Scientific Annals of Economics and Business
70 (1), 2023, 71-81
DOI: 10.47743/saeb-2023-0009

How to Analyze the Association between Two Categorical Variables Based on Census Data with a High Level of Nonresponse

Milan Terek[*], **Eva Muchová**[**], **Peter Leško**[***]

Abstract*:* Statistical surveys are often used in shaping managerial policy and practice. In this paper we study, how to analyze the association between two categorical variables based on census data with a high level of nonresponse. The purpose is to discuss the suggested approach to the investigation. We used the census data from the survey executed at one Slovak University for testing the new process. The proposed process offers the methods of analysis of the association between two categorical variables based on pseudo-population estimated from the census data with a high level of nonresponse. We recommend using the process in the surveys in which the costs of survey execution by the census are practically not different from sample survey costs, and the connections to all units of the population are available.

Keywords: census data; nonresponse; association between two categorical variables; residuals; odds ratio.

JEL classification: C18.

[*] School of Management in Bratislava, Slovak Republic; e-mail: *milan.terek1@gmail.com, mterek@vsm.sk* (corresponding author).

[**] University of Economics in Bratislava, Slovak Republic; e-mail: *eva.muchova@euba.sk*.

[***] University of Economics in Bratislava, Slovak Republic; e-mail: *peter.lesko@euba.sk*.

Article history: Received 17 April 2022 | Accepted 13 February 2023 | Published online 25 March 2023

To cite this article: Terek, M., Muchová, E., Leško, P. (2023). How to Analyze the Association between Two Categorical Variables Based on Census Data with a High Level of Nonresponse. *Scientific Annals of Economics and Business*, 70(1), 71-81. https://doi.org/10.47743/saeb-2023-0009.

1. INTRODUCTION

Statistical surveys are often used in forming managerial policy and practice. In this paper we study, how to analyze the association between two categorical variables based on census data, with a high level of nonresponse.

In the academic year 2020/2021, a second questionnaire survey was planned at the University of Economics in Bratislava (the first one was realized in the academic year 2019/2020). As in the first stage, the objective of the project was to work up an interactive and multimedia framework for teaching The Principles of Economics 1 (ET1) and The Principles of Economics 2 (ET2), the core courses of the study programs at the undergraduate level and, also to study some issues related to the subjective well-being of the students. The sample survey and census were available. The high level of nonresponse, quite common in current surveys, can significantly impair the quality and explanatory power of the survey results (see Cochran, 1977; Tillé, 2001; Särndal & Lundström, 2005; Levy & Lemeshow, 2008; Chaudhuri, 2014; Lohr, 2019; Terek, 2020; Tillé, 2020). If the same response rate in census and sample survey is supposed, then the number of responses is higher in the census. If the connection to all population units is available and the costs of survey execution by the census are practically no different from sample survey costs, then the census should be preferred. Therefore, a census was chosen in the survey at the University of Economics in Bratislava.

Many studies focus on analyzing the association between two categorical variables based on sample survey data (Agresti, 2010, 2013; Agresti & Finlay, 2014; Terek, 2016, 2017; Agresti, 2018; Terek, 2019a). The purpose of this study is to suggest the approach to the analysis of the association between two categorical variables based on census data with a high level of nonresponse. We propose a new procedure of the analysis of association between two categorical variables if the data from census collected in a questionnaire survey with a high level of nonresponse are available. We recommend using of the approach in all surveys in which the costs of the survey by census practically do not differ from the sample survey costs and the connections to all population units are available.

We realize the research in two steps. Firstly, the estimation of the population frequency distribution based on weights modified by compensation for nonresponse is studied. The result is obtaining the frequency distribution of pseudo-population. In the second step, we suggest the methods of the analysis of the association between two categorical variables on the obtained pseudo-population. We verify the proposed approach on census data from the mentioned survey realized at the University of Economics in Bratislava in 2020/2021.

