Determination of the Sampling Technique to Estimate the Performance at Grade Five Scholarship Examination

W.M.R. Kumari and S. Samita[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Peradeniya, Sri Lanka

ABSTRACT. *Sampling techniques have been given much emphasis since the precision of estimates greatly depends on it. Different methods of sampling have to be investigated to find out the best sampling technique for a given situation. This study was conducted to determine the best sampling technique to estimate performance at the grade five-scholarship examination. The study was done using results of the grade five-scholarship examination held in years 2001, 2002, 2003 and 2004. Simple random sampling and stratified random sampling with different allocations were investigated in the study. Stratification of the population was based on district, gender, medium and two-way stratification by gender and district. The random samples were drawn from each stratum by three types of allocations; proportional allocation, equal allocation and Neyman allocation. The variance of sample mean was calculated for each sampling procedure for the sample sizes 1, 2, 3, 4, 5 and 10% of the population. There was 2-10% gain in precision from simple random sampling to stratified random sampling except for equal allocation. Two-way stratification (by gender and district) was more effective (on average 3% increase in precision) than one-way stratification. Overall, the proportional allocation with two-way stratification was the most appropriate sampling technique to estimate the performance.*

INTRODUCTION

The precision of an estimate made from a sample depends both on the procedure by which the estimate is calculated and on the plan of sampling (Cochran, 1977). Since, there are various sampling procedures, investigation of these techniques is important to find out the best sampling technique for a given situation.

Large number of sample surveys have been conducted in the area of education in Sri Lanka. Some of these surveys were to study students' performance at national examinations. Estimates from those sample surveys had been useful to implement various new educational policies (www.moe.lk). In Sri Lanka, about 4 million children attend government schools. Sri Lanka has near universal gender equity in access to primary and secondary education. The primary education cycle takes student through grades 1-5 while the junior secondary and senior secondary cycles involve grades 6-8 and 9-13, respectively. The grade five-scholarship examination is one of the national examinations conducted by Department of Examinations.

[1] Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.

The purpose of this examination is to offer scholarships to the best-performed students to study in popular schools for their secondary education. Students of government and semi government schools are eligible to sit for this examination. In each year there are about 300,000 students from all 25 districts who sit for this examination. It consists of two papers to evaluate students' knowledge on Mathematics and General Intelligence. At present it has become a highly competitive examination in the country.

The role of sampling theory is that it can be used to develop a method of sample selection, which is economical and easy to operate. Thereby it makes sampling more efficient and provides estimates that are precise enough for our purposes (Barnett, 1974). Basically there are two approaches to sampling *i.e.* probability sampling and non-probability sampling. Probability sampling makes conclusions that apply to the population as a whole. Simple random sampling and stratified random sampling are some common probability sampling procedures. Therefore, these sampling procedures were given much emphasis in the present study. The objective of the study was to determine the most appropriate sampling procedure to study the performance at the grade-five scholarship examination.

MATERIALS AND METHODS

Secondary data

The present study was carried out using grade five scholarship examination results (complete population data) of years 2001, 2002, 2003 and 2004. The data set consisted of the variables; school number, serial number, district number, marks of subject-1, marks of subject-2, gender and medium. Table 1 shows the basic information. There is virtually an equal male to female ratio while the Sinhala to Tamil students' ratio is about 4:1. The dataset was summarized and analyzed using MINITAB statistical software (Minitab, 1991).

Table 1. Basic information about the data.

Year	Total No.	No. of males	No. of females	No. of Sinhala	No. of Tamil	No. of absentees
2001	292579	144613	147966	230875	61704	17957
2002	306252	153042	153239	240764	65517	22839
2003	311044	156045	154999	242570	68474	18599
2004	306234	154166	152068	237375	68859	16459

Sampling techniques investigated

In the present study, sampling methods were investigated under two main techniques. They were simple random sampling and stratified random sampling. Both techniques were investigated for the sample sizes 1, 2, 3, 4, 5, and 10% of the population. Since there was a significant positive correlation (r=0.8; $p<0.05$) between subject 1 and subject 2, the total of the both subjects was used in the analysis.

Under simple random sampling, samples were drawn from each of the year 2001 to 2004 using MINITAB statistical software. Sampling units were drawn from population without replacement.

The precision of estimates under simple random sampling was investigated using the measure, variance of sample mean, $V(\bar{y})$, where;

$$V(\bar{y}) = \frac{(1-f)}{n} S^2 \tag{1}$$

and

$$S^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1} \tag{2}$$

$\bar{Y}$ is population mean, n is the sample size, N is the size of the population and f is sampling fraction (n/N).

