**Simulating a sample of households based on aggregated data published by a census**

Author: Edward McNeil (August 2021)

The simulation is based on a modified version of algorithm 3 by Gargiulo et al (2010).

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008828>

Preliminary notes:

1. The program assigns individuals to households according to probabilities proportional to size based on aggregated census data stratified by 6 household sizes (1 – 6), 8 household types2 and 9 age groups (0-9, 10-19, …, 80+). The number of children (age <15 years) and elders (age 65+ years) stratified by household size and type are also used in calculating the composition of each household.
2. The program allows for only one household head per household, although more than one is possible.
3. The sex of each household member is not considered although the Hong Kong census does provide this information.

The program has six steps.

Step 1. Household size

A sample of *n* households of size 1 to 6 is selected from the population of households with probability proportional to size. If the discrepancy between the sample and the population is more than a desired threshold, then the whole sample is abandoned and a new sample is selected. This is repeated until the discrepancy is less than the threshold. The discrepancy between the sample and the population was calculated by taking the sum of squared differences in the proportions of households of each size.

Based on the accepted sample, steps 2 – 6 below attempt to build households by assigning the ages of each household member iteratively. If at each iteration, the discrepancy in the distribution of ages between the sample and the population is more than a desired threshold, then the current household is abandoned and a new one built.

Step 2. Age of household head

The age group of the household head is selected with probability proportional to size in the population.

Step 3. Household type

If the size of the household is 1 (type = 7), then the type2 of household is perfectly determined and building the current household is complete. Otherwise, the type of household is selected according to the age of the household head with probability proportional to size. If the age of the household head is 80+ years and the household contains parents (types 4 and 5), then the household type is resampled repeatedly until it does not contain parents.

Step 4. Age of partner

If the household contains a couple (household types 1, 2, 4 or 5)2, then the age of the partner is considered to be in the same or adjacent 10-year age group as the household head with probabilities equal to 0.75 and 0.25, respectively but cannot exceed the two extreme adult age groups (20-29 and 80+ years).

Step 5. Age(s) of household head’s child(ren) and parent(s)

If the household contains unmarried children (household types 2, 3 or 5)2, then the number of children is determined from the household size and type. For example, if the household contains 4 members and contains a couple and at least 1 child and at least 1 parent (household type 5), then the number of parents must be 1 and the number of children must also be 1. Otherwise, the number of parents is determined to be either 1 or 2 with equal probability\* and the number of children then determined with probability proportional to size. The ages of children within the same household are determined from the age of the household head for single parent households and the age of the partner for households with couples. The age of each child follows a normal distribution with mean equal to the age of the parent less 30 years1 and standard deviation equal to 10 years but not more than their age less 10 years and not less than their age less 50 years. In other words, each child was deemed to have been born when one of their parents was aged between 10 and 50 years. The age(s) of the household head’s parent(s), if any, also follows a normal distribution but with mean equal to the age of the household head plus 20 years3 and standard deviation 10 years but not more than their age plus 50 years and not less than their age plus 10 years.

Step 6. The ages of the remaining household members, if any, are determined based on the household type and size with probabilities proportional to the size in the population based on the most recent census.

***Footnotes***

|  |
| --- |
| 1 Based on the most recent census which states that the median age of women at first childbirth is 31.8 years. |
| 2 Description of household types: |
| 1. A household comprising a married couple without any other related persons. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household comprising a couple and their unmarried child(ren) without any other related persons. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household comprising a father or mother and his/ her unmarried child(ren) without any other related persons. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household comprising a couple and at least one of their parents (including the parent(s) of the wife and/ or husband) without any other related persons. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household comprising a couple, at least one of their parents (including the parent(s) of the wife and/ or husband) and their unmarried children without any other related persons. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household comprising a group of related persons but not being classified in the above categories. It may or may not include other unrelated persons (e.g. domestic helpers). |
| 1. A household with only one person. |
| 1. A household comprising unrelated person(s). |
| 3 Based on trends in the median age of women at first childbirth. |
| \* This can be modified according to the age of the household head if required. |
|  |
| Source: |
| 2016 Population By-census Office |
| Census and Statistics Department |
| The Government of the Hong Kong Special Administrative Region |
| Enquiry telephone: 3547 1800 |
| Email: bycensus2016@censtatd.gov.hk |
|  |
| Release date: 4 July 2017 |

**One paragraph summary**

The program assigns individuals to households according to probabilities proportional to size based on aggregated census data stratified by 6 household sizes (1 – 6), 8 household types2 and 9 age groups (0-9, 10-19, …, 80+). The number of children (age <15 years) and elders (age 65+ years) stratified by household size and type are also used in calculating the composition of each household. Currently, the program allows for only one household head per household. The sex of each household member is also not considered. Briefly, a sample of *n* households of size 1 to 6 is selected from the population of households with probability proportional to size and accepted if the discrepancy with the population is less than a desired threshold. The discrepancy between the sample and the population was calculated by taking the sum of squared differences in the proportions of households of each size. An attempt to build each household is then done iteratively and in a stepwise fashion by assigning the ages of each individual member according to the age of the household head (selected based on the size of the household) and other household information. For example, the type of household is selected with probability proportional to size according to the age of the household head and the size of the household. The ages of other household members (if any), such as the household head’s partner, children, parents, and other non-related individuals, are selected based on information provided by the most recent census. If, at each iteration, the discrepancy in the distribution of ages between the sample and the population is more than a desired threshold, then the current household is abandoned and a new one built.

No

No

No

Notes: *s* = household size (1-6), *t* = household type, ah = age of household head, ap = age of household head’s partner,   
ac = age of child, ao = age of other household members.

P(s), P(ah | s), P(t | ah, s), P(ap | ah), P(child | s, t), P(ac | ah), and P(ao | t, s) come from the census.   
P(|ap = ah| < 10) = 0.75 and P(10 < |ap ˗ ah| < 20) = 0.25.

The ages of the couple’s children are normally distributed with mean ah - 30 and standard deviation 10.   
The ages of the couple’s parents are normally distributed with mean ah + 20 and standard deviation 10.