2. MATERIAL AND METHODS

If the data from the sample survey with a high nonresponse rate are available, it is possible to estimate the frequency distribution with compensation for nonresponse, using the modified sampling weights. In without-replacement sampling, the sampling weight w_{Bi} for observation i is always the reciprocal of the probability π_i that the observation i is included in the sample. If the considering of nonresponse is needed, the adjustment factor to the base weight is used. The final weight for the i^{th} observation is then

$$w_i = w_{Bi} \cdot w_{NRi}$$

where w_{NRi} is the nonresponse adjustment factor (for more details, see Terek *et al.*, 2021). The probability that a unit selected for the sample will respond φ_i (unknown but assumed positive) is called the response propensity for the i^{th} unit. If responding of unit *i* is independent of the indicator variable for presence in the selected sample, then the probability that unit *i* will be selected in the sample and responds is equal to $\pi_i\varphi_i$.

In Lohr (2019) the methods of estimation the response propensity φ_i are presented. One of them is based on poststratification using weights (see Terek *et al.*, 2021, for more details on this method and other possibilities of estimating the response propensity). After taking a simple random sample, units are grouped into *H* different poststrata (for more on poststratification, see Lohr (2019); Levy and Lemeshow (2008); Terek *et al.* (2021)). The population has N_h units in h^{th} poststratum; of these, n_h were selected for the sample, and n_{hR} responded. The response propensity for every respondent *i* in poststratum *h* is estimated by

$$RR_{w_h} = \frac{\sum_{i=1}^{n_{hR}} w_{Bi}}{N_h} \tag{1}$$

and nonresponse adjustment factor is

$$w_{NRi} = \frac{1}{RR_{w_h}} \tag{2}$$

In Eltinge and Yansaneh (1997); Gelman and Carlin (2001); Little and Vartivarian (2003); Vartivarian and Little (2003); Terek *et al.* (2021) is stated when the collapsing of poststrata is needed. The weights w_i can be used to construct the estimators of population quantities.

If the decision to apply the poststratification using weights is taken, the decision on which poststratification variables should be used in poststratification must be solved. It is known that the bias of estimators caused by nonresponse can be minimized by finding poststratification variables that are strongly correlated with the response propensity. In Terek *et al.* (2021) is advised the use of correlation ratio $\eta_{(Z|X)}$ for measuring that correlation.

In Terek *et al.* (2021), the problem of estimation with compensation for nonresponse in statistical analyses of census data is studied and discussed. The suggested approach is based on the idea that the census in which all population units are selected can be understood as a without-replacement sampling of size *N* with the only difference that the last unit is selected non-randomly. Then the poststratification using weights can be used (if MAR data[1] are assumed).

The final weight w_i for unit *i* is then

$$w_i = \frac{1}{RR_{w_h}} = \frac{N_h}{N_{hR}} \tag{3}$$

where N_h is the number of units in poststratum *h;* of these, N_{hR} responded.

If a sample is non-self-weighting, i. e., all final weights are not equal, the use of weights in constructing estimators of population quantities is needed. The estimate of the frequency of class *j* is then (Terek, 2019b).

$$n_j = \sum_{i \in S} w_i u_{i,j} \tag{4}$$

where $u_{i,j}$ = 1 if observation *i* is in class *j* and 0 otherwise, and *S* denotes the selected sample. The frequencies n_j define the pseudo-population (Terek & Muchova, 2017; Lohr, 2019) and, further, are called the frequencies of pseudo-population.

If certain values of one variable tend to go with certain values of the other, there is an association between two variables (Agresti & Finlay, 2014). The data of the analysis of categorical variables are displayed in contingency tables. When the data in the contingency table are analyzed, the corresponding joint, marginal and conditional distributions can be determined (Freund, 1992; Agresti & Finlay, 2014; Miller & Miller, 2014). Two categorical variables are statistically independent if the population conditional distributions on one of them are identical at each category of the other. The variables are statistically dependent if the conditional distributions are not identical (Agresti & Finlay, 2014).

Suppose the contingency table contains the data from a random sample. The questions usually addressed in the analysis of a contingency table are as follows (Agresti & Finlay, 2014):

- Do an association exist? The chi-squared test of homogeneity or independence answers this question (in the case of ordinal variables, also other possibilities exist – see Agresti (2010); Agresti (2013); Agresti and Finlay (2014); Agresti (2018).
- How do the data differ from what is expected under independence? The standardized residuals identify the cells that are different from what the independence predicts.
- What is the strength of association? The difference of proportions and odds ratio are strongly advised for measuring it.

