Population Model Version 3 – Documentation

The aim of this document is to inform of the overall structuring of the model, the finer details and reasoning of the implementation where such may be otherwise illusive. It is expected that this document will be read alongside the projects Javadoc and in the case of the finer details the inline comments and source code.

1 Model Overview

The key focus of this model was to focus on a structure that would be able to take input in the form of summative information about the desired composition of the generated population and then to be able to make a statistical assertion regarding the similarity of the desired and the generated population. The need for a robust statistical comparative approach was deemed necessary due to the complexity of modelling a population across many variables and summative statistics, as was evidenced in the earlier versions of our population models.

Detailed below is what is involved and expected in each stage of the model.

1.1 Data Input

Associated classes: ***DesiredPopulationConpositionFactory, DesiredPopulationComposition***

Inputted data represents the rate at which modelled events occur at in the desired population. The data can be provided for every year in the population. If not, then the approach by which data will be imputed is detailed in section [TO IMPLEMENT].

The events modelled are:

- Birth – specifically the proportion of females of a given age in a given year that give birth.
- Marriage – specifically the proportion of males of a given age marrying females of a given age in a given year.
- Death – specifically the proportion of people (divided by gender) of a given age that will die within one year from the given year.

In the case of birth and death the data form used is often described as a Lifetable or as the Kaplan-Meier method in the domain of demography and actuarial sciences.

The input and construction of this data into a collection is handled by the DesiredPopulationCompositionFactory and results in a DesiredPopulationComposition object.

1.2 Desired Population Composition

Associated Classes: ***DesiredPopulationComposition, EventRateTables, StatisticalTables, PopulationInformationCollection***

Once data has been placed into a DesiredPopulationComposition, the model now has an understanding of what characteristics the user wishes for the end population to exhibit. Access to the information identifying the rates and proportions at which events are desired to occur in the population is provided through the interface EventRateTables. The rate data that is returned is expressed in the form of Tables, details of which are explained in section 1.6 and for the association of Table formats with data types the JavaDoc is the best and persistently updated source.

The information regarding the desired information is also needed by our statistical approaches and often in specific formats, the StatisticalTables interface makes provision for this and is implemented by the GeneratedPopulationComposition class as well.

1.3 Generating the Simulated Population

Associated Classes: ***Population, DesiredPopulationComposition, EventRateTables, IPopulation, IPerson, IPartnership***

In the process of generating the simulated population calls are made to the DesiredPopulationComposition to access data about the rates and proportions that modelled events should occur to the generated population.

The simulation approach is to be able to define cohorts within each year of the simulation and then to apply to each cohort the number of events as specified by the DesiredPopulationComposition.

1.4 Deriving Generated Population Composition

Associated classes: ***Population, GeneratedPopulationCompositionFactory, StatisticalTables, GeneratedPopulationComposition***

Once a population has been generated we need to place it into a form that allows for comparative analysis. This is done by passing the Population to the GeneratedPopulationCompositionFactory that produces a GeneratedPopulationComposition object. The GeneratedPopulationComposition also implements the StatisticalTables interface like the DesiredPopulationComposition class and thus allows for statistical comparisons of the two to be made, this is outlined in section 1.5.

The process of creating the GeneratedPopulationComposition involves processing the generated population so as to create the summative data of the population as required by the StatisticalTables interface.

1.5 Statistical Comparison of Desired and Generated Populations

Associated classes: ***ComparativeAnalysis, KaplanMeierAnalysis, StatisticalTables, DesiredPopulationComposition, GeneratedPopulationComposition***

As detailed in the overview, the aim for the model is that the generated population can be statistically verified as being significantly similar to the desired population.

The ComparativeAnalysis class coordinates the retrieval of the required data from the different population composition classes and then makes use of the assisting statistical classes to compare the collected data.

For strict event based data that can be expressed in survivor table is compared using the Kaplan Meier method (Tierney et al., 2007; Kleinbaum & Klein, 2006). The method compares the two survival curves and can be used to calculate a Hazard Ratio which compares the level

of risk in the desired and generated populations which can be used to indicate the similarity of the two populations.

1.6 Tables – Population Data Representation

The handling of data in the model is a regular occurrence, therefore a table approach is provided. The tables can have either one or two look up variables. For example, in terms of death rates for men there is one look up variable, this being age. Whereas for marriage there are two look up variables, these been the age of the male and the age of the female. Example of the structuring of these tables and example tables for each event can be seen in the JavaDoc in the StatisticalTables interface.

2 Implementation Details

3 References

Kleinbaum, D. G., & Klein, M. (2006). *Survival analysis: a self-learning text*. Springer Science & Business Media.

Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S., & Sydes, M. R. (2007). Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*, *8*(1), 16.