

## Exercises 10.30

Import data

First, Set the environment and import data. *Note that i wrote the code using R*

```
rm(list=ls())
library(dplyr) #dataframe management
library(purrr) #dataframe management
library(haven) #library for importing data set
library(psych) #library for trace matrix
cps09mar <- read_dta("cps09mar.dta")
```

Table 1: First 10 elements of cps09mar

age	female	hisp	education	earnings	hours	week	union	uncov	region	race	marital
52	0	0	12	146000	45	52	0	0	1	1	1
38	0	0	18	50000	45	52	0	0	1	1	1
38	0	0	14	32000	40	51	0	0	1	1	1
41	1	0	13	47000	40	52	0	0	1	1	1
42	0	0	13	161525	50	52	1	0	1	1	1
66	1	0	13	33000	40	52	0	0	1	1	5
51	0	0	16	37000	44	52	0	0	1	1	1
49	1	0	16	37000	44	52	0	0	1	1	1
33	0	0	16	80000	40	52	0	0	1	1	1
52	1	0	14	32000	40	52	0	0	1	1	1

Filter the data for White male Hispanics who never-married and live in the mid-west region,

```
cps <- cps09mar[cps09mar$hisp == 1 & cps09mar$female == 0 & cps09mar$race == 1 &
                 cps09mar$marital == 7 & cps09mar$region == 2,]
cps %>%
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1     99
```

And create matrix for variables.

```
con <- matrix(1,nrow = nrow(cps),ncol = 1)
edu <- matrix(0,nrow = nrow(cps),ncol = 1)
edu[,] <- cps$education

exp <- matrix(0,nrow = nrow(cps),ncol = 1)
exp[,] <- cps$age - cps$education - 6

expsq <- exp^2/100
```

```
wage <- matrix(0,nrow = nrow(cps),ncol = 1)
wage[,] <- cps$earnings / (cps$hours*cps$week)
lwage <- log(wage)

X <- cbind(edu,exp,expsq,con)
Y <- lwage
```

(a) Estimate  $\theta$  and report standard errors calculated by asymptotic, jackknife and the bootstrap.

Now I can estimate Beta and Error.

```
beta_hat <- solve((t(X) %*% X)) %*% (t(X) %*% Y)
e_hat <- Y - X %*% beta_hat
row.names(beta_hat) <- c('edu','exp','expsq/100','con')
colnames(beta_hat) <- c('coefficient')
colnames(e_hat) <- c('error')
```

Set the variable for convenience.

```
XX <- solve((t(X) %*% X)) #XX is variable name it means inverse matrix of (X'X)
leverage <- diag(X%*%XX%*t(X))
```

Calculate S.E using the HC3 method.

```
u3 <- X * ((e_hat/(1-leverage))%*% matrix(1,1,ncol(X)))
v3 <- XX %*% (t(u3) %*% u3) %*% XX
s3 <- sqrt(diag(v3))
```

Table 2: Coefficient estimates and Robust S.E.

	Coefficient	Robust standard errors
edu	0.0851633	0.0167301
exp	0.0510575	0.0181581
expsq/100	-0.1084199	0.0585962
con	1.1802807	0.2525230

Let  $\theta$  be the ratio of the return to one year of education to the return to one year of experience for experience = 10.

we got  $\log(wage) = \beta_1 education + \beta_2 experience + \beta_3 experience^2/100 + \beta_4$

$$\frac{\partial \log(wage)}{\partial education} = \beta_1$$

$$\frac{\partial \log(wage)}{\partial experience} = \beta_2 + \beta_3 experience/50$$

$$\theta = \frac{\beta_1}{\beta_2 + \beta_3/5}$$

Thus following result is  $\hat{\theta}$ .

```
theta_hat <- beta_hat[1]/(beta_hat[2]+(beta_hat[3]/5))
print(theta_hat)
```

```
## [1] 2.899323
```

### Asymptotic Standard Errors:

First, Calculate  $\mathbf{R} = \frac{\partial r(\beta)'}{\partial \beta}$ .