If the data from the whole population (or pseudo-population) are accessible, the statistical independence can be found directly by comparing conditional distributions.

2.1 Pseudo-population and analysis of association between two categorical variables

If the census takes the form of the questionnaire survey, a high level of nonresponse must be expected, which means that the obtained population data are not complete. Imagine the response rate is 20 %. Is it possible to consider such data set as the studied population and to compare conditional distributions? Certainly not.

If we understand a census of the population of the size *N* be random sampling without replacement of size *N*, can we use the chi-square test and other well-known steps of the analysis of association? We cannot because it supposes that the observations are statistically independent and equally distributed, which can be considered fulfilled if the sample size n is small compared to the size *N* of the population ($\frac{n}{N} \leq 0{,}05$). But, in a census, all units of the population are selected. The corrections for the chi-squared test are needed (see Lohr, 2019). However, the nonresponse is not considered.

We propose the following approach. If we have the data from a sample survey with a high nonresponse rate, we can estimate the frequency distribution with compensation for nonresponse. Then we can analyze the association between two categorical variables on the obtained pseudo-population. If the census is understood as without-replacement random sampling, the same approach is also possible for a census.

After compensating for nonresponse, the final weights are not equal for all observation units. That means that a sample is non-self-weighting. Then the use of sampling weights in estimating the population frequency distribution is needed. The estimate of the frequency of class *ij* is:

$$n_{ij} = \sum_{k \in S} w_k u_{k,ij} \tag{5}$$

where $u_{k,ij}$ = 1 if observation k is in the class ij and 0 otherwise. The frequencies n_{ij} are the frequencies of pseudo-population.

If the contingency table contains the data on the population (or pseudo-population), the conditional distributions can be directly compared. The comparison of conditional distributions can only provide evidence of the association's existence but not on its structure. We propose to analyze the association structure by residuals $d_{ij} = \left(n_{ij} - o_{ij}\right)$, where n_{ij} are the frequencies of pseudo-population and $o_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ are expected frequencies of pseudo-population. The cells different from the situation supposed by the independence of variables can be recognized by residuals. The positive sign of residual indicates that the combination *ij* is more frequent than expected under independence, and the negative sign indicates that the combination *ij* is less frequent than expected under independence.

The strength of association in any sub-table 2 x 2 of the contingency table can be measured by the odds ratio. For a binary response variable, success denotes the outcome of interest and failure the other outcome. The odds of success are defined as follows.

$$\text{Odds} = \frac{Probability\ of\ success}{Probability\ of\ failure}$$

Suppose the data from a random sample are available. The estimated odds for a binary response variable equal the number of successes divided by the number of failures. If the data of finite population (or pseudo-population) are available, the word estimated in the term estimated odds is omitted. The odds ratio measures the association for any 2 x 2 contingency table. It equals the odds in row 1 divided by the odds in row 2. The odds ratio can equal any nonnegative number, and its values farther from 1 in a given direction represent a stronger association (Agresti & Finlay, 2014).

3. ANALYSES OF ASSOCIATION IN THE REALIZED CENSUS

We carried out the census mentioned above in the framework of the project "Learn Economics" at the end of the academic year 2020/2021. We included in the survey six Faculties of the University of Economics in Bratislava – Faculty of National Economy (FNE), Faculty of Business Informatics (FBI), Faculty of Commerce (FC), Faculty of Business Management (FBM), Faculty of International Relations (FIR), and Faculty of Applied Languages (FAL).

The questionnaire consisted of two parts. The first part (questions 1 – 24) concerns the teaching of the courses. The Principles of Economics 1 (ET1), and The Principles of Economics 2 (ET2), the second part of the questionnaire (questions 27 – 38), focus on subjective well-being. We used Google forms software for the execution of the survey and Microsoft Excel in analysis of contingency tables. A total of 1,336 students at the University of Economics completed at least one of the mentioned courses in the academic year in question. We approached the students via the Academic Information System at their e-mail addresses. We contacted only those students who took part in the exam of the courses ET1 and/or ET2 during the academic year 2020/2021. We delivered the questionnaire to each of them. The questions of the first part referred to the content, methodology, format of lectures, and seminars. The questions asked for the assessment of online teaching, preferable way of teaching, identification of most complicated topics in the syllabus, ideas for improving the lectures and seminars, personal experience with the online platform used for education, and others. The second part of the questionnaire was focused on subjective well-being and quality of life. It included 13 questions dealing with life satisfaction, health, family background, university study, daily activities. 258 students returned the completed questionnaire, total response rate achieving 238/1336 = 0.1781, i. e., 17.81%.

