Volume: 09 Number: 01 Category: 02 Page: 53-64 ISSN No: 1303-6319 Cilt: 09 Sayı: 01 Kategori: 02 Sayfa: 53-64

COVARIATES OF UNIT NONRESPONSE ERROR BASED ON PROXY RESPONSE FROM HOUSEHOLD SURVEYS

A. Sinan TÜRKYILMAZ* **H. Öztaş AYHAN****

ABSTRACT

Unit nonresponse error and its related covariates are examined from the results of a sample survey. A procedure is proposed to study unit nonresponse when data are from a two stage household sample survey in which household are the units of the first level and individuals are the units of second level. The individual person responses within the sample survey did not contain information on the nonrespondents. Therefore, household schedule variables which are based on proxy person response information are combined with the binary dependent response/nonresponse variable from the individual survey records. The idea is to estimate a logistic model whose dependent variable is the binary unit response indicator and where individual characteristics at the right hand side are approximated by household information collected at the first level. Among other models, a binary logistic regression model is proposed and the results are analyzed and interpreted by the computed odds ratios. The results have indicated several significant covariates for the model of nonresponse.

Keywords: Binary dependent variable, Covariates of nonresponse, Logistic regression, Nonresponse error components, Proxy response.

1. INTRODUCTION

Unit nonresponse is the failure to obtain the minimum required information from an eligible housing unit or person in the sample. Unit nonresponse occurs when the responsents are unable or unwilling to participate; interviewers are unable to locate addresses or respondents, or when other barriers exist for completing the interview.

Covariates of unit nonresponse error have been a concern of survey researchers as a major part of the total survey error. Components of unit nonresponse error are basically associated with the factors related to the reasons of survey non-participation.

In order to have logical causality measures, one has to identify the direct and indirect factors affecting such relations. In many cases, information on such ideal factors may not be available as a survey variable, due to the current objectives of such a survey. Alternative information can be derived from the other existing survey variables which are naturally available due to the survey objective. Consequently, the researchers have to make sense out of such information, because the ideal information which will explain the causality may not be available.

*Assoc. Prof., Hacettepe University, Institute of Population Studies, 06100 Ankara, e-mail: aturkyil@hacettepe.edu.tr
**Prof., Middle East Technical University, Department of Statistics, 06800 Ankara, e-mail: oayhan@metu.edu.tr

With a limited research budget, one can obtain information only on a reasonably small scale. On the other hand, for a large scale survey, additional questions will also bring extra cost, which may not be tolerable by the survey management. Under the circumstances, another alternative may be to utilize the best of the available information.

The examination of the components of unit nonresponse in a demographic survey have been given by Ayhan (1981), and some of the other recent studies have also been evaluated (Ayhan, 1998). The current study examines the issue by taking an alternative approach. The following sections of this paper cover the methodology used, covariates of nonresponse, proposed models and testing, and the conclusions of the findings from this investigation.

2. SURVEY METHODOLOGY

2.1. Sample Design and Implementation

The sample design and sample size of the *Turkey Demographic and Health Survey* (TDHS) – 2003 (HUIPS, 2004) make it possible to perform analyses for Turkey as a whole, for urban and rural areas and for the five demographic regions of the country. A weighted, multistage, stratified cluster sampling approach was used in the selection of the survey sample. The results of the household and individual questionnaire executions are summarized in Table 1.

Table 1. Results of the household and individual interviews in 2003 Turkey Demographic and Health Survey

Outcomes	Urban	Rural	Total
Household interviews:			
Selected sample households	8718	2941	11659
Households interviewed	7956	2880	10836
Household Nonresponse Rate (*HHRR*)	0.087	0.021	0.071
Individual interviews:			
Eligible women selected	6259	2188	8447
Eligible women interviewed	5976	2099	8075
Individual Nonresponse Rate Component (*IRRC*)	0.045	0.041	0.044
Individual Person Nonresponse Rate (*IPNRR*)*	0.128	0.061	0.112

* Computation of the *IPNRR* = [*1 – HHRR * IRRC*]

The target sample size of the TDHS–2003 was set at 13160 dwelling units. This was expected to yield about 11000 completed household interviews. Out of 11659 selected sample households, 10836 number of households were interviewed. Within this, 8447 number of eligible women was present and 8075 was interviewed during the survey operation. Information is provided on the overall coverage of the sample, including household and individual nonresponse rates.

