Experiment-1

Estimate the total no' of words in a dictionary provided to you. Find the standard error of total no' of words. Also obtain 95% fiducial limit for the total no' of words.

Formulae Used :→ Let N denotes the total no' of pages in the dictionary. Let n be the total no' of pages taken from "RANDOM NUMBER TABLE." as sample.

Let y_i be the number of wordes on ith page.

Estimate of total no' of words $(\hat{y}) = N\bar{y} = \dfrac{N \sum_{i=1}^{n} y_i}{n}$

Standard error of $\hat{Y} = \sqrt{var(\hat{y})}$

$$= \sqrt{\frac{N(N-n)}{n} \hat{S}^2}$$

$$= \sqrt{\frac{N(N-n)}{n}} \cdot \hat{S}$$

$$= \sqrt{\frac{N(N-n)}{n}} \cdot s \quad \left[\begin{array}{l}\text{Replace } S^2 \text{ by its} \\ \text{unbiased estimate } s^2\end{array}\right]$$

where $s^2 = \dfrac{1}{n-1} \sum_{i=1}^{k} (y_i - \bar{y})^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{k} y_i^2 - n\bar{y}^2\right]$

95% fiducial limit means 95% confidence limit for total number of words $= \hat{y} \pm 1.96\ S.E.(\hat{y})$

Solution: A random dictionary be chosen containing 657 pages whose real content start from page 4 to page 653. Then,

$$N = 650$$

We take a random sample of pages of 33 by using below random no table.

$$n = 33$$

We are using RANDOM NUMBER TABLE 4 for collecting the sample →

1027539623	7150543623	5431048298
2841506188	6485272018	8336360556
3421425702	5919189748	8030033098
6181772323	8282115408	5328705896
6115181354	1686202688	9074805509
0414121509	2678254747	9176216464
3971650962	9476621189	0097790806
0524677007	8497508746	3646183494
4407395543	4234439228	8898996050
1453905117	5201630159	0437598721
4491133297	7531626654	8480327577
3730285985	5356685340	0174395973
7520802777	7891691600	0843187368
6595794294	9362408996	4356477166
0502032417	4797815651	9234860182

Calculations:

$$N = 650$$

$$n = 33$$

$$\sum_{i=1}^{33} y_i = 631$$

$$\sum_{i=1}^{33} y_i^2 = 14199$$

Sr. No.	Page No.	No. of words(y_i)	y_i^2
1	102	9	81
2	415	23	529
3	232	29	841
4	151	30	900
5	543	13	169
6	272	33	1089
7	189	28	784
8	282	20	400
9	620	17	289
10	310	5	25
11	636	22	484
12	33	10	100
13	589	15	225
14	74	31	961
15	141	22	484
16	397	21	441
17	467	23	529
18	440	15	225
19	145	6	36
20	267	25	625
21	476	22	484
22	344	25	625
23	520	24	576
24	618	13	169
25	216	2	4
26	80	24	576
27	605	19	361
28	437	25	625
29	28	24	576
30	491	22	484
31	208	3	9
32	595	18	324
33	502	13	169
	Total	631	14199

$$\bar{y} = \frac{\sum_{i=1}^{33} y_i}{n} = \frac{631}{33} = 19.12$$

$$\hat{y} = N\bar{y} = 650(19.12) = 12428$$

$$S.E.(\hat{y}) = \sqrt{\frac{N(N-n)}{n}} \cdot s$$

where $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{33} y_i^2 - n\bar{y}^2\right]$

$$= \frac{1}{33-1}\left[14199 - 33(19.12)^2\right]$$

$$= \frac{1}{32}[14199 - 12063.96]$$

$$= \frac{1}{32}(2315.04)$$

$$= 66.72$$

$$s = \sqrt{66.72} = 8.17$$

$$S.E(\hat{y}) = \sqrt{\frac{650(650-33)}{33}} \cdot (8.17)$$

$$= \sqrt{12153.03} \cdot (8.17)$$

$$= 110.24(8.17)$$

$$= 900.67$$

95% fiducial limit $= \hat{y} \pm 1.96\,(S.E.(\hat{y}))$

$$= 12428 \pm 1.96(900.67)$$

$$= 12428 \pm 1765.3132$$

$$= (14193.3132,\ 10662.6868)$$

Observation & Result :

The total no. of words in the dictionary as estimated are 12428 came out with an error of 900.67 for total number of words. We can say with 95% confidence that words in a dictionary will be lie in range (10662.68, 14193.31).

Experiment-2

A petition was collected on 600 sheets. Each sheet has enough space for 50 signature per sheet. A Random sample of 40 sheets was taken with the following details:

Signature per sheet :	50	48	45	43	41	39	35	32	30	29	25	22	16	4	3
No of sheets :	5	7	3	2	3	2	2	2	1	1	3	2	1	2	4

Estimate the total number of signatures. Also, obtain 99% confidence limit for the total.

Formulae Used

Let 'N' be the total number of sheets (i.e Population size) and n be the sample size.

Let 'y_i' be the number of signatures on ith sheet and 'f_i' be the number of sheets.

Population Mean $(\bar{Y}) = \frac{\sum_{i=1}^{N} f_i y_i}{N}$ where $N = 600$

Sample mean $(\bar{y}) = \frac{\sum_{i=1}^{k} f_i y_i}{n}$ where $n = \sum_{i=1}^{k} f_i$

Mean square for the sample $(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{k} f_i y_i^2 - n\bar{y}^2\right]$

Standard Error $(S.E(\hat{Y})) = \sqrt{\frac{N(N-n)}{n} \cdot s^2}$

Signature Per Sheet(y_i)	No. Of Sheets(f_i)	f_iy_i	$f_iy_i^2$
50	5	250	12500
48	7	336	16128
45	3	135	6075
43	2	86	3698
41	3	123	5043
39	2	78	3042
35	2	70	2450
32	2	64	2048
30	1	30	900
29	1	29	841
25	3	75	1875
22	2	44	968
16	1	16	256
4	2	8	32
3	4	12	36
Total	40	1356	55892

99% Confidence Limit $= \hat{y} \pm 2.58\ S.E(\hat{y})$ where $\hat{y} = N\bar{y}$

