

Statistical standard for age

Rationale

Age is a key variable for determining the demographic characteristics of a population.

Definition

Age is the length of time a person has been alive measured in complete, elapsed years. It is measured as the difference between 'date of birth' and 'data collection date'.

Operational issues

The reliability of age data may be affected by the type of question asked. Asking for day, month and year of birth has been shown to yield data that are more precise because the question is so specific it discourages answers in approximate terms.

Responses to the question on age at last birthday are likely to be less accurate than responses to the question on date of birth. Independent of their method of reckoning age, respondents may be unsure whether the age wanted is that at last, next or nearest birthday. Other causes of error in responses to a direct question on age are:

- a tendency to report ages ending in certain preferred digits, such as zero and five, hence age clustering at five and ten year intervals;
- exaggeration, especially in older age groups;
- understatement due to a fear of aging or reaching a new decade;
- ignorance of current age;
- carelessness in reporting.

Explanatory notes

Date of birth is often used as an input variable for age, rather than collecting age information directly. However, date of birth is not a recognised output variable. For this reason, both variables are discussed within the one standard.

Classification criteria

The criterion used to place a person into the classification is the elapsed time between birth and a data collection, calculated in whole years.

Classification

Age classification

Age is a flat classification with single-year categories from 0 years to 119 years inclusive, plus a category for 120 years and over.

No provision is made for residual categories as, in line with international practice, it is Statistics New Zealand policy to impute missing values for age data. The reasoning for this is due to the practicalities of data presentation. Age is a fundamental variable used in a variety of published cross-tabulated outputs. Due to space limitations in hardcopy publications, and to simplify the presentation of data for users (eliminating an over abundance of residual categories in tabulated outputs), residual categories are removed by imputing any missing age data.

Date of birth classifications

Classifications for date of birth are required because age is derived from the information collected by such a question. Date of birth is derived from three flat classifications; date of birth day, date of birth month and date of birth year. Day has single categories 01 to 31 (representing the days of the month). Month has single categories 01 to 12 (representing the months of the year). Year has single categories (representing 120 years back from the current year). For example the 2007 code structure for the year classification ranges from 1887 to 2007. All three classifications also have residual categories.

The residual categories are defined in [Glossary](#).

Coding process

Age should be derived from date of birth, see the standard derivation routine below.

If values are missing, age is to be imputed.

Derivation of age

1. Where day, month and year of birth are all specified, the respondent's true age is calculated as a whole year with respect to the reference day.
2. Where only month and year of birth are specified, and the month specified is not the reference month, age can again be calculated as a whole year with respect to the reference day.
3. Where only month and year of birth are specified, and the month equals the reference month, compute $X = (\text{reference day}) / (\text{number of days in reference month})$. Generate a random number between 0 and 1. If the random number is less than X , day of birth is assumed to fall before reference day. If the random number is greater than or equal to X , the day of birth is assumed to fall on or after reference day.
4. Where only year of birth is specified, a process similar to step 3 is followed, with X being computed as $(\text{reference day, specified as number of days since 1 January in reference year}) / (\text{total number of days in reference year})$.
5. Where only day and year of birth are specified, a random number between 1 and 12 inclusive is generated for the month. Age can then be calculated as a whole year, with respect to the reference date.
6. Where only day and month of birth are specified or none of the three fields are given or the values specified are later than the reference date, then age is imputed.

Age

Classification	Age – New Zealand Standard Classification
Abbreviation	AGE

Version	V1.0
Effective date	26/01/1995

Date of birth – day

Classification	Date of Birth – Day Classification
Abbreviation	BIRTHDAY
Version	V1.0
Effective date	30/03/1999

Date of birth – month

Classification	Date of Birth – Month Classification
Abbreviation	BIRTHMON
Version	V2.0
Effective date	30/03/1999

Date of birth – year

Classification	Date of Birth – Year Classification
Abbreviation	BIRTHYR
Version	V12.0
Effective date	01/01/2007

Questionnaire module

Requirements

Organisations such as the United Nations recommend that information on age may be secured either by obtaining the date of birth (day, month, year) or by asking directly for a persons age at their last birthday. The first method yields more precise information and, in the interests of data comparability and integration, is the recommended standard which should be used whenever possible.

Although a direct question on age is likely to yield less accurate results, there may be certain circumstances where it will be necessary or appropriate to ask a direct question on age. Asking age information by proxy is a good example where the person supplying the information is more likely to know the respondent's age than their date of birth.

