The Bootstrap Methods to Test the Equality of Two Means

ILIR PALLA
LORENC EKONOMI
Department of Mathematics, Informatics and Physics
Faculty of Natural and Human Sciences
University of Korça, Albania

Abstract:

One of the most commonly applied hypothesis test procedures in applied research is the comparison of two population means. This article contains the bootstrap methods to test the equality of means of two random samples. Such a problem is called a two-sample problem. In considering a bootstrap hypothesis test for comparing the two means, there is no compelling reason to assume equal variances and hence we don't make that assumption. These methods can be used even when both random samples have not a normal distribution. Another way to test the difference of means between the two independent random variables is through the permutation test. Permutation test is useful when we do not know how to compute the distribution of a test statistic. We have given the algorithms for computation of the achieved significance level of the test, which are constructed from B. Efron and R.J. Tibshirani 1993. A way to judge the acceptance or reject the null hypothesis we use the achieved significance level (ASL), otherwise called the p-values. We used R program to estimate type I errors by simulating data of two known random samples to compare results with results that are given from bootstrap methods.

Key words: bootstrap tests, p-values, two-sample problem, null hypothesis, degrees of freedom.

INTRODUCTION

In the first part of the article is given the known cases of statistics, when random variables are independent and have a normal distribution. In the second part of the article is given the achieved significance level and bootstrap methods for hypothesis test. Finally we check for the performance of two bootstrap methods, one in case we have assumptions that the two independent random variables have the same dispersion and the second case when we do not take into consideration the assumption that random variables have equal dispersions.

We observe two independent random variables X and Y drawn from possibly different probability distributions F and G. Suppose that the observed values are $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_m)$. We want to test the hypothesis H_0 that there is no difference between their means.

1. HYPOTHESIS TESTING FOR THE DIFFERENCE OF MEANS

1.1 Hypothesis testing for the difference of means when they have normal distribution.

Suppose we have two independent random variables with normal distribution and we want to test:

$$H_0: \mu_X = \mu_Y \text{ versus } H_A: \mu_X \neq \mu_Y. \tag{1}$$

with significance level α.

Situation 1: If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, the variances are known. In this case we can use test statistic:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \tag{2}$$

when null hypothesis H_0 is true, Z has standard normal distribution. Allowed area is $]-z_{\alpha/2}, z_{\alpha/2}[$, where $\Phi(z_{\alpha/2}) = (1-\alpha)/2$, Φ the Laplace's function, α the level of significance.

Situation 2: If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, the variances are unknown but equal we can use test statistic:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} \tag{3}$$

where $s_X^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)$, $s_Y^2 = \sum_{i=1}^{m}(y_i - \bar{y})^2 / (m-1)$. When H_0 is true $T \sim S(n+m-2)$. Allowed area is $]-t_{\alpha/2}, t_{\alpha/2}[$, where $t_{\alpha/2}$ is found in table of student distribution with $\alpha/2$ with $n+m-2$ degree of freedom.

Situation 3: If X and Y are paired, then we put random variable $Z = X - Y$, where $Z_i = X_i - Y_i$, $i = 1,...,n$. Now we have $Z \square N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$ which can treated as in the case "one sample", student criterion.

Situation 4: Variances are unknown and unequal. We can use Welch criterion:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m}}} \tag{4}$$

When H_0 is true, $T \sim S(n_0)$, where $n_0 = \dfrac{\left(s_X^2/n + s_Y^2/m\right)^2}{\left(s_X^2/n\right)^2/(n-1) + \left(s_Y^2/m\right)^2/(m-1)}$. The approximation n_0 dates back to a series of papers by Welch 1938, 1947.

The achieved significance level (ASL)

In all cases treated with the above assumptions we had about random variables to come from a normal distribution. In this part we describe how bootstrap methods can be used to produce significance tests. The basic idea to test the hypothesis with the bootstrap method is to test the hypothesis, without normality assumption.

