Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie

11 (947)

ISSN 1898-6447
Zesz. Nauk. UEK, 2015; 11 (947): 53–66
DOI: 10.15678/ZNUEK.2015.0947.1104

Ondrej Dúžik
Department of Statistics
University of Economics in Bratislava, Slovakia

Analysis of the Dependence of the Gross Money Income of Single-member Households in Slovakia*

Abstract

The article focuses on modelling the dependence of the gross money income of Slovak single-member households using relevant factors. The main aim is to construct an appropriate regression model through which to quantify the influence of relevant factors on the gross money income of these households. The article shows the results of an analysis gained from the Household Budget Survey conducted by the Statistical Office of the Slovak Republic in 2012. The basic aim of family accounts is to provide information for the analysis and monitoring of the social situation of households, mainly the structure of their incomes and expenses. The 2012 survey looked at 4,704 Slovak households, 1,098 of which – or 23.3% – were single-member households. Multiple analysis of variance, tests for the equality of marginal means, regression and correlation analysis were all used with SAS Enterprise Guide.

Keywords: household budget survey, gross money income, marginal means, regression analysis.

* Realisation of the paper was supported by a grant agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic VEGA. Project: VEGA no. 1/0548/16 "The progress of the SR in achieving of the strategy Europe 2020 in the area of poverty and social exclusion".

1. Introduction

This article evaluates the results of the analyses of the dependence of the gross money income (GMI) of single-member households using relevant factors. The analyses were done using the database of cross-sectional data from the Household Budget Survey conducted in 2012. The Household Budget Survey (HBS) is a regular selection inquiry into incomes and expenses of private households. The basic aim of family accounts is to provide information for the analysis and monitoring of the social situation of households, mainly of the structure of their incomes and expenses. 4704 Slovak households took part in the 2012 survey, out of which 1098 (or 23.3% of the total number) were single-member households.

The main aim of my research, the results of which are presented in this article, was to quantify the influence of relevant factors on the gross income of single--member households. For this purpose, it was necessary, with the help of statistical tools, to select from a set of potential factors those which significantly influenced the gross incomes of single-member households, to construct an adequate regression model depicting the dependence of incomes on selected factors and on the basis of this type of model to quantify their influence or, as the case may be, to estimate the rate of their influence on household incomes. The analyses were done with the PROC GLM and PROC REG procedures in SAS Enterprise Guide.

2. Selecting the Relevant Factors

These independent variables[1] were included in the analysis:

Table 1. List of Independent Variables

Variables	Description
REGION	administrative division of Slovakia (region)
SIZE_MU	size of municipality; the population
TYPE_MU	type of municipality: county seat, other cities or towns, village
AGE	quantitative variable
GENDER	gender
STATUS	marital status
EDUC	education
CEA	current economic activity

[1] In more detail – see Appendix.

Table 1 cnt'd

Variables	Description
EMP_ST	employment status
LEN_WT	length of working time
TYPE_WT	type of working time

Source: the author's own construction.

I used the Backward Elimination Method to select the relevant regressors from point of view of their benefit to the explanation of the variability of the gross money incomes of single-member households. Those variables which were statistically insignificant on the selected significance level of 0.1 were gradually excluded from the full model. This made it possible to determine that single-member household incomes had not been influenced by age, size or type of municipality where the household (person) lived, their employment status or type of working time. The other variables, which were left in the regression model, significantly helped explain the variability of single-member household incomes. The test results of the significance of the influence of the relevant factors are shown in Table 2.

