GRAHAM KALTON
VASJA VEHOVAR

VZORČENJE V ANKETAH

Ljubljana 2001

Naslov izvirnika (prvega dela slovenske izdaje)
INTRODUCTION TO SURVEY SAMPLING

The International professional Publishers
Newbury Park, London, New Delhi

Graham KALTON, Vasja VEHOVAR
VZORČENJE V ANKETAH

Knjižna zbirka PROFESIJA
Izdajatelj FAKULTETA ZA DRUŽBENE VEDE, UNIVERZA V LJUBLJANI
Za založbo Ivan HVALA

Oblikovanje naslovnice Jaka KRAMBERGER
Slika na naslovnici iz programa CI (avtor Andrej BLEJEC)
Prelom B&V Co.
Tisk Tiskarna LITTERA PICTA

Izdajo knjige je omogočilo Ministrstvo za šolstvo, znanost in šport Republike Slovenije

CIP - Kat ložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

316.334.3(497.4:4-014)(082)

KALTON, Graham
Vzorčenje v anketah / Graham Kalton, Vasja Vehovar. – Ljubljana : Fakulteta za družbene vede, 2001. – (Knjižna zbirka Profesija)

ISBN 961-235-050-7
1. Vehovar, Vasja
110486784

VSEBINA

PREDGOVOR

Vzorčenje je obsežno statistično področje, za katero obstaja vrsta odličnih učbenikov v angleškem jeziku. Pri pripravi publikacije, ki v problematiko uvaja slovenske bralce, je zato smiselno, da prevedemo in razširimo obstoječo publikacijo. Knjiga *Introduction to Survey Sampling* profesorja Grahama Kaltona je za to nadvse primerna, saj predstavlja prodoren in razumljiv uvod v razmeroma zahtevno problematiko. O kvaliteti in popularnosti navedene knjige pričajo tudi številni ponatisi založbe *Sage*, ki knjigo izdaja v okviru serije *Quantitative Applications in the Social Sciences*.

Pričujoča publikacija je torej delo dveh avtorjev:

- Poglavja (1–10) so prevod oziroma priredba Kaltonove monografije. Pri tem je slovenski avtor – poleg zadnjih treh odstavkov v podpoglavju *Obravnava neodgovorov* (9.5) – dodal še podpoglavji *Administrativni vzorčni okviri* (8.5) ter *Uteži in vzorčna varianca* (10.3).
- Poglavja (11–14) je napisal slovenski avtor in s tem bistveno razširil obravnavo praktičnih vprašanj. Iz Kaltonove publikacije so v tem delu vključena le podpoglavja: *Velikost vzorca in standardna napaka* (11.1), *Prostorski vzorci v ZDA* (13.1) in *Telefonski vzorci v ZDA* (14.2).
- Predzadnje poglavje o neverjetnostnih vzorcih (15) je ponovno prevod angleškega originala. Dodano pa je še podpoglavje *Kvotno vzorčenje v praksi* (15.4), ki je priredba zapisa profesorja Leslie Kisha iz leta 1998. Tudi zaključno poglavje (16) izhaja iz angleške publikacije; slovenski avtor je dodal le zadnji odstavek o novejših bibliografskih enotah.

Za nastanek pričujočega dela gre zahvala založbama *Sage* in Z*aložbi FDV* kot tudi *Ministrstvu za znanost, šolstvo in šport* ter *Statističnemu uradu Republike Slovenije,* ki sta podprla izid publikacije.

Posebna zahvala gre tudi profesorju Grahamu Kaltonu za sodelovanje, profesorju Robertu Grovesu za idejo za nastanek take publikacije ter profesorju Leslie Kishu za dovoljenje objave zapisa o kvotnih vzorcih.

Dr. Vasja Vehovar

1. UVOD

Raziskovanje, ki temelji na vzorcu populacije, je danes splošno sprejet pristop pri zbiranju statističnih podatkov. Vzorčenje uporabljamo na najrazličnejših področjih v raziskovalne, poslovne, upravne in administrativne namene. Tako so na podlagi proučevanja vzorcev ciljne populacije na mnogih znanstvenih področjih razvili, preverili ali redefinirali številne raziskovalne hipoteze, posebej v sociologiji, demografiji, političnih vedah, ekonomiji, izobraževanju, socialni psihologiji in zdravstvu. Vzorčne raziskave rutinsko uporabljajo tudi vladne in druge javne institucije za ugotavljanje razmer na področju brezposelnosti, dohodkov, življenjskih stroškov, stanovanjskih razmer, izobrazbe, prehrane, zdravja, potovanj ipd., vladne službe pa redno naročajo anketne raziskave podjetij in drugih organizacij s področja industrije, trgovine, kmetijstva in šolstva. Najobsežnejšo uporabo vzorčnih anket najdemo v marketinškem raziskovanju, kjer številne agencije raziskujejo tržišče, potrošnikovo obnašanje, storitve in medije. Izredno odmevne so tudi javnomnenjske ankete, ki spremljajo odnos do javnih oseb, političnih strank in siceršnje razpoloženje javnosti glede aktualnih vprašanj.

Glede na široko in intenzivno uporabo anketnih raziskav (*angl. surveys*), ki temeljijo na vzorcih (*angl. samples*), je pravzaprav presenetljivo, da ima vzorčno raziskovanje razmeroma kratko zgodovino, omejeno na 20. stoletje. Resnejše raziskave, ki temeljijo na vzorcih, so se namreč pojavile šele v tridesetih letih tega stoletja, saj so statistiki še na začetku stoletja razglabljali, ali je kakovostno raziskavo sploh mogoče izvesti samo na delu populacije (O'Muricheartaigh in Wong, 1981). Šele v letih pred 2. svetovno vojno opazimo pomembnejše premike v statistični teoriji in tudi v raziskovalni praksi, s čimer so postala načela verjetnostnega vzorčenja splošno sprejet način zbiranja anketnih podatkov. V razmeroma kratkem času so se razvile številne metode vzorčenja, ki se nenehno izpopolnjujejo še danes in omogočajo vse učinkovitejšo uporabo vzorčnih tehnik.

Načrtovanje anketnih raziskav (*angl. survey design*) je nadvse kompleksen proces in vključuje številne dejavnike, kot so izbira načina anketiranja (osebno, telefonsko, samoanketiranje), zasnovo vprašalnika, oblikovanje vprašanj, metode za obdelavo podatkov in izdelavo vzorčnega načrta (Moser in Kalton, 1971; Warwick in Lininger, 1975). V pričujočem delu se osredotočamo le na izdelavo in analizo vzorca, ki je specifičen vidik anketnega raziskovanja in ena ključnih komponent za uspešno opravljeno anketno raziskavo.

Populacija

Prvi korak pri načrtovanju anketne raziskave je opredelitev populacije, ki jo nameravamo proučevati. Na tem mestu bomo izraz »populacija« (*angl. population*) uporabljali v smislu množice vseh elementov, na katere se nanašajo naše ugotovitve. Za to množico bomo na podlagi anketne raziskave opravili določeno statistično sklepanje *(angl. statistical inference)*. Populacija je sestavljena iz posameznih »elementov«, ki so osnovna enota naše analize. Elementi so lahko osebe, lahko pa tudi gospodinjstva, kmetije, šole, podjetja ipd. Opredelitev populacije je nadvse pomembna, saj bodo razlage rezultatov močno odvisne od definicij, ki smo jih uporabili pri samem načrtovanju vzorca.

Oglejmo si primer. Denimo, da v manjšem mestu izvajamo anketo, s katero ugotavljamo podporo prebivalcev za nov sistem mestnega avtobusnega prometa. Pri opredelitvi populacije se pojavlja vrsta vprašanj. Ali naj anketo izvedemo le na osebah, ki živijo v mejah mesta? Kakšna naj bo najnižja starost osebe, ki naj odgovarja na anketo? Ali naj v anketo vključimo prebivalce, ki nimajo volilne pravice? Ali naj iz raziskave izključimo vse, ki v tem mestu živijo le začasno? Pri opredelitvi populacije se torej pojavlja množica vprašanj, zato je lahko njeno razmejevanje izredno težavno.

Priporočljivo je, da proučevano populacijo na samem začetku opredelimo v idealni obliki kot želeno ciljno populacijo (*angl. target population*). V praksi namreč zaradi določenih operativnih razlogov namesto ciljne populacije pogosto proučujemo le tako imenovano anketirano populacijo (*angl. survey population*). V številnih anketah v ZDA so npr. izključeni ljudje, ki v času anketiranja živijo zunaj osrednjega dela ZDA, na Havajih ali Aljaski, kot tudi institucionalna populacija – osebe v vojski, zaporih, hotelih ali drugih ustanovah. Podobno tudi v Sloveniji npr. pri volilnih anketah ne vključujemo volivcev, ki živijo zunaj Slovenije, niti volivcev, ki stanujejo v ustanovah (domovi za ostarele, bolnice ipd.). Seveda pa med anketiranjem običajno nastopijo še dodatne težave pri določanju vzorčnega okvira in zbiranju odgovorov, saj nekatere izbrane osebe ne želijo ali pa ne morejo odgovarjati na anketo. Anketirana populacija se torej pogosto razlikuje od ciljne populacije, zato je natančnost pri njeni opredelitvi – kot tudi pri spremljanju morebitnih odstopanj – izredno pomembna.

Samo če začnemo opredeljevanje z idealno proučevano populacijo in nato postopno izvedemo odstopanja, se s tem tudi jasno zavedamo, kakšne posledice imajo odstopanja za interpretacijo rezultatov.

Vzorec

Ko je populacija opredeljena, se lahko začnemo ukvarjati z vzorcem. Ena od možnosti je, da vključimo v raziskavo vse elemente v populaciji in jih enostavno popišemo *(angl. Census)*, vendar je to zaradi velikosti populacije običajno neprimerno. Zbiranje podatkov iz dela populacije je namreč bistveno cenejše, hkrati pa lahko zagotovi povsem zadostno natančnost ocen. Prav tako lahko pri uporabi vzorčnih raziskav pridobimo na času – od zasnove ankete do končnih rezultatov – ki je pri proučevanju cele populacije seveda bistveno daljši. Z izvedbo raziskave na vzorcu pa lahko zaradi večje kakovosti anketiranja pridobimo celo boljše podatke v primerjavi s popisom celotne populacije. Zaradi navedenih razlogov se pri anketah večjih populacij skoraj vedno odločamo za uporabo vzorca.

Vzorec je torej del ciljne populacije, na podlagi katerega izvedemo sklepanje o celotni populaciji. Celoten proces zasnove in izdelave vzorca se tako nanaša na vprašanje, kako izbrati del populacije, ki bo vključen v anketno raziskavo. Pri tem je temeljno načelo, za katero se moramo odločiti že na začetku vzorčenja, opredelitev za slučajno izbiranje elementov v vzorec. Posebej pomembni so verjetnostni vzorci *(angl. probability samples)*, kjer ima vsak element v populaciji vnaprej znano in neničelno verjetnost, da se pojavi v vzorcu. Samo pri verjetnostnih vzorcih namreč lahko uporabimo statistično sklepanje, ki je podlaga za izračun intervalov zaupanja *(angl. confidence intervals)*.

