



ŠTATISTICKÁ ANALÝZA VYBRANÉHO EKONOMICKÉHO ČASOVÉHO RADU

THE STATISTICAL ANALYSIS OF SELECTED ECONOMIC TIME SERIES

Eva OSTERTAGOVÁ

Abstract

Content of this paper consists of practical use of simple exponential smoothing (SES) method. It is one of the most popular forecasting methods that uses weighted moving average of past data as the basis for a forecast. This technique is based on the premise that the level of time series should fluctuate about a constant level or change slowly over time. The accuracy of the SES method strongly depends on the optimal value of the smoothing constant. To determine the optimal value of the smoothing constant was used the optimization method based on the lowest root mean square error.

Key words: simple exponential smoothing, smoothing constant, prediction, mean square error, mean absolute percentage error

Úvod

Dôležitou úlohou štatistickej analýzy ekonomických javov je skúmanie ich dynamiky. Empirické pozorovania v ekonomickej oblasti sú často usporiadané do časového radu, pod ktorým rozumieme dáta (výsledky pozorovaní), ktoré sú chronologicky usporiadané. Máme ale na mysli tzv. štatistické (stochastické) rady, ktoré sú zaťažené neistotou. Cieľom analýzy časového radu je určenie modelu (mechanizmu), podľa ktorého sú generované sledované dáta. Poznanie tohto modelu umožňuje predvídať budúci vývoj systému a do určitej miery aj riadiť a optimalizovať chovanie systému vhodnou voľbou vstupných parametrov a začiatočných podmienok. Výber metódy analýzy časového radu závisí na mnohých faktoroch, ku ktorým patrí: účel analýzy, typ sledovaného časového radu, skúsenosť štatistika, dostupná databáza, softwarové a hardwarové vybavenie.

Základné charakteristiky vybraného ekonomického časového radu

V tabuľke 1 sú údaje o tržbách za vlastné výkony a tovar v priemyselnej výrobe potravín, nápojov a tabakových výrobkov (y_t) v tisícoch € na Slovensku v rokoch 2000 – 2011.

Tab. 1 Tabuľka vstupných hodnôt

Rok	2000	2001	2002	2003	2004	2005
y_t	3514951,5	3933272,0	4095828,2	3961038,3	3949011,3	3938238,1
Rok	2006	2007	2008	2009	2010	2011
y_t	4145873,9	4209879,3	4464052,7	3676812,6	3757702,8	4276256,3

Pre daný intervalový časový rad najprv zostrojíme *spojnicový graf*, ktorý je jedným zo základných prostriedkov prezentácie časových radov. Na horizontálnu os sa vynáša časová



premenná, v našom prípade $t = 1, 2, \dots, 12$. Na vertikálnu os sa vynesú hodnoty časového radu y_t . Získaný spojnicový graf vývoja tržieb je na obrázku 1a).

Majme časový rad y_t , $t = 1, 2, \dots, n$, intervalového ukazovateľa. Celková úroveň hodnôt ukazovateľa časového radu sa charakterizuje najčastejšie pomocou *priemerov*. Priemerná hodnota intervalového časového radu (ak sú intervaly rovnako dlhé) sa vypočíta

pomocou *prostého aritmetického priemeru* $\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$.

Pre uvažovaný časový rad dostaneme $\bar{y} = 3993576,4$. Teda priemerná hodnota tržieb za vlastné výkony a tovar v priemyselnej výrobe potravín, nápojov a tabakových výrobkov v rokoch 2000 – 2011 na Slovensku je 3993576,4 tis. €.

Pre charakterizovanie *dynamiky vývoja* časového radu, t.j. pre skúmanie rýchlosti zmien hodnôt sledovaného ukazovateľa v závislosti na čase, je možné použiť rôzne štatistické charakteristiky. Najjednoduchšou mierou dynamiky je *absolútny prírastok (prvá diferenciacia)*, pre ktorý platí vzťah $\Delta y_t = y_t - y_{t-1}$, $t = 2, 3, \dots, n$. Táto charakteristika vyjadruje absolútny prírastok alebo úbytok skúmaného ukazovateľa v čase t oproti času $t - 1$. Veľmi zjednodušene povedané, prvé diferencie majú pre priebeh časového radu podobný význam ako prvé derivácie pre priebeh spojitej funkcie. Ak došlo oproti predchádzajúcemu obdobiu k nárastu, tak je $\Delta y_t > 0$. Ak došlo oproti predchádzajúcemu obdobiu k poklesu, tak je $\Delta y_t < 0$.