Table 2. Verification of Statistical Significance of the Benefit of Variables to the Explanation of the Variability of the Gross Money Incomes of Single-member Households

Source	DF	Type III SS	Mean Square	*F*-value	*p*-value
REGION	1	4 081 309.63	4 081 309.63	49.12	< 0.0001
GENDER	1	756 574.91	756 574.91	9.10	0.0026
STATUS	2	1 072 155.06	536 077.53	6.45	0.0016
EDUC	4	10 102 794.30	2 525 698.57	30.39	< 0.0001
CEA	2	3 459 673.02	1 729 836.51	20.82	< 0.0001
LEN_WT	2	2 582 668.85	1 291 334.42	15.54	<0 .0001

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

While analysing the gross money incomes of single-member households, I also concentrated on regional disparities. On the basis of the tests for the equality of marginal means of the single-member household incomes (Table 3) I discovered that single-member household incomes in the Bratislava Region, with significance levels of 0.1 and 0.05, are statistically significantly different from those in the other regions of Slovakia (all *p*-values given in the bottom row of Table 3 are lower than the level of significance). Furthermore, a statistically significant difference in the marginal means of single-member household incomes was not confirmed between any other pair of regions in Slovakia (Fig. 1). The 90% interval estimates of marginal means

and the confirmation of statistical significance of the marginal means between individual pairs of regions are shown in Fig. 2. This figure also clearly confirms that average incomes of single-member households in the Bratislava Region are significantly higher than those in all the other regions. With a significance level of 0.1, there is no statistically significant difference in marginal means of gross money incomes of single-member households between any other pair of regions.

Table 3. *p*-values for the Test for the Statistical Significance of the Equality of Marginal Means for the Variable REGION

Least Squares Means for Effect REGION *p*-value for H0: LSMean (*i*) = LSMean (*j*) Dependent Variable: Gross Money Income								
i/j	KE (Košice)	PO (Prešov)	BB (Banská Bystrica)	ZA (Žilina)	NR (Nitra)	TN (Trenčín)	TT (Trnava)	BA (Bratislava)
KE		0.8010	0.3422	0.4116	0.3208	0.2534	0.6422	0.0001
PO	0.8010		0.5037	0.5707	0.4728	0.3778	0.8461	0.0002
BB	0.3422	0.5037		0.9721	0.9511	0.7830	0.6105	< 0.0001
ZA	0.4116	0.5707	0.9721		0.9293	0.7812	0.6765	< 0.0001
NR	0.3208	0.4728	0.9511	0.9293		0.8316	0.5705	< 0.0001
TN	0.2534	0.3778	0.7830	0.7812	0.8316		0.4544	< 0.0001
TT	0.6422	0.8461	0.6105	0.6765	0.5705	0.4544		< 0.0001
BA	0.0001	0.0002	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	

Remark: administrative division of Slovakia (region) – see description in Appendix.

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

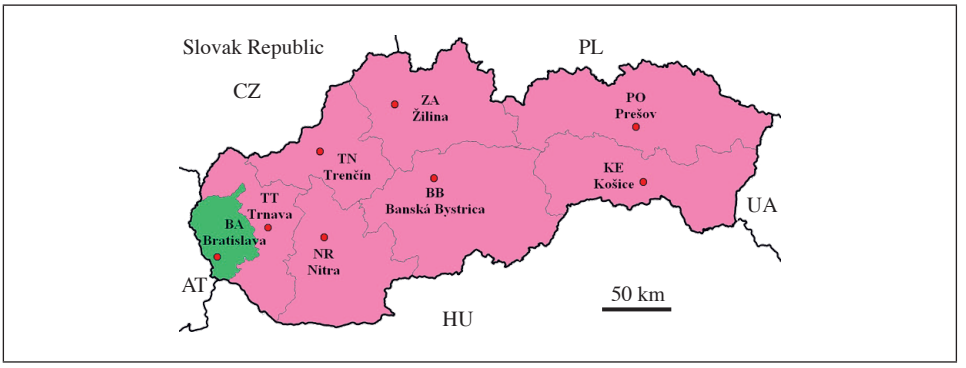

Fig. 1. Differentiation According to Slovak Regional Disparities in Incomes of Single-member Households

Source: the author's own construction.