Vzorci, ki ne zadostijo pogojem o vnaprej znani in neničelni verjetnosti izbora, se imenujejo neverjetnostni vzorci (*angl. nonprobability samples*) in tudi v okviru tovrstnih vzorcev obstaja širok spekter praktičnih postopkov vzorčenja. Večina teh postopkov temelji na anketarjevem subjektivnem izbiranju elementov, čeprav lahko vključujejo tudi določene sestavine slučajnosti. Seveda pa subjektivnost izbora elementov pri neverjetnostnih vzorcih načeloma onemogoča izvedbo statističnega sklepanja iz vzorca na populacijo. Prvi predpogoj za korektno statistično sklepanje je namreč verjetnostna izbira elementov, to je izbira, za katero je med drugim značilno tudi to, da je neodvisna od anketarja. Verjetnostne vzorce zaradi objektivnosti izbire in kvantificiranja potencialnega tveganja za napako včasih imenujemo tudi znanstveni vzorci (*angl. scientific samples*).

Neverjetnostno vzorčenje torej ne omogoča statističnega sklepanja niti oblikovanja strokovnih trditev oziroma statističnih interpretacij, pri katerih je tveganje za napako natančno opredeljeno. Zaradi navedenih razlogov se bomo v nadaljevanju omejili predvsem na verjetnostne vzorce.

Prvi pogoj za izvedbo verjetnostnega vzorčenja je obstoj vzorčnega okvira (*angl. sampling frame*), iz katerega izberemo elemente v vzorec. V preprostih primerih, ko imamo na voljo celoten seznam elementov v populaciji, lahko za vzorčni okvir uporabimo kar tak seznam. Če seznama nimamo, lahko tak okvir predstavlja tudi poseben nadomestni postopek, ki zagotavlja uvrstitev elementov v populaciji. Pri gospodinjstvih, stanovanjih ali poslovnih prostorih se tak postopek imenuje prostorsko vzorčenje (*angl. area sampling*), ki poskrbi za vzpostavitev posebnega nadomestka za vzorčni okvir. Pri tem vse elemente v populaciji najprej povežemo v prostorske enote (npr. naselja), ki jih nato izbiramo v vzorec. Šele v drugem koraku izbiramo posamezne ali kar vse elemente v izbranih prostorskih enotah oziroma območjih. Problematiko vzorčnih okvirov in prostorskih vzorcev bomo v nadaljevanju podrobneje obravnavali v posebnih poglavjih.

Za učinkovito načrtovanje vzorcev se v praksi uporabljajo številne tehnike verjetnostnega vzorčenja. Med najbolj razširjenimi so:
- sistematično vzorčenje,
- stratifikacija,
- večstopenjsko vzorčenje in
- vzorčenje z verjetnostjo, ki je sorazmerna velikosti enot.

Navedenim tehnikam se bomo posvetili v posebnih poglavjih, kar velja tudi za zahtevnejše pristope v primeru osebnega anketiranja geografsko razpršenih populacij. Pri tem z osebnim anketiranjem razumemo terensko osebno anketiranje *(angl. face-to-face interviews).* Uvodoma pa začenjamo obravnavo s preprostimi oblikami vzorčenja, ki so običajno primerne za majhne vzorce in strnjene populacije.

2. ENOSTAVNO SLUČAJNO VZORČENJE

Enostavno slučajno vzorčenje (*angl. simple random sampling* – SRS) je naravno izhodišče za obravnavo verjetnostnih vzorcev, ne zato, ker bi bilo tako razširjeno, ampak ker je najpreprostejša metoda, ki vsebuje vsa načela kompleksnejših metod. Pri tem kot kompleksne vzorčne načrte *(angl. complex sample design)* obravnavamo vse vzorce, ki niso SRS vzorci.

Oznake

Osnovne oznake, ki jih uvajamo v SRS vzorcih, so nadvse pomembne, saj jih bomo uporabljali tudi pri bolj zapletenih vzorčnih načrtih. Tako bomo z n vedno označevali velikost vzorca, z N pa velikost populacije.

Pri SRS vzorčenju ima vsaka podmnožica n različnih elementov iz populacije z N elementi enako verjetnost za izbor. Hkrati s tem pa ima tudi vsak element v populaciji enako verjetnost za izbor v vzorec, kar bomo označevali s kratico EPSEM (*angl. Equal Probability Selection Method*). Kot bomo videli, imajo poleg SRS vzorcev lastnost EPSEM tudi nekateri kompleksni vzorčni načrti. Posebnost SRS vzorcev je torej v tem, da imajo – poleg siceršnje lastnosti EPSEM – vsi SRS vzorci določene velikosti tudi enako verjetnost, da jih izberemo. Slednje, kot rečeno, ne velja za vse EPSEM vzorčne načrte.

Tipičen primer EPSEM vzorca, ki ni SRS, je npr. vzorec za anketo *General Social Survey* v ZDA, v Sloveniji pa vzorec za anketo *Slovensko javno mnenje*, kjer imajo vsi odrasli državljani enako verjetnost za vključitev, vendar je zaradi gostitve vzorca v nekaj več kot 100 izhodiščnih točkah (krajih) nemogoče, da bi vzorec vključeval osebe iz samo enega kraja. V primeru SRS vzorca pa načeloma obstaja verjetnost – čeprav izredno majhna – da slučajno izberemo npr. vse elemente iz enega samega kraja.

SRS vzorčenje si bomo ogledali na nekaj posebnih primerih. Predpostavimo, da izvajamo anketo o dejavnostih dijakov v prostem času. Na voljo je seznam 1,872 dijakov, urejen po identifikacijskih številkah. Številke segajo od 0001 do 1917, z nekaj neustreznimi številkami, ker se je manjše število dijakov iz šole izpisalo. Vzemimo, da za raziskavo potrebujemo SRS vzorec velikosti 250 elementov. (O določanju velikosti vzorca bomo podrobneje govorili v posebnem poglavju).

Eden od načinov za izbiro vzorca je lahko metoda loterije. To pomeni, da številke dijakov napišemo na 1,872 enakih kroglic. Vse kroglice damo v posodo in nato iz nje slučajno izberemo 250 kroglic. Če bi bil izbor izveden brez napake, bi izbrane kroglice določale ravno SRS vzorec 250 dijakov. Čeprav je to načeloma videti preprosto, pa je tak postopek nadvse neroden, zato ga v praksi skoraj ne uporabljamo.

Pogosteje uporabljen način izbire SRS vzorca je tabela slučajnih števil, ki je sestavljena tako, da imajo vsa enomestna, dvomestna itd. števila enako pogostost pojavljanj. Primer slučajnih števil je naveden v tabeli 1, ki sta jo izdelala Kendall in Smith (1939).

Tabela 1: Tabela slučajnih števil

67 28	96 25	68 36	24 72	03 85	49 24
85 86	94 78	32 59	51 82	86 43	73 84
40 10	60 09	05 88	78 44	63 13	58 25
94 55	89 48	90 80	77 80	26 89	87 44
11 63	77 77	23 20	33 62	62 19	29 03

Ker je v našem primeru identifikacijska številka dijakov štirimestna, moramo izbrati števila v nizu po štiri. V praksi se izbira začne slučajno, tu pa bomo zaradi enostavnosti začeli v zgornjem levem kotu, nadaljevali pa navzdol.

Števila, ki presegajo razpon identifikacijskih številk dijakov, običajno izpustimo. V našem primeru prve štiri številke (6728, 8586, 4010, 9455) ne določajo nobenega dijaka. Šele peta številka (1163) določa veljavno identifikacijsko številko.

Ker bi tako morali pregledati veliko slučajnih števil, je priporočljivo, da posamezni identifikacijski številki dijaka pripišemo tudi druga slučajna števila. V našem primeru bomo zato vsaki številki dijaka prišteli po 2000, vse dokler ne presežemo števila 9999. Ker bodo vsa ta števila določala istega dijaka, bomo izbor opravili bistveno hitreje. Prvi izbrani element, torej dijak z identifikacijsko številko 0001, bo tako določen s katerim koli slučajnim številom 2001, 4001, 6001 in 8001; dijak 0002 z 2002, 4002, 6002, 8002 in tako naprej do dijaka z identifikacijsko številko 1917, ki ga predstavljajo števila 3917, 5917, 7917 in 9917.

Na tej podlagi prve štiri številke prinesejo naslednjo izbiro: 6728 je dijak s številko 0728, 8586 je dijak z zaporedno številko 0586, 4010 je dijak s številko 0010, 9455 pa dijak s številko 1455.

Pri izbiranju zaporednih številk iz tabele slučajnih števil je lahko število seveda izbrano večkrat. To pri metodi loterije ni bilo možno. Ko kroglico enkrat izberemo, jo odstranimo iz posode in zato nima možnosti ponovnega izbora. Vzorčenje s ponavljanjem bi lahko izvedli pri loterijskem načinu tako, da bi izbrane kroglice vračali v posodo. Če vzorčimo brez ponavljanja, vzorec vsebuje n različnih elementov, s ponavljanjem pa lahko pri izbiri n slučajnih števil izberemo manj kot n različnih elementov. Vzorčenje s ponavljanjem imenujemo tudi vzorčenje brez omejitev (*angl.*

without restrictions). Enostavno slučajno vzorčenje (SRS) pa brez dodatnih oznak razumemo kot vzorčenje brez ponavljanja. V primeru vzorčenja s ponavljanjem to lastnost jasno poudarimo v smislu izrecne navedbe »SRS vzorčenje s ponavljanjem« (*angl. SRS with replacement*).

V našem primeru lahko SRS izbiro (brez ponavljanja) dosežemo z enostavnim neupoštevanjem ponovljenih elementov. Ker pa SRS vzorčenje brez ponavljanja omogoča natančnejše ocene kot vzorčenje s ponavljanjem, se bomo v nadaljevanju omejili na SRS vzorčenje brez ponavljanja.

Ko smo izbrali SRS vzorec velikosti 250 dijakov, za začetek predpostavimo, da smo zbrali tudi vse odgovore dijakov, ki smo jih vključili v vzorec (neodgovorom se bomo podrobneje posvetili v posebnem poglavju). Naslednji korak v analizi je obravnava vrednosti posameznih odgovorov. Na tej osnovi izračunamo ocene za lastnosti populacije (parametre), npr. aritmetično sredino za število ur gledanja televizije ali delež dijakov, ki berejo določeno knjigo.

Na tem mestu moramo vpeljati nove oznake:

$Y_1, Y_2, ..., Y_N$ označujejo vrednosti spremenljivke y za N elementov v populaciji,

$y_1, y_2, ..., y_n$ označujejo vrednosti spremenljivke y za n elementov v vzorcu.

V splošnem je torej vrednost spremenljivke y za i-ti element v populaciji označena z Y_i (i=1,2,...N), vrednost spremenljivke y za i-ti element v vzorcu pa z y_i (i=1,2,...n).

