University of Edinburgh, School of Mathematics
Statistical Research Skills (MATH11188)
Assignment 3 ,Group Scientific Report

E. Antypas, S. Liu, C.Giannikos, Y. Gu

07/04/2023

Introduction

This report performs exploratory analysis in the data of potential victims as recorded by four different types of organisations and obtains an estimation regarding the scale of modern slavery in the UK. The analysis is conducted through the implementation of the multiple systems estimation technique in the given lists of recorded victims.

Because the data have been collected through different resources, the victim coverage is only partial. The figure of 2428 victims, as recorded in the four given lists, contains only cases which were encountered and recorded by at least one of the organisations. As a result, it does not include the potential victims which were not encountered by any of the four. Therefore, a multiple systems estimation approach can be used to estimate the unknown number of cases that have not come to attention.

Statistical Model/Methodology/Analysis

1. Model Selection

Since $y_{0000}|\mu_{0000} \sim Poisson(\mu_{0000})$ and $log(\mu_{000}) = \beta_0$, the estimate of the true total of potential victims is

$$\hat{N} = exp(\hat{\beta}_0) + n.$$

For the purposes mentioned above, a Generalised Linear Model from the Poisson family with a log-link function was fitted to the observed data in R. At first, a glm considering the main effects from all four lists, as well as all the two-way interactions between them, was fitted. Then, the routine stepAIC was used to fit log-linear models, firstly considering the full model, and then deducting interaction effects stepwise, at each stage removing the term which offers the least information, i.e the lowest AIC, stopping at the point where further removal would result in a worse fit according to the AIC.

The model finally chosen for the four-list data contains all the main effects of the four lists and three of their six possible interactions, namely $LA \times NG$,$LA \times PF$ and $NG \times GO$. It must be noted that both a forward stepwise routine based on the AIC, as well as stepwise model selection routines based on the BIC (Bayesian Information Criterion) yield the exact same model.

2.Testing Goodness of Fit

The deviance of a glm is an appropriate measure for goodness of fit, since it demonstrates the difference between the log-likelihood of said model and of a model that perfectly fits the data. Here, the deviance of the fitted model was 11.3 on 7 degrees of freedom. Additionally, in the Deviance Residual Plot for the model (Figure 1), the deviance residuals are evenly distributed around the x-axis, and do not demonstrate a systematic pattern, thus the goodness of fit hypothesis is not violated.

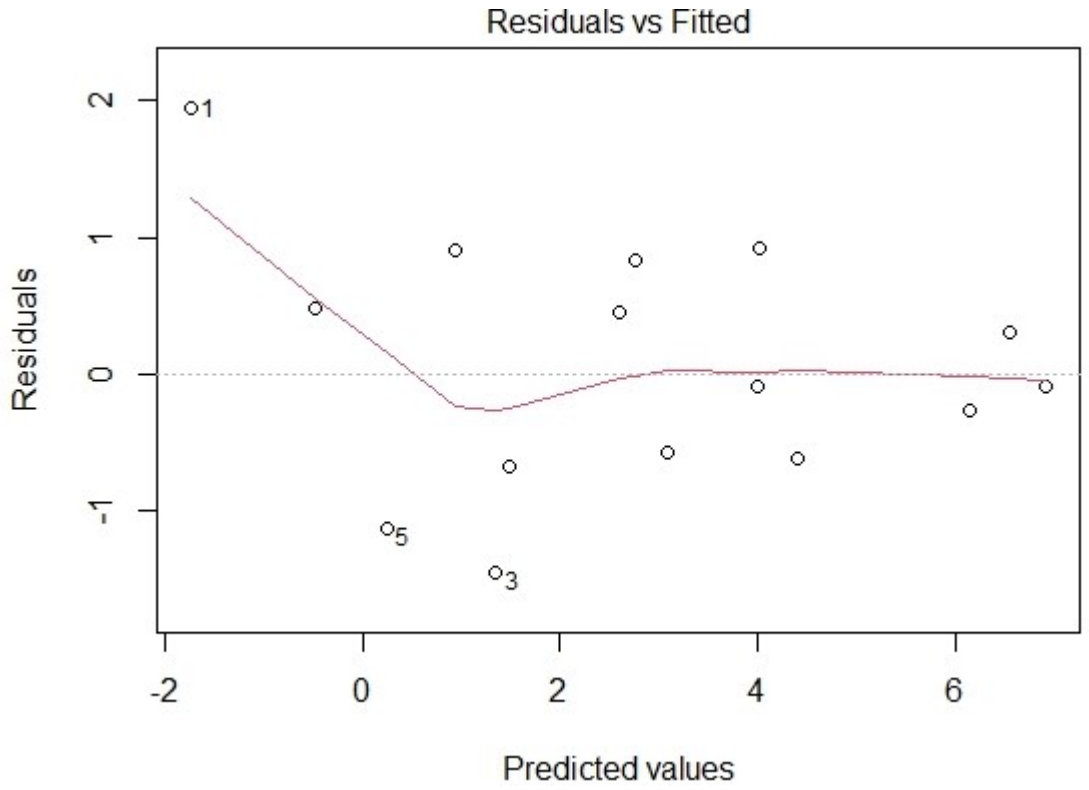

Figure 1: Deviance Residuals Plot

Result Summary

The model finally chosen for the four-list data provides an estimated total population size of 11015 with a standard error of 800. A 95% confidence interval is given by $[9447, 12583]$. Additionally, as demonstrated in the result output below, there is positive correlation between LA and each of NG and PF. This implies that being known to the local authority increases the chance of being known to NGOs or the police. However, there is a negative correlation between NG and GO, which implies that there exists a percentage of cases known to NGOs but unknown to Government agencies.

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.73059  -0.60592  -0.08187   0.62588   1.34699

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  9.05803    0.09315  97.242  < 2e-16 ***
LA          -5.05793    0.15418 -32.805  < 2e-16 ***
NG          -2.90609    0.09509 -30.563  < 2e-16 ***
PF          -2.14163    0.08810 -24.308  < 2e-16 ***
GO          -2.51454    0.09122 -27.564  < 2e-16 ***
LA:NG        1.49843    0.27687   5.412 6.23e-08 ***
LA:PF        0.89569    0.26283   3.408 0.000655 ***
NG:GO       -0.56185    0.22398  -2.508 0.012125 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 6047.262  on 14  degrees of freedom
Residual deviance:   11.341  on  7  degrees of freedom
AIC: 96.83

Number of Fisher Scoring iterations: 5
```

Figure 2: R Model Output