METODIKA TESTOVANIA NEZÁVISLOSTI MEDZI KVALITATÍVNYMI ZNAKMI

Alena Košťálová*, Oľga Zajasenská*

Úvod

Cieľom predkladaného článku je predstaviť komplexne spracovanú metodiku testovania nezávislosti medzi hodnotami dvoch kvalitatívnych znakov štatistického súboru. Táto metodika bola využitá pri spracovaní dizertačnej práce zameranej na oblasť marketingového prieskumu poštových a telekomunikačných služieb v regionálnom podnikateľskom prostredí. Vedecký problém spracovania údajov získaných marketingovým prieskumom a následný rozbor výsledkov sa opieral o uplatnenie metodiky testovania nezávislosti medzi hodnotami dvoch kvalitatívnych znakov štatistického súboru, ktoré reprezentujú názory (odpovede) respondentov tvoriacich štatistický súbor.

Metódy spracovania údajov

Základnou úlohou spracovania údajov získaných marketingovým prieskumom je pretransformovať získané údaje do podoby vhodnej pre analýzu za účelom riešenia výskumného problému, od čoho závisí aj výber optimálnej metódy. Výber správnej štatistickej metódy analýzy údajov závisí od viacerých faktorov ako sú: typy údajov, ktoré sa skúmajú; nezávislosť výberu, teda akým spôsobom bol zo základného súboru vytvorený výberový súbor; aký je počet premenných, ktoré sú predmetom skúmania; veľkosť výberového súboru, teda koľko meraní máme k dispozícií; aké sú podmienky použitia danej metódy. Ak sú štatistické znaky štatistického súboru kvalitatívne – nominálne (slovné) a ordinárne (poradové), je to dôvod pre použitie metód **induktívnej štatistiky** pri triedení a kvantifikácii získaných údajov (najmä metódy testovania štatistických hypotéz).

Induktívne štatistické metódy sa členia na parametrické a neparametrické [1]:

- ✓ **parametrické** štatistické metódy narábajú s parametrami základných súborov, pričom vychádzajú zo špecifických predpokladov o rozdeleniach pravdepodobnosti základných súborov a výberových štatistík (napr. Fisherova analýza rozptylu),
- ✓ **neparametrické** štatistické metódy sa nespoliehajú na odhad parametrov charakterizujúcich rozdelenia premennej v základnom súbore. Preto sa tieto metódy niekedy označujú ako metódy s voľnými rozdeleniami. Neparametrické metódy pracujú s početnosťami (napr. chí-kvadrát test nezávislosti), alebo s poradovými číslami, ktoré boli pridelené pôvodným údajom (napr. Kruskal-Wallisov test).

* Ing. Alena Košťálová, PhD., Žilinská univerzita v Žiline, FPEDAS, Katedra spojov, Univerzitná 8215/1, 010 26 Žilina, tel.: +421 41 513 3143, e-mail: Alena.Kostalova@fpedas.uniza.sk

* doc. Ing. Oľga Zajasenská, CSc., Žilinská univerzita v Žiline, FPEDAS, Katedra spojov, Univerzitná 8215/1, 010 26 Žilina, tel.: +421 41 513 3138

Predtým, ako sa budeme venovať metóde testovania nezávislosti medzi hodnotami dvoch kvalitatívnych znakov štatistického súboru, je potrebné ozrejmiť metódu jednostupňového a dvojstupňového triedenia štatistického súboru. Výsledky triedenia pomocou týchto metód sú totiž východiskom metódy testovania nezávislosti.

Metóda jednostupňového triedenia štatistického súboru

Pri zameraní sa na jeden znak štatistického súboru je hlavným nástrojom analýzy jeho hodnôt **jednostupňové triedenie**. Úlohou jednostupňového triedenia je získať absolútne alebo relatívne početnosti výskytu jednotlivých hodnôt. Kumulatívne početnosti pri nominálnych znakoch vzhľadom na neusporiadateľnosť ich hodnôt nemajú zmysel, taktiež nelogicky pôsobí priemerná hodnota kvalitatívneho znaku [2].

