Modeling the Accidental Deaths

Mariyam Hafeez
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore, Pakistan

Muhammad Qaiser Shahbaz
Department of Mathematics
COMSATS Institute of Information Technology, Lahore, Pakistan

Abstract

The model for accidental deaths in the city of Lahore has been developed by using a class of Generalized Linear Models. Various link functions have been used in developing the model. The diagnostic checks have been carried out to see the validity of the fitted model.

1. Introduction

The accidents have been a major cause of deaths in many countries of the world. Lots of planning has been made across the world to reduce the impact of accidents on human life loss in many countries of the world. Almost all the countries have established special agencies to cope up with the accidental impacts. These countries have neatly monitored the data of injuries and also deaths from various types of accidents.

The recorded data for the accidents have been analyzed by lots of people across the world. The recorded data of accidents has been helpful for lot of other agencies to establish their policies. The actuaries are major beneficiary of the data related to accidental deaths. Jaffe (1998) have studied the impact of accidental deaths on actuarial policies by using a new table of accidental benefits given by Brodie and November (1959). Various factors have been associated with accidental deaths from time to time by various analysts. Edwards (1995) have given a comprehensive study of effect of depression on accidental deaths. Penttinen (2001) have studied the effect of mental disorder on accidental deaths. He has also studied the effect of said factor on suicide. The two studies conducted by Edwards (1995) and Penttinen (2001) have studied the effect of psychological factors on accidental deaths in the municipalities of Finland. The deaths caused by electricity in Sweden have been studied by Lindstorm et al (2006). The data in the study was the deaths caused by electricity in Sweden from 1975 to 2000. The study constitutes the descriptive analysis of the data collected. Same sort of study has also been conducted by Cawley and Homce (2003) in USA alongside the recommendations to reduce the accidental deaths in United States. Almost all the studies mentioned above have been limited to the descriptive analysis of the data or to some extend the tests of associations have been carried out to see the significant factors. None of these studies have attempted to model the accidental deaths by using some statistical model. In this paper we have modeled the accidental deaths by using statistical modeling techniques. The theoretical framework used in the modeling in given in section 2 of the paper and the data analysis have been carried out in section 3.

2. Theoretical Framework

Various modeling techniques are available in literature that can be used to obtain the models for prediction or forecasting of some response variable. These models range from classical regression models to complex time series models. The comprehensive review of these models can be found in Box and Jenkins (1976) and Neter et al (1996). The classical regression models are generally used to model a normal response variable. The time series models are used to forecast a sequence of dependent values.

The present study aims at modeling the number of deaths on the basis of various factors. Since the number of deaths is a discrete variable so this can be modeled by using the class of Generalized Linear Models, described by McCullagh and Nelder (1989). These models contain several models as the special case. The response variable in this class of models can belong to any member of the exponential family of distributions.

The general setup of the generalized linear models is given as:

$$\left.\begin{aligned} Y_i &= \eta_i + \varepsilon_i \\ \eta_i &= g\left[E\left(Y_i\right)\right] = g\left(\mu_i\right) = \mathbf{x}_i^{/}\boldsymbol{\beta} \\ E\left(Y_i\right) &= g^{-1}\left(\eta_i\right) = g^{-1}\left(\mathbf{x}_i^{/}\boldsymbol{\beta}\right) \end{aligned}\right\} \qquad (2.1)$$

The function $g\left(\mu_i\right)$ is any real valued function of the response mean and is generally called the link function. Model (2.1) contains several specialized models as a special case. Some of these models include Binary Logistic model, Multinomial Logistic model, Gamma Regression model, Negative Binomial model. Specifically the Poisson model to model the counts is given as a special case of (2.1) where the response variable has the Poisson distribution and the link function $g\left(\mu_i\right)$ is given as $g\left(\mu_i\right) = ln\left(\mu_i\right)$ or $g\left(\mu_i\right) = \mu_i^{\lambda}$ where λ is a positive real number. The selection of link function leads us to different Poisson model.

In the following section we have fitted the Poisson model to accidental deaths for data given in Appendix–A.

3. Analysis

In this section the analysis of the data has been carried out. The data constitute information about following variables from the city of Lahore. The data has been obtained from sources of Rescue 1122.

- Y : Expiry from the accident
- X1: Number of Accidents
- X2: Number of Fire cases
- X3: Number of Building Collapses
- X4: Number of Bomb Blasts

Initially we have applied the Kolmogrov – Simirnov test on the response variable to test whether the response variable follow the Poisson distribution or not. The computed value of the statistic turned out to be 0.963 with a p–value of 0.312 indicating that the response variable follows the Poisson distribution.

After deciding about the distribution we have fitted the Poisson model for prediction of the accidental deaths in Lahore. We have fitted two Poisson models for modelling of accidental deaths in the city of Lahore. These models have been fitted by using different link functions in (2.1) and assuming that the response variable follows the Poisson distribution.