The simplest situation involves a simple null hypothesis H_0 which completely specifies the probability distribution of the data. Thus, we are dealing with a single sample $X = (X_1, X_2, ..., X_n)$ from a population with CDF (cumulative distribution function) F, then H_0 specifies that $F = F_0$, where F_0 contains no unknown parameters, for example "exponential with mean 1". The more usual situation in practice is that H_0 is a composite null hypothesis, which means that some aspects of F are not determined and remain unknown when H_0 is true. An example: "normal with mean 5", the variance of the normal distribution being unspecified.

Suppose t_{obs} is the observed value of a statistic T with cumulative distribution F_T under the null hypothesis.

The achieved significance level (ASL) or p-value of the test is defined to be the probability of observing at least that large a value when the null hypothesis is true (Bradley Efron and R.J Tibshirani 1993, 203).

$$ASL = \mathrm{P}(T \geq t_{obs} | \mathrm{H}_0) \qquad (5)$$

The quantity t_{obs} is fixed at its observed value. The random variable T has the null hypothesis distribution; the distribution of T if H_0 is true. Let we wish to perform a test at level α that reject the null hypothesis when t_{obs} is in the upper tail. Then the ASL calculate

$$ASL = 1 - F_T(t_{obs}) \qquad (6)$$

Where ASL is less than 0.1, we have borderline evidence against H_0, if ASL<0.05 reasonably strong evidence against H_0, if ASL<0.025 strong evidence against H_0, if ASL<0.01 very strong evidence against H_0.

If we are in **situation 1** the variances are known

$$ASL = \mathrm{P}(Z > \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_X^2 / n + \sigma_Y^2 / m}}) = 1 - \Phi(\frac{\bar{x} - \bar{y}}{\sqrt{\sigma_X^2 / n + \sigma_Y^2 / m}}) \qquad (7)$$

when the $\Phi(.)$ is the Laplace function, probability density function of standard normal distribution.

ASL calculation, when we know variances using R program:

ASL<-1-pnorm((mean(x)-mean(y))/sqrt(var(x)/n+var(y)/m)

The variances are known, and equal: $F = N(\mu_X, \sigma^2)$, $G = N(\mu_Y, \sigma^2)$

then $H_0 : T \sim N(0, \sigma^2(\frac{1}{n} + \frac{1}{m}))$ and ASL calculate in this way:

$$ASL = \mathrm{P}(Z > \frac{\bar{x} - \bar{y}}{\sigma\sqrt{1/n + 1/m}}) = 1 - \Phi(\frac{\bar{x} - \bar{y}}{\sigma\sqrt{1/n + 1/m}}) \qquad (8)$$

If the variances are unknown, bat equal as in situation 2, the test statistic T has Student's t-distribution with *n+m-2* df (degrees of freedom), then ASL calculate in this way:

$$ASL = \mathrm{P}(t_{df} > \frac{\bar{x} - \bar{y}}{\bar{\sigma}\sqrt{1/n + 1/m}}) \qquad (9)$$

where standard estimate for σ is $\bar{\sigma} = \sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2]/[n + m - 2]}$.

The bootstrap achieved significance level (ASL$_{boot)}$

A bootstrap hypothesis test is based on a test statistic T. To emphasize that a test statistic need not be an estimate of a parameter, we denote it here by $T(Z)$. In our example we have $T(Z) = \bar{X} - \bar{Y}$. The achieved significance level is:

$$ASL = P(T(Z) \geq t_{obs} | H_0) \tag{10}$$

The quantity t_{obs} is fixed at its observed value and the random variable Z has a distribution specified by the null hypothesis H_0. Call this distribution F_T. Bootstrap hypothesis testing uses a "plug-in" style estimate for F_0. Denote the combined sample by Z and let its empirical distribution be $\hat{F}_0$, putting probability $1/(n+m)$ on each member of Z. Under H_0, $\hat{F}_0$ provides a nonparametric estimate of the common population that gave rise to both x and y.

Algorithm 1. (B. Efron and R.j Tibshirani, 1993, p. 221).