Metóda dvojstupňového triedenia štatistického súboru

Širšie možnosti poskytuje analýza hodnôt dvoch nominálnych (alebo ordinárnych) znakov, ich výskytu a vzájomnej závislosti použitím **dvojstupňového triedenia.** Jeho výsledkom sú kontingenčné tabuľky. Kontingenčná tabuľka obsahuje [2]:

- ✓ **legendu**, kde sú jednotlivé varianty znaku A (označenie je $\mathbf{a_i}$, pre i =1, 2, ..., r),
- ✓ **hlavičku**, kde sú jednotlivé varianty znaku B (označenie je $\mathbf{b_j}$, pre j =1, 2, ..., s),
- ✓ **vlastné pole tabuľky**, do ktorého sa zapisujú tzv. simultánne (združené) početnosti, ktoré vyjadrujú, koľkokrát sa v súbore vyskytujú kombinácie variantov oboch znakov, teda výsledky dvojstupňového triedenia (označenie združených početností je $\mathbf{(a_i\ b_j)}$),
- ✓ **posledný riadok**, v ktorom sú súčty stĺpcových združených početností, tzv. marginálne početnosti (označenie je $\mathbf{(a_i)}$),
- ✓ **posledný stĺpec** v ktorom sú súčty riadkových združených početností, tzv. marginálne početnosti (označenie je $\mathbf{(b_j)}$),
- ✓ **priesečník posledného riadku a posledného stĺpca**, kde je rozsah súboru (označenie je **n**).

Tabuľka 1. Dvojrozmerné rozdelenia združených početností

A	B						(a_i)
	b_1	b_2	...	b_j	...	b_s	
a_1	$(a_1\ b_1)$	$(a_1\ b_2)$	...	$(a_1\ b_j)$	...	$(a_1\ b_s)$	(a_1)
a_2	$(a_2\ b_1)$	$(a_2\ b_2)$	...	$(a_2\ b_j)$	...	$(a_2\ b_s)$	(a_2)
...	...	...	...	...	...	...	...
a_i	$(a_i\ b_1)$	$(a_i\ b_2)$	...	$(a_i\ b_j)$	...	$(a_i\ b_s)$	(a_i)
...	...	...	...	...	...	...	...
a_r	$(a_r\ b_1)$	$(a_r\ b_2)$	...	$(a_r\ b_j)$	...	$(a_r\ b_s)$	(a_r)
(b_j)	(b_1)	(b_2)	...	(b_j)	...	(b_s)	n

Zdroj: PACÁKOVÁ, V. a kol.: Štatistika pre ekonómov. [3]

Simultánne (združené) početnosti $\mathbf{(a_i\ b_j)}$ sú v prvom rade absolútne početnosti, ktoré udávajú, koľko jednotiek zo súboru má súčasne i-tu hodnotu riadkovej premennej A a zároveň j-tu hodnotu stĺpcovej premennej B. Jednotlivé políčka tabuľky môžu obsahovať aj relatívne početnosti $\mathbf{f_i}$, ktoré môžu mať rozličný základ. Môžu sa počítať ako:

a) podiely z celkového rozsahu súboru **n**, teda podľa vzťahu:

$$\mathbf{f_i = (a_i\, b_j) / n}, \quad (1)$$

b) relatívne stĺpcové početnosti podľa vzťahu:

$$\mathbf{f_i = (a_i\, b_j) / (b_j)}, \quad (2)$$

c) relatívne riadkové početnosti podľa vzťahu:

$$\mathbf{f_i = (a_i\, b_j) / (a_i)}. \quad (3)$$

Riadkové a stĺpcové relatívne početnosti nazývame aj podmienené relatívne početnosti. Pri ich použití vzniká niekoľko problémov [2]. Najviac informácií sa získa, ak v jednotlivých políčkach tabuľky budú súčasne uvádzané tak absolútne početnosti, ako aj celkové, riadkové a stĺpcové relatívne početnosti. Výhodou je úplná informovanosť. Nevýhodou je, že v takejto koncentrovanej podobe je príliš veľa informácií a časť prezentovaného obsahu sa môže v tejto „hmle" početností strácať. Použitie len niektorých z relatívnych početností určuje, čo chceme prezentáciou výsledkov dvojstupňového triedenia získať, či zdôrazniť.

Testovanie nezávislosti medzi hodnotami dvoch znakov

Na skúmanie vzťahov medzi štatistickými znakmi sa využívajú metódy podľa toho, či ide o kvantitatívne, alebo kvalitatívne štatistické znaky. Pri kvalitatívnych znakoch sa na skúmanie vzťahov medzi nimi využíva metóda nazývaná **meranie asociácie**. Pri analýze súboru štatistických jednotiek na základe početností výskytu možných dvojíc hodnôt dvoch nominálnych alebo ordinárnych znakov A a B nastáva dôležitá otázka [2]:

- ✓ či medzi hodnotami týchto kvalitatívnych znakov **je nejaká závislosť**, asociácia (t.j., či s výskytom určitých hodnôt znaku A môžeme predpokladať výskyt niektorých hodnôt druhého znaku B),
- ✓ alebo či hodnoty týchto kvalitatívnych znakov **sú nezávislé** (výskyt hodnôt znaku A nepredurčuje výskyt hodnôt druhého znaku B).