$$\mathbf{R} = \begin{bmatrix} \frac{1}{\beta_2 + \beta_3/5} \\ -\frac{25\beta_1}{(5\beta_2 + \beta_3)^2} \\ -\frac{5\beta_1}{(5\beta_2 + \beta_3)^2} \\ 0 \end{bmatrix}$$

$$\text{Thus } \hat{\mathbf{R}} = \begin{bmatrix} \frac{1}{\hat{\beta}_2 + \hat{\beta}_3/5} \\ -\frac{25\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} \\ -\frac{5\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} \\ 0 \end{bmatrix} = \begin{bmatrix} 34.04429 \\ -98.7054 \\ -19.74108 \\ 0 \end{bmatrix}$$

$$\text{since } s(\hat{\theta}) = \sqrt{\hat{R}' \hat{\mathbf{V}}_{\hat{\beta}} \hat{R}} = \sqrt{\begin{bmatrix} \frac{1}{\hat{\beta}_2 + \hat{\beta}_3/5} & -\frac{25\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} & -\frac{5\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} & 0 \end{bmatrix} \hat{\mathbf{V}}_{\hat{\beta}} \begin{bmatrix} \frac{1}{\hat{\beta}_2 + \hat{\beta}_3/5} \\ -\frac{25\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} \\ -\frac{5\hat{\beta}_1}{(5\hat{\beta}_2 + \hat{\beta}_3)^2} \\ 0 \end{bmatrix}}$$

Calculation is following codes.

```
R_hat <- matrix(c(1/(beta_hat[2]+(beta_hat[3]/5)),
-25*beta_hat[1]/((5*beta_hat[2]+beta_hat[3])^2),
-5*beta_hat[1]/((5*beta_hat[2]+beta_hat[3])^2),0),4,1)
v_theta <- t(R_hat)%*%v3*%R_hat
s_theta <- sqrt(v_theta)
print(s_theta)
```

```
## [1]
## [1,] 0.7954028
```

Thus  $s(\hat{\theta})^{HC3} = 0.7954028$ .

### Calculate Jackknife Standard Errors:

Jackknife estimate of variance for  $\hat{\beta}$  where  $\bar{\theta} = n^{-1} \sum_{i=1}^n \hat{\theta}_{(-i)}$  is

$$\hat{V}_{\hat{\theta}}^{jack} = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(-i)} - \bar{\theta})(\hat{\theta}_{(-i)} - \bar{\theta})'$$

Thus Standard Errors of  $\hat{\theta}$  with jackknife is following codes.

```
bj <- matrix(0,4,99)
thetaj <- matrix(0,99,1)
for (i in 1:99) {
  Yj <- Y[-i]
  Xj <- X[-i,]

  bj[,i] <- solve(t(Xj)%*%Xj)%*%(t(Xj)%*%Yj)
  thetaj[i,] <- bj[1,i]/(bj[2,i]+(bj[3,i]/5))
}
jackv <- 98*(t(thetaj - matrix(mean(thetaj),99,1))%*%
            (thetaj - matrix(mean(thetaj),99,1)))/(99)
jacks <- sqrt(jackv)
```

Thus  $s(\hat{\theta})^{jackknife} = 0.8229674$ .

### Calculate Bootstrap Standard Errors:

Bootstrap estimate of variance for  $\hat{\theta}$  where  $b = 1, \dots, B$  is index of bootstrap samples is

$$\hat{V}_{\hat{\theta}}^{boot} = \frac{1}{B-1} \sum_{i=1}^n (\hat{\theta}^*(B) - \bar{\theta}^*)(\hat{\theta}^*(B) - \bar{\theta}^*)'$$

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(B).$$

calculate  $s(\hat{\theta})^{bootstrap}$  is following codes.

```
B <- 100000;

bb <- matrix(0,4,B)
t1 <- matrix(0,B,1)
for (i in 1:B) {
  boot <- matrix(rdunif(99,1,99),99,1)
  Yb <- Y[boot]
  Xb <- X[boot,]

  bb[,i] <- solve(t(Xb)%*%Xb)%*%(t(Xb)%*%Yb)
  t1[i,] <- bb[1,i]/(bb[2,i]+(bb[3,i]/5))
}
bootv <- (t(t1 - mean(t1)))%*%(t1 - mean(t1))/(B-1)
s_thetab <- sqrt(bootv)
```

Thus  $s(\hat{\theta})^{bootstrap} = 39.2278948$ .

