Example 5.1

An assignment was given to four students attending a sample survey course. The problem was to estimate the average time per week devoted to study in Punjab Agricultural University (PAU) library by the students of this university. The university is running undergraduate, master's degree and doctoral programs. Number of students registered for the three programs is 1300, 450, and 250 respectively. Since the value of the study variable is likely to differ considerably with the program, the investigator divided the population of students into 3 strata: undergraduate program (stratum I), master's program (stratum II), and doctoral program (stratum III). First of the four students selected WOR simple random samples of sizes 20, 10, and 12 students from strata I, II, and III respectively, so that, the total sample is of size 42. The information about weekly time devoted in library is given in table 5.1.

Table 5.1 Time (in hours) devoted to study in the university library during a week

Stratum I			Stratum II		Stratum III	
0	1	9	12	6	10	24
4	4	4	9	10	14	15
3	3	6	11	9	20	14
5	6	1	13	11	11	18
2	8	2	8	7	16	19
0	10	3			13	20
3	2					

Estimate the average time per week devoted to study by a student in PAU library. Also, build up the confidence interval for this average.

```
> #  Read data CSV file in R as
> LBhour= read.csv("~/Desktop/Stat 331/LBhour.csv", header = T)
> #please change the code here to your local address where you save the file
> #hint: use "/" if "\" shows in the local address and does not work.
> LBhour
> y<-LBhour$HOUR
> stratum<-LBhour$PROGRAM
> y
> stratum
> #Following this helps you look at mean and variance across stratum.
> #sample size, mean and var for strata 1
> y1<-y[stratum==1]
> y1
 [1] 0 4 3 5 2 0 3 1 4 3 6 8 10 2 9 4 6 1 2 3
```

```
> n1=length(y1)
> n1
[1] 20
> y1_bar=mean(y1)
> y1_bar
[1] 3.8
> v1=var(y1)
> v1
[1] 7.957895
>
> #sample size, mean and var for strata 2
> y2<-y[stratum==2]
> y2
 [1] 12  9 11 13  8  6 10  9 11  7
> n2=length(y2)
> n2
[1] 10
> y2_bar=mean(y2)
> y2_bar
[1] 9.6
> v2=var(y2)
> v2
[1] 4.933333
>
> #sample size, mean and var for strata 3
> y3<-y[stratum==3]
> y3
 [1] 10 14 20 11 16 13 24 15 14 18 19 20
> n3=length(y3)
> n3
[1] 12
> y3_bar=mean(y3)
> y3_bar
[1] 16.16667
> v3=var(y3)
> v3
[1] 17.06061
>
> N1=1300
> N2=450
> N3=250
> N=N1+N2+N3
> N
[1] 2000
>
> #determine an estimate of the population mean by stratified sample
> W1=N1/N; W2=N2/N ; W3=N3/N
> W1; W2; W3
```

```
[1] 0.65
[1] 0.225
[1] 0.125
> str_mean=(W1*y1_bar+W2*y2_bar+W3*y3_bar)
> str_mean
[1] 6.650833
>
> ##OR We can use following code
> Wh=c(N1/N,N2/N,N3/N)
> Wh[1]*y1_bar+ Wh[2]*y2_bar+ Wh[3]*y3_bar
[1] 6.650833
>
> #Variance of sample mean =sum(((1-fh)/nh)*(Wh^2*Vh^2), fh=nh/Nh
> var_str=((W1)^2)*((N1-n1)/N1)*(v1/n1)+((W2)^2)*((N2-n2)/N2)*(v2/n2)+((W3)^2)*((N3-
n3)/N3)*(v3/n3)
> var_str
[1] 0.2110923
>
> CIU=str_mean+2*(sqrt(var_str))
> CIU
[1] 7.569729
> CIL=str_mean-2*(sqrt(var_str))
> CIL
[1] 5.731938
> #estimation for total and its variance
> t_hat=N*str_mean
> t_hat
[1] 13301.67
>
> Var_t=(N^2)*var_str
> Var_t
[1] 844369
>
> #95%CI for  total estimation
> CI_totU=t_hat+2*sqrt(Var_t)
> CI_totU
[1] 15139.46
> CI_totL=t_hat+2*sqrt(Var_t)
> CI_totL
[1] 15139.46
>
```

5.7 An insurance company's records show that out of the total of 500 claims, 280 are major claims (from Rs 1000 to Rs 2500) and 220 are minor (below Rs 1000). A WOR simple random sample of 10 claims was drawn from each category (stratum), and claim amounts were recorded as :

Stratum I : 1200, 1600, 1800, 1400, 1980, 2110, 2440, 1660, 1790, 1910
Stratum II : 720, 880, 760, 660, 790, 840, 550, 960, 640, 800

Estimate the total amount of all the 500 claims, and construct the confidence interval for it.