1. Draw R samples of size $n + m$ with replacement from Z. Call the first n observations X^* and the remaining m observation Y^*.
2. Evaluate $T(Z^{*r}) = \bar{X}^* - \bar{Y}^*$, $r = 1,...,R$.
3. Approximate ASL_{boot} by

$$A\hat{S}L_{boot} = \frac{number\{T(Z^{*r}) \geq t_{obs}\}}{R} \tag{11}$$

When we wish to perform a two-tailed test, and suppose that, the distribution of T is symmetrically distribution around zero, then, bootstrap ASL calculated in this way:

$$A\hat{S}L_{boot} = \frac{number\{|T(Z^{*r})| \geq t_{obs}\}}{R} \tag{12}$$

If we are not willing to make this assumption, we can use the equal-tail bootstrap ASL (MacKinnon 2007, 4)

$$A\hat{S}L_{boot} = 2\min\left\{\frac{number\{T(Z^{*r}) \geq t_{obs}\}}{R}, \frac{number\{T(Z^{*r}) < t_{obs}\}}{R}\right\} \quad (13)$$

A similar definition for p-value has also given A.C. Davison and D.V. Hinkley 1997,

$p_{boot} = Prob(T^{*} \geq t | \hat{F}_{0})$. We can use this in nonparametric bootstrap test, p_{boot} to approximate by $p = \dfrac{1 + number\{t^{*r} \geq t\}}{R+1}$ using $t^{*1}, t^{*2}, ..., t^{*R}$ from R bootstrap samples.

More accurate testing can be obtained through the use of a studentized statistic. If we are not willing to assume that the variances in the two populations are equal, we could base the test statistic (4). With normal populations, the quantity (4) no longer has a Student's t distribution.

The equal variance assumption is attractive for the t-test. In considering a bootstrap hypothesis test for comparing the two means, there is no compelling reason to assume equal variances and hence we don't make that assumption. The algorithm (B. Efron and R.J. Tibshirani, 1993, p. 224) for computation of the bootstrap test statistic for testing equality of means:

1. Let $\hat{F}$ put equal probability on the points $\tilde{X}_{i} = X_{i} - \bar{X} + \bar{Z}$, $i = 1,...,n$, and $\hat{G}$ put equal probability on the points $\tilde{Y}_{i} = Y_{i} - \bar{Y} + \bar{Z}$, $i = 1,...,m$, where $\bar{X}$ and $\bar{Y}$ are the group means and $\bar{Z}$ is the mean of the combined sample.
2. Form R bootstrap data sets (X^{*}, Y^{*}) where X^{*} is sampled with replacement from $\tilde{X}_{1},...,\tilde{X}_{n}$ and Y^{*} is sampled with replacement from $\tilde{Y}_{1},...,\tilde{Y}_{m}$.
3. Evaluate $T(.)$ defined by (3) on each data set,

$$T(Z^{*r}) = \frac{\bar{X}^* - \bar{Y}^*}{\sqrt{s_X^{2*}/n + s_Y^{2*}/m}}, \quad r = 1, \ldots, R \tag{14}$$

4. Approximate ASL_{boot} by $\hat{ASL}_{boot} = number\{T(Z^{*r}) \geq t_{obs}\}/R$, where t_{obs} is the observed value of the statistic.

2. SIMULATIONS IN R

2.1 Bootstrap methods for hypothesis test in R program

Suppose we have two independent random samples, a random sample of size 50 is taken from $X \sim N(10,5)$ and a random sample of size 40 is taken from $Y \sim N(10,8)$. We know that difference of means is 0 and the variances are unequal in this case.

The null and alternative hypothesis:

$$H_0 : \mu_X - \mu_Y = 0 \quad \text{vs.} \quad H_a : \mu_X - \mu_Y \neq 0$$

We use the methods bootstrap in R for test the hypothesis. The commands for this simulation will be mentioned in the appendix A1.
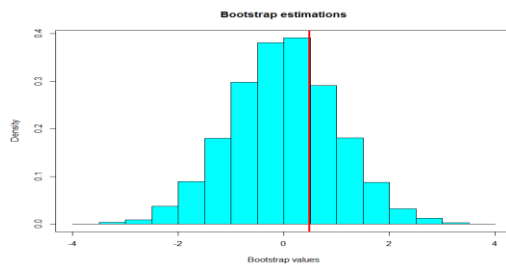
**Figure 1 Histogram of bootstrap replications of statistics T(x).
Red line is observation t-value.**

We calculated ASL_{boot} = 0.6317. This is bigger than 0.05, then we come to the conclusion that null hypothesis is true with significant level $\alpha = 0.05$, as we expected.