The study department at the University of Economics provided some auxiliary information about the population of 1,336 students. The frequency distribution by Faculty and Gender is presented in Table no. 1 (with corresponding numbers of responding students in parentheses). Then, we calculated the correlation ratio $\eta_{(Z|X)}$ between the response propensity and variables (for the calculation details, see Terek *et al.*, 2021). The results are presented in Table no. 2. Table no. 2 shows that the Faculty – Gender reaches maximum correlation ratio, and thus, these will serve as poststratification variables. We define the poststrata by combinations of categories of Faculty and Gender. There are 2 x 6 = 12 poststrata.

Table no. 1 –The distribution of students accomplishing at least one of the courses in the academic year in question by Faculty and Gender

	FNE	FBI	FC	FBM	FIR	FAL	Total
Males	153 (26)	138 (17)	109 (9)	135 (12)	42 (6)	5 (0)	582 (70)
Females	186 (39)	136 (38)	164 (42)	159 (25)	78 (18)	31(6)	754 (168)
Total	339 (65)	274 (55)	273 (51)	294 (37)	120 (24)	36 (6)	1 336 (238)

Table no. 2 – The values of the correlation ratio

	$\eta_{(Z|X)}$
Faculty	0,075
Gender	0,133
Faculty – Gender	0,169

Next, the association between the Faculty of the students studies and the answers to the No. 31 question of the questionnaire: "How satisfied are you with the studies at the University of Economics?" will be analyzed. The listed answers to the question include – "completely unsatisfied", "unsatisfied", "I can't decide", "satisfied", "completely satisfied". 233 students answered the question. Based on the requirement to gain at least 20 responding units in each poststratum, we collapsed some columns. The second condition – $w_{NRi} \leq 2$, advised in Lohr (2019), cannot be met, so we must be reckoned with less stability of weights. After collapsing there are 2 x 3 = 6 poststrata. Table no. 3 shows the resulting structure of poststrata (the number of addressed and responding students, in parentheses).

Table no. 3 – The structures of poststrata after collapsing with the number of responding students in parentheses

	FNE	FC, FBM	FBI, FIR, FAL	Total
Males	153 (25)	244 (21)	185 (22)	582 (68)
Females	186 (38)	323 (65)	245 (62)	754 (165)
Total	339 (63)	567 (86)	430 (84)	1336 (233)

The final weights calculated from Table no. 3, following relation (3), are presented in Table no. 4.

Table no. 4 −Final weights

	FNE	FC, FBM	FBI, FIR, FAL
Males	6.120	11.619	8.406
Females	4.895	4.969	3.952

The frequencies of responding students linked to poststrata are in Table no. 5. After weighing the frequencies by final weights in Table no. 4 and after joining males and females, the final frequency distribution n_{ij} of the pseudo-population is acquired as presented in Table no. 6 (n_{ij} are calculated following the relation (5)).

Table no. 5 −Frequencies linked to poststrata

		FNE	FC, FBM	FBI, FIR, FAL	Total
Completely unsatisfied	Males	1	0	0	1
	Females	0	2	0	2
Unsatisfied	Males	1	0	0	1
	Females	2	6	3	11
I can't decide	Males	0	8	2	10
	Females	5	8	6	19
Satisfied	Males	18	10	18	36
	Females	23	44	38	105
Completely satisfied	Males	5	3	2	10
	Females	8	5	15	28
Total		63	86	84	233

Table no. 6 −The final frequency distribution of pseudo-population

	FNE	FC, FBM	FBI, FIR, FAL	Total
Completely unsatisfied	6.12	9.938	0	16.058
Unsatisfied	15.91	29.814	11.856	57.58
I can't decide	24.475	132.704	40.524	197.703
Satisfied	222.745	334.826	301.484	859.055
Completely satisfied	69.76	59.702	76.092	205.554
Total	339	567	430	1336

For identifying if there is an association between responses to No. 31 question and faculties, the conditional distributions of answers to No. 31 question on faculties are calculated (dividing each frequency in the referring column in Table no. 6 by its column total) and presented in Table no. 7. The conditional distributions in the columns of Table no. 7 are not identical. Thus, there is an association between faculties and answers to the No. 31 question.

