9. PRIMENA METODA UZORKA U ANKETAMA

9.1. PREDNOSTI PRIMENE METODA UZORKA

U savremenom društvu stalno raste potreba za obaveštenjima iz raznih oblasti života. Zbog toga se sve više pribegava anketama na principima metoda uzorka. Primena anketa ima više prednosti u odnosu na popis. Tako dok se u popis uključuju sve jedinice ispitivanja, anketa obuhvata samo jedan njihov deo. Na taj se način postiže velika ušteda u troškovima. Pored toga, anketom se znatno ranije dolazi do rezultata jer je obrada podataka brža. I najzad, anketa angažuje relativno manji broj lica za prikupljanje podataka i na taj je način moguće veću pažnju posvetiti njihovom izboru. Veća stručnost anketara je i veća garancija uspeha ankete.

Nedostatak ankete u odnosu na popis je da ona ne obuhvata sve jedinice skupa pa se tako smanjuje vrednost njenih rezultata. Istina, adekvatnom primenom metoda uzorka može se utvrditi preciznost ocenjenih rezultata i u znatnoj meri sagledati njihova upotrebljivost u praktične ciljeve.

Varijacije kod uzorka, rezultat različitosti njegovih jedinica, se izražavaju putem standardne devijacije, odnosno standardne greške. Standardna devijacija uzorka zavisi samo od varijabiliteta jedinica, dok standardna greška zavisi i od veličine uzorka. Kako preciznost ocene zavisi od veličine standardne greške, razumljivo je nastojanje da se ona smanji. S obzirom na to da se ne može uticati na varijabilitet jedinica, standardna greška može da se smanji povećanjem broja jedinica uzorka o čemu je već ranije bilo reči. Na smanjenje standardne greške može da se utiče i izborom podesnog *plana uzorka*.

Prost slučajan uzorak, gde svaka jedinica skupa ima jednaku mogućnost da u njega bude uključena, obično ne osigurava najbolju preciznost ocene. Drugim planovima uzorka se može u znatnijoj meri uticati na smanjenje standardne greške uz iste troškove. Normalno, što je standardna greška manja, preciznost je veća. Teorija metoda uzorka je od velike pomoći pri utvrđivanju preciznosti ocene na osnovu odgovarajućeg plana. Međutim, ona ne može otkloniti pristranost pod kojom se podrazumeva sistematska greška. Ova greška nastaje u slučajevima kada je na sistematski način podcenjena ili precenjena vrednost jedinica u uzorku.

U primeni metoda uzorka normalno se izračunavaju i utvrđuju preciznosti za četiri sledeće karakteristike: *total skupa, aritmetička sredina, proporcija i odnos.*

9.2. METODI IZBORA UZORKA

Izbor uzorka može da se izvede na više načina. U društveno-ekonomskim ispitivanjima osnovni preduslov je da se raspolaže listom jedinica skupa ili mapom pomoću koje izabrane jedinice mogu da se identifikuju.

Izbor prostog slučajnog uzorka može da se izvede na više načina. Jedan, dosta jednostavan, je pomoću tablice slučajnih brojeva (tabela IX u prilogu). Pre nego što se pristupi izboru, jedinice skupa treba da su numerisane od 1 do N. Neka se skup sastoji od 920 radnika jedne fabrike, i u uzorak treba izabrati 20 radnika. Pošto je skup trocifreni broj, u tablicama slučajnih brojeva uzimaju se trocifreni brojevi, nakon što je utvrđeno odakle će se početi i u kom pravcu. Uzmimo da se počinje od kolone 17—20 i reda 11 za prvu hiljadu. Kao trocifreni broj tu se nalazi 260. Ako se ide po koloni, drugi broj je 030, treći 377 itd. U uzorak se uzimaju jedinice skupa sa tim rednim brojevima. U ovom slučaju preskače se trocifreni broj iznad 920 kada se na njega naiđe. Kod *uzorka bez ponavljanja* preskače se broj koji se ponavlja. Ako se ostavi mogućnost da se ista jedinica bira u uzorak dva i više puta, onda je to *uzorak sa ponavljanjem*. Kod takvog uzorka ponovljena jedinica se javlja onoliko puta koliko se puta javila u tablici slučajnih brojeva prilikom izbora.