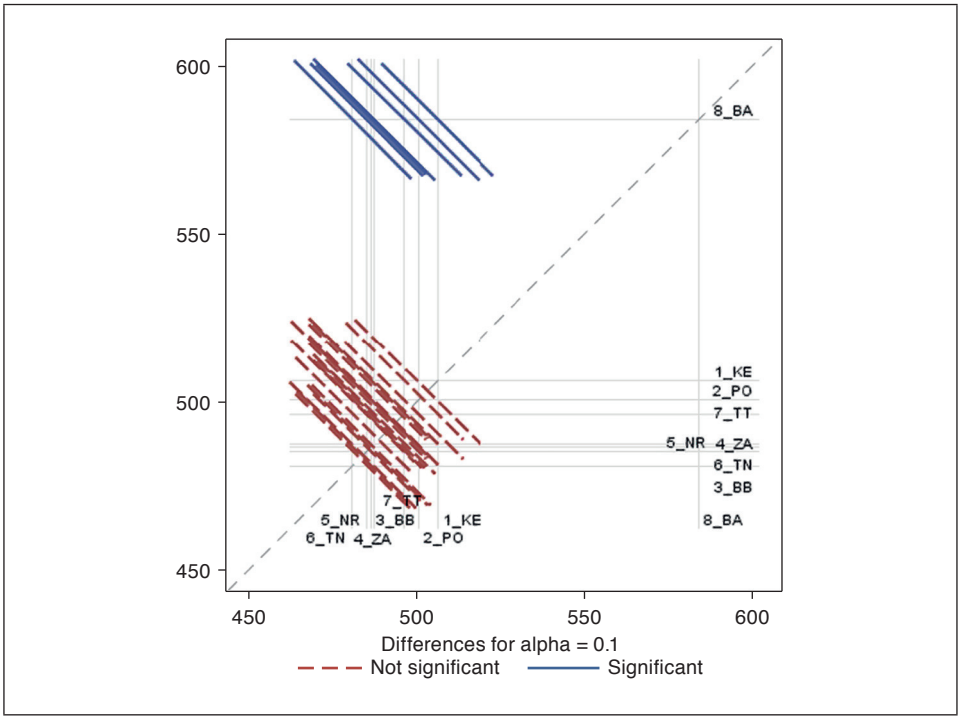

Fig. 2. 90% Interval Estimates of Marginal Mean Values of the Gross Money Incomes of Single-member Households in Individual Slovak Regions in 2012

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

Table 4. Point Estimates and Test for the Equality of Marginal Means of the Gross Money Income of Slovak Single-member Households in 2012 for the Bratislava Region and the Other Regions

Region	GMI LSMean	H0:LSMean1 = LSMean2
		p-value
Bratislava	608.94	< 0.0001
Other regions	542.16	

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

Proceeding from these findings, a variable REGION with two variants was created: BA (Bratislava) and other regions. On the basis of the point estimates of marginal means shown in Table 4, it is evident that in 2012, provided that all the other independent variables included in the regress model are fixed, the average gross money incomes of single-member households in the Bratislava Region were

higher on average by 66.78 euro than in the other regions. On the basis of the *p*-value shown, this difference is statistically significant on each commonly used level of significance.

3. Estimating the Regression Model

To avoid distorted results owing to outlying and influential observations (see Bollen & Jackman 1990, Cook 1997, Cook & Weisberg 1982, Vojtková & Labudová 2010), I used Cook's statistics and DFFIT statistics to detect and then delete such observations. From the original file, I deleted 47 statistical units where Cook's statistics and DFFIT statistics both proved that the observation is influential and thus it can distort the results to a considerable extent. These 47 observations represented 4.28% of the whole range of the file and the range was reduced to 1051 statistical units.

Another factor which can negatively influence the quality of the model, and especially the interpretability of regression coefficients in an estimated regression model, is a high level of multicollinearity. The variables *current economic activity* and *length of working time* contributed to multicollinearity to the greatest extent. These were identified through the variance inflation factor (VIF) and proportion of variance. This is understandable, since persons without working time (who did not work, in other words) comprised the majority of those to whom the variable *length of working time* within the Household Budget Survey could not have been applied, and, at the same time, had the status of unemployed or economically inactive within the variable CEA. They were then combined, and a new variable, *current economic activity*, was created with the modifications given in Table 5. In further analysis, a variable thus defined replaced two of the original variables.