Populacijska aritmetična sredina je osnovni populacijski parameter in je podana z izrazom:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i .$$

Podobno je opredeljena tudi vzorčna aritmetična sredina:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

Populacijska elementarna varianca (*angl. population element variance*) spremenljivke y je v populaciji N elementov definirana kot:

$$S^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 , \qquad [1]$$

elementarna varianca v vzorcu (*angl. sample element variance*) pa je enaka:

$$s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2 . \qquad [1a]$$

Če v izrazu za S^2 uporabimo imenovalec N in ne *(N-1),* nima to zaradi velikih vrednosti števila N nobenih posledic, poenostavi pa lahko mnoge statistične izraze. V tem primeru populacijsko elementarno varianco označujemo s σ^2:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, \qquad [1b]$$

od koder sledi:

$$\sigma^2 = \frac{(N-1)S^2}{N}.$$

Oba izraza, σ^2 in S^2, se uporabljata enakovredno. Iz teoretične opredelitve variance sicer izhaja oblika σ^2, zato pa je za statistično uporabo primernejša oblika S^2. V nadaljevanju pa bomo populacijsko elementarno variacijo krajše imenovali kar »elementarna varianca«.

Vzorčna varianca

Predpostavimo torej, da v našem primeru z anketo ocenjujemo povprečno število ur gledanja televizije $\bar{Y}$ za vse dijake v šoli. Tu se pojavi osnovno vprašanje vzorčenja: kako dobra cenilka (*angl. estimator*) populacijske aritmetične sredine $\bar{Y}$ je vzorčna aritmetična sredina $\bar{y}$. Cenilko, to je funkcijo, ki izbranemu vzorcu vsakič dodeli določeno vrednost, včasih imenujemo tudi statistika (*angl. statistics*). Ker populacijske aritmetične sredine $\bar{Y}$ običajno ne poznamo, je videti, da na podlagi enega samega vzorca ni mogoče odgovoriti na vprašanje o kakovosti cenilke $\bar{y}$. Za odgovor bi namreč potrebovali ponavljajoče se izbiranje vzorcev.

Da bi ločevali rezultate na podlagi posameznega vzorca od rezultatov na podlagi ponavljajočih se vzorcev, moramo vpeljati nov pojem. Ocena (*angl. estimate*) naj zato označuje določeno vrednost (npr. vzorčno aritmetično sredino) iz posameznega vzorca, medtem ko izraz cenilka (*angl. estimator*) označuje funkcijo – to je slučajno spremenljivko (*angl. random variable*) – ki izhaja iz splošnega postopka za izračun ocene populacijskega parametra in zavzema določeno vrednost pri vsaki ponovitvi vzorca. V našem primeru lahko na podlagi vseh elementov v izbranem vzorcu izračunamo oceno:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 2.20,$$

kar pomeni, da dijaki, ki so bili izbrani v vzorec, v povprečju gledajo TV več kot dve uri dnevno. Cenilka $\bar{y}$ je torej v našem vzorcu zavzela oceno oziroma vrednost 2.20.

Statistična teorija je odkrila zakonitosti pri obnašanju cenilk, neposredno pa ne daje informacije o kvaliteti posamične ocene. Kljub temu lahko pod določenimi pogoji uporabimo statistično teorijo tudi v primeru, ko razpolagamo z oceno iz enega samega vzorca. V nadaljevanju bomo zato na kratko navedli najpomembnejše ugotovitve splošne teorije statističnega sklepanja z vidika SRS vzorčenja. Za podrobnejši vpogled v obravnavano problematiko je na voljo več odličnih besedil, npr. Blalock (1972).

Pri obravnavi vzorčnih cenilk moramo upoštevati rezultate, ki jih dobimo z neomejenim številom ponovitev postopka vzorčenja. Seveda pri tem – preden izbiramo naslednji vzorec – vsak predhodno izbran vzorec vrnemo nazaj v populacijo. V našem primeru bi to pomenilo, da neskončnokrat izberemo vzorec 250 elementov iz množice 1,872 dijakov ter vsakič izračunamo aritmetično sredino za izbrano spremenljivko na vseh elementih v vzorcu. Na ta način generiramo novo množico – množico vzorčnih ocen $\bar{y}$ – v kateri so vse možne aritmetične sredine, ki smo jih izračunali pri posameznih vzorcih. Dobljeni niz vzorčnih aritmetičnih sredin oziroma ocen je torej slučajna spremenljivka in ima posebno porazdelitev, ki jo imenujemo vzorčna porazdelitev (*angl. sampling distribution*). Kadar velikost vzorca n ni premajhna – v določenih primerih je lahko dovolj že 20 elementov – je statistična teorija ugotovila, da je porazdelitev vzorčnih aritmetičnih sredin približno normalna, in to ne glede na porazdelitev spremenljivke v osnovni populaciji. Navedena ugotovitev je izredno pomembna in izhaja iz statistične zakonitosti, ki jo imenujemo »centralni limitni izrek«, včasih pa tudi »zakon velikih števil«. Z drugimi besedami, vzorčne ocene povprečij na vseh vzorcih določene populacije se porazdeljujejo normalno. Poleg tega je aritmetična sredina teh povprečij enaka populacijski aritmetični sredini $\bar{Y}$. Pri tem je treba še enkrat ponoviti, da v primeru vzorčne porazdelitve opazujemo porazdelitev vzorčnih aritmetičnih sredin v populaciji vseh vzorcev in ne porazdelitve elementov v osnovni populaciji. Seveda pa to nikakor ne velja za cenilke vseh parametrov. Tako npr. aritmetična sredina minimumov na vzorcih nikakor ni enaka populacijskemu minimumu.

Kadar je aritmetična sredina vzorčnih ocen – v jeziku verjetnostnega računa bi temu rekli tudi matematično upanje oziroma pričakovana vrednost (*angl. expected value*) vzorčne cenilke – enaka populacijskemu parametru, ki ga ocenjujemo, potem je taka cenilka nepristranska *(angl. unbiased estimator)* za ocenjevanje populacijskega parametra. Tako je pri SRS vzorcu cenilka $\bar{y}$ nepristranska cenilka populacijskega parametra $\bar{Y}$, cenilka s^2 pa nepristranska cenilka parametra S^2.

Čeprav je središče vzorčne porazdelitve slučajne spremenljivke $\bar{y}$ enako $\bar{Y}$, pa se bodo posamezne ocene, ki jih zavzame cenilka v določenem vzorcu, seveda razlikovale od populacijske aritmetične sredine. Zato potrebujemo mero, ki bo izražala variiranje posameznih ocen okoli $\bar{Y}$, torej variiranje slučajne spremenljivke $\bar{y}$. V statistiki je splošno sprejeta mera variabilnosti slučajnih spremenljivk varianca oziroma standardni odklon (kvadratni koren variance), zato je primerno, da jo uporabimo tudi tokrat. Izračunati moramo torej običajno varianco in standardni odklon vzorčnih aritmetičnih sredin $\bar{y}$ v tej nekoliko posebni vzorčni porazdelitvi. V njej so namreč posamezni elementi $\bar{y}_k$, k=1...K, aritmetične sredine vzorcev v populaciji K vseh možnih vzorcev velikosti n iz populacije z N elementi. Varianco na osnovi vzorčne porazdelitve aritmetičnih sredin imenujemo vzorčna varianca (*angl. sampling variance*):

$$V(\bar{y}) = \frac{1}{K}\sum_{k=1}^{K}(\bar{y}_k - E(\bar{y}))^2 . \qquad [1c]$$

Da bi jasno ločevali standardni odklon vzorčne variance na osnovi vzorčnih aritmetičnih sredin v populaciji vseh vzorcev od siceršnjega elementarnega standardnega odklona (*angl. standard deviation – SD)* za spremenljivko v osnovnem vzorcu (koren elementarne variance S^2), bomo standardnemu odklonu vzorčne porazdelitve rekli standardna napaka (*angl. standard error – SE*). Imamo torej:

$\bar{y}_0$ – vzorčna aritmetična sredina SRS vzorca (indeks 0 označuje SRS izbiro),
S^2 – elementarna varianca $S^2 \cong \sigma^2$,
$SD(y)$ – elementarni standardni odklon, $SD(y) = S = \sqrt{S^2} \cong \sqrt{\sigma^2}$,
$V(\bar{y}_0)$ – varianca vzorčne aritmetične sredine oziroma vzorčna varianca,
$SE(\bar{y}_0)$ – standardni odklon vzorčne porazdelitve oziroma standardna napaka, pri tem seveda velja zveza: $SE(\bar{y}_0) = \sqrt{V(\bar{y}_0)}$.

Na osnovi statistične teorije je mogoče varianco cenilke aritmetične sredine v SRS vzorcu izračunati analitično. Vzorčna varianca je v tem primeru podana z izrazom $Var(\bar{y}_0)$ oziroma krajše $V(\bar{y}_0)$:

$$Var(\bar{y}_0) = V(\bar{y}_0) = \frac{N-n}{N-1} \times \frac{\sigma^2}{n} \qquad [1d]$$

oziroma enakovredno z izrazom:

$$V(\bar{y}_0) = \frac{N-n}{N} \times \frac{S^2}{n} = (1-\frac{n}{N}) \times \frac{S^2}{n} = (1-f) \times \frac{S^2}{n}. \qquad [2]$$

kjer je:

$$f = \frac{n}{N}$$

vzorčni delež (*angl. sampling fraction*). Takoj moramo poudariti, da se izraz (1d) nanaša samo na cenilko aritmetične sredine v SRS vzorcih. V kompleksnih vzorcih je namreč vzorčna varianca $V(\bar{y})$ cenilke $\bar{y}$ seveda drugačna. Dodati tudi velja, da ima aritmetična sredina kot cenilka razmeroma enostaven analitičen izraz za vzorčno varianco. Cenilke drugih parametrov (npr. regresijski koeficient, razmerje) imajo bistveno bolj zapleteno obliko, včasih – posebej v kompleksnih vzorcih – pa vzorčne variance praktično sploh ni mogoče analitično izraziti.

Zgornji izraz (2) kaže, da je varianca $V(\bar{y}_0)$ odvisna od treh dejavnikov. Prvi dejavnik je:

$$FPC = 1 - f = 1 - \frac{n}{N} = \frac{N-n}{N} \approx \frac{N-n}{N-1}$$

in se imenuje popravek končne populacije (*angl. finite population correction – FPC*). Razlika med *(N-n)/(N-1)* in (1-*f*) je pri velikem *N* seveda povsem zanemarljiva.

