

Investigating Dental Care Status with Multilevel Modelling

*Elif Çoker*¹ – *Meral Yay*² – *Ömer Uysal*³

The aim of this study is to explore the status of dental care. For this purpose, the data set is taken from a real data set which the survey was carried out in 2003 at Istanbul University Cerrahpaşa School of Medicine, Turkey. The population is defined to be people who are over the age of 18, living in Istanbul. According to the Turkish Statistical Institute indicators, Istanbul is divided into three regions. In these regions, there are 25 towns in total and all of them are included. From these 25 towns, 285 districts are selected randomly. In total, the analyses are performed with the participation of 931 individuals. Taking into account the nested structure of the data set (individuals are nested within districts, districts within towns and towns within regions), multilevel modelling approaches are investigated.

Keywords: Dental care, Multilevel modelling, Nested data, Gifi

1. Introduction

Statistical methods are commonly used in social research. The usage of statistical methods in social research can be explained in two parts: the first half of it introduces descriptive statistics and inferential methods (confidence intervals and significance tests), the second half of it introduces bivariate methods (contingency table analysis, regression) and advanced regression methods (multiple regression, analysis of variance, logistic regression and its extensions). In this study we focused on an extension of logistic regression called multilevel multinomial logistic regression and its application in social research.

Generally social research is based on individuals. However, since individuals live in social groups, they can not be considered independently from them. Despite

¹ Elif Çoker, PhD student, Mimar Sinan Fine Arts University Faculty of Science and Letters, Department of Statistics (Istanbul)

² Meral Yay, PhD student, Mimar Sinan Fine Arts University Faculty of Science and Letters, Department of Statistics (Istanbul)

³ Ömer uysal, PhD student, Istanbul University Cerrahpaşa School of Medicine, Department of Biostatistics (Istanbul)

this structure, many social researchers aim to explain variability in behaviour and attitudes of individuals, here we are specifically interested in the use of statistical models to analyze quantitative data. Multilevel modelling which has an important place in statistical methods aims to redress the balance, by emphasizing both individuals and their social contexts.

The aim of this study is to describe the Multilevel Multinomial Logistic Regression Models and to apply these models to a data set collected in Turkey on Dental Care Status and then use the Gifi transformation to re-analyze the data set.

2. Logistic Regression Models

A great many variables in social sciences are categorical. It is hardly surprising that social scientist frequently wants to estimate regression models in which the response variable is categorical. In this context logistic regression is a statistical modeling method that can be useful. It describes the relationship between the categorical response variable and one or more continuous and/or categorical explanatory variables. Logistic regression is used when explanatory variables are either continuous or categorical and response variables are categorical. Categorical variables have two main types of measurement scales. These are nominal and ordinal scales. Nominal categorical variables have categories that have no natural order to them. Ordinal categorical variables have a natural order. The goal of logistic regression is to correctly predict the category of response for individual cases using the most parsimonious model.

Early uses were in biomedical studies, but the past 20 years have also seen much use in social science research and marketing (Agresti 2002). In this context there are two main uses of logistic regression. The first is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables.

Logistic regression does not make the typical assumptions: the responses, conditional on the explanatory variables do not have to be normally distributed, they don't have to be linearly related, and we don't require equal variance within each group. Suppose there is a single quantitative explanatory variable X . For a binary response variable Y , recall $\pi(x)$ denotes the "success" probability at value x . This probability is the parameter for the binomial distribution. Logistic regression has a linear form for the logit of this probability,

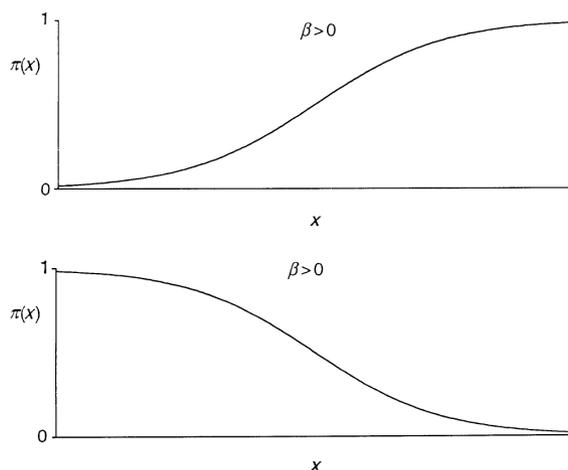
$$\text{logit}[\pi(x)] = \log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta x \quad (1)$$