Table 5. New Categories of Current Economic Activity (CEA)

Type	Description
CEA Unemployed	unemployed person
CEA Inactive	economically inactive person
CEA FT_empl	full-time employed person
CEA PT_empl	part-time employed person
CEA NA_empl	employed person with inapplicable length of working time (NA – not applicable)

Source: the author's own construction.

The graphic analysis of studentised residuals (Fig. 3) revealed heteroskedasticity of a random component, and therefore the weighted method of least squares was used to estimate the regression model depicting the dependence of gross incomes of single-member households on the aforementioned factors. I therefore applied FGLS method (see Wooldridge 2009), which provides a flexible approach to modelling heteroskedasticity. The results in this model show that gender has no statistically significant influence on the income of a single-member household. However, they also confirm that the income of a single-member household is significantly influenced by the type of municipality the household lives in. The final estimate of the regression model is shown in Table 6.

The model estimated by the weighted method of least squares is, in its entirety, a statistically significant p-value < 0.0001, and all the independent variables incorporated into the regression model at the level of 0.05 statistically significantly influence the single-member household incomes. Variables in the output shown in Table 6 are arranged in descending order according to their benefits to the explanation of variability of the dependent variable (see column *Squared Semi-partial Corr Type I*). The variable *current economic activity* contributes to the greatest extent to the explanation of variability of gross incomes of single-member households (37.7%). It is followed by education, which subsequently contributes to the explanation of variability of gross money incomes of single-member households to a lesser extent (7.3%), etc. Despite the fact that only the categorical variables from the set of independent variables were put into the regression model by the variable selection method, the model explains more than 50% of the variability of gross money incomes of single-member households. To be precise, the model explains 53.37% (see R-squared-coefficient of multiple determination: Budayová, Bolek & Šupšák 2013, Bollen & Jackman 1990, Cook & Weisberg 1982, Garson 2012, Šoltés 2008).

The coefficient for the intercept represents mean gross money incomes (in euro) of single-member households where the member of this household is employed full--time, single or divorced, has an elementary education and lives in a village not located in the Bratislava Region.

If we concentrate on the *current economic activity* of the person of a single--member household, then the highest income can be seen in the single-member households where the person is employed full-time. If the single-member household consists of a part-time employed person, the income is lower on average by 193.39 euro. Naturally, the lowest income can be seen in the single-member households where the person is unemployed. The income is lower on average by 419.03 euro than the income of a single-member household with a full-time employed person. The given interpretations are valid on ceteris paribus condition, i.e. provided that all the other independent variables are constant.

Table 6. Estimate of the Regression Model for Gross Money Incomes of Single-member Households Acquired by Weighted Method of Least Squares

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	*F*-value	*p*-value
Model	12	103 267 615	8 605 635	99.01	< 0.0001
Error	1038	90 221 728	86 919	–	–
Corrected Total	1050	193 489 343	–	–	–

Root MSE	294.8200	*R*-square	0.5337
Dependent Mean	422.1480	Adj *R*-square	0.5283
CoeffVar	69.8381	–	–

Parameter Estimates						
Variable	Param. Est	Stand. Error	*t*-value	*p*-value	Squared Semi--partial Corr Type I	Variance Inflation
Intercept	614.71	14.6051	42.09	< 0.0001	–	0
CEA Full-time empl	0	–	–	–	–	–
CEA Unemployed	–419.03	17.3750	–24.12	< 0.0001	0.0536	1.8529
CEA Inactive	–321.94	13.0994	–24.58	< 0.0001	0.2770	2.5649
CEA Part-time empl	–193.39	21.1943	–9.12	< 0.0001	0.0365	1.4576
CEA Not-applicable empl	–128.31	31.2023	-4.11	< 0.0001	0.0096	1.1759
EDUC Elementary school	0	–	–	–	–	–
EDUC Higher education	222.22	18.4525	12.04	< 0.0001	0.0457	1.1723
EDUC Bachelor's degree	93.73	30.1596	3.11	0.0019	0.0007	1.0741
EDUC Secondary education	71.97	8.2633	8.71	< 0.0001	0.0205	1.7379
EDUC Vocational school	36.53	8.5158	4.29	< 0.0001	0.0057	1.5813
STATUS Single or divorced	0	–	–	–	–	–
STATUS Widowed	85.65	6.8389	12.52	< 0.0001	0.0607	1.2311
REG Other region	0	–	–	–	–	–
REG Bratislava	66.78	12.1282	5.51	< 0.0001	0.0209	1.2914
TYPE_MU Village	0	–	–	–	–	–
TYPE_MU County seat	20.60	9.4341	2.18	0.0292	0.0008	1.6335
TYPE_MU Other cities	16.23	7.2594	2.23	0.0257	0.0022	1.3278