Permutation test gives p-value = 0.6407. This suggests that the null hypothesis is not rejected. The commands for permutation test will be mentioned in the appendix A1.

2.2 Simulations. Estimate of type I error

Type I error is the error of rejecting a null hypothesis when it is actually true.

State the random variables and the parameters in words. X- the first random variable, Y- the second variable, μ_X - mean of first random variable, μ_Y - mean of first random variable.

Null hypothesis and alternative hypothesis:

$$H_0 : \mu_X - \mu_Y = 0 \text{ vs. } H_a : \mu_X - \mu_Y \neq 0$$

H_0: true difference in means is equal to 0 versus H_a: true difference in means is not equal to 0.

We used R program for estimate of type I error and this results are shown in tables below. The commands for those simulations will be mentioned in the appendix A2.

In Table 1 and 2 the random variable are taken from normal distribution. We take 1000 Monte Carlo simulation, for any Monte Carlo simulation we find bootstrap distribution for difference of means and calculate bootstrap ASL for any of them. Several numbers for bootstrap replicate are taken, R=99, R=999 and R=9999. For more Monte Carlo simulation see Pierre Lafaye de Micheaux and Viet Anh Tran 2016.

Table1: Estimate of error type I. $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. **The variances are unknown but equal.** $MC = 1000$.

Simulation with $\mu_X = \mu_Y = 0$, $\sigma_X^2 = \sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 10$, m $= 15$			
Methods	R=99	R=999	R=9999
Two sample t-test	0.057	0.05	0.048
Bootstrap method 1	0.061	0.052	0.051
Simulation with $\mu_X = \mu_Y = 10$, $\sigma_X^2 = \sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 10$, m $= 15$			
Two sample t-test	0.049	0.056	0.042

Bootstrap method 1	0.053	0.060	0.044
Simulation with $\mu_X = \mu_Y = 10$, $\sigma_X^2 = \sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 100$, m $= 80$			
Two sample t-test	0.050	0.055	0.055
Bootstrap method 1	0.049	0.049	0.055

Table 2: Estimate of error type I. $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. **The variances are unknown but unequal.** MC $= 1000$

Simulation with $\mu_X = \mu_Y = 10$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 10$, m $= 15$			
Methods	R=99	R=999	R=9999
Welch two sample t-test	0.058	0.051	0.05
Bootstrap method 2	0.055	0.046	0.05
Simulation with $\mu_X = \mu_Y = 10$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 100$, m $= 80$			
Welch two sample t-test	0.049	0.041	0.056
Bootstrap method 2	0.052	0.041	0.055

In Table 3 the random variable are taken from normal distribution. We take 10000 Monte Carlo simulations, for any Monte Carlo simulation we find bootstrap distribution for difference of means and calculate bootstrap ASL for any of them. Several numbers for bootstrap replicate are taken, *R*=99, *R*=999 and *R*=9999.

Table 3: Estimate of error type I. $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. **The variances are unknown but unequal.** MC $= 10000$.

Simulation with $\mu_X = \mu_Y = 10$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 8$, $\alpha = 0.05$, n $= 10$, m $= 15$			
Methods	R=99	R=999	R=9999
Welch two sample t-test	0.0477	0.0479	0.0514
Bootstrap method 2	0.0472	0.0498	0.0546

In Table 4 the random variable are taken from exponential distribution, while in table 5 the random variable are taken from gamma distribution. We take 1000 Monte Carlo simulation, for any Monte Carlo simulation we find bootstrap distribution for difference of means and calculate bootstrap ASL for any of them. Several numbers for bootstrap replicate are taken, *R*=99, *R*=999 and *R*=9999.

Table 4: Estimate of type error type I. $X \sim E(\lambda)$, $Y \sim E(\lambda)$,
$MC = 1000$, $\alpha = 0.05$, $n = 10$, $m = 15$.

