Model Regresi Zero-and-One-Inflated Poisson = Zero-and-One-Inflated Poisson Regression Model

Rida Martiza, author

Abstrak

Regresi Poisson merupakan generalized linear models (GLM) yang umum digunakan untuk memodelkan hubungan antara variabel respon berbentuk count data dengan satu atau lebih kovariat. Hanya saja, kerap dijumpai count data yang tidak memenuhi asumsi equidispersion sehingga tidak dapat dimodelkan dengan regresi Poisson. Salah satu penyebabnya adalah fenomena overdispersion yang teridentifikasi dengan banyaknya observasi yang bernilai nol (excess zeros) pada count data. Model regresi Zero-Inflated Poisson (ZIP) dapat digunakan untuk memodelkan count data yang mengalami overdispersion akibat excess zeros. Namun, pada beberapa kasus, count data dapat mengandung excess zeros dan excess ones dalam suatu periode waktu tertentu. Oleh karena itu, diperkenalkan solusi atas permasalahan tersebut menggunakan sebuah distribusi baru, yaitu distribusi Zero-and-One-Inflated Poisson (ZOIP), yang dibangun berdasarkan distribusi Bernoulli dan Poisson. Pada skripsi ini, dikonstruksi model regresi ZOIP untuk memodelkan count data yang mengandung excess zeros dan excess ones dalam suatu periode waktu tertentu. Parameter model regresi ZOIP tersebut diestimasi menggunakan metode maksimum likelihood dan algoritma Expectation Maximization (EM). Selanjutnya, diaplikasikan model regresi ZOIP dengan satu kovariat dan tanpa kovariat ke data klaim asuransi mobil. Berdasarkan nilai Akaike Information Criteria (AIC), didapatkan bahwa model regresi tanpa kovariat lebih cocok untuk memodelkan data klaim asuransi mobil yang dipakai.
.....Poisson regression is a generalized linear model (GLM) that is commonly used to model the relationship between response variables in the form of count data with one or more covariates. However, it is often found that count data does not meet the equidispersion assumption, so it cannot be modeled using Poisson regression. One of the causes is the phenomenon of overdispersion which is identified by the number of observations that are zero (excess zeros) in the count data. The Zero-Inflated Poisson (ZIP) regression model can be used to model count data that experiences overdispersion due to excess zeros. However, in some cases, count data may contain excess zeros and excess ones in a certain period of time. Therefore, a solution to this problem was introduced using a new distribution, namely the Zero-and-One-Inflated Poisson (ZOIP) distribution, which was built based on the Bernoulli and Poisson distribution. In this thesis, a ZOIP regression model is constructed to model count data containing excess zeros and excess ones in a certain period of time. The parameters of the ZOIP regression model are estimated using the maximum likelihood method and the Expectation Maximization (EM) algorithm. Furthermore, the ZOIP regression model with a covariate and without covariates were applied to the car insurance claim data. Based on the Akaike Information Criteria (AIC) value, it was found that the regression model without covariates is more suitable for modeling the car insurance claim data used.