Example 5.9

All the 80 farms in a population are stratified by farm size. The expenditure on the insecticides used during the last year by each farmer is presented in table 5.10 below :

Table 5.10 Expenditure (in '00 rupees) on insecticides used

Large farmers		Medium farmers				Small farmers		
75	76	55	40	51	28	35	31	26
65	79	45	38	55	47	28	38	32
86	62	35	33	41	61	36	42	18
57	92	30	43	48	35	40	33	16
45	50	42	53	54	31	25	29	
69	48	38	37	36	23	18	25	
48	77	40	52	44		28	35	
60	60	36	39	47		32	26	
55	64	48	46	39		13	30	
66	58	46	42	41		19	37	

a) Compute the overall population mean $\bar{Y}$ and the population mean square S^2.
b) Verify that population mean equal the weighted average of the means of the strata.
c) Determine the number of units to be selected from each stratum with proportional allocation, equal allocation, and Nyman allocation of a sample of size n=24.
d) Determine the variance of the sample mean for stratified sampling with proportional allocation, equal allocation, and Nyman allocation of a sample of size n=24. Assume that the sampling is WOR.
e) Compare the relative precision of stratified sample mean $\bar{y}_{str}$ based on each of the above mentioned allocations, with respect to the simple random sample mean $\bar{y}_{srs}$.
f) Select a stratified sample of 24 farmers by using Proportional allocation.(use R program)

```
> #Example 5.9
> #convert Excl file to CSV , then read it in R
> Farms= read.csv("~/Desktop/Stat 331/farms.csv", header = T)
>
> Y<-Farms$expenditure
> stratum<-Farms$farms
> table(stratum) #obtain size of each stratum.
stratum
 1  2  3
20 36 24

> ## compute population size, mean for each stratum 1,2 and 3.
> #sample size, mean and var for strata 1
> Y1<-Y[stratum==1]
```

```
> N1=length(Y1)
> Y1_bar=mean(Y1)
> V1=var(Y1)
> #sample size, mean and var for strata 2
> Y2<-Y[stratum==2]
> N2=length(Y2)
> Y2_bar=mean(Y2)
> V2=var(Y2)
> #sample size, mean and var for strata 3
> Y3<-Y[stratum==3]
> N3=length(Y3)
> Y3_bar=mean(Y3)
> V3=var(Y3)
> N=N1+N2+N3 #Population size
> V1;   V2   ; V3
[1] 169.5158
[1] 70.56111
[1] 61.44928
>
> ## a.pop mean
> pop_mean=mean(c(Y1,Y2,Y3))  #or pop_mean=mean(Y)
> pop_mean
[1] 43.7875

> #b.verify that population mean equal the weighted average of the means of the strata
> Wh=c(N1/N,N2/N,N3/N)
> mean_str=Wh[1]*Y1_bar+Wh[2]*Y2_bar+Wh[3]*Y3_bar
> mean_str
[1] 43.7875

> #a. pop variance
> pop_v=var(c(Y1,Y2,Y3))  #or pop_v =var(Y)
> pop_v
[1] 268.6758
>
> # c. determine the variance of the sample mean for SRS with the same size 24
> n=24
> Var_srs=((N-n)/(N*n))*pop_v
> Var_srs
[1] 7.836377
>
> #determine the variance of the sample mean for stratified sampling with Equal allocation
> nh=24/3
> nh
[1] 8
> n1=n2=n3=nh
```

```
>
> #d. variance
> Var_str.E=((Wh[1])^2)*((N1-n1)/N1)*(V1/n1)+((Wh[2])^2)*((N2-
n2)/N2)*(V2/n2)+((Wh[3])^2)*((N3-n3)/N3)*(V3/n3)
> Var_str.E
[1] 2.644647
>
> # e.Relative efficiency
> RE= Var_srs/ Var_str.E*100
> RE
[1] 296.3109
>
> ## c. determine the variance of the sample mean for stratified sampling
> #with proportional allocation
> n1_p=round(Wh[1]*n, digits = 0) #Rounds a value to the nearest integer.
> n2_p=round(Wh[2]*n, digits = 0)
> n3_p=round(Wh[3]*n, digits = 0)
> n1_p;
[1] 6
> n2_p;
[1] 11
> n3_p;
[1] 7
> np=c(n1_p,n2_p,n3_p)
>
> ## d. determine the variance of the sample mean for stratified sampling
> #with proportional allocation of a sample of size n=12
> #variance
> Var_str.p=((Wh[1])^2)*((N1-n1_p)/N1)*(V1/n1_p)+((Wh[2])^2)*((N2-
n2_p)/N2)*(V2/n2_p)+((Wh[3])^2)*((N3-n3_p)/N3)*(V3/n3_p)
> Var_str.p
[1] 2.69774
>
> # e.Relative efficiency
> RE= Var_srs/ Var_str.p*100
> RE
[1] 290.4794
>
> ## c. determine the variance of the sample mean for stratified sampling
> #with Neyman allocation
> Nh=c(N1,N2,N3)
> Vh=sqrt(c(V1,V2,V3))
>   Nunv=sum(Nh[1]*Vh[1]+Nh[2]*Vh[2]+Nh[3]*Vh[3])
> Nunv
[1] 750.9339
> n1_o=round((Nh[1]*Vh[1]*n)/Nunv, digits = 0)
```

```
> n1_o
[1] 8
> n2_o=round((Nh[2]*Vh[2]*n)/Nunv, digits = 0)
> n2_o
[1] 10
> n3_o=round((Nh[3]*Vh[3]*n)/Nunv, digits = 0)
> n3_o
[1] 6
>
> #d.variance
> Var_str.o=((Wh[1])^2)*((N1-n1_o)/N1)*(V1/n1_o)+((Wh[2])^2)*((N2-
n2_o)/N2)*(V2/n2_o)+((Wh[3])^2)*((N3-n3_o)/N3)*(V3/n3_o)
> Var_str.o
[1] 2.517866
>
> #e.Relative efficiency
> RE= Var_srs/ Var_str.o*100
> RE
[1] 311.2309
```

