Q26-WilliamKennedy

William Kennedy

2023-11-09

1.Based on this data set, provide an estimate for the population mean of medv. Call this estimate $\hat{\mu}$.

```
Boston = read.csv("Boston.csv")

medv.sub = Boston$medv
medv.mean = mean(medv.sub)
```

The population mean is $\hat{\mu} \approx 22.532$

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

```
medv.std = sd(medv.sub)
medv.std.error = medv.mean/sqrt(nrow(Boston))
```

The standard error for $\hat{\mu}$ is $se(\hat{\mu}) \approx 1.001705$

3.Now estimate the standard error of $\hat{\mu}$ using the bootstrap. How does this compare to your answer from 2.?

```
library("boot")
```

```
## Warning: package 'boot' was built under R version 4.3.2
```

```
alpha.fn = function (data , index) {
  x= Boston$medv[index]
  mean.two = mean(x)
  return(mean.two)
}

boot(Boston,alpha.fn,R=1000)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston, statistic = alpha.fn, R = 1000)
##
##
## Bootstrap Statistics :
##     original      bias    std. error
## t1* 22.53281 0.02228478   0.4200753
```

The standard error using bootstrapping is 0.4044482, whereas computing the standard error using the population mean is 1.001705. Hence using bootstrapping yields a smaller standard error for the feature medv in the Boston dataset sample.

4.Based on your bootstrap estimate from 3., provide a 95% confidence interval for the mean of medv. Compare it to the results obtained using t.test(Boston$medv).

The 95% confidence interval of $\hat{\mu}$ is $[\hat{\mu} - 2SE(\hat{\mu}), \hat{\mu} + 2SE(\hat{\mu})] = [22.532 - 0.8088964, 22.532 + 0.8088964] = [21.723, 23.3408964]$

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

```
##
##  One Sample t-test
##
## data:  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
```

The confidence interval calculated from the SE for the bootstrap and the t-test are very similar

5.Based on this dataset, provide an estimate, $\hat{\mu}_{med}$ for the median value of medv in the population.

```
median.medv = median(Boston$medv)
```

6.

```
boot.median = function(data,index) {
  x=medv[index]
  x.median = median(x)
  return (x.median)
}

boot(medv,boot.median,R=1000)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = medv, statistic = boot.median, R = 1000)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*     21.2 -0.0078   0.3831527
```

Hence the estimate for $\hat{\mu}_{med} \approx 0.3748417$

7.

```r
perc.ten = quantile(medv, 0.1)
```

8.

```r
percentile.ten = function(data, index){
  X=medv[index]
  Y = quantile(X, 0.1)
  return(Y)
}

boot(Boston,percentile.ten,1000)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Boston, statistic = percentile.ten, R = 1000)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*    12.75  -0.006   0.5132664
```

Hence the standard error is 0.5127017