Calculations:

$N = 600$

$n = 40$

$k = 15$

$$\sum_{i=1}^{15} f_i y_i = 1356$$

$$\sum_{i=1}^{15} f_i y_i^2 = 55892$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{15} f_i y_i = \frac{1356}{40} = 33.9$$

$$\hat{y} = N\bar{y} = 600 \times 33.9 = 20340$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{15} f_i y_i^2 - n\bar{y}^2\right] = \frac{1}{40-1}\left[55892 - 40(33.9)^2\right]$$

$$= \frac{1}{39}\left[55892 - 45968.4\right]$$

$$= 254.45$$

$$S.E(\hat{y}) = \sqrt{\frac{N(N-n)}{n} s^2} = \sqrt{\frac{600(600-40)}{40} \cdot 254.45}$$

$$= \sqrt{\frac{600 \times 560}{40} \times 254.45} = \sqrt{600 \times 14 \times 254.45}$$

$$= \sqrt{84 \times 25445} = \sqrt{2137380} = 1461.98$$

99% confidence limit $= \hat{y} \pm 2.58\ S.E(\hat{y})$

$= 20340 \pm 2.58\,(1461.98)$

$= (17006.6,\ 24111.9)$

Results: Estimated total number of signatures are 20340 and 99% Confidence limit for the total is (17006.6, 24111.9).

Experiment-3

The table shown below gives the no of inhabitents in 201 cities that had population over 50,000. Calculate the standard error of the total no. of inhabitants in all the cities by taking

i) a simple random sample of size 50

ii) a sample that includes 8 largest cities with a random sample of 42 cities out of remaining cities

iii) a sample that includes 10 largest cities with a random sample of 40 cities out of remaining cities

Population (in thousands)	No of cities	Population (in thousands)	No of cities
50-100	24	550-600	6
100-150	18	600-650	2
150-200	15	650-700	1
200-250	13	700-750	3
250-300	22	750-900	2
300-350	32	900-1000	3
350-400	19	1000-1200	2
400-450	17	1200-1500	1
450-500	11	1500-1800	1
500-550	8	1800-2000	1
		Total	201

Sr. no.	Population in thousands	y_i	No. Of Cities(f_i)	f_iy_i	$f_iy_i^2$
1	50-100	75	24	1800	135000
2	100-150	125	18	2250	281250
3	150-200	165	15	2475	408375
4	200-250	225	13	2925	658125
5	250-300	275	22	6050	1663750
6	300-350	325	32	10400	3380000
7	350-400	375	19	7125	2671875
8	400-450	425	17	7225	3070625
9	450-500	475	11	5225	2481875
10	500-550	525	8	4200	2205000
11	550-600	575	6	3450	1983750
12	600-650	625	2	1250	781250
13	650-700	675	1	675	455625
14	700-750	725	3	2175	1576875
15	750-900	825	2	1650	1361250
16	900-1000	950	3	2850	2707500
17	1000-1200	1100	2	2200	2420000
18	1200-1500	1350	1	1350	1822500
19	1500-1800	1650	1	1650	2722500
20	1800-2000	1900	1	1900	3610000
		TOTAL	**201**	**68825**	**36397125**

Formulae Used :

Let y_i be the no. of inhabitants in the cities and f_i denotes the no. of cities.

Let 'N' be the population size and 'n' be the sample size.

In simple random sampling :

Mean square for the population $(S^2) = \frac{1}{N-1}\left[\sum_{i=1}^{N} f_i y_i^2 - \frac{\left(\sum_{i=1}^{N} f_i y_i\right)^2}{N}\right]$

standard error i.e. $S.E(\hat{y}) = \sqrt{\frac{N(N-n)}{n} \cdot S^2}$

Calculations :

$$N = 201$$

i) $n = 50$

$$\sum_{i=1}^{20} f_i y_i = 68825$$

$$\sum_{i=1}^{20} f_i y_i^2 = 36397125$$

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{20} f_i y_i^2 - \frac{\left(\left(\sum_{i=1}^{20} f_i y_i\right)^2\right)}{N}\right]$$

$$= \frac{1}{201-1}\left[36397125 - \frac{(68825)^2}{201}\right]$$

$$= \frac{1}{200}\left[36397125 - 23566570.273\right]$$

$$= 64152.773$$

$$S.E.(\hat{Y}) = \sqrt{\frac{N(N-n)}{n} \cdot s^2} = \sqrt{\frac{201(201-50)}{50} \times 64152.773}$$

$$= \sqrt{38942016.26646}$$

$$= 6240.354$$

ii) Now, a sample that includes 8 largest cities is taken. So, $\Sigma f_i y_i$ of 8 largest cities $= 9950$ and $\Sigma f_i y_i^2$ of 8 largest cities $= 13282500$

$\Sigma f_i y_i^2$ for remaining 193 cities =

$\Sigma f_i y_i^2$ of 201 cities $-$ $\Sigma f_i y_i^2$ for 8 largest cities

$$= 36397125 - 13282500$$

$$= 23114625$$

$\Sigma f_i y_i$ for remaining 193 cities =

$\Sigma f_i y_i$ for 201 cities $-$ $\Sigma f_i y_i$ of 8 largest cities

$$= 68825 - 9950$$

$$= 58875$$

$$s_1^2 = \frac{1}{N_1 - 1}\left[\Sigma f_i y_i^2 - \frac{(\Sigma f_i y_i)^2}{N_1}\right] \quad \text{where } N_1 = 193$$

$$= \frac{1}{193-1}\left[23114625 - \frac{(58875)^2}{193}\right]$$

$$= \frac{1}{192}\left[23114625 - 17959925.518\right]$$

$$= 26847.39$$

$$S.E.(\hat{Y}) = \sqrt{\frac{N_1(N_1-n_1)}{n_1} \cdot s_1^2} \quad \text{where } n_1 = 42$$

$$= \sqrt{\frac{193\,(193-42)}{42} \times 26847.39}$$

$$= \sqrt{186288922.542142}$$

$$= 4316.1200796713 \sim 4316.12$$

iii) Here a sample that includes 10 largest cities is taken. So, $\Sigma f_i y_i$ for 10 largest cities $= 11600$