Ak takáto závislosť neexistuje, znamená to, že ak pri štatistickej jednotke poznáme hodnotu znaku A, nedáva nám to žiadnu informáciu o možnej hodnote znaku B pri tejto jednotke. Podkladom pre skúmanie asociácie sú asociačné, resp. kontingenčné tabuľky a následne pre súhrnné testovanie existencie štatisticky významného vzťahu medzi kvalitatívnymi znakmi sa používa **χ^2 – štvorcová kontingencia** [4].

Pri testovaní nezávislosti medzi hodnotami dvoch kvalitatívnych znakov je **postup rozdelený do štyroch krokov**. Prvým krokom je formulovanie hypotézy H_0 oproti alternatívnej hypotéze H_1. Druhým krokom je stanovenie hladiny významnosti α. Tretím krokom je výpočet testovacej štatistiky a pravdepodobnosti. Štvrtým krokom je porovnanie vypočítanej hodnoty χ^2 s kritickou hodnotou a prijatie štatistického rozhodnutia.

1. krok - Formulovanie hypotéz

H_0: medzi hodnotami znakov A a B nie je závislosť (nie je asociácia)

H_1: medzi hodnotami znakov A a B je závislosť (je asociácia)

Konečným cieľom väčšiny štatistických testov je zhodnotenie vzťahu medzi premennými. Nulová hypotéza vyjadruje nezávislosť premenných. Oproti tomu alternatívna hypotéza, ktorej pravdivosť väčšinou chceme dokázať, najčastejšie vyjadruje štatistickú závislosť premenných. Pravdivosť alternatívnej hypotézy sa dokazuje vždy iba nepriamo a to tak, že ukážeme, že nulová hypotéza je nepravdepodobná a alternatívna (jediná zostávajúca) je teda pravdepodobná.

Nezávislosť je overovaná **chí-kvadrát testom**. Podľa A. Parasuramana sa chí-kvadrát test používa na zistenie (sprostredkujúci prameň [5]):

- ✓ či sa skúmané rozdelenie početností odpovedí nominálne škálovanej premennej štatisticky významne odlišuje od očakávaného rozdelenia,
- ✓ či možno rozdelenie respondentov z hľadiska triedenia podľa určitých kategórií považovať za zhodné s rozdelením základného súboru.

Chí-kvadrát rozdelenie má prvoradý význam pri analýze závislosti v asociačných a kontingenčných tabuľkách [2]. Najčastejšou situáciou použitia chí-kvadrát rozdelenia je testovanie zhody tabuľky s niektorým teoretickým modelom. Ide o porovnávanie empirických (nameraných) početností $(\mathbf{a_i\ b_j})$ a teoretických početností $(\mathbf{a_i\ b_j})_\mathbf{0}$, teda takých, aké by mali empirické početnosti byť, keby boli znaky A a B nezávislé. Pri určovaní teoretických početností vychádzame z vety o nezávislosti náhodných javov A a B [4]:

$$\mathbf{P(A \cap B) = P(A) * P(B).} \tag{4}$$

Ak sú znaky A a B nezávislé, potom platí:

$$\mathbf{P(a_i\, b_j) = P(a_i) * P(b_j).} \tag{5}$$

Na základe relatívnych početností:

$$\frac{(a_i\ b_j)_0}{n} = \frac{(a_i)}{n} * \frac{(b_j)}{n} \tag{6}$$

odhadneme teoretické početnosti $(\mathbf{a_i\ b_j})_\mathbf{0}$:

$$(a_i\ b_j)_0 = \frac{(a_i) * (b_j)}{n}. \tag{7}$$

2. krok – Stanovenie hladiny významnosti α

Hladina významnosti je pravdepodobnosť tzv. chyby prvého druhu, ktorú urobíme ak zamietneme nulovú hypotézu, ktorá v skutočnosti platí. Teda ak dospejeme k záveru, že medzi premennými existuje vzťah, pričom medzi nimi vzťah nie je. Chybu prvého druhu môžeme zmenšiť, ak za α zvolíme menšie číslo. Tým však umožňujeme prijatie hypotézy a súčasne väčšmi umožníme vznik tzv. chyby druhého druhu. Táto spočíva v tom, že nezamietneme nesprávnu hypotézu. Stanovenie hladiny významnosti je teda akýmsi kompromisom medzi týmito dvoma druhmi chýb a jeho voľba závisí od charakteru testovaných skutočností, od skúseností a pod. [6]

Alfa (α) sa tradične stanovuje na hodnotu 5 % (= 0,05), alebo 1 % (= 0,01). Odchýlky, ktoré sa vyskytujú s pravdepodobnosťou menšou ako je zvolená hladina významnosti, sa nazývajú štatisticky významné (signifikantné) pri zvolenej hladine významnosti.

