General Household Survey 2005
Sample Design and Sampling Errors

Introduction

1 The General Household Survey (GHS) is a mid-decade national survey collecting socio-economic information on the population and households in Singapore. It is conducted once in ten years in between the decennial Population Censuses. The GHS 2005 is the second in the series.

2 The GHS 2005 built upon the experiences of the Population Census 2000, combining a register-based approach with a large-scale sample survey. Under the register-based approach, basic demographic data such as age, sex and ethnic group are obtained from the Singapore Department of Statistics (DOS)' Household Registration Database. Additional data items are collected from a sample survey of 90,000 dwelling units.

Sample Design

3 The sampling frame for the selection of sample for GHS 2005 is based on the National Database of Dwellings (NDD) maintained by DOS. The NDD is a register of all residential dwelling units in Singapore, and is updated monthly based on administrative data. As the sample survey of the GHS 2005 covers only private households in residential dwellings, institutions such as military camps, hospitals and hostels are excluded from the frame.

4 The sample for GHS 2005 is selected based on DOS's main sampling methodology which is a 2-stage stratified design. The primary sampling units for the first stage consist of Sampling Divisions, which are geographical locations based on planning areas used by the Urban Redevelopment Authority of Singapore. The dwelling units are the sampling units in the second stage of selection.

First Stage Sample Selection of Sampling Divisions

5 During the first stage selection, the Sampling Divisions are stratified into three strata by its predominant housing types, namely, 'Public Flats', 'Private Houses and Flats' and 'Others', the last group of which included attap/zinc-roofed houses and shophouses. Within each stratum, the Sampling Divisions are arranged by geographical location.

6 From this master list, 200 Sampling Divisions are systematically selected with probability proportional to the number of dwelling units in each Sampling Division.

Second Stage Sample Selection of Dwelling Units

7 The second stage involved the selection of dwelling units from the selected Sampling Divisions. Within each selected Sampling Division, the dwelling units are stratified by detailed housing types such as HDB dwelling units and other public flats, landed properties and private flats/apartments, attap/zinc-roofed houses and shophouses. For every chosen Sampling Division, 450 dwelling units are selected by systematic sampling procedure with a random start.

8 This sampling method yields an overall equal probability and self-weighting sample of 90,000 dwelling units, and ensures good control over the desired sample size. The selected sample is representative of the national dwelling type distribution.

Sampling Errors

9 The accuracy of estimates derived from the sample survey of the GHS 2005 are affected by sampling errors, since the estimates are based on information obtained from a fraction of the population instead of the whole population.

Concept

10 Sampling errors refers to the difference between the estimate based on a sample and its 'true' population value that would result if the whole population has been surveyed. The extent of sampling error of an estimate under a particular sample design is assessed by the variability of the estimate across all possible samples under the design. One common measure of this variability is given by the standard error (SE), which is the standard deviation of the sampling distribution of the estimate.

11 According to probability theory, about 68, 95 and 99 per cent of estimates from all possible samples will fall within the interval defined by one, two or three standard errors respectively on either side of the 'true' population value. Based on this property, it is possible to construct an estimated range of values which is likely to include the 'true' population value with a known level of probability or confidence using the estimate from the particular selected sample and an estimate of its standard error. For a specified confidence level, the smaller the standard error of an estimate, the narrower is the range of possible values for its 'true' population value. By statistical convention, the level of confidence is usually set at 95 per cent.

Computation of Sampling Error Based on GHS 2005 Sample Design

12 The formula to be used for computing standard errors of an estimate from a survey is dependent on the adopted sample design. The GHS 2005 sample is based on a complex sample design, and therefore has complicated sampling error computation formula for sample estimate such as the total number of elements in the population with a given attribute Y (T_Y) and the proportion of the total population with a given attribute Y (P_Y).

13 In practice, it would be inefficient and costly to compute standard errors for estimates using the sampling error computation formula. One way to overcome this problem is to ascertain the relationship between the sampling errors of an estimate under the sample design used and those of a simple random sample design. The formula for standard error of T_Y and P_Y based on simple random sample are:

$$\begin{aligned} SE(T_Y; SRS) &= \sqrt{Var(T_Y;SRS)} \\ &= \sqrt{Var(N \times P_Y;SRS)} \\ &= N \times \sqrt{Var(P_Y;SRS)} \\ &= N \times SE(P_Y; SRS) \\ &= N \times \sqrt{\frac{N-n}{N-1} \times \frac{P_Y(1-P_Y)}{n}} \end{aligned} \quad [1]$$

where N is the total population count;
n is the count of persons covered in the sample.

