

Sample Design and Sampling Variability

Introduction

The Census of Population 2010 adopted a register-based approach, similar to the Census of Population 2000. Basic population count and characteristics were compiled from administrative sources. Additional information not available from administrative data sources were collected from a sample survey of about 200,000 dwelling units.

Sample Design and Selection

The sampling frame used for the selection of the sample for Census 2010 is based on the National Database of Dwellings (NDD) maintained by DOS. The NDD is a register of all residential dwelling units in Singapore. As the sample survey of the Census 2010 covered only households in residential dwellings, institutions such as military camps, hostels and hotels were excluded from the frame.

The sample for the Census 2010 was selected based on a stratified design with proportional allocation. The strata were defined based on the planning areas demarcated by the Urban Redevelopment Authority. Each dwelling unit in the frame was first placed into one of the strata based on its planning area. Within each stratum, the units were sorted into dwelling type groupings. A sample was selected using simple random sampling without replacement from each broad dwelling type grouping. The selected samples across the strata were combined to form the required sample of about 200,000 dwelling units.

Sampling Variability

The precision of estimates derived from the sample survey are affected by sampling errors since the estimates are based on information obtained from a fraction of the population instead of the whole population. Sampling errors refer 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. Another measure is 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.

Computation of Sampling Error

The sampling errors of the estimates under the actual sample design used can be derived based on their relationship with those under a simple random sample (SRS) design. The formula for standard error of T_Y based on simple random sample is:

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

where T_Y is the total number of elements in the population with a given attribute Y

N is the total population count

n is the count of persons covered in the sample

P_Y is the proportion of the total population with a given attribute

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:

$$\text{DEFT}(T_Y) = \sqrt{\text{DEFF}(T_Y)} = \sqrt{\frac{\text{Var}(T_Y)}{\text{Var}(T_Y; SRS)}} = \frac{SE(T_Y)}{SE(T_Y; SRS)}$$

This implies that

$$\begin{aligned}
 SE(T_Y) &= \text{DEFT}(T_Y) \times SE(T_Y; SRS) \\
 &= \text{DEFT}(P_Y) \times SE(T_Y; SRS), \quad \text{since } T_Y = N \times P_Y.
 \end{aligned}$$

Generalized Sampling Errors Table

From Table A1, the DEFT of the selected attributes (T_Y) is about 1. It is impractical to compute and display the sampling error for each and every of the possible estimates such as the total number of elements in the population with a given attribute Y from the Census 2010. Thus, a generalized sampling errors

table is provided instead as a guide to data users for estimating the errors of any estimates.

Table A2 provides the generalised sampling errors of a selected range of estimates with DEFT value of 1. The smaller the estimate, the larger is the RSE. This implies that sample estimates of a rare characteristic would have high RSEs and users would have to be careful in drawing inferences based on the sample estimates.

Table A1 Sampling Errors and DEFT of T_Y for Selected Attributes

	Sample Estimate (T_Y)	Standard Error (T_Y)	Relative Standard Error (T_Y)	95% Confidence Interval		DEFT (T_Y)
				Lower	Upper	
<u>Residents Aged 15 Years & Over</u>						
Single	1,000,467	2,103	0.2	996,345	1,004,589	0.99
Married	1,844,606	2,621	0.1	1,839,468	1,849,744	1.02
Widowed	157,556	892	0.6	155,808	159,304	0.96
Divorced/Separated	103,120	729	0.7	101,691	104,549	0.97
<u>Ever-Married Resident Females</u>						
With no Children Born	131,576	867	0.7	129,876	133,276	1.02
With 1 - 2 Children Born	598,132	1,736	0.3	594,729	601,535	1.01
With 3 - 4 Children Born	302,683	1,263	0.4	300,207	305,159	1.00
<u>Resident Students</u>						
Attending Primary and Below	343,642	1,339	0.4	341,018	346,266	1.00
Attending Secondary	231,992	1,103	0.5	229,829	234,155	0.99
Attending Polytechnic	79,708	649	0.8	78,436	80,980	0.98
Attending University	70,386	612	0.9	69,186	71,586	0.98
<u>Resident Non-Students</u>						
With Below Secondary Qualifications	900,147	2,011	0.2	896,205	904,089	0.98
With Secondary Qualifications	526,359	1,596	0.3	523,231	529,487	0.98
With Polytechnic Qualifications	250,213	1,131	0.5	247,996	252,430	0.98
With University Qualifications	634,098	1,777	0.3	630,616	637,580	1.00
<u>Residents Aged 5 Years & Over</u>						
Speaking English at Home	1,097,443	2,163	0.2	1,093,204	1,101,682	0.98
Speaking Mandarin at Home	1,211,505	2,242	0.2	1,207,110	1,215,900	0.98
Speaking Malay at Home	414,475	1,403	0.3	411,724	417,226	0.96
Speaking Tamil at Home	110,667	790	0.7	109,119	112,215	1.01
<u>Residents Aged 15 Years & Over</u>						
Not Literate	128,661	805	0.6	127,083	130,239	0.96
Literate in one language only	878,214	2,001	0.2	874,292	882,136	0.99
Literate in two languages only	1,896,268	2,630	0.1	1,891,114	1,901,422	1.02
Literate in three or more languages	202,606	1,072	0.5	200,506	204,706	1.02

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	78.79	2,188	0.05	3,995,711	4,004,289
3,500,000	68.94	2,477	0.07	3,495,145	3,504,855
3,000,000	59.09	2,632	0.09	2,994,841	3,005,159
2,500,000	49.24	2,676	0.11	2,494,754	2,505,246
2,000,000	39.40	2,616	0.13	1,994,873	2,005,127
1,500,000	29.55	2,442	0.16	1,495,213	1,504,787
1,000,000	19.70	2,129	0.21	995,827	1,004,173
750,000	14.77	1,900	0.25	746,277	753,723
500,000	9.85	1,595	0.32	496,874	503,126
250,000	4.92	1,158	0.46	247,730	252,270
100,000	1.97	744	0.74	98,542	101,458
75,000	1.48	646	0.86	73,734	76,266
50,000	0.98	529	1.06	48,964	51,036
25,000	0.49	375	1.50	24,266	25,734
10,000	0.20	237	2.37	9,535	10,465
7,500	0.15	206	2.74	7,097	7,903
5,000	0.10	168	3.36	4,671	5,329
2,500	0.05	119	4.75	2,267	2,733
1,000	0.02	75	7.51	853	1,147
500	0.01	53	10.62	396	604
200	0.00	34	16.80	134	266
HOUSEHOLDS					
1,150,000	99.85	89	0.01	1,149,825	1,150,175
850,000	73.80	1,023	0.12	847,995	852,005
550,000	47.76	1,162	0.21	547,722	552,278
250,000	21.71	959	0.38	248,120	251,880
100,000	8.68	655	0.66	98,716	101,284
75,000	6.51	574	0.77	73,875	76,125
50,000	4.34	474	0.95	49,071	50,929
25,000	2.17	339	1.36	24,335	25,665
10,000	0.87	216	2.16	9,577	10,423
7,500	0.65	187	2.50	7,133	7,867
5,000	0.43	153	3.06	4,700	5,300
2,500	0.22	108	4.33	2,288	2,712
1,000	0.09	69	6.85	866	1,134
500	0.04	48	9.69	405	595
200	0.02	31	15.33	140	260