3. krok - Výpočet testovacej štatistiky χ^2 a pravdepodobnosti

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{[(a_i\ b_j) - (a_i\ b_j)_0]^2}{(a_i\ b_j)_0} \tag{8}$$

Pomocou testovacej štatistiky podľa vzťahu (8) sa zistí štatistická hodnota testu chí-kvadrát pre empirické a teoretické početnosti. Vypočíta sa zo vzorky, ktorá má za predpokladu pravdivosti nulovej hypotézy príslušné rozdelenie pravdepodobnosti (chí-kvadrát). Okrem testovacej štatistiky sa vyčísľuje aj pravdepodobnosť, tzv. **P-hodnota** (Probability Level), pre ktorú platí nasledovné:

- ✓ predstavuje pravdepodobnosť, že testovacia štatistika za predpokladu nulovej hypotézy dosiahne prinajmenšom tak extrémnu hodnotu ako je hodnota vypočítaná zo vzorky,
- ✓ je to pravdepodobnosť, že vzťah zistený zo získaných údajov je iba dôsledkom „nešťastne zvolenej“ vzorky a ak by bola vybratá ďalšia náhodná vzorka, vzťah z nových údajov by nemusel byť zistený,
- ✓ je to najnižšia hodnota hladiny významnosti, ktorá vedie k zamietnutiu nulovej hypotézy,
- ✓ je to odhadovaná pravdepodobnosť zamietnutia pravdivej nulovej hypotézy,
- ✓ čím menšia je P-hodnota, tým viac možno byť presvedčený, že nulová hypotéza nie je pravdivá a mala by byť zamietnutá.

K. Pearson, ktorý ako prvý navrhol testovaciu štatistiku χ^2 (sprostredkujúci prameň [2]), dokázal, že ak teoretické hodnoty $\mathbf{(a_i\, b_j)_0}$ nebudú príliš malé (vyžaduje sa $(a_i\, b_j)_0 > 5$), rozdelenie premennej χ^2 bude približne takým istým ako chí-kvadrát rozdelenie, pričom počet stupňov voľnosti **d** sa rovná:

$$\mathbf{d = (r - 1) * (s - 1),} \tag{9}$$

kde **r** je počet riadkov a **s** je počet stĺpcov kontingenčnej tabuľky (znak A má **r** – úrovní, obmien a znak B má **s** – úrovní, obmien). Podľa W. G. Cochrana [7], [8], [9], teoretických početností menších ako 5 môže byť najviac 20 %, ale aj tieto musia byť väčšie ako 1:

$$\mathbf{P[(a_i\, b_j)_0 < 5] \leq 0{,}2 \;\wedge\; (a_i\, b_j)_0 > 1.} \tag{10}$$

V prípade, že podmienka (10) nie je splnená, je potrebné dodatočne nájsť príbuzné triedy (do ktorých bol štatistický súbor roztriedený), ktoré spolu vecne súvisia a ktoré by bolo možné zlúčiť, čím sa následne zmení aj počet stupňov voľnosti. [10]

Podmienka (10) je predpokladom, alebo tiež obmedzením pre použitie Pearsonovho chí-kvadrát testu a výpočet testovacej štatistiky χ^2 v tabuľkách s viac ako 2 riadkami a viac ako 2 stĺpcami. Pre použitie Pearsonovho chí-kvadrát testu v tabuľkách s rozmermi 2 x 2 platia nasledujúce obmedzenia [7], [11]:

- ✓ všetky teoretické početnosti musia byť väčšie alebo rovné 10,
- ✓ ak je niektorá z teoretických početností menšia ako 10, ale väčšia alebo rovná 5, niektorí autori odporúčajú použiť Yatesovu korekciu spojitosti[1],
- ✓ ak je niektorá z teoretických početností menšia ako 5, použije sa niektorý z iných testov (napr. Fisherov exaktný test pre kontingenčné tabuľky s rozmermi 2 x 2).