Firstly, the model has been fitted by using the "log" link. The fitted model turned out to be:

$$\left.\begin{array}{l} ln\left(\hat{\mu}_i\right) = 1.324 + 0.003X_1 + 0.259X_3 + 0.261X_4 \\ \qquad\qquad (0.161)\;(0.001)\qquad (0.022)\qquad (0.062) \\ p - values: \qquad (< 0.001)\;\; (0.006)\quad (< 0.001) \end{array}\right\} \quad (3.1)$$

Initially the model has been fitted by using all the explanatory variables. It was found that the Fire cases have insignificant role in prediction of the number of accidental deaths so this variable was dropped from the model. The revised model was fitted without using the fire cases. The value of overall Chi–Square statistic is 46.265 with 3 degrees of freedom. This value is highly significant. The Deviance of the model is 96.781 which at 68 degrees of freedom indicate high adequacy of the model. All the Leverage Values and Cook's distances are within the normal range indicating no influential observation or outlier.

We have again fitted the Poisson model by using the power link. The fitted model this time is:

$$\left.\begin{array}{l} \hat{\mu}_i^{0.5} = 1.692 + 0.004X_1 + 0.105X_3 + 0.414X_4 \\ \qquad (0.224)\ (0.001) \quad (0.036) \quad (0.098) \\ p-values: \quad (<0.001)\ (0.003) \qquad (<0.001) \end{array}\right\} \qquad (3.2)$$

Again it was found that the Fire cases have insignificant role in prediction of the number of accidental deaths so this variable was dropped again. The revised model is fitted without using the fire cases as explanatory variable. The value of overall Chi–Square statistic is 47.659 with 3 degrees of freedom. This value is highly significant. The Deviance of the model is 95.387 which at 68 degrees of freedom indicate high adequacy of the model. All the Leverage Values and Cook's distances are within the normal range indicating no influential observation or outlier.

4. Conclusion

The Poisson model is fitted to the data of accidental deaths. Two different models were fitted by using two different link functions. It is found the Accidents, Building Collapses and Bombs play significant role in accidental deaths. It is also found that the power link has smaller deviance value as compared with the "log" link functions and consequently this model is more to predict the accidental deaths. The suggested model to predict the accidental deaths is:

$$E(Y_i) = \hat{\mu}_i = (1.692 + 0.004X_1 + 0.105X_3 + 0.414X_4)^2 \qquad (4.1)$$

This model can be used to predict the accidental deaths by using the significant response variables.

References

1. Box, G.E.P. and Jenkins, G.M. (1976). *Time Analysis, Forecasting and Control*, 2nd ed., Holden Day, San Francisco.
2. Brodie, N., and Noven, B. (1959). A New Table for Accidental Death Benefits," *Tran. Soci. Actur.* : 749.
3. Cawley J. C., Homce G. T. (2003). Occupational electrical injuries in the United States, 1992–1998, and recommendations for safety research. *J Safety Res,* 34, 241–248.
4. Edwards J. G. (1995). Depression, antidepressants, and accidents, *Br Med J.*, 311, 887–888.

5. Jeffe, J. M. (1998). Accidental death experience: A Review of Recent experience for the practicing actuary and the 1996 Accidental Death Benefits Mortality table, *Tran. Soci. Actur*., 159 – 178.

6. Lindstorm, R., Byland, P. O. and Erikkson, A. (2006) Accidental Deaths Caused by Electricity in Sweden, 1975–2000, *J. Forensic Sci*., 50(6), 1383–1388.

7. McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, Chapman & Hall, New York.

8. Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996) *Applied Linear Statistical Models*, Irwin Book, Chicago.

9. Penttinen, J. (2001). Risk of Suicide and Accidental Death among Subjects Visiting a Doctor because of Mental Disorder: A Matched Case-Control Study in Finnish Farmers, *J. Occup Health*, 43, 107 – 110.

Appendix – A

Acci-dents	Fire Cases	Building Collapse	Bomb Blast	Deaths	Acci-dents	Fire Cases	Building Collapse	Bomb Blast	Deaths
166	5	5	0	14	233	10	1	0	11
154	3	1	1	12	220	5	3	0	10
158	6	2	0	10	222	10	3	1	6
166	4	1	1	10	187	5	8	1	10
178	5	0	0	7	207	10	3	1	7
141	8	2	0	8	276	7	5	0	7
144	6	1	0	5	284	4	2	0	5
171	16	0	0	6	284	9	2	0	9
147	10	0	0	4	282	4	6	0	15
135	8	1	0	9	288	8	1	0	11
170	7	0	0	2	293	5	3	1	12
64	12	0	0	4	264	8	0	1	18
105	4	0	0	2	274	5	2	0	16
140	6	1	1	5	306	6	5	0	13
144	6	1	1	7	275	5	3	2	11
98	2	1	1	9	336	5	1	0	9
128	5	0	0	3	280	4	2	0	9
96	1	3	2	11	276	3	1	1	15
130	6	3	0	8	292	5	0	2	24
132	7	1	1	6	262	8	0	0	7
152	5	2	0	6	261	8	2	0	17
160	5	2	1	7	306	10	1	0	18
177	10	2	0	11	269	4	1	1	17
215	4	1	0	8	282	3	0	0	7
215	14	3	0	5	253	10	1	0	8
208	31	1	0	11	226	8	0	0	9
207	17	1	1	7	287	10	0	0	6
206	21	1	1	10	233	7	2	2	15
173	6	1	1	7	260	7	7	0	20
196	33	1	1	5	213	7	2	1	10
241	14	0	1	8	209	4	1	1	8
183	12	1	1	14	265	5	3	1	7
230	11	1	0	5	258	11	0	1	10
264	17	1	0	4	265	6	3	0	9
212	12	0	0	4	313	7	3	0	9
185	11	1	1	16	315	12	1	0	5