Using Multinomial Logistic Regression Analysis in Artificial Neural Network: An Application

Murat Kayri* and Ömay Çokluk**
*Yüzüncü Yıl University, Faculty of Education
**Ankara University, Faculty of Educational Sciences
*E-mail address for correspondence: mkayri@yyu.edu.tr

Abstract: Determination of Artificial Neural Networks’ classification and parameter estimation with Multinomial Logistic Regression Analysis was examined in this study. One of the modeling types suggested in case of having the dependent variables in categorized/classified structure and the independent variables in different structures such as nominal, ordinal, and intervals etc. in a research pattern is “Multinomial Logistic Regression (MLR)” method. MLR and Artificial Neural Networks (ANN) based MLR Analyses’ findings were studied comparatively in the model, where the dependent variable performed categorical structure and the independent variables performed mixed (continuous-discrete) structure. For the research, real data that were gathered in the context of the study entitled “Studying Primary School Students’ Views on their Communications with the Teachers and the Expected Situation” were used by the “Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale” developed by Doğan (2009). Within the context of this study, the total score obtained from the scale was assigned as the dependent variable, and variables such as the school type (public-private), gender, grade, mother’s profession, father’s profession, mother’s educational status, father’s educational status, number of brothers/sisters, monthly income, and internet usage time were assigned as independent variables. ANN has classified the dependent variable in high correctness level and showed the model’s fit in a higher level than MLR. Moreover, ANN has obtained parameter coefficients unlike MLR. It was considered that the model studied was estimated more consistently and correctly with ANN.

Key words: Multinomial Logistic Regression Analysis, Classification, Artificial Neural Network

Introduction

Scientific researches not only aim at describing a current situation, but also constantly bringing up cause and effect relations between actions. From this point of view, it can be stated that the scientific researches try to present the cause and effect relations in a model and that they realize these models in different methods according to the structures of the data. Result(s) is taken as dependent variable, and the set of causes affecting the result as independent variables in cause and effect based researches. Modelings also vary according to the data structures of the dependent and independent variables.

Linear regression models (univariate linear, multiple, and multivariate linear) are commonly used models in social sciences; these models require the dependent and independent variables to be constant variables. Moreover, they also require meeting assumptions such as the distribution normality of dependent and independent variable sets, and error distribution normality of observation values.

On the contrary, the results obtained in social sciences researches are generally categorical and so it is not possible to study them within linear regression models’ context. Accordingly, one of the regression models that can be used in such cases is logistic regression. Logistic regression is an analysis type appropriate for the situations where the dependent variable is not continuous or quantitative, in other words, it is appropriate for categorical or nominal situations (Long, 1997; Mertler & Vannatta, 2005). As a result of this, logistic regression does not require meeting basic assumptions in linear regression models.
Three different types of logistic regression analysis can be stated according to the type of scale in which the dependent variable is measured, and the category (level / choice) number of the dependent variable. If the dependent variable is a categorical variable with two choices, it is named “Binomial / Binary Logistic Regression Analysis” (Stephenson, 2008).

If the dependent variable has only two categories, dichotomous, in other words, generally what is observed is likelihood of an action to occur or not or the existence / absence of a feature (Long, 1997). For instance, situations such as a student’s admission / nonadmission in an academic program or a student with / without learning difficulty are appropriate for this situation. If the dependent variable is a variable with more than two categories of classification, then it is named “Multinomial Logistic Regression Analysis (MLR)” (Stephenson, 2008). For instance, multinomial logistic regression analysis is applied for estimation of a dependent variable consisting of students attending five different faculties. Having the dependent variable with more than two categories, polytomous, in other words, is a situation frequently faced in application. However, the most important point to be considered here is whether the categories are ordinal or not since some models are only appropriate for ordinal categories whereas some other models can be used when the categories are ordinal or not. If the dependent variable is obtained by ordinal scale, in other words, if the categories are ordinal, then “Ordinal Logistic Regression Analysis” is used (Garson, 2005; Stephenson, 2008). For instance, a dependent variable in which the assertiveness levels of students are grouped as “low”, “mid” and “high” requires the application of ordinal logistic regression analysis. Eventually, the dependent variable in logistic regression analysis cannot be a continuous variable. However, level (type) of explanatory variables is not important and it is also possible to build mixed models.