4. krok - Prijatie štatistického rozhodnutia o hypotéze H_0:

Pri prijímaní štatistického rozhodnutia o hypotéze H_0 je potrebné sa riadiť nerovnosťou medzi vypočítanou hodnotou testovacieho kritéria a príslušnou kritickou hodnotou, ktorou je vymedzená kritická oblasť pre chí-kvadrát:

$$\chi^2 \geq \chi^2_{1-\alpha}\,[(r-1)*(s-1)] \tag{12}$$

---

[1] **Yatesova korekcia spojitosti** spočíva v tom, že pri výpočte testovacej štatistiky χ^2 podľa vzťahu (8), sa pred umocnením rozdielu empirickej a teoretickej početnosti odčíta hodnota 0,5 od absolútnej hodnoty tohto rozdielu [12], (táto korekcia je však kontroverzná a neodporúčajú ju všetci autori).

$$\chi^2(Yates) = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{[\,|\,(a_i\, b_j) - (a_i\, b_j)_0\,| - 0{,}5\,]^2}{(a_i\, b_j)_0} \tag{11}$$

Pri porovnávaní vypočítanej hodnoty testovacieho kritéria s príslušnou kritickou hodnotou χ^2 pre hladinu významnosti α (α = 0,05) so stupňami voľnosti d = (r - 1) * (s - 1), ktorá je tabelovaná, môžu nastať dva prípady [6]:

a) Vypočítaná hodnota testovacieho kritéria je väčšia alebo rovná ako kritická hodnota, t.j. vzťah (12) platí. To však znamená, že nastal prípad, ktorý sme za predpokladu správnosti nulovej hypotézy očakávali s takou malou pravdepodobnosťou (α = 0,05),že ho môžeme považovať za takmer nemožný. Ak aj napriek malej pravdepodobnosti nastal, predpokladáme, že testovaná odchýlka nemá charakter náhodný a že teda jej veľkosť je výrazom skutočného rozdielu medzi vlastnosťami, vyjadrenými porovnávanými štatistickými charakteristikami. Preto **nulovú hypotézu zamietame** pri zvolenej hladine významnosti (hypotéza H_1 je preukázaná) a tvrdíme, že rozdiel medzi početnosťami zistenými vo vzorke a očakávanými početnosťami je príliš veľký na to, aby bol iba dôsledkom náhodného výberu. **Sledovaný rozdiel je štatisticky významný. Medzi premennými teda existuje vzťah.** Čím je α menšie, tým má zamietnutie H_0 väčší význam, tzn. jej zamietnutie je zriedkavejšie nesprávne.

b) Vypočítaná hodnota testovacieho kritéria je menšia ako kritická hodnota, t. j. vzťah (12) neplatí. To znamená, že ide o prípad, ktorý sme za predpokladu, že nulová hypotéza je správna, očakávali s takou veľkou pravdepodobnosťou (0,95), že jeho výskyt môžeme považovať za istý. Predpokladáme preto, že testovaná odchýlka má charakter náhodný. **Nulovú hypotézu v tomto prípade nemožno zamietnuť** pri zvolenej hladine významnosti a tvrdíme, že rozdiel medzi početnosťami zistenými vo vzorke a očakávanými početnosťami môže byť dôsledkom náhodného výberu. **Sledovaný rozdiel nie je štatisticky významný.** To však neznamená, že nulová hypotéza je správna. Správne rozhodnutie je, že nemáme dostatočné dôkazy na to, aby sme nulovú hypotézu zamietli. **Teda nemáme dostatok dôkazov na to, aby sme mohli tvrdiť, že medzi premennými existuje vzťah.** Hypotézu H_1 možno alebo považovať za nepreukázanú a zdržať sa záveru o testovanej H_0, alebo možno H_0 prijať, čo však vyžaduje mať pod kontrolou pravdepodobnosť chyby II. druhu.

Na testovanie nezávislosti medzi hodnotami dvoch kvalitatívnych znakov štatistického súboru hodnôt získaných marketingovým prieskumom je vhodné využiť funkciu CHITEST v prostredí Microsoft Excel. Výsledkom funkcie CHITEST je hodnota pravdepodobnosti chí-kvadrát rozdelenia pre danú testovaciu štatistiku a príslušný počet stupňov voľnosti (P-hodnota), ktorá bude využitá na rozhodnutie o hypotéze H_0 [2]:

- ✓ ak je hodnota pravdepodobnosti chí-kvadrát rozdelenia **v intervale od 0 po α**, (P-hodnota < α), potom **hypotézu H_0 zamietame** voči príslušnej alternatívnej hypotéze; tj. **medzi znakmi A a B je štatisticky významná závislosť** (asociácia),
- ✓ ak je hodnota pravdepodobnosti chí-kvadrát rozdelenia **v intervale od α po 1**, (P-hodnota $\geq$ α), potom **hypotézu H_0 reprezentujúcu nezávislosť znakov A a B nemožno zamietnuť**.

