Sampling Design and Survey Practice Lab #3

TA - Seungkyu Kim

2022-10-26

Install and load packages

```
name_pkg <- c("survey", "sampling", "SDAResources")
name_pkg <- unique(name_pkg)
bool_nopkg <- !name_pkg %in% rownames(installed.packages())
if (sum(bool_nopkg) > 0) {
  install.packages(name_pkg[bool_nopkg])
}
invisible(lapply(name_pkg, library, character.only = T))
```

1. 비추정 (Ratio Estimation)

```
data(agsrs)
n<-nrow(agsrs) #300
agsrs$sampwt <- rep(3078/n,n) # 3078 = nrow(agpop)
agdsrs <- svydesign(id = ~1, weights=~sampwt, fpc=rep(3078,300), data = agsrs)
agdsrs
```

```
## Independent Sampling design
## svydesign(id = ~1, weights = ~sampwt, fpc = rep(3078, 300), data = agsrs)
```

모비 $\beta = \tau_y/\tau_x$ 에 대한 추정

두 변수간의 correlation이 클 때 비추정량의 정확도가 높아진다. 비추정량은 svyratio 명령어를 통해 구할 수 있다.

```
cor(agsrs$acres87,agsrs$acres92)
```

```
## [1] 0.995806
```

```
# estimate the ratio acres92/acres87
sratio<-svyratio(numerator = ~acres92, denominator = ~acres87,design = agdsrs)
sratio
```

```
## Ratio estimator: svyratio.survey.design2(numerator = ~acres92, denominator = ~acres87,
##     design = agdsrs)
## Ratios=
##           acres87
## acres92 0.9865652
## SEs=
##             acres87
## acres92 0.005750473
```

```
confint(sratio, df=degf(agdsrs))
```

```
##                       2.5 %    97.5 %
## acres92/acres87 0.9752487 0.9978818
```

모합 τ_y 에 대한 추정

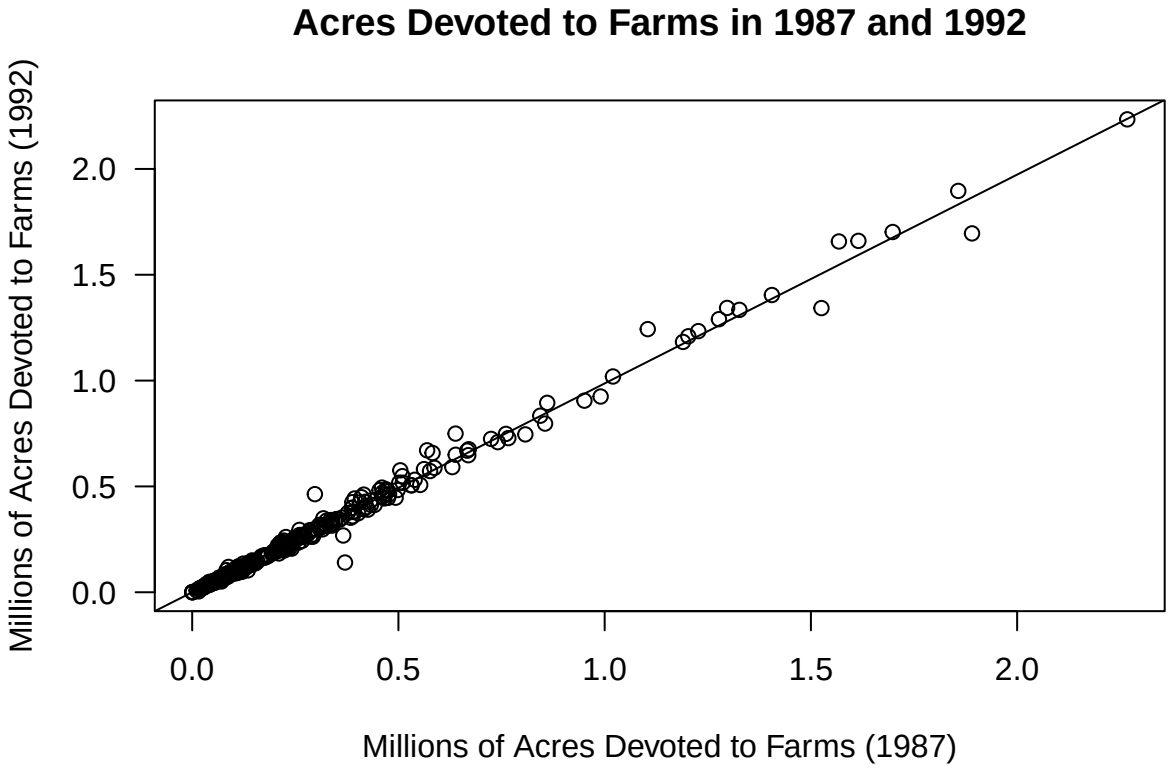
```
# provide the population total of acres87
xpoptotal <- 964470625
# Ratio estimate of population total
predict(sratio,total=xpoptotal)
```

```
## $total
##           acres87
## acres92 951513191
##
## $se
##         acres87
## acres92 5546162
```

```
# Ratio estimate of population mean
predict(sratio,total=xpoptotal/3078)
```

```
## $total
##          acres87
## acres92 309133.6
##
## $se
##          acres87
## acres92 1801.872
```

