1976 CENSUS

OF POPULATION AND

HOUSING

SAMPLING ERRORS ASSOCIATED WITH CENSUS ESTIMATES (REVISED)

CATALOGUE No 2129.0

SAMPLING ERRORS ASSOCIATED WITH CENSUS ESTIMATES

INTRODUCTION

Unlike the preliminary data, the final data from the main processing phase is based on the processing of all Census schedules from non private dwellings, all schedules from the Northern Territory and a 50% sample of private dwellings in the other States and the ACT. Any estimate for the Northern Territory from either the preliminary or main processing phase is not subject to sampling error since all schedules for the Northern Territory were processed. Counts of the total number of males, total number of females, total number of persons for a CD or LGA from the final processing phase were constrained to agree with those from the preliminary processing phase. Therefore, these estimates of total are not subject to sampling error.

WHAT IS SAMPLING ERROR

Since only a 50% sample of private dwelling schedules was processed, it is likely that the estimates derived from the Census would differ from figures which would have been obtained if all schedules were included. These differences are called sampling errors. The sampling error associated with any estimate can be estimated from the sample results and one measure so derived is the standard error. The particular 50% sample selected was one of a large number of possible 50% samples. Each possible 50% sample would have yielded different estimates. The standard error measures the variation of all the possible 50% sample estimates around the figures which would have been obtained if all schedules had been processed.

Given an estimate and the standard error on that estimate, there are about two chances in three that the sample estimate will differ by less than one standard error from the figure that would have been obtained if all schedules had been processed, and about nineteen chances in twenty that the difference will be less than two standard errors.

Another measure of the sampling error is the relative standard error, which is obtained by expressing the standard error as a percentage of the estimate; that is

$$\text{Relative Standard Error} = \frac{\text{Standard Error}}{\text{Estimate}} \times 100$$

Both standard error and relative standard error are used in the following discussion of the reliability of the estimates. An example of their application is as follows:

> EXAMPLE. If an estimate of 70 has a relative standard error of 10%, then the standard error of that estimate is 10% of 70 or 7. Thus there are two chances in three that the figure that would have been obtained if all schedules had been processed will be in the range $70 \pm (1 \times 7)$ or 63 to 77, and nineteen chances in twenty that this figure is between $70 \pm (2 \times 7)$ or 56 to 84.

An estimate of less than 10 persons or dwellings should be treated with caution as it will be subject to such high sampling error (relative standard error generally in excess of 40%), as to detract seriously from its value for most reasonable uses.

PRESENTATION OF SAMPLING ERRORS

It would have been impracticable to publish standard errors of all Census estimates because difficulties in presentation would have been encountered with the large number of estimates. In addition, computer production of all standard errors would have been costly.

Consequently, graphs and tables which relate the relative standard error of an estimate to the size of the estimate are given at the end of this extract. As can be seen from the graphs, the larger an estimate the greater its reliability and thus the smaller the relative standard error. The graphs are not intended to give a precise measure of the error for a particular estimate, but provide an indication of the likely magnitude of the relative standard error for estimates of any particular size.

HOW TO DETERMINE THE SAMPLING ERROR ON AN ESTIMATE

There is no sampling error on an estimate if

- the estimate is total males, total females or total persons in a CD, LGA or aggregations of these areas.
- the estimate relates to the Northern Territory.

If the estimate is greater than 1000 persons or dwellings, then the relative standard error will be less than 2.5% and so the sampling error can be assumed to be negligible for most practical purposes.

The relative standard error or standard error for any other estimate may be found by reference to the graphs and tables given at the end of this extract. A complete description of the methods to be used to obtain the relative standard error for any estimate is given in the sections below.

SAMPLING ERRORS ON DWELLING AND PERSON ESTIMATES

Sampling errors depend on the type of estimate concerned.

(1) For dwelling estimates the relative standard errors are given by LINE D (refer GRAPH 1).

(2) For person estimates refer to GRAPH 2. Sampling errors on person estimates depend on the particular topic of interest. Two groups of topics have been identified:

. Use LINE A if the estimate involves any of the following topics:

Year of arrival; period of residence; birthplace (if overseas); country of citizenship (if overseas); religion; languages regularly used; racial origin.

. Use LINE B for all other topics related to persons.

