Comparing demographic populations using Log Linear Models

Tom Dalton

22/02/2017

This ideas underlying this approach were first discussed in the meeting on 18/2/2017 between Monique MacKenzie, Graham Kirby and Tom Dalton.

The purpose of this document is to layout the statistical approach to assert the similarity of two demographic populations. The input is taken to be in the form of a contingency table. This contingency table is derived from the simulated population and the set of summary input tables.

Example data files

Through out we will make use of several different example data files, print outs of these can be seen in *Appendix A*.

```
close <- read.table("fake-pop-close-match.dat", header = T)
wayward <- read.table("fake-pop-wayward.dat", header = T)
```

This is the top of one of these data tables:

```
head(close)
```

```
##    yob sex age died source   freq
## 1 1990   m   0   no     in 900.00
## 2 1990   m   0  yes     in 100.00
## 3 1990   m   1   no     in 891.00
## 4 1990   m   1  yes     in   9.00
## 5 1990   m   2   no     in 882.09
## 6 1990   m   2  yes     in   8.91
```

Each of the columns hold descriptors of people in the populations and the frequency column indicates how may people in the population have these properties.

The models we talk about are focused on understanding the interactions between all the other columns and how they impact on frequency.

Approach 1 - Identify if a powerful model exists that does not contain interactions with source

- Here we create a log linear model using all values in a contingency table
- We then select the most parsimonious model based on AIC
- We then consider the interactions in the model
- If no source interactions remain in the selected model and the model exhibits good predictive power then we can conclude that the 'in' and the 'sim' populations can be considered one

```
library("MASS")
 dT <- close
 model.sat = loglm(freq ~ yob * age * sex * died * source, data = dT)
 model.step.result = step(model.sat, direction = "backward")
 model.sel = eval(parse(text=model.step.result["call"]))
 cat("If this below model does not contain any source interactions
       then we can assert that the sim and input populations are of
       the same specified statistical properties\n")
 model.sel
```

```
## If the below model does not contain any interactions of source
##         then we can assert that the sim and input populations are of
##         the same specified statistical properties

## Call:
## loglm(formula = freq ~ age + sex + died + age:sex + age:died +
##     sex:died + age:sex:died, data = dT, evaluate = FALSE)
##
## Statistics:
##                       X^2 df P(> X^2)
## Likelihood Ratio 0.2552486 12        1
## Pearson          0.2549084 12        1
```

We can see here that we have a model with high explanative power (high value of P) which makes no reliance on the variable source in the model. Therefore we can assert that the people derived from the input summary tables and those counted in the simulation population can be seen as originating from the same population.

Question: is this sufficient to claim that our two populations can be considered to conform to the same summary input properties that we use to inform our simulation?

We can repeat this with a populations that are a poorer match:

```
 dT <- wayward
```

```
 model.sel
```

```
## Call:
## loglm(formula = freq ~ age + sex + died + source + age:sex +
##     age:died + sex:died + age:source + sex:source + died:source +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source, data = dT, evaluate = FALSE)
##
## Statistics:
##                  X^2 df P(> X^2)
## Likelihood Ratio   0  0        1
## Pearson            0  0        1
```

Again we have selected the most parsimonious model which also has strong predictive ability. However, as can be seen the model's formula we see that source interactions are required to create a model that is able to predict well. Therefore we can assert that the two sets of people ('in' and 'sim') are distinctly different and thus say that in this contingency table the two populations do not follow the same input summary data.

Full code output traces for approach 1 can be seen in appendix B.

Approach 2 - Train two models and compare

This second approach has been born out of a desire to find a way to perform the comparison and get a singular value that indicates similarity.

The approach entails:

- Sub-setting the contingency table by source
- We then create a log linear model (excluding source) based on the rows with source 'in'
- And a second model (excluding source) based on the rows with source 'sim'
- We then select the most parsimonious models based on AIC
- We then compare these models

Question: is there a way to take a meaningful value from this comparison?

```
library("MASS")
d <- close
d.in = subset(d, source == "in")
d.sim = subset(d, source == "sim")

model.in.sat = loglm(freq ~ yob * age * sex * died, data = d.in)
model.in.step.result = step(model.in.sat, direction = "backward")
model.in.sel = eval(parse(text=model.in.step.result["call"]))

model.sim.sat = loglm(freq ~ yob * age * sex * died, data = d.sim)
model.sim.step.result = step(model.sim.sat, direction = "backward")
model.sim.sel = eval(parse(text=model.sim.step.result["call"]))

model.in.sel.glm = glm(model.in.sel, data = d, family = poisson)
model.sim.sel.glm = glm(model.sim.sel, data = d, family = poisson)
```

```
anova(model.in.sel.glm, model.sim.sel.glm, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
## Model 2: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        16     62.983
## 2        16     62.983  0        0
```

```
AIC(model.in.sel.glm, model.sim.sel.glm)
```

```
##                    df AIC
## model.in.sel.glm    8 Inf
## model.sim.sel.glm   8 Inf
```

Question: Here we want to identify if the two models are statistically similar. Is this a legitimate way to do so? Or is there another metric we should investigate? My understanding is that the anova analysis assumes nested models are therefore isn't sutable in this setting. AIC doesn't have this

underlying assumption but doesn't seem to yield values that support the findings for the respective datasets under approach 1.