Methods	R=99	R=999	R=9999
Two sample t-test	0.056	0.042	0.045
Bootstrap method 1	0.054	0.049	0.047
Two sample t-test	0.044	0.047	0.043
Bootstrap method 2	0.044	0.052	0.051

Table 5: Estimate of type error type I.
$X \sim Gamma(shape = 3, scale = 2)$, $Y \sim Gamma(shape = 3, rate = 2)$,
$MC = 1000$, $\alpha = 0.05$, $n = 10$, $m = 15$.

Methods	R=99	R=999	R=9999
Two sample t-test	0.035	0.062	0.041
Bootstrap method 1	0.039	0.065	0042
Welch two sample t-test	0.056	0.045	0.053
Bootstrap method 2	0.055	0.047	0.055

CONCLUSION

The main practical difficulty with hypothesis test comes in calculating the ASL (5). In order to actually calculate the ASL, we had to either approximate the null hypothesis variance as in (9), Student's method, but only applied to the normal situation. In considering a bootstrap hypothesis test for comparing the two means, there is no compelling reason to assume equal variances and hence we don't make that assumption. Bootstrap methods performance is given in the above tables.

APPENDIX

A1. Calculate ASL with bootstrap method

```
sim.x <-rnorm(50,10,5); n.x=length(sim.x);
mean.x=mean(sim.x);var.x=var(sim.x)
sim.y <- rnorm(40,10,8); n.y <- length (sim.y); mean.y <- mean(sim.y)
var.y <- var(sim.y);
t.obs <- (mean.x-mean.y)/sqrt(var.x/n.x + var.y/n.y)
total <- c(sim.x,sim.y); mean.tot <- mean(total)
x.tilde = sim.x - mean.x + mean.tot; y.tilde = sim.y - mean.y + mean.tot;
R=9999
mean.x.star <- var.x.star <- numeric(); mean.y.star <- var.y.star <- numeric()
t.star <- numeric()
```
 # Start the bootstrap procedure
```
for (r in 1:R){
x.tilde.star <- sample(x.tilde, replace=TRUE)
mean.x.star[r] <- mean(x.tilde.star)
var.x.star[r] = var(x.tilde.star)
y.tilde.star <- sample(y.tilde, replace=TRUE)
mean.y.star[r] <- mean(y.tilde.star)
var.y.star[r] <- var(y.tilde.star)
t.star[r] <- (mean.x.star[r]-mean.y.star[r])/sqrt(var.x.star[r]/n.x +
var.y.star[r]/n.y)
}                 # finish the bootstrap procedure
```
 # Calculate the approximate ASL
```
ASL.star <- sum( abs(t.star) > abs(t.obs))/(R+1)
hist(t.star,freq = FALSE,col = "5", main="Bootstrap estimations",
xlab="Bootstrap values")
box()
abline(v=t.obs,col="red",lwd=3,lty=1); paste("ASL.star =",ASL.star)
```
##Permutation test
```
group = rep (c ("g1", "g2"), c (length (sim.x), length (sim.y)))
lengths = data.frame( lengths = c(sim.x, sim.y), group)
n.x <- length (sim.x)
n.y <- length (sim.y)
pool <- lengths$lengths
obs.diff.p <- mean (sim.x) - mean (sim.y)
 iterations <- 10000
sampl.dist.p <- NULL
for (i in 1 : iterations) {
 resample <- sample (c(1:length (pool)), length(pool))
g1.perm = pool[resample][1 : n.x]
g2.perm = pool[resample][(n.x+1) : length(pool)]
```

```
sampl.dist.p[i] = mean (g1.perm) - mean (g2.perm)
}
p.permute <- (sum (abs (sampl.dist.p) >= obs.diff.p) + 1)/ (iterations+1)
```

A.2 Testing hypothesis. Two sample t-test, Welch two sample t-test, bootstrap methods

H0: true difference in means is equal to 0. Type I error estimates.

