Chapter5Homework

Jose Chavarria

4/24/2020

A-F

A.

```
library(MASS)
attach(Boston)
mu.hat <- mean(medv)
mu.hat
```

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

B.

```
se <- sqrt(var(medv)/dim(Boston)[1])
se
```

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

C.

```
library(boot)
set.seed(1)
bootstrap <- function(data, index) {
    mu<- mean(data[index])
    return (mu)
}
estimate<-boot(medv, bootstrap, 1000)
estimate
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = medv, statistic = bootstrap, R = 1000)
##
##
## Bootstrap Statistics :
##     original       bias    std. error
## t1* 22.53281 0.008517589   0.4119374
```

With a bootstrap the given standard error is 0.4119374 compared to 0.4088611 from $\hat{\mu}$

D.

```
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
```

```
ci<-c(22.53281-2*0.4119374,  22.53281+2*0.4119374)
ci
```

```
## [1] 21.70894 23.35668
```

E.

```
med.hat <- median(medv)
med.hat
```

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

F.

```
set.seed(1)
boot2 <- function(data, index) {
    med <- median(data[index])
    return (med)
}
estimate2<-boot(medv, boot2, 2500)
estimate2
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = medv, statistic = boot2, R = 2500)
##
##
## Bootstrap Statistics :
##     original    bias    std. error
## t1*     21.2 -0.01376   0.3789064
```

The standard error for median using the bootstrap is 0.3789064, compared to the mean bootstrap, smaller.

> College Football Data Set

```
library(readxl)
cfb <- read_excel("CFB2018completeISLR.xlsx")
```

Leave One Out Cross Validation

-Models

```
- model1 <- lm(Zsagarin ~ lysagarin + Fr5star + coachexp_school,data=cfb)

- model2 <- lm(Zsagarin ~ lysagarin + Fr5star + coachexp_school + I(Fr5star^2) + I(coachexp_scho
ol^2), data=cfb)
```

```
set.seed(1)
train <- sample(857,428)

model1 <- lm(Zsagarin ~ lysagarin + Fr5star + coachexp_school,data=cfb, subset=train)
mean((cfb$Zsagarin-predict(model1,cfb))[-train]^2)
```

```
## [1] 0.5201733
```

```
model2 <- lm(Zsagarin ~ lysagarin + Fr5star + coachexp_school + I(Fr5star^2) + I(coachexp_school
^2), data=cfb, subset=train)
mean((cfb$Zsagarin-predict(model2,cfb))[-train]^2)
```

```
## [1] 0.5173948
```

```
set.seed(1)
glm.fit1<-glm(Zsagarin ~ lysagarin + Fr5star + coachexp_school, data = cfb)
coef(glm.fit1)
```

```
##     (Intercept)       lysagarin         Fr5star coachexp_school
##    -4.314682869     0.055105498     0.157257312     0.009994924
```

```
cv.err1<-cv.glm(cfb,glm.fit1)
cv.err1$delta
```

```
## [1] 0.5206591 0.5206565
```

```
set.seed(1)
glm.fit2<-glm(Zsagarin ~ lysagarin + Fr5star + coachexp_school + I(Fr5star^2) + I(coachexp_schoo
l^2), data = cfb)
coef(glm.fit2)
```

```
##          (Intercept)            lysagarin              Fr5star
##        -4.2687433204         0.0539339823         0.2534698353
##      coachexp_school         I(Fr5star^2) I(coachexp_school^2)
##         0.0212820250        -0.0228142347        -0.0004000457
```

```
cv.err2<-cv.glm(cfb,glm.fit2)
cv.err2$delta
```

```
## [1] 0.5192124 0.5192087
```

Our second, quadratic, model appears to be a better fit than our first model.

K Fold Cross Validation

```
set.seed(1)
glm.one<-glm(Zsagarin ~ lysagarin + Fr5star + coachexp_school, data = cfb)
cv.glm.one <- cv.glm(cfb,glm.one,K=10)
cv.glm.one$delta
```

```
## [1] 0.5205424 0.5203103
```

```
set.seed(1)
glm.two<-glm(Zsagarin ~ lysagarin + Fr5star + coachexp_school + I(Fr5star^2) + I(coachexp_school
^2), data = cfb)
cv.glm.two <- cv.glm(cfb,glm.two,K=10)
cv.glm.two$delta
```

```
## [1] 0.5182963 0.5180113
```

Quadratic model performs better than a regular linear model.