Čestu primenu ima *sistematski uzorak*. Za ilustraciju njegovog izbora uzmimo da se skup sastoji od 1.000 domaćinstava i da u uzorak treba da se izabere 20. Polazi se od *koraka odabiranja* koji se dobija podelom 1.000 sa 20 što iznosi 50. Nasumce se zatim izabere jedan broj od 1 do 50, recimo 17. Domaćinstvo sa tim brojem se uzima u uzorak, a zatim domaćinstvo s brojem kada se na 17 doda 50. Broj 67 je tako druga jedinica u uzorku. Dalje se sistematski na toj osnovi uzimaju u uzorak sledeće jedinice: $67 + 50 = 117$; $117 + 50 = 167$ itd. sve dok se ne nađe i 20-a jedinica.

Teorijske osnove sistematskog uzorka su nešto drugačije nego prostog slučajnog uzorka. Da li će sistematski uzorak pružiti veću preciznost od prostog slučajnog zavisi od rasporeda jedinica u skupu. Zbog toga je česta dilema koji od ova dva uzorka da se primeni. Kada jedinice skupa nisu sistematski raspoređene, primena bilo kojeg od ova dva načina izbora neće bitnije uticati na visinu standardne greške. Međutim, kada su jedinice skupa rangirane prema veličini, primena sistematskog uzorka je bolje rešenje jer će se na taj način u njega uključiti jedinice sa različitim vrednostima. Neka su, u ekstremnom slučaju, vrednosti skupa raspoređene na sledeći način:

$$1,1,1,1,\ 2,2,2,2,\ 3,3,3,3,\ 4,4,4,4,\ 5,5,5,5.$$

Broj jedinica u skupu je $N = 20$; treba uzeti uzorak od $n = 5$ jedinica te je korak odabiranja $k = 4$. Prema tome, uzorak će uključivati po jednu jedinicu sa vrednostima od 1 do 5. Kao druga krajnost neka su jedinice skupa raspoređene na način:

$$1,2,3,4,5,\ 1,2,3,4,5,\ 1,2,3,4,5,\ 1,2,3,4,5,\ 1,2,3,4,5.$$

Za uzorak od 5 jedinica, bilo koji da je početak, uzorak će se sastojati od jedinica iste vrednosti što isključuje adekvatnu ocenu preciznosti. Ove krajnosti se ne susreću često u praksi ali raspored jedinica skupa sa sličnim tendencijama nije tako redak kao saobraćaj u toku dana, promet u trgovini u toku sedmice itd.

U praksi metoda uzorka često se jedinice razlikuju po veličini, kao gradovi, sela itd. Ako bi se primenio prost slučajan ili sistematski uzorak standardna greška ne bi pružala zadovoljavajuću preciznost. U takvim situacijama se primenjuje izbor *s verovatnoćom proporcionalnom veličini jedinice*. Za to se uzima pomoćna promenljiva koja odražava veličinu jedinice. Tako, ako su gradovi jedinice broj njihovih stanovnika služi kao pomoćna promenljiva. Gradu sa većim brojem stanovnika daje se veća verovatnoća da bude uključen u uzorak, nego onom sa manjim.

Za izbor uzorka s verovatnoćom proporcionalnom veličini jedinice potrebno je raspolagati brojem pomoćnih jedinica, za svaku edinicu, koje služe kao osnova izvlačenja. To nije jednostavno kod velikog skupa, recimo kada su glavne jedinice

domaćinstva a u uzorak se izvlače prema broju njihovih članova. Zbog toga se ovakav izbor primenjuje kada postoje velike razlike u veličini jedinica skupa.

Izbor uzorka sa verovatnoćom proporcionalnom veličini jedinice može da se izvede *sa ponavljanjem i bez ponavljanja*. Uzorak sa ponavljanjem se češće primenjuje jer se do njegove standardne greške relativno jednostavnije dolazi.

Tabela za izbor jedinica u uzorak sa verovatnoćom proporcionalnoj veličini jedinice

Tabela 9.1.