2.2. Questionnaire Design

The data collection for household sample surveys have been executed in two stages; the completion of the household schedule, and the individual survey. The *household schedule* is completed by a selected adult member of the household, as a proxy respondent for the other members of the household, and a self respondent for him/herself.

For the *individual survey*, data are only collected from the eligible women as a self respondent, and no information is available for the non-respondents. On the other hand, household schedule also contains some more additional information about other characteristics of the respondents and non-respondents of the individual survey.

For the responding households, generally the household schedule contains full information on all household members. On the other hand, the selected household member for the individual survey may or may not respond to the individual person's interview. Consequently, we will have two possible groups for the individual survey; respondents and non-respondents.

This study combines the household based proxy information for selected variables, and response-nonresponse outcome information of the individual person's survey from the same household.

3. COVARIATES OF NONRESPONSE

The following household information is obtained from the household schedule by proxy interviews;

A. Independent survey variables: (*Based on household survey information*)

1. ***Stratification variables used as survey variables***:
    - Region
    - Type of place of residence
2. ***Household based proxy individual variables***:
    - Gender
    - Age groups
    - Place of birth
    - Maternal and paternal survival
    - Migration and mobility
    - Literacy and education status
    - Work status
    - Marital status

3. *Housing characteristics*:

- Household ownership
- Safe water access
- Sanitary toilet
- Number of rooms
- Household durability
- Household facilities
- Household income

B. Dependent survey variable: (*Based on individual survey information*)

- Binary nonresponse information

Some of the household based current and generated variables, their response options, and their frequencies are given in Table 2.

4. PROPOSED MODELS AND TESTING

4.1. Search for Models

In the literature, multinomial logistic regression models are grouped into two distinct types as generalized and cumulative logit models. Generalized logit models are usually employed when the response categories are unordered whereas cumulative logit models should be employed when response categories are ordered. Both classical and Bayesian methodologies are available to estimate the model parameters. Moreover, multinomial logistic regression models are developed to analyze categorical response data occuring in matched case-control studies.

For the analysis of data occuring in matched case-control studies, conditional logistic regression likelihood functions are developed to adjust the analysis for the nuisance parameters that are of high dimension. There is a vast literature on multinomial logistic regression models and analysis. For instance Hosmer and Lemeshow (2000) and Agresti (2002) provide the basics, extensions, as well as related special topics including logistic regression analysis for correlated data.

Besides well established multinomial logistic regression models, novel developments emerged in recent years motivated by categorical response data with interesting features that occur especially in epidemiological studies. Of the recent developments, Chatterjee (2004) developed a two stage multinomial logistic regression approach to analyze data with multivariate classification information and derived the asymptotic properties of the test statistics.

Table 2. Current and generated variables, options and their frequencies

Name of variables	Code and Explanation	Weighted percent
Response and Nonresponse	1 Nonresponse	4.7
	0 Response	95.3
hv017- Number of visits to household	1	79.7
	2	14.9
	3	5.4
v024 – Regions	1 West	40.7
	2 South	12.7
	3 Central	23.1
	4 North	7.3
	5 East	16.2
hv025 - Type of place of residence	1 Urban	71.2
	2 Rural	28.8
hv270 - Wealth index	1 Poorest	15.6
	2 Poorer	18.1
	3 Middle	20.2
	4 Richer	22.4
	5 Richest	23.6
hv102 - Usual resident	0 No	3.6
	1 Yes	96.4
sh26 - Currently working	0 No	75.1
	1 Yes	24.9
SANITATE- Sanitary toilet	0 No	90.7
	1 Yes	9.3
SAFEWAT – Safewater	0 No	92.4
	1 Yes	7.6
CROWD – Number of persons per room	0 less than 3	80.5
	1 more than 3 and over	19.5
Educ – Education level	1 No education / Primary incomplete	22.1
	2 Primary complete/ secondary incomplete	60.7
	3 Secondary +	17.2
hv116 - Marital status	1 Currently married	94.7
	2 Formerly / ever married	5.3
agegroup – Age groups	1 15-19	3.0
	2 20-24	12.9
	3 25-29	18.2
	4 30-34	18.3
	5 35-39	17.5
	6 40-44	16.5
	7 45-49	13.5

In this study, individual survey respondent's related household schedule characteristics are used as possible covariates for the non-response error. The possible covariates are evaluated under several alternative statistical models. For this purpose, several generalized linear models have been examined. As possible alternatives, *loglinear model, logit model, probit model,* and *logistic regression model*s have been evaluated.

