

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

4/2014

ročník/volume 24

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 2

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 12 – 27

Dátum vydania/Publication date: 15. október 2014/October 15, 2014



Erik ŠOLTÉS

Katedra štatistiky Fakulty hospodárskej informatiky, Ekonomická univerzita v Bratislave

Ondrej DÚŽIK

Katedra štatistiky Fakulty hospodárskej informatiky, Ekonomická univerzita v Bratislave

**MODELOVANIE ZÁVISLOSTI HRUBÝCH PEŇAŽNÝCH PRÍJMOV
JEDNOČLENNÝCH DOMÁCNOSTÍ NA SLOVENSKU
OD RELEVANTNÝCH FAKTOROV**

**MODELLING DEPENDENCE OF GROSS MONEY INCOME OF SINGLE PERSON
HOUSEHOLDS IN SLOVAKIA FROM RELEVANT FACTORS**

ABSTRAKT

V príspevku sa zameriavame na modelovanie závislosti hrubých peňažných príjmov slovenských jednočlenných domácností od relevantných faktorov. Hlavným cieľom je zostrojiť vhodný regresný model, prostredníctvom ktorého môžeme kvantifikovať vplyv relevantných faktorov na hrubé peňažné príjmy týchto domácností. Článok poskytuje výsledky našej analýzy, ktoré sme získali zo zisťovania Štatistika rodinných účtov, ktoré realizoval Štatistický úrad SR v roku 2012. Na naše účely sme použili viacfaktorovú analýzu rozptylu, testy zhody marginálnych stredných hodnôt, regresnú a korelačnú analýzu. Tieto štatistické metódy boli aplikované prostredníctvom SAS Enterprise Guide.

ABSTRACT

In this article we focus on modelling dependence of gross money income of Slovak single person households from relevant factors. The main goal is to construct the appropriate regression model through which we can quantify the influence of relevant factors on gross money income of these households. The article provides results of our analysis obtained from the Household Budget Survey realized by the Statistical Office of the Slovak Republic in 2012. For our purposes we used multiple analysis of variance, tests for the equality of marginal means, regression and correlation analysis. These statistical methods were applied using the SAS Enterprise Guide.

KLÚČOVÉ SLOVÁ

štatistika rodinných účtov, hrubé peňažné príjmy domácností, marginálne stredné hodnoty, regresná analýza

KEY WORDS

household budget survey, gross money household income, marginal means, regression analysis

1. ÚVOD

Príspevok vyhodnocuje výsledky analýz závislosti hrubých peňažných príjmov (HPP) jednočlenných slovenských domácností od relevantných faktorov. Analýzy boli urobené na databáze prierezových údajov zo zisťovania Štatistika rodinných účtov z roku 2012. Štatistika rodinných účtov (ďalej „ŠRÚ“) je pravidelné výberové

zistovanie o príjmoch a výdavkoch súkromných domácností. Základným cieľom rodinných účtov je poskytovať informácie na analýzy a monitorovanie sociálnej situácie domácností, predovšetkým na štruktúru ich výdavkov a príjmov. Použitú databázu poskytol Štatistický úrad SR. Do zisťovania v roku 2012 bolo zapojených 4 704 náhodne vybraných slovenských domácností. V článku sme sa zamerali len na jednočlenné domácnosti, ktorých bolo 1 098 a predstavovali 23,3 % z celkového počtu domácností zapojených do ŠRÚ 2012.

Hlavným cieľom výskumu, ktorého výsledky sú prezentované v príspevku, bolo kvantifikovať vplyv relevantných faktorov na hrubé peňažné príjmy jednočlenných domácností. Na tento účel bolo potrebné štatistickými nástrojmi vyselektovať z množiny potenciálnych faktorov tie, ktoré v roku 2012 signifikantne vplývali na hrubé peňažné príjmy jednočlenných domácností, skonštruovať adekvátny regresný model vystihujúci závislosť príjmov od vybraných faktorov a na základe takéhoto modelu kvantifikovať ich vplyv, prípadne odhadnúť mieru, akou ovplyvňujú príjmy domácností. Analýzy sa realizovali prostredníctvom procedúr PROC GLM a PROC REG v aplikácii Enterprise Guide štatisticko-analytického softvéru SAS.

2. VÝBER RELEVANTNÝCH FAKTOROV

Do analýzy sme zahrnuli tieto vysvetľujúce premenné¹:

Tabuľka č. 1: Zoznam uvažovaných vysvetľujúcich premenných

Označenie	Vysvetlenie
KRAJ	administratívne členenie SR, resp. VÚC
VELKOBCE	členenie obcí podľa počtu obyvateľov
TYPOBCE	členenie obcí: <i>krajské mesto; ostatné mestá; obec</i>
VEK	kvantitatívna premenná
POHL	pohlavie
RSTAV	rodinný stav
VZDEL	najvyššie ukončené vzdelanie
SEA	súčasná ekonomická aktivita
POST_ZAM	postavenie v zamestnaní
DPD	dĺžka pracovnej doby
TPD	typ pracovnej doby

Zdroj údajov: vlastné spracovanie

Na výber relevantných regresorov sme použili metódu postupnej eliminácie. Z úplného modelu, v ktorom sa uvažovali všetky vysvetľujúce premenné uvedené v tabuľke č. 1, sa postupne vylučovali premenné, ktorých prínos na vysvetlenie variability hrubých peňažných príjmov jednočlenných domácností bol štatisticky nevýznamný na zvolenej hladine významnosti 0,1. Takto sme zistili, že príjmy jednočlenných domácností neboli ovplyvnené vekom, veľkosťou obce a typom obce, v ktorej domácnosť (osoba) žije, postavením v zamestnaní a zároveň ani typom pracovnej doby zamestnanca. Pri ostatných premenných, ktoré boli ponechané

¹ Podrobnejšie v prílohe A.

v regresnom modeli, sa potvrdilo, že významnou mierou prispievajú k vysvetleniu variability príjmov jednočlenných domácností. Výsledky testov významnosti vplyvu relevantných faktorov sú uvedené v tabuľke č. 2.