f) Select a stratified sample of 24 farmers by using Proportional allocation.

```
> RNGkind(sample.kind = "Rejection")
> set.seed(200)
> sample1=sample(Y1,6)
> sample1
[1] 69 60 50 60 48 79
> s1=var(sample1)
> sample2=sample(Y2,11)
> sample2
 [1] 36 39 44 35 38 48 39 42 55 61 36
> s2=var(sample2)
> sample3=sample(Y3,7)
> sample3
[1] 40 18 42 18 26 36 31
> s3=var(sample3)
> str_mean=Wh[1]*mean(sample1)+Wh[2]*mean(sample2)+Wh[3]*mean(sample3)
> str_mean
[1] 43.64286
>
> svar_str.p=((Wh[1])^2)*((N1-n1_p)/N1)*(s1/n1_p)+((Wh[2])^2)*((N2-
n2_p)/N2)*(s2/n2_p)+((Wh[3])^2)*((N3-n3_p)/N3)*(s3/n3_p)
> svar_str.p
[1] 2.79218
```

5.14 The list of all the 50,000 adults in a town was available. In order to estimate proportion of literate adults (educated up to at least 8th grade), the population was stratified into 3 strata with respect to age. A WOR random sample of size 500 persons was drawn using proportional allocation. The sample size allocation to each stratum and the number of literate persons recorded in sample of size n_h, h = 1, 2, 3, are given below :

Age group (years)	Persons	n_h	Literate persons
20-40	25600	256	243
40-60	18100	181	144
60 and over	6300	63	31

Compute the estimate of proportion of literate persons in the town, and construct confidence interval for it.

stratum I	stratum II	stratum III
$N_1 = 25600$	$N_2 = 18100$	$N_3 = 6300$
$n_1 = 256$	$n_2 = 181$	$n_3 = 63$
$x_1 = 243$	$x_2 = 144$	$x_3 = 31$
$w_1 = \frac{25600}{50000} = 0.512$	$w_2 = \frac{18100}{50000} = 0.362$	$w_3 = \frac{6300}{50000} = 0.126$
$p_1 = \frac{243}{256} = 0.9492$	$p_2 = \frac{144}{181} = 0.7956$	$p_3 = \frac{31}{63} = 0.4921$

$N = N_1 + N_2 = 50000$

The estimate of the required overall population proportion is

$$p_{str} = \sum_{h=1}^{L} w_h p_h \quad = w_1 p_1 + w_2 p_2 + w_3 p_3$$

$$= (0.512)(0.9492) + (0.362)(0.7956) + (0.126)(0.4921) = 0.8360$$

The estimate of variance is

$$v(p_{str}) = \sum_{h=1}^{L} w_h^2 \frac{p_h q_h}{n_h - 1} \frac{N_h - n_h}{N_h}$$

$$= 4.0414$$

The confidence interval for population proportion is

$$p_{str} \mp Z_{1-\frac{\propto}{2}} se(p_{str})$$

$$0.8360 \mp 2\left(\sqrt{4.0414}\right)$$

$$[0.8057\ , 0.8663\ ]$$

To summarize, the sample estimate of proportion indicates that about 83.6 percent is literate with CI between

$$[0.8057\ , 0.8663\ ]$$

```
> #Q2
> N1=25600
> N2=18100
> N3=6300
> N=N1+N2+N3
> n1=256
> x1=243
> (W1=N1/N)
[1] 0.512
> (p1=x1/n1)
[1] 0.9492188
> (q1=1-p1)
[1] 0.05078125
>
>
> n2=181
> x2=144
> (W2=N2/N)
[1] 0.362
> (p2=x2/n2)
[1] 0.7955801
> (q2=1-p2)
[1] 0.2044199
>
> n3=63
> x3=31
> (W3=N3/N)
[1] 0.126
> (p3=x3/n3)
[1] 0.4920635
> (q3=1-p3)
[1] 0.5079365
> str_mean_prop=W1*p1+W2*p2+W3*p3
> str_mean_prop
```

```
[1] 0.836

> Var_str_prop=((W1^2)*(N1-n1)*p1*q1/(N1*(n1-1))+((W2^2)*((N2-n2)*p2*q2/N2*(n2-1)))
+ +((W3^2)*((N3-n3)*p3*q3/N3*(n3-1))))
> Var_str_prop
[1] 0.0002296334
```