The recommended questionnaire module must obtain the date of birth: Day–Month–Year.

Example

The questionnaire module below is an example that meets the requirements documented in this statistical standard. Other questionnaire modules may vary in format but should conform to the requirements contained in this statistical standard.

Your age on your last birthday?

7 When were you born?
 day month year you were
 (e.g. 29) (e.g. 11) born (e.g. 1963)

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

Standard output

Age

Age should always be stored as single years to provide maximum flexibility at the output stage.

Single years

This is a complete classification for age by single years where the output categories are the same as the classification categories.

- Less than 1 Year
- 1 Year
- 2 Years
- :
- 119 Years
- 120 Years and Over

Five yearly groupings

This groups age into five year groups, except for the final category which aggregates all persons 100 years and over.

0–4 Years
5–9 Years
10–14 Years
15–19 Years
20–24 Years
25–29 Years
30–34 Years
35–39 Years
40–44 Years
45–49 Years
50–54 Years
55–59 Years
60–64 Years
65–69 Years
70–74 Years
75–79 Years
80–84 Years
85–89 Years
90–94 Years
95–99 Years
100 Years and Over

Ten yearly groupings

This third standard output set groups single year of age into ten yearly age ranges, but aggregates person 90 years and over into one category.

Less than 10 Years
10–19 Years
20–29 Years
30–39 Years
40–49 Years
50–59 Years
60–69 Years
70–79 Years
80–89 Years
90 Years and Over

Persons 15 years and over

This standard output set provides a compact set of ranges for use in cross-tabulations. This grouping has been used in the 1991 and 1996 Census of Population and Dwellings National Summaries.

15–17 Years
18–19 Years
20–24 Years
25–29 Years
30–34 Years
35–39 Years
40–44 Years
45–49 Years
50–54 Years
55–59 Years
60–64 Years
65 Years and Over

Person of all ages

This standard output set provides a compact set of ranges for use in cross-tabulations.. This grouping has been used in the 1991 and 1996 Census of Population and Dwellings National Summaries.

Less than 5 Years
5–14 Years
15–19 Years
20–29 Years
30–39 Years
40–49 Years
50–59 Years
60–64 Years
65 Years and Over

Date of Birth

The standard output variable categories for date of birth classifications day and year are the same as the classification. For the date of birth month classification the 01 to 12 categories that represent each month are output as their respective months January to December.

The residual categories may be output separately or combined. Where a combination item of residuals is to be used in output, this item should be labelled 'not elsewhere included' and should have a footnote indicating its composition.

The residual categories are defined in [Glossary](#).

Related classifications and standards

New Zealand

Age, along with sex is one of the most important social variables collected by Statistics New Zealand. They are more frequently cross-tabulated with other characteristics of the population than any other variables. Age is also used as a fundamental input in the creation of a large range of derived output variables. For example:

- family type
- household composition
- child dependency status
- age of youngest child
- age of reference person
- age of parent
- number of persons aged 0-2 years
- Maori electoral population.

The standard is currently used by other parts of the Official Statistics System in New Zealand.

International

The classification and standard is not based on an international standard or the standard of another country, but follows the recommendations of the United Nations.

Glossary

Residual categories

Don't know

Use of this category is discretionary. The use of a category capturing don't know responses is most applicable to household surveys where don't know may be a legitimate response to certain questions.

Refused to answer

This category is only used when it is known that the respondent has purposefully chosen not to respond to the question. Use of this residual category in processing is optional. Its use is most applicable in face-to-face or telephone interviews, but may be used in self-completed questionnaires if the respondent has clearly indicated they refuse or object to answering the question.

Response unidentifiable

This category is used when there is a response given, but:

1. the response is illegible, or
2. it is unclear what the meaning or intent of the response is – this most commonly occurs when the response being classified contains insufficient detail, is ambiguous or is vague, or
3. the response is contradictory eg, both the yes and no tick boxes have been ticked, or
4. the response is clear and seemingly within the scope of the classification but can not be coded because no suitable option (particularly other residual category options such as 'not elsewhere classified' or 'not further defined') exists in the classification or codefile.

Response Outside Scope

This category is used for responses that are positively identified (ie the meaning and the intent are clear) but which clearly fall outside the scope of the classification/topic as defined in the standard.

Not stated

This category is only used where a respondent has not given any response to the question asked, ie it is solely for non-response.