Bootstrapping to Compute SE of a statistic, Example 2

2023-16-08

A generic code for bootstrapping to estimate the standard error of a statistic

Example uses RExam data set and illustrates the Bootstrap estimate for the standard error

of the sample first quartile (Q1)

RExam data is analysized in DSUR

```
Rexam <-
RExam <-
  read.delim("C:/Users/criss/Desktop/STAT414/Assignments/10-19/RExam.dat")
```

Plotting the origina data: From RExam, we are going to consider the computer scores. The objective is to estimate the population quratile (Q1) for this population and estimate its standard error by bootstrapping.

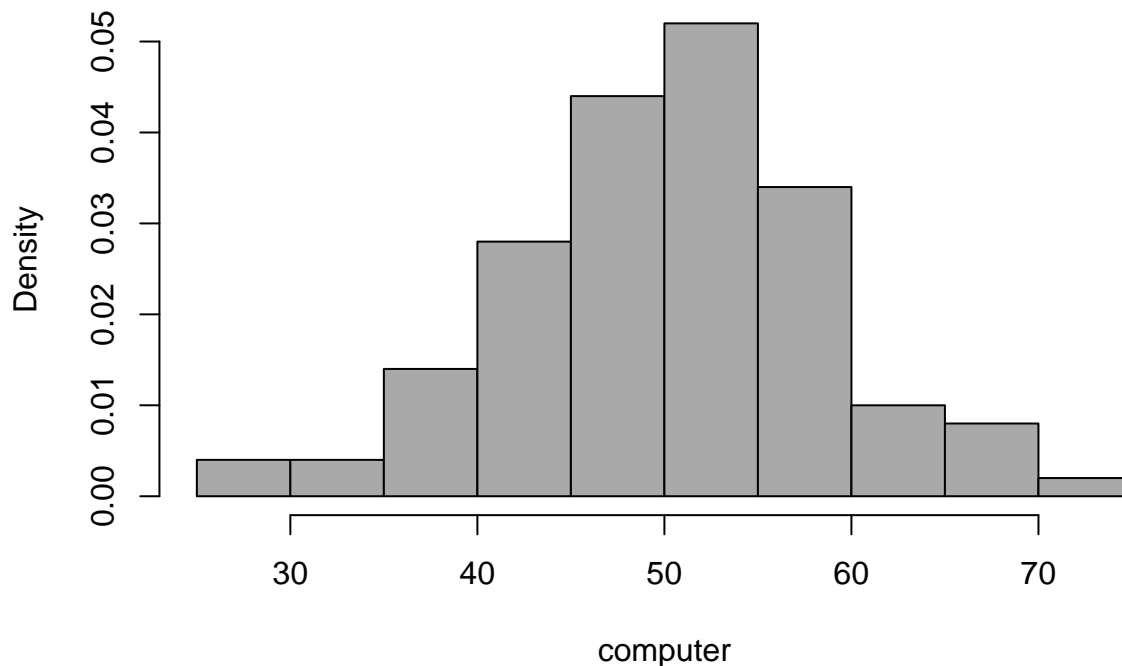## Validity of bootstrapping requires only an assumption that

* the data is collected according to a SRSWR design

No specific distributional asusmption is made about the collected data

Step 1. Estimate of the population quartile by using the sample quartile:

```
with(Rexam, hist(computer, prob=T, breaks="Sturges",
col="darkgray"))
```

Histogram of computer


```
sampQ1 <- quantile(Rexam$computer,0.25)
```

Step 2. Perform bootstrapping (draw samples from the observed data using the same sampling design as the one used to obtain the original data)

Generate 10000 Bootstrap samples of each size n=100 from Rexam$computer

Resulting BS samples are save in a matrix called results

```
BS.Samples <- as.data.frame(matrix(sample(Rexam$computer,size=10000*100, replace=TRUE),ncol=100))
rownames(BS.Samples) <- paste("BSsample", 1:10000, sep="")
colnames(BS.Samples) <- paste("obs", 1:100, sep="")
```

Create a function to compute the sample quantile.

You can use any built in R function or build your own.

```
funquant <- function(x) {
quantile(x,c(0.25))
}
```

Apply the function to each row of

the ensamble of bootstrap samples.

```
results <- apply(BS.Samples,FUN=funquant,1)
```

Compute summary statistics

Step 3. Obtain bootstap estimates of the SE and the sampling distribution of sampQ1

The standard deviation reported here is the bootstrap standard error of estimated quartile (sampQ1) The histogram reported below is the bootstrap estimate of sampling distribution of sample quartile

```
summary(results)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   41.00   44.00   45.75   45.55   46.75   50.00
```

```
sd(results)
```

```
## [1] 1.585754
```

```
hist(results,prob=T,breaks=8)
```

Histogram of results