Artificial Neural Networks (ANN) is an artificial intelligence application developed from the neural (neurological) pattern of human brain’s learning spot. First studies on ANN started with modeling the nerve cells forming the brain, and the application of these modelings into the computer systems. Afterwards, it has become common in many fields in parallel with the development in computer systems. ANN has efficient usage in many fields such as medicine, industry, biology, electronic systems, optimization, and social sciences (Golden, 1996).

Artificial Neural Networks could conduct both linear and non-linear model approaches together and so it can get the correlation between variables on a more valid basis (Eğrioğlu and Aladağ, 2005). ANN is accepted as a strong method that learns the structures of the current data, establishes a new relations network in the real world and conducts many statistical processes such as making parameter estimation, classification, optimization, and time series in this relations network in a determined way (Badr, Nasr & John, 2003; Elmas 2003; Fausett, 1994; Uzun and Erdem, 2005). This method analyzes the data set in three stages. At the first stage, a considerable portion of the data set is used for “training process”. ANN tries to detect the relationships between the variables of the data set and so it tries to determine the characteristic of the research pattern. At the second stage, based on the learnings of the first stage, it tries to perceive the model and this process is named the “perceptron process / hidden process”. In the perceptron process, the ideal functions that belong to the model are produced and weights (W_i) of explanatory variable(s) upon the dependent variables are obtained. The third stage is the new model estimation the ANN produces for the real world, and this process is named the “output process”. This three-staged process of ANN is presented in Figure 1.

![Figure 1. General structure / logic of ANN (Manel, Dias & Ormerod, 1999).](image)

X variables in Figure 1 show explanatory variables in the model, y variable shows dependent variables, and “b” variables show the set of functions ANN produced during perceptron process.

ANN tries to learn current data structure within the frame of some learning algorithms. Learning algorithms discussed in the literature are “Single Layer Perceptron – SLP”, “Additive Linear Element – ALE”, “Multi-Layer Perceptron – MLP”, “Perceptron Neural Network – PNN”, “General Regression Neural Network – GRNN”, and
“Radial Basis Function Networks – RBFN” methods (Gardner & Dorling, 1998; Wieland & Mirschel, 2008). According to the data set being linear or not, ANN not only suggests the ideal learning algorithm, but also shows possibilities of correct definitions for model by different learning algorithms. These learning algorithms base Feed-Forward Back-Propagation (FFBP) algorithm that minimizes Mean Square of Error – MSE. It is pointed out in many studies in the literature that FFBP algorithm is an effective method that minimizes MSE (Gardner & Dorling, 1998; Haykin, 2008; Kurt and Türe, 2005).

Considering the discussions mentioned above, the main aim of this study consists of the usage of MLR method together with Artificial Neural Networks (ANN). More clearly, the main aim in this study is to exemplify the applicability of ANN, which takes the human brain functioning as the basis and conducts this functioning in computers as artificial intelligence applications, into different statistical methods (Haykin, 2008) using MLR method.

MATERIALS AND METHOD

Sample - Instrument
This study is a correlational research. Correlational researches are the researches in which the relationship between two or more variables is analyzed without any intervention to these variables. Correlational researches are effective researches that reveal the relationship between variables, determine the level of this relationship, and provide further essential hints for higher quality researches (Büyüköztürk et. al, 2008). The actual aim of the study is to put forth the use of ANN in statistical analyses, and show its applicability upon a sample data set. Instead of forming this data set hypothetically, real data that were gathered in the context of the study entitled “Studying Primary School Students’ Views on their Communications with the Teachers and the Expected Situation” were used by the “Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale” developed by Doğan (2009). This scale has a 5-point Likert type response and it consists of 21 items grouping under a single factor. The scale was applied to 459 female and 364 male students (with a total of 823 students) who attended primary education’s second stage (6th-8th grades). The lowest score that can be obtained from the scale is 21 and the highest is 105. Cronbach-Alphä internal consistency coefficient of the scale is α =.94.

Process
The normality test of the total scores obtained from the “Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale” was analyzed by the Kolmogorov-Smirnov Test, and it was seen that the distribution was normal (p>0.05). Similarly, the homogeneity of the distribution was analyzed by the Levene Test and it was found homogeneous (F=.43, p=.84).

Within the context of this study, the total score obtained from the scale was assigned as the dependent variable, and variables such as the type of school (public-private), gender, grade, mother’s profession, father’s profession, mother’s educational status, father’s educational status, number of brothers/sisters, monthly income, and internet usage time were assigned as independent variables. This model formed in the research was tested both with MLR and ANN based MLR analysis.