Faktor *(1-f)* kaže, da je ciljna populacija končna in ne neskončna, kot običajno predpostavlja klasična obravnava v statistični teoriji. Izraža pa tudi dejstvo, da je bilo vzorčenje izvedeno brez ponavljanja. Pri neskončni populaciji ali pri vzorčenju s ponavljanjem zato popravek *FPC* ne nastopa in takrat imamo samo:

$$V(\bar{y}_0) = \frac{S^2}{n}.$$

Faktor (1-*f*) torej kaže prednost vzorčenja brez ponavljanja pred vzorčenjem s ponavljanjem. Za vzorce velikosti 2 ali več je faktor (1-*f*) manjši od 1, torej je ocena $\bar{y}$, izračunana iz SRS vzorca, natančnejša oziroma ima manjšo varianco kot vzorčna varianca za $\bar{y}$ iz SRS vzorca s ponavljanjem enake velikosti.

Razlika med vzorčenjem s ponavljanjem ali brez njega je pri velikih populacijah povsem zanemarljiva, saj je pri izbiranju s ponavljanjem verjetnost ponovnega izbora istega elementa izredno majhna. Če je vzorčni delež *f* npr. 1/10, je faktor *FPC* enak 0.9 in njegov učinek na standardno napako je enak:

$$\sqrt{1-f} = 0.95.$$

Podobno pri *f* = 0.05 sledi:

$$\sqrt{1-f} = 0.97.$$

Očitno je pri majhnih vzorčnih deležih *f* faktor *FPC* blizu 1 in ima zanemarljiv učinek na standardno napako. Faktor *FPC* zato običajno zanemarjamo oziroma poenostavljamo na vrednost 1, če je le vzorčni delež manjši kot npr. *f* =1/20, včasih pa celo pri velikosti *f* =1/10. Očitno je torej, da lahko v raziskovalni praksi faktor *f* zanemarimo, saj je vzorec običajno bistveno manjši od populacije; izjema so le nekatere raziskave manjših populacij, npr. anketiranje oseb v podjetjih, organizacijah in društvih.

Drugi faktor v izrazu (2) za vzorčno varianco $V(\bar{y}_0)$ je velikost vzorca *n*. Očitno se velikost $V(\bar{y}_0)$ manjša z velikostjo vzorca; pri večjem vzorcu je zato vzorčna varianca manjša. Manj očitno pa je – v prevladujočem primeru, ko je *f* majhen – da ostaja pri velikih populacijah velikost *N* enako nepomembna kot pri majhnih populacijah, saj v izrazu (2) vrednost *N* sploh ne nastopa. Zato je vzorec z velikostjo *n* =2,000, ki ga izberemo iz populacije npr. milijarde ljudi ($N = 1{,}000{,}000{,}000$), enako natančen kot vzorec $n = 2{,}000$, ki ga izberemo iz populacije prebivalcev manjšega mesta velikosti $N = 40{,}000$. Pri tem seveda predpostavljamo, da sta elementarni varianci S^2 v obeh populacijah enaki. Iz tega izhaja, da se vzorčenje seveda veliko bolj izplača v velikih populacijah. Določena izjema so lahko majhne populacije, kjer prednosti vzorčenja niso pomembne, čeprav bi imel faktor *FPC* določen učinek na varianco. Če imamo npr. populacijo velikosti 200, je namreč skoraj enako ali celo enostavnejše, če anketiramo vseh $N = 200$ elementov kot pa, denimo, vzorec velikosti $n = 175$.

Tretji faktor v izrazu (2) za vzorčno varianco $V(\bar{y}_0)$ je elementarna varianca S^2 za spremenljivko y v populaciji. Pri tem lahko S^2 povsem enakovredno nadomestimo s σ^2. Očitno bo v primeru, ko vsi dijaki gledajo TV v približno enakem obsegu, vzorčna aritmetična sredina praktično enaka populacijski v vseh vzorcih k, $k = 1...K$, ki nastanejo pri izbiranju vzorcev velikost n iz populacije z N elementi. Elementarna varianca in tudi vzorčna varianca bosta v tem primeru izredno majhni. Če pa se ure gledanja izraziteje razlikujejo, bo aritmetična sredina posameznega vzorca običajno drugačna od populacijske. Tako S^2 kot σ^2 sta populacijska parametra, na katera ne moremo vplivati, in sta običajno neznani količini. Ena od prednosti izraza (2) za $V(\bar{y}_0)$, kjer uporabljamo S^2 (in ne σ^2), je zato dejstvo, da je vzorčna cenilka:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

nepristranska cenilka za parameter S^2 (ne pa tudi za σ^2). Zato lahko vrednosti za $V(\bar{y}_0)$ in $SE(\bar{y}_0)$ enostavno ocenimo z naslednjima izrazoma:

$$v(\bar{y}_0) = (1-f)\frac{s^2}{n}, \qquad [3]$$

$$se(\bar{y}_0) = \sqrt{\frac{(1-f)s^2}{n}}, \qquad [4]$$

kjer male črke pri $v(\bar{y}_0)$ in $se(\bar{y}_0)$ označujejo vzorčne cenilke za populacijske vrednosti $V(\bar{y}_0)$ in $SE(\bar{y}_0)$.

Interval zaupanja

Na osnovi ocene za standardno napako lahko izračunamo tudi interval zaupanja (*angl. confidence interval*) za populacijsko aritmetično sredino. Statistična teorija je pokazala, da je pri dovolj velikem vzorcu 95-odstotni interval zaupanja za $\bar{Y}$ enak:

$$\bar{y}_0 \pm 1.96 \times se(\bar{y}_0), \qquad [4a]$$

kjer smo vrednost $z_{\alpha/2} = 1.96$ dobili iz tabele za standardizirano normalno porazdelitev pri:

$$\alpha = (1-0.95) = 0.05.$$

Pri vsaki normalni porazdelitvi je namreč 95 % vseh elementov znotraj intervala, ki ga v obe smeri začrtuje ±1.96 standardnega odklona od aritmetične sredine. Če bi torej izbirali neskončno število vzorcev iz določene populacije in vsakič izračunali aritmetično sredino in standardno napako (4), bi zgoraj opredeljeni interval zaupanja – ki bi bil v vsakem vzorcu seveda nekoliko drugačen – vseboval populacijsko vrednost $\bar{Y}$ v 95 % vseh primerov vzorcev. V našem primeru zato s 5-odstotnim tveganjem trdimo, da interval zaupanja, ki ga vsakič oblikujemo na zgoraj opisani način, vključuje pravo populacijsko vrednost.

Meje intervala zaupanja so torej slučajna spremenljivka, zato je pomembno, da interval zaupanja, ki opredeljuje populacijski parameter, ločujemo od verjetnostnega intervala, ki določa gibanje cenilke. Tako bi lahko zapisali, da se npr. s 95 % verjetnostjo vzorčna aritmetična sredina nahaja v verjetnostnem intervalu:

$$P\left[\bar{y}_0 \in \left(\bar{Y} \pm 1.96 \times SE(\bar{y}_0)\right)\right] = 0.95\,. \qquad [4b]$$

Pri tem velja poudariti, da so – za razliko od intervala zaupanja – meje verjetnostnega intervala tokrat fiksne, saj so odvisne samo od populacijskih vrednosti $\bar{Y}$ in $SE(\bar{y}_0)$.

Tveganje 5 % ($\alpha = 0.05$) je pri statističnem sklepanju običajno in pomeni, da se pri eni od 20 tako postavljenih trditev motimo. Pri manj pomembnih ocenah uporabimo tudi tveganje $\alpha = 0.1$ in takrat imamo (namesto vrednosti $z_{\alpha/2} = 1.96$, ki določa širino intervala zaupanja pri $\alpha = 0.05$) manjšo vrednost $z_{\alpha/2} = 1.65$ in zato tudi natančnejšo oceno oziroma ožji interval zaupanja. Bolj kot smo pripravljeni tvegati in manj kot so naše trditve strokovne, bolj smo namreč lahko natančni pri naših ocenah. Po drugi strani pa včasih želimo pri najpomembnejših ocenah manjše tveganje, npr. $\alpha = 0.01$, kar znatno razširja interval zaupanja ($z_{\alpha/2} = 2.56$).

Nadaljujmo z našim primerom. Denimo, da je povprečno število ur gledanja televizije na dan za 250 dijakov v našem vzorcu enako $\bar{y}_0 = 2.192$ ure, ocena elementarne variance pa je $s^2 = 1.008$. Sledi 95-odstotni interval zaupanja:

$$2.192 \pm 1.96 \times \sqrt{\left(1 - \frac{250}{1{,}872}\right)\frac{1.008}{250}} = 2.192 \pm 0.116\,.$$

V tem primeru torej s 5-odstotnim tveganjem trdimo, da je populacijska aritmetična sredina v intervalu od 2.076 do 2.308.

Poleg aritmetičnih sredin pogosto ocenjujemo tudi odstotke oziroma deleže populacije z določeno lastnostjo, kot npr. odstotek dijakov, ki trenutno bere izbrano knjigo. Izrazi za ocenjevanje deležev izhajajo neposredno iz izrazov za aritmetične sredine. Pri tem P označuje delež elementov z določeno lastnostjo v populaciji, p pa odgovarjajoče ocene v vzorcu. Ker pa spremenljivka y v tem primeru zavzema na posameznem elementu samo vrednosti od 0 do 1 (posedovanje ali neposedovanje določene lastnosti), se izraza za S^2 in s^2 poenostavita:

$$S^2 = \frac{NPQ}{N-1},$$

$$s^2 = \frac{np_0q_0}{n-1},$$

kjer je $Q = (1 - P)$ in $q_0 = (1 - p_0)$. Z uporabo teh poenostavitev dobimo:

$$V(p_0) = (1-f)\frac{NPQ}{(N-1)n}, \qquad [5]$$

$$v(p_0) = (1-f)\frac{p_0q_0}{(n-1)}. \qquad [6]$$

Če je n dovolj velik in lahko *FPC* zanemarimo, se cenilka vzorčne variance nadalje poenostavi v obliko:

$$v(p_0) = \frac{p_0q_0}{n}.$$

Zgornji izrazi veljajo za deleže, seveda pa jih lahko priredimo tudi za odstotke, če upoštevamo $Q = (100 - P)$ in $q_0 = (100 - p_0)$.

Recimo, da je 165 od 250 dijakov v našem vzorcu bralo določeno knjigo. Iz tega izhaja, da je $p_0 = 66$ %. Torej je 95-odstotni interval zaupanja za populacijski odstotek P enak:

$$66.0 \pm 1.96 \times \sqrt{\left(1-\frac{250}{1,872}\right)\frac{66\times 34}{249}} = 66.0 \pm 5.5\%.$$

S 5-odstotnim tveganjem torej trdimo, da je populacijska vrednost P za odstotek dijakov, ki berejo določeno knjigo, v intervalu med 60.5 % in 71.5 %.

Ponoviti velja, da smo v tem poglavju opisovali korake statističnega sklepanja samo za primer aritmetične sredine na podlagi SRS vzorcev in v tem okviru smo predstavili tudi izračunavanje intervalov zaupanja. Povsem enak pristop je mogoče uporabiti tudi za ocenjevanje drugih populacijskih parametrov pri SRS vzorčenju, podobna načela pa veljajo tudi pri kompleksnih vzorcih.

