**MODERN COLLEGE OF ARTS,SCI. & COMM. PUNE-05. (Autonomous)**

**DEPARTMENT OF STATISTICS.**

**M.Sc.( I )- ST-17**

**EXPT.NO. 15 Date:**

**Sub. date:**

**TITLE : Cluster sampling with equal cluster size**

1. For studying the cultivation practices and yield of apple ,a pilot sample survey was conducted in a district of Himachal Pradesh (India).The yield (in Kilogram’s ) of 15 clusters of 4 trees each, selected at random out of 308 bearing trees in a village , are given below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster  No. | Tree No. | | | |
| 1 | 2 | 3 | 4 |
| 1 | 5.53 | 4.84 | 0.69 | 15.79 |
| 2 | 26.11 | 10.93 | 10.08 | 11.18 |
| 3 | 11.08 | 0.65 | 4.21 | 7.56 |
| 4 | 12.66 | 32.52 | 16.92 | 37.02 |
| 5 | 0.87 | 3.56 | 4.81 | 27.54 |
| 6 | 6.40 | 11.68 | 40.05 | 5.12 |
| 7 | 54.21 | 34.63 | 52.55 | 37.20 |
| 8 | 1.24 | 35.97 | 29.54 | 25.28 |
| 9 | 37.94 | 47.07 | 19.64 | 28.11 |
| 10 | 54.92 | 17.69 | 26.24 | 6.77 |
| 11 | 25.52 | 38.10 | 24.74 | 1.90 |
| 12 | 45.98 | 5.17 | 1.17 | 6.53 |
| 13 | 7.13 | 34.35 | 12.18 | 9.86 |
| 14 | 14.23 | 16.89 | 28.93 | 21.70 |
| 15 | 3.53 | 40.76 | 5.15 | 1.25 |

Estimate

1. The average yield per tree as well as the production of apple in the village and their standard errors.
2. The intra-cluster correlation coefficient between trees within clusters.
3. The efficiency of cluster sampling as compared to simple random sampling.

Solution :

Formulas/ Notations

1. N= Total numbers of clusters in population
2. M=cluster size
3. n=sample size
4. ӯi.= mean of ith cluster
5. (ӯn.)cl= =A.M.of n clusters.
6. Var((ӯn.)cl)= Variance of clustere sampling .=( )\*{1+(M-1)ρhat}

ρhat= The intra class correlation between the elements within a cluster.

=

sum of square between the cluster mean in the sample.

= =

Sw2=Mean sum of square within clusters in sample

=

1. E= (Relative Efficiency of cluster sampling with SRSWOR )

=Var(ӯSRS)/Var((ӯn.)cl)

If E>1;(Cluster sampling is better than SRSWOR.)

If E<1;(SRSWOR is better than Cluster sampling.)

If E=1;(Both are equally likely.)

STEPS:-

1. Divide the whole population into clusters according to well define rule.
2. Treat the clusters as sampling units.
3. Choose sample of clusters by using SRSWOR.

**Q1**. x=read.table("clipboard",sep="\t",header=TRUE);x

or x=read.csv(file.choose(),header =TRUE,sep=',');x

. x1 x2 x3 x4

1 5.53 4.84 0.69 15.79

2 26.11 10.93 10.08 11.18

3 11.08 0.65 4.21 7.56

4 12.66 32.52 16.92 37.02

5 0.87 3.56 4.81 27.54

6 6.40 11.68 40.05 5.12

7 54.21 34.63 52.55 37.20

8 1.24 35.97 29.54 25.28

9 37.94 47.07 19.64 28.11

10 54.92 17.69 26.24 6.77

11 25.52 38.10 24.74 1.90

12 45.98 5.17 1.17 6.53

13 7.13 34.35 12.18 9.86

14 14.23 16.89 28.93 21.70

15 3.53 40.76 5.15 1.25

> N=77

> n=15

> M=4

> m=as.matrix(x);m

x1 x2 x3 x4

[1,] 5.53 4.84 0.69 15.79

[2,] 26.11 10.93 10.08 11.18

[3,] 11.08 0.65 4.21 7.56

[4,] 12.66 32.52 16.92 37.02

[5,] 0.87 3.56 4.81 27.54

[6,] 6.40 11.68 40.05 5.12

[7,] 54.21 34.63 52.55 37.20

[8,] 1.24 35.97 29.54 25.28

[9,] 37.94 47.07 19.64 28.11

[10,] 54.92 17.69 26.24 6.77

[11,] 25.52 38.10 24.74 1.90

[12,] 45.98 5.17 1.17 6.53

[13,] 7.13 34.35 12.18 9.86

[14,] 14.23 16.89 28.93 21.70

[15,] 3.53 40.76 5.15 1.25

> y\_bar\_i=rep(0,n);y\_bar\_i

[1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> for(i in 1:15)