### Trimmed Bootstrap Standard Errors:

Since we have nonlinear estimators, The untrimmed estimator is unreliable as we can see in the results above.  $s(\hat{\theta})^{bootstrap}$  well exceed other values of standard errors calculated by other methods. Thus it is more

desirable to calculate the trimmed bootstrap standard errors and report it. Following codes are calculation of trimmed bootstrap standard errors.

```
bbt <- matrix(0,4,B)
t1t <- matrix(0,B,1)
tau <- 2
thetat <- 0
for (i in 1:B) {
  boott <- matrix(rdunif(99,1,99),99,1)
  Ybt <- Y[boott]
  Xbt <- X[boott,]

  bbt[,i] <- solve(t(Xbt)%*%Xbt)%*%(t(Xbt)%*%Ybt)
  thetab <- bbt[1,i]/(bbt[2,i]+(bbt[3,i]/5))
  thetat <- thetab *(abs(thetab-theta_hat) <= tau) +
    (theta_hat-tau) * (thetab < theta_hat-tau) +
    (theta_hat+tau)*(thetab > theta_hat +tau)
  t1t[i,] <-thetat
}
bootvt <- (t(t1t - mean(t1t)))*%(t1t - mean(t1t))/(B-1)
s_thetabt <- sqrt(bootvt)
```

Thus  $s(\hat{\theta})_{trimmed}^{bootstrap} = 0.8129487$ .

Standard Errors of  $\hat{\theta}$  calculated various methods:

Table 3: S.E. of  $\hat{\theta}$  calculated by various methods.

$s(\hat{\theta})^{HC3}$	$s(\hat{\theta})^{jackknife}$	$s(\hat{\theta})^{bootstrap}$	$s(\hat{\theta})_{trimmed}^{bootstrap}$
0.7954028	0.8229674	39.22789	0.8129487

## (b) Explain the discrepancy between the standard errors.

Asymptotic standard errors assume that distribution of  $\sqrt{n}(\hat{\theta} - \theta)$  follows the normal distribution as  $n \rightarrow \infty$ . where basic assumptions are there. and  $\theta$  is function of  $\beta$ . But the asymptotic distribution is never the same as the exact distribution.

Other two estimation methods are not based on the asymptotic. Others are using resampling methods. They use the computational power to estimate the distribution. While jackknife estimator is calculated without reference to asymptotic theory, jackknife produces an estimator which is asymptotically similar to one produced by asymptotic methods. As show in the [Table 3], we can check that indeed Standard error calculated by HC3 and jackknife is similar in results.

$$\begin{aligned}
\hat{V}_{\hat{\theta}}^{jack} &= \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(-i)} - \bar{\theta})(\hat{\theta}_{(-i)} - \bar{\theta})' \\
&\simeq \frac{n-1}{n} \hat{R}' (X'X)^{-1} \left( \sum_{i=1}^n X_i X_i' \tilde{e}_i^2 - n \tilde{\mu} \tilde{\mu}' \right) (X'X)^{-1} \hat{R} \\
&= \hat{R}' \hat{V}_{\hat{\beta}}^{jack} \hat{R} \\
&\simeq \hat{R}' \hat{V}_{\hat{\beta}}^{HC3} \hat{R}
\end{aligned}$$

Bootstrap is other method of resampling. bootstrap distribution is obtained by estimation on independent samples created by i.i.d. from original dataset. Discrepancy between the untrimmed bootstrap S.E. and other two values are very large. Because of the  $\theta$  is nonlinear. As trimming the values with  $\tau$  that satisfying  $\tau_n = O(e^{(n/8)})$ , we can observe that S.E. calculated by trimmed bootstrap method is more reliable.<sup>1</sup>

**(c) Report confidence intervals for  $\theta$  using the BC percentile method.**

Let  $\alpha = 0.05$ . Following codes are calculation of biased corrected percentile confidence interval of  $\theta$ .

```
z0 <- qnorm(mean(t1<=theta_hat))
z1 <- qnorm(0.025)
z2 <- qnorm(0.975)
xa1 <- pnorm(z1+2*z0)
xa2 <- pnorm(z2+2*z0)
cl <- quantile(t1,probs = c(xa1,xa2))[[1]]
ch <- quantile(t1,probs = c(xa1,xa2))[[2]]
```

Thus  $\hat{C}^{BC} = [1.7456499, 5.3108549]$ .

---

<sup>1</sup>Hansen, Econometrics, p. 276.