$\Sigma f_i y_i^2$ for 10 largest cities $= 14643750$

$\Sigma f_i y_i$ for remaining 191 cities =

$\Sigma f_i y_i$ for 201 cities $-$ $\Sigma f_i y_i$ for 10 largest cities

$= 68825 - 11600$

$= 57225$

$\Sigma f_i y_i^2$ for remaining 191 cities =

$\Sigma f_i y_i^2$ for 201 cities $-$ $\Sigma f_i y_i^2$ for 10 largest cities

$= 36397125 - 14643750$

$= 21753375$

$$S_2^2 = \frac{1}{N_2-1}\left[\Sigma f_i y_i^2 - \frac{(\Sigma f_i y_i)^2}{N_2}\right] \quad \text{where } N_2 = 191$$

$$= \frac{1}{191-1}\left[21753375 - \frac{(57225)^2}{191}\right]$$

$$= \frac{1}{190}\left[21753375 - 17145029.45\right]$$

$$= \frac{1}{190}\left[4608345.55\right]$$

$$= 24254.45$$

$$S.E(\hat{Y}) = \sqrt{\frac{N_2(N_2-n_2)}{n_2} \cdot S_2^2} \quad \text{where } n_2 = 40$$

$$= \sqrt{\frac{191(191-40)}{40} \times 24254.45}$$

$$= \sqrt{17488064.81125}$$

$$= 4181.873$$

Result: Standard error of the total no of inhabitants in 201 by taking

i) a simple random sample of size 50 is 6~~7~~240.354

ii) a sample that includes 8 largest cities with a random sample of 42 cities out of remaining cities is 4316.12.

iii) a sample that includes 10 largest cities with a random sample of 40 cities of remaining cities is 4181.873.

Experiment - 4

Estimate the total number of words in a dictionary using stratified random sampling (take 5% of the total number of pages or 50 words which is greater than as the sample size). Also find the standard error of the estimate and 99% confidence limit for it.

Formulae Used : Let N be the total number of pages (with words in the dictionary) and n_h be the sample taken from h th strata (which is of size N_h)

Estimate of the total no of words in dictionary is $\hat{Y} = N\overline{y}_{st}$ where

$$\overline{y}_{st} = \sum_{h=1}^{L} \frac{N_h \overline{y}_h}{N}$$

$$S.E(\hat{Y}) = N(S.E.\ of\ \overline{y}_{st}) = N\sqrt{Var(\overline{y}_{st})}$$

$$= N\sqrt{\sum_{h=1}^{L} \frac{1-f_h}{n_h} W_h^2 \hat{S}_h^2}$$

$$= N\sqrt{\sum_{h=1}^{L} \frac{1-f_h}{n_h} W_h^2 s_h^2}$$

where, $E(S_h^2) = s_h^2$

and $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \overline{y})^2$

$$= \frac{1}{n_h - 1}\left[\sum_{i=1}^{n_h} y_{hi}^2 - \frac{1}{n_h}\left(\sum_{i=1}^{n_h} y_{hi}\right)^2\right]$$

and, $W_h = \dfrac{N_h}{N}$, $f_h = \dfrac{n_h}{N_h}$

99% confidence limit for $\hat{y}$ are -

$$\hat{y} \pm 2.58 \; S.E(\hat{y})$$

We have taken little Hindi-to-sanskrit (Eng) dictionary.

The total no. of pages in the dictionary are 650. i.e. $N = 650$

Samples are drawn with the help of random no table and the sample size is 51

$n = 51$

We are taken five strata from the population.

Calculation :

We are taken, $N_1 = 185$, $n_1 = 15$

$N_2 = 162$, $n_2 = 12$

$N_3 = 124$, $n_3 = 10$

$N_4 = 92$, $n_4 = 8$

$N_5 = 87$, $n_5 = 6$

Sr. no.	Stratum	No. of Pages	Sample size	Page no.
1	A-G	185	15	102, 150, 142, 34, 43, 92, 89, 165, 123, 67, 03, 56, 128, 178,96
2	H-L	162	12	198, 203, 224, 267, 278, 289,301,345,188, 322, 275,239
3	M-R	124	10	382, 377, 393, 407, 422, 436, 455,461,484,356
4	S-V	92	8	477, 498, 487, 504, 561, 524, 547, 558
5	W-Z	87	6	567, 632, 604, 647, 598, 633
		650	51	

Stratum 1 (A-G)			
Sr. no.	Page no.	No. Of Words(y_{1j})	y_{1j}^2
1	102	9	81
2	150	23	529
3	142	13	169
4	34	30	900
5	43	11	121
6	92	28	784
7	89	4	16
8	165	6	36
9	123	5	25
10	67	10	100
11	3	15	225
12	56	21	441
13	128	27	729
14	178	22	484
15	96	25	625
	Total	249	5265

STRATUM 2 (H-L)			
Sr. no.	Page no.	No. Of Words(y_{1j})	y_{2j}^2
1	198	13	169
2	203	18	324
3	224	3	9
4	267	22	484
5	278	24	576
6	289	25	625
7	301	19	361
8	345	24	576
9	188	2	4
10	322	13	169
11	275	23	529
12	239	5	25
	Total	191	3851

STRATUM 3 (M-R)			
Sr. no.	Page no.	No. Of Words(y_{1j})	y_{3j}^2
1	382	17	289
2	377	5	25
3	393	22	484
4	407	10	100
5	422	15	225
6	436	31	961
7	455	22	484
8	461	25	625
9	484	24	576
10	356	13	169
	Total	184	3938

STRATUM 4(S-V)			
Sr. no.	Page no.	No. Of Words(y_{1j})	y_{4j}^2
1	477	15	225
2	498	5	25
3	487	7	49
4	504	25	625
5	561	22	484
6	524	6	36
7	547	13	169
8	558	28	784
	Total	121	2397