After the examination of the current available variables, *multiple logistic regression model* has been selected. Summary measures of goodness-of-fit are provided as output with any fitted model and give an overall indication of the fit of the model (Hosmer and Lemeshow, 1980, and Lemeshow and Hosmer, 1982).

The present model takes non-response as the binary dependent variable which is associated with the other household covariates. In order to test our model, the latest TDHS – 2003 data is used. Questions and topics which are listed in Section 3 were asked during the household interviews. The household survey and individual person's survey data sets are combined under the weighted, stratified cluster design, for the survey analysis. The *SPSS 13.0's "complex samples" feature* were used to perform *binary logistic regression,* where the sample design was naturally taken into account.

4.2. Inferences from Binary Logistic Regression

A binary logistic regression model has been proposed to explain the effect of covariates on survey unit nonresponse for this study. After the regression diagnostics, such as outlier detection and collinearity tests were performed the following model and results were obtained. Some variables were not taken into account, such as work type, since only a portion of women are working. Moreover, only variables available for "all cases" were included to increase the number of cases in model.

The hypothesis to be tested is

$$H_0 : \beta_i = 0 \quad \text{versus} \quad H_a : \beta_i \neq 0 .$$

The binary logistic regression prediction equation for an *S*–shaped curve for the desired probability *p* is

$$p = \exp\left(\hat{\alpha} + \sum_{i=1}^{k} \hat{\beta}_i x_i\right) \Big/ \left[1 + \exp\left(\hat{\alpha} + \sum_{i=1}^{k} \hat{\beta}_i x_i\right)\right] . \qquad (1)$$

Within the *S*–shaped regression model, the probability *p* falls between 0 and 1 for all possible *x* values. Test statistics for the regression model coefficients are

$$t_i = \left(\hat{\beta}_i - \beta_i\right) \big/ se\,(\hat{\beta}_i) . \qquad (2)$$

4.3. The Odds Ratio

The odds ratio ($\boldsymbol{\theta}$) is a measure of association which has found wide use in many disciplines. It approximates how much more likely (or unlikely) it is for the outcome to be present among those with $x = 1$ than among those with $x = 0$ (Lemeshow and Hosmer, 1983). The odds ratio is usually the parameter of interest in a logistic regression due to its ease of interpretation. The interpretation given for the odds ratio is based on the fact that in many instances it approximates a quantity called the *relative risk* (Hosmer and Lemeshow, 2000). Along with the point estimate of a parameter, it is a good idea to use a confidence interval estimate to provide additional information about the parameter value.

The odds ratio is used to interpret the computed coefficients of the binary logistic regression prediction equation, in terms of relative comparative risks. The data layout structure of the odds related variables are given in Table 3, below.

Table 3. The data layout structure for odds

Variables	Nonresponse	Response	Total
Variable *option A*	n_{11}	n_{12}	n_{1+}
Variable *option* A^c	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

The desired (success) probabilities for the two groups are;

π_1 is estimated by $p_1 = n_{11}/n_{1+}$,

π_2 is estimated by $p_2 = n_{21}/n_{2+}$.

In 2×2 contingency tables, the *relative risk* is the ratio of the desired probabilities for the two groups.

$$\text{The Relative Risk} = \pi_1/\pi_2 \tag{3}$$

The ratio of odds from two rows is given by

$$\theta = \frac{\pi_1(1-\pi_1)}{\pi_2(1-\pi_2)} = \frac{\pi_{11}\ \pi_{22}}{\pi_{12}\ \pi_{21}} . \tag{4}$$

Sample odds (cross–product) ratio is

$$\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}\ n_{22}}{n_{12}\ n_{21}} . \tag{5}$$

The odds ratio can be equal to any nonnegative number.

The odds ratio can be interpreted as;

(1) When $1 < \theta < \infty$, the odds of success are higher in row 1 than in row 2.

(2) When X and Y are independent, $\pi_1 = \pi_2$, so that

$$\theta = [\text{odds}_1 / \text{odds}_2] = 1 .$$

(3) When $0 < \theta < 1$, a success is likely in row 1 than in row 2, that is $\pi_1 < \pi_2$.

Generalized linear models yield fitted coefficients that are commonly used to estimate odds ratio or other measures of association. Standard fitting techniques such as maximum likelihood and estimating equation methods yield consistent estimators with first order asymptotically normal sampling distributions (Cox and Oakes 1984; Agresti 2002; Lyles, Guo and Greenland 2012).