The difference between the relative standard errors for different person estimates arises because some characteristics are generally similar for persons in the same dwelling, but differ from persons in different dwellings. That is these characteristics are clustered by dwelling (for example, religion and racial origin). The sampling scheme used involved the inclusion of ALL persons in selected dwellings rather than selection of every second person in a dwelling. Hence for characteristics which are clustered by dwelling, there is a greater chance that such persons would have been either undersampled or oversampled. Thus estimates of number of persons classified by characteristics which are clustered by dwelling will have somewhat higher relative standard errors.

If an estimate is known to include a large number of persons from non private dwellings where all schedules were processed (for example an estimate of males ten to fifteen years of age in a CD with a large boarding school for boys), then the relative standard error as read from the graph will overestimate the true relative standard error.

EXAMPLE. Consider an estimate of the number of female university graduates in an LGA. The relative standard error will be derived from LINE B of GRAPH 1. If the estimated number of female university graduates in the LGA is 50, then reading from this line, the relative standard error is approximately 12%. The standard error on the estimate is 50 x 12/100 = 6. Therefore, there are nineteen chances in twenty that the number of female university gradutes in the LGA is in the range $30 \pm (2 \times 6)$ or 18 to 42.

Relative standard errors for dwelling and person estimates may also be determined from Table 1 which sets out relative standard errors for selected estimate sizes.

SAMPLING ERRORS ON ESTIMATES OF PROPORTIONS AND PERCENTAGES

Proportions and percentages formed from the ratio of two Census estimates are also subject to sampling errors. The size of the error depends on the accuracy of both the numerator and the denominator. The formula for the relative standard error of a proportion is given below:

Relative Standard Error (x/y)

$$= \sqrt{\{\text{Relative Standard Error } (x)\}^2 - \{(\text{Relative Standard Error } (y)\}^2}$$

The relative standard error on a percentage is the same as for the corresponding proportion. Thus the relative standard error on an estimate of 58% is the same as that on the proportion .58.

EXAMPLE. Consider an estimate of the labour force participation rate for persons born overseas for a particular LGA. If the number of persons born overseas who are in the labour force is 100 and the total number of persons born overseas is 160, then the estimated proportion is 100/160 = 0.63. The relative standard errors for both the numerator and denominator will be derived from GRAPH 2 LINE A. Reading from this line, the relative standard error of the numerator (ie the number of persons born overseas who are in the labour force) is approximately 14%, and the relative standard error of the denominator (ie the number of persons born overseas) is approximately 11%. The relative standard error of the estimate of the proportion is therefore

$$\sqrt{14^2 - 11^2} = 8.6\%$$

The standard error on the proportion is 8.6 x 0.63/100 = .05. Therefore, there are nineteen chances in twenty that the labour force participation rate for persons born overseas for the LGA is in the range 0.63 $\pm$ (2x0.05) or 0.53 to 0.73.

As can be seen from the above formula, the relative standard error of a proportion or percentage will always be less than the relative standard error of the numerator. However, whenever a proportion or percentage is small (ie the denominator considerably greater than the numerator), it will be reasonable to approximate the relative standard error of the proportion or percentage by the relative standard error of the numerator.

For proportions or percentages where the denominator is the total number of males, females or persons in a CD or group of CDs, the relative standard error of the denominator is zero because these totals were derived from the preliminary processing phase. In these cases, the relative standard error of the proportion or percentage is given simply by the relative standard error of the numerator.

EXAMPLE. Consider an estimate of the percentage of persons born overseas for a particular CD. If the number of persons born overseas in the CD is 300 and the total number of persons in the CD is 1000, then the estimated percentage is (300/1000)x 100 = 30%. The relative standard error on the denominator is zero since estimates of total persons in a CD are not subject to sampling error. The relative standard error on the numerator can be obtained from GRAPH 2 LINE A. This graph gives the relative standard error on the numerator as approximately 8.1%. Therefore, the relative standard error on the percentage is also 8.1% and hence the standard error on the estimate of percentage is 8.1 x 30/100 = 2.5 percentage points. Therefore, there are nineteen chances in twenty that the percentage of persons born overseas in the CD will lie within the range 30 $\pm$ (2 x 2.5) or 25% to 35%.

Relative standard errors for estimates of proportions or percentages may also be determined from Table 2 which sets out relative standard errors for selected percentages or proportions.

SAMPLING ERRORS ON ESTIMATES OF DIFFERENCES

The relative standard error on differences between two estimates of numbers or between two estimates of proportions (or percentages) can also be derived from the graphs of relative standard errors.

