



VÝPOČET VZORKY RESPONDENTOV V MARKETINGOVOM PRIESKUME PRI JEDNODUCHOM A VIACSTUPŇOVOM DELENÍ

Ivan Stríček¹

Abstrakt

Článok je zameraný na určenie efektívneho rozdelenia vzorky v marketingovom výskume pri jednoduchom a viacstupňovom delení. Na praktickom príklade si ukážeme výpočet vzorky a vplyv jej rozdelenia do vrstiev na konečnú chybu výberu.

Úvod

Súčasný marketing sa orientuje na uspokojenie cieľových skupín zákazníkov prostredníctvom čo najlepšej alokácie zákazníckych potrieb a požiadaviek. Z tohto dôvodu firmy vynakladajú veľké finančné prostriedky do rôznych výskumov zameraných na prieskum trhu zákazníkov. Druhy výskumov, využívané pre zistenie potrebných informácií, sa líšia na základe rôznych aspektov a hĺbky štruktúry údajov, získaných z jednotlivých prieskumov. Preto variabilitou jednotlivých druhov výskumov vzniká aj rôznorodosť prístupov pri výpočte reprezentatívnej vzorky respondentov, ktorá určuje vierohodnosť a zmyslupnosť výsledkov uskutočneného výskumu pre ich uplatnenie na širšiu masu zákazníkov.

Keďže existuje viacero spôsobov výpočtu vzorky, často sa stretávame s ich zámenou, prípadne nevhodným využitím vzhľadom na cieľ výskumu. Preto si objasníme základné spôsoby výpočtu a ich aplikáciu na vzorovom príklade.

Výpočtová základňa

Pred samotným výpočtom si predstavíme vzorce používané pre jednotlivé delenia a ich premenné.

Základný najuniverzálnejší vzorec pre výpočet veľkosti n , ktorý používa väčšina autorov zaoberajúcich sa danou problematikou je

$$n = \frac{(z_{1-\alpha/2}^2 * \sigma^2)}{E^2} \quad (1)$$

kde:

- $z_{1-\alpha/2}$ je kritická hodnota určená z tabuliek (pri 95% spoľahlivosti 1,96),
- σ^2 je rozptyl vypočítaný zo smerodajnej odchýlky $p*(1-p)$,
- E je výberová chyba (ktorú si zvolíme).

¹ Ing. Ivan Stríček, Žilinská univerzita v Žiline, Fakulta Prevádzky a ekonomiky dopravy a spojov, Katedra spojov, Univerzitná 1, 010 26 Žilina, tel.: +4215133144, fax: +4215655615, e-mail: ivan.stricek@fpedas.uniza.sk

Ak však N nie je dostatočne veľké, teda poznáme, alebo vieme odhadnúť jeho veľkosť a tá je nižšia ako 100 000 respondentov, je nutné pre lepšiu presnosť vzorky použiť nasledovné výpočty.

Pri neznámom rozdelení základného súboru použijeme vzťah

$$n = \frac{z_{1-\alpha/2}^2 * \sigma^2 * \left[\frac{N}{(N-1)} \right]}{E^2 + \left[\frac{z_{1-\alpha/2}^2 * \sigma^2}{(N-1)} \right]}, \quad (2)$$

kde:

- $z_{1-\alpha/2}^2$ je kritická hodnota určená z tabuliek (pri 95% spoľahlivosti 1,96),
- σ^2 je rozptyl vypočítaný zo smerodajnej odchýlky $p*(1-p)$,
- E je výberová chyba (ktorú si zvolíme),
- N je veľkosť základného súboru.

Chyba pri vrstvenom výbere

Pri výpočte chyby vrstveného výberu je potrebné zahrnúť do výpočtu podielového delenia jednotlivých vrstiev a preto nepostačujú vzorce pre výpočet vzorky jednoduchého náhodného výberu. Znova si môžeme vzťahy rozdeliť podľa známosti rozdelenia. Potom pri neznámom (odhadovanom) rozdelení použijeme vzťah

$$E = \frac{z_{1-\alpha/2}}{N} * \sqrt{\sum_{h=1}^k \left[N_h^2 * \left(1 - \frac{n_h}{N_h} \right) * \left(\frac{p_h * (1 - p_h)}{n_h - 1} \right) \right]}, \quad (3)$$

kde:

- n_h je veľkosť vzorky pre vrstvu h ,
- N_h je veľkosť populácie vo vrstve h ,
- p_h je podiel želaného javu vo vrstve h ,
- N je veľkosť populácie,
- k je počet vrstiev,
- $z_{1-\alpha/2}$ je kritická hodnota určená z tabuliek (pri 95% spoľahlivosti 1,96).

