Computer Exercise 11 for Sample Survey

Important points

- You will stratify the population into strata according to certain variable.

- You will draw a stratified random sample using *Neyman* allocation.

- You will estimate the mean and its standard error from the stratified random sample and compare them to those using a SRS.

Useful R commands

- To draw a SRS of size n from a vector y of observations, the R command is

  ```
  sample(y,n,replace = FALSE, prob = NULL)).
  ```

- To stratify a population into subpopulations, ys1, ys2, ..., of sizes, N1, N2, ..., in ascending order of y, the R commands are

  ```
  ys=sort(y)
  N=length(y)
  N1=length(which(ys<=10000))
  N2=length(which(ys>10000 & ys<=20000))
  N3=length(which(ys>20000 & ys<=30000))
  N4=length(which(ys>30000 & ys<=40000))
  Nv=c(N1,N2,N3,N4)
  Nv
  sum(Nv)
  Wv=Nv/N
  indy=rep(1:4,Nv)
  ys1=ys[indy==1]
  ys2=ys[indy==2]
  ys3=ys[indy==3]
  ys4=ys[indy==4]
  ```

Practice Problems

1. Stratify the population into subpopulations certain variable.

(a) Read the file `p33.dat` into `y` which contains a finite population of $N = 1000$ incomes (Y_i).

(b) Sort the population in ascending order of Y_i and stratify the resulting population `ys1, ys2, ...` into four strata according to levels of income Y_i in (0,10000], (10000,20000], (20000,30000] and (30000,40000].

(c) Report the vector of strata sizes `Nv`, means `Meanv` and variances `S2v`.

2. Calculate the vector of required sample sizes `nv` and `nv2` for stratified samples if the total sample size is $n = 100$ under:

(a) Neyman allocation;

(b) proportional allocation.

Compare the two sets of sample sizes.

3. Draw SRSs (use `set.seed(12345)`) of sizes `nv` from each stratum under *Neyman* allocation.

(a) Report the stratified random samples `sam1, sam2, ...`.

(b) Report the vector sample means `meanv` and variances `s2v` across strata.

(c) Estimate the mean $\widehat{\bar{Y}}_{st}$ and its standard error $se(\widehat{\bar{Y}}_{st})$ using the stratified random samples.

4. Draw one SRS (no need to reset seed) of size 100 from the overall population `y` and store it in `sam`.

(a) Estimate the mean $\widehat{\bar{Y}}_{srs}$ and its standard error $se(\widehat{\bar{Y}}_{srs})$.

(b) Compare the estimates with those from Question 3(c) and comment.