

# Using the information in outputs from the Scottish Household Survey

## How data is displayed in tables

Tables are generally presented in the format 'dependent variable by independent variable' where the independent variable is being used to examine or explain variation in the dependent variable. Thus, a table titled 'housing tenure by household type' shows how housing tenures vary among different household types. Tables generally take three forms; column percentages (the dependent variable is in the rows), row percentages (the dependent variable is in the columns) and cell percentages which may show agreement or selection of a statement with one or a number of statements.

All tables have a descriptive and numerical base showing the population or population sub-group examined in it. While all results have been calculated using unrounded weighted data, the bases shown provide the unweighted counts. These have been rounded to the nearest 10 to comply with statistical disclosure control principles and the Code of Practice for Official Statistics. Where base numbers are less than five, the bases shown have been rounded down to zero. It is therefore not possible to calculate how many respondents gave a certain answer based on the results and bases presented.

## Reporting conventions

In general, percentages in tables are rounded to the nearest whole number. Zero values are shown as a dash (-), values greater than 0 per cent but less than 0.5 per cent are shown as 0 per cent and values of 0.5 per cent but less than 1 per cent are rounded up to 1 per cent. Columns or rows may not add to exactly 100 per cent because of rounding, where 'don't know/refused' answers are not shown<sup>1</sup> or where multiple responses to a question are possible.

In some tables, percentages are removed and replaced with '\*'. This is where the base on which percentages would be calculated is less than 50 and this data is judged to be insufficiently robust for publication.

---

<sup>1</sup> Missing responses are not included within any analysis. Similarly 'don't know/refused' options are not shown as a separate category in some tables.

## Variations in base size for totals

As the questionnaire is administered using computer assisted personal interviewing (CAPI), item non-response is kept to a minimum. Bases do fluctuate slightly due to small amounts of missing information (where, for example, the age or gender of household members has been refused and where derived variables such as household type use this information).

Some questions are asked of a reduced sample and the bases are correspondingly lower. From January 2012, the redesigned survey asked questions typically of full or one-third sample allocation. This concept of streaming was first introduced to the SHS in 2007, when some questions were streamed or changed in the course of the year and again the base size is lower. Further changes to streaming have been made in subsequent years.

## Statistical significance

All proportions produced in a survey have a degree of error associated with them because they are generated from a sample survey of the population rather than a survey of the entire population (e.g. Census). Any proportion measured in the survey has an associated confidence interval (within which the 'true' proportion of the whole population is likely to lie), usually expressed as  $\pm x$  per cent. As a general rule of thumb, the larger the sample size for a given question, the smaller the confidence interval around that result will be thus making it easier to detect real change year-on-year and differences between sub-groups.

It is possible with any survey that the sample achieved produces estimates that are outside this range. If the survey were to be run multiple times on the same population in the same year (i.e. under repeated sampling), the number of times out of a 100 surveys that the result achieved would be expected to lie within the confidence interval is also quoted; conventionally the level set is 95 out of 100, or 95 per cent. Technically, all results should be quoted in this way however, it is less cumbersome to simply report the percentage as a single percentage, the convention adopted in outputs from the Scottish Household Survey.

Where sample sizes are small or comparisons are made between sub-groups of the sample, the sampling error needs to be taken into account. There are formulae to calculate whether differences are statistically significant (i.e. they are unlikely to have occurred by chance), and our guidance on confidence intervals and statistical significance (see the supporting files) provides a simple way to calculate whether differences are significant. Our guidance also provides further explanation on statistical significance and on how confidence intervals can be interpreted.

The [SHS Data Explorer](#) incorporates a tool which highlights cells that are significantly different from the comparator figure. Comparisons can be made between local authorities, a local authority can be compared with national level data and data can be compared for different years.