Assignment26

John Bute

2024-11-07

R Markdown

```
library(MASS)
data("Boston")
```

We now consider the Boston housing data set. 1. Based on this data set, provide an estimate for the population mean of medv. Call this estimate ˆµ.

```
medv_mean <- mean(Boston$medv)
medv_mean
```

```
## [1] 22.53281
```

2. Provide an estimate of the standard error of ˆµ. 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.

```
standard_error_of_mean <- sd(Boston$medv)/sqrt(length(Boston$medv))

standard_error_of_mean
```

```
## [1] 0.4088611
```

3. Now estimate the standard error of ˆµ using the bootstrap. How does this compare to your answer from 2.?

```
n <- length(Boston$medv)
bootstrap_means <- replicate(1000, mean(sample(Boston$medv, n, replace =
TRUE)))
se_bootstrap <- sd(bootstrap_means)

se_bootstrap
```

```
## [1] 0.4204812
```

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).

Hint: You can approximate a 95% confidence interval using the formula [ˆµ −2SE(ˆµ), ˆµ + 2SE(ˆµ)].

```
ci_bootstrap <- c(medv_mean - 2 * se_bootstrap, medv_mean + 2 * se_bootstrap)
ci_bootstrap
```

```
## [1] 21.69184 23.37377
```

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

```
## [1] 21.72953 23.33608
## attr(,"conf.level")
## [1] 0.95
```

The bootstrap 95% C.I. for the mean is (21.92652 23.13909) while it is very close to the same as the t-test C.I. of (21.72953 23.33608)

5. Based on this dataset, provide an estimate, ˆµmed for the median value of medv in the population.

```
median_medv <- median(Boston$medv)
median_medv
```

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

6. We now would like to estimate the standard error of ˆµmed. 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.

```
bootstrap_median <- replicate(1000, median(sample(Boston$medv, n, replace =
TRUE)))
se_median_bootstrap <- sd(bootstrap_median)

se_median_bootstrap
```

```
## [1] 0.3667246
```

The standard error is very small, which suggests that any confience interval will be fairly tight too.

7. Based on this data set, provide an estimate for the tenth percentile of medv in Boston suburbs. Call this quantity ˆµ0.1. (You can use the quantile() function.)

```
percentile_10 <- quantile(Boston$medv, 0.1)
percentile_10
```

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

8 Use the bootstrap to estimate the standard error of ˆµ0.1. Comment on your findings

```
bootstrap_percentiles <- replicate(1000, quantile(sample(Boston$medv, n,
replace = TRUE), 0.1))
se_percentile_10 <- sd(bootstrap_percentiles)
se_percentile_10
```

```
## [1] 0.5211287
```

The standard error is small, but not quite the same as the median or the mean.