STRATUM 5(W-Z)			
Sr. no.	Page no.	No. Of Words(y_{1j})	y_{5j}^2
1	567	19	361
2	632	2	4
3	604	13	169
4	647	18	324
5	598	22	484
6	633	20	400
	Total	94	1742

sr. no.	n_h	$\sum Y_h$	$Y_h = \sum Y_h / n_h$	N_h	$N_h Y_h$	$(1-f_h)/n_h$	W_h	S_h^2	$[(1-f_h)/n_h)*W_h^2 S_h^2$
1	15	249	16.6	185	3071	0.061261	0.284615	80.82	0.401071243
2	12	191	15.916667	162	2578.5	0.07716	0.249231	73.72	0.353332544
3	10	184	18.4	124	2281.6	0.091935	0.190769	61.37	0.205331673
4	8	121	15.125	92	1391.5	0.11413	0.141538	80.98	0.185151905
5	6	94	15.666667	87	1363	0.155172	0.133846	53.88	0.149780024
Total					10685.6				1.294667389

we have $N = 650$ and $n = 51$

$$S_1^2 = \frac{1}{n_1 - 1}\left[\sum y_{hj}^2 - \frac{1}{n_1}\left(\sum y_{hj}\right)^2\right]$$

$$= \frac{1}{14}\left[5265 - \frac{1}{15}(249)^2\right] = 80.82$$

$$S_2^2 = \frac{1}{n_2 - 1}\left[\sum y_{hj}^2 - \frac{1}{n_2}\left(\sum y_{hj}\right)^2\right]$$

$$= \frac{1}{11}\left[3851 - \frac{(191)^2}{12}\right] = 73.72$$

$$S_3^2 = \frac{1}{n_3 - 1}\left[\sum y_{hj}^2 - \frac{1}{n_3}\left(\sum y_{hj}\right)^2\right]$$

$$= \frac{1}{9}\left[3938 - \frac{1}{10}(184)^2\right] = 61.37$$

$$S_4^2 = \frac{1}{n_4 - 1}\left[\sum y_{hj}^2 - \frac{1}{n_4}\left(\sum y_{hj}\right)^2\right]$$

$$= \frac{1}{7}\left[2397 - \frac{1}{8}(121)^2\right] = 80.98$$

$$S_5^2 = \frac{1}{n_5 - 1}\left[\sum y_{hj}^2 - \frac{1}{n_5}\left(\sum y_{hj}\right)^2\right]$$

$$= \frac{1}{5}\left[1742 - \frac{1}{6}(94)^2\right]$$

$$= 53.88$$

Now, $\hat{Y} = N \cdot \sum_{h=1}^{L} \frac{N_h \bar{y}_h}{N}$

$$= 650 \sum_{h=1}^{5} \frac{N_h \bar{y}_h}{650}$$

$$= 10685.6$$

$$S.E(\hat{Y}) = N \sqrt{\sum_{h=1}^{5} \frac{1-f_h}{n_h} W_h^2 S_h^2}$$

$$= 650 \times \sqrt{1.294667389}$$

$$= 650 \times 1.137834$$

$$= 739.5924$$

99% confidence for $\hat{Y}$ are $\rightarrow$ $\hat{Y} \pm 2.58 (S.E(\hat{Y}))$

$$= 10685.6 \pm 2.58 (739.5924)$$

$$= 10685.6 \pm 1908.1483$$

$$= (8777.4517, 12566.7)$$

Result :-

- Estimate of total no of words in the dictionary is $10685.6 \sim 10686$

- S.E. of the total estimates of the words is 739.5924

- 99% CL for the estimate is $(8777.4517, 12566.7)$

Experiment - 5

The following table gives the no of inhabitants in 64 cities. The cities are divided into two strata, first containing 16 large cities and second containing remaining 48 cities. The total no of inhabitants in 64 cities are to be estimated using a sample size 24. Find the standard error of estimate of this total use in

i) Simple random sampling

ii) Stratified random sampling with proportional allocation.

iii) Stratified random sampling with 12 cities drawn from each strata that is equally proportional.

Population (in thousands)

678	888	807	764	310	764	119
905	887	887	876	245	790	112
983	690	691	655	145	410	113
679	456	453	482	142	640	234
683	578	453	452	252	390	
896	506	610	556	123	370	
678	500	496	482	231	340	
674	570	490	562	121	555	
675	467	452	562	157	345	
676	466	564	562	234	520	

Formulae Used :

Let N be the total number of cities (population size) and n be the sample size. Let y_i denotes the no. of inhabitants in the ith city. In case of Simple Random sampling,

S.E. of estimate is $\sqrt{\frac{N(N-n)}{n} \cdot \hat{S}^2}$

where $\hat{S}^2 = s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N}\right]$

Now in Stratified Random sampling with proportional allocations :

N_h = Population size of h^{th} stratum

n_h = sample size of h^{th} stratum

$$\frac{n_h}{N_h} = \frac{n}{N} \Rightarrow f_h = f$$

Here, S.E. of estimate is $\sqrt{\frac{1-f}{n} \sum_{h=1}^{N} W_h S_h^2}$

where $W_h = \frac{N_h}{N}$, $f_h = \frac{n_h}{N_h}$

and $S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N} (y_{hi} - \bar{y})^2$

In stratified Random sampling,

$$V(\bar{y}_{st}) = \sum_{h=1}^{N} \frac{1-f_h}{n_h} W_h^2 S_h^2$$

Stratum 1		
S.No.	y_i	y_i^2
1	983	966289
2	905	819025
3	896	802816
4	888	788544
5	887	786769
6	887	786769
7	876	767376
8	807	651249
9	790	624100
10	764	583696
11	764	583696
12	691	477481
13	690	476100
14	683	466489
15	679	461041
16	678	459684
Total	12868	10501124

Stratum 2		
S.No.	y_i	y_i^2
1	678	459684
2	676	456976
3	675	455625
4	674	454276
5	655	429025
6	640	409600
7	610	372100
8	578	334084
9	570	324900
10	564	318096
11	562	315844
12	562	315844
13	562	315844
14	556	309136
15	555	308025
16	520	270400
17	506	256036
18	500	250000
19	496	246016
20	490	240100
21	482	232324
22	482	232324
23	467	218089
24	466	217156
25	456	207936
26	453	205209
27	453	205209
28	452	204304
29	452	204304
30	410	168100
31	390	152100
32	370	136900
33	345	119025
34	340	115600
35	310	96100
36	252	63504
37	245	60025
38	234	54756
39	234	54756
40	231	53361
41	157	24649
42	145	21025
43	142	20164
44	123	15129
45	121	14641
46	119	14161
47	113	12769
48	112	12544
Total	20185	9977775