Druhé diferencie $\Delta^2 y_t$ získame ako rozdiel dvoch susedných prvých diferencií. Platí $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, $t = 3, 4, \dots, n$. Druhé diferencie charakterizujú absolútne zrýchlenie, resp. spomalenie vývoja v skúmanom časovom rade. Opäť veľmi zjednodušene povedané, druhé diferencie majú pre priebeh časového radu podobný význam ako druhé derivácie pre priebeh spojitej funkcie. Hodnota $\Delta^2 y_t > 0$ znamená, že priebeh príslušnej trojice hodnôt (v poradí y_{t-2}, y_{t-1}, y_t) má konvexný charakter, kým hodnota $\Delta^2 y_t < 0$ znamená, že priebeh príslušnej trojice hodnôt má konkávny charakter. Analogicky je možné stanoviť diferencie vyšších stupňov (tretieho, štvrtého, atď.).

Veľmi dôležitou mierou dynamiky časových radov je *koeficient rastu* $k_t = \frac{y_t}{y_{t-1}}$, $t = 2, 3, \dots, n$. Charakterizuje relatívnu postupnú rýchlosť zmien hodnôt v časovom rade. Ak došlo oproti predchádzajúcemu obdobiu k nárastu, tak je $k_t > 1$. Ak došlo oproti predchádzajúcemu obdobiu k poklesu, tak je k_t medzi 0 a 1. Ak koeficient rastu vyjadríme v percentách, hovoríme o *tempe rastu* $\tau_t = k_t \cdot 100$ %.

Diferencie a koeficienty rastu sa okrem priameho použitia pre charakterizovanie dynamiky časového radu používajú aj ako jedno z kritérií pre nájdenie vhodnej trendovej funkcie pri modelovaní trendu časového radu. Ak prvé diferencie kolíšu okolo konštanty, dá sa povedať, že sledovaný časový rad má lineárny trend. V špeciálnom prípade, ak kolíšu okolo nuly, rad je bez trendu, t.j. trvale ani nerastie ani neklesá. Ak sú druhé diferencie v podstate konštantné, je možné použiť kvadratický trend. Ak koeficienty rastu časového radu kolíšu okolo konštanty, je pre daný časový rad charakteristický exponenciálny trend.

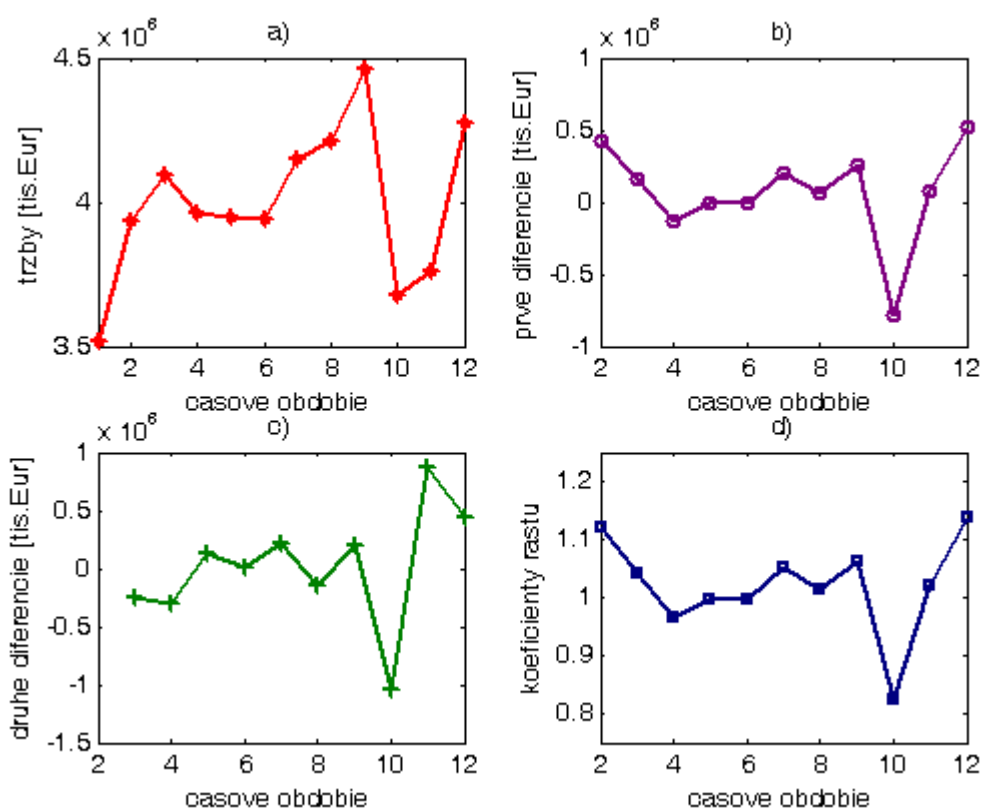
Základné charakteristiky dynamiky pre sledovaný časový rad sú uvedené v tabuľke 2.