14 The ratio of the variance of the estimate based on the sample design used and that of a simple random sample of the same size is known as the "design effect" (deff). This ratio gives the net effect of the various complexities of the design used on the variance relative to a simple random sample design. The square root of the design effect (deft) gives the ratio of the standard error of the estimate under the sample design used to that of a simple random sample:

$$deft(T_Y) = \sqrt{deff(T_Y)} = \sqrt{\frac{Var(T_Y)}{Var(T_Y;SRS)}} = \frac{SE(T_Y)}{SE(T_Y;SRS)}$$

This implies that

$$\begin{aligned} SE(T_Y) &= deft(T_Y) \times SE(T_Y;SRS) \\ &= deft(P_Y) \times SE(T_Y;SRS), \end{aligned} \quad [2]$$

since $T_Y = N \times P_Y$.

15 Estimate of deff (P_Y) is obtained directly with the use of statistical software packages such as IVEware[1]. The IVEware would take into account the various complex features of the GHS 2005 sample design during the estimation process.

Results on Sampling Errors for Selected T_Y

16 Estimated sampling errors and deft of T_Y for selected attributes in the GHS 2005 are given in Table A1 in Annex. The size of SE generally increases with the size of sample estimates of T_Y, so that the larger the sample estimate, the larger is the SE. As such, a better indicator of the precision of a sample estimate would be the relative standard error (RSE), which is obtained by expressing the standard error as a percentage to the estimate. The smaller the RSE, the more precise is the estimate. In this connection, it may be noted that the RSE generally decreases as the size of the estimate increases. This implies that sample estimate of the population with a rare characteristic would have high RSE and users would have to be careful in drawing inferences based on the sample estimate.

Generalized sampling errors tables

17 From Table A1, the deft of T_Y for selected attributes ranged between 1.4 and 6.2, and the average deft was about 3. As it is impractical to compute and display the sampling error for each and every of the possible estimates from the GHS 2005, generalized sampling errors tables which provide standard errors of a selected range of estimates (totals and proportions) with deft value of 1, 3 and 5 are shown instead (Table A2 – A4 in Annex).

18 Table A2 (deft = 1) relates to Simple Random Sampling and is presented mainly for theoretical interest. Since most of the selected attributes in Table A1 have deft between 2 and 4, data users should generally use Table A3 (deft = 3) to determine sampling errors for the attribute of their own interest. For attributes with larger design effect, Table A4 (deft = 5) could be used. From these tables, it can be noted that the smaller the size of estimates, the larger is the RSE.

Singapore Department of Statistics
May 2007

[1] The IVEware is an 'Imputation and Variance Estimation Software' developed by the Survey Research Centre of the Institute of Social Research in the University of Michigan (http://www.isr.umich.edu/src/smp/ive/).