3. SISTEMATIČNO VZORČENJE

Ko smo v prejšnjem poglavju uporabili tabele slučajnih števil za izbiro SRS vzorca 250 dijakov, je računanje zahtevalo precej dela. Še več dela bi imeli pri večji populaciji ali v primeru, ko dijaki ne bi bili razporejeni po identifikacijskih številkah. V takih primerih lahko metoda sistematičnega vzorčenja *(angl. systematic sampling)* precej poenostavi delo pri izbiri vzorca. Sistematično vzorčenje je namreč nadvse enostavno za uporabo, saj izberemo v vzorec vsak k-ti element, pri čemer smo predhodno izbrali določen slučajni začetek.

Oglejmo si primer, ko potrebujemo vzorec $n = 250$ dijakov iz šole, ki šteje $N = 2{,}000$ dijakov. Vzorčni delež je $f = 250/2{,}000 = 0.125$, vzorčni interval pa 1:8. Sistematični vzorec velikosti 250 dobimo tako, da določimo slučajno številko med 1 in 8 – s tem določimo prvega dijaka v vzorcu – nato pa vključujemo vsakega osmega dijaka. Če bi slučajno izbrali, denimo, začetno številko 5, so dijaki v vzorcu določeni s številkami 5,13, 21 in tako naprej.

Uporaba sistematičnega vzorčenja v primeru iz prejšnjega poglavja je malce bolj zapletena, ker imamo vzorčni delež 250/1,872 oziroma korak 1:7.488. Interval tokrat ni celo število, kar bi lahko poenostavljeno rešili tudi tako, da število enostavno zaokrožimo. V našem primeru bi z razmerjem 1:7 dobili vzorec velikosti 267, z razmerjem 1:8 pa 234. Očitno torej je, da zaokroževanje ne prinese želene velikosti vzorca. Imamo pa tudi druge rešitve. Ena možnost je, da začnemo s poljubnim dijakom med 1 in 1,872, nakar izbiramo nadaljnje dijake s predpisanim celoštevilčnim vzorčnim intervalom toliko časa, dokler ne dobimo vzorca velikosti 250, pri tem pa seznam obravnavamo kot cikličen, kar pomeni, da se konec in začetek seznama stikata.

Druga rešitev je decimalni korak vzorčenja s celoštevilčnim zaokroževanjem navzdol. Denimo, da je slučajno začetno število v intervalu, ki ga določa tisočkratnik vzorčnega koraka (1,000–7,488), ravno število 3,654 – gre torej za izbiro slučajnega števila v tem intervalu. Ko vstavimo ustrezno decimalno vejico, dobimo slučajni začetek 3.654, ki mu prištevamo vzorčni korak, to je 7.488. Dobimo števila: 11.142, 18.630, 26.118 itd. Z zaokroževanjem navzdol (*angl. truncation*) izberemo v vzorec dijake s številkami 3, 11, 18, 26 itd. Interval med dijaki je torej včasih 7, včasih pa 8.

Tako kot pri SRS vzorčenju ima tudi pri sistematičnem vzorčenju vsak element v populaciji enako verjetnost izbora v vzorec. Imamo torej vzorec enakih verjetnosti (EPSEM), ki se od SRS vzorca razlikuje po tem, da verjetnost za vključitev v vzorec ni enaka za vse množice elementov. Povedano drugače, vsi vzorci niso enako verjetni. Za ponazoritev: v našem primeru izbiranja dijakov iz spiska zaporednih številk je verjetnost, da bosta v vzorec hkrati izbrana elementa (številki) 1 in 2, enaka 0. Po drugi strani pa je verjetnost, da bosta v vzorec hkrati izbrana elementa 1 in 9, enaka 1/8. Števili 1 in 9 se namreč hkrati pojavita natanko v enem od skupno osmih možnih vzorcev. Čeprav imajo vsi elementi enako verjetnost za izbor, pa očitno nimajo vsi elementi enake verjetnosti, da se pojavijo v vseh vzorcih. Zato tudi niso vsi vzorci enako verjetni. EPSEM lastnost sistematičnega vzorčenja sicer daje vtis, da je vzorčna aritmetična sredina dobra cenilka populacijske aritmetične sredine, vendar neenake verjetnosti vzorcev pomenijo, da izrazi za izračun standardne napake SRS vzorcev brez dodatnih predpostavk pri sistematičnem vzorčenju niso uporabni. Oglejmo si to nekoliko podrobneje.

Pri sistematičnem izbiranju s korakom 1:8 je nadvse enostavno določiti porazdelitev vzorčnih aritmetičnih sredin. Populacija vseh vzorcev je namreč majhna, saj obstaja samo 8 različnih vzorcev. Ker ima vseh 8 vzorcev enako verjetnost, da jih izberemo, je vzorčna porazdelitev sestavljena iz osmih vzorčnih aritmetičnih sredin, ki so enako verjetne – verjetnost vsake od njih je enaka 1/8. Pri sistematičnem vzorčenju imamo očitno na voljo izredno skromno število elementov za ocenjevanje vzorčne variance. Vzorčne aritmetične sredine se zato lahko močno razlikujejo in vzorčna varianca je lahko izredno velika.

Za ponazoritev vzemimo delež dijakov, ki trenutno berejo določeno knjigo. Predpostavimo, da 1,500 od 2,000 dijakov bere knjigo in da so dijaki urejeni v zaporedje šestih bralcev, ki jim sledita dva nebralca, nato pa spet šest bralcev itd. Vzorci, ki bi se slučajno začeli z začetkom 1, 2, 3, 4, 5 ali 6, bi tako vsebovali ravno vse bralce. Ocenjeni delež bralcev je v tem primeru $p = 100\,\%$. Vzorca z začetkom 7 in 8 pa bi vsebovala vse nebralce ($p = 0\,\%$). Vzorčna varianca je v takem primeru izredno velika, saj so tudi razlike med vzorčnimi deleži izjemno velike. Izračun nam zato daje standardno napako:

$$se(p) = 0.433,$$

kar pomeni, da je običajni interval zaupanja v vsako smer širok ±85 %. Poleg tega znotraj posameznega vzorca vrednosti sploh ne variirajo, saj npr. vsi dijaki v izbranem vzorcu bodisi berejo bodisi ne berejo, zato je v vsakem izbranem vzorcu variabilnost enaka $s^2 = 0$. Na podlagi enega vzorca zato vzorčne variance sploh ne bi mogli oceniti, kar je še dodatna neugodnost.

Zgornji primer je seveda generiran umetno, kljub temu pa ponazarja možne nevarnosti sistematičnega vzorčenja. Če vzorčno varianco ocenjujemo na podlagi sistematičnega vzorčenja, moramo torej sprejeti dodatne predpostavke o urejenosti elementov v populaciji. Samo kadar je seznam elementov slučajno urejen glede na ciljne spremenljivke, lahko tak vzorec obravnavamo kot SRS vzorec. Na srečo je v praksi skoraj vedno tako, da pojavi niso urejeni v povsem enakih intervalih.

Včasih je seznam elementov, ki jih izbiramo v vzorec, urejen po skupinah enakih elementov, npr. dijaki z enako oceno ali osebe v določenih regijah. V tem primeru lahko s sistematičnim vzorcem ravnamo kot s stratificiranim vzorcem. Gre za tako imenovano implicitno stratifikacijo, ki je nadvse ugodna, saj izboljšuje natančnost podobno kot proporcionalna stratifikacija, o čemer bomo podrobno govorili v naslednjem poglavju. Pri vsaki izdelavi vzorcev lahko zato seznam elementov pred izborom sistematičnega vzorca predhodno uredimo (sortiramo) glede na najpomembnejše razpoložljive spremenljivke in s tem pridobimo večino prednosti proporcionalne stratifikacije.

Povzamemo lahko, da je sistematično vzorčenje sporno samo, če je seznam glede na proučevani pojav urejen ciklično, naš vzorčni interval pa sovpada z večkratnikom dolžine cikla. Sama ciklična urejenost elementov v seznamu torej ni problematična, če ne sovpada z večkratnikom vzorčnega intervala. Ker pa je tako sovpadanje izredno redko, se sistematično vzorčenje v praksi uporablja brez posebnih zadržkov. Med izračunavanjem vzorčne variance se zato za večino praktičnih potreb sistematično vzorčenje obravnava povsem enako kot SRS vzorčenje.

4. STRATIFIKACIJA

Pri načrtovanju vzorcev pogosto vnaprej poznamo določene informacije o populaciji, ki jo proučujemo. Največkrat imamo na podlagi popisnih podatkov na voljo informacije o izbranem geografskem območju, kot so npr. tip naselja (mesto ali primestje), število oseb in sociodemografske značilnosti prebivalstva. Tovrstne informacije lahko uporabimo že pri načrtovanju vzorcev in z njihovo pomočjo izboljšamo vzorčni načrt. Po drugi strani pa z njimi lahko v fazi analize izboljšujemo natančnost ocenjevanja. Seveda je najbolje, če lahko z dodatnimi informacijami izboljšamo tako vzorčni načrt kot cenilko. V nadaljevanju bomo najprej govorili o informacijah, ki jih uporabimo pri načrtovanju vzorca s pomočjo postopkov stratifikacije *(angl. stratification)*.

Stratifikacija temelji na predhodni razdelitvi populacije na podpopulacije oziroma stratume, kar opravimo s pomočjo dodatnih populacijskih informacij. Pri tem je posebej pogosta stratifikacija na osnovi geografskih informacij, kjer regije obravnavamo kot stratume. V postopku izbiranja vzorca nato opravimo vzorčenje ločeno in neodvisno v vsakem stratumu posebej. Ena od prednosti stratifikacije je zato v tem, da so velikosti vzorcev v stratumih nadzorovane in niso določene slučajno v samem procesu vzorčenja. Pogosto so velikosti vzorca v stratumih enostavno sorazmerne ustrezni velikosti populacije. V takem primeru imamo proporcionalno stratifikacijo (*angl. proportional stratification*), kar pomeni, da v vseh stratumih uporabimo enak vzorčni delež. Poznamo pa tudi disproporcionalno stratifikacijo (*angl. disproportional stratification*), kjer imajo stratumi različne vzorčne deleže, zato v takem primeru struktura stratumov v vzorcu ne odseva strukture stratumov v populaciji.

Ne glede na vzorčni delež, s katerim izbiramo končne elemente v stratumih, pa v vsakem stratumu izbiramo elemente povsem neodvisno. Čeprav imamo za izbiro elementov znotraj stratumov na voljo različne načine verjetnostnega vzorčena, se bomo uvodoma omejili na uporabo SRS vzorčenja znotraj stratumov, kasneje pa si bomo ogledali tudi druge metode.

