Chapter 3, stratified random sampling

```
#############Chapter 3: Stratified random sampling#################
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
library(SDaA)
data(agstrat)
head(agstrat)   #take a look of the first 6 obsns
```

```
##              county state acres92 acres87 acres82 farms92 farms87 farms82
## 1     PIERCE COUNTY    NE  297326  332862  319619     725     857     865
## 2   JENNINGS COUNTY    IN  124694  131481  139111     658     671     751
## 3      WAYNE COUNTY    OH  246938  263457  268434    1582    1734    1866
## 4  VAN BUREN COUNTY    MI  206781  190251  197055    1164    1278    1464
## 5    OZAUKEE COUNTY    WI   78772   85201   89331     448     483     527
## 6 CLEARWATER COUNTY    MN  210897  229537  213105     583     699     693
##   largef92 largef87 largef82 smallf92 smallf87 smallf82 region   rn
## 1       54       54       42       58       67       48     NC  805
## 2       14       13       14       42       36       38     NC  241
## 3       20       19       16      175      186      184     NC  913
## 4       23       17        9       56       66       55     NC  478
## 5        6        5        5       56       49       48     NC 1028
## 6       34       32       23        8       19       13     NC  496
##     weight
## 1 10.23301
## 2 10.23301
## 3 10.23301
## 4 10.23301
## 5 10.23301
## 6 10.23301
```

```
nrow(agstrat)   #number of rows
```

```
## [1] 300
```

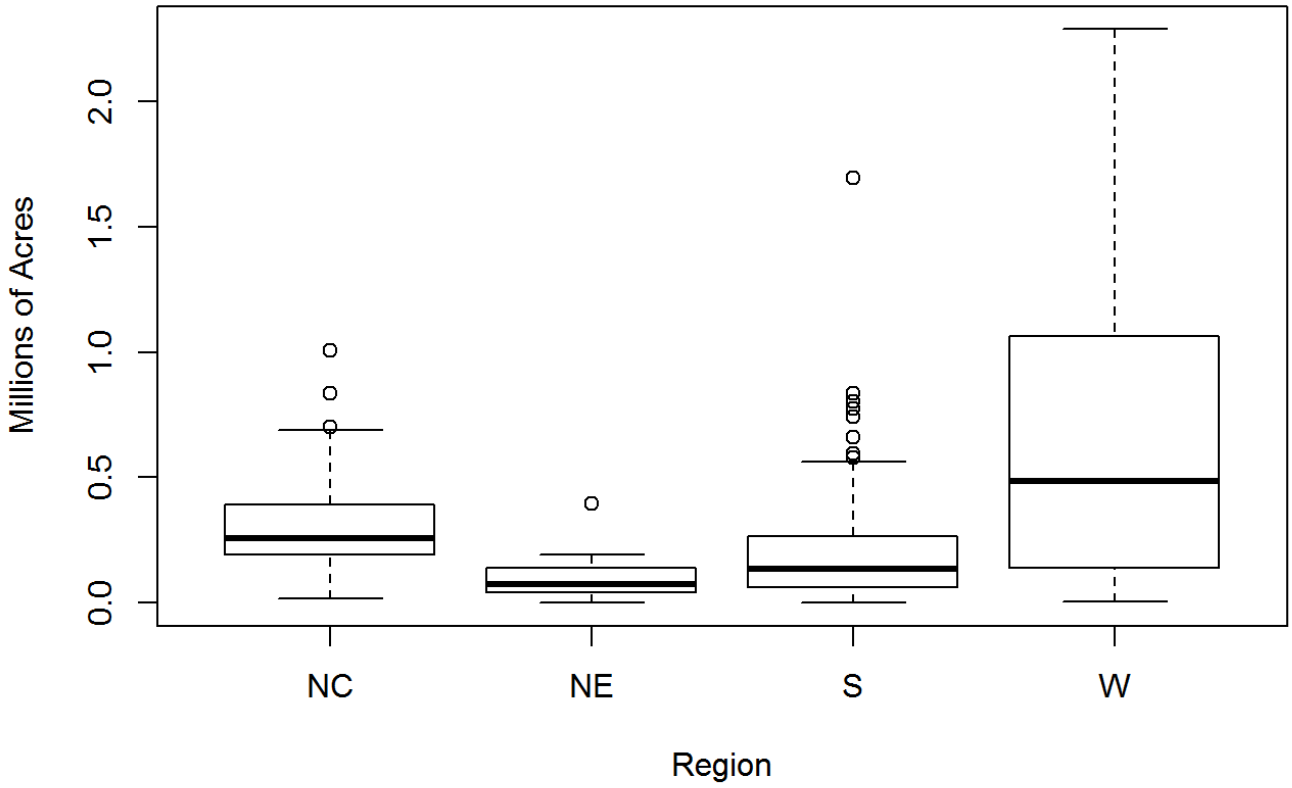
```
agstrat$region #take a look of the four regions, NC, NE, S, W
```

```
##   [1] NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC
##  [24] NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC
##  [47] NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC
##  [70] NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC NC
##  [93] NC NC NC NC NC NC NC NC NC NC NC NE NE NE NE NE NE NE NE NE NE NE NE
## [116] NE NE NE NE NE NE NE NE NE S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [139] S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [162] S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [185] S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [208] S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [231] S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
## [254] S  S  S  S  S  S  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W
## [277] W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W  W
## [300] W
## Levels: NC NE S W
```

```
table(agstrat$region)
```

```
##
##  NC  NE   S   W
## 103  21 135  41
```

```
boxplot(acres92/10^6 ~ region, xlab = "Region", ylab = "Millions of Acres",data =
agstrat)
```




```
#notice the large variability in western region
dstr <- svydesign(id = ~1, strata = ~region, fpc = c(rep(1054,103),rep(220,21),rep
(1382,135),rep(422,41)), weight = ~weight, data = agstrat)
dstr2 <- svydesign(id = ~1, strata = ~region, fpc = c(rep(1054,103),rep(220,21),re
