Exercise 2

Jan-Philipp Kolb, Stefan Zins and Matthias Sand

1 Februar 2016

Example A

Estimation for the European Social Survey under stratified design

- Download ESS for Sweden and Denmark
- Import data to R and merge the two datasets
- Define a `survey` object (stratified design)
- Calculate the unbiased totals for the tv consumption

Download and Import ESS

- Download the ESS dataset for Denmark (Sampling Data and Country File) of the 5th round

Packages for data import

- Use the package `foreign` or `memisc` for import

```
library(foreign)
```

```
library(memisc)
```

Load the ESS dataset and the country file

```
library(foreign)
DK <- read.spss("ESS5DK.sav",to.data.frame=T)
SE <- read.spss("ESS5SE.sav",to.data.frame=T)
```

```
DK <- as.data.frame(DK)
DK$N <- DK$pweight*10000*nrow(DK)

SE <- as.data.frame(SE)
SE$N <- SE$pweight*10000*nrow(SE)
```

```
DK_tv <- data.frame(tvtot=as.character(DK$tvtot),
                    N=DK$N,
                    cntry=as.character(DK$cntry))
SE_tv <- data.frame(tvtot=as.character(SE$tvtot),
                    N=SE$N,
                    cntry=as.character(SE$cntry))


NE <- rbind(DK_tv,SE_tv)
```

Define a survey object

```
library(survey)
```

Define a survey object:

```
svydes_NE <- svydesign(id=~1,strata=~cntry, fpc=~N, data=NE)
```

```
svytable(~tvtot,svydes_NE)
```

Example B

- Load the survey package and the `api` datasets.
- Compute the mean of the Academic Performance Index (2000), asuming SRS
- Use other allocations
- Select a StrSRS from apipop for each allocations.
- Estimate the mean of api00 from different samples (equal, proportional, optimal).

The survey library

Load `survey` library and dataset apistrat

```
library(survey)
```

The dataset apistrat is a sample of schools from apipop stratified by stype.

```
data(api)
```

```
head(apistrat)
```

cds	stype	name	sname	snum
19647336097927	E	Open Magnet: Ce	Open Magnet: Center for Individual (Char	2077
19647336016018	E	Belvedere Eleme	Belvedere Elementary	1622
19648816021505	E	Altadena Elemen	Altadena Elementary	2236
19647336019285	E	Soto Street Ele	Soto Street Elementary	1921
56739406115430	E	Walnut Canyon E	Walnut Canyon Elementary	6140
56726036084917	E	Atherwood Eleme	Atherwood Elementary	6077
56726036055800	E	Township Elemen	Township Elementary	6071
15633216109078	E	Thorner (Dr. Ju	Thorner (Dr. Juliet) Elementary	904

Stratified designs

Assuming the selection within the strata was done by SRS, define a svydesign object that enables you to make unbiased point and variance estimates.

- Estimate the mean of variable `api00`.

```
mean(apistrat$api00)
```

```
## [1] 652.82
```

Allocations

Now you should try different allocations.

Using `stype` as a stratification variable calculate the allocation of a sample of 60 schools from apipop. Use

- equal allocation
- proportional allocation (proportional to nr. of schools)
- optimal allocation (with regard to `api99` allocation)

Select a StrSRS from apipop for each of your allocations.

Equal allocation

```
library(sampling)
Nh <- table(apistrat$stype)

s_equal <- strata(apistrat,"stype",
         size=c(20,20,20),
         method="srswor")

ind <- match(s_equal$stype,names(Nh))
s_equal$N <- Nh[ind]
s_equal$api00 <- apistrat$api00[s_equal$ID_unit]
```

Proportional allocation

```
strSRsample <- function(strind, nh, replace=FALSE){
  Nh <- table(strind)[names(nh)]
  h.id <- split(1:sum(Nh), strind)[names(nh)]


  sam <- mapply( function(x,y) sample(x, y, replace=replace)
                , Nh, nh, SIMPLIFY = F)
  unlist(mapply(function(x,y) x[y]
               , h.id
               , sam, SIMPLIFY = F)
         ,use.names = FALSE)
}
```

```
n <- 2000

nh <- tapply(apipop$api99,apipop$stype,function(x)sum(x)/sum(apipop$api99)*n)
samp <- strSRsample(apipop$stype,nh)

ap <- apipop[samp,]
```

Estimation

- Estimate again the mean of api00 from your three different samples.

```
svy_equal <- svydesign(id=~1,strata=~stype, fpc=~N, data=s_equal)

svymean(s_equal$api00,svy_equal)
```

```
##        mean     SE
## [1,] 684.95 12.444
```

```
library(devtools)

install_github("BernStZi/SamplingAndEstimation/r/sampaest",ref="short")
```