Tab. 2 Tabuľka základných charakteristík dynamiky časového radu

t	y_t	Δy_t	$\Delta^2 y_t$	k_t	τ_t [%]
1	3514951,5	—	—	—	—
2	3933272,0	418320,5	—	1,1190	111,90
3	4095828,2	162556,2	-255764,3	1,0413	104,13
4	3961038,3	-134789,9	-297346,1	0,9671	96,71
5	3949011,3	-12027,0	122762,9	0,9970	99,70
6	3938238,1	-10773,2	1253,8	0,9973	99,73
7	4145873,9	207635,8	218409,0	1,0527	105,27
8	4209879,3	64005,4	-143630,4	1,0154	101,54
9	4464052,7	254173,4	190168,0	1,0604	106,04
10	3676812,6	-787240,1	-1041413,5	0,8236	82,36
11	3757702,8	808902,0	868130,3	1,0220	102,20
12	4276256,3	518553,5	437663,3	1,1380	113,80

Na obrázku 1b) je zachytený vývoj prvých diferencií, na obrázku 1c) vývoj druhých diferencií a na obrázku 1d) vývoj koeficientov rastu sledovaného časového radu.



Obr. 1a) – d) Grafické zobrazenie hodnôt y_t , Δy_t , $\Delta^2 y_t$, k_t za roky 2000 – 2011



Výber vhodného modelu pre daný časový rad a možnosti predikcie

Z grafickej analýzy daného časového radu je zrejmé, že nevykazuje sezónnosť a ani žiaden výrazný trend. Pre modelovanie časového radu so značne nerovnomerným vývojom je vhodné aplikovať *Brownov model jednoduchého exponenciálneho vyrovňovania* (SES – *Simple Exponential Smoothing*).

Exponenciálne vyrovňovanie sa zakladá na myšlienke, že pri konštrukcii extrapoláčnej prognózy sa najvyššia váha prideli poslednej hodnote časového radu, pretože táto obsahuje najviac informácií o možných zmenách vo vývoji sledovaného ukazovateľa a očakáva sa aj ich čiastočný prenos do budúcnosti. Odhad trendu je získavaný vo forme lineárnej kombinácie všetkých predchádzajúcich pozorovaní, pričom váhy skorších pozorovaní exponenciálne klesajú.

Brownov model jednoduchého exponenciálneho vyrovňovania vychádza z predstavy, že každé pozorovanie analyzovaného časového radu je možné vyjadriť v tvare súčtu konštantnej deterministickej zložky a náhodnej (reziduálnej) zložky.

Model je teda možné zapísať v tvare $y(t) = \beta(t) + \varepsilon(t)$, kde $\beta(t) = \beta$ je konštanta o ktorej sa predpokladá, že v každom úseku časového radu zostáva relatívne stabilná, ale v čase môže prechádzať pomalými zmenami; $\varepsilon(t)$ je náhodná zložka. Úlohou je nájsť odhad konštanty β , ktorá sa v tomto prípade rovná vyrovnanej hodnote $\hat{y}$.