This is called the logistic regression function (logit function). Equation (1) implies equation (2) for the probability $\pi(x)$, using the exponential function $\exp(\alpha + \beta x) = e^{\alpha + \beta x}$,

$$P(Y = 1 / X = x) = \pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} = \frac{1}{1 + e^{-(\alpha + \beta x)}} \quad (2)$$

In this function the parameter β indicates the rate of decrease or increase of the curve in the Figure 1. When $\beta > 0$, $\pi(x)$ increases as x increases (Agresti 2007).

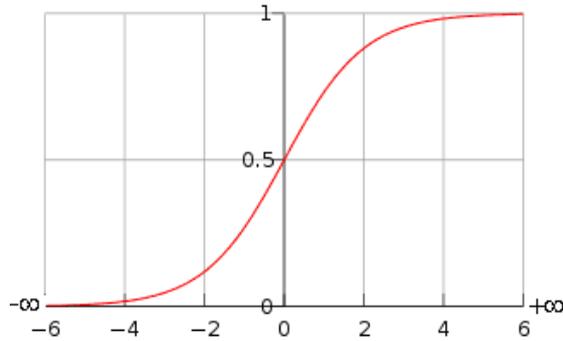
Figure 1. Logistic regression functions



Source: Agresti 2007

The shape of the logistic regression function, which describes the mathematical form of the logistic model can be seen in Figure 2. This shows that $\pi(x)$ increases or decreases as an S shaped function of x . The change in the $P(Y / X = x)$ per unit change in x becomes progressively lower as the conditional mean gets closer to zero or 1 (Hosmer-Lemeshow 2000).

Figure 2. Linear approximation to logistic regression curve



Source: Hosmer-Lemeshow 2000

3. Multinomial Logistic Regression Models

Multinomial logistic regression can be used with a categorical response variable that has more than two categories, but the categories can be ordered or unordered. It compares multiple groups through a series of binary logistic regressions. The group comparisons are equivalent to the comparisons for a dummy-coded response variable, with the category with the highest numeric score used as the baseline category. Like logistic regression, multinomial logistic regression does not make any assumptions of normality, linearity, or variance homogeneity. The multinomial logit model also assumes that the response variable cannot be perfectly predicted from the explanatory variables for any case.

Suppose Y to be a categorical response with C categories and $\{\pi_1, \pi_2, \dots, \pi_C\}$ denotes the response probabilities, satisfying $\sum_{c=1}^C \pi_c = 1$.

With n independent observations, the probability distribution for the number of outcomes of the C types is multinomial. This distribution defines the probability for n observations into C categories. Multinomial logit models simultaneously use all pairs of categories by specifying the odds of outcome in one category instead of another (Agresti 2002). Logit models pair each response category with the baseline category. The model can be expressed as in equation (3):

$$\log \frac{\pi_c(x)}{\pi_1(x)} = \alpha_c + \beta_c x, \quad c = 2, \dots, C \tag{3}$$

The model has $C-1$ equations, with separate parameters for each. The effects vary according to the category paired with the baseline. If $C=2$, then this model has a single equation, reducing to ordinary logistic regression for binary responses.

4. Multilevel Multinomial Logistic Regression Model

Multilevel multinomial logistic regression models (MM-LRM) are developed for data sets which have a nested structure. These models are also known as mixed-effects multinomial logistic regression models or multilevel logistic regression models for polytomous data (Hedeker 2003; Skrondal-Rabe-Hesketh 2003).

For the terminology of multilevel analysis, let i denote the level-1 units (individuals) and j denote the level-2 units (clusters). Suppose that there are $j=1, \dots, N$ level-2 units and $i=1, \dots, n_j$ level-1 units nested within each level-2 unit. Thus the

total number of level-1 units across level-2 units is $n = \sum_{j=1}^N n_j$.

