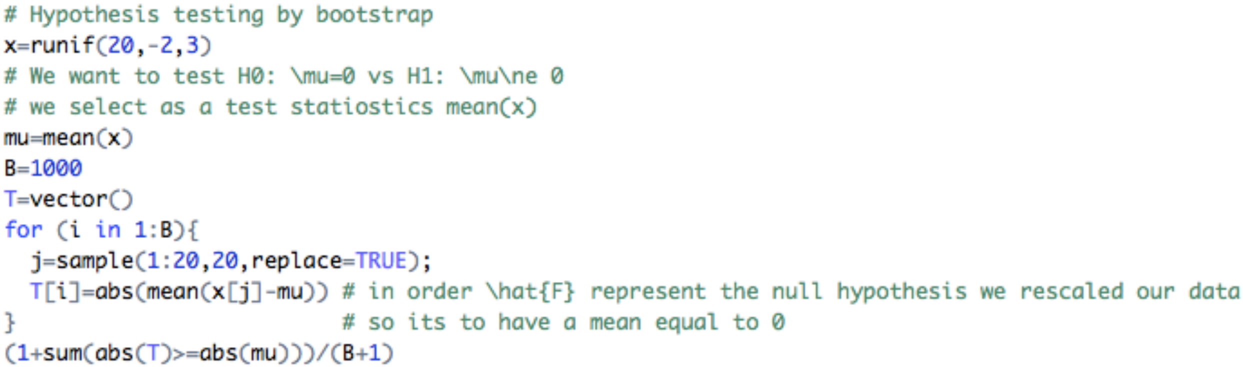
Estimate the p-value of the test as

$$\widehat{p} = \frac{1}{B} \sum_{i=1}^{B} \mathbb{1}(T^*_{(i)} \geq t_{obs}) \text{ or } \tilde{p} = \frac{1}{B+1} \left(1 + \sum_{i=1}^{B} \mathbb{1}(T^*_{(i)} \geq t_{obs}) \right).$$

Other test statistics are applicable, as for example t-statistics.

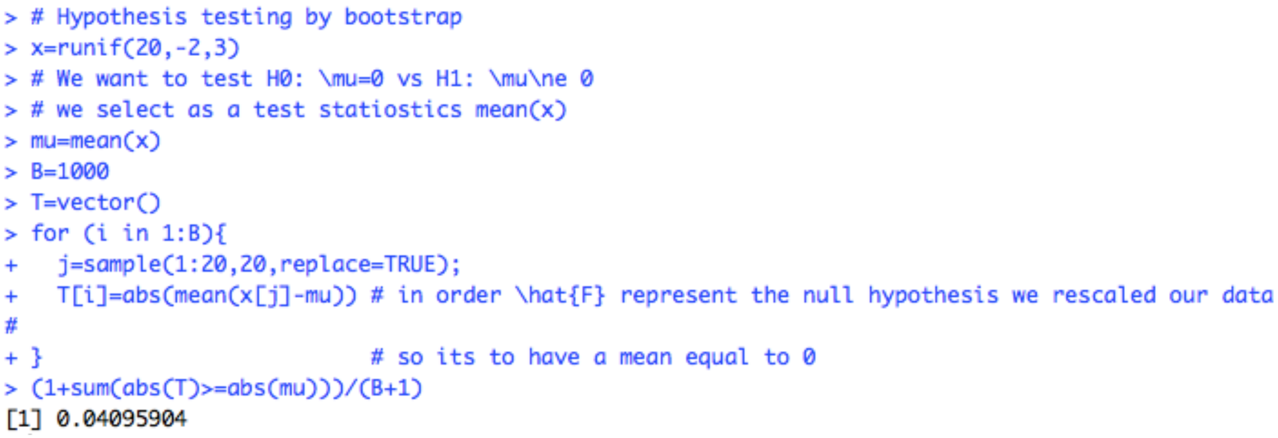
Bootstrap hypotheses testing. One-sample problem.