Under stratified random sampling, the population was first divided into strata. Strata in this study were based on the criteria, district (25 strata), gender (2 strata), medium (2 strata) and two-way stratification by gender and district (50 strata). In order to choose stratum sample size, three types of allocations were considered namely, proportional allocation, equal allocation and *Neyman* allocation. The notations used in the computation under stratified random sampling were N_h-Total number of units in h^{th} stratum, N-Population size, n_h-Number of units in sample in the h^{th} stratum, n-Sample size, y_{hi}-Value obtained from the i^{th} sampling unit in the h^{th} stratum,

$$f_h = \frac{n_h}{N_h}$$ - sampling fraction for the h^{th} stratum, (3)

$$W_h = \frac{N_h}{N}$$ - stratum weight, (4)

$$\bar{y}_h = \sum_{h=1}^{n_h} \frac{y_{hi}}{n_h}$$ - sample mean in the h^{th} stratum, (5)

$$\bar{Y}_h = \sum_{h=1}^{N_h} \frac{y_{hi}}{N_h}$$ - true mean of the h^{th} stratum, (6)

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2}{N_h - 1}$$ - True variance of the h^{th} stratum. (7)

The precision of estimates under stratified random sampling were investigated using the measure $V(\bar{y}_{st})$, where;

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 V(\bar{y}_h) \tag{8}$$

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{9}$$

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{10}$$

L is the number of strata, $V(\bar{y}_h)$ is the variance of sample mean from h^{th} stratum given by;

$$V(\bar{y}_h) = \frac{S_h^2}{n_h}\left(\frac{N_h - n_h}{N_h}\right) \tag{11}$$

In the proportional allocation, the specific number of sampling units was drawn from each stratum proportional to the stratum size *i.e.*;

$$n_h = \frac{nN_h}{N}.$$

Variance of sample mean under proportional allocation is given by;

$$V(\bar{y}_{st}) = \left(\frac{1-f}{n}\right)\sum W_h S_h^2 \tag{12}$$

In the case of equal allocation, the same number of sample units were drawn from each stratum, *i.e.* $n_1 = n_2 = \ldots\ldots\ldots = n_L$ Variance of the sample mean for equal allocation is given by;

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S^2{}_h}{n_h}(1 - f_h) \tag{13}$$

The *Neyman* allocation is a special case of optimum allocation. If cost per sample unit is the same for all strata, then optimum allocation for fixed cost reduces to fixed sample size. Accordingly,

$$n_h = n\frac{N_h S_h}{\sum N_h S_h} \tag{14}$$

and the variance of sample mean under *Neyman* allocation is given by,

$$V(\bar{y}_{st}) = \frac{\left(\sum W_h S_h\right)^2}{n} - \frac{\sum W_h S_h^2}{N} \tag{15}$$

RESULTS AND DISCUSSION

Population statistics

The basic statistics of marks for subject-1, subject-2 and subject-total are shown in Table 2. The population mean of the subject-1 during 2001-2004 varied from 41 to 48. For subject 2 mean varied from 26 to 31. It is clear that marks for subject-1 was greater than that of subject-2. The maximum marks for subject-1 was 100 in all four years and for subject-2 the maximum was around 97. The population mean of subject-total was around 75. The maximum marks for subject-total varied 191 to 193. Overall population statistics were consistent in all four years.

Table 2. Population statistics of grade 5-scholarship examination results.

Year	Variable	Mean	Median	CV	Minimum	Maximum
2001	Subject 1	45.10	45	37.01	0	100
	Subject 2	30.99	26	66.53	0	98
	Subject Total*	76.10	73	46.58	0	191
2002	Subject 1	48.63	53	45.44	0	100
	Subject 2	26.02	21	69.55	0	96
	Subject Total*	74.66	73	51.39	3	192
2003	Subject 1	41.38	43	45.53	0	100
	Subject 2	31.71	28	68.34	0	98
	Subject Total*	73.1	70	53.43	0	193
2004	Subject 1	45.45	48	38.78	0	100
	Subject 2	30.48	27	65.57	0	97
	Subject Total*	75.94	74	47.31	0	192

Note: CV-Coefficient of variation, *out of 200

Basic statistics of marks by gender

The basic statistics of marks (subject total) according to gender is given in Table 3. Mean mark had varied from 69 to 73 for male students, whereas for female it had varied from 78 to 80 during the period of 2001-2004. Mean marks of both subjects were higher for female students than that for the male counterparts in all four years. However, subject-total maximum was higher for male students in 2001 and 2004.