Case 1.1: First case variables are normal. Variances are unknown but equal.

```
# Testing with two sample t-test and with bootstrap method1.
rm(list=ls())
library(bootstrap)
#We define the function as follows
diff.means <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.1 must be less than length(x)",call.=FALSE,domain=NA),
mean(x[1:n1]) - mean(x[(n1+1):length(x)]))}
alpha<-0.05 # Ha: true difference in means is not equal to 0 (two-sided)
x.var=8; x.mean=10; x.n=10; y.var=8; y.mean=10; y.n=15
decision.mc <- numeric(); decision.boot <- numeric(); p.value.boot <- numeric()
MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <- numeric()
R.ind=0
for (R in R.vec) {
R.ind = R.ind+1
for(mc in 1:MC) {
x<-rnorm(x.n,x.mean,sqrt(x.var)); y<-rnorm(y.n,y.mean,sqrt(y.var));
total=c(x,y)
decision.mc[mc]=ifelse(t.test(x,y,alternative ="two.sided", var.equal =
TRUE)$p.value < alpha,0,1)
t.obs = mean(x)-mean(y)
# Bootstrap method 1
d.boot <-bootstrap(total,nboot=R,theta=diff.means,n1 = x.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} # finish Monte Carlo cycle
error1.boot[R.ind] = 1-sum(decision.boot)/MC; error1.mc[R.ind] = 1-
sum(decision.mc)/MC
# paste ("error1.Monte Karlo:",error1.mc)
# paste ("error1.bootstrap:",error1.boot)
}
```

```
Results <- matrix(c(R.vec,error1.mc, error1.boot),nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr","Two sample t-test","Bootstrap
method"))
Results
```

Case 1.2: **Test with normality assumption. Variances are unknown and unequal.**
Testing with Welch two sample t-test and with bootstrap method 2.

```
rm(list=ls())
library(bootstrap)
#We define the function as follows
diff.means.sd <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.sample must be less than
length(x)",call.=FALSE,domain=NA),
(mean(x[1:n1]) - mean(x[(n1+1):length(x)])
)/sqrt((var((x[1:n1]))/n1+var(x[(n1+1):length(x)])/(length(x)-n1))))
}
alpha<-0.05 # Ha: true difference in means is not equal to 0 (two-sided)
x1.var=4; x1.mean=10; x1.n=10; y1.var=8; y1.mean=10; y1.n=15
decision.mc <- numeric(); decision.boot <- numeric(); p.value.boot <- numeric()
MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <- numeric()
R.ind=0
for (R in R.vec) {## 1
R.ind = R.ind+1
for(mc in 1:MC) {## 2
x1<-rnorm(x1.n,x1.mean,sqrt(x1.var)); y1<-rnorm(y1.n,y1.mean,sqrt(y1.var))
total=c(x1,y1); x<-x1-mean(x1)+mean(total); y<-y1-mean(y1)+mean(total)
decision.mc[mc]=ifelse(t.test(x1,y1,alternative ="two.sided", var.equal =
FALSE)$p.value < alpha,0,1)
t.obs = (mean(x1)-mean(y1))/sqrt(var(x1)/x1.n+var(y1)/y1.n)
# Bootstrap method 2
d.boot <-bootstrap(total,nboot=R,theta=diff.means.sd,n1 = x1.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} # finish Monte Carlo cycle
error1.mc[R.ind] = 1-sum(decision.mc)/MC; error1.boot[R.ind] = 1-
sum(decision.boot)/MC
# paste ("error1.Monte Carlo:",error1.mc)
#  paste ("error1.bootstrap:",error1.boot)
```

```
}
Results <- matrix(c(R.vec,error1.mc, error1.boot),nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr","Welch two sample t-test","Bootstrap
method"))
Results
```

Case 2.1 Difference means of two exponential distribution variables.

Two sample t-test and bootstrap method 1. Ha: true difference in means is not equal to 0.

```
rm(list=ls())
diff.means <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.sample must be less than length(x)", call.=FALSE,
domain=NA),
mean(x[1:n1]) - mean(x[(n1+1):length(x)]))}
rate.x.exp=0.5; x.exp.n=10;rate.y.exp=0.5; y.exp.n=15
decision.mc <- numeric(); decision.boot <- numeric(); p.value.boot <- numeric()
alpha<-0.05; MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <-
numeric()
R.ind=0
for (R in R.vec) { ## 1
R.ind = R.ind+1
for(mc in 1:MC) { ### start Monte Carlo cycle
x<-rexp(x.exp.n,rate = rate.x.exp); y<-rexp(y.exp.n,rate.y.exp); total=c(x,y)
decision.mc[mc]=ifelse(t.test(x,y,alternative ="two.sided", var.equal =
TRUE)$p.value < alpha,0,1)
t.obs = mean(x)-mean(y)
# Bootstrap method 1
d.boot <-bootstrap(total,nboot=R,theta=diff.means,n1 = x.exp.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} #finish  Monte Carlo cycle
error1.mc[R.ind] = 1-sum(decision.mc)/MC; error1.boot[R.ind] = 1-
sum(decision.boot)/MC
#  paste ("error1.Monte Carlo:",error1.mc); # paste
("error1.bootstrap:",error1.boot)
}
Results <- matrix(c(R.vec,error1.mc, error1.boot),nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr", "Welch two sample t-test", "Bootstrap
method"))
```