p(1382,135),rep(422,41)), data = agstrat)
dstr  #
```

```
## Stratified Independent Sampling design
## svydesign(id = ~1, strata = ~region, fpc = c(rep(1054, 103),
##     rep(220, 21), rep(1382, 135), rep(422, 41)), weight = ~weight,
##     data = agstrat)
```

```
dstr2 #without weight info, it should give the same results
```

```
## Stratified Independent Sampling design
## svydesign(id = ~1, strata = ~region, fpc = c(rep(1054, 103),
##     rep(220, 21), rep(1382, 135), rep(422, 41)), data = agstrat)
```

```
##calculate mean, SE and confidence interval
smean<-svymean(~acres92, dstr)
smean
```

```
##           mean    SE
## acres92 295561 16380
```

```r
smean2<-svymean(~acres92, dstr2)
smean2
```

```
##           mean    SE
## acres92 295561 16380
```

```r
confint(smean, level=.95,df=296) #note df =n-H = 300-4
```

```
##          2.5 %   97.5 %
## acres92 263325 327796.5
```

```r
##calculate total, SE and CI
stotal<-svytotal(~acres92, dstr)
stotal
```

```
##             total       SE
## acres92 909736033 50417248
```

```r
confint(stotal, level=.95,df=296)
```

```
##             2.5 %     97.5 %
## acres92 810514348 1008957719
```

```r
#calculate S^2 for stratum NC
newdata1<-agstrat[which(agstrat$region=='NC'),]
var(newdata1$acres92)
```

```
## [1] 29618183543
```

```r
##you can calculate the quantities stratum by stratum together
tapply(agstrat$acres92,agstrat$region,mean) #calculate mean for each region
```

```
##        NC        NE         S         W
## 300504.16  97629.81 211315.04 662295.51
```

```r
tapply(agstrat$acres92,agstrat$region,var) #calculate variance for each region
```

```
##           NC           NE            S            W
##  29618183543   7647472708  53587487856 396185950266
```

```r
tapply(agstrat$acres92,agstrat$region,sd) #calculate sd for each region
```

```
##        NC        NE         S         W
## 172099.34  87449.83 231489.71 629433.04
```

```
##calculate var(\bar y_h) for the four regions
svyby(~acres92, ~region, dstr, svymean, keep.var = TRUE)
```

```
##    region   acres92       se
## NC     NC 300504.16 16107.59
## NE     NE  97629.81 18149.49
## S       S 211315.04 18925.35
## W       W 662295.51 93403.65
```

```
#calculate var(\bar y_NC) by direct formula
sqrt(1-103/1054)*172099.34/sqrt(103)
```

```
## [1] 16107.59
```

```
##calculate var(\bar t_h) for the four regions
svyby(~acres92, ~region, dstr, svytotal, keep.var = TRUE)
```

```
##    region   acres92       se
## NC     NC 316731379 16977399
## NE     NE  21478558  3992889
## S       S 292037390 26154840
## W       W 279488705 39416342
```

```
#############Select a stratified random #sampling#################
#install.packages("sampling")
library(sampling)
```

```
##
## Attaching package: 'sampling'
```

```
## The following objects are masked from 'package:survival':
##
##     cluster, strata
```

```
 data(agpop)
 agpop2<-agpop[order(agpop$region),]