Table 3. Basic statistics of the results by gender.

Year	Gender	Mean	Median	CV	Minimum	Maximum
2001	Male	73	69	49.99	0	191
	Female	79	76	43.22	0	189
2002	Male	71	69	55.57	3	192
	Female	78	77	47.28	3	192
2003	Male	69	64	57.82	0	187
	Female	78	76	48.97	0	193
2004	Male	73	71	50.88	0	192
	Female	80	77	43.67	0	188

Basic statistics of marks by medium

The grade five-scholarship examination is held in both Sinhala and Tamil media. Basic statistics of marks of subject-total is given in Table 4. The mean marks of subject-total varied from 77 to 79 and 63 to 67 for Sinhala and Tamil media respectively during four years. In addition mean marks for each subject were also higher for Sinhala medium than that of the Tamil medium for all years.

Table 4. Basic statistics of results by medium.

Year	Medium	Mean	Median	CV	Minimum	Maximum
2001	Sinhala	79	76	44.85	1	191
	Tamil	65	59	51.52	0	186
2002	Sinhala	78	77	49.63	3	192
	Tamil	63	59	56.21	3	177
2003	Sinhala	77	75	50.89	0	193
	Tamil	59	53	60.57	0	182
2004	Sinhala	78	76	45.32	0	192
	Tamil	67	64	53.7	0	182

Mean and variance of subject total by district

The mean and variance of subject-total by district is given in Table 5. The highest mean of subject-total was recorded in Colombo in 2001 and 2002 and Gampaha in 2003 and 2004. The lowest mean of subject total was recorded in Killinochchi district in all four years.

Table 5. Basic statistics of marks by district.

District	2001			2002			2003			2004		
	Total count	Mean	Variance	Total count	Mean	Variance	Total count	Mean	Variance	Total count	Mean	Variance
Colombo	29679	86	1233.0	30095	86	1379	30369	83	1482	29858	83	1208
Gampaha	27673	83	1312.0	28150	83	1436	28438	85	1569	27760	85	1255
Kuluthara	15718	80	1415.0	15680	78	1629	15854	77	1670	15859	79	1411
Kandy	21244	76	1235.0	21647	76	1464	21580	74	1517	21236	77	1206
Matele	7099	66	1231.0	7664	65	1416	7647	65	1408	7540	68	1193
N_eliya	10166	62	1015.0	10883	60	1283	11236	59	1264	11119	65	1124
Galle	17325	81	1285.0	17678	79	1559	16962	77	1619	17343	77	1363
Matara	12684	79	1227.0	13846	76	1496	13249	77	1532	13207	78	1339
Hamban	9738	76	1238.0	10370	74	1545	10477	73	1424	10429	75	1280
Jaffna	7919	73	1254	8228	69	1321	9200	67	1400	9514	74	1334
Killinoch	2065	60	998.7	2294	53	978.3	2038	47	1104	2032	53	1107
Mannar	1190	77	1033	1265	73	1117	1471	65	1009	1500	73	1011
Vuwniya	2500	73	1204	2603	68	1139	2857	66	1242	2763	73	1182
Mulathie	2077	66	1151	2228	61	1129	1674	55	1054	1741	59	1174
Baticoloa	6864	69	1311	7787	65	1401	7955	61	1445	8372	66	1510
Ampara	9826	73	1090	10609	72	1367	11448	68	1362	11052	73	1200
Trincomal	5408	65	1240	5887	65	1466	6847	62	1448	6545	69	1426
Kurunegal	23269	80	1168	24216	79	1413	24960	79	1481	23932	82	1194
Putlam	10565	71	1076	11972	69	1315	12243	66	1286	12219	69	1152
Anuradha	13663	69	1040	14542	69	1249	15582	68	1320	14159	72	1115
Polonnaru	6161	71	1165	7045	70	1312	7250	69	1283	6689	72	1157
Badulla	12673	74	1111	13877	70	1315	13734	67	1383	13789	72	1192
Monraga	7674	64	1136	8091	61	1365	8771	59	1367	8677	63	1189
Ratnepura	17012	73	1327	17440	70	1571	17106	69	1585	16791	72	1332
Kegalle	12387	78	1181	12155	77	1399	12096	76	1479	12108	81	1233

Precision of estimates from simple random sampling

Figure 1 shows the $V(\bar{y})$ against the sample size (sampling percentage) from the population. The variances are almost similar in all years but decreased with the increase of sample size as expected.