Vzorové riešenie

Aj keď existuje priamy vzťah na výpočet vzorky pri viacstupňovom delení, jedná sa o vzťah prepočítavajúci postupne jednotlivé vrstvy zvlášť. Preto je jednoduchšie vzorku vypočítať jedným zo základných vzťahov (1), (2) a stanoviť ju nanovo po prepočítaní chyby vzťahom (3). Dá sa povedať, že ide o tzv. strieľanie na slepo, no to posluži najmä ak pravdepodobnosti želaných javov vo vrstvách len odhadujeme.

V nasledovnom príklade si ukážeme výpočet vzorky pri jednoduchom delení základného súboru a jej následný prepočet po zohľadnení diverzity aspektu želaného javu pri viacstupňovom delení.

Příklad:

Chceme vykonať prieskum využitia poštových služieb v malých a stredných podnikoch, pôsobiacich v okrese Bytča. Podniky si rozdelíme podľa počtu zamestnancov a výskum by mal mať max. 3 % odchýlku pri 95 % intervale spoľahlivosti. Potrebujeme teda vypočítať, aká bude celková vzorka a koľko podnikov v jednotlivých podskupinách potrebujeme zaradiť do prieskumu, aby sme zaručili stanovenú chybu výberu.

Riešenie:

Pre výpočty celkovej vzorky sa použije vzťah (1) pre malý základný súbor. Keďže ide o prieskum využitia poštových služieb, ktoré využíva prakticky každý podnik, môžeme dať odhad pravdepodobnosti želaného stavu na 80 %. Zo štatistického úradu sme si zistili, že v bytčianskom okrese sa nachádza 303 malých a stredných podnikov.

Výpočet vzorky potom bude vyzeráť nasledovne:

$$n = \frac{z_{1-\frac{\alpha}{2}}^2 * \sigma^2 * \left[\frac{N}{(N-1)} \right]}{E^2 + \left[\frac{z_{1-\frac{\alpha}{2}}^2 * \sigma^2}{(N-1)} \right]} = \frac{1,96^2 * 80 * (100 - 80) * \left[\frac{303}{(303 - 1)} \right]}{3^2 + \left[\frac{1,96^2 * 80 * (100 - 80)}{(303 - 1)} \right]} = \frac{6\,166,9129}{9 + 20,3529}$$

$$= 210,1$$

Z výpočtu vidíme, že pre daný prieskum nám postačí vzorka 210 respondentov. Môžeme si ukázať aj alternatívny vývoj vzorky pri inak zvolenej spoľahlivosti odhadu, alebo chybe (viď. Tab. 1.). Pravdepodobnosť želaného stavu ponecháme na 80 %.

Tabuľka 1. Vývoj veľkosti vzorky pri zmene spoľahlivosti a chyby výberu

Spoľahlivosť v [%]	Hodnota z	Chyba v [%]	Veľkosť základného súboru	Veľkosť vzorky
75	1,15	1	303	265,2
75	1,15	2	303	192,9
75	1,15	3	303	132,6
90	1,65	1	303	283,4
90	1,65	2	303	237,2
90	1,65	3	303	186,6
95	1,96	1	303	288,8
95	1,96	2	303	253,2
95	1,96	3	303	210,1
99	2,58	1	303	294,6
99	2,58	2	303	272,1
99	2,58	3	303	241,4

Zdroj: vlastné spracovanie

Pre dosiahnutie lepšej vypovedacej hodnoty vzorky pre daný prieskum je potrebné danú vzorku účinne rozdeliť podľa členenia základného súboru. V našom prípade uplatníme členenie podľa počtu zamestnancov.

Na základe informácií zo štatistického úradu bolo zistené, aké je delenie podnikov podľa počtu zamestnancov v danej oblasti. Následne sme si vypočítali podiely podnikov

v jednotlivých skupinách a podľa vypočítaných podielov sme rozdelili vzorku do týchto skupín (viď. Tab. 2.).

Tabuľka 2. Výpočet veľkosti vzorky v jednotlivých podskupinách

	Spolu	Skupina			
		0 - 9	10 - 19	20 - 49	50 - 249
Počet podnikov	303	261	17	14	11
Podiel	100%	86%	6%	5%	4%
Veľkosť vzorky	210	181	12	10	8

Zdroj: vlastné spracovanie

Po prepočte jednotlivých skupín si môžeme overiť chybu výberu prostredníctvom vzťahu (3) používanom pri vrstevnom výbere. Ten môže určiť odlišnú chybu výberu, keďže zahŕňa odhadovanú pravdepodobnosť želaného javu pre jednotlivé podskupiny. Ak berieme želaný stav v jednotlivých skupinách, je zo skúseností zrejmé, že využívanie poštových služieb (teda želaného javu) v podnikoch s 0-9 zamestnancami, bude určite nižšie ako v podnikoch s 50-249 zamestnancami. Preto chybu zo vzťahu (3) prepočítame aj so zahrnutím týchto predpokladov.