If the nominal response variable has c categories, the multilevel multinomial logit model can be defined in terms of a mixed Generalized Linear Model (Grilli-Rampichini, 2007):

$$\eta_{ij}^{(c)} = \alpha^{(c)} + \beta^{(c)'} x_{ij} + \xi_j^{(c)} + \delta_{ij}^{(c)}, \quad c = 2, \dots, C \tag{4}$$

It should be noted that there are no category-specific explanatory variables in equation (4), although this is possible. Each equation in this model may have a different intercept ($\alpha^{(c)}$) and regression coefficients ($\beta^{(c)}$). Also ξ_j and δ_{ij} are vectors of random error terms which show unobserved heterogeneity at the cluster and individual level, respectively. We assume the errors are distributed normally ($\xi_j \sim N(0, \Sigma_\xi)$ and $\delta_{ij} \sim N(0, \Sigma_\delta)$), the errors for different levels are assumed to be independent from each other.

The multinomial logit link is defined as in equation (5):

$$P(y_{ij} = c \mid x_{ij}, \xi_j, \delta_{ij}) = \frac{\exp(\eta_{ij}^{(c)})}{1 + \sum_{c=2}^C \exp(\eta_{ij}^{(c)})} \tag{5}$$

We consider the response variable y_{ij} to follow a multinomial distribution spanning the set of categories $c = 2, \dots, C$. We use $c = 1$ as the baseline category for which all the parameters and the random error terms are set to zero. Thus, the conditional

probability of $y_{ij} = 1$ is $\left[1 + \sum_{c=2}^C \exp(\eta_{ij}^{(c)}) \right]^{-1}$.

The likelihood function of the multilevel multinomial logistic regression model is given in equation (6):

$$L(\theta) = \prod_{j=1}^N \int \prod_{i=1}^{n_j} \left\{ \int P(y_{ij} / x_{ij}, \xi_j, \delta_{ij}) f(\delta_{ij}) d\delta_{ij} \right\} f(\xi_j) d\xi_j \quad (6)$$

where $\theta' = (\alpha^{(2)}, \dots, \alpha^{(C)}, \beta^{(2)}, \dots, \beta^{(C)}, \Sigma_\xi, \Sigma_\delta)$. We must use integral approximations to maximize the likelihood, since the integrals do not have closed-form solutions (Grilli-Rampichini 2007). Thus, several methods are proposed and implemented in various software packages for the estimation of these models. But the most frequently used methods are Marginal quasi-likelihood (MQL), Predictive or Penalized quasi-likelihood (PQL) and Full Information Maximum Likelihood (FIML). MQL involves expansion around the fixed part of the model and tends to underestimate the values of both the fixed and random parameters. PQL involves expansion around both the fixed and random part of the model and is more accurate than MQL but computationally less stable (Hedeker, 2008; Pickery-Loosveldt 2002). PQL and MQL are used in MLwiN (Rasbash et al. 2005). FIML uses Gauss-Hermite quadrature for the approximation of the likelihood function's integral and is available in Supermix, SAS PROC NL MIXED (SAS/Stat 2004), Stata (StataCorp 2005), LIMDEP (Greene 2002) etc.

5. Gifi

Gifi is a transformation proposed by Albert Gifi (1989). For a data set which is a combination of continuous and categorical variables (called mixed), the Gifi transformation converts the non-linear categorical variables to a linear scale. Once the non-linear variables are transformed to a linear scale, several classical multivariate techniques can be applied to the transformed continuous data.

Although Albert Gifi wrote a book about the Gifi transformation, it did not receive much interest for a long time. Michailidis and de Leeuw (1996) applied this transformation on a pure categorical data set and then used the classical multivariate techniques on the transformed scale to determine the patterns in the data set. Following this study, Suman Katragadda (2009) used the Gifi transformation in a mixed data set which is more complex than a pure categorical data set. After implementing the transformation, the data set was composed of only continuous

variables. Thus, he applied classical multivariate techniques in the transformed continuous space and identified useful patterns.