Results

Case 2.2 Difference means of two exponential distribution variables.
Two sample test and bootstrap method 2.

```
rm(list=ls())
diff.means.sd <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.sample must be less than
length(x)",call.=FALSE,domain=NA),
(mean(x[1:n1]) - mean(x[(n1+1):length(x)])
)/sqrt((var((x[1:n1]))/n1+var(x[(n1+1):length(x)])/(length(x)-n1))))
}
rate.x1.exp=0.5; x1.exp.n=10; rate.y1.exp=0.5; y1.exp.n=15; decision.mc <-
numeric()
decision.boot <- numeric(); p.value.boot <- numeric()
alpha<-0.05; MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <-
numeric()
R.ind=0
for (R in R.vec) { # 1
R.ind = R.ind+1
for(mc in 1:MC) { # start Monte Carlo cycle
x1<-rexp(x1.exp.n,rate = rate.x1.exp); y1<-rexp(y1.exp.n,rate=rate.y1.exp);
total=c(x1,y1)
x<-x1-mean(x1)+mean(total); y<-y1-mean(y1)+mean(total)
decision.mc[mc]=ifelse(t.test(x1,y1, alternative ="two.sided", var.equal =
FALSE)$p.value < alpha,0,1)
t.obs = (mean(x1)-mean(y1))/sqrt(var(x1)/x1.exp.n+var(y1)/y1.exp.n)
# Bootstrap method 2
d.boot <-bootstrap(total,nboot=R,theta=diff.means.sd,n1 = x1.exp.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} # finish Monte Carlo cycle
error1.boot[R.ind] = 1-sum(decision.boot)/MC; error1.mc[R.ind] = 1-
sum(decision.mc)/MC
# paste ("error1.Monte Carlo:",error1.mc); # paste
("error1.bootstrap:",error1.boot)
}
Results <- matrix(c(R.vec,error1.mc, error1.boot), nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr", "Welch two sample t-test", "Bootstrap
method"))
Results
```

Case 3.1 Difference means of two gamma distribution variables.

```
#Ha: true difference in means  is not equal to 0.
rm(list=ls())
diff.means <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.sample must be less than
length(x)",call.=FALSE,domain=NA),
mean(x[1:n1]) - mean(x[(n1+1):length(x)]))}
x.shape=3; x.scale=2; x.gamma.n=10; y.shape=3; y.scale=2; y.gamma.n=15
decision.mc <- numeric(); decision.boot <- numeric(); p.value.boot <- numeric()
alpha<-0.05; MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <-
numeric()
R.ind=0
for (R in R.vec) {# 1
R.ind = R.ind+1
for(mc in 1:MC) {# start Monte Carlo cycle
x<-rgamma(x.gamma.n,shape = x.shape, scale=x.scale)
y<-rgamma(y.gamma.n,shape = y.shape, scale=y.scale)
total=c(x,y)
decision.mc[mc]=ifelse(t.test(x,y,alternative ="two.sided", var.equal =
TRUE)$p.value < alpha,0,1)
t.obs = mean(x)-mean(y)
#Bootstrap method 1
d.boot <-bootstrap(total,nboot=R,theta=diff.means,n1 = x.gamma.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} ### finish Monte Carlo cycle
error1.boot[R.ind] = 1-sum(decision.boot)/MC; error1.mc[R.ind] = 1-
sum(decision.mc)/MC
# paste ("error1.Monte Carlo:",error1.mc); # paste
("error1.bootstrap:",error1.boot)
}
Results <- matrix(c(R.vec,error1.mc, error1.boot), nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr", "Welch two sample t-test", "Bootstrap
method"))
Results
```