Observation :

$N = 64$ and $n = 24$

64 cities are divided into 2 strata where $n_1 = 16$ and $N_2 = 48$. By proportional allocation, we get

$$\frac{n_1}{N_1} = \frac{n}{N} \quad \text{and} \quad \frac{n_2}{N_2} = \frac{n}{N}$$

So, $n_1 = \frac{n}{N} \times N_1 = \frac{24}{64} \times 16 = 6$

and $n_2 = \frac{n}{N} \times N_2 = \frac{24}{64} \times 48 = 18$

Calculations :

i) $N = 64$, $n = 24$, $\sum_{i=1}^{64} y_i = 33053$

$$\sum_{i=1}^{64} y_i^2 = 20478899$$

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{64} y_i^2 - \frac{\left(\sum_{i=1}^{64} y_i\right)^2}{N}\right] = \frac{1}{64-1}\left[20478899 - \frac{(33053)^2}{64}\right]$$

$$= \frac{1}{63}[20478899 - 17070325.14]$$

$$= 54104.34$$

$$S.E(\hat{y}) = \sqrt{\frac{N(N-n)}{n} \cdot s^2} = \sqrt{\frac{64(64-24)}{24} \times 54104.34}$$

$$= \sqrt{5771129.6} = 2402.317$$

ii) $N_1 = 16$, $N_2 = 48$, $n_1 = 6$, $n_2 = 18$, $f_1 = f_2 = f = \frac{6}{16}$

$W_1 = \frac{N_1}{N} = \frac{16}{64} = \frac{1}{4}$, $W_2 = \frac{N_2}{N} = \frac{48}{64} = \frac{3}{4}$

In stratum 1 :

$$\sum_{i=1}^{16} y_i = 12868 \quad , \quad \sum_{i=1}^{16} y_i^2 = 10501124$$

In stratum 2,

$$\sum_{i=1}^{48} y_i = 20185 \quad , \quad \sum_{i=1}^{48} y_i^2 = 9977775$$

$$S_1^2 = \frac{1}{N_1-1}\left[\sum y_i^2 - \frac{(\sum y_i)^2}{N_1}\right] = \frac{1}{16-1}\left[10501124 - \frac{(12868)^2}{16}\right]$$

$$= 10135.66$$

$$S_2^2 = \frac{1}{N_2-1}\left[\sum y_i^2 - \frac{(\sum y_i)^2}{N_2}\right] = \frac{1}{48-1}\left[9977775 - \frac{(20185)^2}{48}\right]$$

$$= 31692.808$$

$$S.E(\hat{y}) = \sqrt{\frac{1-f}{n}\sum W_h S_h^2} = \sqrt{\frac{1-f}{n}\left[W_1 S_1^2 + W_2 S_2^2\right]}$$

$$= \sqrt{\frac{1-\frac{6}{16}}{24}\left[\frac{16}{64}\times 10135.66 + \frac{48}{64}\times 31692.808\right]}$$

$$= \sqrt{\frac{10}{16\times 24}\left[2533.915 + 23769.606\right]}$$

$$= \sqrt{\frac{263035.21}{16\times 24}} = \sqrt{684.9875} = 26.172265$$

iii) $N = 64, \quad n = 24, \quad N_1 = 16, \quad N_2 = 48, \quad n_1 = n_2 = 12$

$$f_1 = \frac{12}{16} = 0.75 \quad , \quad f_2 = \frac{12}{48} = 0.25$$

$$W_1 = \frac{N_1}{N} = \frac{16}{64} = 0.25 \quad , \quad W_2 = \frac{N_2}{N} = \frac{48}{64} = 0.75$$

$$S_1^2 = 10135.66 \quad , \quad S_2^2 = 31692.808$$

$$Var(\bar{y}_{st}) = \sum_{h=1}^{2}\left(\frac{1-f_h}{n_h}\right) W_h^2 S_h^2$$

$$= \left(\frac{1-f_1}{n_1}\right) W_1^2 S_1^2 + \left(\frac{1-f_2}{n_2}\right) W_2^2 S_2^2$$

$$= \left(\frac{1-0.75}{12}\right)(0.25)^2(10135.66) + \left(\frac{1-0.25}{12}\right)(0.75)^2(31692.808)$$

$$= \frac{(0.25)^3}{12} \times (10135.66) + \frac{(0.75)^2}{12} \times (31692.808)$$

$$= 1127.1785$$

$$S.E(\bar{y}_{st}) = \sqrt{1127.1785}$$

$$= 33.573$$

Result :

The total number of inhabitants in 64 cities are estimated.

i) Standard error of estimate using Simple Random sampling is 2402.317

ii) Standard error of estimate using stratified random sampling with proportional allocation is 26.172265

iii) Standard error of estimate using stratified Random sampling with equal proportional is 33.573.

Experiment - 6

The data given below are for small artificial population which exhibits a fairly steady rising trend. Each column represents a systematic sampling and the 4 rows are the strata. The data is for 10 systematic sample with sample $n = 4$, $k = 10$ and $N = nk = 40$. Compare the precision of systematic sampling, random sampling and stratified sampling.

Strata	Systematic Sample Number									
	1	2	3	4	5	6	7	8	9	10
A	0	1	1	2	5	4	7	7	8	6
B	6	8	9	10	13	12	15	16	16	17
C	18	19	20	20	24	23	25	28	29	27
D	26	30	31	31	33	32	35	37	38	38

Formulae Used :

For $N = 40$, $n = 4$, $k = 10$

$$Var(y_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{io} - \bar{y}_{oo})^2 = \frac{1}{k}\left[\sum_{i=1}^{k}\bar{y}_{io}^2 - k\,\bar{y}_{oo}^2\right]$$

$$= \frac{1}{n^2k}\left[\sum_{i=1}^{k}(n\bar{y}_{io})^2 - n^2k\,\bar{y}_{oo}^2\right]$$

$Var(\bar{y}_{st})^2 = \frac{k-1}{nk} \cdot S_{swt}^2$ where S_{swt}^2 is within mean sum of squares. $S_{swt}^2 = \frac{1}{n(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{io})^2$

$Var(\bar{y}_n)_{Ran} = \left(\frac{1}{n} - \frac{1}{N}\right)S^2$ where,

$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{oo})^2$ is the total mean sum of squares.

