

<http://www.stats.govt.nz/tools_and_services/university-students/2006-census-birth-cohort.aspx>

**2006 Census birth cohort SURF**

**Data summary**

* 10,000 observations (synthesized newborn children)
* 29 variables – 20 categorical and nine numeric variables

*NOTE: This is a large data set suitable for exploratory use with Achievement Standards  
91035 (1.10), 91036 (1.11), 91264 (2.9), 91581 (3.9) and 91582 (3.10)*

**OVERVIEW**

This data set contains 10,000 records on the newborn child, their household and their mother and father (if applicable).

More details on the background of the data set can be found at the [end](#aboutbirth) of this document. Variable details are given on the following pages.

**VARIABLES**

The 2006 Census birth cohort SURF contains the following variables:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Variable type** | **Code** | **Description** |
| **random\_ID** | ID number |  |  |
| **Ch\_gender** | Categorical | male female | Gender of newborn child |
| **Deprivation** | Categorical | Scale from 1 to 10 | NZDep2006 Deprivation index decile from 1 (least deprived) to 10 (most deprived)\*\* |
| **M\_age** | Numerical | n/a | Mother’s age in years |
| **M\_work** | Categorical | full\_time  part\_time  unemployed  not in LF | Mother employed full time  Mother employed part time  Mother unemployed  Mother not in the labour force |
| **F\_age** | Numerical |  | Father’s age in years |
| **F\_work** | Categorical | full\_time  part\_time  unemployed  not in LF | Father employed full time  Father employed part time  Father unemployed  Father not in the labour force |
| **No\_bedrooms** | Numerical |  | Number of bedrooms in the child’s main dwelling |
| **M\_years\_current\_addr** | Numerical |  | Number of years mother has lived at the current address |
| **M\_owns\_residence** | Categorical | owns  not\_owned | Mother owns usual residence  Mother does not own usual residence |
| **M\_education** | Categorical | none  level\_1  level\_2\_3  level\_4  diploma  degree  post\_grad | Mother’s highest qualification:  none  Level 1 certificate  Level 2 or 3 certificate  Level 4 certificate  Level 5 or 6 diploma  Bachelor degree  Post graduate degree including honours, masters, doctorate degrees |
| **M\_work\_hours** | Numerical |  | Mother’s total hours in work per week |
| **M\_smoke** | Categorical | smoker  ex\_smoker  never | Mother is:  a smoker  an ex smoker  has never smoked |
| **F\_years\_current\_addr** | Numerical |  | Number of years father has lived at the current address |
| **F\_owns\_residence** | Categorical | owns  not\_owned | Father owns usual residence  Father does not own usual residence |
| **F\_education** | Categorical | none  level\_1  level\_2\_3  level\_4  diploma  degree  post\_grad | Father’s highest qualification:  none  Level 1 certificate  Level 2 or 3 certificate  Level 4 certificate  Level 5 or 6 diploma  Bachelor degree  Post graduate degree including honours, masters, doctorate degrees |
| **F\_work\_hours** | Numerical |  | Father’s total hours in work per week |
| **F\_smoke** | Categorical | smoker  ex\_smoker  never | Father is:  a smoker  an ex smoker  has never smoked |
| **Cellphone** | Categorical | cellphone  none | Access to a cellphone  No access to a cellphone |
| **Landline** | Categorical | landline  none | Access to a landline telephone  No access to a landline telephone |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable name** | **Variable type** | **Code** | **Description** | | |
| **Internet** | Categorical | internet  none | Access to the internet  No access to the internet | | |
| **Hhld\_income\_cat** | Categorical |  | | Total annual gross income for the household | |
| Loss | |
| Zero Income | |  | |
| $1 - $5,000 | |  | |
| $5,001 - $10,000 | |  | |
| $10,001 - $15,000 | |  | |
| $15,001 - $20,000 | |  | |
| $20,001 - $25,000 | |  | |
| $25,001 - $30,000 | |  | |
| $30,001 - $35,000 | |  | |
| $35,001 - $40,000 | |  | |
| $40,001 - $50,000 | |  | |
| $50,001 - $70,000 | |  | |
| $70,001 - $100,000 | |  | |
| $100,001 or more | |  | |
| **Hhld\_income\_value** | Numerical |  | Total annual gross income for the household (taken from middle of income categories above) | | |
| *Income category* | | *Income value* |
| 1 Loss | | 🡪 NA |
| 2 Zero Income | | 🡪 $0 |
| 3 $1 - $5,000 | | 🡪 $2,500 |
| 4 $5,001 - $10,000 | | 🡪 $7,500 |
| 5 $10,001 - $15,000 | | 🡪 $12,500 |
| 6 $15,001 - $20,000 | | 🡪 $17,500 |
| 7 $20,001 - $25,000 | | 🡪 $22,500 |
| 8 $25,001 - $30,000 | | 🡪 $27,500 |
| 9 $30,001 - $35,000 | | 🡪 $32,500 |
| 10 $35,001 - $40,000 | | 🡪 $37,500 |
| 11 $40,001 - $50,000 | | 🡪 $45,000 |
| 12 $50,001 - $70,000 | | 🡪 $60,000 |
| 12 $70,001 - $100,000 | | 🡪 $85,000 |
| 14 $100,001 or More | | 🡪 $100,000 |
| **No\_usual\_resdnt** | Numerical |  | Usual number of residents in the household | | |
| **Parents** | Categorical | two  mother  father | Two parent family  Single mother family  Single father family | | |
| **Ch\_Ethnicity** | Categorical | Maori  Pacific  Asian  MELAA  Other  NZ European | Ethnicity of newborn child\*\* | | |
| **M\_Ethnicity** | Categorical | Maori  Pacific  Asian  MELAA  Other  NZ European | Ethnicity of mother\*\* | | |
| **F\_Ethnicity** | Categorical | Maori  Pacific  Asian  MELAA  Other  NZ European | Ethnicity of father\*\* | | |
| **M\_benefit** | Categorical | benefit  none | Mother receives government benefit  Mother does not receive government benefit | | |
| **F\_benefit** | Categorical | benefit  none | Father receives government benefit  Father does not receive government benefit | | |