Zgrade (redni broj)	Stanovi, M_i	Kumulacija od M_i	Rang stanova od 1 do M
1	5	5	1—5
2	8	13	6—13
3	10	23	14—23
4	7	30	24—30
5	15	45	31—45
6	3	48	46—48

Do uzorka s verovatnoćom proporcionalnom veličini jedinice obično se dolazi na bazi *spiska jedinica*. Neka se skup sastoji od $N = 6$ stambenih zgrada, i u uzorak treba da se izaberu $n = 2$ zgrade. Za pomoćnu promenljivu uzet je broj stanova u svakoj zgradi (M_i). Za tu svrhu korisno je sačiniti pregled prikazan u tabeli 9.1. koja sadrži kumulaciju stanova rangiranih po jedinicama izbora. Uzimaju se dva broja na slučajan način manja od 48 koliko skup ima stanova. Neka je prvi broj 15, i s obzirom na to da se nalazi rangiran u grupu od 14 do 23, zgrada sa rednim brojem 3 je prva jedinica uzorka. Verovatnoća izvlačenja te zgrade u uzorak je 10/48. Svaki izvučeni broj od 14 do 23 uključuje u uzorak zgradu pod rednim brojem 3. Na isti način se postupa i kod izbora druge zgrade u uzorak.

9.3. STRATIFIKOVANI SLUČAJNI UZORAK

U praksi se često skup deli na grupe ili stratume iz kojih se biraju jedinice na slučajan način. Motivi za stratifikaciju mogu da budu dvojaki. Prvi je želja da se dobiju posebna obaveštenja za svaki stratum pored obaveštenja za ceo skup, a drugi je da se poveća preciznost ocene. Stratifikacija može da se izvede po različitim osnovama, kao što su geografska područja, zanimanje stanovnika, starost, prihodi itd. Ako je osnovni cilj stratifikacije da se poveća preciznost ocene smanjenjem standardne greške, podelu treba izvoditi tako da jedinice u stratumima budu što *homogenije*.

Prema tome, u primeni stratifikacije skup je podeljen na stratume i ukupan broj jedinica je $N_1 + N_2 + \ldots + N_L = N$ $(h = 1, 2, \ldots, L)$. Uzorak se sastoji od $n_1 + n_2 + \ldots + n_L = n$ jedinica. Aritmetička sredina uzorka u h-tom stratumu je

$$\bar{X}_h = \frac{1}{n_h} \sum_i X_{hi}, \quad (i = 1, 2, \ldots, n_h), \tag{9.1}$$

dok je njena ocena za ceo skup

$$\bar{X}_{st} = \frac{1}{N} \sum_h N_h \bar{X}_h. \quad (h = 1, 2, \ldots, L). \tag{9.2}$$

Ocena totala skupa je

$$\hat{X}_{st} = \sum_h N_h \bar{X}_h = N\bar{X}_{st}. \tag{9.3}$$

Varijansa greške aritmetičke sredine stratifikovanog uzorka je

$$s^2_{\bar{X}.\,st} = \frac{1}{N^2} \sum_h N_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}, \tag{9.4}$$

gde je

$$s_h^2 = \frac{1}{(n_h - 1)} \sum_i (X_{hi} - \bar{X}_h)^2. \tag{9.5}$$

Standardna greška aritmetičke sredine je

$$s_{\bar{X}.\,st} \sqrt{s^2_{\bar{X}.\,st}}. \tag{9.6}$$

Za ilustraciju uzmimo da se osnovni skup sastoji iz dva stratuma od po 5 jedinica sa sledećim vrednostima:

Stratum 1 : 5, 9, 6, 4, 7

Stratum 2 : 15, 12, 10, 7, 13

$N_1 + N_2 = 5 + 5 = 10.$

Uzmimo u uzorak po dve jedinice iz svakog stratuma: iz prvog stratuma jedinice 5 i 7, a iz drugog 12 i 10. Ukupna veličina uzorka je $n_1 + n_2 = 2 + 2 = 4$. Ocenjene aritmetičke sredine su

$$\bar{X}_1 = \frac{5+7}{2} = 6; \quad \bar{X}_2 = \frac{13+7}{2} = 10; \quad \bar{X}_{st} = \frac{6+10}{2} = 8.$$

Za standardnu grešku izračunava se

$$s_1^2 = \frac{1}{2-1} [(5-6)^2 + (7-6)^2] = 2$$

$$s_2^2 = \frac{1}{2-1} [(13-10)^2 + (7-10)^2] = 18$$

$$s^2_{\bar{X}.\,st} = \frac{1}{10^2} \left[(5^2) \left(\frac{5-2}{5} \right) \frac{2}{2} + (5^2) \left(\frac{5-2}{5} \right) \frac{18}{2} \right] = 1{,}5$$

$$s_{\bar{X}.\,st} = \sqrt{1{,}5} = 1{,}22.$$

Interval poverenja za parametar skupa utvrđuje se na način objašnjen u odeljku 7.3.

