Stats approach

Tom Dalton

08/09/2017

This documents details the approach taken to verify the genalogical populations that we create.

For efficency we have produced five contingency tables which are each concerned with one of the input distrbtions efffecting genalogical structure.

Lets load these in:

```
path = paste("/Users/tsd4/OneDrive/cs/PhD/code/population-model/validated/src/main",
               "/resources/results/review/20170908-120243:448/tables/", sep = "")
data.death = read.csv(paste(path, "death-CT.csv", sep = ""), sep = ',', header = T)
data.obirth = read.csv(paste(path, "ob-CT.csv", sep = ""), sep = ',', header = T)
data.mbirth = read.csv(paste(path, "mb-CT.csv", sep = ""), sep = ',', header = T)
data.partner = read.csv(paste(path, "part-CT.csv", sep = ""), sep = ',', header = T)
data.sep = read.csv(paste(path, "sep-CT.csv", sep = ""), sep = ',', header = T)
```

Column abbriviations:

- NPCIAP - Number of previous children in any partnership
- CIY - Children in year (Yes/No)
- NCIY - Number of children in year
- NPA - New partners age
- NCIP - Number of children in partnership

These tables are as follows (data will be cleaned later):

```
head(data.death, 2)
```

```
##   Source    Sex Age Died Date         freq
## 1   STAT FEMALE 109   NO 1800 2.896647e-02
## 2   STAT FEMALE  52   NO 1995 1.067993e+03
```

```
head(data.obirth, 2)
```

```
##   Source    Age NPCIAP CIY Date     freq
## 1   STAT 20to24     0+ YES 1760 710.3867
## 2   STAT 20to24     0+ YES 1755 710.4510
```

```
head(data.mbirth, 2)
```

```
##   Source    Age NCIY Date     freq
## 1   STAT 15to49    0 1757 41045.67
## 2   STAT 15to49    0 1756 41065.84
```

```
head(data.partner, 2)
```

```
##   Source    Age    NPA Date     freq
## 1   STAT 25to29 25to29 1869 254.6565
## 2   STAT 25to29 25to29 1868 254.7067
```

```
head(data.sep, 2)
```

```
##   Source NCIP Separated Date freq
## 1    SIM    3        NO 1975  398
```

```
## 2    SIM    3        NO 1974  369
```

Death Analysis

```
# Standardise the data
data.death$freq <- round(data.death$freq)
data.death <- data.death[which(data.death$freq != 0), ]
data.death <- data.death[which(data.death$Date >= 1855) , ]
data.death <- data.death[which(data.death$Date < 2015) , ]

# Analysis
library("MASS")
model = loglm(freq ~ Date + Sex + Age + Died + Sex:Age + Sex:Died + Age:Died
              + Sex:Age:Died, data = data.death)
model
```

```
## Call:
## loglm(formula = freq ~ Date + Sex + Age + Died + Sex:Age + Sex:Died +
##     Age:Died + Sex:Age:Died, data = data.death)
##
## Statistics:
##                       X^2     df P(> X^2)
## Likelihood Ratio 30155.3 118411        1
## Pearson          27926.5 118411        1
```