We want to test $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. What is the appropriate way to estimate the null distribution? The empirical distribution $\widehat{F}$ is not an appropriate estimation, because it does not obey H_0. We can use the empirical distribution of the points: $x_i = x_i - \bar{x} + \mu_0, i = 1, \ldots, n$. Which has a mean of μ_0.



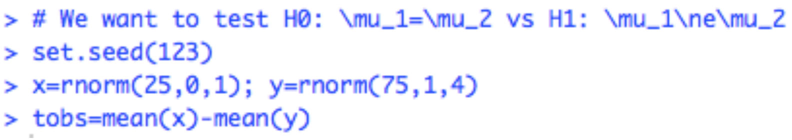
```
# Hypothesis testing by bootstrap
x=runif(20,-2,3)
# We want to test H0: \mu=0 vs H1: \mu\ne 0
# we select as a test statiostics mean(x)
mu=mean(x)
B=1000
T=vector()
for (i in 1:B){
  j=sample(1:20,20,replace=TRUE);
  T[i]=abs(mean(x[j]-mu)) # in order \hat{F} represent the null hypothesis we rescaled our data
}                         # so its to have a mean equal to 0
(1+sum(abs(T)>=abs(mu)))/(B+1)
```

Bootstrap hypotheses testing. One-sample problem.

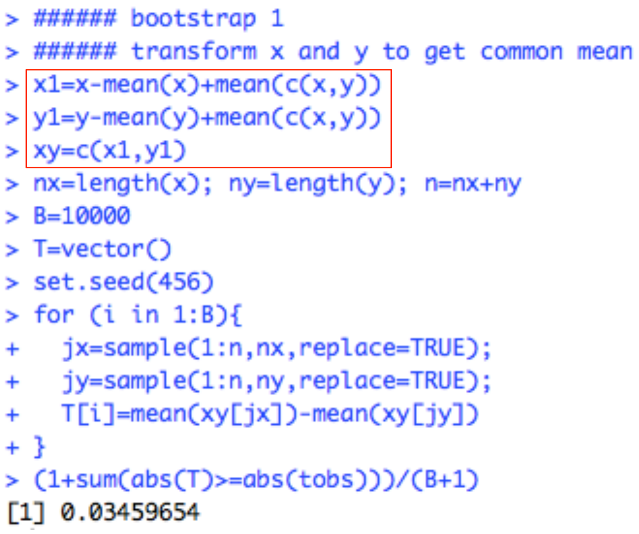


```
> # Hypothesis testing by bootstrap
> x=runif(20,-2,3)
> # We want to test H0: \mu=0 vs H1: \mu\ne 0
> # we select as a test statiostics mean(x)
> mu=mean(x)
> B=1000
> T=vector()
> for (i in 1:B){
+   j=sample(1:20,20,replace=TRUE);
+   T[i]=abs(mean(x[j]-mu)) # in order \hat{F} represent the null hypothesis we rescaled our data
#
+ }                         # so its to have a mean equal to 0
> (1+sum(abs(T)>=abs(mu)))/(B+1)
[1] 0.04095904
```

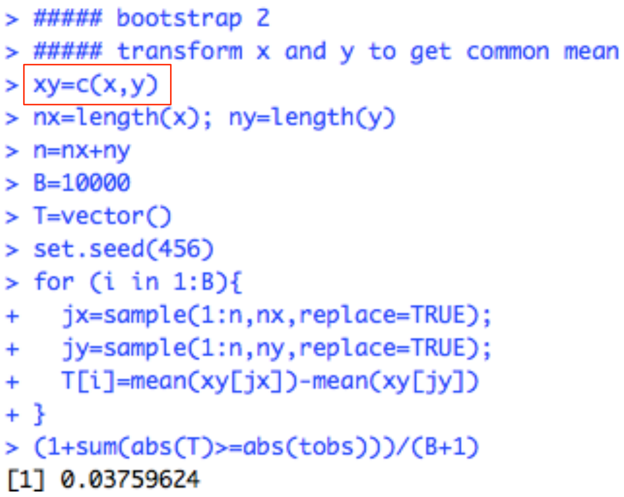
Bootstrap hypotheses testing. Two-sample problem.



```
> # We want to test H0: \mu_1=\mu_2 vs H1: \mu_1\ne\mu_2
> set.seed(123)
> x=rnorm(25,0,1); y=rnorm(75,1,4)
> tobs=mean(x)-mean(y)
```