Odhad trendu, t.j. vyrovnanú hodnotu časového radu, je možné vypočítať podľa rekurentného vzorca

$$\hat{y}_2 = y_1, \hat{y}_t = \hat{y}_{t-1} + \alpha \cdot (y_{t-1} - \hat{y}_{t-1}), \text{ pre } t = 3, \dots, n;$$

kde $0 < \alpha < 1$ je zvolená konštanta, ktorú nazývame *vyrovňavacia konštanta*, y_{t-1} – empirická hodnota ukazovateľa v časovom období $t-1$, $\hat{y}_{t-1}$ – vyrovnaná hodnota ukazovateľa v časovom období $t-1$, ktorá bola prognózovaná na toto obdobie; $\hat{y}_t$ – nová vyrovnaná hodnota ukazovateľa, ktorá je prognózou na časové obdobie t . Hodnota $\hat{y}_1$ nie je definovaná.

Po úprave dostaneme rekurentný vzorec v tvare $\hat{y}_t = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot \hat{y}_{t-1}$.

Hodnoty $e_t = y_t - \hat{y}_t$, $t = 1, 2, \dots, n$ sa nazývajú *rezíduá časového radu*. Sú to odhady náhodnej zložky časového radu. Vo všeobecnosti sa za najlepší považuje ten model, ktorý dáva najmenšie rezíduá.

Určenie optimálnej hodnoty vyrovňavacej konštanty α je zásadným problémom aplikácie uvedeného modelu. V reálnych aplikáciách exponenciálneho vyrovňovania sa hodnota vyrovňavacej konštanty α určuje spravidla experimentálne pomocou „metódy skúšok a omylov“. Táto metóda spočíva v postupnom skúšaní rôznych hodnôt α , napr. 0,1; 0,2; ... ; 0,9 (prípadne aj s jemnejším krokom). Potom za najvhodnejšiu hodnotu vyrovňavacej konštanty α vyberáme tú, ktorá minimalizuje vybrané reziduálne charakteristiky, napr. *MSE* (*Mean Square Error*), *RMSE* (*Root Mean Square Error*), *MAPE* (*Mean Absolute Percentage Error*) a pod.

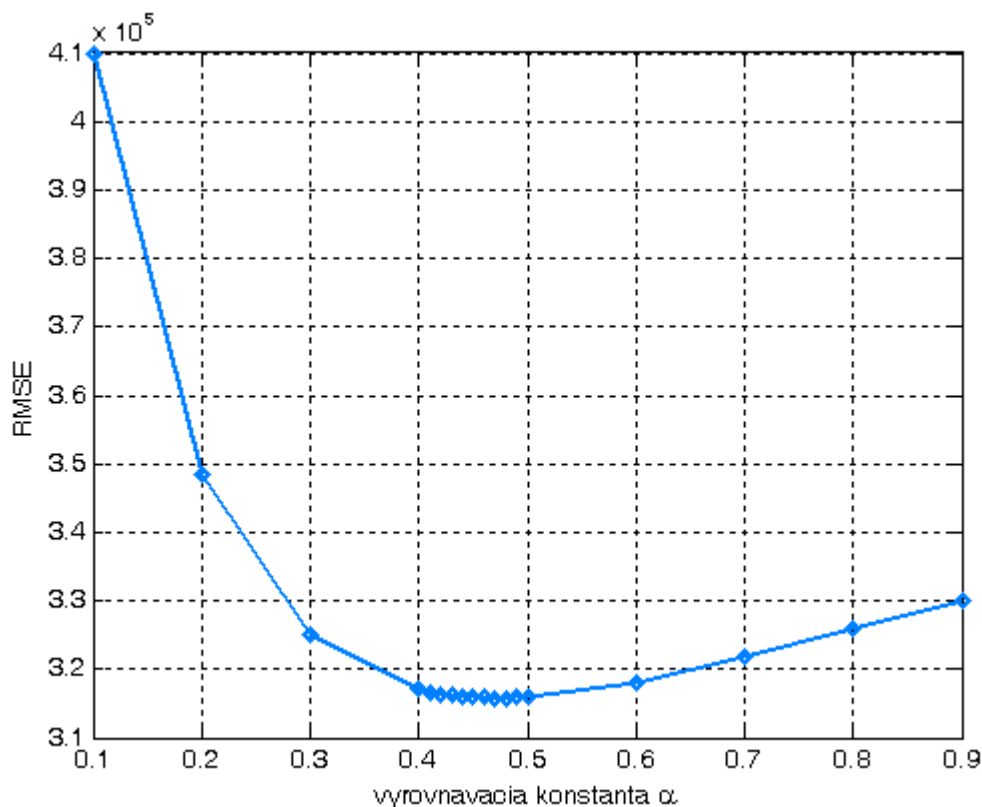
Pre reziduálne charakteristiky budeme v prípade uvažovaného modelu aplikovať

$$\text{vzt'ahy: } MSE = \frac{1}{n} \sum_{t=1}^n e_t^2, \quad RMSE = \sqrt{MSE}, \quad MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{y_t} \cdot 100 \, \%.$$



