THE PROCEDURE OF POISSON REGRESSION MODEL

BY :

ZURAIRA BINTI LIBASIN
SHAMSUNARNIE BINTI MOHAMED ZUKRI
SURYAEFIZA BINTI KARJANTO

MARCH 2011

Contents

5. Report

5.1 Proposed Executive Summary

The Poisson regression model is one of the nonlinear regression models where the response outcomes are discrete (Frome *et al,* 1973). Therefore all the theories including the model development, model building, diagnostics and inferences that have been used in the analysis of Poisson regression model are carried out in a similar fashion as for the nonlinear regression models so that the accurate explanation will be obtained (Neter *et al,* 2003). However, most researchers take a simple way to model the Poisson regression without concerning the assumptions that corresponding to the model before doing any further analysis. This problem also happens when modelling the linear regression model (Suryaefiza *et al,* 2007). Failure to fulfil the assumptions will cause an inaccurate model as well as insignificant study. The objectives are: (1)To identify the existance of departures in Poisson regression model through the appropriate diagnostic tools, (2) To fix the departures by using the appropriate goodness measurement. The Poisson regression model is frequently used to analyze count data (Heinzl & Mittlbock, 2003). It is also widely used in the prediction of counts on potential independent variables (Choi *et al,* 2005). According to Maura (2000), interest often lies in predicting a rate or incidence such that bacteria counts per unit volume or cancer deaths per person-months of exposure to a carcinogen, and determining its relationship to a set of explanatory variables. Poisson regression became popularized as an analysis method in the 1970s and 1980s (research done by Charnes *et al,* 1976; and Frome, 1983). Later in 1990s, Poisson regression is a widely used modelling technique such that a homicide incidence study (Shahpar & Guohua, 1999), a study of injuries incurred by electrical utility workers (Loomis *et al,* 1999), and an evaluation of the risk of endometrial cancer as related to occupational physical activity (Moradi *et al,* 1998). Currently, the applications of Poisson regression models are widely used in a various type of field such as biomedicine (Waltoft, 2009), accident analysis and prevention (El-Basyouny & Sayed, 2009), insurance (Morata, 2009), biostatistics and epidemiology (Shults, 2005), environmental sciences (Agarwal et al, 2002), criminology (Osgood, 2000) and agriculture (Hall, 2000). The Poisson regression model is one of the nonlinear regression models where the response outcomes are discrete; therefore all the theories including the model development, model building, diagnostics and inferences that have been used in the analysis of Poisson regression model are carried out in a similar fashion as for the nonlinear regression models (Neter *et al,* 2003). Estimation of the parameters of a nonlinear regression model is usually carried out by the method of least squares or the method of maximum likelihood (Spiegelman & Hertzkmark, 2005; Linda & Julio, 2001). Unlike in linear regression, it is usually not possible to find analytical expressions for the least squares and maximum likelihood estimators for nonlinear model regression models. Inferences about the regression parameters in nonlinear regression are usually based on large-sample theory. This theory states that the least squares or maximum likelihood estimators for nonlinear regression models with normal error terms, when the sample size is large, are approximately normal distributed and almost unbiased, and have almost minimum variance. In Poisson regression model, a large-sample test of a single regression parameter can be constructed by using Wald test (Yang *et al,* 2009) whether a large-sample tests concerning of several regression parameters can be constructed by using likelihood ratio test (Hardin *et al,* 2007). The model building process of the regression model considers the selection of variables, diagnostic tools and remedial measures. The automatic selection procedures that have been used are stepwise method and *enter* method (Kleinbaum *et al,* 1998). The model building process for nonlinear regression models often differs somewhat from that for linear regression models. The reason is that the functional form of many nonlinear models is less suitable for adding or deleting predictor variables and interaction effects in the direct fashion that is feasible for linear regression models. Use of diagnostic tools to examine the appropriateness of a fitted model plays an important role in the process of building a nonlinear regression model. Plots of residuals can be helpful in diagnosing departures from the assumed model. Two types of the goodness of fit tests can be determined are Pearson chi-square (Neter *et al,* 2003) and the deviance goodness of fit test (Wang *et al,* 2007). If unequal error variances are found to be present, weighted least squares can be used in fitting the nonlinear regression model (Bender & Heinemann, 1995). Alternatively, transformations of the response variable can be investigated that may stabilize the variance of the error terms and also permit use of a regression model (Neter *et al,* 2003). Multicollinearity can be verified by using the collinear quantity and condition index that can be obtained from the eigen system (Lazaridis, 2007). Plots of deviance residuals help to identify the outliers that indicate in the model (Shrestha, 2007).

5.2 Enhanced Executive Summary

This study considers an analysis using a Poisson regression model where the response outcome is a count, with large outcomes being rare events. Estimates of the parameters are obtained by using the maximum likelihood estimates. Inferences about the regression parameters are based on Wald test and likelihood ratio test. In the model building process, the stepwise selection method were used to determine important predictor variables, diagnostic tools were used in detecting multicollinearity, non-constant variance, outliers, and also analysis of residual were used to measure the goodness fit of the model. Applications of these methods are illustrated by employing a case study of lower respiratory illness data in infants which took repeated observations of infants over one year (LaVange et at, 1994). Six explanatory variables involve the number of weeks during that year for which the child is considered to be at risk, crowded conditions occur in the household, family's socioeconomic status, race, passive smoking, and age group. We found that the explanatory variables which contribute significantly are passive smoking and crowding. Social economic status and race do not appear to be influential, and neither does age group.

Keywords: Poisson distribution, Poisson Regression Model, Nonlinear Model, Model Building