Assignment 3: Resampling Methods

1. Consider the Boston housing dataset in the MASS library.

```
# Loading the Boston housing dataset
library(MASS)
str(Boston)
```

```
## 'data.frame':    506 obs. of  14 variables:
##  $ crim   : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
##  $ zn     : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
##  $ indus  : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
##  $ chas   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ nox    : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.5
24 ...
##  $ rm     : num  6.58 6.42 7.18 7 7.15 ...
##  $ age    : num  65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...
##  $ dis    : num  4.09 4.97 4.97 6.06 6.06 ...
##  $ rad    : int  1 2 2 3 3 3 5 5 5 5 ...
##  $ tax    : num  296 242 242 222 222 222 311 311 311 311 ...
##  $ ptratio: num  15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...
##  $ black  : num  397 397 393 395 397 ...
##  $ lstat  : num  4.98 9.14 4.03 2.94 5.33 ...
##  $ medv   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
```

(a) Provide an estimate of the population mean of the variable medv. Denote it by $\hat{\mu}$.

Solution:

```
mu.hat <- mean(Boston$medv)
sprintf("%s is %0.4f", "An estimate of the population mean of the variable me
dv", mu.hat)
```

```
## [1] "An estimate of the population mean of the variable medv is 22.5328"
```

Therefore, $\hat{\mu}$ of the variable medv is 22.5328.

(b) Estimate the standard error of $\hat{\mu}$ by the sample standard error and bootstrap method.

Solution:

The sample standard error method

```
# The sample standard error method
n <- nrow(Boston)
se <- sd(Boston$medv)/sqrt(n)
sprintf("%s is %0.7f", "The standard error of the population mean of the vari
able medv", se)
```

```
## [1] "The standard error of the population mean of the variable medv is 0.4
088611"
```

By using the sample standard error method, the standard error of $\hat{\mu}$ of the variable medv is 0.4088611.

The bootstrap method

```
# The bootstrap method
library(boot)
set.seed(1)

# Define boot function
boot.fn <- function(data, index) {
  mean(data[index])
}
# Estimate the standard error using bootstrap
se.boot.mu <- boot(data = Boston$medv, statistic = boot.fn, R = 1000)
print(se.boot.mu)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston$medv, statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##     original      bias    std. error
## t1* 22.53281 0.007650791   0.4106622
```

By using the bootstrap method, the standard error of $\hat{\mu}$ of the variable medv is 0.4106622.

Therefore, the results are very close as the standard error of the mean using the sample standard error method is 0.408861, and the standard error of the mean using the bootstrap method is 0.4106622.

However, the standard error of the mean using the bootstrap can be different if we do not set a command of set.seed(1) and change the number of bootstrap replicates (R) to different numbers than 1,000.

(c) Provide an estimate for the median of the variable medv and its standard error using the bootstrap method.

Solution:

An estimate for the median of the variable medv

```
# Estimate for the median of the variable medv
med.hat <- median(Boston$medv)
```

```
sprintf("%s is %0.2f", "An estimate of the median of the variable medv", med.
hat)

## [1] "An estimate of the median of the variable medv is 21.20"
```

From the result, the estimated median of the variable medv is 21.2.

Standard error of the estimate for the median of the variable medv using the bootstrap method.

```
set.seed(1)
# Define boot function
boot.fn <- function(data, index) {
  median(data[index])
}
# Estimate the standard error using bootstrap
se.boot.med <- boot(data = Boston$medv, statistic = boot.fn, R = 1000)
print(se.boot.med)

##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston$medv, statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*     21.2 0.02295   0.3778075
```

By using the bootstrap method, the standard error of the estimated median of the variable medv is 0.3778075, which is smaller than the standard error of the estimated mean of the variable medv.

However, the standard error of the mean using the bootstrap can be different if we do not set a command of set.seed(1) and change the number of bootstrap replicates (R) to different numbers than 1,000.

