Calculating estimates for stratified random samples

Here is some R code that finds the stratified estimator of the population mean under simple random sampling within each stratum and an estimate of its standard deviation.

I begin by calling a very simple stratified population that I have constructed.

```
> library("RCurl")
> x<-getURL("http://users.stat.umn.edu/~gmeeden/classes/5201/moredata/stratpop1.txt")
> X<-read.table(textConnection(x), header=T)
> names(X)

[1] "y"     "strat"

> X

       y strat
1  10.21     b
2   6.05     a
3   8.80     b
4   3.35     a
5   5.94     a
6  11.22     b
7   7.87     b
8   5.57     a
9  13.02     b
10  9.69     b
11  6.60     a
12  9.23     b

> y<-X[,1]
> strat<-X[,2]
> foo<-split(y,strat)
```

You see there are two strata a and b, the first has 5 units and the second has 7 units. To make it easier to work with the strata I created foo using the split command. In R terminology foo is a list where each element of the list is a stratum. The sapply commend lets me find information about each stratum in the population.

```
> foo

$a
[1] 6.05 3.35 5.94 5.57 6.60

$b
[1] 10.21  8.80 11.22  7.87 13.02  9.69  9.23
```

```
> foo[[1]]

[1] 6.05 3.35 5.94 5.57 6.60

> stratsizes<-sapply(foo,length)
> stratsizes

a b
5 7

> popsize<-length(X[,1])
> popsize

[1] 12

> WW<-stratsizes/popsize
> WW

        a         b
0.4166667 0.5833333

> stratmns<-sapply(foo,mean)
> stratmns

       a        b
 5.50200 10.00571
```

Next I find a random sample of size 2 from stratum a and size 3 from stratum b. I then use the sapply command to find the strata means for the sample. Then I get the point estimate of the population mean. Next I find the estimated variance of our estimate and its standard error. Finally I find an approximate 95% confidence interval for the mean. Note this makes no sense here but it is included for completeness.

```
> smpsizes<-c(2,3)
> foo.smp<-list()
> set.seed(1962)
> for(i in seq(along=foo)){
+  foo.smp[[i]]<-sample(foo[[i]],smpsizes[i])
+ }
> names(foo.smp)<-names(foo)
> foo.smp

$a
[1] 6.05 5.94

$b
[1] 7.87 9.23 9.69
```

```
> smpmns<-sapply(foo.smp,mean)
> smpmns

    a     b
5.995 8.930

> est<-sum(WW*smpmns)
> est

[1] 7.707083

> smpvar<-sapply(foo.smp,var)
> smpvar

      a       b
0.00605 0.89560

> ratiosmptopop<-smpsizes/stratsizes
> ratiosmptopop

        a         b
0.4000000 0.4285714

> estvarofest<-sum((WW^2)*(1-ratiosmptopop)*(1/smpsizes)*smpvar)
> SEofest<-sqrt(estvarofest)
> SEofest

[1] 0.2415849

> lwbd<-est - 1.96*SEofest
> upbd<-est + 1.96*SEofest
> c(lwbd,upbd)

[1] 7.233577 8.180590
```

These commands should help you do HW assignment 4. Note for some of the bigger problems you will not want to look at all of X.