+ {

+ y\_bar\_i[i]=mean(m[i,])

+ }

> d=data.frame(y\_bar\_i);d

y\_bar\_i

1 6.7125

2 14.5750

3 5.8750

4 24.7800

5 9.1950

6 15.8125

7 44.6475

8 23.0075

9 33.1900

10 26.4050

11 22.5650

12 14.7125

13 15.8800

14 20.4375

15 12.6725

> ybar=mean(y\_bar\_i);ybar

[1] 19.3645

>

**#Average yield of apple is 19.3645**

> a=(y\_bar\_i-ybar)^2;a

[1] 160.073104 22.939310 181.966610 29.327640 103.418730 12.616704

[7] 639.230089 13.271449 191.144450 49.568640 10.243200 21.641104

[13] 12.141740 1.151329 44.782864

> Sbsq=(sum(a))/(n-1);Sbsq

[1] 106.6798

> ## For Swbar\_sq

> Si=rep(0,n)

> for (i in 1:15)

+ {

+ Si[i]=sum((m[i,]-y\_bar\_i[i])^2)

+ }

> Si

[1] 123.5761 178.0733 60.0041 418.3992 456.8261 807.4587 309.7065

[8] 689.6883 424.6258 1274.6169 681.8411 1319.0685 467.6338 124.8343

[15] 1059.5545

> Swbar\_sq=(sum(Si))/(n\*(M-1));Swbar\_sq

[1] 186.5757

**## To optain ira class corelation**

> b=((n-1)\*M\*Sbsq)-(n\*Swbar\_sq)

> c=((n-1)\*M\*Sbsq)+(n\*(M-1)\*Swbar\_sq)

> rho\_c=b/c;rho\_c

[1] 0.2209769

> Var\_Clu=((N-n)\*Sbsq\*(1+((M-1)\*rho\_c)))/(N\*(N-1)\*M);Var\_Clu

[1] 0.4698764

> SE=sqrt(Var\_Clu);SE

[1] 0.6854753

> dt=scan('clipboard');dt

Error in scan("clipboard") :

scan() expected 'a real', got 'dt=scan('clipboard');dt'

> dt=scan('clipboard');dt

Read 60 items

[1] 5.53 4.84 0.69 15.79 26.11 10.93 10.08 11.18 11.08 0.65 4.21 7.56

[13] 12.66 32.52 16.92 37.02 0.87 3.56 4.81 27.54 6.40 11.68 40.05 5.12

[25] 54.21 34.63 52.55 37.20 1.24 35.97 29.54 25.28 37.94 47.07 19.64 28.11

[37] 54.92 17.69 26.24 6.77 25.52 38.10 24.74 1.90 45.98 5.17 1.17 6.53

[49] 7.13 34.35 12.18 9.86 14.23 16.89 28.93 21.70 3.53 40.76 5.15 1.25

> var\_srs=var(dt)

> var\_srs

[1] 243.5589

> **##Efficiency of cluster sampling over SRSWOR**

> E=Var\_SRS/Var\_Clu;E

Error: object 'Var\_SRS' not found

> E=var\_srs/Var\_Clu;E

[1] 518.3467

> if (E>1)

+ {print("Cluster sampling is more efficient than SRSWOR")

+ }else

+ {

+ print("SRSWOR is more efficient than cluster sampling")

+ }

[1] "Cluster sampling is more efficient than SRSWOR"

>

1. **Average Yield of apple is 19.3645 kilogram with the standard error of 0.6854753 kg**
2. **Intra-class correlation coefficient between the trees within the cluster is 0.2209769**

**iii) Efficiency of cluster sampling over SRSWOR is 7.026598** (E>1)

**"Cluster sampling is more efficient than SRSWOR"**

**2. In the following data given below the number of standards of pepper in 15 clusters of 4 fields each, selected by simple random sampling without replacement out of 300 fields (i) estimate the average number of standards per**

**field along with its standard error.(ii) Estimate intra-cluster correlation coefficient between fields belonging to the same clusters with respect to simple random sampling without replacement.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **Field** | | | |
|  | **1** | **2** | **3** | **4** |
| **1** | **22** | **18** | **27** | **28** |
| **2** | **53** | **47** | **38** | **29** |
| **3** | **43** | **29** | **37** | **47** |
| **4** | **50** | **47** | **41** | **51** |
| **5** | **73** | **62** | **58** | **47** |
| **6** | **65** | **71** | **69** | **59** |
| **7** | **71** | **75** | **31** | **21** |
| **8** | **24** | **49** | **43** | **75** |
| **9** | **21** | **72** | **47** | **72** |
| **10** | **36** | **43** | **51** | **39** |
| **11** | **72** | **49** | **56** | **69** |
| **12** | **68** | **64** | **76** | **57** |
| **13** | **59** | **72** | **67** | **76** |
| **14** | **43** | **35** | **71** | **40** |
| **15** | **76** | **58** | **47** | **34** |

