Bruce Campbell ST-617 Homework 2

Wed Jul 06 06:33:37 2016

Chapter 5

Problem 9

We will now consider the Boston housing data set, from the MASS library. ### a) Based on this data set, provide an estimate for the population mean of medv. Call this estimate $\hat{\mu}$

```
library(MASS)
attach(Boston)
mu_hat <- mean(Boston$medv)
```

$\hat{\mu} =$ `22.5328063`

b)

Provide an estimate of the standard error of $\hat{\mu}$. Interpret this result.

Hint: We can compute the standard error of the sample mean by dividing the sample standard deviation by the square root of the number of observations.

```
se_sm <- sd(Boston$medv)/nrow(Boston)
```

Our estimate of the standard error of the sample mean based on the standard deviation is `0.0181761`

c)

Now estimate the standard error of $\hat{\mu}$ using the bootstrap. How does this compare to your answer from (b)?

```
library(boot)
alpha.fn <- function(data, index) {
    D = data[index, ]
    result <- mean(D$medv)
    return(result)
}

se_bootstrap <- boot(Boston, alpha.fn, 100)

se_bootstrap
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
```

```
## boot(data = Boston, statistic = alpha.fn, R = 100)
##
##
## Bootstrap Statistics :
##     original      bias    std. error
## t1* 22.53281 -0.01673518   0.4333896
```

d)

Based on your bootstrap estimate from (c), provide a 95% confidence interval for the mean of medv. Compare it to the results obtained using t.test(Boston$medv).

Hint: You can approximate a 95% confidence interval using the formula $[\hat{\mu} - 2SE(\hat{\mu}), \hat{\mu} + 2SE(\hat{\mu})]$.

```
left_ci <- mu_hat - 2 * sd(se_bootstrap$t)
right_ci <- mu_hat + 2 * sd(se_bootstrap$t)
```

Our 95% CI as estimated from the boostrap calculation is [`21.6660271` , `23.3995856`]

The results of the t-test are comparable to the bootstrap estimate, but the one sample t-test provides a smaller estimate of the 95% CI.

```
t.test(Boston$medv)
```

```
##
##  One Sample t-test
##
## data:  Boston$medv
## t = 55.111, df = 505, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  21.72953 23.33608
## sample estimates:
## mean of x
##  22.53281
```

e)

Based on this data set, provide an estimate, med, for the median value of medv in the population.

We can use the sample median for an estimate of the population median which is :

```
median(Boston$medv)
```

```
## [1] 21.2
```

f)

We now would like to estimate the standard error of $\hat{\mu_m}ed$. Unfortunately, there is no simple formula for computing the standard error of the median. Instead, estimate the standard error of the median using the bootstrap. Comment on your findings.

```r
alpha.fn <- function(data, index) {
    D = data[index, ]
    result <- median(D$medv)
    return(result)
}
se_median_bootstrap <- boot(Boston, alpha.fn, 100)

se_median_bootstrap
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston, statistic = alpha.fn, R = 100)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*     21.2 -0.0335   0.3801286
```

The estimate of standard error of the median as calculated by the bootstrap is

```r
sd(se_median_bootstrap$t)
```

```
## [1] 0.3801286
```

```r
left_ci <- mu_hat - 2 * sd(se_median_bootstrap$t)
right_ci <- mu_hat + 2 * sd(se_median_bootstrap$t)
```

Our 95% CI as estimated from the boostrap calculation is [`21.7725492` , `23.2930635`]

We're not sure if this estimate is valid for the median so let's calculate the 5th and 95th quantile from the bootstrap sample

```r
left_ci <- quantile(se_median_bootstrap$t, 0.05)
right_ci <- quantile(se_median_bootstrap$t, 0.95)
```

[`20.6` , `21.705`]

This is a much tighter CI, so we wonder if the estimate above applied to the median.

g)

Based on this data set, provide an estimate for the tenth percentile of medv in Boston suburbs. Call this quantity $\hat{\mu_0}.1$. (You can use the quantile() function.)

The sample tenth percentile of medv is

```r
quantile(Boston$medv, 0.1)
```

```
##   10%
## 12.75
```

h)

Use the bootstrap to estimate the standard error of $\hat{\mu_0}.1$. Comment on your findings.

```
alpha.fn <- function(data, index) {
    D = data[index, ]
    result <- quantile(D$medv, 0.1)
    return(result)
}
se_quantile_bootstrap <- boot(Boston, alpha.fn, 100)

se_quantile_bootstrap
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston, statistic = alpha.fn, R = 100)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*    12.75  0.0095   0.5113705
```

The estimate of standard error of the tenth percentile as calculated by the bootstrap is

```
sd(se_quantile_bootstrap$t)
```

```
## [1] 0.5113705
```