(1) For differences between estimates from the 1976 Census and those from the 1971 Census, the standard error of the difference will be identical to the standard error of the 1976 estimate alone, since 1971 estimates are not subject to sampling error.

EXAMPLE. If estimates for the 1971 and 1976 Census are 500 dwellings and 800 dwellings respectively, then the difference is estimated as 300 dwellings. The 1971 estimate has no relative standard error, whilst the 1976 estimate has a relative standard error (as read from GRAPH 1) of approximately 3% and hence a standard error of 3% of 800 or 24. The standard error of the difference is 24 and there are nineteen chances in twenty that if all schedules from the 1976 Census had been processed, the observed difference would be within the range $300 \pm (2 \times 24)$ or 252 to 348.

(2) For differences between two 1976 Census estimates the standard error of the difference may be approximated by the following formula:

$$\text{Standard Error } (z\text{-}y) = \sqrt{\{\text{Standard Error } (z)\}^2 + \{\text{Standard Error } (y)\}^2}$$

This approximation will be exact for differences between estimates of the same characteristic in two different areas (eg LGAs, CDs), or for differences between separate and uncorrelated characteristics in the same area. If, however, there is positive correlation between the characteristics (eg comparison of number of lawyers with number of persons with law degrees), the above approximation will overestimate the true standard error. If there is a negative correlation between the characteristics (eg comparison of the percentage of persons who travel to work by train and by car), it will underestimate the true standard error.

EXAMPLE. If the estimates for two LGAs of the total number of occupied dwellings are 1000 and 800 and the number of occupied dwellings with outer walls of brick are 250 and 650 respectively, then the percentage of occupied dwellings with brick walls in each of these LGAs is $(250/1000) \times 100 = 25\%$ and $(650/800) \times 100 = 81.2\%$ respectively. The difference between these estimated percentages is therefore 56.2%. The calculation of the standard error of this difference requires the standard error of each of the percentages to be calculated. The relative standard errors on each of the estimates of numbers (1000, 800, 250 and 650) can be derived from GRAPH 1 LINE D. Using the formula previously given for the relative standard error of a percentage then relative standard errors on the estimated percentages are:

$$\sqrt{6.0^2 - 3.0^2} = 5.2\% \text{ and } \sqrt{3.7^2 - 3.3^2} = 1.7\%$$

The standard errors on each of the percentages are

$$5.2 \times 25/100 = 1.3 \text{ and } 1.7 \times 81.2/100 = 1.4$$

Finally, the standard error on the difference is

$$\sqrt{1.3^2 + 1.4^2} = 1.9 \text{ percentage points.}$$

Therefore, there are nineteen chances in twenty that the difference between the percentage of occupied dwellings with brick walls in the different LGAs will be within the range $56.2 \pm (2 \times 1.9)$ or 52.4% to 60.0%.

TABLE 1. RELATIVE STANDARD ERRORS OF DWELLING AND PERSON ESTIMATES

ESTIMATE	2	5	10	15	20	30	40	50	75	100	500	1000
A-LINE	80	53	38	32	28	23	20	18	15	13	6.4	4.7
B-LINE	62	39	27	22	19	15	13	12	9.6	8.3	3.6	2.5
D-LINE	70	44	31	25	22	18	15	14	11	9.6	4.2	3.0

TABLE 2. RELATIVE STANDARD ERRORS ON PERCENTAGES OR PROPORTIONS

TABLE 2A. CLUSTERED PERSON TOPICS (A LINE)

	PERCENTAGE					
	15	20	30	45	60	75
DENOMINATOR						
50	40	34	26	19	14	10
100	29	25	19	14	10	7.3
200	21	18	14	10	7.5	5.3
500	14	12	9.1	6.7	5.0	3.5
750	11	9.8	7.6	5.5	4.1	2.9
1000	10	8.6	6.6	4.9	3.6	2.6

TABLE 2B. UNCLUSTERED PERSON TOPICS (B LINE)

	PERCENTAGE					
	15	20	30	45	60	75
DENOMINATOR						
50	29	24	19	13	9.9	7.0
100	20	17	13	9.4	6.9	4.9
200	14	12	9.1	6.6	4.8	3.4
500	8.9	7.5	5.7	4.1	3.0	2.1
750	7.3	6.1	4.6	3.3	2.5	1.7
1000	6.3	5.2	4.0	2.9	2.1	1.5