Strata	1	2	3	4	5	6	7	8	9	10	Srata Total T_{0i}
A	0	1	1	2	5	4	7	7	8	6	41
B	6	8	9	10	13	12	15	16	16	17	122
C	18	19	20	20	24	23	25	28	29	27	233
D	26	30	31	31	33	32	35	37	38	38	331
Sample Total $n\bar{y}_{0j}$	50	58	61	63	75	71	82	88	91	88	$\Sigma T_{0i} = 727$

Source of Variation	d.f.	S.S.	M.S.S.
Between Strata	4-1=3	4828.275	485.5/36=13.486
Within Strata	40-1=36	485.5	
Total	39	5313.775	5313.775/39=136.251

$$\text{Total S.S.} = \sum\sum y_{ij}^2 - C.F \quad \text{where } C.F = \frac{G^2}{N} = \frac{(\sum\sum y_{ij})^2}{N}$$

$$\text{Between strata S.S.} = \sum_{j=1}^{n} \frac{T_{oj}^2}{k} - C.F$$

$$\text{Within strata S.S} = \text{Total S.S.} - \text{Between S.S.}$$

$$S_{swt}^2 = \frac{\text{Within strata S.S}}{n(k-1)}$$

$$S^2 = \frac{\text{Total S.S.}}{N-1}$$

Calculations :

$$Var(y_{sys}) = \frac{1}{n^2 k}\left[\sum_{i=1}^{k}(n\bar{y}_{io})^2 - n^2 k \bar{y}_{oo}^2\right]$$

$$= \frac{1}{(4^2)10}\left[54713 - 160\left(\frac{727}{40}\right)^2\right]$$

$$= \frac{1}{160}\left[54713 - 52823.824\right]$$

$$= 11.807$$

$$\text{Total S.S} = 18527 - C.F \qquad \left[C.F = \frac{G^2}{N} = \frac{528529}{40} = 13213.225\right]$$

$$= 18527 - 13213.225$$

$$= 5313.775$$

$$\text{Between strata S.S} = \frac{(41)^2 + (122)^2 + (233)^2 + (331)^2}{10} - 13213.225$$

$$= \frac{180415}{10} - 13213.225$$

$$= 4828.275$$

$$\text{Within strata S.S} = 5313.775 - 4828.275$$

$$= 485.5$$

$$S_{wst}^2 = \frac{\text{Within strata S.S}}{n(k-1)} = \frac{485.5}{36} = 13.486$$

$$s^2 = \frac{\text{Total S.S}}{N-1} = \frac{5313.775}{39} = 136.251$$

$$Var(\bar{y}_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_{wst} = \left(\frac{1}{4} - \frac{1}{40}\right)(13.486)$$

$$= \frac{9}{40} \times 13.486$$

$$= 3.034$$

$$Var(\bar{y}_n)_{Ran} = \left(\frac{1}{n} - \frac{1}{N}\right) s^2 = \left(\frac{1}{4} - \frac{1}{40}\right)(136.251)$$

$$= \frac{9}{40} \times 136.251$$

$$= 30.656$$

Here $Var(\bar{y}_{sys})$ = Variance in systematic sampling

$Var(\bar{y}_{st})$ = Variance in stratified random sampling

$Var(\bar{y}_n)_{Ran}$ = Variance in simple random sampling.

Result :

Here $3.034 \leq 11.807 \leq 30.656$

$$V(\bar{y}_{st}) \leq V(\bar{y}_{sys}) \leq Var(\bar{y}_n)_{Ran}$$

$$\frac{1}{V(\bar{y}_{st})} \geq \frac{1}{V(\bar{y}_{sys})} \geq \frac{1}{V(\bar{y}_n)_{Ran}}$$

precision of $\bar{y}_{st}$ $\geq$ precision of $\bar{y}_{sys}$ $\geq$ precision of $(\bar{y}_n)_{Ran}$

Hence precision of stratified random sampling is more than other two.

Experiment - 7

For testing the variety effect in a completely randomised experiment, the data shown below in following table.

Perform the ANOVA

320	372	400	340	430	383
340	455	355	375	358	383
398	417	334	320	378	375
360	420	331	325	395	308
350	358	370	358	328	400

Formulae Used :-

Let y_{ij} denotes the jth observation of ith class where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$

- Raw sum of squares (S.S.) = $\sum_i \sum_j y_{ij}^2$
- Grand total (G) = $\sum_i \sum_j y_{ij}$
- Correction factor (C.F.) = $\frac{G^2}{n}$ where $n = \sum_{i=1}^{k} n_i$
- Total SS = Raw SS − C.F.
- SS due to class effect (SST) = $\frac{T_{10}^2}{n_1} + \frac{T_{20}^2}{n_2} + \ldots + \frac{T_{k0}^2}{n_k} - C.F.$

  where $T_{i0} = \sum_{j=1}^{n_i} y_{ij}$; $1 \leq i \leq k$

Class 1		
	y_i	y_i^2
320	-40	1600
340	-20	400
398	38	1444
360	0	0
350	-10	100
372	12	144
Total	**-20**	**3688**

Class 4		
	y_i	y_i^2
325	-35	1225
358	-2	4
430	70	4900
358	-2	4
378	18	324
395	35	1225
Total	**84**	**7682**

Class 2		
	y_i	y_i^2
455	95	9025
417	57	3249
420	60	3600
358	-2	4
400	40	1600
355	-5	25
Total	**245**	**17503**

Class 5		
	y_i	y_i^2
328	-32	1024
383	23	529
383	23	529
375	15	225
308	-52	2704
400	40	1600
Total	**17**	**6611**

Class 3		
	y_i	y_i^2
334	-26	676
331	-29	841
340	-20	400
370	10	100
375	15	225
320	-40	1600
Total	**-90**	**3842**