Ak je vypočítaná hodnota $MAPE < 10\%$, je možné daný model hodnotiť ako excelentný z hľadiska použitia pre tvorbu prognóz; pre hodnoty $MAPE$ medzi $10 - 20\%$ ako dobrý pre predikciu, medzi $20 - 50\%$ ako akceptovateľný a nad 50% ako nevhodný pre predikciu.

Pre uvažovaný časový rad daný tabuľkou 1 boli vypočítané hodnoty $RMSE$ pre zvolené hodnoty α . Výber optimálnej hodnoty vyrovnávacej konštanty α sa uskutočnil na základe kritéria minimalizácie reziduálnej charakteristiky $RMSE$. Na obrázku 2 je znázornená závislosť hodnôt $RMSE$ od voľby konštanty α .



Obr. 2 Grafické zobrazenie hodnôt $RMSE$ pre rôzne hodnoty α

Na základe výpočtov, ktoré sú graficky ilustrované na obr. 2, zvolíme $\alpha = 0,47$. Pre túto hodnotu vyrovnávacej konštanty α bola dosiahnutá minimálna hodnota reziduálnej charakteristiky $RMSE$, a to konkrétne $RMSE = 315825$.

Pre zvolenú hodnotu $\alpha = 0,47$ je reziduálna charakteristika $MAPE = 6,4837\%$. t.j. vybraný model jednoduchého exponenciálneho vyrovnávania poskytuje kvalitnú predikciu.

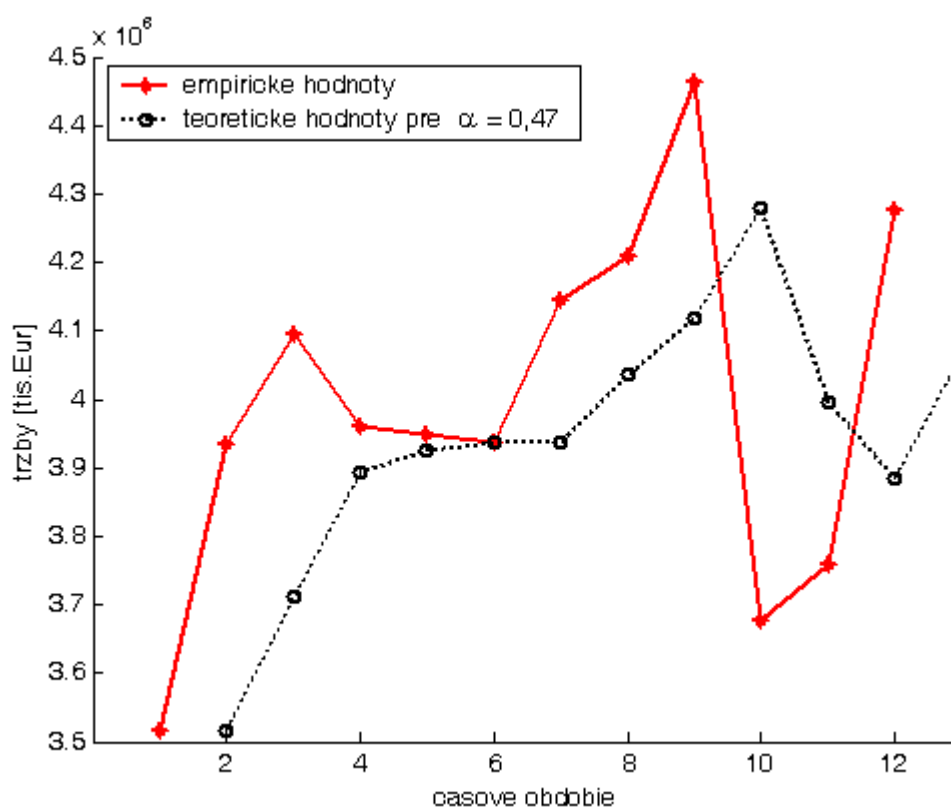
Empirické hodnoty y_t a vypočítané teoretické (vyrovnané) hodnoty $\hat{y}_t$ pre $\alpha = 0,47$ sú prehľadne usporiadané v tabuľke 3. Grafické porovnanie empirických a teoretických hodnôt časového radu je na obrázku 3.