V nasledujúcej tabuľke 2. sú uvedené situácie, ktoré môžu nastať pri testovaní hypotéz v súvislosti s výskytom chyby I. a II. druhu.

Tabuľka 2. Situácie, ktoré môžu nastať pri testovaní hypotéz

Skutočnosť	Rozhodnutie	
	H_0 nezamietnutá	H_0 zamietnutá
H_0 je pravdivá	Správne rozhodnutie ($p = 1 - \alpha$)	Chyba I. druhu ($p = \alpha$)
H_0 je nepravdivá	Chyba II. druhu ($p = \beta$)	Správne rozhodnutie ($p = 1 - \beta$)

Zdroj: RIMARČÍK, M.: Testy štatistických hypotéz [13]

Vysvetlivky: p ... pravdepodobnosť nastatia danej situácie
α ... hladina významnosti (Significance Level)
$1 - \alpha$ ... spoľahlivosť (Confidence Level)
$1 - \beta$ ... sila testu (Power)

Miery stupňa závislosti (asociácie) medzi dvomi znakmi

Pomocou chí-kvadrát testu je možné získať informáciu, či vzájomnú závislosť medzi hodnotami kvalitatívnych znakov A a B je možné považovať za štatisticky významnú alebo nie. V prípade štatistickej významnosti tejto závislosti (asociácie) však stupeň tejto závislosti možno hodnotiť len nepriamo, pomocou hodnoty pravdepodobnosti chí-kvadrát rozdelenia (výsledok funkcie CHITEST). Čím je hodnota pravdepodobnosti bližšia k nule, tým je stupeň vzájomnej závislosti vyšší.

Autori Chajdiak – Komorník – Komorníková [2] uvádzajú nasledovné **štatistické miery vzájomnej závislosti nominálnych znakov**, ktoré sú odvodené od vypočítanej hodnoty testovacieho kritéria - χ^2 hodnoty:

- **Phi koeficient** (Phi je anglickým prepisom gréckeho písmena φ)
  - predstavuje druhú odmocninu z pomeru χ^2 hodnoty k rozsahu súboru n,
  - nadobúda hodnoty, ktoré sa rovnajú, alebo sú vyššie ako 0 (nevýhodou Phi koeficientu je práve neexistencia horného ohraničenia),
  - využíva sa len v tabuľkách s rozmermi 2 x 2 [14],
  - interpretácia - čím viac sa jeho hodnota blíži k 0, tým je vyšší stupeň nezávislosti medzi hodnotami znakov A a B.

$$\text{Phi koeficient} = \sqrt{\frac{\chi^2}{n}} \tag{13}$$

- **Kontingenčný koeficient (Contingency Coefficient, Pearsonov koeficient C)**
  - predstavuje druhú odmocninu z pomeru χ^2 hodnoty k súčtu χ^2 hodnoty a rozsahu súboru n,
  - nadobúda hodnoty z intervalu <0,1),
  - predstavuje upravený Phi koeficient určený pre použitie v tabuľkách s rozmermi väčšími ako je 2 x 2,
  - jeho maximum sa blíži, ale nikdy úplne nedosiahne hodnotu 1, iba vo veľkých tabuľkách, preto ho niektorí autori odporúčajú iba pre tabuľky s rozmermi 5 x 5 a väčšie [14] (jeho maximálna hodnota v tabuľke s rozmermi 2 x 2 je 0,707),
  - interpretácia - čím viac sa jeho hodnota blíži k 0, tým je vyšší stupeň nezávislosti medzi hodnotami znakov A a B.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \tag{14}$$

- **Cramerovo V (Cramer´s V)**
  - vypočíta sa ako druhá odmocnina z pomeru χ^2 hodnoty k súčinu rozsahu súboru n a hodnoty h; kde h je minimum z čísel (r - 1) a (s - 1),
  - nadobúda hodnoty z intervalu <0,1>,
  - je to najčastejšie používaná miera asociácie, pretože nadobúda hodnoty od 0 (žiadny vzťah) po 1 (dokonalý vzťah) bez ohľadu na veľkosť tabuľky [14],
  - má tendenciu dosahovať nižšiu hodnotu ako Phi koeficient alebo kontingenčný koeficient,
  - interpretácia - čím viac sa jeho hodnota blíži k 0, tým je vyšší stupeň nezávislosti medzi hodnotami znakov A a B.