2. Repeat the above question using the log-returns in your second homework. Comment on your findings.

```
# Load data
AAPL <- getSymbols("AAPL",src = 'yahoo', from = '2012-01-01',to = "2022-02-11
AAPL$Log.return <- diff(log(AAPL$AAPL.Close))
AAPL <- AAPL[-1]
head(AAPL$Log.return)
```

```
##              Log.return
## 2012-01-04  0.005359694
## 2012-01-05  0.011040828
## 2012-01-06  0.010399504
## 2012-01-09 -0.001587396
## 2012-01-10  0.003574057
## 2012-01-11 -0.001631621
```

(a) Provide an estimate of the population mean of the log returns. Denote it by $\hat{\mu}$.

Solution:

```
mu.hat <- mean(AAPL$Log.return)
sprintf("%s is %0.7f", "An estimate of the population mean of the log returns
", mu.hat)

## [1] "An estimate of the population mean of the log returns is 0.0009675"
```

Therefore, $\hat{\mu}$ of the log return is 0.0009675.

Because the log returns include both positive and negative returns, the population mean of the log returns is very small.

(b) Estimate the standard error of $\hat{\mu}$ by the sample standard error and bootstrap method.

Solution:

The sample standard error method

```
# The sample standard error method
n <- nrow(AAPL)
se <- sd(AAPL$Log.return)/sqrt(n)
sprintf("%s is %0.7f", "The standard error of the population mean of the log
returns", se)

## [1] "The standard error of the population mean of the log returns is 0.000
3552"
```

By using the sample standard error method, the standard error of $\hat{\mu}$ of the log returns is 0.0003552.

The bootstrap method

```
# The bootstrap method
library(boot)
set.seed(1)

# Define boot function
boot.fn <- function(data, index) {
  mean(data[index])
}
# Estimate the standard error using bootstrap
```

```
se.boot.mu <- boot(data = AAPL$Log.return, statistic = boot.fn, R = 1000)
print(se.boot.mu)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = AAPL$Log.return, statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##       original        bias     std. error
## t1* 0.00096747 -1.672345e-05 0.0003440686
```

By using the bootstrap method, the standard error of $\hat{\mu}$ of the log returns is 0.0003440686.

Therefore, there is a small difference between these two methods as the standard error of the mean using the sample standard error method is 0.0003552, and the standard error of the mean using the bootstrap method is 0.0003441.

(c) Provide an estimate for the median of the log returns and its standard error using the bootstrap method.

Solution:

An estimate for the median of the log returns

```
# Estimate for the median of the log returns
med.hat <- median(AAPL$Log.return)
sprintf("%s is %0.7f", "An estimate of the median of the log returns", med.ha
t)
```

```
## [1] "An estimate of the median of the log returns is 0.0007606"
```

From the result, the estimated median of the log returns is 0.0007606.

Standard error of the estimate for the median of the log returns using the bootstrap method.

```
set.seed(1)
# Define boot function
boot.fn <- function(data, index) {
  median(data[index])
}
# Estimate the standard error using bootstrap
se.boot.med <- boot(data = AAPL$Log.return, statistic = boot.fn, R = 1000)
print(se.boot.med)
```

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

```
##
##
## Call:
## boot(data = AAPL$Log.return, statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##         original       bias     std. error
## t1* 0.0007606322 4.235446e-06 0.0003202219
```

By using the bootstrap method, the standard error of the estimated median of the log returns is 0.0003202219, which is smaller than the standard error of the estimated mean of the log returns.

Conclusion:

From the results, we can observe that even though we used the different datasets, we still found that:

- The estimated means ($\hat{\mu}$) are greater than the estimated medians from both the variable medv and log returns.
- Also, the standard errors of the mean obtained by using the sample standard method are greater than the standard errors of the mean obtained by using the bootstrap method for both the variable medv and log returns.
- However, the results are very close. In addition, the standard errors of the median are less than the standard errors of the mean for both the variable medv and log returns.