**STRATIFIED SAMPLING**

**Question\_1:**

A car manufacturing company has sold 2000 cars to the public through licensed dealers. The company is now interested in finding out the average distance travelled per week by a car manufactured by the company. This information is likely to be helpful in fixing the warranty period for certain parts of the car. The addresses and telephone numbers, if installed, of all the buyers along with their occupations are available at the head office of the company. Since the distance travelled by a car is likely to vary with the profession of the buyer, the investigator divides the population into 3 groups - the businessmen (stratum I), employees (stratum II), and others (stratum III) which includes farmers, etc. Out of 2000 buyers, 825 are businessmen, 700 employees, and 475 others. The average per unit cost for collecting information is expected to be $ 4 for businessmen, $ 5.5 for employees, and $ 6.5 for persons from other category. The total budget at hand is $ 1550 which includes the overhead cost of $1000. On using optimum allocation formula given in (5.11), the investigator arrived at allocation of sample size nl = 53 buyers to stratum I, n2 = 34 buyers to stratum II, and n3 = 23 buyers to stratum III (procedure of determining these allocations is explained in the solution). The observations on the study variable obtained from these three WOR simple random samples are given in table 1.

**Table\_1:** Average distance (in km) per week covered by cars included

in the sample.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Stratum I** | |  | **Stratum II** | | | **Stratum III** | |
| 655 | 300 | 574 | 666 | 471 | 280 | 685 | 714 | 236 |
| 400 | 871 | 524 | 715 | 350 | 411 | 493 | 678 | 824 |
| 526 | 813 | 310 | 691 | 625 | 240 | 206 | 664 | 385 |
| 774 | 861 | 650 | 480 | 388 | 636 | 579 | 318 | 650 |
| 780 | 722 | 470 | 680 | 566 | 422 | 358 | 840 | 585 |
| 812 | 705 | 460 | 841 | 421 | 517 | 385 | 421 | 496 |
| 805 | 831 | 483 | 825 | 398 | 451 |  | 666 | 704 |
| 525 | 748 | 310 | 488 | 881 | 380 |  | 848 | 569 |
| 401 | 446 | 489 | 330 | 434 | 326 |  | 410 | 614 |
| 806 | 856 | 576 | 580 | 405 | 595 |  | 549 |  |
| 828 | 387 | 615 | 811 | 693 | 401 |  | 602 |  |
| 746 | 399 | 704 |  | 615 | 612 |  | 253 |  |
| 560 | 635 | 774 |  | 375 | 564 |  | 777 |  |
| 475 | 560 | 533 |  | 469 | 343 |  | 411 |  |

**SOLUTION:-**

C = $ 1550, Co = $1000, C1 = $ 4, C2 = $ 5.5, C3 = $ 6.5,

n1=53, n2=34, n3=23, N1= 825, N2=700, N3=475, N = 2000,

W1 =0.4125, W2 = 0.3500, W3 = 0.2375,

yst = h yh

st  =[N11+N22+N33]

1 = ∑y1/n1 2 =∑y2/n2 3 =∑y3/n3

1 = 618.9811 2 = 470.7352 3 = 574.5217

st = [825(618.9811)+700(470.7352)+475(574.5217)]

**st = 556.5359**

Var(st) = W12S12(N1-n1)/N1n1 + W22S22(N2-n2)/N2n2 + W32S32(N3-n3)/N3n3

S12 = (1/n1-1)(∑y12-n112)

S12 = (1/52)(21791262-53(383137.6022)

S12 = 28557.0978

S22 = (1/n2-1)(∑y22-n222)

S22 = 20509.0797

S32 = (1/n3-1)(∑y32-n332)

S32 = 32902.12605

Var(st) = 85.7901+70.3039+76.3238

**Var(st)** **= 232.4202**

NOW,

Confidence Interval,

st ± 2st)

556.5359 ± 2(15.2453)

556.5359 ±30.4906

526.0453, 587.0265

**Question:-2**

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 2.

**Table-2:-** **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 |  |

Select a stratified sample of 24 farmers by using equal allocation, proportional allocation, and Neyman allocation. Compute the overall population mean and the population mean square S2. Work out the relative efficiency of stratified sample mean st based on each of the above mentioned allocations, with respect to the simple random sample mean Y for the same total sample size. Assume that the sampling is WOR.

**Solution:-**

Given,

n = 24, N1 = 20, N2 = 36, N3 = 24

W1 = 0.25, W2 = 0.45, W3 = 0.30,

Compute overall population mean,

= ∑Y/N

= (75+65+86+57+…………+16)

80

**= 43.7873**

Population mean square,

S2 = 1 [2 –N()2]

N-1

S2 = 1 [ (75)2+ (65)2+………..+(16)2 – (80)(43.7873)2]

79

= 268.6758

Now, sample variance

Var(Y) = N-n S2

Nn

= (80-24) (268.6758)

(80)(24)

= **7.8373**

Now, stratum mean squares

Var(sh2) = 1 [∑yh2- nh(h)2]

nh-1

S12 = 169.5158, S22 = 70.5611, S32 = 61.4493

Now, obtain value of var (st) using equal allocation method.