V tabuľke 3 je zvýraznená prognóza na 13. časové obdobie, t.j. na rok 2012. Teda na základe zvoleného modelu je možné na rok 2012 predpovedať hodnotu tržieb za vlastné výkony a tovar v priemyselnej výrobe potravín, nápojov a tabakových výrobkov na Slovensku vo výške 4068542,7 tis. €.



Tab. 3 Empirické a teoretické hodnoty

t	y_t	$\hat{y}_t$
1	3514951,5	—
2	3933272,0	3514951,5
3	4095828,2	3711562,1
4	3961038,3	3892167,2
5	3949011,3	3924536,6
6	3938238,1	3936039,7
7	4145873,9	3937072,9
8	4209879,3	4035209,4
9	4464052,7	4117304,3
10	3676812,6	4280276,0
11	3757702,8	3996648,2
12	4276256,3	3884343,9
13	—	4068542,7



Obr. 3 Grafické zobrazenie empirických hodnôt a teoretických hodnôt pre $\alpha = 0,47$



Súhrn

V článku je prezentovaná konkrétna aplikácia jednej z adaptívnych metód vyrovnávania časových radov, a to jednoduché exponenciálne vyrovnávanie. Všetky výpočty a grafy boli realizované v prostredí MATLABu.

Adaptívne metódy vyrovnávania časových radov nepredpokladajú nemennosť parametrov modelu ani stabilitu analytického tvaru ani stabilitu štrukturálnych parametrov v priebehu sledovaného obdobia. Taktiež nepožadujú ani spojitosť trendovej funkcie. Jednoduché exponenciálne vyrovnávanie nachádza svoje uplatnenie najmä pre krátkodobé predpovede. Metóda je vhodná pre pozorovania, ktorých hodnoty kolíšu okolo priemeru, t.j. nie je pri nich zjavný trend. Výhodou metódy je aj jej výpočtová nenáročnosť a nízke náklady.

Kľúčové slová: jednoduché exponenciálne vyrovnávanie, vyrovnávací konštanta, predikcia, priemerná kvadratická chyba, priemerná absolútna percentuálna chyba

Príspevok bol pripravený v rámci riešenia grantového projektu VEGA č. 1/0102/11 Metódy a techniky experimentálneho modelovania vnútropodnikových výrobných a nevýrobných procesov.

Použitá literatúra

- [1] ACZEL, A. D.: *Complete Business Statistics*. Irwin, 1989.
- [2] ARLT, J., ARLTOVÁ, M., RUBLÍKOVÁ, E.: *Analýza ekonomických časových rad s príklady*. VŠE, Praha, 2002.
- [3] BROWN, R. G.: *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York, 1959.
- [4] CYHELSKÝ, L., KAŇOKOVÁ, J., NOVÁK, I.: *Základy teórie štatistiky pro ekonomy*. SNTL/Alfa, Praha, 1979.
- [5] HANČLOVÁ, J., TVRDÝ, L.: *Úvod do analýzy časových rad*. VŠB-TU, Ostrava, 2003.
- [6] LEWIS, C. D.: *Industrial and Business Forecasting Methods*. London, Butterworths, 1982.
- [7] OSTERTAGOVÁ, E.: *Aplikovaná štatistika v počítačovom prostredí MATLABu*. Equilibria, Košice, 2012.
- [8] OSTERTAGOVÁ, E., OSTERTAG, O.: *The Simple Exponential Smoothing Model*. Proceedings of the 4th International Conference on Modelling of Mechanical and Mechatronic Systems, Technical University of Košice, Slovak Republic, pp. 380-384, 2011.
- [9] <http://www.statistics.sk>

Kontaktná adresa

PhDr. Eva Ostertagová, PhD.
Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky
Katedra matematiky a teoretickej informatiky
Němcovej 32, 040 01 Košice
e-mail: eva.ostertagova@tuke.sk