Bootstrapping to Compute SE of a statistic

2023-06-08

Direct way using the code we already are familiar with from Module04

Create a function to compute the statistic. Here we are using the example $T = \frac{X_1+X_2+\cdots+X_n}{n}$. You can change this part of the program to any other function as you need.

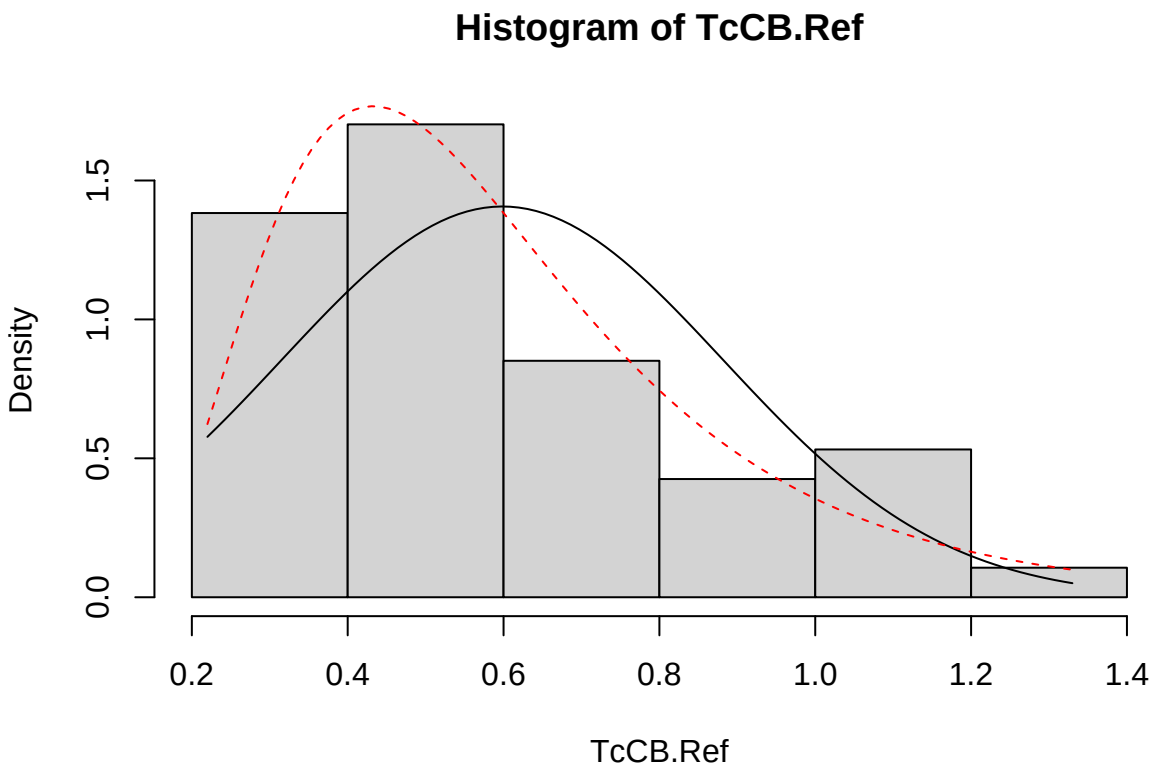
```
fun <- function(x) {
mean(x)
}
```

Below we are picking up the observations in the data set EPA.94b.tccb.df corresponding to the Reference area. It is then followed by the usual preliminary analysis of the data.

```
TcCB.Ref <- EPA.94b.tccb.df$TcCB[EPA.94b.tccb.df$Area=="Reference"]
stat.desc(TcCB.Ref,norm=T)
```

```
##      nbr.val     nbr.null       nbr.na          min          max        range
## 47.000000000  0.000000000  0.000000000  0.220000000  1.330000000  1.110000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 28.130000000  0.540000000  0.598510638  0.041373257  0.083280047  0.080452081
##      std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
##  0.283640761  0.473910977  0.845166520  1.219328422 -0.132652976 -0.097407835
##   normtest.W   normtest.p
##  0.917640846  0.002768207
```

```
hist(TcCB.Ref,prob=T)
curve(dnorm(x,mean=mean(TcCB.Ref),sd=sqrt(var(TcCB.Ref))),
      from=min(TcCB.Ref),to=max(TcCB.Ref), add=T)
curve(dlnorm(x,meanlog=mean(log(TcCB.Ref)),sdlog=sd(log(TcCB.Ref))),
      from=min(TcCB.Ref),to=max(TcCB.Ref), add=T,lty="dashed",col="red")
```



The following code demonstrates the bootstrapping method of estimating the standard error of the statistic computed by the function "fun". In this example, we are using 500 bootstrap replications.

