Census advice note

Research and Consultation Service, Bradford Council

There are no 1s and 2s in Census Output – what do I have to know about this?

Issue

To avoid any charge of disclosure of personal information, all small numbers in tables from the 2001 Census have been adjusted. In effect, every 1 and 2 has been rounded randomly to either 0 or 3. The Registrar General for England and Wales took a strong view of the risk of disclosing confidential information; in Scotland small numbers have not been adjusted in this way.

These errors are additional to those from respondents and from imputation of missing data. Normally the errors from rounding are relatively small. This note outlines the consequences of rounding small numbers and what to beware of in cases where the effect is large.

Consequences

1. There is little impact for most tables, particularly tables released by ONS for larger areas, and tables with fewer categories in them: those with few 0s and 3s.
2. Table subtotals and totals will be affected by rounding even if they are not themselves small numbers, because totals are calculated *after* the small counts in 'internal' cells have been rounded.
3. When the same count of say the number of ill people appears in more than one table, it will differ because each table has been rounded independently.
4. Adding census counts for several areas to create a count for a neighbourhood, will be affected to the extent that there are 0s and 3s in the area tables that have been added.

Advice

1. 1. Where a count is available in more than one table, use the table with least number of 0s and 3s. This usually means, where possible:
   - Use Key Statistics tables and Univariate or CAS rather than Standard Tables.
   - Use Key Statistics or Univariate tables for denominators.
   - For data for Wards, use ward tables, do not add up the Output Area values.
   - If a neighbourhood consists of a Ward minus an output Area, calculate in that way rather than add up many Output Areas.
2. Beware of percentages based on small rounded values.
   This includes some of the data from many detailed tables for wards and Output Areas, and the Special Migration and Commuting statistics due in 2004.
3. Calculate the impact on the count you are interested in. Use this rule of thumb:
   Add the number of 0s and 3s among all the internal cells that have been added to create your count, take the square root of that number and multiply by 0.8. The result is the *average* error that might be expected (one can expect the error to be more than this as often as it less than this), and 5% of the time the error will be more than twice that value. We cannot tell the direction of the error, because it was random. See example 3 below.
4. Where appropriate, add a warning against data that are approximate "This value is approximate, due to rounding of small numbers in census output", or don't use them.

Examples

1. Zeros may not be true zeros. The Census tables for Bradford show no health professionals in agriculture (ST039 – occupation by industry). The census database may really have no such records, or there may be 1 or 2 such records.

The same table shows neither health nor protective services technical workers in agriculture. These figures could also be 1 or 2. Adding up these three cells, the table shows no health or associated workers in agriculture in Bradford, but this figure is approximate and there is a small probability it could be as much as six.

2. Inconsistent numbers. The number of people in households with a long-term illness can be found for Baildon ward on tables ST016 (2,528), ST017 (2,551), ST018 (2,540), ST019 (2,538), ST020 (2,534).

3. Output Area indicators. These are often used in indices, and in statistics for neighbourhoods. To compute unemployment for an Output Area among those of working age, table CAS028 has to be used. The example below shows a total of 15 unemployed, based on adding the 16 values for males and females of different age groups, all of which are 0 or 3. The rule of thumb suggests an error of plus or minus 0.8*√(16) or ±3, and 5% of the time an error of ±6 or more. The 128 economically active is based on adding the 48 values for males and females of different age groups in the three categories of activity. 40 of these are 0 or 3. The unemployment rate is 11.7% (15/128) but the value on the census database should be *expected* to be in error by around ±2.5%, ie be in the range 9.2% - 14.1%, and to be outside the range 6.7% - 16.7% 5per cent of the time. In this case the same value can also be calculated from key statistics where there is only one cell of 0 or 3 and so is likely to be very close to the true value on the census database. The unemployment rate from the Key Statistics table KS09a (below) shows 10/120 = 8.3%.

More information

http://www.statistics.gov.uk/census2001/discloseprotect.asp ONS' short guide to their disclosure protection measures.

http://www.statistics.gov.uk/census2001/op10.asp ONS' evaluation reports give finer details of various aspects of census quality.

http://www.statistics.gov.uk/census2001/cn_22.asp ONS background to adjusting small cells, with link to detailed proposal document.

http://www.ccsr.ac.uk/staff/Ludi/census..htm Discussion of impact of rounding small numbers.

Table CAS028 SEX AND AGE BY ECONOMIC ACTIVITY
Table population : All people aged 16 to 74
Geographical level : Output Area 00CXGC0010

	ALL PEOPLE	Economically Active : Total	Economically Active : Employee	Economically Active : Self-employed	Economically Active: Unemployed	Economically Inactive : Total	Economically Active: Full- time Student	Economically Inactive: Retired	Economically Inactive: Student	Economically Inactive: Other economically inactive
ALL PEOP	198	128	98	12	15	70	3	12	16	42
16 to 17	7	0	0	0	0	7	0	0	7	0
18 to 19	16	10	7	0	0	6	3	0	3	3
20 to 24	13	10	4	0	6	3	0	0	0	3
25 to 34	43	29	26	3	0	14	0	0	3	11
35 to 54	86	70	55	9	6	16	0	0	3	13
55 to 59	6	3	3	0	0	3	0	0	0	3
60 to 64	14	6	3	0	3	8	0	5	0	3
65 to 74	13	0	0	0	0	13	0	7	0	6
Males	93	70	50	8	9	23	3	4	9	10
16 to 17	3	0	0	0	0	3	0	0	3	0
18 to 19	13	7	4	0	0	6	3	0	3	3
20 to 24	3	3	0	0	3	0	0	0	0	0
25 to 34	19	19	16	3	0	0	0	0	0	0
35 to 54	42	35	27	5	3	7	0	0	3	4
55 to 59	3	3	3	0	0	0	0	0	0	0
60 to 64	6	3	0	0	3	3	0	0	0	3
65 to 74	4	0	0	0	0	4	0	4	0	0
Females	105	58	48	4	6	47	0	8	7	32
16 to 17	4	0	0	0	0	4	0	0	4	0
18 to 19	3	3	3	0	0	0	0	0	0	0
20 to 24	10	7	4	0	3	3	0	0	0	3
25 to 34	24	10	10	0	0	14	0	0	3	11
35 to 54	44	35	28	4	3	9	0	0	0	9
55 to 59	3	0	0	0	0	3	0	0	0	3
60 to 64	8	3	3	0	0	5	0	5	0	0
65 to 74	9	0	0	0	0	9	0	3	0	6

These Census tables are for 00CXGC0010, an Output Area near Peel Park in Undercliffe

Table KS09a Economic activity - all people, aged 16-74

	All people aged 16 - 74	Economically active: Employees Part-time*	Economically active: Employees Full-time*	Economically active: Self - employed	Economically active: Unemployed	Economically active: Full-time student	Economically inactive: Retired	Economically inactive: Student
00CXGC0010	187	24	75	11	10	0	11	13

Ludi Simpson, September 2003