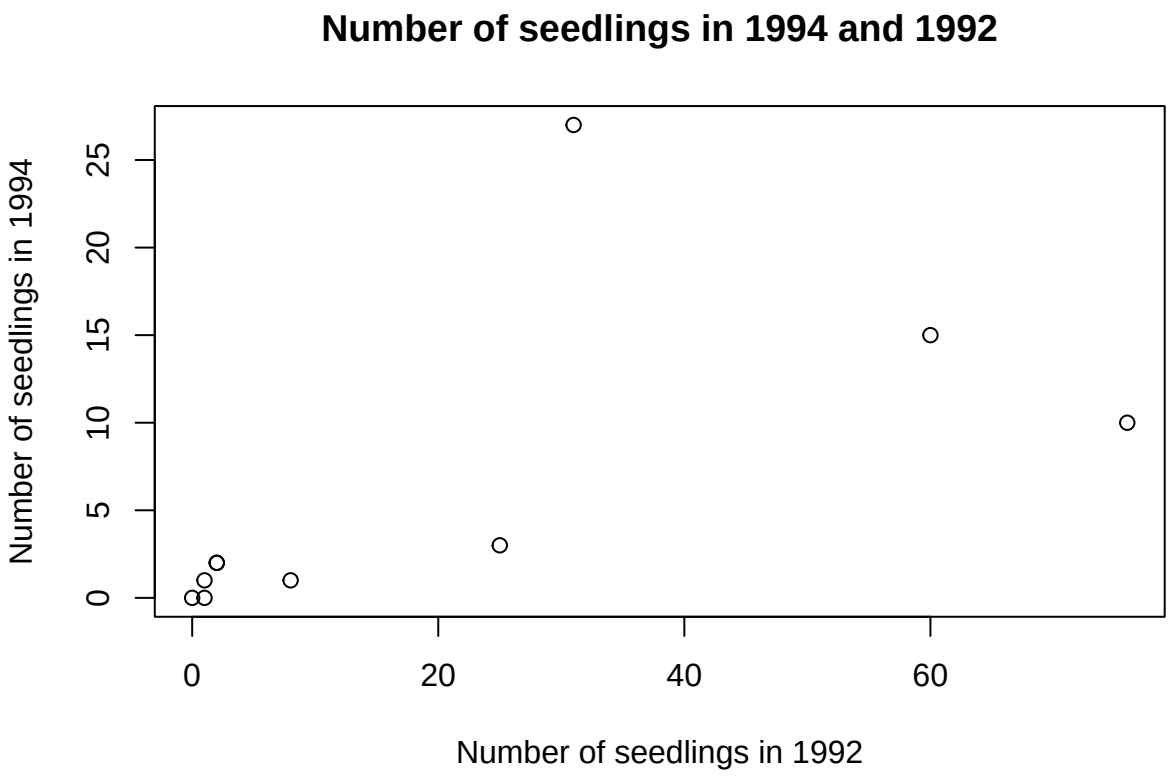
```
# draw the scatterplot
par(las=1) # make tick mark labels horizontal (optional)
plot(x=agsrs$acres87/1e6,y=agsrs$acres92/1e6,
     xlab="Millions of Acres Devoted to Farms (1987)",
     ylab = "Millions of Acres Devoted to Farms (1992)",
     main = "Acres Devoted to Farms in 1987 and 1992")
# draw line through origin with slope bhat
abline(0,coef(sratio))
```



그림에 보이는 직선은 원점을 지나면서 기울기가 비추정량(0.9865652)인 직선이다. 단순회귀분석을 통해 구할 수 있는 직선의 기울기는 앞서 구한 correlation(0.995806)과 같다. 단순회귀분석을 통해 구하는 직선은 원점을 지나는지와의 여부와 상관없이 구하는 것이고, 비추정량과 관련된 직선은 원점을 지난다는 제약조건이 있기 때문에 여기에서 값의 차이가 나타나게 된다.

* Population size를 모를 때 비추정량 구하기

```
# scatterplot and correlation of seed92 and seed94
data(santacruz)
plot(santacruz$seed92,santacruz$seed94,
     main="Number of seedlings in 1994 and 1992",
     xlab="Number of seedlings in 1992",ylab="Number of seedlings in 1994")
```




```
cor(santacruz$seed92,santacruz$seed94)
```

```
## [1] 0.6106537
```

```
nrow(santacruz) #10
```

```
## [1] 10
```

```
santacruz$sampwt <- rep(1,nrow(santacruz))
design0405 <- svydesign(ids = ~1, weights = ~sampwt, data = santacruz)
design0405
```

```
## Independent Sampling design (with replacement)
## svydesign(ids = ~1, weights = ~sampwt, data = santacruz)
```

```
#Ratio estimation using number of seedlings of 1992 as auxiliary variable
sratio3<-svyratio(~seed94, ~seed92,design = design0405)
sratio3
```

```
## Ratio estimator: svyratio.survey.design2(~seed94, ~seed92, design = design0405)
## Ratios=
##           seed92
## seed94 0.2961165
## SEs=
##           seed92
## seed94 0.1152622
```

```
confint(sratio3, df=10-1)
```

```
##                     2.5 %    97.5 %
## seed94/seed92 0.03537532 0.5568577
```