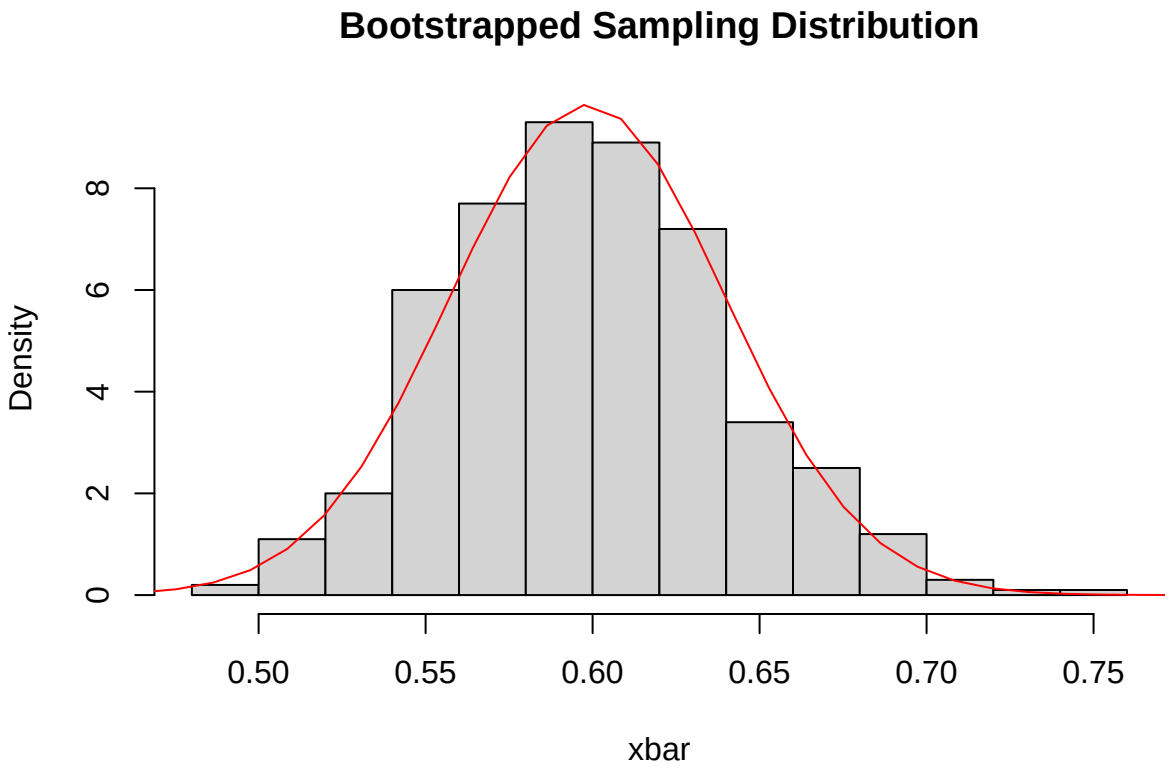
```
bootsize <- 500
sampsize <- length(TcCB.Ref)
set.seed(2)
Pop <- TcCB.Ref
Samples <- as.data.frame(matrix(sample(Pop, bootsize*sampsize, replace=TRUE),ncol=sampsize))
rownames(Samples) <- paste("BS.rep", 1:bootsize, sep="")
colnames(Samples) <- paste("obs", 1:sampsize, sep="")
results <- apply(Samples,FUN=fun,1)
#Samples
```

Now, we plot the histogram of all the bootstrapped values of the statistic T, and compare it to the Normal distribution, with $\mu = \bar{X}$ and $\frac{s}{\sqrt{n}}$.

```
stat.desc(results)
```

```
##      nbr.val     nbr.null       nbr.na          min          max        range
## 5.000000e+02 0.000000e+00 0.000000e+00 4.834043e-01 7.431915e-01 2.597872e-01
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 2.993057e+02 5.964894e-01 5.986115e-01 1.887921e-03 3.709253e-03 1.782122e-03
##      std.dev     coef.var
## 4.221519e-02 7.052185e-02
```

```
hist(results,prob=T,main="",xlab="xbar")
curve(dnorm(x,mean=mean(TcCB.Ref),sd=sd(TcCB.Ref)/sqrt(length(TcCB.Ref))),
      from = min(TcCB.Ref), to = max(TcCB.Ref),add=T,col="red")
title(main="Bootstrapped Sampling Distribution")
```



Alternative way using the package boot

```
fun <- function(x, i) {
mean(x[i])
}
```

```
results <- boot(TcCB.Ref,fun,500)
```