- Sum of square due to errors (SSE) = Total SS − SST
- Mean squares due to class effect (MST) = $\frac{SST}{k-1}$
- Mean squares due to errors (MSE) = $\frac{SSE}{n-k}$
- ratio = F = $\frac{MST}{MSE}$

Calculations :

As the value of F (ratio of variance) is not affected by changing the origin and scale.
Let us subtract 360 from each value.
Now, Grand Total (G) = $\sum_i \sum_j y_{ij}$

$$= (-20) + 245 + (-90) + 84 + 17$$
$$= 236$$

Correction factor (C.F.) = $\frac{G^2}{n} = \frac{(236)^2}{30} = \frac{55696}{30} = 1856.53$

Raw S.S. = $\sum_i \sum_j y_{ij}^2$

$$= 3688 + 17503 + 3842 + 7682 + 6611$$
$$= 39326$$

Total SS = Raw SS − C.F

$$= 39326 - 1856.53$$
$$= 37469.47$$

SST = $\frac{T_{10}^2}{n_1} + \frac{T_{20}^2}{n_2} + \frac{T_{30}^2}{n_3} + \frac{T_{40}^2}{n_4} + \frac{T_{50}^2}{n_5} - C.F$

$$= 322.6 + 10004.16 + 1350 + 1176 + 48.16 - 1856.53$$

$SST = 12900.92 - 1498.13 = $ [illegible]

$SSE = \text{Total } SS - SST$

$= 37827.87 - 11402.79 = 26425.08$

$$MST = \frac{SST}{n-1} = \frac{11402.79}{5-1} = 2761.0975$$

$$MSE = \frac{SSE}{n-k} = \frac{26425.08}{30-5} = 1057.0032$$

$$F = \frac{MST}{MSE} = \frac{2761.0975}{1057.0032} = 2.61$$

ANOVA Table →

Source of variation	S. S.	df	Mean square	F
Between classes	$SST = 11044.39$	$k-1 = 4$	$MST = 2761.0975$	$F = \frac{MST}{MSE}$ $= 2.61$
Within classes	$SSE = 26425.08$	$n-k = 25$	$MSE = 1057.0032$	
Total	$TSS = 37$[illegible]$7.$[illegible]7	$n-1 = 29$		

Tabulated $F_{0.05}(4, 25) = 2.76$ and calculated $F(4, 25) = 2.61$.

$$|F_{cal}| < |F_{tab}|$$

We shall accept the null hypothesis

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$ or $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$

Result: Since $|F_{cal}| < |F_{tab}|$. So we shall accept the null hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_5 = \mu$

$\therefore$ mean effect due to different classes/treatments is equal.

Experiment - 8

An opuiment was carried out on wheat with three treatments in four blocks. The plan and yield per plot in kg are given in following table :

I	II	III	IV
A	C	A	B
8	10	6	10
C	B	B	A
12	8	9	8
B	A	C	C
10	8	10	9

Analyse the data and state the conclusion and obtain the efficiency of this design relative to CRD.

Formulae Used : Let y_{ij} be the jth observation in the ith class where $i = 1, 2, \ldots, p$ & $j = 1, 2, \ldots, q$

Grand total $(G) = \sum_i \sum_j y_{ij}$

Correction factor $(C.F.) = \frac{G^2}{pq}$

Raw S.S. $= \sum_i \sum_j y_{ij}^2$

Total S.S. = Raw S.S − C.F

Sum of squares due to rows (PSSR) $= \frac{T_{10}^2 + \ldots + T_{p0}^2}{q} - C.F.$

Sum of squares due to columns (SSC) $= \frac{T_{01}^2 + \ldots + T_{0q}^2}{p} - C.F.$

Sum of squares due to error (SSE) = Total S.S. − SSR − SSC

		BLOCKS					
		I	II	III	IV	Total(T_{i0})	Square of Total
A	y	8	8	6	8	30	900
	y^2	64	64	36	64	228	
B	y	10	8	9	10	37	1369
	y^2	100	64	81	100	345	
C	y	12	10	10	9	41	1681
	y^2	144	100	100	81	425	
	Total(T_{0j})	30	26	25	27		
	Square of Total	900	676	625	729		

Mean squares due to Rows (MSR) $= \frac{SSR}{p-1}$

Mean squares due to columns (MSC) $= \frac{SSC}{q-1}$

Mean squares due to errors (MSE) $= \frac{SSE}{(p-1)(q-1)}$

$$F_R = \frac{MSR}{MSE} \quad \text{and} \quad F_c = \frac{MSC}{MSE}$$

Efficiency of RBD relative to CRD =

$$\frac{(q-1)\,MSC + q(p-1)\,MSE}{(pq-1)\,MSE}$$

Calculations :

Grand total (G) $= \sum\sum y_{ij}$

$$= 30+37+41 = 108$$

Correction factor (C.F.) $= \frac{G^2}{pq} = \frac{(108)^2}{3\times4} = 972$

Raw S.S. $= \sum\sum y_{ij}^2 = 228+345+425 = 998$

Total S.S = Raw S.S. − C.F. $= 998 - 972 = 26$

$$SSR = \frac{T_{10}^2 + T_{20}^2 + T_{30}^2}{4} - C.F = 15.5$$

$$SSC = \frac{T_{01}^2 + T_{02}^2 + T_{03}^2 + T_{04}^2}{3} - C.F = 4.6666$$

SSE = Total S.S − SSR − SSC $= 5.83333$

$$MSR = \frac{SSR}{3-1} = \frac{15.5}{2} = 7.75$$

$$MSC = \frac{SSC}{4-1} = \frac{4.6666}{3} = 1.5555$$

$$MSE = \frac{SSE}{2\times3} = \frac{5.83333}{6} = 0.97222$$

$$F_R = \frac{MSR}{MSE} = \frac{7.75}{0.9722} = 7.9714$$

$$F_C = \frac{MSC}{MSE} = \frac{1.5555}{0.9722} = 1.6$$

ANOVA Table :

