

Computer Exercise 10 for Sample Survey**Important points**

- You will calculate the population mean and variance for a finite population of Y_i .
- You will draw a simple random sample (SRS) and compare the mean and its standard error estimate from the SRS with the population values.
- You will post-stratify the population and sample according to the level of Y_i and compare the sample proportions (weights) across strata to population proportions (weights).
- You will estimate the overall mean and its standard error from the post-stratified random sample and compare them to those estimates using a SRS and the population values.
- You will estimate the mean and its standard error for stratum 1 from the post-stratified random sample using two methods and compare between them.

Useful R commands

- To draw a SRS of size n from a vector y of observations, the R command is
`sam=sample(y,n,replace = FALSE, prob = NULL)`.
- To post-stratify `sam` into subsamples according to levels: 0, 10000, 20000, 30000, 40000 say, the R commands are

```
sams=sort(sam)
n1=length(which(sams<=10000))
n2=length(which(sams>10000 & sams<=20000))
n3=length(which(sams>20000 & sams<=30000))
n4=length(which(sams>30000 & sams<=40000))
nv=c(n1,n2,n3,n4)
nv
sum(nv)
inds=rep(1:4,nv)
sam1=sams[inds==1]
sam2=sams[inds==2]
sam3=sams[inds==3]
sam4=sams[inds==4]
```

Practice Problems

1. From the file `p33.dat` (set `y=p33.dat`) which contains a finite population of $N = 1000$ incomes (Y_i), plot the skew population distribution using `hist` and calculate
 - (a) the population mean income, \bar{Y} ,
 - (b) the variance of income, S^2 , and
 - (c) the theoretical value of the standard error $SE(\bar{y})$ for a SRS without replacement of size $n = 100$.
2. Perform the following to compare the mean and its standard error estimate from a SRS with the corresponding population values.
 - (a) Draw a SRS of size $n = 100$ without replacement and report the sample. Use `set.seed(12345)` to set 12345 as the seed for drawing sample. Report the sample and plot the sample distribution using `hist`.
 - (b) Estimate the mean \hat{Y}_{srs} and its standard error estimate $se(\hat{Y}_{srs})$ based on the SRS.
 - (c) Compare the mean estimate \hat{Y}_{srs} with the population mean \bar{Y} and the standard error estimate $se(\hat{Y}_{srs})$ with the theoretical value $SE(\bar{y})$ and comment.
3. Perform the following to compare the population and sample proportions (weights) across strata.
 - (a) Post-stratify the population according to the size of income Y_i in $(0,10000]$, $(10000,20000]$, $(20000,30000]$ and $(30000,40000]$. Report the strata sizes and proportions (weights) across strata. Assume these strata sizes are known.
 - (b) Post-stratify the sample according the strata in (a). Report the post-stratified samples, the sample sizes and proportions (weights) across strata.
 - (c) Compare the sample proportions (weights) with population proportions (weights).
4. Based on the post-stratified sample, estimate the overall mean \hat{Y}_{ps} and its standard error estimate $se(\hat{Y}_{ps})$. This estimator adjusts the proportion of sample in each stratum to the population proportion.
5. Based on the post-stratified sample, estimate the mean $\hat{Y}_{ps,1mj}$ and its standard error estimate $se(\hat{Y}_{ps,1mj})$ for stratum 1 using both methods, $j = 1, 2$ (P.17 of lecture note).