8th International Congress on Civil Engineering,
May 11-13, 2009, Shiraz University, Shiraz, Iran

Evaluation of Mathematical Models Used as Aggregate Crash Prediction Models

Jalil Shahi[1], Mahmoud Ahmadinezhad[2], Ali Naderan[3]
1- Associate Professor, Iran University of Science and Technology
2- Assistant Professor, Iran University of Science and Technology
3- Ph.D. Candidate, Iran University of Science and Technology
Address: Tehran, Narmak

Corresponding Author's E-mail: Ali@Naderan.com

Abstract

Aggregate crash prediction models (ACPM) attempt to relate the independent variables of interest aggregated at the zonal level to accidents using a mathematical equation. They predict number of crashes at each traffic analysis zone, and provide a basic tool to assess safety implications of large-scale transportation facilities at the network level for the long-range safety planning. This paper aims to critically review previous researches on the issue and evaluate the methodological approaches adopted to develop ACPM. A detailed analysis will be carried out on the mathematical formulation, goodness-of-fit tests, and the theory behind developing each model.

Keywords: Mathematical models, Aggregate crash prediction, Poisson, Negative binomial

1. INTRODUCTION

Statistical and econometric modeling is a key source of knowledge within many scientific disciplines, especially those in which controlled experiments are difficult or expensive to conduct [1]. This approach is particularly useful in crash analysis and in identifying and evaluating potential countermeasures. It is impractical if not impossible to conduct experimental analysis of traffic crashes, so statistical models should be hired to explain them.

In econometric modeling, theoretical aspects come first. That is, variables should be selected as part of a well-defined theory, rather than a try and error approach using statistically significance and goodness-of-fit tests. For example, if the idea is to use Aggregate Crash Prediction Models (ACPM) to develop safety countermeasures, they should consist of variables that may be influenced by engineering (e.g. improvement of geometric characteristics) or policymaking (road classification/construction). On the other hand, if they are meant purely for forecasting number of crashes in future, demographic and socio-economic variables may be best to use. After defining the structure of the model variables, it is essential to understand the mathematical representation of the issue at hand, i.e. crashes, and then select the best available model formulation.

The following steps constitute the development procedure of a mathematical model [2], [3]:

1. Variable selection: identifying preliminary variables based on theoretical issues,
2. Data collection: Development of a database containing variables used for modeling,
3. Model postulation: Hypothesizing the mathematical form of the relationship between variables,
4. Model calibration: Estimating the parameters of the postulated model based on observed data,
5. Model validation: Determining how well the calibrated model explains the observed data, based on Goodness-of-fit tests.

This paper critically reviews previous researches on the issue and evaluates the methodological approaches adopted to develop ACPM. A detailed analysis will be carried out on the model postulation, calibration, validation, variables, and the theory behind developing each model. The comprehensive judgment of these models is performed based on the following criteria [2], [4]:

1. Overall model fit (coefficient of determination, R^2 or the like),
2. Significance of parameter estimates (*t* test for all β_i, p-value),
3. Equality of all parameters to zero (overall model worthiness, *F* test or the like),
4. Standard error of estimate of the dependent variable,
5. Sign and size of parameter estimates (intuitiveness of parameters)