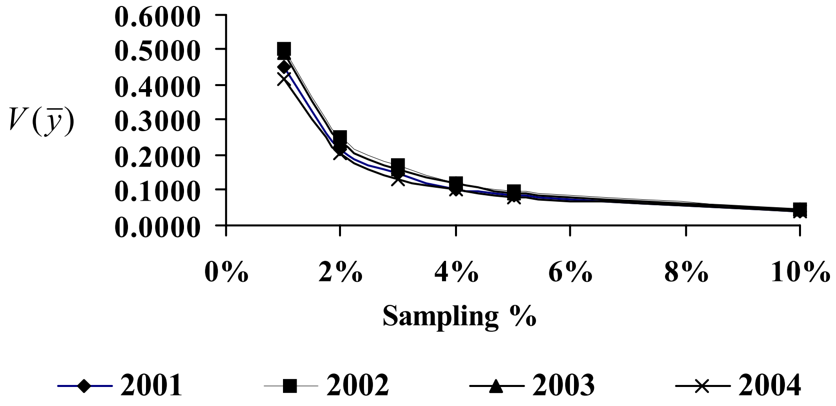

Fig. 1. Change of $V(y)$ against sampling percentage.

The variance reduction with different sample sizes relative to 1% sampling is given in Table 6. It is clear that when the sample size increased from 1% to 2%, the variance reduction was approximately 50%. Therefore, gain in precision from 1% to 2% sampling is about 100%. Similarly, when the sample size was increased further up to 10% $V(\bar{y})$ gradually decreased leading to gradual gain in precision.

Table 6. Variance reduction with respect to 1% sampling.

Sample size	Reduction in variance (%) relative to 1% sampling			
	2001	2002	2003	2004
2% sampling	52	50	52	50
3% sampling	67	66	68	68
4% sampling	77	76	76	75
5% sampling	81	80	81	81
10% sampling	91	91	91	91

Precision of estimates from stratified random sampling

Stratification of the population was done by district, gender, medium and by both district and gender together. Number of strata in each stratification was 25, 2, 2, and 50 respectively. $V(\bar{y}_{st})$ obtained under proportional allocation, equal allocation and *Neyman* allocation for a fixed sample size (1% sampling) is given in Table 7 and Figure 2.

Table 7. $V(\bar{y}_{st})$ **obtained under different types of allocations with each stratification at 1% sampling.**

Stratification variable	Year	Variance of sample mean $V(\bar{y}_{st})$		
		Proportional allocation	Equal allocation	*Neyman* allocation
District	2001	0.4109	0.5915	0.4101
	2002	0.4575	0.6387	0.4566
	2003	0.4646	0.6556	0.4784
	2004	0.4035	0.5543	0.4030
Gender	2001	0.4220	0.4218	0.4259
	2002	0.4723	0.4723	0.4717
	2003	0.4796	0.4797	0.4793
	2004	0.4144	0.4146	0.4139
Medium	2001	0.4142	0.5626	0.4097
	2002	0.4631	0.6317	0.4625
	2003	0.4681	0.6317	0.4675
	2004	0.4102	0.5339	0.4102
District and gender	2001	0.4074	0.5909	0.4062
	2002	0.4534	0.6380	0.4521
	2003	0.4578	0.6457	0.4567
	2004	0.4003	0.5339	0.3992

The relative precision of estimates by types of allocation is discussed under each stratification criterion.

Stratification by gender

Since, there was equal male to female ratio in the population, stratification by gender resulted in equal stratum size. Thus the proportional allocation is equivalent to equal allocation. This was reflected in the precision of estimates obtained under proportional and equal allocations. Both allocations showed more or less similar precision compared to *Neyman* allocation.

Stratification by medium

The population was stratified into two strata based on the medium. Since the Sinhala to Tamil ratio was 4:1, the strata were not equal in size. It was observed that the highest precision was obtained under *Neyman* allocation followed by the proportional allocation. The lowest precision was resulted under equal allocation.

Stratification by district

Since, there are 25 districts in the country, the stratification by district resulted 25 strata that were unequal in size. The *Neyman* and proportional allocations gave almost similar and higher precision than the equal allocation.

Two-way stratification

Two-way stratification by district and gender stratified the population into 50 strata. It was observed that precision obtained under *Neyman* and proportional allocations were almost similar and showed higher precision than equal allocation.