Jedan od problema stratifikovanog uzorka je kako utvrditi njegovu veličinu da se postigne maksimum preciznosti uz što manje troškova. Kada su troškovi po jedinici uzorka isti za sve jedinice, logično bi bilo da se relativno veći broj jedinica uzme iz onog stratuma kod koga je varijabilnost jedinica veća. Ako je, pak, u svim stratumima ista varijabilnost jedinica, a troškovi po jedinici različiti, tada je normalno uzeti veći broj jedinica iz onog stratuma gde su ti troškovi manji.

Stratifikovani uzorak može da se izabere na više načina. Najjednostavnije je da se iz svakog stratuma uzme *isti* broj jedinica koji iznosi $\bar{n} = n/L$, gde je L broj stratuma. Ovakvim izborom se ne uzima u obzir broj jedinica po stratumima što bi bilo normalnije pogotovu kada stratumi nisu jednaki. To se postiže *proporcionalnim rasporedom,* gde broj jedinica u uzorku iz svakog stratuma zavisi od njegove veličine. Na primer, uzorak treba da obuhvati 10% od skupa. Iz svakog stratuma se uzima 10% njegovih jedinica tako da je

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \ldots = \frac{n_L}{N_L} = \frac{n}{N} = 10\%.$$

Još veća preciznost u oceni postiže se primenom *Nejmanovog rasporeda* gde se veličina uzorka po stratumima utvrđuje zavisno od varijabilnosti njihovih jedinica. To zahteva poznavanje standardnih devijacija skupova za svaki stratum. Kako, normalno, one nisu poznate, uzimaju se njihove ocene na osnovu nekog ranijeg ispitivanja, ili u krajnjoj liniji procene, jer velika preciznost nije ni nužna.

Optimalni raspored treba da omogući dobijanje najmanje standardne greške za utvrđeni trošak po jedinici uzorka. On je specifičan slučaj Nejmanovog rasporeda, jer, ako su troškovi po jedinici istovetni, nema razlike između ova dva rasporeda. Inače funkcija troškova je

$$C = c_0 + \sum_h c_h n_h, \tag{9.7}$$

gde su: C = ukupni troškovi, c_0 = opšti troškovi ankete, c_h = troškovi po jedinici u h-tom stratumu. U utvrđivanju optimalnog rasporeda uzima se u obzir samo promenljivi deo troškova.

Kod ocene *proporcije,* njena efikasnost dolazi do izražaja u stratifikovanom uzorku tek kada postoji značajna razlika između proporcija po stratumima.

Ova objašnjenja o izboru stratifikovanog uzorka imala su u vidu samo jednu karakteristiku skupa. Međutim, u primeni metoda uzorka normalno se ispituje više karakteristika. Postavlja se pitanje kako da se postupi s obzirom na to da je moguće da se na osnovu jedne karakteristike dobije jedan raspored jedinica po stratumima, pa čak i ukupna veličina uzorka, a na osnovu drugih drugačiji. S obzirom na značaj ovog problema bilo je više pokušaja da se on podrobnije sagleda. Neko jedinstveno rešenje se nije pronašlo. Normalno je tražiti oslonac na osnovu jedne ili još neke relevantne karakteristike.

Jedan od problema je *utvrđivanje broja stratuma.* Praktično tu nema ograničenja. U krajnjem slučaju taj broj može da se izjednači s brojem jedinica u uzorku. Ne sme se izgubiti iz vida da su za izračunavanje standardne greške potrebne bar dve jedinice stratuma. Veći broj stratuma, po pravilu, doprinosi smanjenju standardne greške, ali i tu ima granice sa gledišta racionalnosti. Naime, smanjenje standardne greške nije ravnomerno tako da posle izvesnog broja stratuma njihovo dalje povećanje neznatno utiče na grešku. Osnovno je pravilo da, gde god je moguće, jedinice unutar stratuma budu homogene, tj. što izjednačenije po varijabilnosti pojedinih obeležja.

Ovde spada i problem *utvrđivanja granica između stratuma.* Iako su činjeni napori da se nađu najbolja rešenja, nije se došlo do jednog opštevažećeg pravila. Rukovodeći se principom homogenosti jedinica, važnu ulogu tu igra iskustvo. Razumljivo je da se taj problem ne postavlja tamo gde stratumi predstavljaju prirodne ili druge celine.