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

In 2012 the lowest gross money incomes were in those single-member households where the members had an elementary education. If the single-member

household consisted of a person with vocational education, the income was higher on average by 36.53 euro; if it consisted of a person with a secondary education, the income was higher on average by 71.97 euro; if it consisted of a person with higher education, the income was higher on average by 222.22 euro.

If we look at the interval estimates of regression coefficients, we can see that the confidence intervals for regression coefficients pertaining to secondary education and bachelor's degree have a large intersection and therefore in 2012 there was no significant difference between the gross money incomes of single-member households of persons with secondary education and those with a bachelor's degree. This finding is confirmed by Fig. 3 as well.

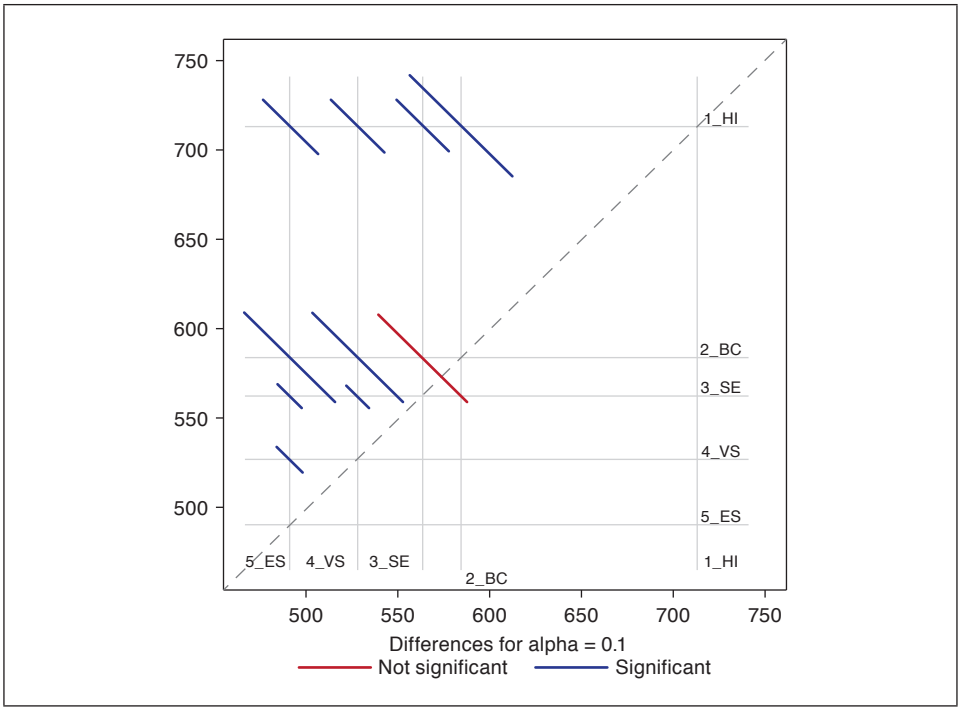

Fig. 3. 90% Interval Estimates for the Marginal Mean Values of the Gross Money Incomes of Slovak Single-member Households in 2012 in Relation to Level of Education Attained

Source: the author's own construction in SAS, data – *Household Budget Survey* (2012). Category description in Appendix.

Naturally, income increases with level of education, and thus the single-member households comprised of members with an elementary education have the lowest incomes and those of members with higher education have the highest incomes.

By comparing the estimates of marginal means shown in Table 7 (e.g. 712.88 euro for higher education and 490.66 euro for elementary education) we obtain the differences corresponding to the regression coefficients interpreted for the education variable (e.g. see the regression coefficient 222.22 euro for the variable *EDUC Higher education*).

