Sampling and Estimation - Exercise 2

Jan-Philipp Kolb, Stefan Zins and Matthias Sand

2 Februar 2016

Exercise 2.A

Estimation under a stratified design

- Download the ESS for Sweden and Denmark (round 5)
- Import data to R and combine the two datasets
- Define a `survey` object (stratified design)
- Calculate the combined total for the tv consumption (`tvtot`) and compare it with the totals in Sweden and Denmark

Solution Exercise 2.A

- Download the ESS for Sweden and Denmark (round 5)

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

Use the package `foreign` for import

```
library(foreign)
```

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)
```

Preparation - finite population correction

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

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

Combine the two datasets

```
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 the survey object:

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

Calculate the total for the tv consumption

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

Exercise 2.B

- Load the `survey` package and the `api` datasets.

```
library(survey)
data(api)
```

- The dataset `apistrat` is a sample of schools from `apipop` stratified by `stype`. Assuming the selection within the strata was done by SRS, define a survey object (`svydesign`) and calculate a point and variance estimate for the mean of `api00`.

Exercise 2.B

- Using `stype` again as a stratification variable try different allocations for stratified sample. Calculate the allocation of a sample of 60 schools from `apipop` using equal, proportional and optimal allocation. The proportional allocation should be proportional to the number of schools within the strata and the optimal alloaction should be optimal with regard to `api99`.

Exercise 2.B

- Select a StrSRS from `apipop` for each allocation.
- Estimate again the mean of `api00` from all three samples and compare the results.