In this study, ANN’s learning models (Linear, Multi-Layer Perceptrons (triad-quart), Radial Based Function, and Perceptron Neural Networks – Generalized Regression Neural Networks) were comparatively analyzed, and the learning algorithms that estimated the model with the highest likelihood were used for parameter estimations. “Feed-Forward Back-Propagation (FFBP)” algorithm was used for the model. Moreover, the network map of the ideal model was obtained.

From 823 individuals’ data in the data set, 413 were used for ANN’s “training process”, 205 for “performance”, and 205 for “test stage”. The individuals were randomly selected for the related processes. Individual numbers divided for training, performance, and test processes are ANN’s default values of the software used, and there was not any interventions made for these values. In other words, the software itself determined the number of individuals divided for each process. The trial versions of SPSS 16.0 and Statistica 7.0 were used for the analyses.
RESULTS

Findings of the research are presented in this section. The presentation of findings was made appropriate for the process steps followed in data analysis. These process steps are:

1. For the objective of getting the categorical dependent variable required for Multinomial Logistic Regression Analysis application, total scores obtained from the Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale were made a categorical dependent variable with cluster analysis.

2. The categorical total scores related to the Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale appeared as the dependent variable whereas variables such as the type of school (public-private), gender, grade, mother’s profession, father’s profession, mother’s educational status, father’s educational status, number of brothers/sisters, monthly income, and internet usage time appeared in the model as independent variables and they were applied MLR analysis.

3. The model that was structured in the research and summarized in the second step was tested with ANN based MLR analysis.

4. Results of Multinomial Logistic Regression analysis and ANN based MLR analysis were comparatively examined.

Correct defining of the individual differences or similarities in the population and, as mentioned before, the process of obtaining a multinomial dependent variable necessary for MLR application were realized with the cluster analysis applied in the first step.

Frequency and percentage distributions of the groups obtained from cluster analysis are presented in Table 1, and descriptive statistics of related groups are presented in Table 2.

Table 1. Percentage-frequency distribution of dependent variable in groups

<table>
<thead>
<tr>
<th>Group</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>274</td>
<td>33.30</td>
</tr>
<tr>
<td>2</td>
<td>264</td>
<td>32.10</td>
</tr>
<tr>
<td>3</td>
<td>285</td>
<td>34.60</td>
</tr>
<tr>
<td>Total</td>
<td>823</td>
<td>100.00</td>
</tr>
</tbody>
</table>

As Table 1 is evaluated, it is seen that the data set, which was homogeneous at first, actually consisted of three sub-groups with $n_1= 274$ (33.30%) students in the first group, $n_2=264$ (32.10%) in the second, and $n_3=285$ (34.60%) in the third group.

Table 2. Descriptive statistics of dependent variable

<table>
<thead>
<tr>
<th>Group</th>
<th>$\bar{X}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102.234</td>
<td>2.881</td>
</tr>
<tr>
<td>2</td>
<td>84.060</td>
<td>4.725</td>
</tr>
<tr>
<td>3</td>
<td>62.326</td>
<td>9.297</td>
</tr>
<tr>
<td>Overall Average</td>
<td>82.817</td>
<td>17.909</td>
</tr>
</tbody>
</table>

As Table 2 is evaluated, it is seen that 274 individuals in group 1 had an average of $102.234\pm2.881$; 264 individuals in group 2 performed a different manner structure than group 1 and had an average of $84.060\pm4.725$; 285 individuals in group 3 performed more different manner characteristics than the other groups and had an average of $62.326\pm9.297$.  

262
“Multinomial Logistic Regression” analysis was applied in the second step. Regression equation that had been previously accepted as the model was presented below. This model was applied a “Multinomial Logistic Regression” analysis.

\[ \text{Total Score} = \beta_0 + \beta_1 \times a_1 + \beta_2 \times a_2 + \beta_3 \times a_3 + \ldots + \beta_7 \times a_7 \]

Multinomial Logistic Regression analysis findings were respectively presented in tables below. Model fitting information obtained from MLR application is presented in Table 3.

### Table 3. Model fitting information

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Fitting Criteria</th>
<th>Likelihood Ratio Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2 Log Likelihood</td>
<td>( \chi^2 )</td>
</tr>
<tr>
<td>Intercept</td>
<td>1673.766</td>
<td>134.091</td>
</tr>
<tr>
<td>Final</td>
<td>1539.676</td>
<td></td>
</tr>
</tbody>
</table>

As Table 3 is evaluated, it is seen that -2 log likelihood value of basic model only with intercept term was 1673.766, this value decreased into 1539.676 with the independent variables’ appearance in the model, and the improvement / change occurred in the model was significant \( \chi^2_{(58)} = 134.091; p < .001 \).