Tabuľka č. 2: Overenie štatistickej významnosti prínosu premenných k vysvetleniu variability hrubých peňažných príjmov jednočlenných domácností

Source	DF	Type III SS	Mean Square	F Value	Pr > F
KRAJ	1	4 081 309.63	4 081 309.63	49.12	<.0001
POHL	1	756 574.91	756 574.91	9.10	0.0026
RSTAV	2	1 072 155.06	536 077.53	6.45	0.0016
VZDEL	4	10 102 794.30	2 525 698.57	30.39	<.0001
SEA	2	3 459 673.02	1 729 836.51	20.82	<.0001
DPD	2	2 582 668.85	1 291 334.42	15.54	<.0001

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Pri analýze hrubých peňažných príjmov jednočlenných domácností sme sa zamerali aj na regionálne rozdiely. Na základe testov zhody marginálnych stredných hodnôt (*least squares means*) príjmov jednočlenných domácností (tab. č. 3) sme zistili, že na hladine významnosti 0,1, ale aj na 0,05 sú príjmy jednočlenných domácností v Bratislavskom kraji štatisticky významne odlišné od príjmov jednočlenných domácností v ostatných krajoch Slovenska (všetky *p*-hodnoty uvedené v poslednom riadku tab. 3 sú menšie ako hladina významnosti). Naopak, medzi žiadnou inou dvojicou krajov SR sa nepotvrdil štatisticky významný rozdiel v marginálnych stredných hodnotách príjmov jednočlenných domácností.

Tabuľka č. 3: Tabuľka *p*-hodnôt pre test štatistickej významnosti zhody marginálnych stredných hodnôt pre faktor KRAJ

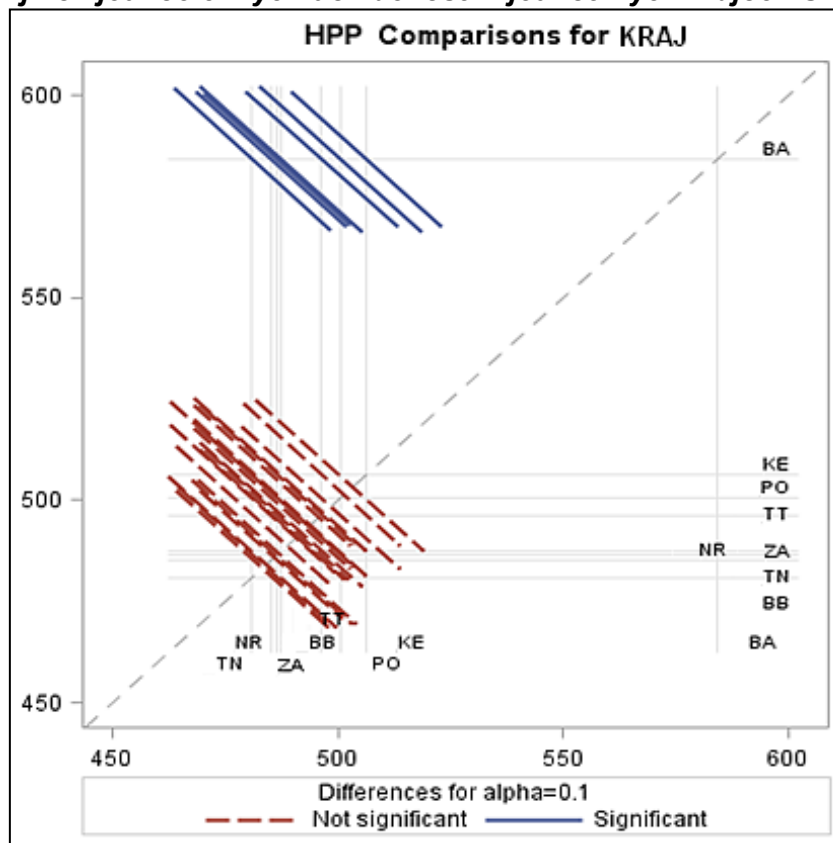
Least Squares Means for effect KRAJ Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: HPP								
i/j	KE	PO	BB	ZA	NR	TN	TT	BA
KE		0.8010	0.3422	0.4116	0.3208	0.2534	0.6422	0.0001
PO	0.8010		0.5037	0.5707	0.4728	0.3778	0.8461	0.0002
BB	0.3422	0.5037		0.9721	0.9511	0.7830	0.6105	<.0001
ZA	0.4116	0.5707	0.9721		0.9293	0.7812	0.6765	<.0001
NR	0.3208	0.4728	0.9511	0.9293		0.8316	0.5705	<.0001
TN	0.2534	0.3778	0.7830	0.7812	0.8316		0.4544	<.0001
TT	0.6422	0.8461	0.6105	0.6765	0.5705	0.4544		<.0001
BA	0.0001	0.0002	<.0001	<.0001	<.0001	<.0001	<.0001	

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

90 % intervalové odhady marginálnych stredných hodnôt a overenie štatistickej významnosti rozdielov týchto stredných hodnôt medzi jednotlivými dvojicami krajov je znázornené v grafe č. 1. Tento graf taktiež jasne potvrdzuje, že priemerné príjmy

jednočlenných domácností v Bratislavskom kraji sú signifikantne vyššie ako vo všetkých ostatných krajoch a medzi žiadnou ďalšou dvojicou krajov už nie je na hladine významnosti 0,1 štatisticky významný rozdiel v marginálnych stredných hodnotách hrubých peňažných príjmov jednočlenných domácností.

Graf č. 1: 90 % intervalové odhady pre marginálne stredné hodnoty hrubých peňažných príjmov jednočlenných domácností v jednotlivých krajoch SR v roku 2012



Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Vychádzajúc z týchto zistení sme vytvorili premennú *KRAJ* s dvomi obmenami: *BA* a *Ostatne_kraje*. Na základe bodových odhadov marginálnych stredných hodnôt, ktoré sú uvedené v tabuľke č. 4, je zrejmé, že za predpokladu, že všetky ostatné vysvetľujúce premenné zahrnuté v regresnom modeli fixujeme, tak priemerné hrubé peňažné príjmy jednočlenných domácností v roku 2012 boli v Bratislavskom kraji oproti ostatným krajom vyššie v priemere o 66,78 €. Na základe uvedenej *p*-hodnoty je tento rozdiel štatisticky významný na každej bežne používanej hladine významnosti.