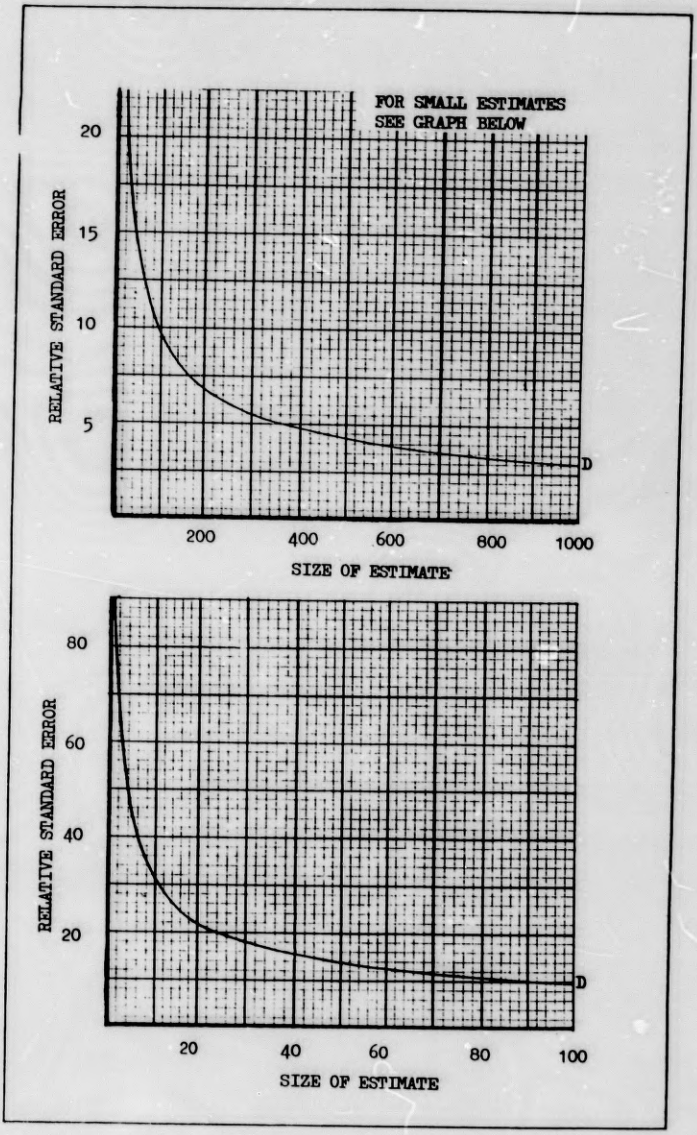
TABLE 2C. NUMERATOR - CLUSTERED PERSON TOPICS (A LINE)
DENOMINATOR - UNCLUSTERED PERSON TOPICS (B LINE)

	PERCENTAGE					
	15	20	30	45	60	75
DENOMINATOR						
50	42	36	29	24	20	17
100	31	27	22	17	15	13
200	23	19	16	13	11	9.5
500	15	13	11	8.5	7.3	6.4
750	12	11	8.8	7.1	6.1	5.3
1000	11	9.5	7.7	6.3	5.3	4.7