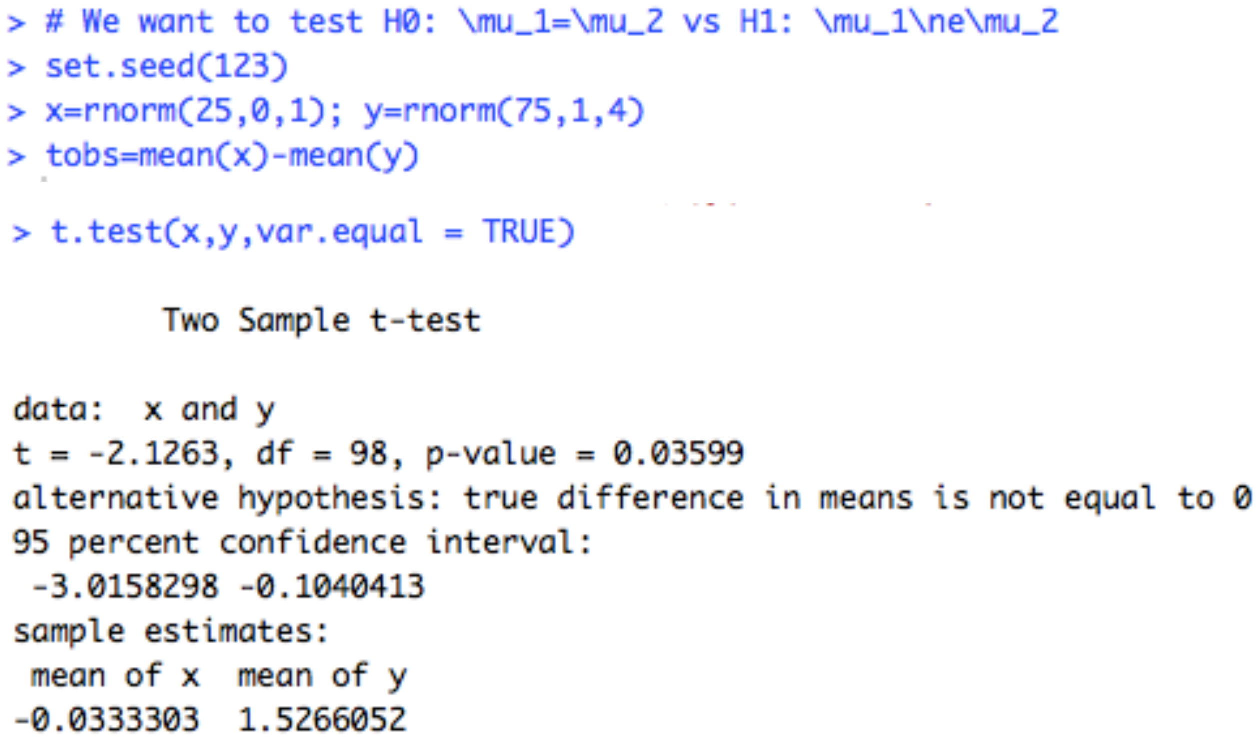


```
> ###### bootstrap 1
> ###### transform x and y to get common mean
> x1=x-mean(x)+mean(c(x,y))
> y1=y-mean(y)+mean(c(x,y))
> xy=c(x1,y1)
> nx=length(x); ny=length(y); n=nx+ny
> B=10000
> T=vector()
> set.seed(456)
> for (i in 1:B){
+   jx=sample(1:n,nx,replace=TRUE);
+   jy=sample(1:n,ny,replace=TRUE);
+   T[i]=mean(xy[jx])-mean(xy[jy])
+ }
> (1+sum(abs(T)>=abs(tobs)))/(B+1)
[1] 0.03459654
```