Table A1. Sampling Errors and deft of T_Y for Selected Attributes

	Sample Estimate	Standard Error	Relative Standard Error	95% Confidence Interval		deft (T_Y)
	T_Y	SE (T_Y)	RSE (T_Y)	Lower	Upper	
Residents Aged 15 & Over:						
Single	858,133	8,682	1.0	841,116	875,150	2.98
Married	1,700,462	14,371	0.8	1,672,296	1,728,628	4.02
Ever-Married Resident Females:						
With 1 - 2 Children Born	334,389	3,308	1.0	327,905	340,874	1.70
With 3 – 4 Children Born	284,789	2,969	1.0	278,969	290,609	1.64
With 5 or more Children Born	95,220	2,183	2.3	90,941	99,499	2.04
Resident Students:						
Attending Primary and Below Level	381,349	5,154	1.4	371,246	391,451	2.49
Attending Upper Secondary Level	55,747	1,274	2.3	53,251	58,244	1.55
Resident Non-students:						
With Secondary Qualifications	534,987	6,367	1.2	522,508	547,466	2.65
With Polytechnic Qualifications	205,066	3,309	1.6	198,580	211,552	2.13
Residents Aged 5 Years & Over:						
Speaking Mandarin at Home	1,133,725	16,715	1.5	1,100,963	1,166,487	5.20
Speaking Tamil at Home	98,417	4,000	4.1	90,577	106,256	3.67
Resident Working Persons:						
In Professional Position	204598	5,719	2.8	193,389	215,807	3.69
In Sales and Services Position	239404	4,609	1.9	230,371	248,438	2.76
In Production Position	110602	2,719	2.5	105,273	115,930	2.36
In Manufacturing	275,176	5,620	2.0	264,160	286,192	3.15
In Financial Services	102,134	2,944	2.9	96,364	107,904	2.66
In Business Services	207,043	2,221	1.1	202,689	211,398	1.43
Employees	1,415,107	7,792	0.6	1,399,835	1,430,379	2.27
Employers	75,507	2,631	3.5	70,351	80,664	2.75
Travelling by Public Bus only	360,571	10,355	2.9	340,274	380,867	5.13
Requiring no transport	125,294	3,322	2.7	118,784	131,804	2.71
Residents Aged 15 & Over Who Have Travelled Overseas in Last 12 Months:						
For Work-Related Purposes	220,498	4,837	2.2	211,017	229,978	3.01
For Holidays	837,497	6,072	0.7	825,596	849,398	2.10
Resident Households:						
In Owner-Occupied Dwelling	958,648	6,168	0.6	946,559	970,737	6.12
In Rented Dwelling	86,619	6,149	7.1	74,566	98,672	6.22
No Family Nucleus	154,106	2,518	1.6	149,172	159,041	1.98
One Family Nucleus	844,822	5,731	0.7	833,589	856,056	4.03
Multi-Family Nucleus	57,054	1,271	2.2	54,563	59,545	1.56

Table A2. Sampling Errors for Square Root of Design Effect (deft) Equals 1

Size of Estimates	Proportion of Total Population (%)	Standard Error	Relative Standard Error (%)	95% Confidence Interval	
				Lower	Upper
			PERSONS		
4,000,000	91.92	1,995	0.05	3,996,089	4,003,911
3,500,000	80.43	2,905	0.08	3,494,306	3,505,694
3,000,000	68.94	3,389	0.11	2,993,358	3,006,642
2,500,000	57.45	3,621	0.14	2,492,904	2,507,096
2,000,000	45.96	3,650	0.18	1,992,847	2,007,153
1,500,000	34.47	3,480	0.23	1,493,178	1,506,822
1,000,000	22.98	3,081	0.31	993,962	1,006,038
750,000	17.24	2,766	0.37	744,579	755,421
500,000	11.49	2,335	0.47	495,423	504,577
250,000	5.75	1,704	0.68	246,660	253,340
100,000	2.30	1,097	1.10	97,849	102,151
75,000	1.72	953	1.27	73,132	76,868
50,000	1.15	780	1.56	48,470	51,530
25,000	0.57	553	2.21	23,915	26,085
10,000	0.23	351	3.51	9,313	10,687
7,500	0.17	304	4.05	6,905	8,095
5,000	0.11	248	4.96	4,514	5,486
2,500	0.06	175	7.02	2,156	2,844
1,000	0.02	111	11.10	782	1,218
500	0.01	78	15.70	346	654
			HOUSEHOLDS		
1,000,000	95.33	758	0.08	998,514	1,001,486
750,000	71.50	1,622	0.22	746,822	753,178
500,000	47.66	1,794	0.36	496,484	503,516
250,000	23.83	1,530	0.61	247,000	253,000
100,000	9.53	1,055	1.05	97,932	102,068
75,000	7.15	925	1.23	73,186	76,814
50,000	4.77	765	1.53	48,500	51,500
25,000	2.38	548	2.19	23,926	26,074
10,000	0.95	349	3.49	9,316	10,684
7,500	0.71	303	4.04	6,907	8,093
5,000	0.48	247	4.95	4,515	5,485
2,500	0.24	175	7.01	2,157	2,843
1,000	0.10	111	11.09	783	1,217
500	0.05	78	15.68	346	654