**Interpreting the variables**

**Deprivation**

<http://www.health.govt.nz/publication/nzdep2006-index-deprivation>

The **Deprivation** variable is based on the NZDep2006 [index of socioeconomic deprivation calculated by researchers](http://www.otago.ac.nz/wellington/research/hirp/otago020194.html) from the Department of Public Health, University of Otago, Wellington (UOW). NZDep2006 combines nine variables from the 2006 Census which reflect eight dimensions of deprivation. In this SURF, NZDep2006 is an ordinal score between 1 and 10 where 1 represents areas with the least deprived scores and 10 the areas with the most deprived scores. NZDep2006 is a decile rating (divides New Zealand into tenths of the distribution), and applies to geographical areas rather than individual people.

**Ethnicity**

A newborn child’s ethnicity response is decided by whoever completes their census form. Please note that ethnicity is different from nationality or race.

People can identify with multiple ethnicities in the census and up to six are recorded in the original SURF. These have been collapsed to one **Ethnicity** variable using the Ministry of Education priority reporting for student ethnicity system (<http://www.educationcounts.govt.nz/data-services/collecting-information/code_sets/ethnic_group_codes>). The MoE priority order of ethnicities is

1. Maori
2. Pacific
3. Asian
4. MELAA (Middle Eastern, Latin American, African)
5. Other
6. NZ European

This method of reducing ethnicity is unlikely to result in a variable that represents the ethnic diversity and structure of the New Zealand population.

Please note that the prioritising of ethnicity is not preferred practice by Statistics New Zealand, but in this case helps simplify the double-counting problem for the data set to be suitable for teaching. For further details see: <http://www.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/~/media/Statistics/browse-categories/population/census-counts/review-measurement-ethnicity/understanding-working-ethnicity-data.pdf>

**Income, benefits, and labour force status**

The labour force status variables (**M\_work** and **F\_work**) are based on official classifications, and may not align with rules for benefits or typical expectations. The census counts someone of working age (15 years and older) as employed if they worked for at least one hour in the reference week (ending 5 March 2006). This work could have been for pay or profit, or without pay if for a family-owned farm or business.

See the [census definitions](http://www.stats.govt.nz/Census/about-2006-census/2006-census-definitions-questionnaires.aspx) for more information. This is the same official definition used in the [Household Labour Force Survey (HLFS)](http://www.stats.govt.nz/browse_for_stats/income-and-work/employment_and_unemployment/HouseholdLabourForceSurvey_HOTPJun13qtr/Definitions.aspx). For example: an unemployment beneficiary could be working a few hours a week, and therefore classified as employed by the census.

See [Key differences between the officially unemployed, registered job seekers and recipients of Unemployment Benefits](http://www.msd.govt.nz/about-msd-and-our-work/publications-resources/statistics/statistical-report/statistical-report-2008/multiple-sections/unemployed-definition.html#differences) for a comparison of our official definitions and criteria used by the Ministry of Social Development.

**Parents**

The dataset contains information about the parents of each newborn. If there is only one parent, information on the first or second parent will be missing in the dataset. The information on a parent may not be the biological parent, but a grandparent, foster parent, or other person. This information can be found in the original SURF dataset if needed.

**BACKGROUND ON DATA SET**

*Synthetic unit-record files (SURF)*

This data set …

* contains 10,000 records on the newborn child, their household and their mother and father (if applicable).
* is based on data for 10,000 randomly selected newborns (aged 0 years) and their parents from the 2006 Census of Population and Dwellings
* represents approximately 19% of all families with newborns in the 2006 Census
* was created by Centre of Methods and Policy Application in the Social Sciences (COMPASS) at the University of Auckland with the support of a grant from the Ministry of Business, Innovation and Employment.
* NOTE: This synthetic data should NOT be used as a source of accurate statistical information, but is largely realistic. Key statistical measures such as mean and variance are similar to those in the real data. Relationships between the variables and the distributions of each variable also imitate the real data.
* Some variables from original 2006 Census birth cohort SURF has been re-coded for use in iNZight – see below for details.
* Details of the variables included in this iNZight version of the 2006 Census birth cohort SURF are on the following pages.