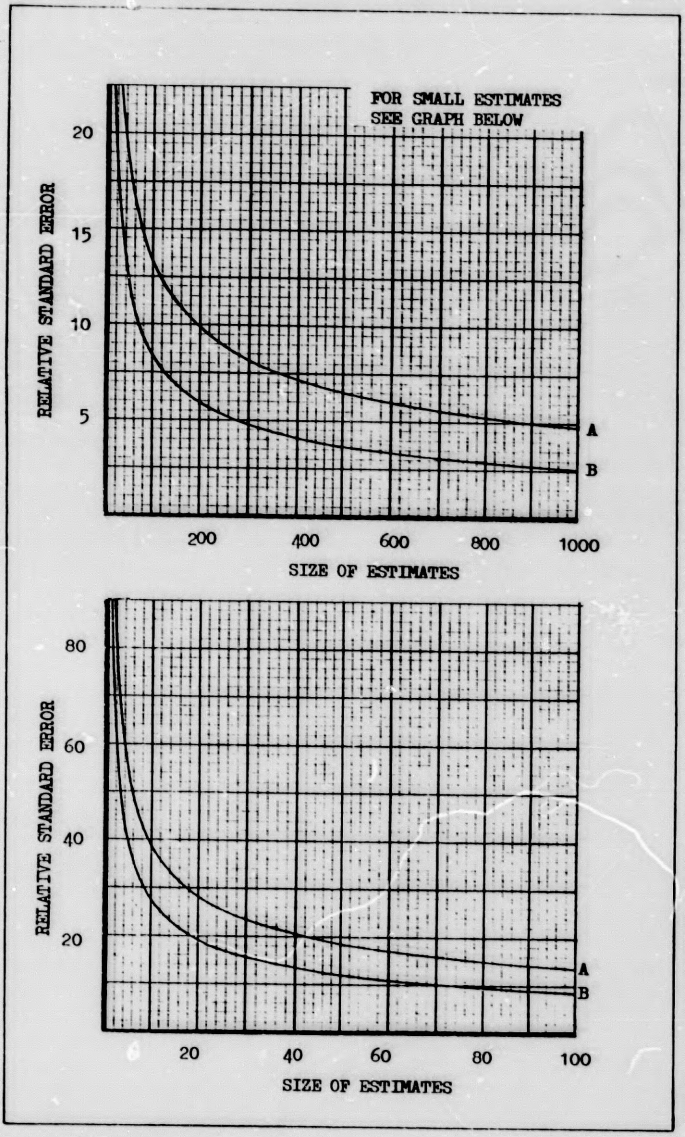
TABLE 2D. DWELLING TOPICS (D LINE)

	PERCENTAGE					
	15	20	30	45	60	75
DENOMINATOR						
50	33	28	21	15	11	7.9
100	23	19	15	11	7.9	5.6
200	16	14	10	7.5	5.5	3.9
500	10	8.6	6.5	4.7	3.5	2.5
750	8.3	7.0	5.3	3.8	2.8	2.0
1000	7.2	6.0	4.6	3.3	2.4	1.7

RELATIVE STANDARD ERRORS FOR DWELLING TOPICS GRAPH 1



RELATIVE STANDARD ERRORS FOR PERSON TOPICS GRAPH 2



APPENDIX D

INQUIRIES

State or Territory	For Personal Inquiries –	Telephone	For Written Inquiries –
NEW SOUTH WALES	Census Liaison Officer Population Census Subsection Level 5 St Andrews House Sydney Square SYDNEY	(02) 236.6111	Deputy Commonwealth Statistician (Attention : Supervisor Demography Section) Box 796, G.P.O. SYDNEY. N.S.W. 2001
VICTORIA	Supervisor, Demography Section Commonwealth Bank Building Cnr Flinders and Elizabeth Streets MELBOURNE.	(03) 652.6211	Deputy Commonwealth Statistician (Attention : Supervisor Demography Section) Box 2796Y, G.P.O. MELBOURNE. VIC. 3001
QUEENSLAND	Supervisor, Social and Demography 345 Ann Street BRISBANE.	(07) 33.5606	Deputy Commonwealth Statistician (Attention : Supervisor, Social and Demography Section) 345 Ann Street BRISBANE. QLD 4000
SOUTH AUSTRALIA	Information Officer Prudential Building 195 North Terrace ADELAIDE.	(08) 228.9439	Deputy Commonwealth Statistician (Attention : Information Officer) Box 2272, G.P.O. ADELAIDE. S.A. 5001
WESTERN AUSTRALIA	Census Liaison Officer 1-3 St George's Terrace PERTH.	(09) 323.5285	Depuxy Commonwealth Statistician (Attention : Census Liaison Officer) 1-3 St George's Terrace PERTH. W.A. 6000
TASMANIA	Census Liaison Officer 188 Collins Street HOBART.	(002) 20.4437	Deputy Commonwealth Statistician (Attention : Census Liaison Officer) Box 66A, G.P.O. HOBART. TAS. 7001
NORTHERN TERRITORY	Census Liaison Officer 7th Floor, M.L.C. Building 81 Smith Street DARWIN.	(089) 80.2612	The Statistician, Northern Territory (Attention : Census Liaison Officer) Box 3796 P.O. DARWIN. N.T. 5794
AUSTRALIAN CAPITAL TERRITORY	Assistant Supervisor User Contact and Dissemination Population Census Branch Wing 4, Level Gb Cameron Offices BELCONNEN.	(062) 52.6006	Australian Statistician (Attention : Population Census Branch) Box 10, P.O. BELCONNEN. A.C.T. 2616