Table A3. Sampling Errors for Square Root of Design Effect (deft) Equals 3

Size of Estimates	Proportion of Total Population (%)	Standard Error	Relative Standard Error (%)	95% Confidence Interval	
				Lower	Upper
			PERSONS		
4,000,000	91.92	5,986	0.15	3,988,268	4,011,732
3,500,000	80.43	8,715	0.25	3,482,918	3,517,082
3,000,000	68.94	10,166	0.34	2,980,075	3,019,925
2,500,000	57.45	10,862	0.43	2,478,711	2,521,289
2,000,000	45.96	10,949	0.55	1,978,541	2,021,459
1,500,000	34.47	10,441	0.70	1,479,535	1,520,465
1,000,000	22.98	9,242	0.92	981,885	1,018,115
750,000	17.24	8,297	1.11	733,737	766,263
500,000	11.49	7,006	1.40	486,268	513,732
250,000	5.75	5,112	2.04	239,980	260,020
100,000	2.30	3,292	3.29	93,548	106,452
75,000	1.72	2,859	3.81	69,396	80,604
50,000	1.15	2,341	4.68	45,411	54,589
25,000	0.57	1,660	6.64	21,746	28,254
10,000	0.23	1,052	10.52	7,938	12,062
7,500	0.17	911	12.15	5,714	9,286
5,000	0.11	744	14.89	3,541	6,459
2,500	0.06	526	21.06	1,468	3,532
1,000	0.02	333	33.30	347	1,653
500	0.01	235	47.10	38	962
			HOUSEHOLDS		
1,000,000	95.33	2,274	0.23	995,543	1,004,457
750,000	71.50	4,865	0.65	740,465	759,535
500,000	47.66	5,382	1.08	489,451	510,549
250,000	23.83	4,591	1.84	241,001	258,999
100,000	9.53	3,165	3.16	93,797	106,203
75,000	7.15	2,776	3.70	69,558	80,442
50,000	4.77	2,296	4.59	45,500	54,500
25,000	2.38	1,644	6.57	21,779	28,221
10,000	0.95	1,047	10.47	7,948	12,052
7,500	0.71	908	12.11	5,721	9,279
5,000	0.48	742	14.84	3,545	6,455
2,500	0.24	525	21.02	1,470	3,530
1,000	0.10	333	33.26	348	1,652
500	0.05	235	47.04	39	961

Table A4. Sampling Errors for Square Root of Design Effect (deft) Equals 5

Size of Estimates	Proportion of Total Population (%)	Standard Error	Relative Standard Error (%)	95% Confidence Interval	
				Lower	Upper
			PERSONS		
4,000,000	91.92	9,977	0.25	3,980,446	4,019,554
3,500,000	80.43	14,526	0.42	3,471,530	3,528,470
3,000,000	68.94	16,943	0.56	2,966,792	3,033,208
2,500,000	57.45	18,103	0.72	2,464,518	2,535,482
2,000,000	45.96	18,248	0.91	1,964,235	2,035,765
1,500,000	34.47	17,402	1.16	1,465,892	1,534,108
1,000,000	22.98	15,404	1.54	969,808	1,030,192
750,000	17.24	13,829	1.84	722,895	777,105
500,000	11.49	11,677	2.34	477,114	522,886
250,000	5.75	8,520	3.41	233,300	266,700
100,000	2.30	5,486	5.49	89,247	110,753
75,000	1.72	4,765	6.35	65,660	84,340
50,000	1.15	3,902	7.80	42,352	57,648
25,000	0.57	2,767	11.07	19,576	30,424
10,000	0.23	1,753	17.53	6,564	13,436
7,500	0.17	1,519	20.25	4,523	10,477
5,000	0.11	1,240	24.81	2,569	7,431
2,500	0.06	877	35.09	780	4,220
			HOUSEHOLDS		
1,000,000	95.33	3,790	0.38	992,571	1,007,429
750,000	71.50	8,108	1.08	734,109	765,891
500,000	47.66	8,970	1.79	482,418	517,582
250,000	23.83	7,652	3.06	235,002	264,998
100,000	9.53	5,274	5.27	89,662	110,338
75,000	7.15	4,627	6.17	65,930	84,070
50,000	4.77	3,826	7.65	42,500	57,500
25,000	2.38	2,739	10.96	19,631	30,369
10,000	0.95	1,745	17.45	6,579	13,421
7,500	0.71	1,513	20.18	4,534	10,466
5,000	0.48	1,237	24.74	2,576	7,424
2,500	0.24	876	35.03	784	4,216