2. 회귀추정 (Regrsession Estimation)

회귀추정량은 svyglm 명령어를 통해 구할 수 있다.

```
data(deadtrees)
head(deadtrees)
```

```
##   photo field
## 1    10    15
## 2    12    14
## 3     7     9
## 4    13    14
## 5    13     8
## 6     6     5
```

```
nrow(deadtrees) # 25
```

```
## [1] 25
```

```
# Fit with survey regression
dtree<- svydesign(id = ~1, weight=rep(4,25), fpc=rep(100,25), data = deadtrees)
myfit1 <- svyglm(field~photo, design=dtree)
summary(myfit1) # displays regression coefficients
```

```
##
## Call:
## svyglm(formula = field ~ photo, design = dtree)
##
## Survey design:
## svydesign(id = ~1, weight = rep(4, 25), fpc = rep(100, 25), data = deadtrees)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.0593     1.3930   3.632   0.0014 **
## photo         0.6133     0.1259   4.870 6.44e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 5.548341)
##
## Number of Fisher Scoring iterations: 2
```

```
confint(myfit1,df=23) # df = 25-2
```

```
##                 2.5 %   97.5 %
## (Intercept) 2.1777362 7.940848
## photo       0.3527717 0.873777
```

```
# Regression estimate of population mean field trees
newdata <- data.frame(photo=11.3)
predict(myfit1, newdata)
```

```
##     link    SE
## 1 11.989 0.418
```

```
confint(predict(myfit1, newdata),df=23)
```

```
##      2.5 %   97.5 %
## 1 11.12455 12.85404
```

```
# Estimate total field tree, add population size in total= argument
newdata2 <- data.frame(photo=1130)
predict(myfit1, newdata2, total=100)
```

```
##     link     SE
## 1 1198.9 41.802
```

```
confint(predict(myfit1, newdata2,total=100),df=23)
```

```
##      2.5 %   97.5 %
## 1 1112.455 1285.404
```

3. 층화표집에서의 비추정량 (Ratio Estimation with Stratified Sampling)

병합비추정량 (Combined ratio estimator)

단순임의표집에서의 비추정과 유사한 방법으로 구할 수 있다. (svydesign 명령어 실행 시 층화임의표접으로 design 하기만 하면 된다.)

```
data(agstrat)
popsize_recode <- c('NC' = 1054, 'NE' = 220, 'S' = 1382, 'W' = 422)
agstrat$popsize <- popsize_recode[agstrat$region]
# input design information for agstrat
dstr <- svydesign(id = ~1, strata = ~region, fpc = ~popsize, weight = ~strwt,
                  data = agstrat)
# now compute the combined estimator of the ratio
combined<-svyratio(~ acres92,~acres87,design = dstr)
combined
```

```
## Ratio estimator: svyratio.survey.design2(~acres92, ~acres87, design = dstr)
## Ratios=
##           acres87
## acres92 0.9899971
## SEs=
##              acres87
## acres92 0.006187757
```

```
# we can get the combined ratio estimator of the population total
# with the predict function
predict(combined,total=964470625)
```

```
## $total
##           acres87
## acres92 954823130
##
## $se
##         acres87
## acres92 5967910
```

분리비추정량 (Separate ratio estimator)

separate = TRUE 옵션을 통해 분리비추정량을 구할 수 있다.

```
separate<-svyratio(~acres92,~acres87,design = dstr,separate=TRUE)
separate
```

```
## Stratified ratio estimate: svyratio.survey.design2(~acres92, ~acres87, design = dstr, separate = TRU
## Ratio estimator: Stratum == "NC"
## Ratios=
##           acres87
## acres92 0.9750666
## SEs=
##              acres87
```

```
## acres92 0.005483458
## Ratio estimator: Stratum == "NE"
## Ratios=
##           acres87
## acres92 0.8956073
## SEs=
##             acres87
## acres92 0.008853011
## Ratio estimator: Stratum == "S"
## Ratios=
##           acres87
## acres92 0.9935483
## SEs=
##            acres87
## acres92 0.01418835
## Ratio estimator: Stratum == "W"
## Ratios=
##          acres87
## acres92 1.011974
## SEs=
##            acres87
## acres92 0.01169809
```

```
#  Define the stratum totals for acres87 as a list:
stratum.xtotals <- list(NC=350474227,NE=22033421,S=280631939,W=311331038)
predict(separate,stratum.xtotals)
```

```
## $total
##           acres87
## acres92 955349448
##
## $se
##         acres87
## acres92 5731438
```