Sirinya Teeraananchai 2009: Generalized Linear Models for Longitudinal Study of Car Insurance Claims in Bangkok, Thailand. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsrisawang, Ph.D. 116 pages.

The goal of this study was to model a number of car insurance claims for car policyholders in Bangkok during the five-year insurance period of 2001-2005. The data used in this study came from the Department of Insurance in Thailand which consisted of 3,635 observations with at least one claim count in the 5-year period. The methodologies of GEE and GLMMs approaches were applied by taking account of correlation and covariance structures of data such as First-order Autoregressive (AR(1)) and Compound Symmetry (CS), respectively. The appropriate model is suggested. The dependent variable was the claim counts with Poisson distribution while independent variables were indemnity, car-age, age, gender, occupation, car-group and engine size. These independent variables were chosen to model the relationship between the claim counts and the potential predictors, using a forward stepwise procedure for varliable selection.

The results showed that the GEE model with AR(1) correlation structure was more appropriate than the model with CS structure, as indicated by the values of Pearson Chi-square of residual/DF 0.64 and 0.63, respectively; and the percentages of correct prediction (R_P^2) with 52.32 and 52.05, respectively. The statistically significant factors at a 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. On the other hand, the GLMMs model with AR(1) covariance structure was also more appropriate than the model with CS structure, as indicated by the values of Generalized Chi-Square/DF 0.18 and 0.15, respectively; and R_P^2 with 53.34 and 53.38, respectively. Moreover, the lower Pseudo-AIC and Pseudo-BIC of the GLMMs model with AR(1) structure was used to indicate the more effective covariance structure. The statistically significant factors at the 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. The GLMMs and GEE for population averaged models had the same performance for estimating the claim counts with R_P^2 about 53 %, but the GLMMs model with AR(1) structure and random intercept effect showed more efficiency in predicting the number of claim counts with 76.28 %.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature