

Štatistický úrad Slovenskej republiky
The Statistical Office of the Slovak Republic

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

vedecký časopis/scientific journal

2/2021
ročník 31



ŠTATISTICKÝ
ÚRAD
SLOVENSKEJ
REPUBLIKY

ISSN 1339-6854 (online)
ISSN 1210-1095 (tlačené vydanie)

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

Recenzovaný vedecký časopis založený v roku 1991. Od roku 2014 jednotlivé čísla časopisu zverejňujeme aj v elektronickej podobe na ssad.statistics.sk. Názory autorov článkov sa nemusia zhodovať s názormi vydavateľa.

Zahraniční poradcovia/Foreign Consultants

Gabriela Czanner

University of Liverpool
Veľká Británia/United Kingdom

Jitka Langhamrová

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Estefanía Mourelle Espasandín

Universidade da Coruña
Španielsko/Spain

Michaela Potančoková

Joint Research Centre,
European Commission
Taliansko/Italy

Hana Řezanková

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Milan Stehlík

Universidad Técnica Federico Santa María,
Čile/Chile
Johannes Kepler University Linz
Rakúsko/Austria

Výkonná redaktorka/Executive Editor

Silvia Hudecová

Jazykové redaktorky/Language Editors

Slovenský jazyk/Slovak Language

Silvia Duchková

Anglický jazyk/English Language

Andrea Okenková

SLOVAK STATISTICS AND DEMOGRAPHY

The scientific peer-reviewed journal founded in 1991. From 2014 individual copies of the journal are available to readers in electronic form at the website ssad.statistics.sk. The opinions of the authors do not necessarily correlate with the opinions of the publisher.

Redakčná rada/Editorial Board

Ľudmila Ivančíková

(predsedníčka/chairwoman)
Štatistický úrad SR
Statistical Office of the SR

Mikuláš Cár

Slovenská štatistická a demografická spoločnosť
Slovak Statistical and Demographic Society

Helena Glaser-Opitzová

Štatistický úrad SR
Statistical Office of the SR

Ján Haluška

INFOSTAT Bratislava

Iveta Stankovičová

Univerzita Komenského v Bratislave
Comenius University in Bratislava

Erik Šoltés

Ekonomická univerzita v Bratislave
University of Economics in Bratislava

Pavol Tišliar

Univerzita Cyrila a Metoda v Trnave
University of Ss. Cyril and Methodius in Trnava
Masarykova univerzita
Masaryk University

Boris Vaňo

INFOSTAT - Výskumné demografické centrum
INFOSTAT - Demographic Research Centre

Adresa redakcie/Address of Editorial Office

Slovenská štatistika a demografia
Štatistický úrad SR
Lamačská cesta 3/C, 840 05 Bratislava 45
Slovenská republika

E-mailová adresa/E-mail address

SSaD@statistics.sk

ssad.statistics.sk
www.statistics.sk

OBSAH/CONTENTS

I. VEDECKÉ ČLÁNKY/SCIENTIFIC ARTICLES

- Eva KOTLEBOVÁ, Martin SLIACKY** **3**
MOŽNOSTI POROVNÁVANIA ÚROVNE HODNÔT ZNAKU VO VIACERÝCH
SÚBOROCH PRI NESPLNENÍ PREDPOKLADOV ANALÝZY ROZPTYLU
POSSIBILITIES OF COMPARING MORE THAN TWO GROUP MEANS IN
A RESPONSE VARIABLE IN THE CASE OF FAILURE OF THE ASSUMPTIONS OF
ANALYSIS OF VARIANCE

II. INFORMATÍVNE ČLÁNKY, NÁZORY, RECENZIE, ROZHOVORY, INFORMÁCIE/ INFORMATIVE ARTICLES, OPINIONS, REVIEWS, INTERVIEWS, INFORMATION

- Mikuláš CÁR** **18**
EPIDEMICKÁ SITUÁCIA Z REGIONÁLNEHO POHĽADU
EPIDEMIC SITUATION FROM A REGIONAL PERSPECTIVE
Názor/Opinion

- Monika KÚŠIKOVÁ** **23**
PRODUKTY NA PODPORU ZVYŠOVANIA ŠTATISTICKEJ GRAMOTNOSTI
PRODUCTS SUPPORTING THE ENHANCEMENT OF STATISTICAL LITERACY
Informácia/Information

- Ľudmila IVANČÍKOVÁ** **27**
KONFERENCIA NOVÉ TECHNIKY A TECHNOLOGIE PRE OFICIÁLNE
ŠTATISTIKY (NTTS 2021)
CONFERENCE ON NEW TECHNIQUES AND TECHNOLOGIES FOR OFFICIAL
STATISTICS (NTTS 2021)
Informácia/Information

- III.PRIPRAVUJEME/COMING SOON** **28**

Eva KOTLEBOVÁ,
Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita
v Bratislave
Martin SLIACKY
Swis Re, Bratislava

MOŽNOSTI POROVNÁVANIA ÚROVNE HODNÔT ZNAKU VO VIACERÝCH SÚBOROCH PRI NESPLNENÍ PREDPOKLADOV ANALÝZY ROZPTYLU

POSSIBILITIES OF COMPARING MORE THAN TWO GROUP MEANS IN A RESPONSE VARIABLE IN THE CASE OF FAILURE OF THE ASSUMPTIONS OF ANALYSIS OF VARIANCE

ABSTRAKT

Pri použití analýzy rozptylu sa v praxi javí ako najproblematickejší predpoklad homoskedasticity (rovnosti rozptylov v porovnávaných súboroch). V takom prípade (ak sú splnené ostatné predpoklady) je možnosť využiť Welchovu analýzu rozptylu, pri ktorej sa vo výpočte testovacej štatistiky rozsahy jednotlivých výberov nahradia váhami, v ktorých sú zahrnuté aj výberové rozptyly. V príspevku sa porovnávajú výhody a nevýhody klasickej a Welchovej analýzy rozptylu prostredníctvom veľkostí chýb 1. a 2. druhu pri rôznych vstupných hodnotách parametrov súborov. Výsledkom sú odporúčania pre voľbu vhodnej metódy pre rôzne kombinácie veľkostí porovnávaných súborov a ich variabilít.

ABSTRACT

When using analysis of variance in practice, the assumption of homoscedasticity (equality of variance in the compared files) appears to be the most problematic. In such case (if the other assumptions are met), it is possible to use Welch analysis of variance where the ranges of individual samples are replaced by weights, in the calculation of test statistics, including sample variances. The paper compares the advantages and disadvantages of classical and Welch analysis of variance by means of error sizes of I and II types at different input values of file parameters, the outcome of which are the recommendations for choosing a suitable method for different combinations of sizes of the compared files and their variabilities.

KLÚČOVÉ SLOVÁ

analýza rozptylu, Welchova analýza rozptylu, hladina významnosti, sila testu, simulácie Monte Carlo

KEY WORDS

analysis of variance, Welch analysis of variance, significance level, power of a test, Monte Carlo simulations

1. ÚVOD

Analýza rozptylu je jednou z najviac používaných štatistických metód. V praxi je totiž často potrebné porovnať hodnoty číselného znaku vo viacerých súboroch, aby sme (ak sa preukážu štatisticky významné rozdiely) mohli bližšie identifikovať tie dvojice skupín, medzi ktorými existujú štatisticky významné rozdiely, a prípadne sa pokúsiť hľadať mechanizmus vplyvu faktorov, ktoré ich spôsobujú.

V príspevku sa zaoberáme najjednoduchším prípadom – jednoduchou analýzou rozptylu, v ktorej sa posudzuje vplyv jedného faktora (slovnej premennej A) na hodnoty spojitej číselnej závislej premennej Y . Faktor (kategoriálna premenná) má niekoľko obmien, podľa ktorých sú vytvorené porovnávané skupiny. Využitie metódy analýzy rozptylu si vyžaduje splnenie troch predpokladov: (1) výbery musia byť nezávislé (táto podmienka obyčajne býva v praxi splnená, vyplýva zo samotnej formulácie riešeného problému), (2) v každej skupine by mala mať závislá premenná normálne rozdelenie (táto podmienka sa na základe centrálnej limitnej vety môže čiastočne obísť, ak sú rozsahy súborov dostatočne veľké, respektíve možno využiť neparametrický Kruskallov-Wallisov test) a napokon (3) je tu najviac limitujúca podmienka rovnosti rozptylov v porovnávaných súboroch (homoskedasticita). Ak nie je splnená, nemôžeme sa totiž na výsledok analýzy rozptylu dostatočne spoľahnúť – najmä p-hodnoty blízke hranici 0,05 môžu vzbudiť isté pochybnosti.

V nasledujúcich častiach spresníme, akým spôsobom môžu rozdielne hodnoty rozptylov v porovnávaných súboroch (tzv. heteroskedasticita) ovplyvniť chyby 1. a 2. druhu pri analýze rozptylu a ako je to pri Welchovej analýze rozptylu. Využijeme pritom výsledky simulácií metódou Monte Carlo, ktoré sa realizovali pre rôzne hodnoty parametrov vstupných výberov (rozsahy, výberové rozptyly, stredné hodnoty).

2. ANALÝZA ROZPTYLU

Hodnoty závislej premennej Y , ktoré sú predmetom analýzy, označujeme ako y_{ij} , pričom $i = 1, 2, \dots, k$ (k je počet skupín – úrovní faktora), $j = 1, 2, \dots, n_i$ (j označuje poradie hodnoty y_{ij} v i -tej skupine, pričom počet údajov v i -tej skupine je n_i). Celkový počet údajov je $n = \sum_{i=1}^k n_i$.

Podstatou metódy je testovanie hypotézy, ktorá tvrdí, že faktor A , podľa obmien (úrovní) ktorého boli údaje rozdelené do skupín, nemá na hodnoty premennej Y vplyv, takže priemerná úroveň hodnôt znaku je vo všetkých skupinách rovnaká. Nulová hypotéza má teda tvar: $H_0: \mu_1 = \mu_2 = \dots = \mu_k$. Alternatívna hypotéza H_1 tvrdí, že aspoň jedna z rovností neplatí, t. j. že aspoň medzi dvomi skupinami existujú štatisticky významné rozdiely v stredných hodnotách.

Princípom testovania je porovnávanie medziskupinovej variability (ktorej zdrojom je faktor) a vnútroskupinovej variability, ktorá je determinovaná inými faktormi, pričom sem zaradujeme aj vplyv náhody. Testovacia štatistika má tvar:

$$F = \frac{MSA}{MSE} = \frac{\frac{SSA}{k-1}}{\frac{SSE}{n-k}} = \frac{\frac{\sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i}{k-1}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n-k}} \sim F_{(k-1, n-k)} \quad (1)$$

(\bar{y}_i označuje priemer v i -tej skupine, \bar{y} je celkový priemer hodnôt). Z jej tvaru je zrejmé, že má Fisherovo rozdelenie s počtami stupňov voľnosti $k - 1$ a $n - k$. Čím je medziskupinová variabilita (MSA) v porovnaní s vnútroskupinovou (MSE) väčšia, tým väčší je vplyv uvažovaného faktora na hodnoty závislej premennej. Znamená to, že čím je hodnota testovacej štatistiky väčšia, tým pravdepodobnejšie je zamietnutie nulovej hypotézy. Presnejšie: ak hodnota testovacej štatistiky presiahne 95. percentil jej rozdelenia, na hladine významnosti 0,05 konštatujeme štatisticky významný vplyv faktora A na závislú premennú Y . Ak takáto situácia nastane, je potrebné zistiť, medzi ktorými dvojicami skupín existujú štatisticky významné rozdiely. Na to slúžia testy

mnohonásobného porovnávania (Sheffého, Tukeyho, Bonferroniho, Fisherov LSD test, Duncanov, atď. [9]).

V súvislosti s testami mnohonásobného porovnávania, ale aj so samotnou analýzou rozptylu, je aktuálna otázka, prečo ich nemožno nahradiť párovými t -testami stredných hodnôt (testovali by sa nulové hypotézy $H_0: \mu_r = \mu_s; r = 1, 2, \dots, k; s = 1, 2, \dots, k; r \neq s$). Argumentom na odmietnutie tejto možnosti nie je samotný počet týchto testov (pre k skupín ich je spolu $\binom{k}{2}$), ale ich realizácia by podstatne zvýšila pravdepodobnosť chyby 1. druhu, t. j. pravdepodobnosť zamietnutia nulovej hypotézy, aj keď v skutočnosti platí. Voláme ju hladina významnosti a označujeme ako α .

Ak by sme pomocou analýzy rozptylu porovnávali len tri skupiny, tak pri každom z troch $\binom{3}{2}$ párových t -testov je pravdepodobnosť zamietnutia platnej nulovej hypotézy α , teda pravdepodobnosť, že ju prijmeme, je $1 - \alpha$. Pravdepodobnosť, že sa v troch nezávislých t -testoch nedopustíme chyby 1. druhu, je $(1 - \alpha)^3$ (pre hladinu významnosti 0,05 je to 0,8574, takže súčasné zamietnutie troch platných nulových hypotéz má pravdepodobnosť až 14,26%, pre $k > 3$ je táto pravdepodobnosť ešte vyššia). Pri realizácii viacerých testov sa zvyčajne uvažuje tzv. skupinová miera chýb 1. druhu (familywise error – FWER), ktorá je definovaná ako pravdepodobnosť toho, že sa pri sérii testov dopustíme aspoň jednej chyby 1. druhu [14]. Toto však nie je predmetom nášho príspevku.

Ako sme už uviedli v úvode, z predpokladov využitia metódy analýzy rozptylu sa ukazuje ako najproblematickejší predpoklad homoskedasticity – rovnosti rozptylov premennej Y v porovnávaných súboroch. Testovanie hypotézy na overenie tohto predpokladu by malo predchádzať samotnej analýze rozptylu. K dispozícii sú známe testy:

Bartlettov test – testovacia štatistika má tvar [8]:

$$B = \frac{1}{c} \left[(n - k) \ln \tilde{s}^2 - \sum_{i=1}^k (n_i - 1) \ln \tilde{s}_i^2 \right], \quad (2)$$

v ktorom \tilde{s}_i^2 označuje výberový rozptyl premennej Y v i -tom súbore,

$$c = 1 + \frac{\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{n - k}}{3(k - 1)} \quad (3)$$

a \tilde{s}^2 označuje združený rozptyl

$$\tilde{s}^2 = \frac{\sum_{i=1}^k \tilde{s}_i^2 (n_i - 1)}{n - k} \quad (4)$$

Pre $n \rightarrow \infty$ má rozdelenie chí-kvadrát s počtom stupňov voľnosti $k - 1$.

Leveneov test – testovacia štatistika má tvar [4]:

$$W = \frac{\frac{\sum_{i=1}^k n_i \cdot (\bar{Z}_i - \bar{Z})^2}{k-1}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}{n-k}}, \quad (5)$$

ktorý je rovnaký ako má testovacia štatistika samotnej analýzy rozptylu (aj s rovnakým rozdelením), ale vstupnými hodnotami sú údaje premennej Z , ktorá je transformáciou analyzovanej premennej Y : $Z_{ij} = |Y_{ij} - \bar{Y}_i|$, pričom podľa okolností sa skupinové priemery \bar{Y}_i môžu nahradiť aj skupinovými mediánmi, prípadne 10%-nými zostrihnutými priermi [2].

O'Brienov test [11] je podobný ako Leveneov test; takisto je analýzou rozptylu transformovanej premennej Z , len transformácia má inú podobu. Hodnoty premennej sa určia pomocou vzťahu:

$$Z_{ij} = \frac{(n_{ij}-1,5) \cdot n_i \cdot (y_{ij} - \bar{y}_i)^2 - 0,5s_i^2 \cdot (n_i-1)}{(n_i-1) \cdot (n_i-2)} \quad (6)$$

kde s_i^2 je výberový rozptyl premennej Y v i -tej skupine

Cochranov test – testovacia štatistika má tvar [8]:

$$C = \frac{\max_{i=1,2,\dots,k} (s_i^2)}{\sum_{i=1}^k s_i^2} \quad (7)$$

Hartleyho test – testovacia štatistika má tvar [8]:

$$H = \frac{\max(s_i^2)}{\min(s_i^2)} \quad (8)$$

Posledné dva testy majú svoje špeciálne rozdelenie, ktoré závisí od počtu porovnávaných skupín. Pre všetky testy homoskedasticity platí, že pre heteroskedastické údaje má testovacia štatistika veľkú hodnotu, takže kritická oblasť leží na pravom chvoste rozdelenia.

Ak podmienka homoskedasticity nie je splnená, možno hodnoty premennej Y transformovať pomocou vhodnej funkcie tak, aby transformované hodnoty vyhovovali tejto požiadavke. Voľba transformačnej funkcie závisí od tvaru rozdelenia hodnôt premennej Y [7]. Aj keď sa transformáciou môže dosiahnuť splnenie podmienky homoskedasticity, treba pripomenúť, že platí pre transformované údaje, takže interpretácia výsledkov analýzy rozptylu môže byť problematická.

3. WELCHOVA ANALÝZA ROZPTYLU

Táto metóda [5, 13] sa môže použiť vtedy, keď sa pri klasickej analýze rozptylu nepotvrdí predpoklad homoskedasticity. V testovacej štatistike tohto testu sa namiesto rozsahov výberov n_i používajú ako váhy veličiny $w_i = \frac{n_i}{s_i^2}$, takže sa do určitej miery eliminujú súbory s veľkou variabilitou. Čitateľ testovacej štatistiky je podobne ako pri klasickej analýze rozptylu priemerným štvorcem medziskupinovej variability (MSA),

pričom váhy w_i sa použijú nielen na vyjadrenie súčtu štvorcov odchýlok, ale pomocou nich sa vypočíta aj celkový priemer:

$$\bar{y}_{WELCH} = \frac{\sum_{i=1}^k \bar{y}_i \cdot w_i}{\sum_{i=1}^k w_i} \quad (9)$$

$$SSA_{WELCH} = \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 \cdot w_i \quad (10)$$

$$MSA = \frac{SSA}{k-1} \quad (11)$$

Testovacia štatistika má tvar:

$$F = \frac{MSA}{1 + \frac{2\Lambda(k-2)}{3}} \quad (12)$$

v ktorom

$$\Lambda = \frac{3 \sum_{i=1}^k \left(\frac{1 - \frac{w_i}{\sum_{i=1}^k w_i} \right)^2}{n_i - 1}}{k^2 - 1} \quad (13)$$

Tak ako pri klasickej analýze rozptylu má Fisherovo rozdelenie s počtami stupňov voľnosti $k - 1$ a $\frac{1}{\Lambda}$. Kritickou oblasťou je pravý chvost rozdelenia, takže nulovú hypotézu na hladine významnosti α zamietneme, ak hodnota testovacej štatistiky presiahne $(1 - \alpha)$ -ty percentil rozdelenia.

Nevýhodou tohto testu v porovnaní s klasickou analýzou rozptylu je jeho menšia sila (pravdepodobnosť zamietnutia nulovej hypotézy, ak v skutočnosti neplatí), preto ho používame len ako náhradnú možnosť pre heteroskedastické údaje. Najmä ak sú rozsahy súborov malé, výsledok testu treba brať s rezervou.

V nasledujúcich častiach porovnáme klasickú a Welchovu analýzu rozptylu z hľadiska pravdepodobností chýb 1. a 2. druhu. Využijeme pritom výsledky simulácií metódou Monte Carlo.

4. SIMULÁCIE MONTE CARLO A ICH VYUŽITIE NA ODHADY CHÝB PRI TESTOVANÍ

Simulácie číselných náhodných veličín s konkrétnymi hodnotami parametrov sú vhodným nástrojom na riešenie širokého spektra teoretických (to je náš prípad), ale aj praktických problémov, keď je potrebné odhadnúť hodnoty kľúčových charakteristík vývoja, hospodárskych výsledkov, investičných stratégií, hroziacich rizík atď. Prakticky všetky softvéry využívané v štatistike umožňujú simulácie hodnôt veličín so známym rozdelením s konkrétnymi hodnotami parametrov.

V induktívnej štatistike majú simulácie kľúčový význam. Umožňujú nám napríklad odhadnúť rozdelenie testovacej štatistiky (ak ho nevieme analyticky vyjadriť), porovnávať výsledky rôznych metód pri rôznych vstupných podmienkach, porovnávať kvalitu odhadov, ale aj chýb pri testovaní, čo sme využili v našom príspevku.

Pri **odhade miery chýb 1. druhu** (pravdepodobnosť zamietnutia platnej nulovej hypotézy) vo všeobecnosti postupujeme takto [6]:

1. Vytvorí sa hypotetická populácia (teoretické rozdelenie) s určitými vlastnosťami tak, aby zodpovedalo nulovej hypotéze.
2. Metódou Monte Carlo [3] sa z rozdelenia vygeneruje náhodný výber dopredu stanoveného rozsahu.
3. Pomocou vygenerovaných údajov realizujeme test.
4. Zistíme, či nastala chyba 1. druhu, teda či bola nulová hypotéza zamietnutá (údaje boli generované z rozdelenia spĺňajúceho podmienky nulovej hypotézy).
5. Výsledok zaznamenáme pomocou binárnej veličiny

$$I_i = \begin{cases} 1 & \text{ak nastala chyba 1. druhu} \\ 0 & \text{ak nenastala chyba 1. druhu} \end{cases}$$

6. Kroky 2. – 5. M × zopakujeme.
7. Pravdepodobnosť chyby 1. druhu odhadneme ako podiel testov, v ktorých nastala chyba, zo všetkých M testov:

$$\hat{\alpha} = \frac{1}{M} \sum_{i=1}^M I_i \quad (14)$$

Je optimálne, ak je odhad blízky hodnote 0,05 (v štatistických testoch sa len výnimočne využíva iná hladina významnosti ako 0,05), resp. ak pochádza z intervalu 0,025 – 0,075. Príliš malé hodnoty (pod 0,025) znamenajú, že test je konzervatívny (je vyššia pravdepodobnosť chyby 2. druhu). Ak je odhad chyby väčší ako 0,075, test je príliš liberálny, t. j. má vyššiu pravdepodobnosť zamietnutia nulovej hypotézy, ako je štandardne nastavená hladina významnosti 0,05.

Mieru chýb druhého druhu (a z nej odvodenú **silu testu**) odhadneme na základe postupu [6]:

1. Vytvorí sa hypotetická populácia (teoretické rozdelenie) s určitými vlastnosťami tak, aby nulová hypotéza neplatila.
2. Metódou Monte Carlo sa z rozdelenia vygeneruje náhodný výber dopredu stanoveného rozsahu.
3. Pomocou vygenerovaných údajov realizujeme test.
4. Zistíme, či nastala chyba 2. druhu, teda či bola nulová hypotéza prijatá (údaje boli generované z rozdelenia, ktoré nespĺňalo podmienky nulovej hypotézy).
5. Výsledok zaznamenáme pomocou binárnej veličiny

$$I_i = \begin{cases} 1 & \text{ak nastala chyba 2. druhu} \\ 0 & \text{ak nenastala chyba 2. druhu} \end{cases}$$

6. Kroky 2. – 5. M × zopakujeme.
7. Pravdepodobnosť chyby 2. druhu odhadneme ako podiel testov, v ktorých nastala chyba, zo všetkých M testov:

$$\hat{\beta} = \frac{1}{M} \sum_{i=1}^M I_i \quad (15)$$

8. Odhadom sily testu je potom $1 - \hat{\beta}$.

Za optimálny výsledok považujeme odhad sily testu blízky hodnote 1 (pravdepodobnosť chyby 2. druhu blízka 0).

Pri odhade miery chýb 1. druhu, ktoré môžu nastať pri realizácii klasickej, resp. Welchovej analýzy rozptylu, preto vytvárame skupiny populácií s rovnakou strednou

hodnotou (tak, aby platila nulová hypotéza) a veľkosť chyby 2. druhu odhadujeme pomocou skupín populácií s odlišnými strednými hodnotami.

Okrem stredných hodnôt (rovnakých, vs. rôznych) sme v nami definovaných rozdeleniach použili aj rôzne rozsahy výberových súborov (vyvážené s rovnakými počtami, resp. nevyvážené) a variabilitu (homoskedastické, vs. nehomoskedastické populácie). Tým sme dostali odpovede na otázku, ktoré z týchto okolností a v akej miere ovplyvňujú pravdepodobnosti chýb.

5. ODHAD CHÝB 1. A 2. DRUHU PRI KLASICKEJ A WELCHOVEJ ANALÝZE ROZPTYLU

Pri odhadoch pravdepodobností chýb obidvoch druhov sme sledovali faktor s tromi úrovňami, takže sme generovali trojice súborov s rozsahmi n_1, n_2, n_3 s určitými vopred stanovenými vlastnosťami. Premenná mala vo všetkých súboroch normálne rozdelenie $N(\mu_i; \sigma_i^2); i = 1, 2, 3$.

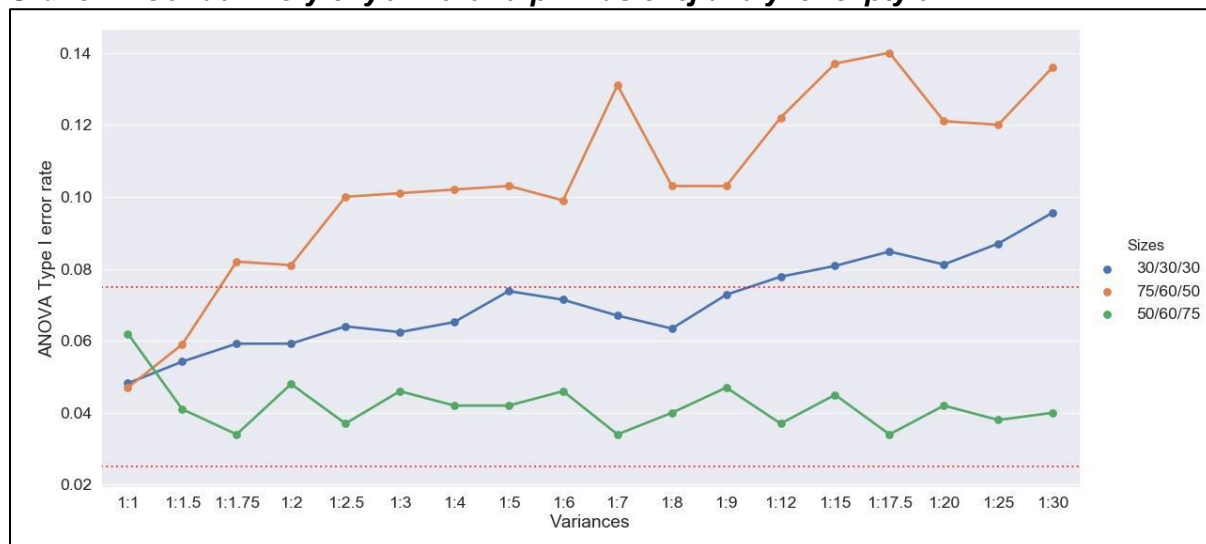
Na odhad **pravdepodobnosti chýb 1. druhu** sme generovali trojice súborov s rovnakou strednou hodnotou (použili sme hodnotu 0), pričom sme pracovali s vyváženými ($n_1 = n_2 = n_3 = 30$) aj nevyváženými súbormi ($n_1 < n_2 < n_3; n_1 : n_2 : n_3 = 50 : 60 : 75$), homoskedastickými aj heteroskedastickými súbormi, v ktorých sa pomer najmenšieho a najväčšieho rozptylu pohyboval v rozmedzí od 1 : 1 až po 1 : 30 (spolu 18 rôznych variantov pomerov). Rozlišovali sme pritom dve možnosti: súbor s najväčšou variabilitou mal najväčší rozsah ($n_1 < n_2 < n_3, \sigma_1^2 < \sigma_2^2 < \sigma_3^2$) alebo to bolo naopak ($n_1 < n_2 < n_3, \sigma_1^2 > \sigma_2^2 > \sigma_3^2$). Celkovo sme tak analyzovali $18 \times 3 = 54$ skupín údajov, pričom pre každú z nich sme vytvorili 5 000 simulácií. Na základe postupu uvedeného v 3. časti sme odhadli pravdepodobnosť chyby 1. druhu pre klasickú analýzu rozptylu aj pre Welchovu analýzu rozptylu. Výsledky sme znázornili graficky.

Na grafe č. 1 sú pomocou spojnicového grafu znázornené odhady miery chýb 1. druhu klasickej analýzy rozptylu, a to pre vyvážené modely (modrá farba), pre modely, v ktorých má súbor s najvyššou variabilitou najväčší rozsah (zelená farba) a oranžovou farbou sú znázornené odhady v prípade, že súbor s najvyššou variabilitou má najmenší rozsah. Na osi x sú znázornené všetky uvažované pomery rozptylov ($\sigma_{min}^2 / \sigma_{max}^2$). Prvý bod zľava pri všetkých lomených čiarach (zodpovedajúci pomeru 1 : 1) tak znázorňuje odhad chyby 1. druhu pre homoskedastické údaje. Optimálna situácia z hľadiska veľkosti chyby 1. druhu je vtedy, keď sa bod lomenej čiary nachádza medzi dvomi prerušovanými čiarami, ktoré sú na úrovni 0,025 a 0,075.

Z grafu č. 2 sú zrejmé nasledujúce skutočnosti: chyba 1. druhu sa nachádza v požadovanom intervale nielen pre údaje spĺňajúce podmienku homoskedasticity, ale aj vtedy, ak je pomer najmenšieho a najväčšieho rozptylu 1 : 1,5, a to pri vyvážených aj nevyvážených modeloch. Smerom k výraznejším rozdielom vo variabilite sa javí ako problematické, keď súbor s najväčšou variabilitou je najmenší (a naopak), ktorý už od pomeru 1 : 1,75 má hodnoty vyššie ako 0,075, čo znamená, že v takýchto prípadoch je vyššie riziko, že súbory budeme považovať z hľadiska úrovne hodnôt znaku za rôzne, aj keď sú v skutočnosti rovnaké. Pri vyvážených modeloch sa odhad chyby už pri pomere 1 : 5 blíži hranici 0,075, presiahne ju až od pomeru 1 : 12. Najpriaznivejší výsledok z hľadiska využiteľnosti klasickej analýzy rozptylu vyšiel vtedy keď najväčšiu variabilitu mal súbor s najväčším rozsahom. Všetky hodnoty vyšli v požadovanom

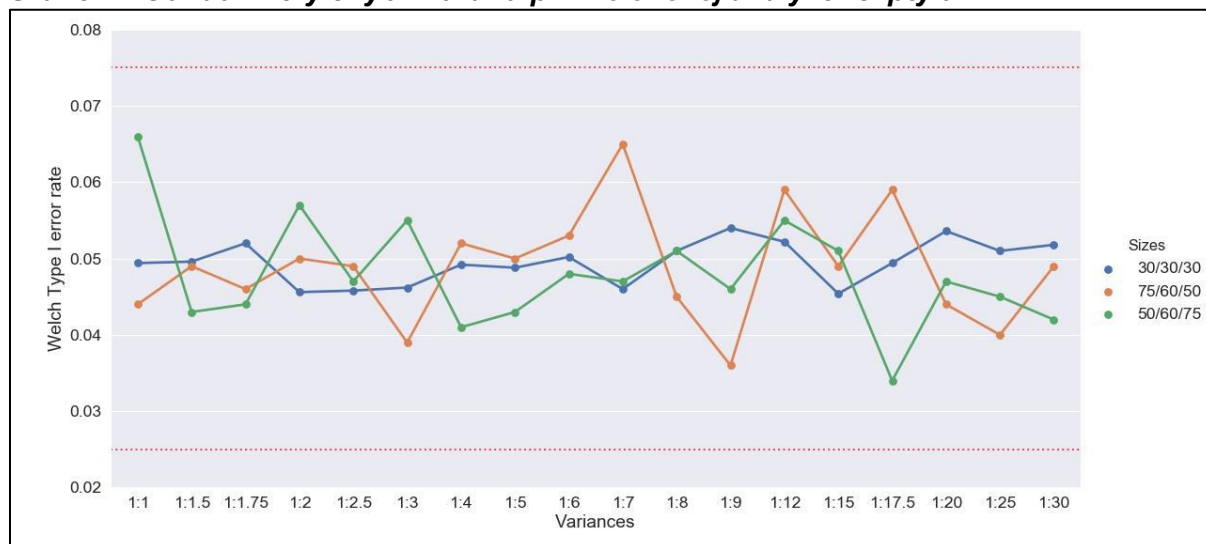
intervale 0,025 až 0,075, aj keď sa pohybujú (s výnimkou pomeru 1 : 1) pod hranicou 0,05. Na grafe č. 2 sú znázornené odhady chýb 1. druhu pre rovnaké spektrum možností vstupných veličín pri Welchovej analýze rozptylu. Vidíme, že pri všetkých kombináciách možností početností súborov a variability sa body lomených čiar nachádzajú vo vnútri požadovaného intervalu (dokonca užšieho), takže žiadna z okolností neovplyvňuje (jedným alebo druhým smerom) pravdepodobnosť chyby 1. druhu.

Graf č. 1: Odhad miery chýb 1. druhu pri klasickej analýze rozptylu



Zdroj: vlastné spracovanie

Graf č. 2: Odhad miery chýb 1. druhu pri Welchovej analýze rozptylu



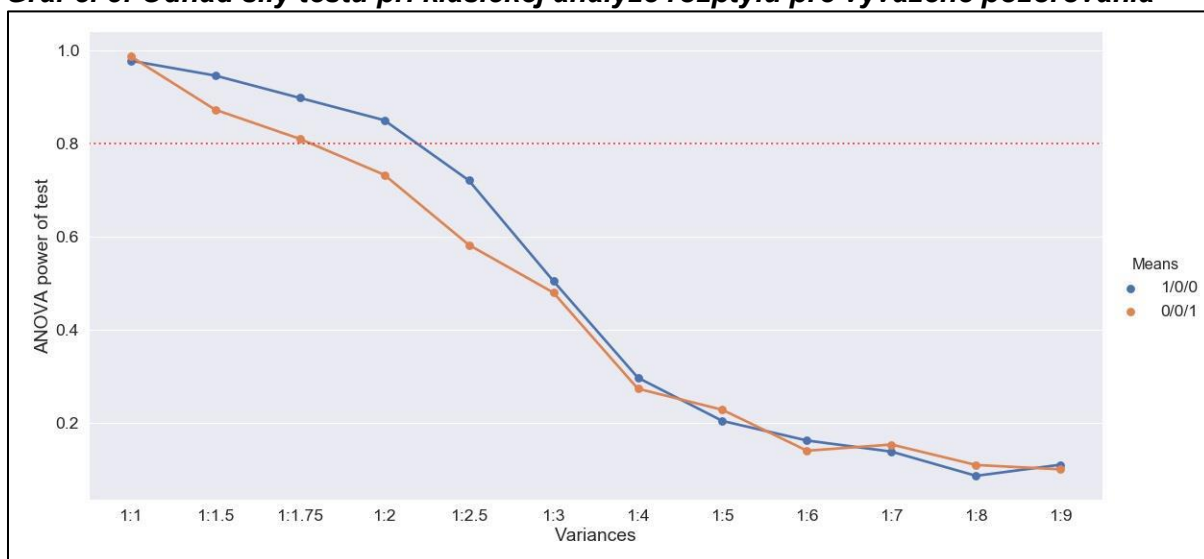
Zdroj: vlastné spracovanie

Na odhad pravdepodobnosti chýb 2. druhu sme generovali trojice súborov s nerovnými strednými hodnotami – pre súbory s početnosťami n_1, n_2, n_3 sme použili trojice stredných hodnôt $(0, 0, 1)$ a $(1, 0, 0)$. Rozsahy súborov sme nastavili tak ako pri odhade pravdepodobnosti chyby 1. druhu: analyzovali sme vyvážené súbory s početnosťami $(30:30:30)$ aj nevyvážené s početnosťami $(50:60:75)$, resp. $(75:60:50)$. Použili sme 12 možností pre pomer najmenšieho aj najväčšieho rozptylu od 1 : 1 do 1 : 9. Celkovo sme analyzovali $2 \times 3 \times 2 \times 9 = 108$ skupín. Pre každú

konkrétnu možnosť zvolených parametrov sme vytvorili 5000 simulácií, pričom pre každú z nich sme realizovali klasickú aj Welchovu analýzu rozptylu. Výsledky sme pre každú kombináciu možností znázornili graficky. Na každom grafe sú na osi x všetky možnosti pomeru rozptylov ($\sigma_{min}^2/\sigma_{max}^2$), y -ové súradnice bodov sú na úrovni sily testu (t. j. $1 - \hat{\beta}$). Hranica na dostatočnú silu (0,8) [12] je znázornená prerušovanou čiarou, takže tie body, ktoré sa nachádzajú nad ňou, predstavujú akceptovateľnú množinu východiskových podmienok z hľadiska chyby 2. druhu.

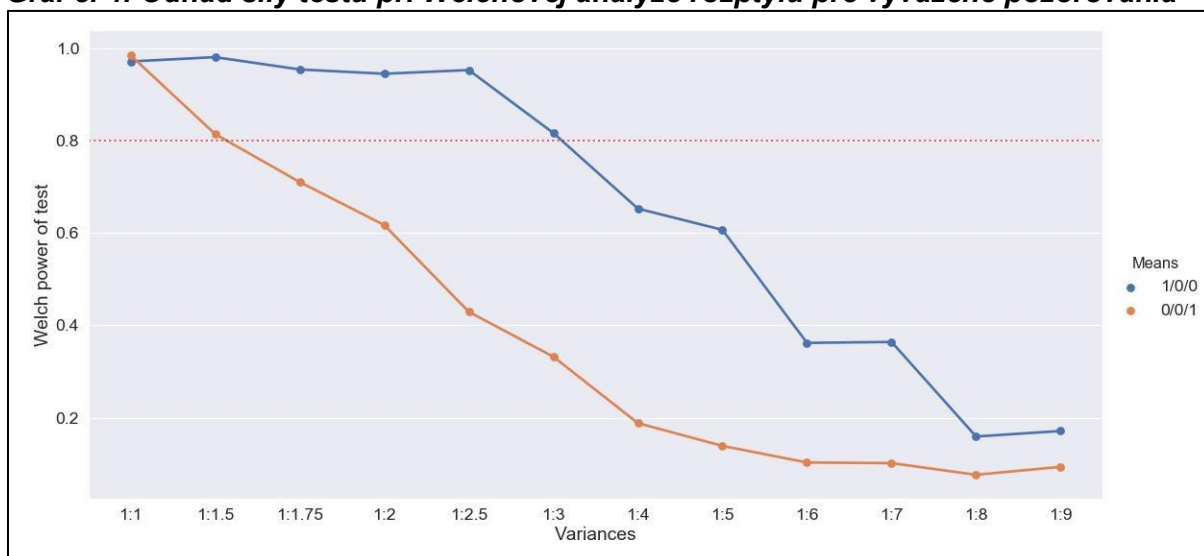
Na grafe č. 3 a č. 4 sú znázornené výsledky pre súbory s rovnakými početnosťami. Farbami sú odlišené možnosti, ktoré vyjadrujú, či skupina s najväčším rozptylom je tá, ktorá má odlišnú hodnotu (oranžová farba), alebo má odlišnú strednú hodnotu skupina s najmenším rozptylom (modrá farba).

Graf č. 3: Odhad sily testu pri klasickej analýze rozptylu pre vyvážené pozorovania



Zdroj: vlastné spracovanie

Graf č. 4: Odhad sily testu pri Welchovej analýze rozptylu pre vyvážené pozorovania



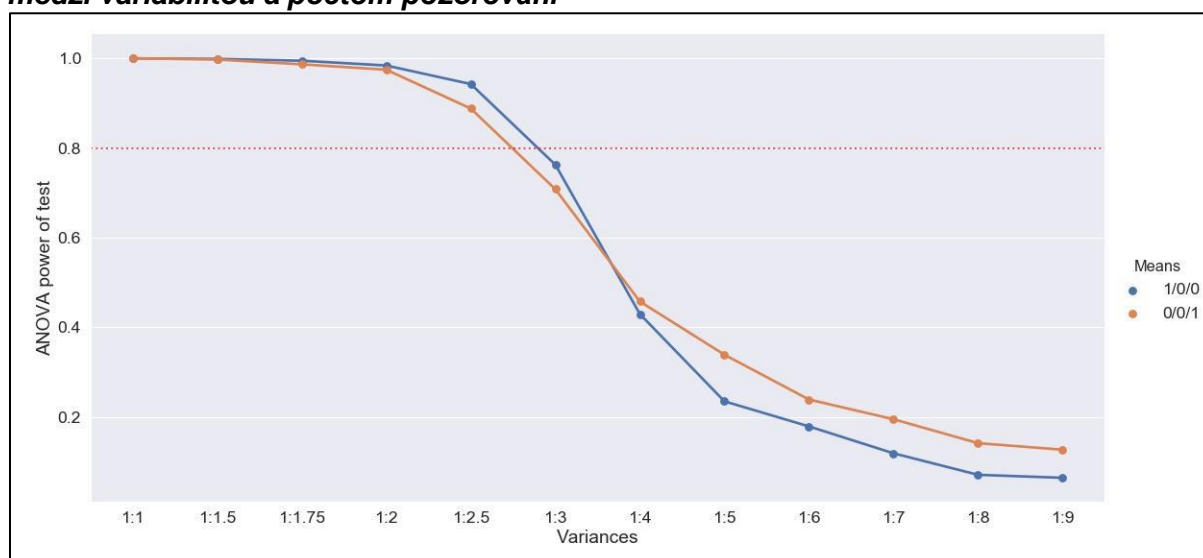
Zdroj: vlastné spracovanie

Ako vidno z grafu č. 3, pri klasickej analýze rozptylu v prípade, že skupina s odlišnou strednou hodnotou je tá, ktorá má najnižšiu variabilitu, spĺňajú podmienku prvé štyri prípady: po pomer rozptylov 1 : 2, ak má odlišnú hodnotu stredná hodnota skupiny s najväčším rozptylom, vyhovujú len prípady do pomeru 1 : 1,75. Smerom doprava (k nižším hodnotám pomerov) sa lomená čiara zvažuje nadol (toto však platí aj pre všetky nasledujúce prípady – veľké disproporcie medzi rozptylmi znižujú silu testu bez ohľadu na ostatné vstupné podmienky).

Pri Welchovej analýze rozptylu badať výraznejší rozdiel medzi prípadmi znázornenými odlišnými farbami: kým modrá lomená čiara je nad hranicou 0,8 až po pomer rozptylov 1 : 3 (dokonca pri predchádzajúcich pomeroch je nad úrovňou 0,9), pre oranžovú platí, že vyhovujúca sila testu je len pre prvé dva prípady: homoskedastickom (pomer 1 : 1) a pomere 1 : 1,5. Znamená to, že z hľadiska sily testu je oveľa priaznivejšia možnosť, ak súbor s odlišnou strednou hodnotou má najmenší rozptyl.

Graf č. 5 znázorňuje odhady sily testu klasickej analýzy rozptylu pre nevyvážené súbory, pričom počty pozorovaní sú v priamom vzťahu s variabilitou ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$). Vidíme, že lomené čiary majú veľmi podobný tvar, v oboch prípadoch (ak je odlišná stredná hodnota súboru s najmenším, resp. s najväčším rozptylom) platí, že až po pomer rozptylov 1 : 2,5 je sila testu nad hranicou 0,8, pri väčších rozdieloch medzi rozptylmi pomerne strmo klesá.

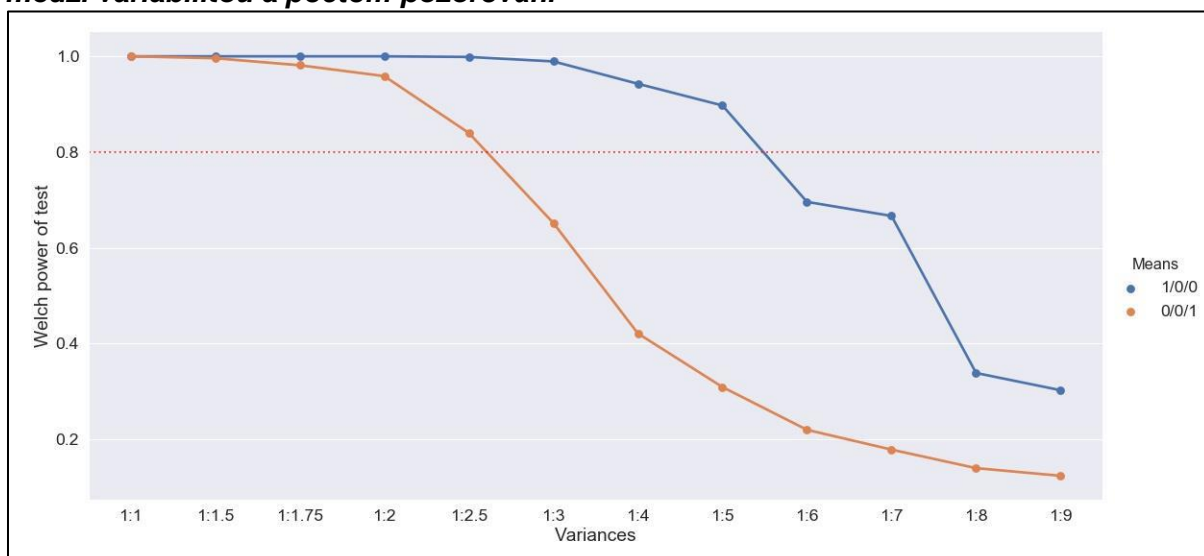
Graf č. 5: Odhad sily testu pri klasickej analýze rozptylu v prípade priameho vzťahu medzi variabilitou a počtom pozorovaní



Zdroj: vlastné výpočty

Na grafe č. 6 sú znázornené odhady sily testu Welchovej analýzy rozptylu pre analogické vzťahy medzi počtami pozorovaní a variabilitou ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$). Oproti klasickej analýze rozptylu tu (podobne ako pri vyvážených súboroch) badať markantný rozdiel medzi odhadovanými silami testov. Kým pri súboroch, v ktorých je najnižšia variabilita v súbore, ktorý má odlišnú strednú hodnotu (modrá čiara), je sila testu nad hranicou 0,8 ešte pre pomery rozptylov 1 : 5, v prípade, že súbor s odlišnou strednou hodnotou má najvyššiu variabilitu, je posledný akceptovateľný pomer len 1 : 2,5.

Graf č. 6: Odhad sily testu pri Welchovej analýze rozptylu v prípade priameho vzťahu medzi variabilitou a počtom pozorovaní

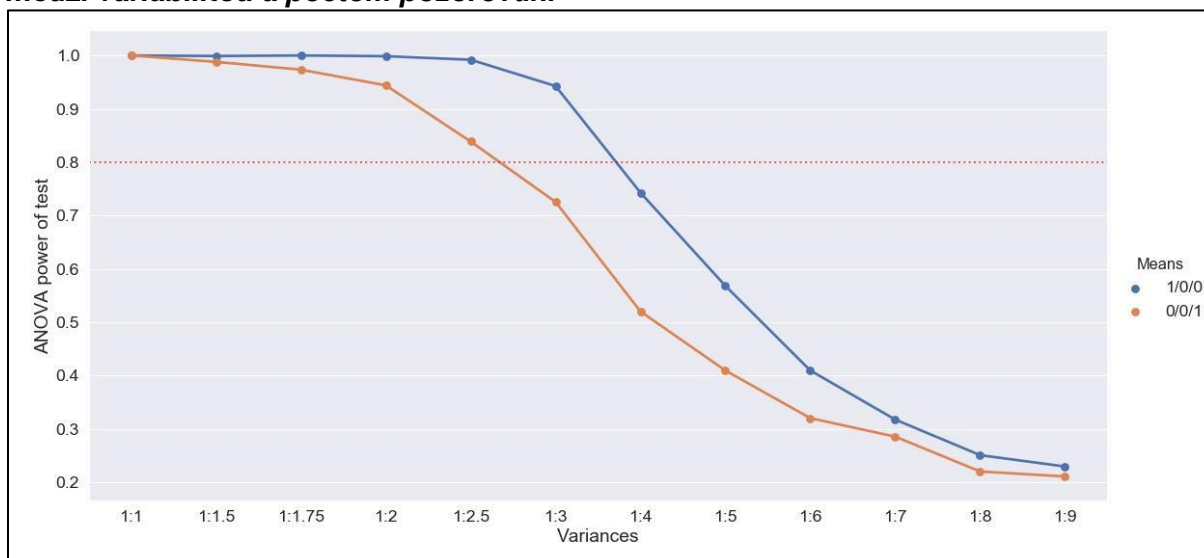


Zdroj: vlastné výpočty

Nakoniec sme sa venovali prípadu, keď medzi rozsahmi súborov a variabilitou existuje nepriamy vzťah ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$). Výsledky sú znázornené na grafe č. 7 a č. 8.

Pri klasickej analýze rozptylu sú znovu z hľadiska sily testu lepšie vlastnosti vtedy, keď najnižšiu variabilitu má súbor s odlišnou strednou hodnotou (modrá lomená čiara) – vyhovujúci pomer rozptylov je ešte 1 : 3, vtedy, keď má súbor s odlišnou strednou hodnotou najvyššiu variabilitu, je tento pomer len 1 : 2,5.

Graf č. 7: Odhad sily testu pri klasickej analýze rozptylu v prípade nepriameho vzťahu medzi variabilitou a počtom pozorovaní

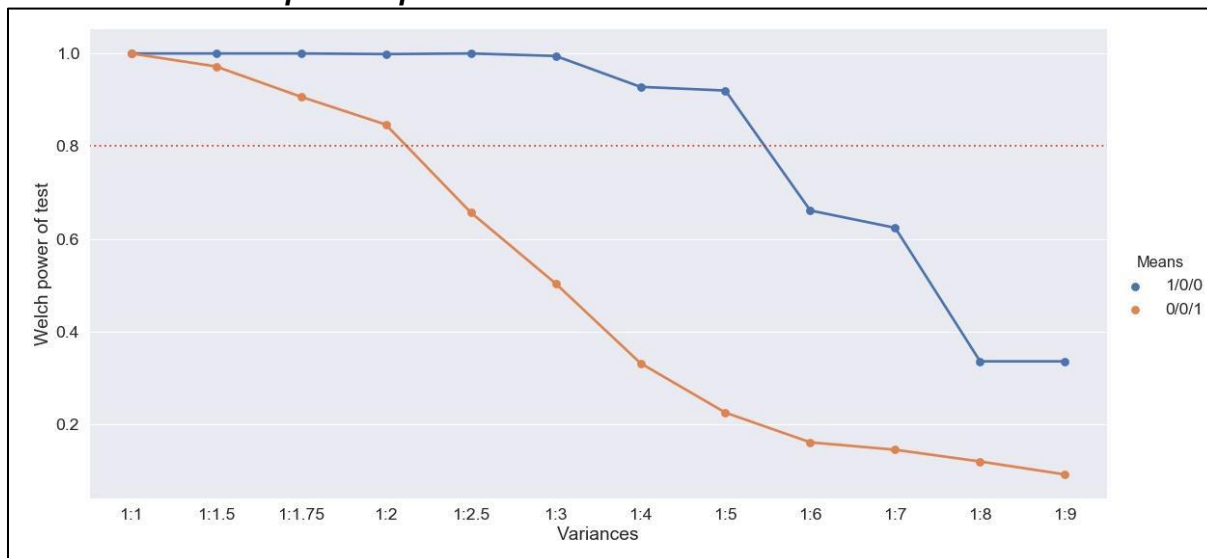


Zdroj: vlastné výpočty

Výraznejší rozdiel medzi lomenými čiarami vidíme pri Welchovej analýze rozptylu (graf č. 8): modrá lomená čiara (súbor s odlišnou strednou hodnotou má najnižšiu variabilitu) sa až po pomer rozptylov 1 : 3 drží veľmi tesne pri hodnote 1, pri ďalších

dvoch pomeroch (1 : 4 a 1 : 5) je ešte nad úrovňou 0,9, hodnoty nižšie ako 0,8 má odhadovaná sila testu až od pomeru 1 : 6. Pre skupiny údajov, kde súbor s odlišnou strednou hodnotou má najvyššiu variabilitu, už od pomeru rozptylov 1 : 2,5 je odhadovaná sila testu pod hranicou 0,8.

Graf č. 8: Odhad sily testu pri Welchovej analýze rozptylu v prípade nepriameho vzťahu medzi variabilitou a počtom pozorovaní



Zdroj: vlastné výpočty

6. ZÁVER

Realizovaním simulácií, v ktorých sme nastavili rôzne vstupné parametre porovnávaných troch súborov, sme pomocou odhadu chýb 1. druhu a 2. druhu (resp. sily testu) porovnali klasickú a Welchovu analýzu rozptylu.

Z hľadiska veľkosti chýb 1. druhu sa Welchova analýza rozptylu ukázala ako vhodnejšia metóda pre každý z nami zvolených pomerov rozptylov (až do pomeru najmenšieho a najväčšieho rozptylu 1 : 30) a to bez ohľadu na vzájomné vzťahy medzi rozsahmi súborov, pretože odhady chýb vo všetkých prípadoch vyšli v intervale od 0,025 po 0,075. Pri klasickej analýze rozptylu sa ukázal rozhodujúci vzťah medzi rozsahmi súborov a ich variabilitou. Ak bol priamy ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$), pre všetky pomery rozptylov bola odhadovaná sila v požadovanom intervale. Pre súbory s rovnakým rozsahom je akceptovateľný ešte pomer rozptylov 1 : 9, ale ak sú rozsahy a rozptyly v nepriamom vzťahu, nevyhovujúce pomery sa začínajú už na úrovni 1 : 1,75.

Pri analýze veľkosti chýb 2. druhu (resp. sily testu) sme skonštatovali, že vo všetkých porovnávaných prípadoch má na odhad sily vplyv najmä to, či má súbor s odlišnou strednou hodnotou najnižšiu alebo najvyššiu variabilitu. Pre silu testov sa ukázala ako priaznivejšia možnosť, ak súbor s odlišnou strednou hodnotou mal najnižšiu variabilitu, pričom rozdiely boli oveľa výraznejšie pri Welchovej analýze rozptylu ako pri klasickej. Vzhľadom na viaceré možnosti vstupných okolností sme prehľadné porovnanie klasickej a Welchovej analýzy rozptylu zapísali do tabuľky č. 1. V nej sú uvedené tie hodnoty pomerov rozptylov, pri ktorých má príslušná metóda za uvedených okolností silu ešte nad hranicou 0,8. Priaznivejšiu možnosť (ak si máme vybrať z porovnávaných metód) sme zvýraznili. Znamená to, že za okolností

zodpovedajúcich záhlaviu tabuľky má metóda z daného riadku výhodu v tom, že aj väčšia disproporcja medzi rozptylmi významne neznižuje silu testu.

Tabuľka č. 1: Porovnanie sily testu klasickej a Welchovej analýzy rozptylu (pomery rozptylov ($\sigma_{min}^2/\sigma_{max}^2$), pre ktoré je odhad sily testu nad úrovňou 0,8)

| metóda \ parametre | $n_1 = n_2 = n_3$ | | $n_1 < n_2 < n_3$ $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ | | $n_1 < n_2 < n_3$ $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ | |
|--------------------|-------------------|----------------|---|-------|---|--------------|
| | M ¹ | O ² | M | O | M | O |
| klasická | 1:2 | 1:1,75 | 1:2,5 | 1:2,5 | 1:3 | 1:2,5 |
| Welchova | 1:3 | 1:1,5 | 1:5 | 1:2,5 | 1:5 | 1:2 |

Zdroj: vlastné spracovanie

Pri overovaní zhody stredných hodnôt v troch súboroch v praxi je v každom prípade potrebné najskôr overiť homoskedasticitu. Ak sa potvrdí, respektíve ak je pomer najmenšieho a najväčšieho rozptylu do 1 : 1,5, možno použiť klasickú metódu analýzy rozptylu. Ak sú medzi rozptylmi výraznejšie rozdiely je vhodné zamerať sa na vzájomné vzťahy medzi rozsahmi a variabilitou, prípadne (na základe charakteristík jednotlivých súborov, ktorými sa odhadujú charakteristiky základných súborov) zistiť, aký je vzťah medzi strednými hodnotami a variabilitou tak, aby sme sa mohli (do určitej miery) riadiť kritériami na základe tabuľky č. 1.

Náš príspevok rieši problematiku porovnania dvoch metód v rámci vstupných podmienok, ktoré sme si vopred stanovili. Treba si uvedomiť, že sme pracovali len s tromi súbormi, premenná mala normálne rozdelenie, pričom sme si zvolili len určité pomery rozsahov súborov. Pri väčšom počte súborov alebo väčších disproporciách medzi rozptylmi môžu byť výsledky iné. (Napríklad [Abdi, 2007]) uvádza, že v prípade ak je najpočetnejší súbor aspoň 4-násobne väčší ako najmenší zo súborov, odporúča sa použiť Welchovu analýzu rozptylu.) Navyše, medzi strednými hodnotami neboli výrazné rozdiely, čo tiež môže zohrať úlohu pri rozhodovaní sa, ktorá z metód je vhodnejšia. Preto možno naše závery aplikovať len za predpokladov, ktoré sú podobné ako tie naše. Na komplexnejšiu analýzu by bolo vhodné pracovať s viacerými súbormi, použiť aj iné rozdelenia, voliť rôzne pomery medzi rozsahmi súborov, prípadne rozšíriť spektrum vzťahov medzi parametrami výberových súborov.

LITERATÚRA

- [1] ABDI, H.: O'Brien test for homogeneity of variance. In: Encyclopedia of measurement and statistics. London: SAGE Publication Ltd. 2007, s. 701 – 704. ISBN 9781412916110.
- [2] BROWN, M. B. – FORSYTHE, A. B.: The ANOVA and multiple comparisons for data with heterogeneous variances. In: Biometrics. Wiley Online Library, 1974, s. 719 – 724. ISSN 0006-3444.
- [3] ĎŘÍMAL, J. – TRUNEC, D.: Úvod do metody Monte-Carlo. 1. vyd. Brno: UJEP. 1989. 122 s. ISBN 80-210-0022-8.
- [4] KOTLEBOVÁ, E. – ŠOLTÉS, E. – SIVAŠOVÁ, D.: Štatistická indukcia v príkladoch. 1. vydanie. Bratislava: EKONÓM, 2015, 222 s. ISBN 978-80-225-4135-0.
- [5] LIU, H.: Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance. A Thesis. Richmond, Virginia: Virginia

¹ M – modrá farba lomenej čiary (súbor s odlišnou strednou hodnotou má najnižšiu variabilitu).

² O – oranžová farba lomenej čiary (súbor s odlišnou strednou hodnotou má najvyššiu variabilitu).

- Commonwealth University, 2015. 46 s. [online]. [cit. 5. 1. 2021]. Dostupné na: <https://scholarscompass.vcu.edu/etd/3985>
- [6] MARTINEZ, W. L. – MARTINEZ A. R.: Computational statistics handbook with MATLAB. 2. vydanie. Boca Raton: FL: Chapman, 2008, 767 s. ISBN 15-848-8566-1.
- [7] MAXWELL, S. E. – DELANEY, H. D. – KELLEY, K.: Designing experiments and analyzing data: A model comparison perspective. Boca Raton: Routledge, 2017. 1080 s. ISBN 978-1138892286.
- [8] PACÁKOVÁ, V. – KOTLEBOVÁ, E. – LABUDOVIČOVÁ, V. – SIPKOVÁ, Ľ.: Štatistická indukcia pre ekonómov. 1. vydanie. Bratislava: EKONÓM, 2012. 317 s. ISBN 978-80-225-3382-9.
- [9] PACÁKOVÁ, V. – LABUDOVIČOVÁ, V. – SIPKOVÁ, Ľ. – STANKOVIČOVÁ, I.: Štatistická indukcia pre ekonómov a manažérov. 1. vydanie. Bratislava: Wolters Kluwer, 2015. 365 s. ISBN 978-80-8168-081-6.
- [10] SLIACKY, M.: Testy hypotéz porovnávajúcich parametre viac ako dvoch súborov. Diplomová práca. Fakulta hospodárskej informatiky, Ekonomická univerzita, Bratislava, 2020. Evidenčné číslo: 103005/I/2020/421000214186
- [11] SOIKLIEW, K. – ARAVEEPORN, A.: Modifications of Levene's and O'Brien's Tests for Testing the Homogeneity of Variance Based on Median and Trimmed Mean, Bangkok: Thai Statistical Association, 2018, s. 106 – 128. ISSN 2351-0676. [online]. [cit. 5.1.2021]. Dostupné na: <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/135555/101285>
- [12] WALMSLEY, A. – L. E., BROWN, M. C.: What Is Power?. [online]. [cit. 5.1.2021]. Dostupné na: <https://www.statisticsteacher.org/2017/09/15/what-is-power/>
- [13] WELCH, B. L.: On the comparison of several mean values: an alternative approach. In *Biometrika*, 1951, s. 330 – 336. ISSN 0006-3444.
- [14] ZAR, J. H.: *Biostatistical Analysis*. 5. vyd. Englewood Cliffs: NJ: Prentice Hall, Inc. 2010. 960 s. ISBN 978-0321656865.

RESUMÉ

Príspevok sa zaoberá možnosťami porovnania stredných hodnôt spojitej kvantitatívnej premennej vo viac ako dvoch nezávislých súboroch (pracovali sme s trojicami súborov), ak je porušený predpoklad homoskedasticity pri analýze rozptylu. Cieľom bolo porovnanie klasickej a Welchovej analýzy rozptylu z hľadiska možností nastania chýb pri testovaní. Pomocou metódy Monte Carlo sme generovali trojice hodnôt súborov s vopred stanovenými vlastnosťami, na ktoré sme aplikovali klasickú aj Welchovu analýzu rozptylu. Výsledky slúžili na odhad miery chýb 1. a 2. druhu, pomocou ktorých sme porovnávali uvedené metódy. Ukázalo sa, že okrem overenia homoskedasticity pred samotným testovaním je vhodné na základe opisných charakteristík súborov zistiť vzťah medzi strednými hodnotami a variabilitou (či v súbore, ktorého stredná hodnota sa najviac odlišuje od zvyšných dvoch, je v porovnaní s ostatnými vysoká alebo nízka variabilita), pretože práve táto okolnosť môže výrazne redukovať silu testu.

Dosiahnuté výsledky môžu byť vhodným podkladom na rozhodnutie, ktorú z dvoch metód treba v danej situácii použiť.

RESUME

The paper deals with the possibilities of comparing the mean values of continuous quantitative variable in multiple independent data sets (we have worked with three data sets), if the presumption of homoscedasticity in the analysis of variance is violated. The aim was to compare the classical and Welch analysis of variance in terms of the

possibility of testing errors. Using the Monte Carlo method, three values of data sets with predetermined properties were generated, on which both classical and Welch analysis of variance were applied. The results were used to estimate the error rate of type I and I type II by means of which the above methods were compared. It has been shown that in addition to verifying homoscedasticity before the actual testing, it is appropriate to determine the relationship between mean values and variability, based on the descriptive characteristics of the data sets (whether there is a high or low variability in the data set whose mean value differs the most from the other two), since it is this circumstance that can significantly reduce the strength of the test. The results obtained can be a suitable basis for deciding which of the two methods should be used in a given situation.

PROFESIJNÝ ŽIVOTOPIS

RNDr. Eva Kotlebová, PhD., je absolventkou Matematicko-fyzikálnej fakulty Univerzity Komenského v Bratislave (vedecký smer matematika – teória systémov). Po ukončení vysokoškolského štúdia bola tri roky na študijnom pobyte na Katedre štatistiky Fakulty riadenia Vysokej školy ekonomickej v Bratislave. Potom pôsobila niekoľko rokov ako stredoškolská učiteľka matematiky na gymnáziu v Bratislave. Od roku 2003 pracuje na Katedre štatistiky Fakulty hospodárskej informatiky v Bratislave. V roku 2008 ukončila doktorandské štúdium. Venuje sa štatistickej indukcii, bayesovskej štatistike a aplikácii štatistických metód v poisťovníctve.

Ing. Martin Sliacky je absolventom Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave (študijný program Štatistické metódy v ekonómii). Pracuje ako dátový analytik v zaistovní Swiss Re.

KONTAKT

eva.kotlebova@euba.sk
martin@sliacky.com

Názor/Opinion

EPIDEMICKÁ SITUÁCIA Z REGIONÁLNEHO POHĽADU**EPIDEMIC SITUATION FROM A REGIONAL PERSPECTIVE**

Dôsledná metodická príprava ukazovateľov a ich skĺbenie do funkčného systému je dôležitým predpokladom na získanie spoľahlivých informácií o konkrétnych sociálno-ekonomických javoch a procesoch. Platí to tak pri získavaní relevantných informácií o vývoji epidemickej situácie v rámci celého Slovenska, ako aj v jednotlivých regiónoch.

Pri hodnotení konkrétnej situácie sa často skĺzava do zjednodušených schém pomocou rutinných a nie vždy odborne zdôvodnených prístupov, čo zvyčajne vedie k získaniu aj nie celkom korektných informácií a poznatkov. Stáva sa to podľa nás aj pri hodnotení vývoja epidemickej situácie na Slovensku.

Už viackrát boli aj na stránkach odborných časopisov publikované názory, že kompetentné inštitúcie i zodpovední vládni predstavitelia používajú pri hodnotení vývoja rizika šírenia koronavírusu a pri manažovaní boja s pandémiou niektoré nekorektné ukazovatele. Aktuálne sa pokúsime poukázať na niektoré diskutabilné otázky v súvislosti s hodnotením epidemickej situácie v regiónoch.

V záujme monitorovania vývoja epidémie a prijímania protiepidemickej opatrení v závislosti od intenzity šírenia koronavírusu, bola v decembri 2020 vládou SR schválená pôvodná verzia covidového automatu. Pred jeho uvedením do praxe od začiatku februára 2021 bolo potrebné upraviť už nevyhovujúce pôvodné parametre pri vybraných ukazovateľoch, čoho výsledkom bol jeho sprísnený variant. Sprísnenie parametrov celonárodných indikátorov spočívalo v znížení hodnôt parametrov pre jednotlivé stupne varovania zhruba o 20 percent.

Na úspešné fungovanie covidového automatu pri hodnotení rizikovosti vývoja koronavírusu na Slovensku je nevyhnutné vybrať správne ukazovatele a dobre nastaviť ich parametre. Medzi štyri kľúčové ukazovatele automatu na hodnotenie epidemickej situácie na regionálnej úrovni však patria aj ukazovatele 14-dňová incidencia a trend 7-dňovej incidence, ktoré sú z nášho pohľadu diskutabilné.

Tabuľka č. 1: Regionálne indikátory na hodnotenie epidemickej situácie

| Regionálne charakteristiky | Monitoring | Stupeň ostražitosti I | Stupeň ostražitosti II | I. stupeň varovania | II. stupeň varovania | III. stupeň varovania | IV. stupeň varovania |
|--|------------|-----------------------|------------------------|---------------------|----------------------|-----------------------|----------------------|
| Vážená 14-dňová incidencia | <40 | 40 - 80 | 80 - 120 | 120 - 240 | 240 - 800 | 800 - 1600 | > 1600 |
| Počet hospitalizovaných na 100 000 obyv. | < 3 | 3 - 5 | 5 - 10 | 1.okt | 20 - 40 | 40 - 60 | > 60 |
| Trend 7-dňovej incidence | 0,5 -0,7 | 0,7 -0,9 | 0,9 -1,0 | 1,0 -1,2 | 1,2 -1,5 | 1,5 -2,0 | > 2 |
| Zhodnotenie situácie na okresnej úrovni (RÚVZ) | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

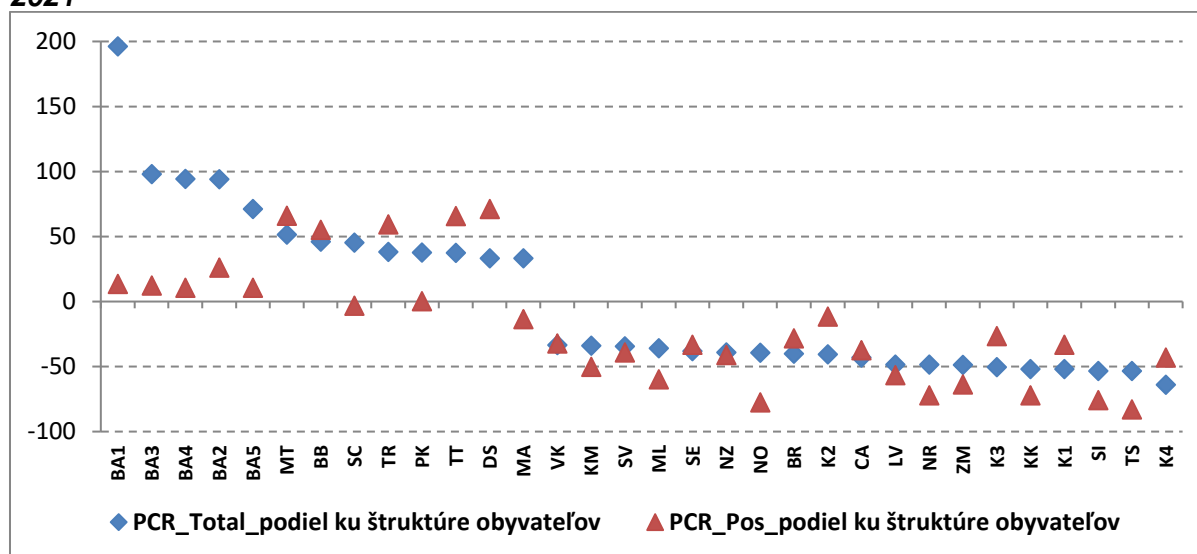
Zdroj: https://www.uvzsr.sk/docs/info/covid19/Covid_automat_2_2_2021.pdf, dátum dostupnosti 9. 3. 2021, vlastné spracovanie

Ukazovatele sedem- a štrnásťdňovej incidence, t. j. počtu pozitívne testovaných na 10-tisíc, či 100-tisíc obyvateľov zostávajú aj v tejto podobe v podstate absolútnymi ukazovateľmi. Naša výhrada proti takto čiastočne upraveným absolútnym ukazovateľom je rovnaká, ako proti jednorazovým i kízavým stredným hodnotám absolútnych ukazovateľov. Po danej úprave sa príslušný ukazovateľ javí ako pomerový, avšak v dostatočnej miere sa neodstráni jeho nekorektnosť spôsobená značnou asymetriou medzi vývojom denného počtu vykonaných testov a počtom pozitívne testovaných osôb. Ukazovatele 7-dňovej a 14-dňovej incidence by bolo možné bez výhrad použiť len v takom prípade, ak by sa priebežne vykonával približne rovnaký počet testov. Taký predpoklad nie je splnený, čo sa pokúsime ukázať na reálnych údajoch.

Na bezproblémové porovnávanie medzi okresmi pomocou ukazovateľa incidence by bolo ideálne, keby bola štruktúra urobených PCR testov podľa okresov zhruba totožná so štruktúrou počtu obyvateľov podľa okresov. Na základe reálnych údajov sa však dá ukázať, že testovanie je stále regionálne značne diferencované a tiež asymetrické z pohľadu počtu vykonaných PCR testov a počtu pozitívne testovaných osôb.

Z údajov za mesiac február 2021 vyplýva, že zhruba v 13 slovenských okresoch sa vykonalo minimálne o tretinu viac testov, ako je podiel obyvateľov daných okresov na celkovom počte obyvateľov na Slovensku. V okrese Bratislava 1 sa vykonalo až trojnásobne viac testov, ako vychádza teoretický prepočet podielu počtu obyvateľov v tomto okrese na celkovom počte obyvateľov na celom Slovensku. Na druhej strane v 19 okresoch sa vykonalo minimálne až o tretinu menej testov, ako by sa ich malo teoreticky urobiť v zmysle rovnomerného testovania. Najhoršie na tom bol vo februári 2021 okres Košice 4, kde sa urobilo až o dve tretiny menej testov v porovnaní s tým, ako sa malo teoreticky urobiť vzhľadom na podiel obyvateľov tohto okresu na celkovom počte obyvateľov Slovenska.

Graf č. 1: Normované ukazovatele o PCR testoch a pozitívite podľa okresov vo februári 2021



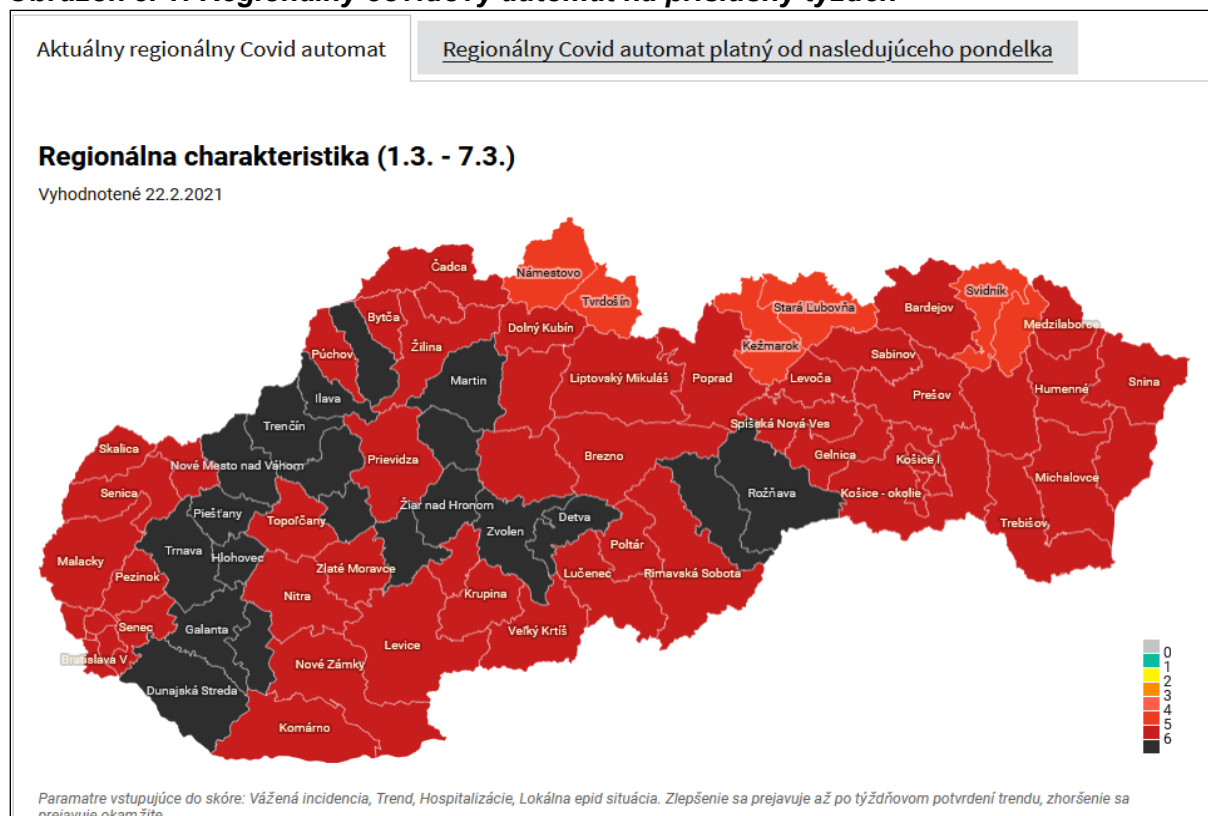
Zdroj údajov: https://raw.githubusercontent.com/Institut-Zdravotnych-Analyz/covid19-data/3917a8e09c729bb96e1e54b64e19a158025db4e4/PCR_Tests/OpenData_Slovakia_Covid_PCRTests_District.csv, dátum dostupnosti 9. 3. 2021, vlastné prepočty

Na posúdenie symetrie medzi počtom vykonaných PCR testov a počtom pozitívne testovaných osôb v jednotlivých okresoch sme použili ich normované ukazovatele vzhľadom na štruktúru obyvateľstva podľa okresov. Z grafu č. 1, v ktorom sú z dôvodu lepšej prehľadnosti zobrazené len okresy z dvoch pólov rebríčka, vyplýva, že významná asymetria medzi počtom PCR testov a počtom pozitívne testovaných osôb je prítomná vo väčšine slovenských okresov a je obojsmerná. Znamená to, že v niektorých okresoch sa v priebehu februára 2021 vykonal „nadmerný“ počet PCR testov oproti zistenej pozitívite osôb a to vzhľadom na ich podiel v štruktúre obyvateľstva. Zhruba v polovici okresov je to zasa naopak.

Výrazne vyššia je asymetria medzi počtom vykonaných PCR testov a počtom pozitívne testovaných osôb v tej skupine okresov, kde je vysoká intenzita testovania pomocou PCR testov. Nižšia, ale stále významná je asymetria medzi sledovanými ukazovateľmi aj v okresoch, kde sa testuje s relatívne nedostatočnou intenzitou. Uvedené poznatky potvrdzujú, že vypovedacia schopnosť ukazovateľa počet pozitívne testovaných napr. na 10-tisíc obyvateľov je v jednotlivých okresoch značne ovplyvnená práve asymetriou medzi počtom vykonaných PCR testov a počtom pozitívne testovaných osôb.

Z vyššie uvedených dôvodov si myslíme, že používanie ukazovateľov o incidencii na 10 tisíc, resp. 100 tisíc obyvateľov je v regionálnom covidovom automate nevhodné. Vzhľadom na už dostupné regionálne údaje o počtoch denne vykonávaných PCR testov a o počtoch pozitívne testovaných sa dá poukázať na diskutabilnú účinnosť regionálneho covidového automatu.

Obrázok č. 1: Regionálny covidový automat na príslušný týždeň



Zdroj: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>, dátum dostupnosti 5. 3. 2021

Podľa oficiálneho regionálneho covidového automatu, pri ktorého zostavovaní sa vychádza aj z podľa nás diskutabilných ukazovateľov o incidencii, bolo do najrizikovejšieho stupňa vývoja koronavírusu v prvom marcovom týždni v roku 2021 zaradených 20 okresov (DS, GA, SA, TT, HC, PN, NM, TN, IL, BN, PE, PB, ZC, ZH, ZV, DT, TR, MT, RA a RV).

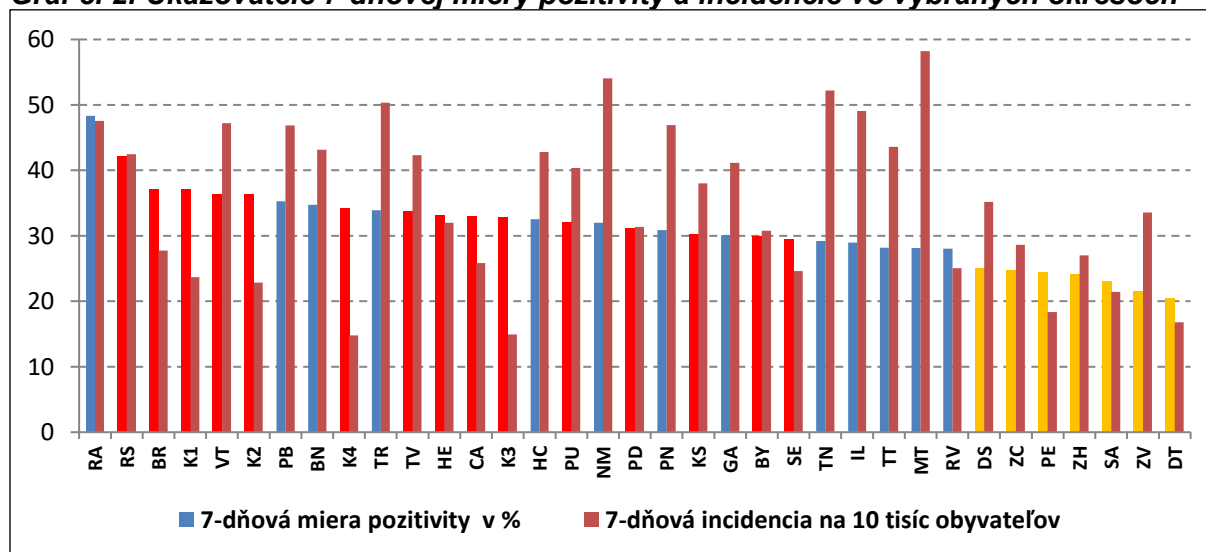
Podľa dostupných údajov o počte vykonaných PCR testov a počte pozitívne testovaných osôb za posledný februárový týždeň v jednotlivých slovenských okresoch možno dospieť k viacerým zaujímavým poznatkom.

Trinášť z 20 okresov, ktoré boli v oficiálnom regionálnom covidovom automate zaradené na základe údajov v poslednom februárovom týždni do najvyššieho stupňa varovania, sa jednoznačne stotožňuje aj s našim zoznamom okresov, vytvoreným podľa 7-dňovej miery pozitívne testovaných. V týchto 13 okresoch dosahovala priemerná miera pozitivity za posledný februárový týždeň hodnoty okolo 30 % a viac.

Najdôležitejší poznatok z nášho pohľadu však spočíva v tom, že medzi najkritickejšie okresy neboli zaradené také okresy, v ktorých miera pozitivity na základe PCR testov za hodnotený týždeň presiahla 30 %. Medzi najkritickejšie okresy z hľadiska zvýšeného rizika šírenia koronavírusu podľa nášho názoru jednoznačne patrili podľa situácie na konci februára 2021 aj okresy Rimavská Sobota, Brezno, všetky 4 košické okresy, Vranov nad Topľou, Trebišov, Humenné, Čadca, Púchov a Prievidza. Tesne pod uvedenou hranicou 30 % miery pozitivity na koronavírus pomocou PCR testov boli aj okresy Bytča a Senica. V priloženom grafe č. 2 je pri týchto vymenovaných okresoch modrý stĺpček zmenený na červený.

Oficiálne zaradenie okresu s vysokým rizikom šírenia koronavírusu v rámci regionálneho covidového automatu do skupiny menej rizikových okresov je nebezpečné, lebo vytvára u obyvateľov falošný pocit väčšej voľnosti a vedie k podceňovaniu nebezpečenstva šírenia pandémie.

Graf č. 2: Ukazovatele 7-dňovej miery pozitivity a incidencie vo vybraných okresoch



Zdroj údajov: ŠÚ SR, [https://github.com/Institut-Zdravotnych-Analyz/covid19-data/blob/main/PCR Tests/OpenData Slovakia Covid PCRTests District.csv](https://github.com/Institut-Zdravotnych-Analyz/covid19-data/blob/main/PCR%20Tests/OpenData%20Slovakia%20Covid%20PCRTests%20District.csv), dátum dostupnosti 9. 3. 2021, vlastné prepočty

Všetky košické okresy sa už v priebehu celého februára radili z nášho pohľadu podľa 7-dňovej miery pozitivity medzi najkritickejšie okresy, kým podľa oficiálneho regionálneho covidového automatu tam neboli zaradené ešte ani v prvom marcovom týždni. Do najkritickejšieho stupňa boli tieto okresy oficiálne zaradené až v týždni od 8. marca 2021, čo je z hľadiska boja proti šíreniu koronavírusu na zamyslenie.

Podľa nás bolo tiež diskutabilné v prvom marcovom týždni oficiálne zaradenie 7 okresov (DS, ZC, PE, ZH, SA, ZV a DT) medzi najkritickejšie okresy s mierou pozitivity v sledovanom období pod 25 %, na úkor okresov, v ktorých zistená miera šírenia rizika koronavírusu presahovala až 35 % (RS, BR, K1, VT a K2).¹

Uvedené poznatky vyvolávajú oprávnené otázky, nakoľko možno zavedený regionálny covidový automat považovať za účinný nástroj prevencie pred šírením koronavírusu, keď maximálne rizikové okresy z pohľadu 7-dňovej miery pozitivity podľa PCR testov boli medzi kritické zaradované minimálne s týždňovým oneskorením.

Sme si vedomí, že pri oficiálnom zaradovaní okresov do regionálneho covidového automatu sa zohľadňujú aj ďalšie ukazovatele, ale určujúcimi by podľa nášho názoru nemali byť diskutabilné ukazovatele incidencie, ale korektnejší ukazovateľ miery pozitivity.

Už krátke fungovanie regionálneho covidového automatu ukázalo, že pri jeho úpravách a priebežnom nastavovaní je potrebné pružne reagovať na meniacu sa reálnu situáciu. Hľadanie funkčných a efektívnych riešení na boj proti pandémie by malo zohľadňovať pripomienky a podnety odborníkov z viacerých vedných odborov.

Ing. Mikuláš CĀR, PhD.

Autor je bývalý dlhoročný člen výboru Slovenskej štatistickej a demografickej spoločnosti.

¹ Aj podľa údajov o týždeň neskôr (za prvý marcový týždeň) boli oficiálne zaradené medzi najrizikovejšie okresy aj okresy Detva, Šaľa a Zvolen, v ktorých bola 7-dňová miera pozitivity pod 20 %, a naopak, chýbali tam okresy Čadca, Vranov nad Topľou, Trebišov a Sobrance, v ktorých bola 7-dňová miera pozitivity vyššia ako 30 %.

Informácia/Information

PRODUKTY NA PODPORU ZVYŠOVANIA ŠTATISTICKEJ GRAMOTNOSTI

PRODUCTS SUPPORTING THE ENHANCEMENT OF STATISTICAL LITERACY

Dynamický a rýchlo sa meniaci svet okolo nás prináša prostredníctvom médií množstvo správ a údajov, z ktorých je niekedy náročné rozlíšiť pravdivé a nepravdivé. Často sa medzi nimi vyskytujú aj tzv. falošné správy („fake news“), ktoré napodobňujú formát spravodajstva alebo iného produktu žurnalistiky, pričom ich tvorcovia zámerne alebo nezámerné zavádzajú svoje publikum.

Vedieť sa v údajoch orientovať a správne ich interpretovať si vyžaduje dostatočnú úroveň štatistickej gramotnosti. Môžeme ju definovať ako schopnosť porozumieť štatistikám a mať zručnosti umožňujúce vyvodiť z nich závery¹.

V rámci Európskeho štatistického spoločenstva (ďalej „EŠS“) prevláda medzi národnými štatistickými úradmi (ďalej „NŠI“) zhoda o potrebe investovania do zvyšovania štatistickej gramotnosti. Prvé „investície“ do tejto oblasti boli realizované v rámci projektu (2016 – 2019) Vízie EŠS 2020 Digitálna komunikácia, analytika pre používateľov a inovatívne produkty (DIGICOM). Jeho hlavným cieľom bola modernizácia komunikácie a šírenia európskej štatistiky. V našej informácii sa budeme venovať produktom, ktoré v súčasnosti na podporu zvyšovania štatistickej gramotnosti ponúka Eurostat, Štatistický úrad SR (ďalej „ŠÚ SR“) a iné NŠI.

Na stránkach Eurostatu sú v rámci podstránky **Statistics for Beginners** (Štatistika pre začiatočníkov) zverejnené články s textami, ktoré prinášajú vybrané štatistické pojmy a témy vysvetlené ľahko zrozumiteľnou a pútavou formou, využívajúc pri tom zaujímavé infografiky. Určené sú pre používateľov, ktorí nemajú žiadne alebo len základné znalosti z oblasti štatistiky, a zároveň pre všetkých, ktorí majú záujem viac porozumieť štatistike.

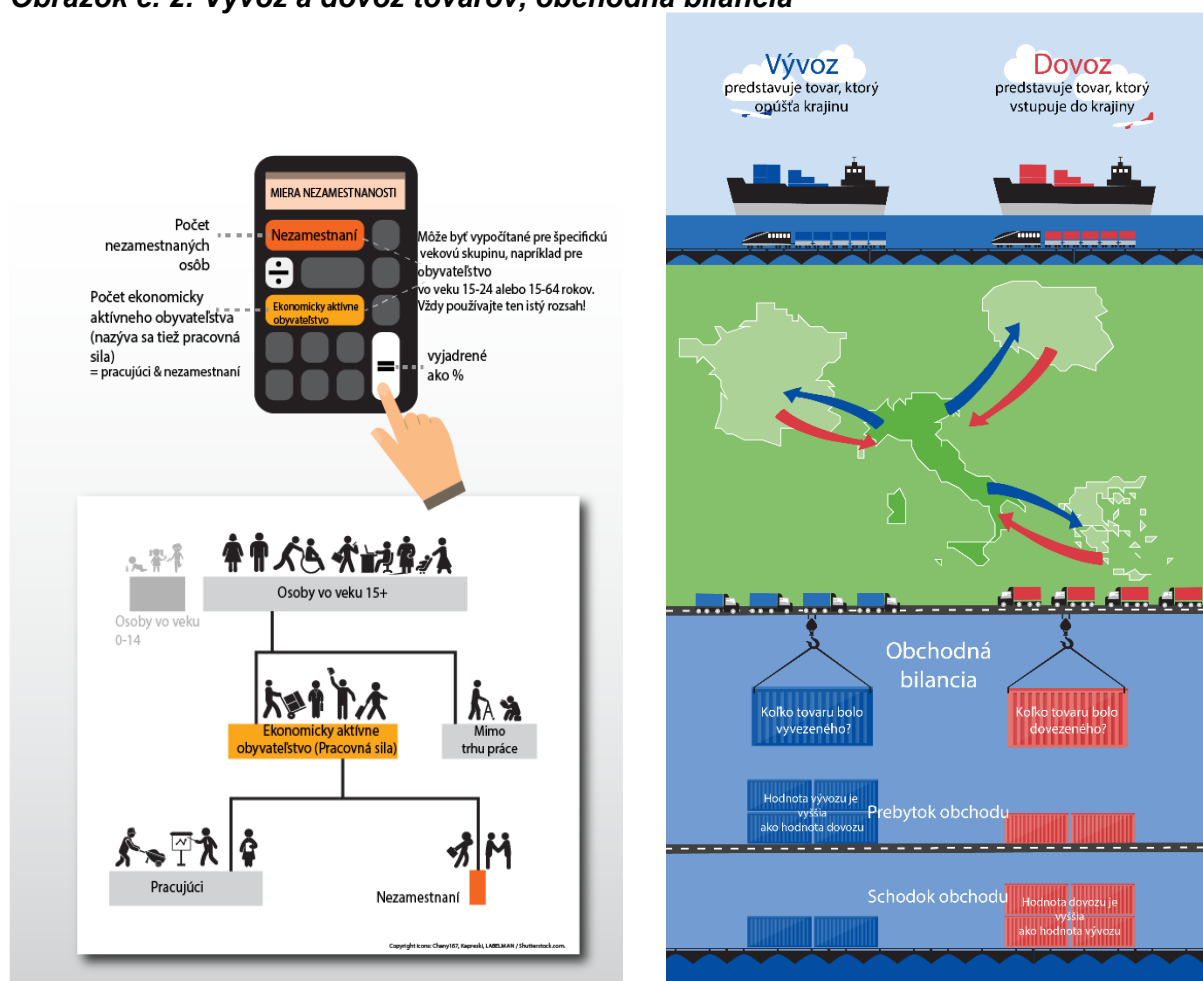
Aktuálne sú k dispozícii samostatné články o štatistických oblastiach týkajúcich sa obyvateľstva, trhu práce, hrubého domáceho produktu, inflácie, verejných financií a medzinárodného obchodu s tovarom. Často používané štatistické pojmy sú tiež definované a vysvetlené v osobitnej rubrike. Pre používateľov sú všetky články k dispozícii aj v slovenskej verzii. V úvode každého článku sa nachádza odkaz na animované video zverejnené na Youtube (dostupné len v anglickej verzii), v ktorom sú atraktívnou jednoduchou formou vysvetlené základné pojmy vzťahujúce sa na danú tému.

Slovenskú verziu všetkých dostupných článkov bloku *Štatistika pre začiatočníkov* pripravil ŠÚ SR v roku 2020 v spolupráci s Eurostatom. V súčasnosti sú v spomenutom bloku zverejnené vybrané články v 7 jazykoch. Odkazy na jednotlivé články v slovenskom jazyku je možné nájsť na portáli Štatistického úradu v časti Služby v rubrike „Štatistika pre začiatočníkov“.

¹ CORSELLI-NORDBLATT, L. – GAUCKLER, B.: *New tools to improve statistical literacy – developments and projects*, 2018, s. 3.

Jedným z článkov zverejnených v Štatistike pre začiatočníkov je napríklad článok o trhu práce s tromi samostatnými časťami (základné definície, zamestnanosť a nezamestnanosť), ktorý ponúka pútavú a zároveň ľahko pochopiteľnú infografiku na vysvetlenie výpočtu miery nezamestnanosti a pojmu ekonomicky aktívne obyvateľstvo (obrázok č. 1). Rovnako prehľadne sú pomocou farebne atraktívnej infografiky v časti o medzinárodnom obchode znázornené a vysvetlené dovoz a vývoz tovarov ako aj schodok a prebytok obchodu (obrázok č. 2).

Obrázok č. 1: Miera nezamestnanosti; Ekonomicky aktívne obyvateľstvo
Obrázok č. 2: Vývoz a dovoz tovarov, obchodná bilancia



Zdroj: Eurostat

Okrem rôznych článkov sa Eurostat zameriava aj na tvorbu **digitálnych publikácií a aplikácií**. Ich prínos spočíva vo vysvetľovaní štatistík jednoduchou formou. Predstavujú kombináciu krátkych textov, interaktívnych nástrojov, vizualizačných prvkov, animácií a ilustrácií. Súčasťou bývajú aj mapy a videá. Neodborný čitateľ (používateľ) tak získa pomerne rýchlo prehľad o zložitej téme a vyberie si údaje, o ktoré má záujem.

V rámci spolupráce s Eurostatom majú NŠI možnosť pripraviť digitálne produkty a aplikácie vo svojich národných jazykoch. ŠÚ SR využil túto príležitosť a pripravil slovenskú verziu viacerých digitálnych publikácií Eurostatu: *Kľúčové údaje o Európe, Život mužov a žien v Európe – štatistický portrét, Ľudia v pohybe – štatistika mobility v Európe a Európske hospodárstvo od začiatku tisícročia – štatistický portrét.*

V slovenskom jazyku sú k dispozícii aj aplikácie Eurostatu *Ty v EÚ* a *Ekonomické trendy*, ktoré sú určené nielen pre začiatočníkov, ale aj pokročilých používateľov štatistických informácií. *Ty v EÚ* prináša používateľovi možnosť porovnať si vybrané aspekty života podľa veku a pohlavia s inými krajinami EÚ. Kvízové otázky sú zamerané na 3 oblasti: „ja a môj život“, „ja a moje životné podmienky“ a „ja a moja práca“. Prostredníctvom aplikácie *Ekonomické trendy* si môže každý záujemca, laik alebo odborník, rýchlo a interaktívne vizualizovať pomocou grafov vývoj jedného z 8 hlavných ukazovateľov ekonomiky: nezamestnanosť, inflácia, priemyselná výroba, verejné financie, verejný dlh, náklady práce, HDP a HDP na obyvateľa.

V úvode spomenutý projekt DIGICOM poukázal na fakt, že v súvislosti so zvyšovaním štatistickej gramotnosti je potrebná **spolupráca NŠI so vzdelávacím sektorom a školami**. Zrodila sa myšlienka usporiadania súťaže na európskej úrovni, do ktorej by boli ochotní na národných úrovniach zapojiť sa v dostatočnom počte študenti i učitelia. Na základe skúseností s organizovaním štatistických súťaží NŠI Španielska a iných európskych NŠI bolo v roku 2018 spustené 1. kolo **Európskej súťaže v štatistike**, do ktorej sa zapojilo 11 krajín, viac ako 11 000 študentov a približne 1 000 učiteľov. Dnes už má súťaž za sebou úspešné 3. kolo so 17 zúčastnenými krajinami.

Súťaž je zameraná na dve vekové kategórie: 14 až 16 ročných a 16 až 18 ročných. Rozdelená je na dve časti: národné kolo a európske finále. Finalisti národných kôl postupujú do európskeho finále, ktoré je v anglickom jazyku. Postúpia maximálne dva tímy z krajiny za každú vekovú kategóriu. Cieľom súťaže je na jednej strane vzbudiť u študentov záujem o oficiálnu štatistiku a podporiť u nich tímovú spoluprácu pri dosahovaní spoločných cieľov, na druhej strane nabádať učiteľov, aby využívali nové formy pri výučbe štatistiky.