Tabuľka č. 4: Bodové odhady a test zhody marginálnych stredných hodnôt hrubých peňažných príjmov jednočlenných slovenských domácností v roku 2012 pre Bratislavský kraj a ostatné kraje

KRAJ	HPP LSMEAN	H0:LSMean1=LSMean2
		Pr > t
BA	608.94	<.0001
Ostatne_kraje	542.16	

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

3. ANALÝZA REZÍDUIÍ A OVERENIE PODMIENOK KLASICKÉHO LINEÁRNEHO REGRESNÉHO MODELU

Aby sme sa vyhli skresleným výsledkom v dôsledku odľahlých a vplyvných pozorovaní (pozri [1], [2], [3], [8]), využili sme Cookovu štatistiku a štatistiku *DFFIT* na detekciu a následné odstránenie takýchto pozorovaní. Z pôvodného súboru sme odstránili 47 štatistických jednotiek, pri ktorých Cookova štatistika a štatistika *DFFIT* zhodne preukázali, že pozorovanie je vplyvné, a teda výraznou mierou môže skresľovať výsledky. Týchto 47 pozorovaní predstavovalo 4,28 % z celého rozsahu súboru a rozsah súboru sa zredukoval na 1 051 štatistických jednotiek.

Ďalší aspekt, ktorý môže negatívne ovplyvniť kvalitu modelu a špeciálne interpretovateľnosť regresných koeficientov v odhadnutom regresnom modeli, je vysoký stupeň multikolinearity. Hoci index podmienenosti mal hodnotu 28,672, čo poukazovalo na slabý stupeň multikolinearity, hodnoty charakteristiky *VIF* (*variance inflation factor*) pri obmenách premennej *súčasná ekonomická aktivita* a *dĺžka pracovnej doby* nadobúdali relatívne vysoké hodnoty (6,025; 12,692 a 9,024). Podľa Marquardtovho kritéria je hodnota väčšia ako 10 signálom vysokej multikolinearity. Niektorí autori považujú za maximálne tolerovanú hodnotu charakteristiky *VIF* hodnotu 5 a niektorí dokonca 4. Silný stupeň závislosti medzi premennými *súčasná ekonomická aktivita* a *dĺžka pracovnej doby* sme diagnostikovali aj na základe charakteristiky *Variance proportion*. Najväčšou mierou k multikolinearite teda prispievali premenné *súčasná ekonomická aktivita* a *dĺžka pracovnej doby*, čo je pochopiteľné, keďže osoby bez pracovnej doby tvorili väčšinu osôb, na ktoré v rámci zisťovania Štatistika rodinných účtov nebolo možné aplikovať premennú *dĺžka pracovnej doby* a súčasne v rámci premennej *SEA* mali status *nezamestnaný* alebo *ekonomicky neaktívny*. Na základe týchto zistení sme z uvedených premenných vytvorili novú premennú *SEA – súčasná ekonomická aktivita* – s obmenami uvedenými v tabuľke č. 5.

Tabuľka č. 5: Nové kategórie premennej *súčasná ekonomická aktivita*

SEA – súčasná ekonomická aktivita	
OBMENA	VYSVETLENIE
SEA_nezamestnany	nezamestnaná osoba
SEA_neaktivny	ekonomicky neaktívna osoba
SEA_zamestn_PD_plna	zamestnaná osoba na plný úväzok
SEA_zamestn_PD_skrat.	zamestnaná osoba na skrátený úväzok
SEA_zamestn_PD_neaplik.	zamestnaná osoba s neaplikovateľnou dĺžkou pracovnej doby

Zdroj údajov: vlastné spracovanie

V ďalšej analýze nahradila takto definovaná premenná pôvodné 2 premenné.

Grafická analýza studentizovaných rezíduí² (graf č. 2) poukázala na problém heteroskedasticity náhodnej zložky. Nesplnenie predpokladu o homoskedasticite

² Na overenie predpokladov o náhodnej zložke regresného modelu sme pred klasickými rezíduami uprednostnili studentizované rezíduá. Výhody použitia studentizovaných rezíduí sú uvedené napr. v publikáciách [5] a [7].

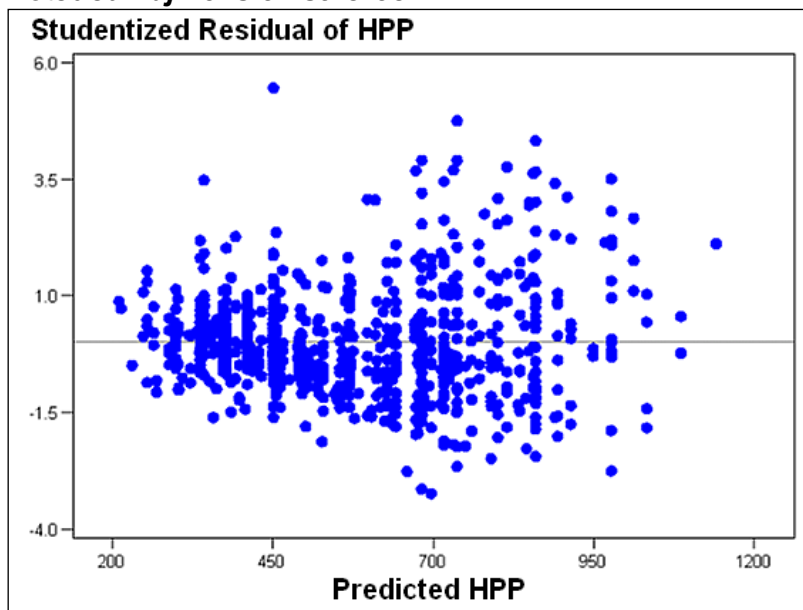
náhodnej zložky sa potvrdilo aj White testom (tab. č. 6), v ktorom sme na základe p -hodnoty ($p < \alpha$) zamietli nulovú hypotézu o homoskedasticite.

Tabuľka č. 6: Overenie homoskedasticity náhodnej zložky

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
64	127.55	<.0001

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Graf č. 2: Grafická analýza studentizovaných rezíduí pre model odhadnutý jednoduchou metódou najmenších štvorcov



Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

4. ODHAD MODELU VÁŽENOU METÓDOU NAJMENŠÍCH ŠTVORCOV

Podľa záveru z predchádzajúcej časti bola náhodná zložka v uvažovanom modeli heteroskedastická, a preto sme na odhad regresného modelu vystihujúceho závislosť hrubých peňažných príjmov jednočlenných domácností od uvedených faktorov použili váženú metódu najmenších štvorcov. Aplikovali sme metódu FGLS (pozri [9]), ktorá poskytuje flexibilný prístup k modelovaniu heteroskedasticity. V tomto modeli sa ukázalo, že pohlavie nemá štatisticky významný vplyv na príjem jednočlennej domácnosti, no na druhej strane sa potvrdilo, že príjem jednočlennej domácnosti je signifikantne ovplyvnený typom obce, v ktorej domácnosť žije. Výsledný odhad regresného modelu je uvedený v tabuľke č. 7.