Table 7. Point and Interval Estimates of Marginal Mean Values of Gross Money Incomes of Slovak Single-member Households in 2012 Depending on the Level of Education the Household Member Attained

Education	Gross Money Income LSMean	90% Confidence Limits	
Higher education – HI	712.88	681.78	743.98
Bachelor's degree – BC	584.39	535.19	633.59
Secondary education – SE	562.64	546.74	578.53
Vocational school – VS	527.19	510.55	543.83
Elementary school – ES	490.66	472.60	508.72

Source: the author's own calculation in SAS EG, data – *Household Budget Survey* (2012).

Concerning the marital status of the person in a single-member household, in 2012 widows had a higher monthly gross money income (85.64 euro, on average) than singles or divorced individuals, again on ceteris paribus condition.

The dominance of incomes in the Bratislava Region was analysed in the 2nd part of this article and the regression coefficient in Table 6 confirmed that in 2012 single-member households in the Bratislava Region had higher gross money incomes on average by 66.78 euro than single-member households in other regions.

Although nowadays a lot of people from villages and smaller towns commute to work in bigger towns, a statistically significant difference was confirmed between the incomes of single-member households in county seat towns and villages. In 2012 single-member households living in county seat towns had higher gross money incomes on average by 20.60 euro than single-member households living in villages.

4. Resume

The aim of the article was to model the dependence of the gross money income of Slovak one-person households using relevant factors and an estimated regression model to quantify the influence of those factors. The analysis is based on the survey: Household Budget Survey from 2012, which was conducted by the Statistical Office of the Slovak Republic.

On the basis of the methods used – multivariate analysis of variance, tests for the equality of marginal means, regression and correlation analysis – I found that the gross income of Slovak one-person households in 2012 was significantly influenced mainly by these qualitative variables: current economic activity, length of working time, the highest attained level of education, marital status, region, and type of municipality where the household lives. Concerning regional disparities, it was found that there was a significant difference in the mean income of one-person households only between households living in the Bratislava Region and any other region of Slovakia.

After deleting the influential observations, I estimated a linear regression model by the weighted least squared method in SAS Enterprise Guide. The model fulfills assumptions about error terms of regression model and has only a slight degree of multicollinearity. On the basis of the estimated partial correlation coefficients, I sought to determine the strength of the relationship between income and the individual explanatory variables. The effects of particular categorical variables on the gross money income of Slovak one-person households were quantified by marginal means as well as regression coefficients.

Appendix

Variable	Count	Category	Description
GMI	Gross Money Income of Households in Slovakia in euro		
REG	Administrative division of Slovakia (region)		
	214	BA	Bratislava
	145	TT	Trnava
	120	TN	Trenčín
	151	NR	Nitra
	97	ZA	Žilina
	155	BB	Banská Bystrica
	105	PO	Prešov
Ref.	111	KE	Košice
SIZE_MU	Size of municipality		Population size
	138	SIZE_MU_1	–999
	101	SIZE_MU_2	1000–1999
	95	SIZE_MU_3	2000–4999
	55	SIZE_MU_4	5000–9999
	105	SIZE_MU_5	10 000–19 999

Appendix cnt'd

Variable	Count	Category	Description
	219	SIZE_MU_6	20 000–49 999
	156	SIZE_MU_7	50 000–99 999
Ref.	229	SIZE_MU_8	100 000–
TYPE_MU	Type of municipality		
	356	TYPE_MU_CS	County seat
	400	TYPE_MU_OT	Other cities or towns
Ref.	342	TYPE_MU_VI	Village
AGE			
GEN	Gender		
	175	M	Men
Ref.	923	W	Women
STATUS	Marital status		
	304	DIV	Divorced
	457	WID	Widowed
Ref.	337	SIN	Single
EDUC	Education		
	146	HI	Higher education
	33	BC	Bachelor's degree
	487	SE	Secondary education
	278	VS	Vocational school
Ref.	154	ES	Elementary school
CEA	Current economic activity		
	36	UNEMP	Unemployed person
	653	INACT	Economically inactive person
Ref.	409	EMPL	Employed person
EMP_ST	Employment status		
	598	NA	Not applicable
	91	UNC	Unclassified
	1	EMR	Employer
	35	SEM	Self-employer
Ref.	373	EME	Employee
LEN_WT	Length of working time		
	725	NA	Not applicable
	50	PT	Part-time
Ref.	323	FT	Full-time