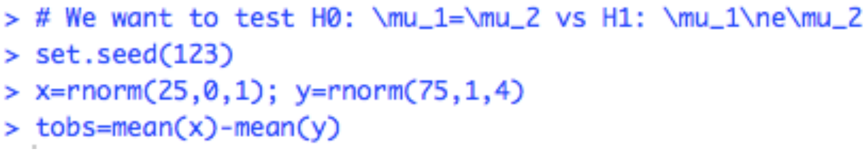
```
> ##### bootstrap 2
> ##### transform x and y to get common mean
> xy=c(x,y)
> nx=length(x); ny=length(y)
> n=nx+ny
> B=10000
> T=vector()
> set.seed(456)
> for (i in 1:B){
+   jx=sample(1:n,nx,replace=TRUE);
+   jy=sample(1:n,ny,replace=TRUE);
+   T[i]=mean(xy[jx])-mean(xy[jy])
+ }
> (1+sum(abs(T)>=abs(tobs)))/(B+1)
[1] 0.03759624
```

Bootstrap hypotheses testing. Two-sample problem.

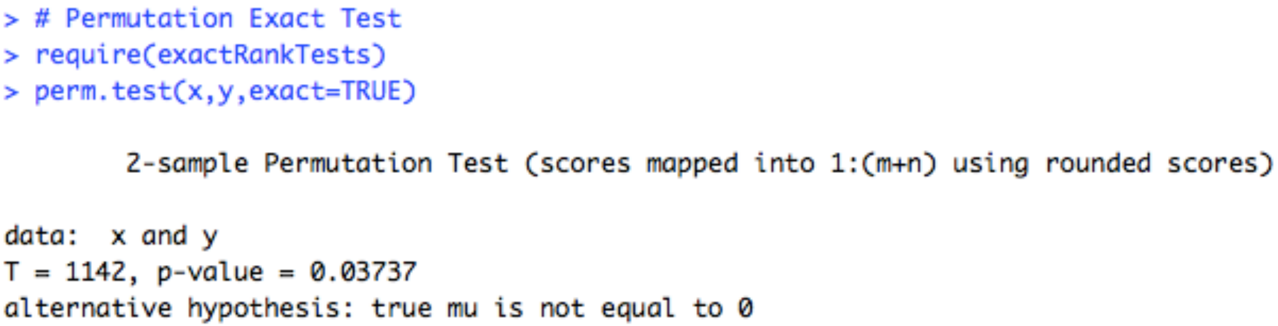


```
> # We want to test H0: \mu_1=\mu_2 vs H1: \mu_1\ne\mu_2
> set.seed(123)
> x=rnorm(25,0,1); y=rnorm(75,1,4)
> tobs=mean(x)-mean(y)

> t.test(x,y,var.equal = TRUE)

	Two Sample t-test

data:  x and y
t = -2.1263, df = 98, p-value = 0.03599
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0158298 -0.1040413
sample estimates:
 mean of x  mean of y 
-0.0333303  1.5266052 
```

Bootstrap hypotheses testing. Two-sample problem.