Table no. 7 − The conditional distributions of answers to No. 31 question on faculties

	FNE	FC, FBM	FBI, FIR, FAL
Completely unsatisfied	0.018	0.018	0
Unsatisfied	0.047	0.053	0.028
I can't decide	0.072	0.234	0.094
Satisfied	0.657	0.591	0.701
Completely satisfied	0.206	0.105	0.177
Total	1	1	1

We analyze the structure of the association by residuals. Firstly, we calculated the expected frequencies $o_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ (in Table no. 8) and then the residuals $d_{ij} = (n_{ij} - o_{ij})$ in Table no. 9. For expected frequencies in Table no. 8, all conditional distributions of answers to No. 31 question on faculties are identical (it means that the variables "Faculty" and "Answers to No. 31 question" are independent).

Table no. 8 − Expected frequencies o_{ij}

	FNE	FC, FBM	FBI, FIR, FAL	Total
Completely unsatisfied	4.075	6.815	5.168	16.058
Unsatisfied	14.610	24.437	18.532	57.58
I can't decide	50.166	83.905	63.632	197.703
Satisfied	217.979	364.584	276.492	859.055
Completely satisfied	52.158	87.237	66.159	205.554
Total	339	567	430	1336

Table no. 9 − Residuals $\mathbf{d_{ij}}$

	FNE	FC, FBM	FBI, FIR, FAL	Total
Completely unsatisfied	2.045	3.123	− 5.168	0
Unsatisfied	1.300	5.377	− 6.676	0
I can't decide	− 25.691	48.799	− 23.108	0
Satisfied	4.766	− 29.758	24.992	0
Completely satisfied	17.602	− 27.535	9.933	0
Total	0	0	0	0

The interpretation of the residuals in Table no. 9 leads to exciting outcomes. The positive sign of a residual indicates that the combination *ij* is more frequent than expected under independence. The negative sign indicates that the combination *ij* is less frequent than expected under independence. Thus, Table no. 9 shows that, for example, the answers "satisfied" and "completely satisfied" of the students of the Faculty of Commerce and Faculty of Business Management are less frequent than expected under independence, and the answers "completely unsatisfied," "unsatisfied," and "I can't decide" of the students at these faculties are more frequent than expected under independence. A different situation is, for example, at the Faculty of Business Informatics, Faculty of International Relations, and Faculty of Applied Languages. The answers "completely unsatisfied," "unsatisfied," and "I can't decide" are less frequent than expected under independence, and the answers "satisfied" and "completely satisfied" are more frequent than expected under independence. At the Faculty of National Economy, the answers "completely unsatisfied," "unsatisfied," "satisfied," and "completely satisfied" are more frequent, and the answers "I can't decide" are less frequent than expected under independence.

Table no. 10 − Sub-table of Table no. 6

	Unsatisfied	Satisfied	Total
FBI, FIR, FAL	11.865	301.484	313.34
FC, FBM	29.814	334.826	364.64

The strength of association can be measured in any sub-table 2 x 2 of the contingency table by odds ratio. The answers "unsatisfied" and "satisfied" at Faculties of Business Informatics, International Relations, and Applied Languages and at Faculties of Commerce and Business Management will be compared, for illustration. Table no. 10 is the corresponding sub-table of Table no. 6.