Source of variation	Sum of squares (S.S)	df	Mean square (M.S)	F
Between rows	SSR = 15.5	$p-1 = 3-1 = 2$	MSR = 7.75	$F_R = \frac{MSR}{MSE} = 7.9714$
Between columns	SSC = 4.6667	$q-1 = 4-1 = 3$	MSC = 1.5555	$F_C = \frac{MSC}{MSE} = 1.6$
due to errors	SSE = 5.8333	$(p-1)(q-1) = 6$	MSE = 0.9722	
Total	TSS = 26	$pq-1 = 11$		

Tabulated $F_{0.05}$ for $(2,6) = 5.143$

Tabulated $F_{0.05}$ for $(3,6) = 4.75$

$\therefore$ $F_{cal} > F_{tab}$ for rows and $F_{tab} > F_{cal}$ for columns

Efficiency of RBD relative to C.R.D $= \frac{(4-1)(1.5555) + 4(2)(0.9722)}{(12-1)(0.9722)}$

$$= \frac{3(1.5555) + 8(0.9722)}{11(0.9722)} = 1.164$$

Result :

- Mean effect due to rows is not significant.
- Mean effect due to columns are equal.
- Efficiency of RBD relative to CRD is 1.164

Experiment - 9

The adjoining table gives the result of Latin square experiment on the effect of five manual treatments A, B, C, D & E. On the yield of significance, test weather the treatments are equally effective and if they are not so compare the departments A & B -

B 405	A 525	E 463	D 441	C 481
C 325	D 445	B 429	A 413	E 493
E 471	B 492	A 472	C 381	D 410
A 552	C 431	D 425	E 572	B 451
D 430	E 469	C 432	B 467	A 460

Formulae Used :

Let y_{ijk} be the observation due to kth treatment in ith row and j^{th} column.

Raw S.S. $= \sum_i \sum_j \sum_k y_{ijk}^2$ and C.F. $= \frac{G^2}{m^2}$

where G = grand total $= \sum_i \sum_j \sum_k y_{ijk}$

Total SS = Raw SS - C.F.

SS due to Rows $= \frac{(T_{1 0})^2 + (T_{2 0})^2 + \dots + (T_{m 0})^2}{m} - $ C.F.

SS due to columns $= \dfrac{(T_{01})^2 + (T_{02})^2 + \dots + (T_{0m})^2}{m} - C.F.$

Efficiency of LSD over RBD of rows $= \dfrac{S_R^2 + (m-1)S_E^2}{m \cdot S_E^2}$

Efficiency of LSD over RBD of columns $= \dfrac{S_c^2 + (m-1)S_E^2}{m \cdot S_E^2}$

SS due to treatments $= \dfrac{(A_{total})^2 + (B_{total})^2 + (C_{total})^2 + (D_{total})^2 + (E_{total})^2}{m} - CF$

$SSE = TSS - SSR - SSC - SST$

$F_R = \dfrac{MSR}{MSE}$, $F_c = \dfrac{MSC}{MSE}$, $F_T = \dfrac{MST}{MSE}$

Efficiency of LSD over CRD is $\dfrac{S_R^2 + S_c^2 + (m-1)S_E^2}{(m+1)S_E^2}$

Calculations : The given values are very large. So, far as computational work concered as the value of F which is the ratio of variances is not affected by changing either origin or scale. Therefore we subtract 400 from each of the given values.

Grand total $= G = 1335$

$m = 5$, $m^2 = 25$

$C.F = \dfrac{G^2}{m^2} = \dfrac{(1335)^2}{25} = \dfrac{1782225}{25} = 71289$

Raw SS $= \sum\sum\sum y_{ijk}^2 = 135969$

Total SS $=$ Raw SS $- CF = 135969 - 71289 = 64680$

$SSR = 11441.2$

$SSC = 3806$

$$SST = \frac{(422)^2 + (244)^2 + (50)^2 + (151)^2 + (468)^2}{5} - 71289 = 25100$$

$$SSE = TSS - SSR - SSC - SST = 64680 - 11441.2 - 3806 - 25100$$
$$= 24332.8$$

$$MSR = \frac{SSR}{m-1} = \frac{11441.2}{5-1} = \frac{114412}{40} = 2860.3$$

$$MSC = \frac{SSC}{m-1} = \frac{3806}{5-1} = 951.5$$

$$MST = \frac{SST}{m-1} = \frac{25100}{5-1} = \frac{25100}{4} = 6275$$

$$MSE = \frac{SSE}{(m-1)(m-2)} = \frac{24332.8}{(5-1)(5-2)} = \frac{24332.8}{4(3)} = 2027.73$$

$$F_R = \frac{MSR}{MSE} = \frac{2860.3}{2027.73} = 1.41059$$

$$F_C = \frac{MSC}{MSE} = \frac{951.5}{2027.73} = 0.4692$$

$$F_T = \frac{MST}{MSE} = \frac{6275}{2027.73} = 3.0945$$

ANOVA Table

Source of variation	df	SS	mean SS	Variance ratio F
due to rows	4	11441.2	2860.3	$F_R = 1.41059$
due to columns	4	3806	951.5	$F_C = 0.4692$
Treatments	4	25100	6275	$F_T = 3.0945$
due to error	12	24332.8	2027.73	
Total	24	64680		

Relative efficiency of LSD over RBD when rows are taken as blocks is $\frac{S_C^2 + (m-1) S_E^2}{m S_E^2}$

$$= \frac{951.5 + 4(2027.73)}{5(2027.73)} = 0.89384$$

Relative efficiency of LSD over RBD when columns are taken as blocks is $\frac{S_R^2 + (m-1) S_E^2}{m S_E^2} = \frac{2806.3 + 4(2027.73)}{5(2027.73)}$

$$= 1.08211$$

Relative efficiency of an LSD over CRD is –

$$\frac{S_R^2 + S_C^2 + (m-1) S_E^2}{(m+1) S_E^2} = \frac{2806.3 + 951.5 + 4(2027.73)}{6(2027.73)} = 0.9755$$

Tabulated $F_{0.05}(4, 12) = 3.259$

$F_R < 3.259$

$F_C < 3.259$

$F_T < 3.259$

Hence, the mean effect due to rows, columns, treatments are equally effective as cal F < tab F in each case.

Result : The treatments A, B, C, D, E all are equally effective.