$$V = \sqrt{\frac{\chi^2}{n * h}} \qquad (15)$$

- **Čuprovov koeficient** [3]
  - vypočíta sa ako odmocnina z pomeru priemernej χ^2 hodnoty (priemernej štvorcovej kontingencie) k odmocnine súčinu čísel (r - 1) a (s - 1), kde r je počet obmien znaku A a s je počet obmien znaku B,
  - nadobúda hodnoty z intervalu <0,1>,
  - hodnotu 1 dosahuje v štvorcových tabuľkách, kde sú riadkové marginálne početnosti identické so stĺpcovými marginálnymi početnosťami [14],
  - využíva sa zriedkavo a nie je podporovaný základným štatistickým softvérom,
  - interpretácia - čím viac sa jeho hodnota blíži k 0, tým je vyšší stupeň nezávislosti medzi hodnotami znakov A a B.

$$\tau = \sqrt{\frac{\phi^2}{\sqrt{(r-1)*(s-1)}}} \ , \ \text{kde } \phi^2 = \frac{\chi^2}{n} \qquad (16)$$

Pri zisťovaní stupňa vzájomnej závislosti medzi hodnotami dvoch kvalitatívnych znakov je vhodné, v zmysle vyššie spomenutých obmedzení, používať jednotlivé miery asociácie nasledovne[2]:

- Phi koeficient – iba v prípadoch, keď má tabuľka rozmery 2 x 2,
- Kontingenčný koeficient (Pearsonov) – pre tabuľky s rozmermi 5 x 5 a väčšie,
- Cramerovo V – pre tabuľky s rozmermi väčšími ako 2 x 2 ale menšími ako 5 x 5,
- Čuprovov koeficient – iba pre štvorcové tabuľky, kde sú riadkové marginálne početnosti identické so stĺpcovými marginálnymi početnosťami.

Pri interpretácii týchto mier asociácie možno použiť škálu, ktorú zaviedol Cohen (1988) pre korelačný koeficient (0,1 až 0,3 značí slabú koreláciu; 0,3 až 0,5 značí strednú koreláciu a 0,5 až 1,0 značí silnú koreláciu [15]). Nie je však nevyhnutné túto škálu striktne dodržiavať, a to najmä v sociálnych vedách.

Interval 0,5 až 1,0 vyjadrujúci silnú koreláciu je pri zisťovaní stupňa vzájomnej závislosti medzi hodnotami dvoch kvalitatívnych znakov veľmi široký a preto je vhodné rozdeliť ho do viacerých menších intervalov, ako je uvedené v tabuľke 3. [16]

[2] všetky spomenuté miery asociácie sú symetrické, tzn. pri ich výpočte nie je dôležité, ktorý zo znakov A a B je stĺpcovou premennou

Tabuľka 3. Interpretácia miery stupňa závislosti medzi dvomi znakmi

Hodnota miery stupňa závislosti	Stupeň závislosti
0,0	úplná nezávislosť
0,0 – 0,1	triviálna závislosť
0,1 – 0,3	slabá závislosť
0,3 – 0,5	stredná závislosť
0,5 – 0,7	silná závislosť
0,7 – 0,9	veľmi silná závislosť
0,9 – 1,0	takmer úplná závislosť
1,0	úplná závislosť

Zdroj: Cohen´s scale for correlation coefficient. [16]

Záver

Úlohou marketingového prieskumu je získavať objektívne, včasné, úplné a spoľahlivé informácie o skúmanom jave. Pri spracovávaní získaných informácií je možné využiť viaceré metódy. Jednou z nich je aj metóda testovania nezávislosti medzi vybranými dvojicami kvalitatívnych znakov štatistického súboru. V prípade potvrdenia štatisticky významnej závislosti medzi hodnotami týchto znakov sa následne stanovuje stupeň ich vzájomnej závislosti pomocou niektorej miery asociácie, za predpokladu dodržania pravidiel pre ich použitie.

Ak je marketingový prieskum zameraný na využívanie poštových a telekomunikačných služieb, tak ako tomu bolo v prípade riešenia spomínanej dizertačnej práce, je vhodné využiť výsledky testovania nezávislosti aj v oblasti marketingovej komunikácie jednotlivých poskytovateľov poštových a telekomunikačných služieb za účelom zvýšenia kvality týchto služieb.