idselected<-strata(agpop2,stratanames=c("region"),size=c(3,2,4,1), method="srswor"
)
 idselected
```

```
##      region ID_unit        Prob Stratum
## 30       NC      30 0.002846300       1
## 537      NC     537 0.002846300       1
## 856      NC     856 0.002846300       1
## 1149     NE    1149 0.009090909       2
## 1160     NE    1160 0.009090909       2
## 1463      S    1463 0.002894356       3
## 1619      S    1619 0.002894356       3
## 1877      S    1877 0.002894356       3
## 1912      S    1912 0.002894356       3
## 2908      W    2908 0.002369668       4
```

```
table(idselected$region)
```

```
##
## NC NE  S  W
##  3  2  4  1
```

```
getdata(agpop2,idselected)
```

```
##                 county state acres92 acres87 acres82 farms92 farms87
## 554   DICKINSON COUNTY    IA  202249  211002  227557     554     593
## 1342 OTTER TAIL COUNTY    MN  821073  876319  924670    2509    2925
## 2042     GEAUGA COUNTY    OH   65266   72766   74347     622     702
## 1975  JEFFERSON COUNTY    NY  300559  338401  368352     894    1058
## 1986   ONONDAGA COUNTY    NY  145329  158276  179015     636     772
## 341  OKEECHOBEE COUNTY    FL  351885  384169  381895     418     400
## 497      THOMAS COUNTY    GA  174020  217103  215350     465     507
## 1515  ISSAQUENA COUNTY    MS  113734  121019  111115     109     107
## 1550    SHARKEY COUNTY    MS  181946  177963  210485     125     127
## 1909    DE BACA COUNTY    NM 1343237 1296777 1165799     191     191
##      farms82 largef92 largef87 largef82 smallf92 smallf87 smallf82 region
## 554      652       30       34       31       38       47       43     NC
## 1342    3245      116      107       96       83      120      101     NC
## 2042     753        1        2        2       50       56       54     NC
## 1975    1245       31       25       22       25       32       40     NE
## 1986     835       24       21       22       56       68       77     NE
## 341      326       75       79       80       75       58       31      S
## 497      540       40       39       45       35       35       41      S
## 1515     134       32       36       31        0        4        4      S
## 1550     162       61       56       67        1        5        0      S
## 1909     204       91       86       88        9       22       23      W
##      ID_unit        Prob Stratum
## 554       30 0.002846300       1
## 1342     537 0.002846300       1
## 2042     856 0.002846300       1
## 1975    1149 0.009090909       2
## 1986    1160 0.009090909       2
## 341     1463 0.002894356       3
## 497     1619 0.002894356       3
## 1515    1877 0.002894356       3
## 1550    1912 0.002894356       3
## 1909    2908 0.002369668       4
```

```r
##method could be srswr, systematic etc



insample<-c(sample(1:1054,103,replace=FALSE),
sample(1055:1274,21,replace=FALSE),
sample(1275:2656,135,replace=FALSE),sample(2657:3078,41,replace=FALSE))
strsample<-agpop2[insample,]
```