Homework 1 Partial Solutions

Spring 2021

4

(a)

```
treatment <- c(94, 197, 16, 38, 99, 141, 23)
control <- c(52, 104, 146, 10, 51, 30, 40, 27, 46)

set.seed(1234)
B <- c(50, 100, 200, 500, 1000, 10000) # number of Bootstrap samples

boots_means <- c()
each_mean <- c()
each_sd <- c()

for (j in 1 : length(B)){
for (i in 1 : B[j]){
boots_means[i] <- mean(sample(treatment, length(treatment), replace = TRUE))
}
each_mean[j] <- mean(boots_means)
each_sd[j] <- sd(boots_means)
}

each_mean
```

```
## [1] 86.98000 83.37571 87.08500 87.87143 86.93029 87.19431
```

```
each_sd
```

```
## [1] 27.41873 21.42969 20.66055 23.69451 23.52464 23.67236
```

(b)

```
treatment <- c(94, 197, 16, 38, 99, 141, 23)
control <- c(52, 104, 146, 10, 51, 30, 40, 27, 46)

set.seed(1234)
B <- c(50, 100, 200, 500, 1000, 10000) # number of Bootstrap samples

boots_med <- c()
```

```r
ind_mean <- c()
ind_sd <- c()

for (j in 1 : length(B)){
for (i in 1 : B[j]){
boots_med[i] <- median(sample(treatment, length(treatment), replace = TRUE))
}
ind_mean[j] <- mean(boots_med)
ind_sd[j] <- sd(boots_med)
}

ind_mean
```

```
## [1] 82.3400 80.4900 81.3950 80.9020 79.8040 79.7709
```

```r
ind_sd
```

```
## [1] 40.94327 35.85253 34.45613 38.67899 38.25900 37.98158
```

(c)

```r
library(tidyverse)
library(knitr)
kable(tibble(B, Mean = each_mean, mean_sd = each_sd, Median = ind_mean, Median_sd = ind_sd))
```

B	Mean	mean_sd	Median	Median_sd
50	86.98000	27.41873	82.3400	40.94327
100	83.37571	21.42969	80.4900	35.85253
200	87.08500	20.66055	81.3950	34.45613
500	87.87143	23.69451	80.9020	38.67899
1000	86.93029	23.52464	79.8040	38.25900
10000	87.19431	23.67236	79.7709	37.98158

(d)

```r
set.seed(1234)

boots_medc <- c()

for (i in 1:10000){
boots_medc[i] <- median(sample(control, length(control), replace = TRUE))
}
mean(boots_medc)
```

```
## [1] 45.6253
```

```r
sd(boots_medc)
```

```
## [1] 12.43127
```

(e)

The estimated standard error of the difference is the square root of the sum of the square of each standard error.

```
var.diff <-  ind_sd[6]^2 + sd(boots_medc)^2
sqrt(var.diff)
```

```
## [1] 39.9642
```

5

Since $\hat{\theta}_{\mathrm{MLE}} = \bar{x}/2$ and $\mathrm{Var}(\bar{x}/2) = \sigma^2/(4n)$, we can compare the bootstrap estiamte of std error (0.214) and the estiamte of std error (0.223). They are pretty close.

```
library(boot)
set.seed(123)
x <- rgamma(200, shape = 2, scale = 5)

boot_theta <- function(x) { mean(x)/2 }

ran_gamma <- function(x, mle) {rgamma(length(x), 2, scale = mle)}
boot(x, statistic = boot_theta, R = 1000, sim = "parametric", ran.gen = ran_gamma,
     mle = mean(x)/2)
```

```
##
## PARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = x, statistic = boot_theta, R = 1000, sim = "parametric",
##     ran.gen = ran_gamma, mle = mean(x)/2)
##
##
## Bootstrap Statistics :
##     original      bias    std. error
## t1* 4.578726 0.005143687   0.2135533
```

```
sd(x)/(2*sqrt(length(x)))
```

```
## [1] 0.2231809
```