Case 3.2: The second case, difference means of two gamma distribution variables.
Welch two sample test and bootstrap method 2.

```
rm(list=ls())
diff.means.sd <- function(x,n1) {
ifelse(length(x) <= n1,
stop("attention !! n.sample must be less than
length(x)",call.=FALSE,domain=NA),
(mean(x[1:n1]) - mean(x[(n1+1):length(x)])
)/sqrt((var((x[1:n1]))/n1+var(x[(n1+1):length(x)])/(length(x)-n1))))
}
x1.shape=3; x1.scale=2; x1.gamma.n=10; y1.shape=3; y1.scale=2;
y1.gamma.n=15
decision.mc <- numeric(); decision.boot <- numeric(); p.value.boot <- numeric()
alpha<-0.05; MC=1000; R.vec <- c(99,999,9999); error1.boot <-error1.mc <-
numeric()
R.ind=0
for (R in R.vec) { ## 1
R.ind = R.ind+1
for(mc in 1:MC) { ### start Monte Carlo cycle
x1<-rgamma(x1.gamma.n,shape = x1.shape, scale = x1.scale)
y1<-rgamma(y1.gamma.n,shape = y1.shape, scale = y1.scale)
total=c(x1,y1)
x<-x1-mean(x1)+mean(total); y<-y1-mean(y1)+mean(total)
decision.mc[mc]=ifelse(t.test(x1,y1,alternative ="two.sided", var.equal =
FALSE)$p.value < alpha,0,1)
t.obs = (mean(x1)-mean(y1))/sqrt(var(x1)/x1.gamma.n+var(y1)/y1.gamma.n)
## Bootstrap method 2
d.boot <-bootstrap(total,nboot=R,theta=diff.means.sd,n1 =
x1.gamma.n)$thetastar
p.value.boot[mc] <- sum(abs(d.boot) > abs(t.obs))/(R+1)
decision.boot[mc]=ifelse(p.value.boot[mc]<alpha,0,1)
} ### finish Monte Carlo cycle
error1.boot[R.ind] = 1-sum(decision.boot)/MC; error1.mc[R.ind] = 1-
sum(decision.mc)/MC
## paste ("error1.Monte Carlo:",error1.mc); ## paste
("error1.bootstrap:",error1.boot)
}
Results <- matrix(c(R.vec,error1.mc, error1.boot),nrow=3,byrow=TRUE)
dimnames(Results) <- list(c("boot.nr","Welch two sample t-test","Bootstrap
method"))
Results
```

REFERENCES

1. Bradley Efron and Robert J Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall / CRC Monographs on Statistics & Applied Probability 57.

2. Bradley Efron. 1987. "Better bootstrap confidence intervals." *Journal of the American Statistical Association* 82, no. 397 (March): 171-185. doi: 10.2307 / 2289144. JSTOR 2289144.

3. Davison A. C. and Hinkley D. V. 1997. *Bootstrap Methods and Their Applications*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press.

4. Hall P. 1988. "Theoretical Comparison of Bootstrap Confidence Intervals". *The Annals of Statistics* 16, no. 3 (Sep): 927-953.

5. James G MacKinnon. 2007. "Bootstrap Hypothesis Testing". *Queen's Economics Department Working Paper* no. 1127 Working Paper, no. 1127 (6), Canada.

6. Pierre Lafaye de Micheaux and Viet Anh Tran. 2016. "PoweR: The Reproducible Research Tool to Ease Monte Carlo Power Simulation Studies for Goodness-of-Fit Tests in R". Journal of Statistical Software. 69 (February), Issue 3, doi: 10.18637 / jss.v069.i03

7. Welch, B.L. 1938. The significance or difference between two means when the population variances are unequal. Biometrics. 29, No.3/4. (Feb.), pp. 350-362.

8. Welch, B.L. 1947. The generalization of 'student's' problem when several different population variances are involved. Biometrics. 34, Issue 1/2, (January), 28-35.