Na podporu zvyšovania štatistickej gramotnosti sa v posledných rokoch začala využívať **gamifikácia**, ktorá predstavuje použitie herného prístupu v nehermom (reálnom) prostredí a javí sa ako vhodný spôsob na motiváciu a prilákanie záujmu študentov o štatistiku. Ku gamifikačným produktom patria mobilné aplikácie a tiež digitálne hry.

Mnohé štatistické úrady rozvíjajú aktivity v tejto oblasti so zámerom priblížiť prostredníctvom pútavých a trendových foriem štatistiku mladej generácii. NŠI Slovinska vyvinulo zaujímavú mobilnú aplikáciu s názvom „Junaki Slovenije“. Ide o geoštatistický kvíz pre viacerých hráčov zameraný na spoznávanie Slovinska a jeho minulosti prostredníctvom postavy doktora Edvarda Valvazorova. Poľský NŠI zverejnil edukačnú štatistickú hru „Stat Mission“ (vo formáte mobilnej aplikácie) určenú pre dospievajúcu mládež. Cieľom hry je nájsť známeho vedca profesora Statistix. Kvíz je zameraný na oficiálnu štatistiku a rozdelený do tematických oblastí – na každej planéte je prezentovaná iná oblasť.

Za zmienku stoja aj online hry pre študentov, ktoré vypracovali NŠI Španielska (INE Spain) a Portugalska. V hre INEcity, zverejnenej na stránke „[explica](#)“ si môžu žiaci vybudovať vlastné mesto na základe demografických a ekonomických štatistik v imaginárnom regióne. Aby daný región prosperoval, musí hráč investovať rozumne. NŠI Portugalska spustilo stránku „[ALE](#)“ určenú na podporu zvyšovania štatistickej

gramotnosti už v roku 1999. Na stránke je dnes k dispozícii osem rôznych štatistických hier.

Forma nástrojov využívaných rôznymi NŠI a Eurostatom na zvyšovanie štatistickej gramotnosti sa postupne modernizuje a preto je dôležité sledovať vývoj v tejto oblasti. ŠÚ SR bude aj naďalej pokračovať v úsilí priblížiť štatistiku mladej generácii využívajúc pri tom už dostupné produkty a nadviazaním užšej spolupráce so strednými školami v jednotlivých regiónoch Slovenska.

Ing. Monika KÚŠIKOVÁ

Autorka pracuje na odbore poskytovania informácií a marketingu Štatistického úradu SR.

Informácia/Information

KONFERENCIA NOVÉ TECHNIKY A TECHNOLOGIE PRE OFICIÁLNE ŠTATISTIKY (NTTS 2021)

CONFERENCE ON NEW TECHNIQUES AND TECHNOLOGIES FOR OFFICIAL STATISTICS (NTTS 2021)

Napriek sťaženým podmienkam spôsobeným epidemiologickou situáciou vzhľadom na pandémiu COVID-19 sa v marci 2021 uskutočnila pravidelná vedecká konferencia **New Techniques and Technologies for Official Statistics** (ďalej v skratke NTTS 2021). Konferencia sa konala online formou počas troch dní (9. 3. – 11. 3. 2021), ktoré boli pre 2400 účastníkov, viac ako 130 prednášajúcich a prezentujúcich z 80 krajín plné inšpirácií a stimulácií. Kľúčový príhovor, ktorý predniesol profesor z Oxfordskej univerzity, Danny Dorling na tému *Nový spôsob nazerania na staré údaje*, uviedol zameranie tém a jednotlivých príspevkov minimálne pre prvý deň, keď rezonovali najmä prezentácie o moderných nástrojoch na politickú podporu (regionálne mikrosimulácie), inovatívnych nástrojoch a metódach zberu v sociálnych štatistikách, nástrojoch na vizualizáciu, otvorených zdrojoch a priestorových modeloch.



Zdroj: https://twitter.com/EU_NTTS/status/1371420661080543235/photo/1

V prvý deň konferencie sa už tradične konalo aj udelenie Hackathon ceny Európske Big Data. Na prvom mieste bol tím z Poľska s prototypom dashboardu na analýzu údajov o využití času obyvateľov, na druhom mieste tím z Talianska s návrhom integrovaného prostredia na analýzu údajov, vizualizáciu a klasifikáciu potravín a na treťom mieste tím zo Spojeného kráľovstva s návrhom systému, ktorý vylepšuje tradičný dotazník o automatické spracovanie fotografií pokrmov.

Druhý deň bol zameraný na 15 samostatných aktuálnych štatistických tém napr. na použitie R v oficiálnych štatistikách, nové dátové zdroje pre štatistiky o mimoriadnych stavoch, sezónne odhady, SDMX, Enterprise architektúru, machine learning, nové dátové zdroje. Zároveň sa v tento deň udeľovala cena EMOS.

Posledný deň bol už tradične venovaný najmä na štatistickej gramotnosti, vizualizácii štatistík a zlepšovaniu štatistických zručností v oblasti šírenia a propagácie údajov.

Na konferencii boli najviac zastúpené národné štatistické úrady, európske inštitúcie, ale aj výskumníci, vládny sektor, súkromný a bankový sektor, medzinárodné inštitúcie a iní. Uvítací príhovor, ako aj uzatvorenie konferencie bolo v rukách generálnej riaditeľky Eurostatu p. Kotzevovej, ktorá zdôraznila predovšetkým dôležitosť inovácie údajov založených na lepších informáciách a boj s dezinformáciami.

PhDr. Ľudmila IVANČÍKOVÁ, PhD.

Autorka je generálna riaditeľka sekcie sociálnych štatistík a demografie Štatistického úradu SR.

PRIPRAVUJEME/COMING SOON

Boris FRANKOVIČ

VÝPOČET RÝCHLYCH ODHADOV V ŠTATISTIKE CESTOVNÉHO RUCHU
CALCULATION OF FLASH ESTIMATES IN TOURISM STATISTICS

Janka SZEMESOVÁ, Marcel ZEMKO, Martin PETRÁŠ

VYHODNOTENIE ŠTATISTICKÉHO ZISŤOVANIA O SPAĽOVACÍCH
ZARIADENIACH A SPOTREBE PALÍV V DOMÁCNOSTIACH
ASSESSMENT OF THE STATISTICAL SURVEY ON HEATING SYSTEMS AND
FUELS USED HOUSEHOLDS

Iveta STANKOVIČOVÁ

ETIKA V PODNIKATEĽSKOM PROSTREDÍ NA SLOVENSKU
ETHICS IN THE SLOVAK BUSINESS ENVIRONMENT

Boris VAŇO

AKÝ MÔŽE BYŤ DLHODOBÝ DEMOGRAFICKÝ VÝVOJ NA SLOVENSKU?
WHAT CAN BE THE LONG-TERM DEMOGRAPHIC DEVELOPMENT IN SLOVAKIA?

Helena GLASER-OPITZOVÁ

SEZÓNNOŠŤ V ÚMRTNOSTI
SEASONALITY IN MORTALITY RATES

Ľudmila IVANČÍKOVÁ – Boris VAŇO

HODNOTENIE KVALITY ADMINISTRATÍVNYCH ZDROJOV ÚDAJOV
QUALITY ASSESSMENT OF ADMINISTRATIVE DATA SOURCES

Róbert VLAČUHA – Silvia GREGORCOVÁ

CHUDOBA A SOCIÁLNE VYLÚČENIE NA SLOVENSKU
POVERTY AND SOCIAL EXCLUSION IN SLOVAKIA

* * *

**ONLINE VERZIA ČÍSLA 2/2021 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE JE
VEREJNE DOSTUPNÁ na internetovej stránke ssad.statistics.sk od 15. APRÍLA
2021.**

**THE ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND
DEMOGRAPHY No 2 (2021) IS PUBLICLY BE AVAILABLE at the website
ssad.statistics.sk from APRIL 15, 2021.**

INFORMÁCIE PRE PRISPIEVATEĽOV

Príspevky prijímame v slovenskom, v českom a v anglickom jazyku. Musia rešpektovať odborné zameranie časopisu a jeho vedecký charakter. Zaslaný príspevok nesmie byť v recenznom konaní v inom časopise, ani uverejnený v odbornej a inej tlači.

Príspevky zasielajte v elektronickej forme vo formáte MS Word alebo Open Office, typ písma Arial, veľkosť 12, riadkovanie 1. Nad titulkom treba uviesť meno autora a jeho pracovisko.

Súčasťou príspevku je abstrakt (základný popis cieľa a spôsobu spracovania faktov v rozsahu do 100 slov), kľúčové slová (maximálne 5), resumé (stručné zhrnutie obsahu článku s dôrazom na jeho prínos a najvýznamnejšie závery v rozsahu do 500 slov), profesijný životopis (v rozsahu do 120 slov) a kontakt (e-mailová adresa autora). Názov článku, abstrakt, kľúčové slová a resumé poskytne autor aj v anglickom jazyku. Zoznam použitej literatúry v abecednom poradí s úplnými bibliografickými údajmi sa uvádza na konci článku. Odkazy na literatúru sa uvádzajú v texte číslami v hranatých zátvorkách. Poznámky s poradovým číslom sú umiestnené pod čiarou na príslušnej strane textu, ku ktorému sa vzťahujú. Podrobnejšie pokyny nájdete autori na ssad.statistics.sk.

Rozsah vedeckých článkov je okolo 15 normostrán, informatívnych článkov 6 normostrán, recenzie, rozhovory a informácie publikujeme v rozsahu maximálne 3 normostrany. Tabuľky, mapy, grafy a obrázky musia mať názov a uvedený zdroj údajov; odporúčame, aby kopírovali šírku textu. Skratky sa používajú len minimálne, pri prvom použití je potrebné skratku v zátvorke rozpísať. Redakcia zabezpečuje jazykovú úpravu textu.

Príspevky sú recenzované. Oponentské konanie je obojstranne anonymné. Konečné rozhodnutie o publikovaní článku vydáva redakčná rada.

Redakcia si vyhradzuje právo zverejniť články schválené redakčnou radou v tlačenej a elektronickej podobe na ssad.statistics.sk.

INFORMATION FOR AUTHORS

Articles are accepted in Slovak, Czech and English languages and must comply with the journal's professional specialisation and scientific nature as well. The submitted articles should not be reviewed by another journal and should not have already been published in any specialised or other press.

Please submit your articles in electronic form, in MS Word or Open Office format, Arial font, size 12 and typed in single spacing. The author's name and workplace should be indicated above the title.

Articles should contain an abstract (general description of the objective and the processing methods used up to 100 words), key words (max. 5), resume (brief summary of the article's content emphasizing its contribution and the most important conclusions up to 500 words), curriculum vitae of the author (no more than 120 words) and the author's contact (e-mail address). The author should submit the article's title, abstract, key words and resume in English language. List of the literature used with full bibliographic data should be given in alphabetical order at the end of an article. Bibliographic citations should be given in square brackets. References are indicated by numbers in a text in square brackets. Footnotes should be numbered in the order of the corresponding page of a text. Authors can find more details at the website ssad.statistics.sk.

Scope of a scientific article is about 15 standard pages, informative articles should be up to 6 standard pages in length, reviews, discussions and information not more than 3 standard pages. Tables, maps, graphs and pictures should have a title and the data source indicated, it is also advised to copy the width of a text. Abbreviations should be used only rarely and should be appropriately explained in parentheses when first used. Language text revisions are provided by the editorial office.

Articles are reviewed. The opponent procedure is mutually anonymous. The final decision on the article's publication is made by the editorial board. The editorial office reserves the right to publish articles approved by the editorial board in printed and electronic form at the website ssad.statistics.sk.

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

je jediný recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov. Propagujeme miesto a význam slovenskej štatistiky v Európskom štatistickom systéme, spoluprácu Eurostatu a národných štatistických úradov pri harmonizácii zisťovaní a multidimenzionálny rozmer štatistiky. Podporujeme rozvoj štatistickej teórie a jej prepojenie s praxou. Naším cieľom je prispievať k využiteľnosti štatistických výstupov v rôznych oblastiach a k zvyšovaniu ich kvality a efektivity.

Publikujeme analytické články, prognózy, názory, diskusné príspevky, recenzie, rozhovory, informácie a oznamy z rôznych oblastí štatistiky (národné účty, produkčné štatistiky, sociálne štatistiky, štatistika životného prostredia a pod.) a demografie (demografická štatistika, teoreticko-metodologické východiská demografie, historická demografia a pod.), vrátane sčítania obyvateľov, domov a bytov ako neodmysliteľnej súčasti demografickej štatistiky.

Vydáva:

Štatistický úrad SR

Identifikačné číslo vydavateľa:

IČO 00166197

Vychádza:

Štyrikrát ročne

Dátum vydania:

15. apríl 2021

Tlač:

Reprografické stredisko
Štatistického úradu SR

Predplatné:

20 € (na rok)

5 € (za jeden výtlačok)

Objednávky prijíma:

Informačný servis
Štatistického úradu SR
Tel.: +4212/502 36 339
+4212/502 36 335
E-mail: info@statistics.sk

SLOVAK STATISTICS AND DEMOGRAPHY

is the only scientific reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures. Our aim is to promote the position and importance of Slovak statistics in the European Statistical System, cooperation between the Eurostat and the national statistical offices in the field of survey harmonisation and the multidimensional character of statistics as well. We support the development of statistical theory and its connection with practice. We aim to contribute to the utility of statistical outputs in various fields and to the improvement of quality and efficiency.

We publish analytic articles, prognoses, views, discussion contributions, reviews, discussions, information and announcements from various statistical fields (national accounts, production statistics, social statistics, environmental statistics etc.) and demography (demographic statistics, theoretical and methodological bases of demography, historical demography etc.) including the population and housing census as an essential part of demographic statistics.

Issued by:

Statistical Office of the SR

Company registration number:

00166197

Published:

Four times a year

Date of issue:

15th April 2021

Press:

Reprographic centre of the
Statistical Office of the SR

Subscription:

€20 (per year)

€5 (for one copy)

Orders are to be addressed to:

Information Service of the
Statistical Office of the SR
Tel.: +4212/502 36 339
+4212/502 36 335
E-mail: info@statistics.sk