As a brief overview of the Gifi transformation, suppose we have m categorical variables and denote these variables as $h_j, j=1, \dots, m$. Each variable is assumed to have k_j categories. Suppose that there are n observations obtained from these m variables. As a result, an $n \times m$ dimensional information matrix H can be defined. Since the transformation process will lead to some information loss, this loss is expressed in a loss function (Gifi 1989):

$$\sigma(X; Y_1, \dots, Y_m) = m^{-1} \sum_{j=1}^m SS(X - G_j Y_j) = m^{-1} \text{tr} \left[(X - G_j Y_j)' (X - G_j Y_j) \right] \quad (7)$$

In equation (7), SS is the sum of squares of the H matrix. For each categorical variable, k_j dummy variables can be composed. Thus $G_j(i, t) = 0$ or 1 can be defined as a $G = [G_1, \dots, G_m]$ vector with $n \times \sum k_j$ dimensions.

The lost function given in equation (7) is the heart of the Gifi system (Michailidis-de Leeuw 1996). The goal is to minimize the function simultaneously over the X and Y_j 's. In this minimizing problem, several restrictions can be imposed. In order to avoid improper solutions corresponding to $X = 0$ and $Y_j = 0$, Gifi (1989) imposed the restrictions given in equation (8) and equation (9).

$$X'X = nI_p \quad (8)$$

$$u'X = 0 \quad (9)$$

In equation (9), u is a $p \times 1$ dimensional vector consisting of all 1's. The first restriction given in equation (8) standardizes the squared length of the observed scores to be equal to n and in addition for two and more dimensions, it requires the columns to be orthogonal. The second restriction given in equation (9) requires the graph plot to be centered around the origin. We can use the Alternating Least Squares algorithm to minimize the loss function.

6. Application

In the twenty-first century, the considerable part of the health services will contain studies about reducing both the extensiveness and the volume of a group of diseases starting cardiovascular system diseases, respiratory diseases, cancer, diabetes and tooth diseases. In this study dental care status is examined in particular.

The data set used in the application is taken from a real data set which the survey was carried out in Istanbul University Cerrahpaşa School of Medicine, Turkey in 2003. The aim of the survey was to examine the dental health of adults. To carry out the research, the target group of the survey was selected as people who

are over the age of 18, living in Istanbul. According to the Turkish Statistical Institute indicators, Istanbul is divided into three regions. In these regions, there are 25 towns in total and all of them are included. Our main interest is on Dental Care Status (DCS) which is measured in a nominal scale from 1 to 6. Here 1 is coded as 'all teeth are present', 2 as 'most teeth are present, no dentures', 3 as 'some dentures are present', 4 as 'all teeth are dentures', 5 as 'neither teeth nor dentures are present' and 6 as 'other'.

The data set includes 1000 individuals, but the analyses are performed with non-missing 931 individuals. The data set has a nested structure: 931 individuals (1st-level) are nested within 285 districts (2nd-level) which are nested within 25 towns (3rd-level) and 3 regions (4th-level) of Istanbul. Since there are so many variables to predict DCS in the survey, as a pre-analysis factor analysis is used for the purpose of data reduction. For the prediction of DCS, explanatory variables gender, age, tooth brushing, me, doctor, chance and environment are used. The last four explanatory variables are composed using factor analysis results. The factor 'me' can be defined as the individual considers herself/himself responsible, 'doctor' as the individual considers the doctor responsible for his/her dental care, 'chance' as the individual thinks the dental care status of his/her is like that by chance and 'environment' as the individual thinks the environment is responsible for his/her dental care. Gender is coded as zero for women and one for men. Tooth brushing variable has a ordinal scale from 1 to 7. For example 1 encodes 'I brush my teeth once a day' and 7 codes 'I never brush my teeth'.

Since we have a nested data structure and our response variable is measured on a nominal scale, the first part of the application is about multilevel multinomial logistic regression models. The application is performed with the Supermix software. To begin with modelling, first of all we have to check that the data set really has a 4-level data structure.