Recently, Allen and Le (2008) introduced the overall odds ratio (OOR) as a new index for quantifying the overall effect size in logistic regression models. The OOR can be interpreted in the same way as the odds ratio of individual independent variables. It is the ratio of the odds of belonging to a category of the dependent variable that a researcher is interested in predicting when the weighted linear combination of the independent variables increases one standard deviation to the odds before such an increase (Le and Marcus 2012).

4.4. Model Based Survey Statistics and Outcomes

Once we have fit a particular multiple (multivariable) logistic regression model, we begin the process of model assessment. The first step in this process is usually to assess the significance of the variables in the model. The likelihood ratio test for overall significance of the p coefficients for the independent variables in the model is performed in exactly the same manner as in the univariate case (Hosmer and Lemeshow, 2000).

Before concluding that any or all of the coefficients are nonzero, we may wish to look at the univariate Wald test statistics. Under the hypothesis that an individual coefficient is zero, these statistics follow the standard normal distribution. In order to obtain the best fitting model while minimizing the number of parameters, the next logical step is to fit a reduced model containing only those variables thought to be significant, and compare it to the full model containing all the variables (Hosmer and Lemeshow, 2000).

The following proposed model is fitted to the TDHS 2003 data.

$$p = \Pr(Y=1) = \exp\left(\hat{\alpha} + \sum_{i=1}^{k} \hat{\beta}_i x_i\right) \Big/ \left[1 + \exp\left(\hat{\alpha} + \sum_{i=1}^{k} \hat{\beta}_i x_i\right)\right] \quad \text{where,} \tag{6}$$

$$\hat{\alpha} + \sum_{i=1}^{k} \hat{\beta}_i x_i = -1.615 + 0.563{*}\text{hv024(1)} + 0.549{*}\text{hv024(2)} + 0.470{*}\text{hv024(3)} + 1.577{*}\text{hv102(0)} - 0.451{*}\text{sh26(0)} - 0.656{*}\text{hv116(1)} - 0.557{*}\text{agegroup(2)} - 0.433{*}\text{agegroup(3)} - 0.469{*}\text{agegroup(4)} - 0.448{*}\text{agegroup(5)} \tag{7}$$

Information on the related correlation measures are given in Table 4. The Nagelgerke R–square is used as a pseudo R-square of linear regression and measures the power of model in terms of how the model explains the variation in dependent variables by independent variables.

Table 4. Several pseudo R square values for the model

Test statistics	R–square
Cox and Snell	0.021
Nagelgerke	0.066
McFadden	0.056

The Nagelgerke R–square is 0.066 so the power of the model is low but the model is significant (with a *p*-value of 0.000, and Wald statistics value = 7.289, df 1 = 25, df 2 = 322).

The results of the test statistics for the model effects are presented in Table 5. Within the logistic regression model, “the number of visits”, “region”, “being usual resident”, “currently working”, “educational level” and “marital status” stands as significant independent variables.

Table 5. Results of the test statistics for model effects

Sources	df 1	df 2	Wald F	Significance	Indicator
(Corrected model)	25	322	7.29	0.00	*
(Intercept)	1	346	54.61	0.00	*
hv017 - Number of visits	2	345	3.12	0.05	*
hv024 – Region	4	343	2.63	0.03	*
hv025 - Type of place of residence	1	346	0.97	0.33	
hv270 - Wealth index	4	343	1.03	0.39	
hv102 - Usual resident	1	346	63.59	0.00	*
sh26 - Currently working	1	346	7.28	0.01	*
SANITATE - Sanitary toilet	1	346	1.09	0.30	
SAFEWAT - Safewater	1	346	0.00	0.96	
CROWD - No of persons per room	1	346	0.30	0.58	
Educ - Education level	2	345	5.43	0.00	*
hv116 – Marital status	1	346	10.35	0.00	*
Age groups	6	341	1.88	0.08	

Finally, the model parameter estimates of the binary logistic regression model are given in detail in Table 6.