We can also do this for the wayward population:

```
  d <- wayward
```

```
  anova(model.in.sel.glm, model.sim.sel.glm, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
## Model 2: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        16     8158.9
## 2        16     8158.9  0        0
```

```
  AIC(model.in.sel.glm, model.sim.sel.glm)
```

```
##                   df      AIC
## model.in.sel.glm   8 8409.813
## model.sim.sel.glm  8 8409.813
```

Full code output traces can be seen in appendix C for option 2.

Appendix

A

```
close <- read.table("fake-pop-close-match.dat", header = T)
close
```

```
##     yob sex age died source   freq
## 1  1990   m   0   no     in 900.00
## 2  1990   m   0  yes     in 100.00
## 3  1990   m   1   no     in 891.00
## 4  1990   m   1  yes     in   9.00
## 5  1990   m   2   no     in 882.09
## 6  1990   m   2  yes     in   8.91
## 7  1990   f   0   no     in 500.00
## 8  1990   f   0  yes     in 500.00
## 9  1990   f   1   no     in 495.00
## 10 1990   f   1  yes     in   5.00
## 11 1990   f   2   no     in 490.05
## 12 1990   f   2  yes     in   4.95
## 13 1990   m   0   no    sim 901.00
## 14 1990   m   0  yes    sim  99.00
## 15 1990   m   1   no    sim 890.00
## 16 1990   m   1  yes    sim  10.00
## 17 1990   m   2   no    sim 882.00
## 18 1990   m   2  yes    sim   9.00
## 19 1990   f   0   no    sim 501.00
## 20 1990   f   0  yes    sim 499.00
## 21 1990   f   1   no    sim 494.00
## 22 1990   f   1  yes    sim   6.00
## 23 1990   f   2   no    sim 491.00
## 24 1990   f   2  yes    sim   4.00
```

```
wayward <- read.table("fake-pop-wayward.dat", header = T)
wayward
```

```
##     yob sex age died source  freq
## 1  1990   m   0   no     in 18023
## 2  1990   m   0  yes     in  1975
## 3  1990   m   1   no     in 17855
## 4  1990   m   1  yes     in   170
## 5  1990   m   2   no     in 17680
## 6  1990   m   2  yes     in   175
## 7  1990   f   0   no     in 11000
## 8  1990   f   0  yes     in  9000
## 9  1990   f   1   no     in 10890
## 10 1990   f   1  yes     in   110
## 11 1990   f   2   no     in 10780
## 12 1990   f   2  yes     in   110
## 13 1990   m   0   no    sim 16000
## 14 1990   m   0  yes    sim  4000
```

```
## 15 1990   m   1   no    sim 18000
## 16 1990   m   1  yes    sim    12
## 17 1990   m   2   no    sim 22000
## 18 1990   m   2  yes    sim   500
## 19 1990   f   0   no    sim 10000
## 20 1990   f   0  yes    sim 12000
## 21 1990   f   1   no    sim  9000
## 22 1990   f   1  yes    sim  1000
## 23 1990   f   2   no    sim  9000
## 24 1990   f   2  yes    sim  1000
```