The relative precision observed under different allocations with each stratification was consistent in all four years (Fig. 2).

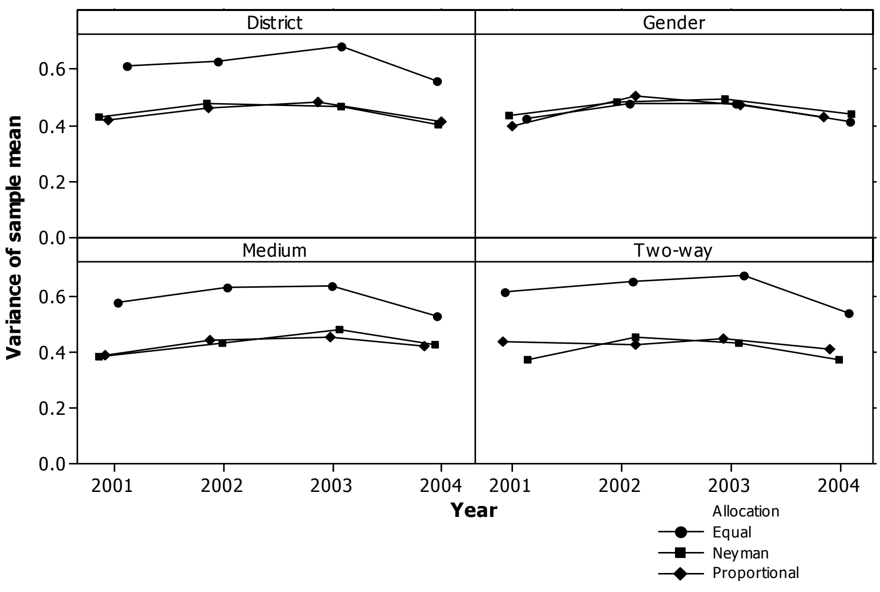

Fig. 2. The change in variance of sample mean $V(\bar{y}_{st})$ over year under different allocations and stratification at 1% sampling.

According to results obtained under each stratification method, (Table 7 and Fig. 2.) *Neyman* and proportional allocations were better than the equal allocation. Since there is no much difference between proportional allocation and *Neyman* allocation and also since the proportional allocation is easier to implement, it can be recommended for sampling.

Precision of stratified random sampling relative to simple random sampling

The percentage gain in precision in stratified random sampling with respect to simple random sampling is shown in Figure 3. It was clear that there was 2-10% gain in precision from simple random sampling to stratified random sampling, except in equal allocation stratified by medium, district and, district and gender. The reason for the gain in precision in equal allocation stratified by gender is that it is equivalent to proportional allocation.

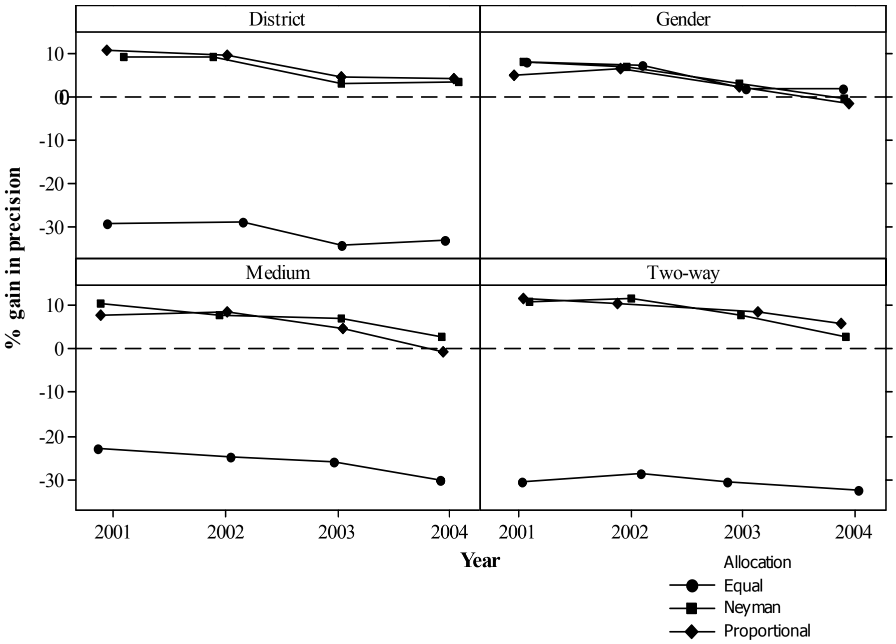

Fig. 3. Percentage gain in precision in Stratified Random Sampling relative to simple random sampling over year at 1% sampling.

