

Count Data Model in Actuarial Applications

SUMMARY

In an insurance sector, it is important to estimate the future cost from which insurers price the policy i.e. premiums collected from the insured. In a non-life insurance claim amount and number of claims are important in knowing the estimates of future cost. Generally, claim details are directly available sometimes we may not have that convenience, so we have considered two datasets. They are Singapore Claims (SC) and US National Medical Expenditure Survey(NMES).

Our objective is to check factors that influence a claim therefore we have taken number of claims (count data) from SC dataset as response variable. Hence, we opted for GLM as a route map to handle non-normal response data. The response variable is count data, we used Poisson regression with log link function for the mean of Poisson response.

From the study we found that claims amount age and gender as the significant factors. After proceeding with diagnostic, we found there are many influential points and these points represent bad data on the contrary, we found assumptions were satisfied in NMES dataset. To remove the influential points in SC dataset, we developed an algorithm and refitted the model. The diagnostic test was done on the refitted model and we observed it was better than the original.

There are limitations in the model. Over dispersion, Sampling techniques, Zero Inflated Poisson(ZIP) and Hurdle model were not attempted. If these techniques and tests are followed, it may give the better model. These methods may be considered for our future modeling attempts with count data.