Model odhadnutý váženou metódou najmenších štvorcov je ako celok štatisticky významný p -value 0,0001 a všetky vysvetľujúce premenné zaradené do regresného modelu na hladine významnosti 0,05 štatisticky významne ovplyvňujú príjem jednočlenných domácností. Premenné vo výstupe uvedenom v tabuľke č. 7 sú zoradené zostupne podľa prínosu k vysvetleniu variability vysvetľovanej premennej (pozri stĺpec *Squared Semi-partial Corr Type I*). Najväčšou mierou k vysvetleniu variability hrubých peňažných príjmov jednočlenných domácností teda prispieva premenná *súčasná ekonomická aktivita*, a to 37,7 %, potom nasleduje *vzdelanie*,

ktoré dodatočne prispieva k vysvetleniu variability príjmov jednočlenných domácností 7,3 %. Parciálny prínos *rodinného stavu* je 6,1 % k vysvetleniu variability príjmov jednočlenných domácností. Parciálne prínosy ďalších relevantných premenných: *kraj*, *typ obce* sú v uvedenom poradí takéto: 2,1 % a 0,3 %. Napriek tomu, že z uvažovanej množiny vysvetľujúcich premenných sa metódami výberu vysvetľujúcich premenných zaradili do regresného modelu len kategoriálne premenné, model vysvetľuje viac ako 50 % (presnejšie 53,37 %, pozri *R-square* – viacnásobný koeficient determinácie) variability hrubých peňažných príjmov jednočlenných slovenských domácností.

Tabuľka č. 7: Odhad regresného modelu pre hrubé peňažné príjmy jednočlenných domácností získaný váženou metódou najmenších štvorcov

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	103 267 615	8 605 635	99.01	<.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-Sq	0.5283
Coeff Var	69.8381		

Parameter estimates						
Variable	Parameter Estimate	Standard Error	t Value	Pr > t	Squared Semi-partial Corr Type I	Variance Inflation
Intercept	614.71	14.6051	42.09	<.0001	.	0
SEA_Nezamestnany	-419.03	17.3750	-24.12	<.0001	0.0536	1.8529
SEA_neaktivny	-321.94	13.0994	-24.58	<.0001	0.2770	2.5649
SEA_Zamestn_PD_skrat.	-193.39	21.1943	-9.12	<.0001	0.0365	1.4576
SEA_Zamestn_PD_neapl	-128.31	31.2023	-4.11	<.0001	0.0096	1.1759
VZDEL_VS	222.22	18.4525	12.04	<.0001	0.0457	1.1723
VZDEL_BC	93.73	30.1596	3.11	0.0019	0.0007	1.0741
VZDEL_SS	71.97	8.2633	8.71	<.0001	0.0205	1.7379
VZDEL_OU	36.53	8.5158	4.29	<.0001	0.0057	1.5813
RS_ovdoveny	85.6450	6.8389	12.52	<.0001	0.0607	1.2311
BA_kraj	66.78	12.1282	5.51	<.0001	0.0209	1.2914
OBEC_Krajske mesto	20.60	9.4341	2.18	0.0292	0.0008	1.6335
OBEC_lne_mesto	16.23	7.2594	2.23	0.0257	0.0022	1.3278

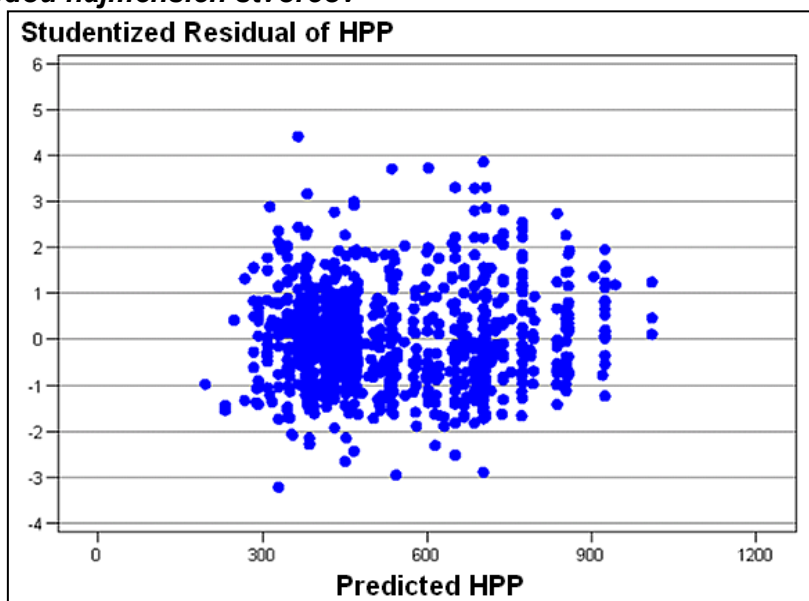
Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Vďaka tomu, že sme z pôvodných premenných *súčasná ekonomická aktivita* a *dĺžka pracovnej doby* vytvorili novú premennú *SEA – súčasná ekonomická aktivita*,

podarilo sa nám eliminovať stupeň multikolinearity. Index podmienenosti sa z pôvodnej hodnoty 28,672 zredukoval na hodnotu 12,126, ktorá poukazuje na veľmi slabý stupeň multikolinearity. Taktiež hodnoty charakteristiky *VIF* (pozri tab. č. 7) sú nízke. Najvyššia hodnota *VIF* (2,565) je pri umelej premennej *SEA_neaktivny* a jej druhá odmocnina nás informuje o tom, že štandardná odchýlka príslušného regresného koeficienta je 1,602-násobne väčšia, ako by bola, ak by táto vysvetľujúca premenná nebola korelovaná s ostatnými vysvetľujúcimi premennými. Keďže žiadna z hodnôt neprekračuje referenčné hodnoty, ktoré by poukazovali na nežiaduco vysoký stupeň multikolinearity (pozri časť 3), skreslenia v dôsledku multikolinearity sú zanedbateľné.