To answer this question two models are composed: a 4-level and a single-level multinomial logistic regression model. Since these models are nested, the deviance statistics is used for comparison. Of course, here, 'nested' indicates that a specific model can be derived from a more general model by removing parameters. For nested models, the difference in the deviances has a chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated in the two models. The deviance test can be used to perform a formal chi-square test, in order to test whether the more general model fits significantly better than the simpler model (Hox 2002). The results suggested that the 4-level multinomial logistic regression model is statistically significant compared to the single-level model ($p < 0.001$). Thus, it is sensible to go on with the multilevel models. Considering that we want a good, but at the same time parsimonious model, we thought to reduce this 4-level model to a 3-level model by including an explanatory variable describing region instead of a level. Besides, it should be kept in mind that four regions is a very low number to have variation for DCS. The comparison of these models

suggested that a 3-level model including the explanatory variable describing the region is better than the 4-level multinomial logistic regression model ($p < 0.001$). Next, the first-level explanatory variables are added to the model and the results can be seen in Table 1.

Table 1. Three-level multinomial logistic regression model including the first-level explanatory variables

Term	1 vs. 2		1 vs. 3		1 vs. 4		1 vs. 5		1 vs. 6	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept	-2,6312	0,3779	-4,593	0,4629	-11,4619	1,4688	-15,6171	3,8366	-4,8639	0,7164
Region 2	-0,2296	0,3038	-0,2734	0,2823	-0,1943	0,5447	-1,1929	1,5509	-0,6153	0,4758
Region 3	-0,0203	0,2512	-0,4276	0,2897	-1,0607	0,6419	-0,1751	1,2643	0,5513	0,5501
Gender	-0,0081	0,1874	-0,3259	0,2216	-0,8629	0,4623	-1,2613	1,1103	-0,0535	0,3381
Age	0,061	0,0088	0,1147	0,0102	0,2097	0,0229	0,2133	0,0507	0,041	0,0149
Tooth brushing	0,3177	0,0727	0,1761	0,0848	0,3505	0,1356	0,8159	0,2508	0,3419	0,1147
Me	0,0595	0,0999	0,237	0,1169	0,006	0,2162	0,4997	0,6008	0,6339	0,2267
Doctor	-0,0837	0,096	0,2176	0,1143	0,1221	0,2378	1,3671	0,6976	0,2582	0,1628
Chance	0,0837	0,0974	0,1243	0,1136	0,0089	0,2274	0,0531	0,5255	-0,1869	0,1909
Environment	0,22	0,1003	0,3579	0,1115	0,576	0,2407	-0,3497	0,5001	-0,0211	0,1545
Level-2 variances	0,0425	0,1027	0,0717	0,1669	1,1376	1,0861	1,1506	2,8945	0,435	0,4306
Level-3 variances	0,3654	0,3834	0,1646	0,1354	0,0742	0,1574	0,8209	1,7352	2,616	0,7196
-2log L	2096,04									
AIC	2216,04									

bold indicates $p < 0.05$ *italic* indicates $0.05 < p < 0.10$

Source: own creation

Looking at Table 1, the estimates of the model suggest that dummy-coded region variables, gender and doctor variables are non-significant for all five equations. After removing these variables, the final model is obtained and the estimates can be seen in Table 2. What we can see from the final model is we have a two-level multinomial logistic regression model where individuals are nested within towns. This means after removing the non-significant variables from the model, the district level also became non-significant. At the end of the table, the ICC values, which indicated how much of the variation of DCS lies within the district level, are also given. Since all of them are higher than 0.05, it can be considered enough for multilevel models (Muthén & Satorra, 1995).