Stratum 1	Stratum 2
$N_1 = 280$	$N_2 = 220$
$n_1 = 10$	$n_2 = 10$
$W_1 = 0.56$	$W_2 = 0.44$
$\bar{y}_1 = 1789$	$\bar{y}_2 = 760$
$s_1^2 = 125410$	$s_2^2 = 14822.22$

$$\bar{y}_h = \frac{\sum_{i=1}^{n} y_{hi}}{n_h};\ s_h^2 = \frac{\sum_{i=1}^{n}(y_{hi} - \bar{y}_{hi})^2}{n_h - 1};\ \mathrm{W_h} = \frac{\mathrm{N_h}}{\mathrm{N}};\ \mathrm{N} = \mathrm{N_1} + \mathrm{N_2} = 500$$

The estimate of average amount of claims is

$\bar{y}_{str} = W_1\bar{y}_1 + W_2\bar{y}_2 = 1336.24$

Also, the variance of the mean is

$$v(\bar{y}_{str}) = \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{n_h}\frac{N_h - n_h}{N_h} = 4066.313$$

we obtain the limits of confidence interval as

$\bar{y}_{str} \pm Z_{1-\propto/2}\sqrt{v(\bar{y}_{str})}$; $Z_{1-\propto/2} \approx 2$

$$1336.24 \pm 2\sqrt{4066.313}$$

$$[1208.705, 1463.775\ ]$$

We are 95% confidence that the average amount of claims is between $[1208.705, 1463.775\ ]$

Note: Multiply these limits by N to obtain the limits for the population total (T). OR

$$y'_{str} = N\,\bar{y}_{str} = 500\ (1336.24) = 668120$$

Estimator of the variance of population total are obtained by

$v(y'_{str}) = N^2\ v(\bar{y}_{str})$

$v(y'_{str}) = (500)^2(4066.313) = 1016578267$

95%CI for total estimation

$y'_{str} \pm Z_{1-\propto/2} \sqrt{v(y'_{str})}$

$$668120 \pm 2 \sqrt{1016578267}$$

$$[604352.4 \ , 731887.6]$$

Code:

```
> #Q5.7
> y1=c(1200,1600,1800,1400,1980,2110,2440,1660,1790,1910)
> n1=length(y1)
> n1
[1] 10
> y1_bar=mean(y1)
> y1_bar
[1] 1789
> v1=var(y1)
> v1
[1] 125410
> y2=c(720,880,760,660,790,840,550,960,640,800)
> n2=length(y2)
> n2
[1] 10
> y2_bar=mean(y2)
> y2_bar
[1] 760
> v2=var(y2)
> v2
[1] 14822.22
> N1=280
> N2=220
> N=N1+N2
```

```
> N
[1] 500
> #determine an estimate of the population mean by stratified sample
> W1=N1/N; W2=N2/N
> W1; W2
[1] 0.56
[1] 0.44
> str_mean=(W1*y1_bar+W2*y2_bar)
> str_mean
[1] 1336.24
> #Variance of sample mean =sum(((1-fh)/nh)*(Wh^2*Vh^2), fh=nh/Nh
> var_str=((W1)^2)*((N1-n1)/N1)*(v1/n1)+((W2)^2)*((N2-n2)/N2)*(v2/n2)
> var_str
[1] 4066.313
> CIU=str_mean+2*(sqrt(var_str))
> CIU
[1] 1463.775
> CIL=str_mean-2*(sqrt(var_str))
> CIL
[1] 1208.705
> #estimation for total and its variance
> t_hat=N*str_mean
> t_hat
[1] 668120
> Var_t=(N^2)*var_str
> Var_t
[1] 1016578267
> #95%CI for  total estimation
> CI_totU=t_hat+2*sqrt(Var_t)
> CI_totU
```

```
[1] 731887.6
> CI_totL=t_hat-2*sqrt(Var_t)
> CI_totL
[1] 604352.4
```