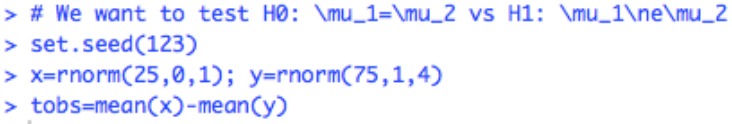


```
> # We want to test H0: \mu_1=\mu_2 vs H1: \mu_1\ne\mu_2
> set.seed(123)
> x=rnorm(25,0,1); y=rnorm(75,1,4)
> tobs=mean(x)-mean(y)
```

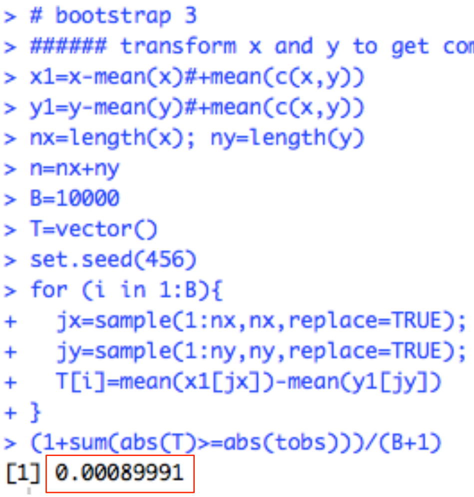


```
> # Permutation Exact Test
> require(exactRankTests)
> perm.test(x,y,exact=TRUE)

        2-sample Permutation Test (scores mapped into 1:(m+n) using rounded scores)

data:  x and y
T = 1142, p-value = 0.03737
alternative hypothesis: true mu is not equal to 0
```

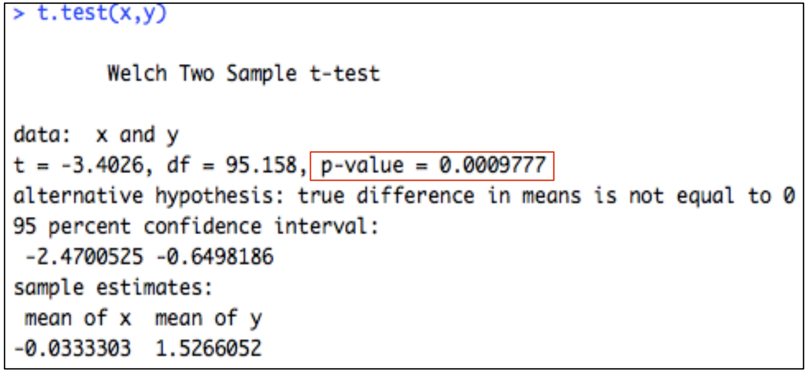
Bootstrap hypotheses testing. Two-sample problem.



```
> # We want to test H0: \mu_1=\mu_2 vs H1: \mu_1\ne\mu_2
> set.seed(123)
> x=rnorm(25,0,1); y=rnorm(75,1,4)
> tobs=mean(x)-mean(y)
```



```
> # bootstrap 3
> ###### transform x and y to get common mean
> x1=x-mean(x)#+mean(c(x,y))
> y1=y-mean(y)#+mean(c(x,y))
> nx=length(x); ny=length(y)
> n=nx+ny
> B=10000
> T=vector()
> set.seed(456)
> for (i in 1:B){
+   jx=sample(1:nx,nx,replace=TRUE);
+   jy=sample(1:ny,ny,replace=TRUE);
+   T[i]=mean(x1[jx])-mean(y1[jy])
+ }
> (1+sum(abs(T)>=abs(tobs)))/(B+1)
[1] 0.00089991
```



```
> t.test(x,y)

        Welch Two Sample t-test

data:  x and y
t = -3.4026, df = 95.158, p-value = 0.0009777
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.4700525 -0.6498186
sample estimates:
 mean of x  mean of y
-0.0333303  1.5266052
```

References.

[CL] Chernick, M. R., anf LaBudde, R. A. (2014). An introduction to bootstrap methods with applications to R. John Wiley & Sons.

[EG] Bradley Efron and Gail Gong. (1983) A Leisurely Look at he Bootstrap, the Jackknife, and Cross-Validation, The Amer. Stat. vol. 37, No. 1.

[DH] Davison, A. C. and Hinkley, D. V. (1997). Bootstrap methods and their application (Vol. 1). Cambridge university press.