We treat the answer to the No. 31 question as to the response variable. If we use the answer "satisfied" as success and "unsatisfied" as a failure, the odds for "satisfied" at the Faculties of Business Informatics, International Relations, and Applied Languages is

$$\frac{301.484}{11.856} = 25.429$$

It means that at the Faculties of Business Informatics, International Relations, and Applied Languages, there were 25.429 satisfied students for every unsatisfied student. The odds for "satisfied" at the Faculties of Commerce and Business Management is

$$\frac{334.826}{29.814} = 11.231$$

It means that at the Faculties of Commerce and Business Management, there were 11.231 satisfied students for every unsatisfied student. The odds ratio is

$$\frac{25.429}{11.231} = 2.264$$

The result can be interpreted, for example, as follows: Suppose one student is randomly selected from the population. If he is from Faculties of Business Informatics, International Relations, and Applied Languages, there are approximately 2.3 times greater odds that he is satisfied at the University of Economics, as if he is from the Faculties of Commerce and Business Management. All possible sub-tables of Table no. 6 can be analyzed accordingly.

4. CONCLUSIONS, LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The purpose of the paper is to propose a procedure of analysis of the association between two categorical variables based on the data collected from a census with a high level of nonresponse. We used the method of estimation with compensation for nonresponse applicable in censuses (proposed in Terek *et al.*, 2021) to estimate the frequency distribution of the population. Then we propose the procedure analyzing the association between two categorical variables on the obtained pseudo-population. The resulting pseudo-population allows the determination of the independence of variables by comparing conditional distributions. If the variables are dependent, we recommend the analysis of the association structure by residuals. Finally, the strength of association for any 2 x 2 sub-table of a contingency table can be measured by odds ratio.

The proposed procedure is designed for the cases in which the costs of survey execution by the census are practically no different from sample survey costs, and some applicable auxiliary information on the population units is available. The procedure also considers nonresponse quite common in current surveys. In the presented study, the response rate was only 17.81%. Therefore, such a data set cannot be regarded as the studied population, and compensation for nonresponse is needed. The interpretation of the analysis results leads to exciting information. The residuals allow identifying if the combination *ij* is more or less frequent than expected under independence. The interpretations of odds and odds ratio are exciting too.

In practice, there are a lot of situations similar to the situation in the presented study. For example, in a large company, they want to analyze employees' attitudes towards various methods of stimulating work, namely whether these attitudes vary depending on gender, employee age category, and the like. Or it may be interesting to determine the attitude and preferences of employees to different forms of further education. Here too, it may be interesting to examine whether these attitudes vary, for example, depending on the employee's age category. These problems came to the forefront during the period of the Covid Pandemic 19 when it was necessary to prefer distance education. At present, it may be interesting to find out whether employees would prefer this form of communication even today when it is no longer necessary or prefer to return to traditional forms of communication. There is no point in similar situations to realize random sampling of respondents because the company has contact with each employee and the cost of sample survey and census is virtually the same. Of course, in similar surveys, it is necessary to count on a large degree of nonresponse, similar to the presented survey. In addition, the company also owns a wealth of information about its employees, which can serve as auxiliary information in finding suitable poststratification variables. The presented procedure can therefore be used only at the level of a company or other organization. In these objects the conditions of the same cost per census and sample survey are practically always met, contact for each unit is known and the suitable auxiliary information should be in disposition.

In the subsequent similar study, in the academic year 2021/22, the original research related to the teaching of the course Basics of Economics and well-being has been extended to investigate the value orientation of students. Based on the obtained new data, the authors plan to use the procedure presented in this paper to the analysis of the association between the type of high school that the student has completed and his satisfaction with the study at the University of Economics. This can be useful to improve the focus of marketing activities aimed at acquiring students to study at that University.

We would like to focus further research on investigating the possibilities of census data analysis with a high level of nonresponse with the use of regression analysis. The procedure will be based on obtaining the frequency distribution of pseudo-population with aid of the procedure presented in this paper. One possibility is then the selection of an ordinal variable as a dependent variable. When we assign a score to each of its values, we can treat it as a quantitative continuous variable and a dependent variable in regression analysis. The parameters of the regression equation can be calculated by the least squares method (for more details, see Lohr, 2019, pp. 436-437). This equation should summarize useful information about the relationship between the dependent variable y and one or more independent variables in the finite population. If we consider the finite population, the population correlation coefficient of x and y, which measures the intensity of the linear relationship between two variables is known (see Lohr, 2019, p. 118). The coefficient of determination in simple linear regression and the multiple correlation coefficient and multiple coefficient of determination in multiple regression can be also calculated and interpreted.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-20-0338

ORCID

Milan Terek http://orcid.org/0000-0001-5638-9287
Eva Muchová https://orcid.org/0000-0001-9758-5313
Peter Leško https://orcid.org/0000-0002-3240-6721

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (2nd ed. ed.): John Wiley & Sons. http://dx.doi.org/10.1002/9780470594001

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed. ed.): Wiley and Sons.