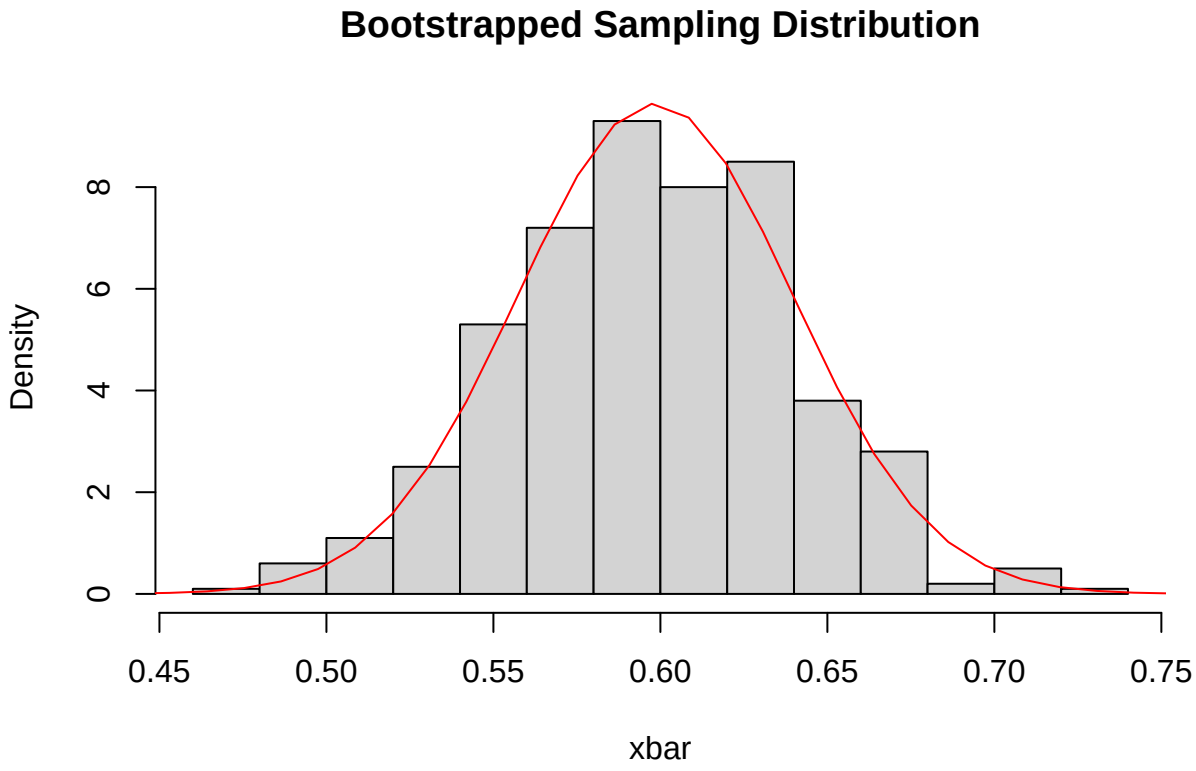
Now, we plot the histogram of all the bootstrapped values of the statistic T, and compare it to the Normal distribution, with $\mu = \bar{X}$ and $\frac{s}{\sqrt{n}}$.

```
stat.desc(results$t)
```

```
##                      V1
## nbr.val      5.000000e+02
## nbr.null     0.000000e+00
## nbr.na       0.000000e+00
```

```
## min          4.748936e-01
## max          7.285106e-01
## range        2.536170e-01
## sum          2.989530e+02
## median       5.961702e-01
## mean         5.979060e-01
## SE.mean      1.870871e-03
## CI.mean.0.95 3.675756e-03
## var          1.750079e-03
## std.dev      4.183395e-02
## coef.var     6.996744e-02
```

```
hist(results$t,prob=T,main="",xlab="xbar")
curve(dnorm(x,mean=mean(TcCB.Ref),sd=sd(TcCB.Ref)/sqrt(length(TcCB.Ref))),
      from = min(TcCB.Ref), to = max(TcCB.Ref),add=T,col="red")
title(main="Bootstrapped Sampling Distribution")
```



If all we want is the bootstapping based estimate of the standard error of the statistic (given by the function fun), it can now be done with single call to the function boot() of a package called boot.

```
boot(TcCB.Ref,fun,500)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
```

```
##
## Call:
## boot(data = TcCB.Ref, statistic = fun, R = 500)
##
##
## Bootstrap Statistics :
##      original      bias    std. error
## t1* 0.5985106 0.001894043  0.04203708
```

```
detach(EPA.94b.tccb.df)
```