Table 6. Binary logistic regression model parameter estimates

Variables	Category	$\hat{\beta}_i$	$se(\hat{\beta}_i)$	t_i	*df*	*p-value*	*deff*	$\hat{\theta}$	Indicator
Intercept		-1.615	0.560	-2.885	346	0.00	1.54	0.20	*
hv017- Number of visits	1	-0.284	0.282	-1.004	346	0.32	1.69	0.75	
	2	0.192	0.296	0.650	346	0.52	1.74	1.21	
	3	0	.	.	.	.	.	1.00	
hv024 – Region	1 West	0.563	0.201	2.803	346	0.01	1.11	1.76	*
	2 South	0.549	0.238	2.309	346	0.02	1.21	1.73	*
	3 Central	0.470	0.224	2.098	346	0.04	1.19	1.60	*
	4 North	0.190	0.284	0.671	346	0.50	1.01	1.21	
	5 East	0	.	.	.	.	.	1.00	
hv025 - Type of place of residence	1 Urban	0.170	0.173	0.983	346	0.33	1.43	1.19	
	2 Rural	0	.	.	.	.	.	1.00	
hv270 - Wealth index	1 Poorest	-0.238	0.277	-0.859	346	0.39	1.76	0.79	
	2 Poorer	-0.358	0.206	-1.735	346	0.08	1.24	0.70	
	3 Middle	-0.264	0.210	-1.258	346	0.21	1.50	0.77	
	4 Richer	-0.343	0.197	-1.739	346	0.08	1.49	0.71	
	5 Richest	0	.	.	.	.	.	1.00	
hv102 - Usual resident	0 No	1.577	0.198	7.974	346	0.00	1.30	4.84	*
	1 Yes	0	.	.	.	.	.	1.00	
sh26 - Currently working	0 No	-0.451	0.167	-2.699	346	0.01	1.83	0.64	*
	1 Yes	0	.	.	.	.	.	1.00	
SANITATE- Sanitary toilet	0 No	-0.280	0.268	-1.042	346	0.30	1.69	0.76	
	1 Yes	0	.	.	.	.	.	1.00	
SAFEWAT - Safewater	0 No	-0.011	0.243	-0.045	346	0.96	1.53	0.99	
	1 Yes	0	.	.	.	.	.	1.00	
CROWD – no of persons per room	0 less than 3	-0.114	0.208	-0.548	346	0.58	1.68	0.89	
	1 more than 3 and over	0	.	.	.	.	.	1.00	
Educ – education level	1 No education/ Primary incomplete	0.335	0.245	1.366	346	0.17	1.58	1.40	
	2 Primary complete/ secondary incomplete	-0.198	0.178	-1.114	346	0.27	1.42	0.82	
	3 Secondary +	0	.	.	.	.	.	1.00	
hv116 - marital status	1 Currently married	-0.656	0.204	-3.217	346	0.00	1.25	0.52	*
	2 Formerly/ ever married	0	.	.	.	.	.	1.00	
Age Group	1 15-19	0.136	0.369	0.368	346	0.71	1.63	1.15	
	2 20-24	-0.557	0.234	-2.384	346	0.02	1.26	0.57	*
	3 25-29	-0.433	0.192	-2.253	346	0.02	1.15	0.65	*
	4 30-34	-0.469	0.197	-2.374	346	0.02	1.20	0.63	*
	5 35-39	-0.448	0.215	-2.083	346	0.04	1.47	0.64	*
	6 40-44	-0.379	0.216	-1.754	346	0.08	1.55	0.68	
	7 45-49	0	.	.	.	.	.	1.00	

For the coefficients of this model, the following results can be summarized in terms of odds ratios. The probabilities of being “non-responder” women are 1.76, 1.73 and 1.60 times higher for women who are in West, South and Central regions when compared to women in East region. Temporary members of the household are 4.84 times more likely to be “non-responders” than the usual members of the household. Non-working women are 1.56 (=1 / 0.64) times better responders compared to working women. Similarly, currently married women are 2 (=1 / 0.52) times better responders. Excluding the youngest age group of reproductive women aged 15-19, all other age groups are about 1.5 times better responders compared to the oldest age group of 45-49.

5. CONCLUSIONS

Since the number of independent variables is limited to questions asked in the household questionnaire and some of them are not included into the model due to small number of cases, the number of significant independent variables is few. However, as expected, the number of visits, the region where the woman lives are significant and the "East" region of Turkey gives smaller odds value; meaning that the response rates are higher than the other regions. In addition, naturally "being a usual resident" and "currently working" are also significant and usual residents and non-working women are better responders. "Being a currently married women" and "middle age women within the reproductive age groups of 15-49" are also significant.

As it is stated earlier the variables that are included into the regression model are based on proxy information and limited to the information collected by household questionnaire. This model can be thought as an indirect way of examining the covariates of non-response when it is not possible to measure the non-response by a well-defined independent module added to the study and applied to non-responders directly. If the number of proxy information is increased, future models may include more independent variables to the model and the power of model may be higher.