B

```
library("MASS")
 dT <- close
 model.sat = loglm(freq ~ yob * age * sex * died * source, data = dT)
 model.step.result = step(model.sat, direction = "backward")
```

```
## Start:  AIC=48
## freq ~ yob * age * sex * died * source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + yob:age:died:source + yob:sex:died:source +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + yob:age:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + age:sex:died:source
##
##
## Step:  AIC=48
```

```
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + age:sex:source + age:died:source +
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + age:sex:died +
##     age:sex:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
```

```
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + age:source + sex:source +
##     died:source + age:sex:died + age:sex:source + age:died:source +
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + age:died + sex:died + age:source + sex:source +
##     died:source + age:sex:died + age:sex:source + age:died:source +
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + age:sex +
##     age:died + sex:died + age:source + sex:source + died:source +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + age:sex + age:died +
##     sex:died + age:source + sex:source + died:source + age:sex:died +
##     age:sex:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:source + sex:source + died:source + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##                         Df    AIC
## - age:sex:died:source  2 44.087
## <none>                   48.000
##
## Step:  AIC=44.09
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:source + sex:source + died:source + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source
##
##                   Df     AIC
## - age:sex:source   2  40.087
## - age:died:source  2  40.253
## - sex:died:source  1  42.087
```

```
## <none>                44.087
## - age:sex:died     2 100.445
##
## Step:  AIC=40.09
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:source + sex:source + died:source + age:sex:died + age:died:source +
##     sex:died:source
##
##                   Df    AIC
## - age:died:source  2 36.254
## - sex:died:source  1 38.087
## <none>               40.087
## - age:sex:died     2 96.448
##
## Step:  AIC=36.25
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:source + sex:source + died:source + age:sex:died + sex:died:source
##
##                   Df    AIC
## - age:source       2 32.254
## - sex:died:source  1 34.254
## <none>               36.254
## - age:sex:died     2 92.615
##
## Step:  AIC=32.25
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     sex:source + died:source + age:sex:died + sex:died:source
##
##                   Df    AIC
## - sex:died:source  1 30.254
## <none>               32.254
## - age:sex:died     2 88.615
##
## Step:  AIC=30.25
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     sex:source + died:source + age:sex:died
##
##                Df    AIC
## - sex:source    1 28.255
## - died:source   1 28.255
## <none>            30.254
## - age:sex:died  2 86.615
##
## Step:  AIC=28.25
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     died:source + age:sex:died
##
##                Df    AIC
## - died:source   1 26.255
## <none>            28.255
## - age:sex:died  2 84.615
##
## Step:  AIC=26.26
```

```
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:sex:died
##
##                Df    AIC
## - source        1 24.255
## <none>            26.255
## - age:sex:died  2 82.616
##
## Step:  AIC=24.26
## freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##
##                Df    AIC
## <none>            24.255
## - age:sex:died  2 80.616
```

```
  model.sel = eval(parse(text=model.step.result["call"]))
  cat("If this below model does not contain any source interactions
        then we can assert that the sim and input populations are of
        the same specified statistical properties\n")
```

```
## If this below model does not contain any source interactions
##         then we can assert that the sim and input populations are of
##         the same specified statistical properties
```

```
  model.sel
```

```
## Call:
## loglm(formula = freq ~ age + sex + died + age:sex + age:died +
##     sex:died + age:sex:died, data = dT, evaluate = FALSE)
##
## Statistics:
##                         X^2 df P(> X^2)
## Likelihood Ratio 0.2552486 12        1
## Pearson          0.2549084 12        1
```

```
 library("MASS")
  dT <- wayward
  model.sat = loglm(freq ~ yob * age * sex * died * source, data = dT)
  model.step.result = step(model.sat, direction = "backward")
```

```
## Start:  AIC=48
## freq ~ yob * age * sex * died * source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
```

```
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + yob:age:died:source + yob:sex:died:source +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + yob:age:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     yob:age:sex:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + yob:age:sex:died +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     yob:died:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + yob:sex:source + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + yob:age:source + age:sex:source + age:died:source +
```

```
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + yob:sex:died +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + yob:age:died + age:sex:died +
##     age:sex:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + yob:age:sex + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + yob:source + age:source +
##     sex:source + died:source + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + yob:died + age:died + sex:died + age:source + sex:source +
##     died:source + age:sex:died + age:sex:source + age:died:source +
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + yob:sex +
##     age:sex + age:died + sex:died + age:source + sex:source +
##     died:source + age:sex:died + age:sex:source + age:died:source +
##     sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + yob:age + age:sex +
##     age:died + sex:died + age:source + sex:source + died:source +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source
##
##
```

```
## Step:  AIC=48
## freq ~ yob + age + sex + died + source + age:sex + age:died +
##     sex:died + age:source + sex:source + died:source + age:sex:died +
##     age:sex:source + age:died:source + sex:died:source + age:sex:died:source
##
##
## Step:  AIC=48
## freq ~ age + sex + died + source + age:sex + age:died + sex:died +
##     age:source + sex:source + died:source + age:sex:died + age:sex:source +
##     age:died:source + sex:died:source + age:sex:died:source
##
##                         Df    AIC
## <none>                      48.00
## - age:sex:died:source  2 910.54
```

```
  model.sel = eval(parse(text=model.step.result["call"]))
  cat("If this below model does not contain any source interactions
        then we can assert that the sim and input populations are of
        the same specified statistical properties\n")
```

```
## If this below model does not contain any source interactions
##         then we can assert that the sim and input populations are of
##         the same specified statistical properties
```

```
  model.sel
```

```
## Call:
## loglm(formula = freq ~ age + sex + died + source + age:sex +
##     age:died + sex:died + age:source + sex:source + died:source +
##     age:sex:died + age:sex:source + age:died:source + sex:died:source +
##     age:sex:died:source, data = dT, evaluate = FALSE)
##
## Statistics:
##                   X^2 df P(> X^2)
## Likelihood Ratio    0  0        1
## Pearson             0  0        1
```