Pseudo \( R^2 \) values of model are presented in Table 4.

### Table 4. Pseudo \( R^2 \) measurements

<table>
<thead>
<tr>
<th>Measurement</th>
<th>( R^2 ) Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox and Snell</td>
<td>.154</td>
</tr>
<tr>
<td>Nagelkerke</td>
<td>.174</td>
</tr>
<tr>
<td>McFadden</td>
<td>.076</td>
</tr>
</tbody>
</table>

As Pseudo \( R^2 \) values presented in Table 4 are examined, explanation ratios of dependent variables upon independent variables are seen. Nagelkerke \( R^2 \) value is the modified form of Cox & Snell coefficient, and so, it is always higher than Cox & Snell \( R^2 \) value (Garson, 2008; Hair et. al, 2006). As values in Table 4 are examined, it is seen that dependent variables define 15.4% of the variance in independent variables according to Cox & Snell \( R^2 \) value, 17.4% according to Nagelkerke \( R^2 \) value, and 7.6% according to McFadden value. Statistics of likelihood ratio tests obtained from MLR are presented in Table 5.
As Table 5 is examined, it is seen that school type variable caused a significant improvement / change in model fitting \( \chi^2 (6) = 26.380; p < .001 \). Other variables’ contributions into the model are not significant.

The data set was examined in the third step with ANN in order to conduct the ANN based MLR analysis application, which is the main aim of the study. From the sample of 823 individuals, the ANN used 413 for training, 205 for perceptron-selection, and 205 for final output-test. These sizes are ANN’s default values. The ideal model was tried to be reached by testing of each of the different learning algorithms in ANN. Correct estimation percentages of different learning algorithms are presented in Table 6.

As Table 6 is examined and education, perceptron, and test processes are evaluated together, it is seen that “Multi-layer Perceptron (MLP)” learning algorithm estimated the independent categorical variable best as a model. In this situation, it is required to take account of MLA’s classification estimation values for individuals in the data set. Function network MLA formed according to the relationship between dependent and independent variables is presented in
As the relationship network presented in Figure 2 is examined and the effect of 21 dependent variables (Students’ Expectations from their Teachers in Teacher-Student Communication Process Scale items) in the model on independent variable is searched, it was observed that dependent variables changed the parameter coefficients ($W_i$) for 104 times during the interaction in “education” and “perceptron” processes in order to improve the model, and in this sense, it was observed that a function with 104 pieces was produced. Afterwards, this polynomial number was decreased into 3, and eventually, dependent variable was attempted to be estimated with the common effect of this three-piece function formed. Findings of Multinomial Logistic Regression analysis applied after the values produced by Artificial Neural Network are presented in tables below. In other words, MLR analysis results presented in this section are ANN based.

**Table 7. Model fitting information**

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Fitting Criteria</th>
<th>Likelihood Ratio Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2 Log Likelihood</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Intercept</td>
<td>1613.402</td>
<td>524.813</td>
</tr>
<tr>
<td>Final</td>
<td>1088.589</td>
<td></td>
</tr>
</tbody>
</table>

As Table 7 is examined, it is seen that -2 log likelihood value related to the model only with the intercept term was 1613.402, and this value decreased into 1088.589 in the final model with the appearance of independent variables in the model, and it is also seen that this improvement / change occurred in the model was significant [$\chi^2_{(58)} = 524.813; \ p < .001$].

Pseudo $R^2$ values of the model are presented in Table 8.
As Pseudo $R^2$ values presented in Table 8 are examined, it is seen that dependent variables define 49.1% of the variance in independent variables according to Cox & Snell $R^2$ value, 55.7% according to Nagelkerke $R^2$ value, and 31.7% according to McFadden value.

Statistics of likelihood ratio tests obtained from MLR are presented in Table 9.

As Table 9 is examined, it is seen that school type [$\chi^2(6) = 209.978; p < .001$], grade [$\chi^2(4) = 12.987, p < .05$], and monthly income [$\chi^2(6) = 14.945; p < .001$] variables caused a significant improvement / change in model fitting.