Var(st) = h2Sh2 [Nh-nh]

Nhnh

=W1S12[N1-n1] + W2S22[N2-n2] + W3S32[N3-n3]

N1n1 N2n2 N3n3

= (0.25)2(169.5158)(20-8) + (0.45)2(70.5611)(36-8) + (0.30)2(61.4493)(24-8)

(20)(8) (36)(8) (24)(8)

= 0.7946 + 1.3892 + 0.4609

= **2.6447**

Relative efficiency of equal allocation,

RE = Var() (100)

Var(st)

= 7.8373 (100)

2.6447

= **296.3**

Now, for proportional allocation

n1 ={ }N1

= {} 20

= 6

n2 = {} 36

= 10.80 ≈ 11

n3= {} 24

= 7.20 ≈ 7

Var(st) = (0.25)2(169.5158)(20-6) + (0.45)2(70.5611)(36-11) + (0.30)2(61.4493)(24-7)

(20) (6) (36) (11) (24) (7)

= 1.2361 + 0.9021 + 0.5596

= **2.6978**

RE = Var() (100)

Var(st)

= 7.8373 (100)

2.6978

= **290.5**

Now, for neyman allocation

nh = n NhSh

∑NhSh

n1 = (24) 260.4

750.9

n1= 8.2 ≈ 8

n2 = 9.7 ≈ 10

n3 = 6

Now,

Var(st) = (0.25)2(169.5158)(20-8) + (0.45)2(70.5611)(36-10) + (0.30)2(61.4493)(24-6)

(20)(8) (36)(10) (24)(6)

**= 2.5179**

RE = 7.8373 (100)

2.5179

**= 311.2**

**Question:-3**

The management of a local newspaper is to decide whether it should continue with the publication of 'Children Column', which had been introduced on experimental basis. For this purpose, it is imperative to estimate the proportion of readers who would favor its continuance. The frame consists of readers who had stayed with the paper for the last six months. The addresses of these readers are available in the office of the newspaper. Since different attitudes are expected from the urban and rural readers, it is reasonable to stratify the population into urban readers and rural readers. Inthe population, there are 73000 urban readers and 30280 rural readers. The total budget at hand is $3000 only. The overhead cost is $820, and the per unit cost for urban and rural readers is expected to be $2 and $2.5 respectively. Using proportional allocation method, explained in solution, the investigator selected WOR simple random samples of 718 respondents from stratum I (urban readers)

and 298 readers from stratum II (rural readers). The number of individuals who favor continuation of the column was 570 from stratum I and 143 from stratum II**.** Estimate the proportion of readers interested in the continuation of the said column. Also, build up confidence interval for the population proportion.

**Solution,**

N1 = 73000, N2 = 30280, N = 103280,

C = $3000, Co = $820,

C1 = $2, C2 = $2.5,

W1= 0.7068, W2= 0.2932,

Now,

n= C-C0

∑WhSh

n= 3000-820

(0.7068)(2)+(0.2932)(2.5)

n = 1015.56≈ 1016

n1 =[ ]N1

= 1016 (73000)

103280

= 718

n2 = 298

p1 = 570/718

p1 = 0.7938

p2 = 143/298

p2 = 0.4798

The estimate of the required overall population proportion is,

Pst= 1/N(N1P1+N2P2)

Pst= 1/103280[(73000)(0.7939)+(30280)(0.4799)]

Pst= 0.7018

Estimation of proportion variance,

Var(pst)= W12p1q1 [N1-n1] + W22p2q2 [N2-n2]

n1-1 N1 n2-1 N2

= (0.7068)2(0.9901)(0.000228)+(0.2932)2(0.9901)(0.000840)

= 0.00018

Confidence Intervals,

pst±2st)

0.7018±2(0.01341)

0.675, 0.7286

**Question:-4**

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 nh,

h = 1, 2, 3, are given below,

|  |  |  |  |
| --- | --- | --- | --- |
| Age group | Persons | nh | Literate |
| (years) |  |  | 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.

**Solution,**

N1 = 25600, N2 = 18100, N3 = 6300, N = 50000

n1 = 256, n2 = 181, n3 = 6300

a1 = 243, a2 = 144, a3 = 31

p1 = a1/n1

p1 = 243/256

p1 = 0.9492

p2 = a2/n2

p2 = 144/181

p2 = 0.7955

p3 = 0.4920

pst = 1/N[N1P1+N2P2+N3P3]

pst = 1/50000[(25600)(0.9492)+(18100)(0.7955)+(6300)(0.4920)]

pst = 0.8359

v(pst)= W12p1q1 [N1-n1] + W22p2q2 [N2-n2] + W32p3q3 [N3-n3]

n1-1 N1 n2-1 N2 n3-1 N3

v(pst)= 0.000049+0.00011+0.000063

v(pst)= 0.00022

Confidence Intervals,

pst±2st)

0.8359±0.0296

0.806 , 0.8655