ISLR Chapter 5, Ex 9
Resampling methods
Tatiana Romanchishina

9.
(a) Based on this data set, provide an estimate for the population mean of medv . Call this estimate $\hat{\mu}$.

```
> mu_medv <- mean(medv)
> mu_medv
[1] 22.53281
```

The estimated population mean of medv is $\approx$ 22.5.

(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.

```
> dim(Boston)
[1] 506  14
> se_mu_medv <- sd(medv) / sqrt(dim(Boston)[1])
> se_mu_medv
[1] 0.4088611
```

The standard error of the sample mean is $\approx$ 0.41.

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

```
> set.seed(1)
> boot.fn <- function(data, index) {
+ mean_medv = mean(data[index])
+ return (mean_medv)
+ }
> boot(medv, boot.fn, 1000)
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = medv, statistic = boot.fn, R = 1000)


Bootstrap Statistics :
      original        bias          std. error
t1* 22.53281 0.008517589    0.4119374
```

The estimated standard error is almost the same as the one estimated in b).

(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 $[\hat{\mu} - 2SE(\hat{\mu}), \hat{\mu} + 2SE(\hat{\mu})]$.

> 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

using the formula $[\hat{\mu} - 2SE(\hat{\mu}), \hat{\mu} + 2SE(\hat{\mu})]$

```
> mean_medv_conf_int <- c(22.53281 - 2*0.4119374, 22.53281 + 2*0.4119374)
> mean_medv_conf_int
[1] 21.70894 23.35668
```

The results are very similar.

(e) Based on this data set, provide an estimate, $\hat{\mu}_{med}$, for the median value of medv in the population.

```
> med_medv <- median(medv)
> med_medv
[1] 21.2
```

The population median of medv is 21.2.

(f) We now would like to estimate the standard error of $\hat{\mu}_{med}$. 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.

```
> boot.fn2 <- function(data, index) {
+ med <- median(data[index])
+ return(med)
+ }
```

```
> boot(medv, boot.fn2, 1000)

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = medv, statistic = boot.fn2, R = 1000)


Bootstrap Statistics :
        original    bias    std. error
t1*       21.2  -0.0025   0.374358
```

The median value is the same with a small standard error.

(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.)

```
> perc_10_medv <- quantile(medv, 0.1)
> perc_10_medv
  10%
12.75
```

(h) Use the bootstrap to estimate the standard error of $\hat{\mu}_{0.1}$. Comment on your findings.

```
> boot.fn3 <- function(data, index) {
+ perc_10 <- quantile(data[index], 0.1)
+ return (perc_10)
+ }
> boot(medv, boot.fn3, 1000)

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = medv, statistic = boot.fn3, R = 1000)


Bootstrap Statistics :
         original  bias    std. error
t1*       12.75  0.0261   0.4912231
```

The same value as obtained in g), relatively small standard error.