Ak sa pozrieme na graf studentizovaných rezíduí (graf č. 3), vidíme, že váhy stanovené FGLS boli vhodné na elimináciu heteroskedasticity (porovnaj s grafom č. 2). Dokonca sa redukovali aj extrémne vysoké a extrémne nízke studentizované rezíduá (väčšie ako 3, resp. menšie ako -3). Napriek tomu, že pre veľké výberové súbory nie je nevyhnutné splnenie normality náhodných chýb (pozri [9]), je vhodné, ak sa ich rozdelenie približuje k normálnemu rozdeleniu. Na základe Kolmogorovovho-Smirnovovho testu ($p = 0,020$) a Cramerovho-von Misesovho testu ($p = 0,014$), ktoré sú uvedené v tabuľke 8, na hladine významnosti 0,01 nezamietame nulovú hypotézu o normálnom rozdelení náhodnej zložky regresného modelu. Durbinova-Watsonova štatistika (pozri tab. č. 9) sa nachádza medzi hornou a dolnou hranicou (pre rozsah výberového súboru $n = 1050$, počet parametrov regresného modelu $p = 15$ a pre hladinu významnosti $\alpha = 0,05$ sú to hodnoty $dw_L = 1,872$ a $dw_U = 1,926$), preto Durbinov-Watsonov test nevedie k jednoznačnému záveru. Podľa testu štatistickej významnosti koeficienta autokorelácie 1. stupňa na hladine významnosti 0,05 nulovú hypotézu o nezávislosti náhodných chýb nezamietame ($p = 0,0645$, vlastný výpočet). Odhadnutý regresný model teda môžeme považovať za adekvátny.

Graf č. 3: Grafická analýza studentizovaných rezíduí pre model, ktorý bol odhadnutý váženou metódou najmenších štvorcov



Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Tabuľka č. 8: Testovanie normálneho rozdelenia náhodnej zložky regresného modelu

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.0471	Pr > D	0.020
Cramer-von Mises	W-Sq	0.7026	Pr > W-Sq	0.014

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Tabuľka č. 9: Durbinova-Watsonova štatistika a výberový koeficient autokorelácie 1. st

Durbin-Watson D	1.884
Number of Observations	1051
1st Order Autocorrelation	0.057

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Tabuľka č. 10: Bodové a intervalové odhady parametrov regresného modelu

Variable	Parameter Estimate	90 % Confidence Limits	
Intercept	614.71	590.6632	638.7525
SEA_Nezamestnany	-419.03	-447.6318	-390.4220
SEA_neaktivny	-321.94	-343.5020	-300.3704
SEA_Zamestn_PD_skrat.	-193.39	-228.2780	-158.4928
SEA_Zamestn_PD_neapl	-128.31	-179.6741	-76.9359
VZDEL_VS	222.22	191.8397	252.5973
VZDEL_BC	93.73	44.0738	143.3785
VZDEL_SS	71.97	58.3686	85.5766
VZDEL_OU	36.53	22.5058	50.5453
RS_ovdoveny	85.64	74.3859	96.9041
BA_kraj	66.78	46.8146	86.7485
OBEC_Krajske mesto	20.60	5.0718	36.1350
OBEC_Ine_mesto	16.23	4.2653	28.1678

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

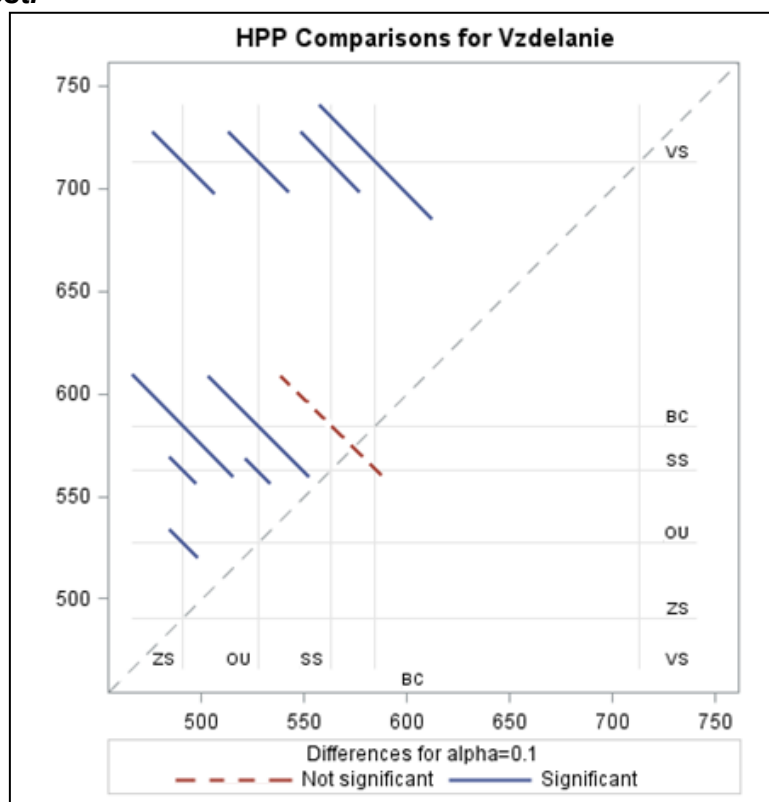
Lokujúca konštanta b_0 614,71 predstavuje priemerné hrubé peňažné príjmy (v €) jednočlenných domácností, v ktorých je člen tejto domácnosti zamestnaný na plný pracovný úväzok, má základné vzdelanie, je slobodný alebo rozvedený a žije na vidieku v inom ako v Bratislavskom kraji.

Ak sa zameriame na súčasnú ekonomickú aktivitu osoby jednočlennej domácnosti, tak najvyšší príjem majú jednočlenné domácnosti, v ktorých je osoba zamestnaná na plný úväzok. Ak jednočlennú domácnosť tvorí zamestnaná osoba na skrátenej úväzok, jej príjem je v priemere o 193,39 € nižší. Prirodzene najnižší príjem majú tie jednočlenné domácnosti, v ktorých je nezamestnaná osoba. Ich príjem je v priemere o 419,03 € nižší ako príjem jednočlenných domácností so zamestnanou osobou na plný úväzok. Uvedené interpretácie platia za podmienky ceteris paribus, t. j. za predpokladu nemennosti hodnôt ostatných vysvetľujúcich premenných v regresnom modeli.

Najnižšie hrubé peňažné príjmy mali v roku 2012 tie jednočlenné domácnosti, v ktorých mala osoba základné vzdelanie. Ak jednočlennú domácnosť tvorila osoba s odborným učňovským vzdelaním, príjem bol v priemere vyšší o 36,53 €, ak ju tvorila osoba so stredoškolským vzdelaním, príjem bol v priemere vyšší o 71,97 €, ak ju tvorila osoba s vysokoškolským vzdelaním, príjem bol v priemere vyšší o 222,22 €.

