522/E1

A comparative study of Generalized Estimating Equations (GEE) and General Regression Neural Networks (GRNN) to assess the production of poultry breeders

J C Hapugoda, D B U S De S Perera* and K A D Deshani
Department of Statistics, University of Colombo, Colombo 03

The analysis of correlated data structures has been a research interest for many years. As a conventional statistical approach, the Generalized Estimating Equations (GEE) method is widely used as a method for handling correlated data. The Artificial Neural Networks (ANNs) technique is a relatively recent development, widely used as an alternative approach to the conventional statistical techniques. General Regression Neural Network (GRNN), for analyzing response variables which are continuous in nature, falls under the category of probabilistic neural networks, where only a fraction of the training samples required for a back-propagation neural network would be needed.

This study had two main objectives. Firstly, the obtained data will be modeled using both Generalized Estimating Equations (GEE) and General Regression Neural Network (GRNN) approaches. Secondly, the resulting models will be compared using the performance measurements; the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the correlation coefficient (R), in order to identify the best model.

The data set used for this study is based on the poultry industry which is considered to be among the most research intensive agricultural industries. The data on the production of chicks were available for nine batches of breeders/hens. The explanatory variables of this study include; 'Hen Age', 'Storage Days', 'Fertility Rate' which are continuous variables and 'Batch' which is an ordinal variable. The response variable is 'Percentage of Salable Chicks' which is also a continuous variable. 'Batch' is the cluster variable.

In constructing the GEE model, the forward selection method based on the Wald test statistic was used to select the final model. In the GRNN model, the holdout and the cross-validation methods were used to obtain the best model based on the lowest sum of squared error (SSE).

The GRNN model was discovered as the best model among the two approaches considered, due to lower RMSE and MAE values, while the R value was also higher than that of the GEE model. The designed GRNN model consists of three neurons in the input layer, one neuron in the output layer and two hidden layers.

udeni@stat.cmb.ac.lk Tel: 0112590111