**Solution :**

x=read.csv(file.choose(),header =TRUE,sep=',');x

x1 x2 x3 x4

1 22 18 27 28

2 53 47 38 29

3 43 29 37 47

4 50 47 41 51

5 73 62 58 47

6 65 71 69 59

7 71 75 31 21

8 24 49 43 75

9 21 72 47 72

10 36 43 51 39

11 72 49 56 69

12 68 64 76 57

13 59 72 67 76

14 43 35 71 40

15 76 58 47 34

> N=77

> n=15

> M=4

> m=as.matrix(x);m

x1 x2 x3 x4

[1,] 22 18 27 28

[2,] 53 47 38 29

[3,] 43 29 37 47

[4,] 50 47 41 51

[5,] 73 62 58 47

[6,] 65 71 69 59

[7,] 71 75 31 21

[8,] 24 49 43 75

[9,] 21 72 47 72

[10,] 36 43 51 39

[11,] 72 49 56 69

[12,] 68 64 76 57

[13,] 59 72 67 76

[14,] 43 35 71 40

[15,] 76 58 47 34

> y\_bar\_i=rep(0,n);y\_bar\_i

[1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> for(i in 1:15)

+ {

+ y\_bar\_i[i]=mean(m[i,])

+ d=data.frame(y\_bar\_i);d

+ d=data.frame(y\_bar\_i);d

+ y\_bar\_i[i]=mean(m[i,])

+ }

> d=data.frame(y\_bar\_i);d

y\_bar\_i

1 23.75

2 41.75

3 39.00

4 47.25

5 60.00

6 66.00

7 49.50

8 47.75

9 53.00

10 42.25

11 61.50

12 66.25

13 68.50

14 47.25

15 53.75

> ybar=mean(y\_bar\_i);ybar

[1] 51.16667

> ## Average yield of apple is 51.1667

> a=(y\_bar\_i-ybar)^2;a

[1] 751.673611 88.673611 148.027778 15.340278 78.027778 220.027778

[7] 2.777778 11.673611 3.361111 79.506944 106.777778 227.506944

[13] 300.444444 15.340278 6.673611

> Sbsq=(sum(a))/(n-1);Sbsq

[1] 146.8452

> ## For Swbar\_sq

> Si=rep(0,n)

> for (i in 1:15)

+ {

+ Si[i]=sum((m[i,]-y\_bar\_i[i])^2)

+ }

> Si

[1] 64.75 330.75 184.00 60.75 346.00 84.00 2267.00 1330.75 1782.00

[10] 126.75 353.00 188.75 161.00 784.75 948.75

> Swbar\_sq=(sum(Si))/(n\*(M-1));Swbar\_sq

[1] 200.2889

> ## To obtain intra-class correlation coefficient

> b=((n-1)\*M\*Sbsq)-(n\*Swbar\_sq)

> c=((n-1)\*M\*Sbsq)+(n\*(M-1)\*Swbar\_sq)

> rho\_c=b/c;rho\_c

[1] 0.3027906

> Var\_Clu=((N-n)\*Sbsq\*(1+((M-1)\*rho\_c)))/(N\*(N-1)\*M);Var\_Clu

[1] 0.7422501

> SE=sqrt(Var\_Clu);SE

[1] 0.8615394

> dt=scan('clipboard');dt

Read 60 items

[1] 22 18 27 28 53 47 38 29 43 29 37 47 50 47 41 51 73 62 58 47 65 71 69 59 71

[26] 75 31 21 24 49 43 75 21 72 47 72 36 43 51 39 72 49 56 69 68 64 76 57 59 72

[51] 67 76 43 35 71 40 76 58 47 34

> Var\_SRS=var(dt)

> Var\_SRS

[1] 292.1412

> ## Efficiency of cluster sampling over SRSWOR

> E=Var\_SRS/Var\_Clu;E

[1] 393.5887

> if (E>1)

+ {

+ print("Cluster sampling is more efficient than SRSWOR")

+ }else

+ {

+ print("SRSWOR is more efficient than cluster sampling")

+ }

[1] "Cluster sampling is more efficient than SRSWOR"

> **i) Average Yield of apple is 51.16667 with the standard error of 0.8702831**

1. **Intra-class correlation coefficient between the trees within the cluster is 0.3027906**

**iii) Efficiency of cluster sampling over SRSWOR is 5.21246** (E>1)

**"Cluster sampling is more efficient than SRSWOR"**