Ak sa pozrieme na intervalové odhady regresných koeficientov, tak zisťujeme, že intervaly spoľahlivosti pre regresné koeficienty prislúchajúce k stredoškolskému vzdelaniu a k bakalárskemu vzdelaniu majú veľký prienik, a teda v roku 2012 nebol významný rozdiel v hrubých peňažných príjmoch jednočlenných domácností so stredoškolsky vzdelanou osobou a domácnosťami s osobou s bakalárskym vzdelaním. Toto zistenie potvrdzuje aj graf 4.

Graf č. 4: 90 % intervalové odhady pre marginálne stredné hodnoty hrubých peňažných príjmov jednočlenných domácností v roku 2012 v závislosti od vzdelania člena domácnosti



Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Tabuľka č. 11: Bodové a intervalové odhady marginálnych stredných hodnôt hrubých peňažných príjmov jednočlenných slovenských domácností v roku 2012 v závislosti od vzdelania člena domácnosti

Vzdelanie	HPP LSMEAN	90% Confidence Limits	
VS	712.88	681.78	743.98
BC	584.39	535.19	633.59
SS	562.64	546.74	578.53
OU	527.19	510.55	543.83
ZS	490.66	472.60	508.72

Zdroj údajov: vlastné výpočty v SAS EG na základe údajov ŠRÚ 2012

Pochopiteľne, s nárastom vzdelania príjem rastie, a teda najnižší príjem majú jednočlenné domácnosti, ktorých člen má základné vzdelanie, a najvyšší príjem majú tie, v ktorých má člen vysokoškolské vzdelanie. Porovnaním odhadov marginálnych stredných hodnôt (napr. pre vysokoškolské vzdelanie 712,88 € a pre základné vzdelanie 490,66 €) získame rozdiely, ktoré korešpondujú s regresnými koeficientmi, ktoré sme interpretovali pre premennú *vzdelanie* (napr. pozri regresný koeficient 222,22 € pri premennej *VZDEL_VS*).

Čo sa týka rodinného stavu osoby v jednočlennej domácnosti, tak v porovnaní so slobodnými a rozvedenými mali ovdovení v roku 2012 v priemere o 85,64 € vyšší hrubý príjem, opäť za podmienky *ceteris paribus*.

Dominanciu príjmov v Bratislavskom kraji sme rozoberali v 2. časti článku a regresný koeficient v tabuľke 10 potvrdil, že v roku 2012 mali jednočlenné domácnosti v Bratislavskom kraji v priemere o 66,78 € vyššie hrubé príjmy ako jednočlenné domácnosti z iných krajov.

Aj keď v súčasnosti mnoho ľudí z vidieka a menších miest dochádza za prácou do väčších miest, potvrdilo sa, že je štatisticky významný rozdiel medzi príjmami jednočlenných domácností v krajských mestách a na vidieku. V roku 2012 mali jednočlenné domácnosti žijúce v krajských mestách v priemere o 20,60 € vyššie hrubé príjmy ako jednočlenné domácnosti žijúce na vidieku.

5. ZÁVER

Cieľom analýzy, ktorej výsledky sú prezentované v článku, bolo zistiť, ktoré faktory sledované v rámci zisťovania Štatistika rodinných účtov v roku 2012 štatisticky významne ovplyvňovali hrubé peňažné príjmy domácností, a následne kvantifikovať ich vplyv na tieto príjmy. Keďže z pohľadu príjmov majú rôzne typy domácností svoje špecifiká, rozhodli sme sa zamerať len na jeden typ domácností, a to jednočlenné domácnosti. Príjmy domácností sú ovplyvňované viacerými faktormi, preto bolo nevyhnutné siahnuť po viacrozmerných štatistických metódach. Na dosiahnutie cieľa sme použili hlavne viacfaktorovú analýzu rozptylu a viacnásobnú regresnú analýzu. Tieto metódy nám umožnili kvantifikovať vplyv jednotlivých relevantných faktorov za podmienky, že všetky ostatné faktory (zahrnuté v modeli) sú fixované. Vďaka odhadom marginálnych stredných hodnôt a regresných koeficientov sme teda získali porovnateľné hodnoty, t. j. také hrubé peňažné príjmy pre jednotlivé kategórie faktorov, ktoré uvažujú nemenné hodnoty ostatných faktorov – *ceteris paribus*.

Z množiny potenciálnych regresorov sme metódou postupnej eliminácie vyseletovali tie, ktoré štatisticky významne ovplyvňovali hrubé peňažné príjmy jednočlenných domácností v SR. Z dôvodu nesplnenia podmienky homoskedasticity náhodných chýb sme na odhad regresného modelu použili váženú metódu najmenších štvorcov, pričom váhy sme vypočítali metódou FGLS (*feasible general least squares*). Tieto váhy sme použili aj v modeli viacfaktorovej analýzy rozptylu.

Na základe odhadnutého regresného modelu sme zistili, že hrubé peňažné príjmy jednočlenných domácností v roku 2012 boli ovplyvnené hlavne statusom ekonomickej aktivity, dĺžkou pracovnej doby a vzdelaním člena domácnosti. Najnižší hrubý mesačný príjem mali, pochopiteľne, domácnosti s nezamestnanou osobou so základným vzdelaním. Jednočlenná domácnosť s nezamestnanou osobou mala

v priemere o 419 € nižší hrubý mesačný príjem ako jednočlenná domácnosť, ktorej člen bol zamestnaný na plný pracovný úväzok (*ceteris paribus*).

Potvrdilo sa, že s vyšším stupňom dosiahnutého vzdelania príjem domácnosti rastie. Jednočlenné domácnosti, ktorých člen mal vysokoškolské vzdelanie 2. alebo 3. stupňa, mali v roku 2012 v priemere o 222 € vyšší hrubý mesačný príjem ako jednočlenné domácnosti, v ktorých mal člen základné vzdelanie. Nepotvrdil sa štatisticky významný rozdiel medzi príjmami jednočlenných domácností, kde mal člen domácnosti ukončený bakalársky stupeň vysokoškolského vzdelania, a príjmami domácností so stredoškolsky vzdelanou osobou. Tento výsledok pravdepodobne súvisí s tým, že bakalárske štúdium na slovenských vysokých školách nie je orientované na praktické zručnosti, ale viac-menej poskytuje teoretické poznatky a pripravuje na 2. stupeň vysokoškolského vzdelania. Zamestnávateľa preto nemajú veľa dôvodov na vyššie finančné ohodnotenie bakalárov v porovnaní so zamestnancami so stredoškolským vzdelaním.