Najprej moramo oznake, ki smo jih uvedli pri SRS vzorčenju, razširiti na stratificirane vzorce. To storimo tako, da k siceršnjim simbolom dodajamo indeks *h*, ki označuje ustrezno količino v stratumu *h*. Imamo torej:

N_h velikost populacije v stratumu *h*,
n_h velikost vzorca v stratumu *h*,
$N = \sum N_h$ skupna velikost populacije,
$n = \sum n_h$ skupna velikost vzorca,

$f_h = \frac{n_h}{N_h}$ vzorčni delež v stratumu h,

$\bar{Y}_h$ populacijska aritmetična sredina v stratumu h,
$\bar{y}_h$ vzorčna aritmetična sredina v stratumu h,
S_h^2 elementarna varianca v stratumu h,
s_h^2 ocena elementarne variance v vzorcu v stratumu h.

K zgornjim oznakam lahko dodamo še delež populacije v stratumu h:

$$W_h = \frac{N_h}{N},$$

pri čemer velja:

$$\sum W_h = 1.$$

Poudariti velja, da vsi sumacijski simboli v poglavju o stratifikaciji predpostavljajo seštevanje po vseh stratumih h.

Če smo za izbor vzorca znotraj stratumov uporabili SRS vzorčenje, lahko rezultate SRS vzorčenja uporabimo znotraj vsakega stratuma. Cenilke $\bar{y}_h$ so zato nepristranske za $\bar{Y}_h$ in tudi njihove variance in standardne napake lahko ocenimo z znanimi izrazi SRS vzorčenja (1,2,3,4,5,6). Dodatna pri stratificiranem vzorcu je le kombinacija aritmetičnih sredin stratumov, ki vodi k cenilki populacijske aritmetične sredine $\bar{Y}$ in tudi k oceni vzorčne variance. Pri tem izhajamo iz dejstva, da se populacijska aritmetična sredina lahko izrazi kot:

$$\bar{Y} = \frac{\sum N_h \bar{Y}_h}{N} = \sum W_h \bar{Y}_h . \qquad [6a]$$

Cenilko populacijske aritmetične sredine dobimo, če v zgornjem izrazu zamenjamo neznane populacijske aritmetične sredine $\bar{Y}_h$ z vzorčnimi aritmetičnimi sredinami $\bar{y}_h$. Nepristranska cenilka, ki ji dodamo indeks »*st*« za stratifikacijo, ima zato obliko:

$$\bar{y}_{st} = \sum W_h \bar{y}_h . \qquad [6b]$$

Če v stratumih izbiramo vzorce ločeno in med seboj neodvisno, je mogoče s pomočjo statistične teorije pokazati, da je varianca cenilke $\bar{y}_{st}$ enaka:

$$V(\bar{y}_{st}) = \sum W_h^2 V(\bar{y}_h). \qquad [7]$$

Če smo v stratumih uporabili SRS vzorčenje, lahko namesto $V(\bar{y}_h)$ vstavimo ustrezni izraz (2) in dobimo naslednjo obliko vzorčne variance:

$$V(\bar{y}_{st}) = \sum W_h^2 \frac{(1-f_h)S_h^2}{n_h}. \qquad [8]$$

Cenilko za $V(\bar{y}_{st})$ dobimo z zamenjavo s_h^2 za neznano vrednost S_h^2 v izrazu (8):

$$v(\bar{y}_{st}) = \sum W_h^2 \frac{(1-f_h)s_h^2}{n_h}. \qquad [9]$$

Proporcialna stratifikacija

Zgornji izrazi veljajo za katerokoli razporeditev vzorca v stratumih. Pri proporcionalni stratifikaciji oziroma enotnem vzorčnem deležu:

$$f = \frac{n_h}{N_h} = \frac{n}{N} \qquad [9a]$$

pa se izrazi nadalje poenostavijo. Proporcionalna stratifikacija je namreč EPSEM vzorčni načrt in cenilka zato preide v enostavno vzorčno aritmetično sredino:

$$\bar{y}_{ps} = \sum W_h \bar{y}_h = \frac{1}{n}\sum_h \sum_i y_{hi} ,$$

kjer oznaka y_{hi} pomeni vrednost spremenljivke y za i-ti element v stratumu h. Varianca iz izraza (8) se zato poenostavi:

$$V(\bar{y}_{ps}) = (1-f)\frac{\sum W_h S_h^2}{n} = (1-f)\frac{S_w^2}{n}. \qquad [10]$$

Pri tem je:

$$S_w^2 = \sum W_h S_h^2$$

utežena aritmetična sredina elementarnih varianc znotraj stratumov (*angl. within stratum variance*). Varianco $V(\bar{y}_{ps})$ torej lahko ocenimo:

$$v(\bar{y}_{ps}) = (1-f)\frac{\sum W_h s_h^2}{n}. \qquad [11]$$

Varianca vzorčne aritmetične sredine, ki temelji na proporcionalno stratificiranem vzorcu (10), je očitno podobna varianci v SRS vzorcu (2). Razlika je le v tem, da je populacijska elementarna varianca S^2 v izrazu za SRS vzorčenje tokrat zamenjana z uteženo aritmetično sredino elementarnih varianc znotraj stratumov S_w^2. Z običajno analizo variance lahko pokažemo, da za dovolj velike N_h velja izraz:

$$S^2 \cong S_w^2 + \sum W_h(\bar{Y}_h - \bar{Y})^2 . \qquad [11a]$$

Ker v zgornjem izrazu vsota kvadratov v drugem členu na desni ni nikoli manjša od 0, proporcionalno stratificirani vzorec ne more biti manj natančen kot SRS vzorec. Razberemo tudi, da je učinek proporcionalne stratifikacije večji pri večjih razlikah med aritmetičnimi sredinami stratumov. Pri proporcionalni stratifikaciji si torej prizadevamo za čim bolj heterogene stratume, ki so notranje kar najbolj homogeni.

Za primerjavo vzorčnih načrtov pogosto uporabljamo izraz vzorčni učinek (*angl. design effect*), ki ga na kratko označimo z *Deff*. To je splošna mera za primerjavo variance enostavnega slučajnega vzorčenja (SRS) z varianco kompleksnih vzorcev enakih velikosti, kjer primerjamo dve varianci za isto spremenljivko:

$$\text{vzorčni učinek} = \frac{\text{varianca cenilke kompleksnega vzorca}}{\text{varianca cenilke SRS vzorca enake velikosti}}.$$

Pri splošni cenilki *z* parametra *Z* lahko zapišemo:

$$Deff(z) = \frac{Var(z)}{Var(z_0)},$$

kar pri ocenjevanju aritmetične sredine proporcionalno stratificiranega vzorca pomeni:

$$Deff(\bar{y}_{pr}) = \frac{S_w^2}{S^2}.$$

V zgornjem primeru je vzorčni učinek zaradi $S_w^2 < S^2$ seveda manjši od 1. Včasih pa vzorčni učinek raje izražamo kot razmerje med standardno napako cenilke v kompleksnem vzorcu ter standardno napako cenilke v SRS vzorcu, kar je kvadratni koren zgoraj definiranega vzorčnega učinka. Takrat količino označimo z *Deft*:

$$Deft = \sqrt{Deff}.$$

Koren vzorčnega učinka *Deft* ima posebej nazorno interpretacijo – pomeni namreč faktor, s katerim se širi oziroma zoži siceršnji interval zaupanja, ki bi ga izračunali na osnovi predpostavke SRS vzorčenja.

Vzorčni učinek *Deff* bi lahko opredelili tudi s primerjavo kompleksnega vzorca ter ustreznega SRS vzorca s ponavljanjem. Ker pa je razlika v varianci pri SRS vzorcu s ponavljanjem ter SRS vzorcu brez ponavljanja le v faktorju *FPC*, ki ga običajno zanemarimo, je razlika povsem nepomembna.

Za ponazoritev proporcionalne stratifikacije se vrnimo k primeru iz prejšnjih poglavij. Denimo, da je seznam dijakov razdeljen v štiri ločene sezname, ki so določeni s šolskimi razredi (prvi, drugi, tretji in četrti). Razredi torej določajo stratume, iz katerih izbiramo posamezne vzorce. Stolpca 2 in 3 v tabeli 2 prikazujeta število in razmerja populacije v vsakem razredu. Stolpec 4 prikazuje velikost vzorca, ki smo ga dobili iz vsakega stratuma s proporcionalno razmestitvijo in z vzorčnim deležem 250/1,872 oziroma s korakom 1:7.488. Stolpci 5, 6 in 7 prikazujejo vsoto, vzorčno aritmetično sredino ter elementarno varianco v vzorcu v vsakem stratumu za obseg gledanja TV na dan (izraženo v urah). Stolpca 8 in 9 prikazujeta število in odstotek dijakov v vzorcu, ki trenutno bere določeno knjigo.

Tabela 2: Proporcionalno stratificiran vzorec dijakov

1	2	3	4	5	6	7	8	9
Stratum	N_h	W_h	n_h	$\sum_i y_{hi}$	$\bar{y}_h$	s_h^2	r_h	p_h
1. razred	524	0.28	70	168	2.40	0.941	35	50%
2. razred	487	0.26	65	169	2.60	1.088	39	60%
3. razred	449	0.24	60	123	2.05	0.804	45	75%
4. razred	412	0.22	55	88	1.60	0.643	44	80%
Skupaj	1,872	1.00	250	548			163	

Vzorčno aritmetično sredino za število ur gledanja TV lahko izračunamo iz splošnega izraza (6b). Ker pa je vzorec proporcionalen, lahko izračunamo tudi enostavneje:

$$\bar{y}_{ps} = \sum_h \sum_i \frac{y_{hi}}{n} = \frac{548}{250} = 2.192.$$

Na enak način lahko izračunamo tudi odstotek dijakov, ki berejo določeno knjigo:

$$p_{ps} = 100\frac{\sum_h r_h}{n} = 100\frac{163}{250} = 65.2\%.$$

Vzorčno varianco ocenimo s pomočjo izraza (11):

$$v(\bar{y}_{ps}) = \left(1 - \frac{250}{1{,}872}\right)\frac{0.8808}{250} = 0.003053$$

in sledi:

$$se(\bar{y}_{ps}) = 0.0553.$$

Običajni interval zaupanja (95 %) je torej enak:

$$\bar{y}_{ps} \pm 1.96 \times se(\bar{y}_{ps}) = 2.192 \pm 1.96 \times 0.055 = 2.192 \pm 0.1084$$

in se razteza od 2.08 do 2.30.

Ocenjeno varianco za p_{ps} lahko prav tako dobimo iz izraza (11), pri čemer je:

$$s_h^2 = \frac{n_h}{n_h - 1} p_h q_h.$$

Torej je:

$$v(p_{ps})=\left(1-\frac{250}{1,872}\right)\frac{2,160}{250}=7.486.$$

Na tej podlagi je:

$$se(p_{ps})=2.736\,\%,$$

zato običajni interval zaupanja sega od 59.8 % do 70.6 %.