It is important to note that the highest gain in precision was obtained under two-way stratification. Thus it is possible that further gain in precision could be obtained with more than two-way stratification.

Effect of sample size on precision of estimates under stratified random sampling with different allocations

Figure 4 shows the change in $V(\bar{y}_{st})$ against sample size in year 2001. It clearly shows that precision increases with the increase of sample size under all stratification methods and types of allocation. However, it can be observed that the precision obtained at large sample sizes (10%) for all allocation methods give almost similar precision. Similar pattern was observed in all four years.

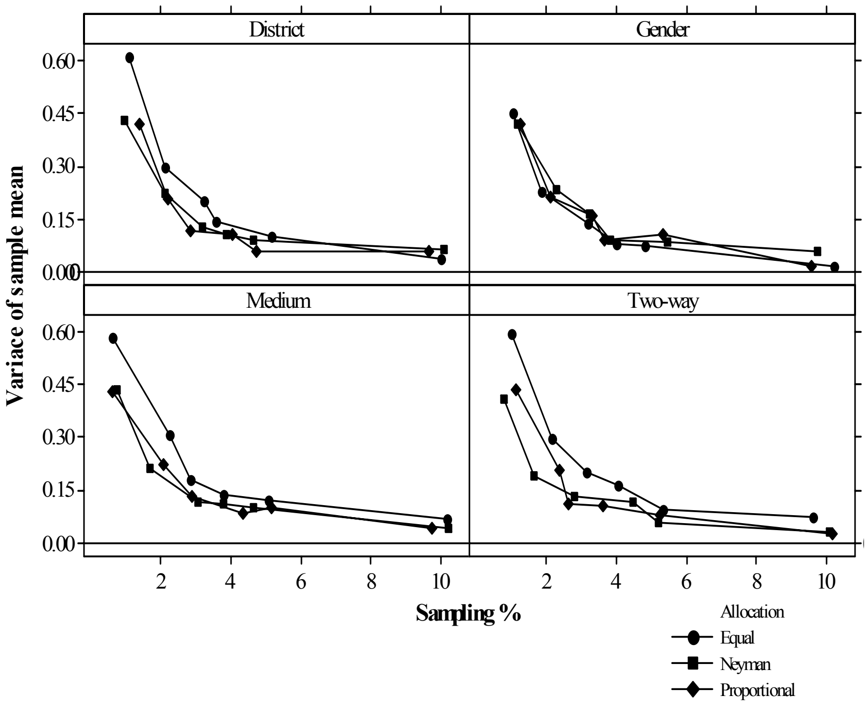

Fig. 4. The change in variance of sample mean $V(\bar{y}_{st})$ against sample size under different allocations and stratification in year 2001.

Although there was not much of a difference in precision among these three types of allocations at 10% sampling, which were about 30,000 units, it is not much of a use since 30,000 would be too much for a study. Thus, the information with respect to small sample sizes such as 1% and 2% would be more useful. Accordingly, stratified random sampling with proportional and *Neyman* allocations are the appropriate sampling techniques for similar studies.

CONCLUSIONS

Surveys of this nature are quite common but not many studies have been carried out to determine the most appropriate sampling procedure for these studies. Hence there is a requirement for such studies since sampling method plays an important role in precision of estimates. In the study, the methodology looked into important aspects such as types of stratification criteria, types of allocation and sample sizes. Hence, findings are applicable to possible different situations. In general, stratified random sampling was found to be superior to simple random sampling. Within stratified random sampling, proportional and *Neyman* allocations gave higher precision than equal allocation. The difference between *Neyman* and proportional allocation was marginal. Since the proportional allocation is more convenient compared to *Neyman* allocation, proportional allocation is recommended for studies of this nature. In addition, it can be concluded that two-way stratification produces more precise estimates than one-way stratification.

ACKNOWLEDGEMENTS

The GTZ-Basic education sector program of Ministry of Education through MG Consultancies, Colombo is acknowledged for the source of data.

REFERENCES

Barnett, V. (1974). Element of Sampling Theory, The Universities Press Ltd, St. Paul's House, Warwick Lane, London.

Cochran, W.G. (1977). Sampling Techniques, Wiley, 2nd edition, New York.

Minitab. (1991). Minitab Statistical Software, copyright Minitab Inc., Texas, USA.

Ministry of Education, Sri Lanka. [online] (cited on 15.12.2005). http://www.moe.lk.