Analýzou sa potvrdili regionálne disparity v príjmoch jednočlenných domácností medzi Bratislavským krajom a zvyškom Slovenska. Jednočlenné domácnosti žijúce v Bratislavskom kraji mali v roku 2012 v priemere o 67 € vyšší hrubý mesačný príjem ako jednočlenné domácnosti z iných krajov. Toto porovnanie je za predpokladu rovnakého statusu ekonomickej aktivity, rovnakého vzdelania, rovnakého rodinného stavu, ako aj rovnakého typu obce, v ktorej domácnosť žije. Poznamenajme, že pri použití „klasických“ aritmetických priemerov by sme získali iné výsledky, ktoré však nie sú porovnateľné, keďže v rôznych krajoch je rôzne zastúpenie domácností z hľadiska vzdelania, ekonomickej aktivity člena domácnosti, ako aj ďalších relevantných faktorov.

Malú, ale štatisticky významnú časť variability príjmov jednočlenných domácností vysvetľovali ešte premenné *rodinný stav člena domácnosti* a *typ obce*, v ktorej domácnosť žila. Vplyv týchto faktorov bol podobne ako vplyv predchádzajúcich regresorov kvantifikovaný prostredníctvom regresných koeficientov odhadnutého modelu.

Odhadnutým regresným modelom sa podarilo vysvetliť 53,37 % variability príjmov jednočlenných domácností v SR v roku 2012. Treba však poznamenať, že v modeli nebolo zohľadnené odvetvie hospodárstva, v ktorom je člen jednočlennej domácnosti zamestnaný, a jeho pracovná pozícia. Obidve premenné ovplyvňujú príjem a ich zaradenie do regresného modelu by prispelo k vysvetleniu variability príjmov jednočlenných domácností.

PRÍLOHA A

Premenná	Početnosť	Katégorie	Popis/Označenie v databáze ŠRU 2012
HPP	Hrubé peňažné príjmy domácností na Slovensku v €		
KRAJ	Administratívny región		
	214	BA	Bratislavský kraj
	145	TT	Trnavský kraj
	120	TN	Trenčiansky kraj
	151	NR	Nitriansky kraj
	97	ZA	Žilinský kraj
	155	BB	Banskobystrický kraj
	105	PO	Prešovský kraj
REF.	111	KE	Košický kraj
VELKOBCE	Veľkosť obce		Počet obyvateľov
	138	VELKOBCE_1	do 999
	101	VELKOBCE_2	1 000 - 1 999
	95	VELKOBCE_3	2 000 - 4 999
	55	VELKOBCE_4	5 000 - 9 999
	105	VELKOBCE_5	10 000 - 19 999
	219	VELKOBCE_6	20 000 - 49 999
	156	VELKOBCE_7	50 000 - 99 999
REF.	229	VELKOBCE_8	100 000 a viac
TYPOBCE	Typ obce		
	356	TYPOBCE_KRM	Krajské mesto
	400	TYPOBCE_OSTM	Ostatné mestá
REF.	342	TYPOBCE_OBEC	Obec
VEK			
POHL	Pohlavie		
	175	M	Muž
REF.	923	Z	Žena
RSTAV	Rodinný stav		Označenie v pôvodnej databáze SRU
	304	ROZV	5. Rozvedený/á (a znovu neženatý/nevydatá); 6. Ženatý/Vydatá (bez partnera v domácnosti)
	457	VDOV	4. Vdovec/Vdova (a znovu neženatý/nevydatá)
REF.	337	SLOB	1. Slobodný/á (nikdy neženatý/nevydatá)
VZDEL	Najvyššie dokončené vzdelanie		Označenie v pôvodnej databáze SRU
	146	VS	13. Vysokoškolské 2. stupeň 14. Vysokoškolské 3. stupeň
	33	BC	12. Vysokoškolské 1. stupeň
	487	SS	08. Úplné stredné všeobecné 09. Úplné stredné odborné 10. Vyššie odborné (pomaturitné) 11. Vyššie odborné neuniverzitné štúdium
	278	OU	03. Učňovské bez maturity kratšie ako dvojročné 04. Učňovské bez maturity dvojročné a dlhšie 05. Stredné bez maturity kratšie ako 2 roky 06. Stredné bez maturity dvojročné a dlhšie 07. Učňovské s maturitou
REF.	154	ZS	00. Bez vzdelania 01. Základné 1. stupeň 02. Základné 2. stupeň

PRÍLOHA A – POKRAČOVANIE

Premenná	Početnosť	Katégorie	Popis/Označenie v databáze ŠRU 2012
SEA	Súčasná ekonomická aktivita		
	36	NEZAM	Označenie v pôvodnej databáze SRU 3. Nezamestnaný
	653	ENEAO	4. Nepracujúci starobný dôchodca 5. Študent, učeň 6. Ekonomicky neaktívny, žena v domácnosti 7. Neschopný práce 9. Neaplikovateľné (nezaopatrené dieťa, nezaradené do 5.)
REF.	409	PRAC	1. Pracujúci (plný aj čiastočný úväzok) 2. Zamestnaný, ale dočasne mimo práce 8. Pracujúci starobný dôchodca
POST_ZAM	Postavenie v zamestnaní		
	598	neaplikovateľne	8. Neaplikovateľné (vzhľadom na vek)
	91	neklasifikovane	6. Osoby neklasifikované podľa postavenia
	1	zamestnavateľ	1. Zamestnávateľ
	35	samozamestnaný	2. Samozamestnaný
REF.	373	zamestnaný	3. Zamestnanec
DPD	Dĺžka pracovnej doby		
	725	neaplik	8. Neaplikovateľné
	50	skratena	2. Skrátená pracovná doba
REF.	323	plny_uv	1. Plný úväzok
TPD	Typ pracovnej doby		
	725	Neaplik	8. Neaplikovateľné
	9	prilez_praca	3. Príležitostná práca bez kontraktu
	45	urcita_doba	2. Zmluva na dobu určitú
REF.	319	neurcita_doba	1. Zmluva na neurčitú dobu