Vzorčna učinka za $\bar{y}_{ps}$ in p_{ps} lahko ocenimo s pomočjo razmerja s_w^2 / s^2. Pri $\bar{y}_{ps}$ je ustrezna vrednost s^2, ki jo lahko ocenimo na osnovi izraza (11a), enaka $s^2 = 1.008$ (Cochran, 1977, poglavje 5.A11) in sledi:

$$deff(\bar{y}_{ps})=0.87.$$

Z drugimi besedami, potrebujemo SRS vzorec velikosti $n = 250/0.87 = 287$, da bi dosegli enako natančnost kot pri proporcionalno stratificiranem vzorcu velikosti 250. Izboljšanje natančnosti v primerjavi s SRS vzorcem torej izhaja iz razlik med aritmetičnimi sredinami števila ur gledanja televizije med stratumi, kar smo izkoristili pri stratificiranem vzorcu.

V primeru ocenjevanja deleža p_{ps} je ustrezna vrednost $s^2 = 2,278$, tako da znaša ocena za vzorčni učinek *Deff*(p_{ps}):

$$deff(p_{ps})=0.95.$$

Izboljšanje v varianci je s stratifikacijo v tem primeru majhno, posebej če upoštevamo razmeroma velike razlike med stratumi v stolpcu 9 tabele 2. Relativno majhna izboljšava pa je nasploh običajna pri računanju deležev oziroma odstotkov, razen če imamo hkrati stratume z izjemno visokimi (npr. nad 90 %) in izjemno nizkimi (npr. pod 10 %) odstotki.

Disproporcionalna stratifikacija

Proporcionalna stratifikacija se pogosto uporablja, ker so cenilke v tem primeru relativno enostavne in ker jamči, da ocene niso manj natančne kot pri SRS vzorcu enake velikosti. Seveda pa se v praksi pojavljajo primeri, ko je koristna in uporabna tudi disproporcionalna stratifikacija.

Eden od ciljev disproporcionalne stratifikacije (*angl. disproportional stratification*) je doseči razmestitev, ki kar najbolj poveča natančnost cenilke populacijske aritmetične sredine v okviru danih finančnih (ali drugih) omejitev. Izkaže se, da dosežemo optimalno razmestitev (*angl. opti-*

mal allocation) takrat, kadar je vzorčni delež v stratumih sorazmeren z elementarnim standardnim odklonom in obratno sorazmeren s kvadratnim korenom stroškov anketiranja v določenem stratumu:

$$f_h \propto \frac{S_h}{\sqrt{c_h}}.$$

Pri tem simbol »∝« označuje sorazmernost, stroške anketiranja izbranega elementa v stratumu *h* pa smo označili s c_h. Bolj heterogeni stratumi in stratumi z nižjimi stroški bodo torej vzorčeni z višjimi vzorčnimi deleži. Pogosto pa se stroški med stratumi ne razlikujejo in v tem primeru optimalno razmestitev omejimo na:

$$f_h \propto S_h,$$

kar imenujemo tudi Neymanova razmestitev (*angl. Neyman allocation*).

Praktična težava pri uporabi optimalne razmestitve je nepoznavanje količin S_h in c_h. Na srečo nam zadostujejo že grobe ocene, saj je izguba natančnosti zaradi manjših odstopanj običajno zanemarljiva.

Disproporcionalna stratifikacija lahko – v primerjavi s proporcionalno – prinese izjemno velike izboljšave v natančnosti, po drugi strani pa lahko v nekaterih spremenljivkah prinese tudi manj natančne ocene kot SRS vzorec enake velikosti. Slednje se pri proporcionalni stratifikaciji ne more zgoditi, zato je proporcionalna stratifikacija varna za vse spremenljivke.

Disproporcionalna stratifikacija se pogosto uporablja pri načrtovanju vzorcev, ki zagotavljajo dovolj natančne ocene za vsak stratum. Vzorčne ocene namreč pogosto potrebujemo ne le za celotno populacijo, ampak tudi za podpopulacije, ki jih imenujemo domene proučevanja (*angl. domains of study*). Kadar pa majhen stratum predstavlja lastno domeno proučevanja, se lahko zgodi, da s proporcionalno stratifikacijo izberemo v njem premajhen vzorec, da bi zagotovili natančne ocene. Ena od rešitev je, da v takem stratumu vzorčimo s povečanim vzorčnim deležem.

Disproporcionalno razmestitev uporabljamo tudi, kadar nas zanimajo primerjave ocen med posameznimi stratumi. Tako je v primeru dveh stratumov optimalna razmestitev vzorca za ocenjevanje razlike med aritmetičnima sredinama dveh stratumov naslednja:

$$\frac{n_1}{n_2} \propto \frac{S_1/\sqrt{c_1}}{S_2/\sqrt{c_2}}.$$

Pogosto pa sta standardna odklona in stroški enaki v obeh stratumih in takrat se optimalna razmestitev poenostavi na $n_1=n_2$. Očitno je torej, da je za primerjavo med stratumi velikost populacije v stratumih nepomembna; pomembna postane šele pri izračunu ocene za celotno populacijo.

Kadar pa hkrati potrebujemo oceno za celotno populacijo ter oceno za razlike aritmetičnih sredin med stratumi, lahko stratumi, ki se po velikosti razlikujejo, povzročijo protislovne rešitve. Za primerjave stratumov je namreč velikost populacije N_h v stratumu nepomembna, medtem ko je pri izračunu cenilke za celotno populacijo to nadvse pomembna količina. Če sta npr. v dveh stratumih enaka elementarna standardna odklona in so enaki tudi stroški anketiranja, pri čemer prvi stratum zajame 90 % populacije, drugi pa 10 % populacije, je optimalna razmestitev vzorca n = 500 naslednja:

- Za ocenjevanje aritmetične sredine celotne populacije je optimalna izbira v prvem stratumu $n_1 = 450$, v drugem stratumu pa $n_2 = 50$.
- Optimalna razmestitev za ocenjevanje razlike med aritmetičnima sredinama spremenljivke v dveh stratumih pa je $n_1 = n_2 = 250$ za vsak stratum.

Očitno se optimalni razmestitvi močno razlikujeta. Optimalna razmestitev za en namen je namreč v drugem primeru neustrezna. V praksi zato skušamo običajno najti nekakšen približen in sprejemljiv kompromis.

Za nadaljnjo ponazoritev disproporcionalnega vzorca se vrnimo k našemu primeru. Podatki v tabeli 3 so enaki podatkom v tabeli 2, razen v stolpcu 4. Izbrana razmestitev namreč tokrat razporeja vzorec 250 elementov enakomerno v štiri stratume. V tem primeru smo torej sledili razmestitvi, ki jo potrebujemo, kadar je vsak stratum samostojna domena, torej samostojen predmet proučevanja. Pri tem smo predpostavili, da so elementarna varianca in stroški anketiranja v vseh stratumih enaki.

Tabela 3: Disproporcionalno stratificiran vzorec dijakov srednje šole

1	2	3	4	5	6	7
Stratum	N_h	W_h	n_h	$\sum_i y_{hi}$	$\bar{y}_h$	s_h^2
1. razred	524	0.28	63	151.2	2.40	0.941
2. razred	487	0.26	63	163.8	2.60	1.088
3. razred	449	0.24	62	127.1	2,05	0.804
4. razred	412	0.22	62	99.2	1.60	0.643
Skupaj	1,872	1.00	250			

Skupno aritmetično sredino za število ur gledanja TV izračunamo na podlagi izraza:

$$\bar{y}_{st} = \sum W_h \bar{y}_h = 2.192,$$

kar je enak rezultat kot v prejšnjem primeru. Seveda pa enostavna aritmetična sredina:

$$\sum\sum \frac{y_{hi}}{n} = 2.165$$

tokrat ni pravilna cenilka za $\bar{Y}$. Višji razredi, ki imajo – glede na število dijakov v populaciji – tokrat v vzorcu preveč elementov, namreč poročajo o manjšem številu ur gledanja televizije, kar je tudi razlog za zgornje podcenjevanje. Varianco $\bar{y}_{st}$ ocenimo s pomočjo izraza (9) in znaša:

$$v(\bar{y}_{st}) = \sum W_h^2 \frac{(1-f_h)s_h^2}{n_h} = 0.003117.$$

Izračunajmo še standardno napako:

$$se(\bar{y}_{st}) = 0.0558$$

in jo primerjajmo z rezultatom pri proporcionalni stratifikaciji $se(\bar{y}_{st})$ = 0.0553. Primerjava kaže na malce slabšo oceno populacijske aritmetične sredine pri disproporcionalni razmestitvi, kar je tudi razumljivo, saj smo disproporcionalno stratifikacijo izbrali tako, da ustreza ocenjevanju posameznih domen ter razlik med njimi. Če pa bi disproporcionalno stratifikacijo izbirali tako, da bi optimizirala natančnost ocenjevanja populacijske aritmetične sredine, bi vsekakor dosegli večjo natančnost kot pri proporcionalni stratifikaciji.

Izbiranje stratumov

Za učinkovito uporabo postopka stratifikacije morata biti izpolnjena dva osnovna pogoja:
- znana morajo biti populacijska razmerja W_h,
- možno je ločeno in neodvisno izbiranje vzorcev po stratumih.

Če zgornja pogoja nista izpolnjena, stratifikacijo pa želimo kljub temu opraviti, si lahko pomagamo s poststratifikacijo ali dvofaznim vzorčenjem, kar bomo obravnavali v posebnih poglavjih. Pogosto pa želimo pri stratifikaciji zadostiti še eni omejitvi: v vsakem stratumu mora biti v vzorec izbran vsaj en element. V nasprotnem primeru namreč ni mogoče izračunati nepristranske cenilke za aritmetično sredino v celotni populaciji. Če torej želimo izračunati oceno standardne napake, moramo v vsakem stratumu izbrati vsaj dva elementa.

V praksi je pogosto precej informacij, ki jih lahko uporabimo pri stratifikaciji in so nam lahko v veliko pomoč pri izbiri primernih stratumov. Za čim večjo natančnost cenilke pri ocenjevanju parametrov za celotno populacijo pa velja upoštevati osnovno načelo, da morajo biti stratumi oblikovani tako, da so glede ciljne spremenljivke notranje čim bolj homogeni. Hkrati s tem morajo med stratumi obstajati čim večeje razlike v aritmetičnih sredinah $\bar{Y}_h$, kar omogoča učinkovito proporcionalno stratifikacijo, ter v variancah S_h^2, kar je pogoj za dodatne izboljšave v vzorčni varianci pri disproporcionalni stratifikaciji.

Kadar potrebujemo cenilke za majhne domene proučevanja, mora biti vsaka od domen v svojem stratumu ali v skupini stratumov. Pri premajhnih stratumih zato izbiramo elemente s povečanim vzorčnim deležem, da bi dobili velikost vzorca, ki zagotavlja potrebno natančnost. Včasih je tudi primerno, da oblikujemo stratume, v katerih uporabimo različne vzorčne načrte.

Vrnimo se spet k srednješolskemu primeru in si oglejmo hkratno uporabo večih stratifikacijskih spremenljivk. Predpostavimo, da poleg delitve po razredih obstajajo še druge spremenljivke:

- spol,
- povprečna ocena šolskega uspeha dijakov, opredeljena s tremi kategorijami (odlično, prav dobro, dobro),
- tip naselja, v katerem dijak živi (mestno, primestno, vaško).