In the fourth step of the study, when the model fitting information obtained as a result of the MLR application and ANN based MLR application comparatively examined, it was seen in MLR analysis that -2 log likelihood value related to the model only with the intercept term was 1673.766, and this value decreased into 1539.676 in the final model with the appearance of independent variables in the model, and it is also seen that this improvement / change occurred in the model was significant [$\chi^2(58) = 134.091; p < .001$]. As results of ANN based MLR application are examined, it was seen that -2 log likelihood value related to the model only with the intercept term was 1613.402, and this value decreased into 1088.589 in the final model with the appearance of independent variables in the model, and it is also seen that this improvement / change occurred in the model was significant [$\chi^2(58) = 524.813; p < .001$].

As results of the two analyses were compared, it was seen that -2 log likelihood value obtained from the ANN based MLR application in the starting model with the intercept term was lower. Moreover, although the decrease in the -2 log likelihood value, which was an indicator of the improvement occurred in the model by the appearance of independent models in the analysis, was 134.09, it was found 524.813 in ANN based MLR application.
As the Pseudo $R^2$ values obtained from Multinomial Logistic Regression analysis and ANN based MLR application are comparatively examined, it is seen that the dependent variable explains 15.4% of the variance in independent variable according to Cox & Snell $R^2$ value, 17.4% according to Nagelkerke $R^2$ value, and 7.6% according to McFadden value. When results of ANN based MLR application are examined, it is seen that dependent variables explain 49.1% of the variance in independent variable according to Cox & Snell $R^2$ value, 55.7% according to Nagelkerke $R^2$ value, and 31.7% according to McFadden value. There were significant changes in variance ratios explained as a result of the ANN based MLR application.

When the significant dependent variables obtained from the Multinomial Logistic Regression analysis application and ANN based MLR application were compared, it was found that only the school type variable in MLR and school type, grade, and monthly income variables in ANN based MLR application results caused significant improvements in model fitting. In other words, the effect of grade and monthly income variables that could not be observed in MLR application appears in ANN based application.

**DISCUSSION**

It is quite important to thoroughly define the effect of the dependent variables on independent variables in testing the models formed for cause and effect based researches. Another important point is to choose the right modeling according to the characteristics of the research’s data set. One of the modeling types suggested for the situations in which the dependent variables are of categorical/classified structure, and independent variables are different structures such as classified, ordinal, interval, etc. in a research pattern is “Multinomial Logistic Regression” analysis. In this way, relationship between independent variable(s) and dependent variable could be defined, and the effects of independent variables on categories of dependent variable and the level of these effects could be examined. The model examined with MLR in this study was also examined with ANN based MLR, and findings of the two analyses were compared.

ANN is a strong method in learning the current sample (characteristics of the current data), with the feature to know new possible samples and make efficient estimations (Elmas, 2003). There can be statistical operations such as linear – non-linear regression analyses, classification, and time series in ANN. In this study, the aim was to make parameter estimations for independent variables that have effects on a multinomial dependent variable. In this way, the validity level of classification of the dependent variable in the research pattern, and model explanation levels of independent variables after classification were examined. Model formed in the study was tested both with Multinomial Logistic Regression analysis and ANN based Multinomial Logistic Regression analysis, and results were comparatively evaluated.

In Multinomial Logistic Regression analysis application, only the effect of “school type” independent variable on categorical dependent variable was found significant. Contributions of other independent variables in the model were not significant. As a result of ANN based MLR analysis, it was found that “school type”, “grade”, and “monthly income” independent variables provided significant contributions into the model. It is important that ANN, unlike other method, determined the grade level and monthly income of family variables of the individuals in sample as significant independent variables. In many studies in the literature, it is stated that models suggested by ANN are more consistent (Badr, Nasr & John, 2003; Durmuş and Meriç, 2005; Kurt and Türe, 2005; Uzun and Erdem, 2005; Wieland & Mirschel, 2008). Besides, model explanation ratios of independent variables in MLR and ratios in ANN considerably differ. Thus, it can be claimed that reclassification of dependent variable in ANN would provide significant contributions into the improvement of model regarding the relationship between dependent and independent variables. Moreover, the correlation between results of individual classifications of cluster analysis and ANN’s own classification levels was examined with Spearman’s rho method and found significant ($r_s = 0.974; p<0.01$). Despite the fact that ANN conducted the classification differently than cluster analysis, the correlation relationship ($r_s$ ) shows ANN’s consistency in this subject.

The parameter estimation of Artificial Neural Network and its efficient usage in forming models such as classification and time series will provide a sound base for obtaining right inferences from application to theory. High level statistical methods of this kind are suggested to be used for cause and effect based researches generally in the fields of social sciences and particularly in educational sciences. From this aspect, this study is considered to be an example for the future researches in the field.
REFERENCES