LITERATÚRA

- [1] BOLLEN, K. A. – JACKMAN, R. W.: Regression diagnostics: An expository treatment of outliers and influential cases. In: Fox, John; and Long, J. Scott (eds.); Modern Methods of Data Analysis. Newbury Park, CA: Sage, 1990. S. 257 – 291.
- [2] COOK, R. D. – WEISBERG, S.: Residuals and influence in regression. New York, NY: Chapman & Hall, 1982. 230 s. ISBN 0-412-24280-0. [toto by malo byť za bibl. č. 3]
- [3] COOK, R. D.: Detection of Influential Observation in Linear Regression. In: Technometrics. Vol. 19, No. 1. American Statistical Association, 1977. S. 15 – 18. Dostupné na internete: <http://www.ime.usp.br/~abe/lista/pdfWiH1zqnMHo.pdf>
- [4] GARSON, G. D.: General Linear Models: Multivariate GLM, MANOVA and MANCOVA. Asheboro, N. C. Statistical Associates Publishers, 2012. 61 s. ISBN (ASIN) B0092WUSQS.
- [5] HEBÁK, P. – HUSTOPECKÝ, J. – MALÁ, I.: Vícerozmerné statistické metódy (2). Praha: Informatorium, 2005. 240 s. ISBN 80-7333-036-9.
- [6] MELOUN, M. – MILITKÝ, J.: Statistická analýza experimentálných dat. 3. vydanie. Statistická analýza experimentálných dat. Praha: Academia, 2004. 980 s. ISBN 80-200-1254-0.
- [7] ŠOLTÉS, E.: Regresná a korelačná analýza s aplikáciami. Bratislava: Iura Edition, 2008. 287 s. ISBN 978-80-8078-163-7.
- [8] VOJTKOVÁ, M. – LABUDOVA, V.: Regionálna analýza výdavkov a príjmov domácností v Slovenskej republike. Ekonomický časopis, roč. 58, č. 8, Bratislava. Ekonomický ústav SAV: Prognostický ústav SAV, 2010, s. 802 – 820.

- [9] WOOLDRIDGE, J. M.: Introductory Econometrics: A modern approach. 4. vydanie. Mason: South-Western Cengage Learning, 2009. 865 s. ISBN 978-0-324-66054-8.

RESUMÉ

Cieľom článku bolo modelovať závislosť hrubých peňažných príjmov jednočlenných slovenských domácností od relevantných faktorov a kvantifikovať ich vplyv prostredníctvom odhadnutého regresného modelu. Analýza vychádza zo zisťovania Štatistika rodinných účtov 2012, ktoré realizoval Štatistický úrad SR.

Na základe použitých metód – viacfaktorová analýza rozptylu, testy zhody marginálnych stredných hodnôt, regresná a korelačná analýza – sme zistili, že hrubý peňažný príjem jednočlenných slovenských domácností bol v roku 2012 štatisticky významne ovplyvnený týmito kvalitatívnymi premennými: súčasná ekonomická aktivita, dĺžka pracovnej doby, najvyššie dosiahnuté vzdelanie, rodinný stav, kraj a typ obce, kde domácnosť žije. Čo sa týka regionálnych rozdielov, zistili sme, že štatisticky významný rozdiel v stredných hodnotách príjmov jednočlenných domácností bol jedine medzi Bratislavským krajom a ostatnými krajmi.

Po odstránení vplyvných pozorovaní sme vážnou metódou najmenších štvorcov odhadli lineárny regresný model v SAS Enterprise Guide. Model spĺňa predpoklady na náhodnú zložku regresného modelu a vykazuje len slabý stupeň multikolinearity. Na základe odhadnutých parciálnych koeficientov korelácie sme posúdili silu vzájomného vzťahu medzi príjmami a jednotlivými vysvetľujúcimi premennými. Vplyv jednotlivých kategoriálnych premenných na hrubé peňažné príjmy jednočlenných slovenských domácností bol kvantifikovaný prostredníctvom marginálnych stredných hodnôt, ako aj regresných koeficientov.

RESUME

The aim of the article was to model dependence of gross money income of Slovak single person households from relevant factors and to quantify the influence of relevant factors by means of the estimated regression model. The analysis is based on a 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 we found out that the gross money income of the Slovak single person households in 2012 was significantly influenced mainly by the following qualitative variables: current economic activity, length of working time, the highest attained level of education, marital status, region, type of municipality where the household lives. Concerning the regional disparities we found out that there was a significant difference in the average income of single person households only between households living in Bratislava region and other regions of.

After eliminating the influential observations we estimated a linear regression model by the weighted least squared method in the SAS Enterprise Guide. The model fulfils assumptions of error terms of regression model and shows only a low degree of multicollinearity. On the basis of the estimated partial correlation coefficients we judged the strength of relationship between income and individual explanatory variables. The effects of particular categorical variables on gross money income of the Slovak single person households were quantified by marginal means as well as by regression coefficients.

PROFESIJNÝ ŽIVOTOPIS

Doc. Mgr. Erik Šoltés, PhD., od roku 1998 pôsobí na Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave. Titul PhD. získal v roku 2004 vo vednom odbore štatistika na FHI EU. V roku 2010 získal vedecko-pedagogickú hodnosť docent pre vedný odbor kvantitatívne metódy v ekonómii. V súčasnosti pôsobí na Fakulte hospodárskej informatiky EU v Bratislave ako prodekan pre rozvoj. V rámci pedagogickej činnosti sa venuje výučbe štatistiky a regresnej a korelačnej analýzy. Jeho vedecká činnosť sa zameriava na aplikáciu teórie kredibility v podmienkach slovenského neživotného poistenia a analýzu sociálno-ekonomických javov s využitím viacrozmerných štatistických metód.

Ing. Ondrej Dúžik, od roku 2008 je študentom na Fakulte hospodárskej informatiky Ekonomickej univerzity v Bratislave. Bakalárske štúdium absolvoval v študijnom programe účtovníctvo. V roku 2013 získal titul Ing. na Katedre štatistiky v študijnom programe štatistické metódy v ekonómii a úspešne ukončil doplňujúce pedagogické štúdium odborných ekonomických predmetov na Katedre pedagogiky Národohospodárskej fakulty Ekonomickej univerzity v Bratislave. Od septembra 2013 je interným doktorandom v študijnom programe kvantitatívne metódy v ekonómii na Katedre štatistiky FHI EU, kde sa aj v rámci svojej dizertačnej práce zaoberá štatistickou analýzou sociálnej situácie obyvateľstva na Slovensku, špeciálne využitím metód analýzy rozptylu. Od roku 2012 je členom Academic tímu v spoločnosti SAS Slovakia.

KONTAKT

solt.es.euba@gmail.com

ondrej.duzik@gmail.com