Literatúra

[1] RIMARČÍK, M. *Testy normality*. [online]. [cit. 2009-05-07]. Dostupné na internete: <http://rimarcik.com/navigator/normal.html>.

[2] CHAJDIAK, J. - KOMORNÍK, J. - KOMORNÍKOVÁ, M. *Štatistické metódy*. Bratislava : STATIS, 1999. ISBN 80-85659-13-1.

[3] PACÁKOVÁ, V. a kol. *Štatistika pre ekonómov*. Bratislava : Edícia EKONÓMIA, 2003. ISBN 80-89047-74-2.

[4] SOJKOVÁ, Z. *Materiál na prednášky z predmetu Ekonomická štatistika*. Slovenská poľnohospodárska univerzita Nitra, Fakulta ekonomiky a manažmentu, Katedra štatistiky a operačného výskumu. [online]. [cit. 2009-04-07]. Dostupné na internete: <http://www.fem.uniag.sk/ACADEMIC/depart/ksov/predmety/statistika>.

[5] LESÁKOVÁ, D.- RAJT, Š. *Marketingové analýzy a prognózy*. Bratislava : EKONÓM, 1996. ISBN 80-225-0743-1.

[6] LINCZÉNYI, A. *Inžinierska štatistika*. Bratislava : ALFA, 1974. 63-025-74.

[7] COCHRAN, W. G. Some methods for strengthening the common χ^2 tests. In *Biometrics,* 1954. 10: pp 417–451.

[8] *Pearson's chi-square test*. [online]. [s.a.]. [cit. 2009-08-28]. Dostupné na internete: <http://en.wikipedia.org/wiki/Pearson's_chi-square_test>.

[9] YATES, D. - MOORE, D. - McCABE, G. *The Practice of Statistics*. 1st Ed. New York : W.H. Freeman, 1999.

[10] HINDLS, R. - HRONOVÁ, S. - SEGER, J. *Statistika pro ekonomy*. Druhé vydání. Praha : Professional Publishing, 2002. ISBN 80-86419-30-4.
[11] COCHRAN, W. G. The χ^2 test of goodness of fit. In *Annals of Mathematical Statistics,* 1952. 25:315–345.
[12] YATES, F. Contingency table involving small numbers and the χ^2 test. In *Supplement to the Journal of the Royal Statistical Society*. JSTOR Archive for the journal, 1934**.** 1(2): 217-235.
[13] RIMARČÍK, M. *Testy štatistických hypotéz*. [online]. [cit. 2009-05-07]. Dostupné na internete: <http://rimarcik.com/navigator/hypotezy.html>.
[14] *Nominal Association: Phi, Contingency Coefficient, Tschuprow´s T, Cramer´s V, Lambda, Uncertainty Coefficient*. [online]. [s.a.]. [cit. 2009-08-28]. Dostupné na internete: <http://faculty.chass.ncsu.edu/garson/PA765/assocnominal.htm>.
[15] COHEN, J. *Statistical power analysis for the behavioral sciences*. 2nd ed., Hillsdale, New Jersey: Lawrence Erlbaum, 1988.
[16] *Cohen´s scale for correlation coefficient.* [online]. [s.a.]. [cit. 2009-08-31]. Dostupné na internete: <http://www.sportsci.org/resource/stats/effectmag.html>**.**
[17] KOŠŤÁLOVÁ, A. *Marketingový prieskum využívania poštových a telekomunikačných služieb v regionálnom podnikateľskom prostredí* : doktorandská dizertačná práca. Žilina : Žilinská univerzita v Žiline. 2009. 123 s.
[18] PAĎOUROVÁ, A. - KOŠŤÁLOVÁ, A. - PAVLIČKO, M. - ŠVÁBOVÁ, L. *Využívanie poštových a telekomunikačných služieb v regionálnom podnikateľskom prostredí so zameraním na malé a stredné podniky v regióne Žilina* : Inštitucionálny projekt č. I16-07-140. Žilina : Žilinská univerzita v Žiline, FPEDAS, 2008.

Grantová podpora

SUROVEC, P. a kol: Socio-ekonomické dáta v prognóze a modelovaní dopravy pri napĺňaní ekonomickej funkcie regiónu. Grantový projekt VEGA MŠ SR a SAV č. 1/0344/08, Žilinská univerzita v Žiline, FPEDAS, 2008.

JANKALOVÁ, M. a kol: Poskytovanie verejnej telefónnej služby a spôsoby jej hodnotenia v procese globalizácie. Grantový projekt VEGA MŠ SR a SAV č.1/0709/08, Žilinská univerzita v Žiline, FPEDAS, 2008.