Exercise 3

Stefan Zins, Matthias Sand and Jan-Philipp Kolb

2 Februar 2016

Exercise 3a

1. Download the dataset for Germany of the 5th ESS-Round (SDDF File and Sampling Data)
2. Create a `svydesign` object to estimate the mean of the variable `agea`
3. To acknowledge that the sample has been collected by a multi stage design, estimate the design effect of your estimate above using the PSU-Indicator variable (Use the model based approach described on slide 20 of today's lecture)

   Advice: the variable `PSU` has to be a factor

4. Calculate the effective sample size

Obtaining MSB, MSW and b^*

```
Ger.d <- read.spss("ESS5DE.spss/ESS5DE.sav",
                   to.data.frame = TRUE,
                   use.value.labels = TRUE)
Ger.ctry <- read.spss("ESS5_DE_SDDF.spss/ESS5_DE_SDDF.por",
                      to.data.frame = TRUE,
                      use.value.labels = TRUE)

colnames(Ger.d)[5] <- "IDNO"
Ger <- merge(Ger.d,Ger.ctry,by="IDNO", all.x = TRUE)
Ger$PSU <- as.factor(Ger$PSU)
n <- nrow(Ger)
L <- length(unique(Ger$PSU))
```

```
## deffc
b.star <- sum(tapply(Ger$dweight,Ger$PSU,
                function(x)sum(x)^2))/sum(Ger$dweight^2)
# Calculate an anova for the regression model Age by PSU
# (Could also be any other Variable)
lin.mod <- lm(as.numeric(Ger$agea)~Ger$PSU)
SS <- anova(lin.mod)
#  MSB and MSW are the means of SSB and SSW
MSB <- SS$`Mean Sq`[1]
MSW <- SS$`Mean Sq`[2]
```

- Execute the following R-Script: to generate a Multistage- and a Cluster- Sample for the belgianmunicipalities data set
- Your workspace now contains the datasets `income`, `Data.be` and `Data.be2`. `income` resembles a dataset that
- Estimate the mean income from both samples using the `survey` package and compare the results