6. REFERENCES

Agresti, A., (2002). Categorical Data Analysis. 2nd edition. John Wiley, Hoboken, NJ.

Allen, J. and Le, H., (2008). An Additional Measure of Overall Effect Size for Logistic Regression Models. Journal of Educational and Behavioral Statistics 33, 416 – 441.

Ayhan, H. Ö., (1981). Sources of Nonresponse Bias in 1978 Turkish Fertility Survey. Turkish Journal of Population Studies 2–3, 104–148.

Ayhan, H. Ö., (1998). Survey Nonresponse Models and Applications in Turkey. In Official Statistics in a Changing World. Stockholm: Statistics Sweden Press, pp. 155–158.

Chatterjee, N., (2004). A Two Stage Regression Model for Epidemiological Studies with Multivariate Disease Classification Data. Journal of the American Statistical Association 99, 127 – 138.

Cox, D. R. and Oakes, D., (1984). Analysis of Survival Data. Chapman and Hall, London.

Hosmer, D. W., and Lemeshow, S., (1980). A Goodness-of-Fit Test for the Multiple Logistic Regression Model. Communications in Statistics A 10, 1043–1069.

Hosmer, D. W., and Lemeshow, S., (2000). Applied Logistic Regression (Second Edition). New York: John Wiley and Sons, Inc.

HUIPS., (2004). Turkey Demographic and Health Survey, 2003. Hacettepe University Institute of Population Studies, Ministry of Health General Directorate of Mother and Child Health and Family Planning, State Planning Organization and European Union. Ankara, Turkey. 309 pp.

Le, H. and Marcus, J., (2012). The Overall Odds Ratio as an Intuitive Effect Size Index for Multiple Logistic Regression: Examination of Further Refinements. Educational and Psychological Measurement 76(6), 1001 – 1014.

Lemeshow, S., and Hosmer, D. W., (1982). The Use of Goodness-of- Fit Statistic in the Development of Logistic Regression Models. American Journal of Epidemiology 115, 92–106.

Lemeshow, S., and Hosmer, D. W., (1983). Estimation of Odds Ratio with Categorical Scaled
Covariates in Multiple Logistic Regression Analysis. American Journal of Epidemiology 119, 147–151.

Lyles, R. H., Guo, Y., and Greenland, S., (2012). Reducing Bias and Mean Squared Error Associated with Regression-Based Odds Ratio Estimators. Journal of Statistical Planning and Inference 142, 3235 – 3241.

HANEHALKI ARAŞTIRMALARINDA YERİNE CEVAPLAYICIDAN ELDE EDİLEN BİRİM CEVAPLANMAMA HATASI ORTAK DEĞİŞKENLERİNİN BİLEŞENLERİ

ÖZET

Birim cevaplanmama hatası ve ortak değişkenlerinin bileşenleri, yapılan bir örneklem araştırmasının sonuçlarına dayanarak incelenmiştir. Birinci aşaması hanehalkı ve ikinci aşaması kişi düzeyinde gerçekleşen iki aşamalı bir çalışmanın verilerde birim cevaplanmama hatasını çalışmak için bir prosedür önerilmiştir. Bu çalışmadaki kişi düzeyinde cevaplanmama ile ilgili bilgiler bulunmamaktadır. Bu nedenle, hanehalkı araştırmasında bulunan seçilmiş değişkenlerle ilgili bilgiler yerine cevaplayıcıdan elde edilmiş ve bu bilgiler aynı kişiye ait olan kişi araştırmasının sonuçlarındaki ikili cevaplama/cevaplamama bağımlı değişkeniyle birleştirilmiştir. Düşünce, cevaplanmama göstergelerini açıklamak için lojistik regresyon modeli geliştirilmesi ve modelin sağ tarafı kişi özelliklerinin ilk aşamada toplanan hanehalkı bilgileriyle yakınsamalarıdır. Diğer modellerin yanında,bir lojistik regresyon önerilmiş ve sonuçlar hesaplanan ihtimaller oranı ile analiz edilmiş ve yorumlanmıştır. Elde edilen sonuçlar,cevaplanmama modelini etkileyen bazı önemli ortak değişkenlerin mevcut olduğunu göstermektedir.

Anahtar Kelimeler: Cevaplanmama hatası bileşenleri, Cevaplanmama ortak değişkenleri, Kesikli bağımlı değişken, Lojistik regresyon, Yerine cevaplama.