Agresti, A. (2018). *Statistical Methods for the Social Sciences* (5th ed. ed.): Pearson.

Agresti, A., & Finlay, B. (2014). *Statistical Methods for the Social Sciences* (4th ed. ed.): Pearson.

Chaudhuri, A. (2014). *Modern Survey Sampling*. New York: Chapman and Hall. http://dx.doi.org/10.1201/b17087

Cochran, W. G. (1977). *Sampling Techniques*. New York: Wiley and Sons.

Eltinge, J. L., & Yansaneh, I. S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey. *Survey Methodology, 23*(1), 33-40.

Freund, J. E. (1992). *Mathematical Statistics* (5th ed. ed.): Prentice-Hall.

Gelman, A., & Carlin, J. B. (2001). Poststratification and Weighting Adjustments *Survey Nonresponse* (pp. 289–302). New York: Wiley and Sons.

Levy, P. S., & Lemeshow, S. (2008). *Sampling of Populations. Methods and Applications* (4th ed. ed.): Wiley and Sons. http://dx.doi.org/10.1002/9780470374597

Little, R. J., & Vartivarian, S. (2003). On weighting the rates in non-response weights. *Statistics in Medicine, 22*(9), 1589-1599. http://dx.doi.org/10.1002/sim.1513

Lohr, S. L. (2019). *Sampling: Design and Analysis* (2nd ed. ed.): CRC Press Taylor & Francis Group. http://dx.doi.org/10.1201/9780429296284

Miller, I., & Miller, M. (2014). *John E. Freund's Mathematical Statistics with Applications* (8th ed. ed.): Pearson Education Limited.

Särndal, C. E., & Lundström, S. (2005). *Estimation in Surveys with Nonresponse*: Wiley and Sons. http://dx.doi.org/10.1002/0470011351

Terek, M. (2016). Information channels effectiveness assessment on the basis of data from statistical survey. *Scientific Annals of Economics and Business, 63*(2), 225–235. http://dx.doi.org/10.1515/saeb-2016-0118

Terek, M. (2017). *Interpretácia štatistiky a dát. 5. doplnené vydanie*: Equilibria.

Terek, M. (2019a). *Dotazníkové prieskumy a analýzy získanych dát. 1. vydanie*: Equilibria.

Terek, M. (2019b). Obtaining the information about incomes from EU-SILC data and market analysis. *Journal of Eastern European and Central Asian Research (JEECAR), 6*(2), 205-219. http://dx.doi.org/10.15549/jeecar.v6i2.313

Terek, M. (2020). Možnosti riešenia problému neodpovedania v analýzach dát pri vyčerpávajúcom skúmaní prostredníctvom dotazníkových zisťovaní. *Slovenská štatistika a demografia, 30*(4), 28-41.

Terek, M., & Muchova, E. (2017). The structure of Incomes Analysis in Slovak Republic and Regions of the Slovak republic Based on EU-SILC Data. *International Journal of Economic Research, 14*(20), 425-434.

Terek, M., Muchova, E., & Lesko, P. (2021). How to make estimates with compensation for nonresponse in statistical analysis of census data. *Journal of Eastern European and Central Asian Research (JEECAR), 8*(2), 149-159. http://dx.doi.org/10.15549/jeecar.v8i2.619

Tillé, Y. (2001). *Théorie de sondages. Echantillonnage et estimation en populations finies*: Dunod.

Tillé, Y. (2020). *Sampling and Estimation from Finite Populations*: Wiley and Sons. http://dx.doi.org/10.1002/9781119071259

Vartivarian, S., & Little, R. (2003). *On the Formation of Weighting Adjustment Cells for Unit Nonresponse.* Paper presented at the The University of Michigan Department of Biostatistics Working Paper Series.

Notes

[1] More about MAR, MCAR and NMAR data, see Lohr (2019).