C

```
  library("MASS")
  d <- close
  d.in = subset(d, source == "in")
  d.sim = subset(d, source == "sim")

  model.in.sat = loglm(freq ~ yob * age * sex * died, data = d.in)
  model.in.step.result = step(model.in.sat, direction = "backward")
```

```
## Start:  AIC=24
## freq ~ yob * age * sex * died
```

```
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     yob:sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + age:sex + age:died +
##     sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + age:sex + age:died + sex:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##
##                Df  AIC
## <none>            24.0
## - age:sex:died  2 48.8
```

```
model.in.sel = eval(parse(text=model.in.step.result["call"]))

model.in.sel
```

```
## Call:
## loglm(formula = freq ~ age + sex + died + age:sex + age:died +
##     sex:died + age:sex:died, data = d.in, evaluate = FALSE)
```

```
##
## Statistics:
##                   X^2 df P(> X^2)
## Likelihood Ratio   0  0        1
## Pearson            0  0        1
```

```
model.sim.sat = loglm(freq ~ yob * age * sex * died, data = d.sim)
model.sim.step.result = step(model.sim.sat, direction = "backward")
```

```
## Start:  AIC=24
## freq ~ yob * age * sex * died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     yob:sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + age:sex + age:died +
##     sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + age:sex + age:died + sex:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
```

```
##
##                Df    AIC
## <none>            24.000
## - age:sex:died  2 51.644
```

```
  model.sim.sel = eval(parse(text=model.sim.step.result["call"]))

  model.in.sel.glm = glm(model.in.sel, data = d, family = poisson)
```

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 882.090000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 8.910000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 490.050000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 4.950000
```

```
  model.sim.sel.glm = glm(model.sim.sel, data = d, family = poisson)
```

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 882.090000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 8.910000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 490.050000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 4.950000
```

```
  anova(model.in.sel.glm, model.sim.sel.glm, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
## Model 2: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        16     62.983
## 2        16     62.983  0        0
```

```
  AIC(model.in.sel.glm, model.sim.sel.glm)
```

```
##                   df AIC
## model.in.sel.glm   8 Inf
## model.sim.sel.glm  8 Inf
```

```r
library("MASS")
d <- wayward
d.in = subset(d, source == "in")
d.sim = subset(d, source == "sim")

model.in.sat = loglm(freq ~ yob * age * sex * died, data = d.in)
model.in.step.result = step(model.in.sat, direction = "backward")
```

```
## Start:  AIC=24
## freq ~ yob * age * sex * died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     yob:sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + age:sex + age:died +
##     sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + age:sex + age:died + sex:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##
```

```
##                Df   AIC
## <none>             24.0
## - age:sex:died  2 489.6
```

```
  model.in.sel = eval(parse(text=model.in.step.result["call"]))

  model.in.sel
```

```
## Call:
## loglm(formula = freq ~ age + sex + died + age:sex + age:died +
##     sex:died + age:sex:died, data = d.in, evaluate = FALSE)
##
## Statistics:
##                   X^2 df P(> X^2)
## Likelihood Ratio    0  0        1
## Pearson             0  0        1
```

```
  model.sim.sat = loglm(freq ~ yob * age * sex * died, data = d.sim)
  model.sim.step.result = step(model.sim.sat, direction = "backward")
```

```
## Start:  AIC=24
## freq ~ yob * age * sex * died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     yob:sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + yob:age:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + yob:age:sex + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     yob:died + age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + yob:age + yob:sex + age:sex +
##     age:died + sex:died + age:sex:died
##
##
## Step:  AIC=24
```

```
## freq ~ yob + age + sex + died + yob:age + age:sex + age:died +
##     sex:died + age:sex:died
##
##
## Step:  AIC=24
## freq ~ yob + age + sex + died + age:sex + age:died + sex:died +
##     age:sex:died
##
##
## Step:  AIC=24
## freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##
##                Df    AIC
## <none>             24.00
## - age:sex:died  2 571.03
```

```
model.sim.sel = eval(parse(text=model.sim.step.result["call"]))

model.in.sel.glm = glm(model.in.sel, data = d, family = poisson)
model.sim.sel.glm = glm(model.sim.sel, data = d, family = poisson)

anova(model.in.sel.glm, model.sim.sel.glm, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
## Model 2: freq ~ age + sex + died + age:sex + age:died + sex:died + age:sex:died
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        16     8158.9
## 2        16     8158.9  0        0
```

```
AIC(model.in.sel.glm, model.sim.sel.glm)
```

```
##                   df      AIC
## model.in.sel.glm   8 8409.813
## model.sim.sel.glm  8 8409.813
```