Table 2. Final Model: Two-level multinomial logistic regression model including the significant first-level explanatory variables

Term	1 vs 2		1 vs 3		1 vs 4		1 vs 5		1 vs 6	
	Estimate	S.E.								
Intercept	-2.7117	0.3293	-4.8958	0.4044	-10.926	0.9257	-14.816	2.3016	-4.6028	0.5531
Age	0.0626	0.0088	0.1138	0.0099	0.1932	0.0173	0.1925	0.0352	0.0439	0.014
Tooth brushing	0.3144	0.0709	0.1596	0.0818	0.2897	0.1197	0.7327	0.2257	0.309	0.1079
Me	0.0775	0.0939	0.2968	0.1105	0.0274	0.1823	0.5898	0.5422	0.6149	0.2079
Doctor	-0.0542	0.0922	0.2003	0.1073	0.0999	0.2158	1.3055	0.8265	0.3024	0.1488
Environment	0.2118	0.0981	0.3324	0.1075	0.4862	0.2108	-0.2719	0.4254	0.0085	0.1438
Level-2 variances	0.4892	0.3095	0.1804	0.124	0.1855	0.2113	1.0641	1.9177	2.2127	0.4788
-2log L	2130.14									
AIC	2200.14									
ICC	0.129		0.052		0.053		0.244		0.402	

Source: own creation

The second part of the application is about using the Gifi transformation. The Gifi transformation is employed for the response variable DCS and the explanatory variable tooth brushing which has an ordinal scale. Since the response variable becomes continuous after the transformation, an ordinary multilevel regression modelling approach is used to predict DCS. Hox’s (2002) 5-step modelling approach is used throughout modelling the multilevel regression model. According to this 5-step approach, a four-level random-intercept model (M1) is composed and compared with a single-level random-intercept model (M2) as a first step in order to check if multilevel modelling approach is appropriate. A random-intercept model is a model which has no explanatory variables. The estimates of these models can be found in Table 3.

Table 3. First step comparing several models to identify optimal structure

Model	M1: Four-level random-intercept model		M2: Single-level random-intercept model		M3: Three-level random-intercept model	
Fixed part						
Predictor	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Intercept	-0.010	0.040	0.000	0.026	-0.010	0.040
Random part						
σ_e^2	0.000	0.000	0.632	0.029	0.016	0.011
σ_u^2	0.016	0.011			0.049	0.021
σ_v^2	0.049	0.021			0.573	0.030
σ_f^2	0.573	0.030				
Deviance	2202.867		2214.904		2202.867	

Source: own creation

For the comparison of the 4-level and the single-level models, the difference between the deviances is 12.04 with 3 degrees of freedom following a chi-square distribution and the related significance level $p=0.003<0.01$. The null hypothesis is rejected here, which means that there is some error variance in the 4-level model for regions. So now we are sure that we're dealing with a 4-level multilevel model in which individuals are nested within districts, districts are nested within towns and towns are nested within regions. Next, we estimated a 3-level regression model (M3) and compared it with the 4-level regression model. Since the deviance statistics are exactly the same, the 3-level model, which has one fewer parameters is preferred over the 4-level model.

As a second step, the first-level explanatory variables are added to the model and thus (M4) is composed and the results are given in Table 4, which shows that all the factors (me, doctor, chance and environment) are non-significant. After removing them, the the model is re-estimated and (M5), which is the final model, is obtained.

Table 4: Second step comparing several models to identify optimal structure

Model	M4: + First-level predictors		M5: The final model	
Fixed part				
Predictor	Coefficient	S.E.	Coefficient	S.E.
Intercept	0.498	0.081	0.500	0.080
Gender	0.064	0.050	0.064	0.049
Age	-0.015	0.002	-0.015	0.002
Tooth Brushing	0.213	0.031	0.211	0.031
Me	-0.021	0.027		
Doctor	-0.027	0.026		
Chance	0.012	0.026		
Environment	0.019	0.025		
Random part				
σ_e^2	0.505	0.026	0.500	0.026
σ_u^2	0.032	0.017	0.037	0.017
σ_v^2	0.010	0.008	0.015	0.010
Deviance	2066.693		2068.781	

Source: own creation

The third step would be to add higher-level explanatory variables. Since there wasn't any higher-level explanatory variables, the first-level variables are aggregated to the second level but seen that none of them were significant. The fourth step, which is assessing whether any of the slopes of explanatory variables have a significant variance component between the groups, is also evaluated. Again, there wasn't any significant results. Since we couldn't find any significant variable which has a

random-slope, in order to explain this random slope, adding cross-level interactions would be final step. Thus, this step is passed.