Predpokladáme, že v podnikoch s 0-9 zamestnancami bude využívanie služieb len 65 %, v podnikoch s 10-19 zamestnancami 70 %, s 20-49 zamestnancami 80 % a v podnikoch s 50-249 zamestnancami 80 %. Pomocou vzťahu (3) si teda overíme, či rozdelenie vzorky do podsúborov s rôznym p , zachovalo maximálnu chybu výberu.

$$\begin{aligned}
 E &= \frac{z_{1-\alpha/2}}{N} * \sqrt{\sum_{h=1}^4 \left[N_h^2 * \left(1 - \frac{n_h}{N_h}\right) * \left(\frac{p_h * (1 - p_h)}{n_h - 1}\right) \right]} = \\
 &= \frac{1,96}{303} * \sqrt{\sum_{h=1}^4 \left[261^2 * \left(1 - \frac{181}{261}\right) * \left(\frac{65 * (100 - 65)}{181 - 1}\right) \right] + \dots} \\
 &+ \left[17^2 * \left(1 - \frac{12}{17}\right) * \left(\frac{70 * (100 - 70)}{12 - 1}\right) \right] + \dots + \left[11^2 * \left(1 - \frac{8}{11}\right) * \left(\frac{80 * (100 - 80)}{8 - 1}\right) \right] = \\
 &= \frac{1,96}{303} * \sqrt{\sum_{h=1}^4 [264419,3 + 17276,31 + 11059,84 + 8971]} = 3,55\%
 \end{aligned}$$

Z výpočtu vidíme, že po zohľadnení meniacej sa pravdepodobnosti želaného javu v jednotlivých podsúboroch sa celková výberová chyba zväčšila na 3,55 %. Preto je nutné pôvodnú vzorku prepočítať s novou chybou výberu, ktorá bude po odrátaní vzniknutej odchýlky (3 % – 0,55 %) rovná 2,45 %.

$$n = \frac{z_{1-\alpha/2}^2 * \sigma^2 * \left[\frac{N}{(N-1)} \right]}{E^2 + \left[\frac{z_{1-\alpha/2}^2 * \sigma^2}{(N-1)} \right]} = \frac{1,96^2 * 80 * (100 - 80) * \left[\frac{303}{(303-1)} \right]}{2,45^2 + \left[\frac{1,96^2 * 80 * (100 - 80)}{(303-1)} \right]} =$$

$$= \frac{6\,166,9129}{6 + 20,3529} = 234$$

Ak teda chceme uplatniť presnejšie opytovanie cez podsúbory pri zachovaní chyby 3 % (po opätovnom prepočítaní vzťahom (3) približne 2,89 %), pôvodnú vzorku musíme navýšiť o 24 respondentov na 234.

Záver

Zo vzorového príkladu je zrejmé, že pri výpočte vzorky pri prieskume musíme vždy brať na vedomie rozdelenie základného súboru a nepodceňovať variabilitu názorov medzi viacerými skupinami respondentov. Samozrejme nie vždy je jednoduché odhadnúť aké odpovede môžeme v skupine očakávať, no dobrý marketingový pracovník, vie svojím odhadom (ak už nie priamym prepočtom) obmedziť chybu opytovania a tým ušetriť možné straty pri navýšení vzorky, prípadne z nedostatku respondentov.

Literatúra

- [1] KOTLER, P. a kol. *Principles of Marketing*. New Jersey, USA: Prentice Hall, 1998. ISBN 0-13-262254-8.
- [2] *Statistics tutorial*. [online]. [s.a.]. [cit. 2011-01-17]. Dostupné na internete: <<http://stattrek.com/Lesson6/SampleSize.aspx?Tutorial=Stat>>.
- [3] *Calculating the sample size*. [online]. [s.a.]. [cit. 2011-03-27]. Dostupné na internete: <http://www.ifad.org/gender/tools/hfs/anthropometry/ant_3.htm>.
- [4] ŠTOFKOVÁ, J. a kol., Manažment podniku, EDIS, Žilinská univerzita v Žiline, 2007

Grantová podpora

Príspevok je publikovaný v rámci riešenia projektu IV 9/KS/2012 znalostný manažment ako súčasť podnikovej stratégie, Žilinská univerzita v Žiline.