Appendix cnt'd

Variable	Count	Category	Description
TYPE_WT	Type of working time		
	725	NA	Not applicable
	9	CW	Casual work
	45	FTC	Fixed-term contract
Ref.	319	IDP	Contract for an indefinite period

Source: the author's own construction, data – *Household Budget Survey* (2012).

Bibliography

Bollen K. A., Jackman R. W. (1990), *Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases* (in:) J. Fox, J. Scott Long (eds), *Modern Methods of Data Analysis*, Sage, Newbury Park, CA.

Budayová J., Bolek V., Šupšák R. (2013), *Analysis of Demand for Financial Services by Income of Households* (in:) *Ekonomika, financie a manažmentpodniku*, Zborník vedeckých statípriprí ležitosti 60. Výročia založenia Fakulty podnikové homanažmentu EU v Bratislave, Bratislava.

Cook R. D. (1997), *Detection of Influential Observation in Linear Regression*, "Technometrics", vol. 19, nr 1, http://www.ime.usp.br/~abe/lista/pdfWiH1zqnMHo.pdf, accessed: 30.11.2014.

Cook R. D., Weisberg S. (1982), *Residuals and Influence in Regression*, Chapman & Hall, New York.

Garson G. D. (2012), *General Linear Models: Multivariate GLM, MANOVA and MANCOVA*, Asheboro, N. C. Statistical Associates Publishers.

Hebák P., Hustopecký J., Malá I. (2005), *Vícerozměrné statistické metody* (*2*), Informatorium, Praha.

Household Budget Survey (2012), Statistical Office of the Slovak Republic.

Meloun M., Militký J. (2004), *Statistická analýza experimentálních dat*, Academia, Praha.

Šoltés E. (2008), *Regresná a korelač ná analýza s aplikáciami*, Iura Edition, Bratislava.

Vojtková M., Labudová V. (2010), *Regionálna analýza výdavkov a príjmov domácností v Slovenskej republike*, "Ekonomický časopis", vol. 58, nr 8, Bratislava.

Wooldridge J. M. (2009), *Introductory Econometrics: A Modern Approach*, South-Western Cengage Learning, Mason.

Analiza zależności przychodów jednoosobowych gospodarstw domowych na Słowacji od wybranych zmiennych

(Streszczenie)

W artykule skupiono się na modelowaniu zależności przychodów pieniężnych gospodarstw jednoosobowych na Słowacji od pewnych zmiennych. Główny cel prac badaw-

czych stanowiła budowa odpowiedniego modelu regresji, za pomocą którego możliwe byłoby ilościowe określenie stopnia wpływu wybranych zmiennych na przychody badanych gospodarstw. W artykule zaprezentowano wyniki badań opartych na danych z Badań Budżetów Gospodarstw Domowych, przeprowadzonych przez Urząd Statystyczny Republiki Słowackiej w 2012 r. Podstawowym celem monitorowania budżetów domowych jest dostarczenie niezbędnych informacji na potrzeby analizy i kontrolowania sytuacji społecznej gospodarstw, w tym zwłaszcza struktury ich dochodów i wydatków. W 2012 r. badaniem objęto 4704 słowackich gospodarstw domowych, przy czym 1098 z nich stanowiły gospodarstwa jednoosobowe (23,3% całkowitej liczby gospodarstw badanych w 2012 r.). Do badań wykorzystano analizę wariancji, testy równości średnich brzegowych oraz analizę korelacji i regresji. Niezbędne obliczenia przeprowadzono za pomocą oprogramowania SAS Enterprise Guide.

Słowa kluczowe: badanie budżetów gospodarstw domowych, przychód pieniężny brutto, średnie brzegowe, analiza regresji.