After finding the final models for both multilevel multinomial logistic regression and ordinary multilevel regression models, Akaike Information Criterion (AIC) is used for comparison. The AIC for the two-level multinomial logistic regression model is 2200,14. For the ordinary three-level regression model, the AIC is calculated with the formula $AIC = Deviance + 2p$ where p is the number of estimated parameter in the model and is calculated as 2082,78. Since the lower value of AIC is better, we can conclude that the ordinary three-level regression model is preferred.

7. Discussion

If the economical conditions of the investment of treatment services are taken into consideration, the minimization of the health budgets is noteworthy. One of the most important parts of the health budgets is left to dental health. From this point of view, dental care status is explored. For this reason, a multilevel multinomial logistic regression and an ordinary multilevel regression model through Gifi transformation is examined for dental care status and compared. In the multilevel multinomial logistic regression model, we ended up with a two-level model in which individuals are nested within towns. And for the prediction of DCS, the final model includes age, tooth brushing, me, doctor and environment explanatory variables. The most fundamental variable of the model is ‘tooth brushing’. This means, the lower the frequency of tooth brushing, the worse the status of dental care is, which is reasonable. In the ordinary multilevel regression model, we ended up with a three-level model in which individuals are nested within districts and districts are nested within towns. Here, the final model includes gender, age and tooth brushing explanatory variables. And the most important variable of this models seems to be gender. With regards to AIC, the ordinal three-level regression model is selected as a final model over the two-level multinomial logistic regression model.

References

- Agresti, A. 2007: *Introduction to Categorical Data Analysis*. Wiley, New Jersey.
Agresti, A. 2002: *Categorical Data Analysis*. Wiley, New Jersey.
Gifi, A. 1989: *Nonlinear Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics, New Jersey.
Greene, W. H. 2002: *LIMDEP Version 8.0 User's Manual. 4th edition*, Econometric Software, Plainview, New York.

- Grilli, L. – Rampichini, C. 2007: A Multilevel Multinomial Logit Model for the Analysis of Graduates' Skills. *Stat. Methods&Applications*, 16, 381-393. p.
- Hedeker, D. 2008: Multilevel models for ordinal and nominal variables. In J. de Leeuw & E. Meijer (Eds.): *Handbook of Multilevel Analysis*. Springer, New York.
- Hedeker, D. 2003: A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433-1446. p.
- Hosmer, D. W. - Lemeshow, S. 2000: *Applied Logistic Regression*. John Wiley & Sons, Canada.
- Hox, J. 2002: *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, New Jersey.
- Katragadda, S. 2008: Multivariate Mixed Data Mining with Gifi System using Genetic Algorithm and Information Complexity. *Unpublished Doctorate Thesis*, University of Tennessee, Knoxville.
- Michailidis, G. - de Leeuw, J. 1996: *The Gifi System of Descriptive Multivariate Analysis*, UCLA Statistical Series Preprints 204. Univ. California, Los Angeles.
- Muthén, B. O. - Satorra, A. 1995: Complex sample data in structural equation modeling. In P.V. Marsden (Ed.): *Sociological methodology*. Blackwell Publishing, Oxford, 267-316. p.
- Pickery, J. – Loosveldt, G. 2002: A Multilevel Multinomial Analysis of Interviewer Effects on Various Components on Unit Nonresponse. *Quality&Quantity*, 36, 427-437. p.
- Rasbash, J. - Charlton, C. - Browne, W.J. - Healy, M. - Cameron, B. 2009: *MLwiN Version 2.1*. Centre for Multilevel Modelling, University of Bristol.
- SAS/Stat, 2004: *SAS/Stat User's Guide. version 9.1*, SAS Institute, Cary, NC.
- Skrondal, A. – Rabe-Hesketh, S. 2003: Multilevel Logistic Regression for Polytomous Data and Rankings. *Psychometrika*, 68, 267-287. p.
- StataCorp, 2005: *Stata Statistical Software: Release 9*. College Station, TX.
- Supermix, <http://www.ssicentral.com/supermix/index.html>