Za vsako od teh spremenljivk predvidevamo, da so povezane z gledanjem TV in v tem primeru lahko opredelimo $2\times3\times3=18$ stratumov. Seveda pa so stratumi lahko oblikovani tudi povsem subjektivno, če menimo, da je to potrebno. Dokler znotraj vsakega stratuma uporabljamo verjetnostno vzorčenje, nobeno izbiranje stratumov ne more povzročiti pristranskosti v cenilkah. Merilo pri oblikovanju stratumov je zato predvsem njihova notranja homogenost, ki ugodno vpliva na standardno napako cenilk v stratificiranih vzorcih. Če pa se kateri od stratumov izkaže za premajhnega, kar pomeni, da bi v njem izbrali manj kot dva elementa, ga lahko združimo s sosednjim stratumom.

Težava pri oblikovanju velikega števila stratumov s proporcionalno razmestitvijo je lahko tudi to, da so v nekaterih manjših stratumih velikosti vzorca določene decimalno. Če moramo npr. vzorčiti v razmerju 1 : 7.488 pri 19 dijakih v stratumu, to pomeni izbor 2.54 dijaka. Čeprav ima zaokroževanje na najbližje celo število pri velikih n zanemarljiv učinek na verjetnost izbora, pa to ne velja za majhne velikosti vzorcev. Običajni način, da se temu problemu izognemo, je uporaba implicitne stratifikacije *(angl. implicit stratification)* namesto eksplicitne stratifikacije (*angl. explicit stratification*). Pri implicitni stratifikaciji, ki smo jo omenili že pri sistematičnem vzorčenju, seznam najprej uredimo po stratumih in nato sistematično izberemo vzorec. S tem v celoti rešimo problem decimalne velikosti vzorca. V našem primeru bi tako v stratumu, ki vključuje 19 dijakov, lahko izbrali 2 ali 3 dijake, odvisno od slučajnega začetka pri sistematičnem izboru.

5. SKUPINE IN VEČSTOPENJSKO VZORČENJE

Vzorčenje v skupinah

Pri večini vzorčnih načrtov lahko privzamemo, da je populacija sestavljena iz različnih podmnožic, v katere se elementi naravno združujejo. Primer take podmnožice so stratumi, kar smo obravnavali v prejšnjem poglavju. Pri stratifikaciji smo v vzorec vključevali vse stratume in v vsakem izbrali neodvisni (pod)vzorec elementov. Nekoliko drugačen primer, pri katerem namesto elementov najprej izbiramo njihove množice, pa so skupine (*angl. clusters*). Tipični primeri skupin elementov (npr. oseb) so podjetja, šole, razredi, družine, naselja ipd. Kadar so v vzorec vključeni vsi elementi izbrane skupine, govorimo o vzorčenju v skupinah (*angl. cluster sampling*). Če pa izberemo v vsaki skupini vzorec elementov, govorimo o dvostopenjskem vzorčenju (*angl. two stage sampling*) oziroma dvostopenjskem vzorčenju v skupinah (*angl. two stage cluster sampling*).

Pogosto uporabljamo hierarhijo skupin in najprej izberemo večje skupine, potem v njih manjše in tako naprej, dokler ne pridemo do stopnje, kjer izbiramo elemente. Pri raziskavi med dijaki bi lahko najprej izbrali vzorec šol, v izbranih šolah vzorec razredov in šele nato vzorec dijakov. V splošnem tovrstno izbiro imenujemo večstopenjsko vzorčenje (*angl. multistage sampling*), saj zaporedno vzorčimo na več stopnjah.

Čeprav so tako stratumi kot skupine množice elementov, pa se njihova vloga v vzorčenju precej razlikuje. Vsi stratumi so namreč vedno vključeni v vzorec in običajno povečujejo natančnost, pri skupinah pa je nekoliko drugače. Prvič, v vzorec ne vključimo vseh skupin, ampak le njihov vzorec, zato morajo izbrane skupine predstavljati tudi preostale (neizbrane) skupine. Pri vzorčenju v skupinah je zato ugodno, če so skupine notranje heterogene in ne homogene tako kot pri stratumih. Nadaljnja razlika med skupinami in stratumi pa je v vplivu na natančnost. Proporcionalna stratifikacija v primerjavi s SRS vzorcem enake velikosti namreč vedno izboljša natančnost, podobno velja tudi za večino drugih postopkov stratifikacije. Po drugi strani pa vzorčenje v skupinah običajno manjša natančnost v primerjavi s SRS vzorcem iste velikosti. Kljub slabši natančnosti se zaradi velikih prihrankov pri stroških anketiranja vzorčenje v skupinah pogosto uporablja predvsem pri osebnem anketiranju. Pri tem velja ponoviti, da z osebnim anketiranjem razumemo terensko osebno anketiranje, ne pa telefonsko anketiranje, kjer bi tudi lahko govorili o določenem osebnem stiku med anketarjem in anketirancem.

Za enostavnejšo predstavitev temeljnih načel vzorčenja v skupinah in večstopenjskega vzorčenja bomo uvodoma privzeli poenostavitev – ki v praksi večinoma ni realna – da imajo vse skupine v populaciji enako velikost B. Skupine neenakih velikosti bomo obravnavali v naslednjem poglavju.

Denimo torej, da imamo v populaciji A skupin z enako velikostjo B. Pri tem a skupin – vključno z vsemi elementi – izberemo v vzorec s SRS vzorčenjem. Tako skupine kot tudi elementi v skupinah so seveda označeni in oštevilčeni. V populaciji se oznaka $Y_{\alpha\beta}$ tako nanaša na element β znotraj skupine α. Pri tem je velikost populacije $N = AB$, velikost vzorca pa $n = aB$. Vzorčni delež je torej:

$$f = \frac{n}{N} = \frac{aB}{AB} = \frac{a}{A},$$

saj v vsaki skupini vključimo v vzorec vse elemente. Populacijsko aritmetično sredino spremenljivke y v skupini α označimo z naslednjim izrazom:

$$\bar{Y}_\alpha = \frac{1}{B}\sum_{\beta=1}^{B} Y_{\alpha\beta},$$

in podobno tudi populacijsko aritmetično sredino:

$$\bar{Y} = \frac{1}{N}\sum_{\alpha=1}^{A}\sum_{\beta=1}^{B} Y_{\alpha\beta} = \frac{1}{A}\sum_{\alpha=1}^{A}\bar{Y}_\alpha.$$

V vzorcu imamo naslednje ocene:

$$\bar{y}_\alpha = \frac{1}{B}\sum_{\beta=1}^{B} y_{\alpha\beta}.$$

Seveda za skupine, ki so bile izbrane v vzorec, tokrat izjemoma velja enakost $\bar{y}_\alpha = \bar{Y}_\alpha$. Cenilko na osnovi vzorca v skupinah pa izrazimo:

$$\bar{y}_c = \frac{1}{n}\sum_{\alpha=1}^{a}\sum_{\beta=1}^{B} y_{\alpha\beta} = \frac{1}{a}\sum_{\alpha=1}^{a}\bar{y}_\alpha.$$

Pri skupinah enakih velikosti je populacijska aritmetična sredina $\bar{Y}$ torej enostavno povprečje aritmetičnih sredin $\bar{Y}_\alpha$ vseh A skupin, vzorčna aritmetična sredina $\bar{y}_c$ pa povprečje vzorčnih aritmetičnih sredin $\bar{y}_\alpha$ v a izbranih skupinah. Vzorec skupin zato lahko obravnavamo kot SRS vzorec a aritmetičnih sredin iz populacije A skupin. Cenilka populacijske aritmetične sredine na osnovi vzorca v skupinah, ki smo jo označili z $\bar{y}_c$, je zato nepristranska cenilka za $\bar{Y}$. Njena varianca izhaja iz izraza (2), v katerem kot elementi tokrat nastopajo skupine:

$$V(\bar{y}_c) = \left(1 - \frac{a}{A}\right)\frac{S_a^2}{a}, \qquad [12]$$

kjer je:

$$S_a^2 = \sum_{\alpha=1}^{A} \frac{(\bar{Y}_\alpha - \bar{Y})^2}{(A-1)} \qquad [12a]$$

elementarna varianca vzorčnih aritmetičnih sredin med skupinami. Ocena vzorčne variance:

$$v(\bar{y}_c) = \left(1-\frac{a}{A}\right)\frac{s_a^2}{a} \qquad [13]$$

je nepristranska cenilka za $V(\bar{y}_c)$, kjer je:

$$s_a^2 = \frac{1}{a-1}\sum_{\alpha=1}^{a}(\bar{y}_\alpha - \bar{y}_c)^2 .$$

Če primerjamo varianco nepristranske cenilke $V(\bar{y}_c)$ z varianco cenilke aritmetične sredine iz SRS vzorca velikosti $n=aB$, lahko vzorčni učinek $Deff(\bar{y}_c)$ izračunamo:

$$Deff(\bar{y}_c) = \frac{\frac{S_a^2}{a}}{\frac{S^2}{aB}} = \frac{BS_a^2}{S^2} . \qquad [13a]$$

Vzorčni učinek $Deff(\bar{y}_c)$ je torej odvisen od razmerja S_a^2 in S^2, ki je predvsem posledica načina oblikovanja skupin. Če je število skupin A v populaciji veliko in skupine sestavljamo povsem naključno, je S_a^2 ravno vzorčna varianca SRS vzorca velikosti B:

$$V(\bar{y}_\alpha) = S_a^2 = \frac{S^2}{B} .$$

Pri takšnih pogojih je $Deff(\bar{y}_c)$ seveda enak 1. Če pa so skupine notranje bolj homogene kot pri slučajni razmestitvi – kar se zgodi skoraj vedno – bodo aritmetične sredine skupin bolj heterogene, zato bo S_a^2 večji od S^2/B in $Deff(\bar{y}_c)$ bo večji od 1.

Vzorčni učinek lahko izrazimo tudi takole:

$$Deff(\bar{y}_c) \doteq 1 + (B-1)\rho , \qquad [14]$$

kjer ρ predstavlja posebno mero za homogenost skupin, ki jo imenujemo intraklasna korelacija (*angl. intraclass correlation*). Intraklasna korelacija torej meri stopnjo podobnosti elementov (Kish, 1965: poglavje 5.4) v skupinah. Nekoliko poenostavljeno lahko tudi rečemo, da je ρ delež elementarne variance, ki jo lahko pojasnimo s pripadnostjo skupini. Kadar so pri velikih populacijah skupine oblikovane povsem slučajno, je zato $\rho \doteq 0$ in $Deff(\bar{y}_c) \doteq 1$, kar je seveda enakovredno SRS vzorcu.

Negativna vrednost ρ bi pomenila, da so skupine notranje bolj heterogene, kot če so oblikovane povsem slučajno. Takšne vrednosti ρ so sicer možne, vendar nadvse redke. Ena takih izjem je npr. spolna struktura populacije, ki v skupinah običajno variira manj, kot bi sicer variiral delež okoli 50 % v SRS vzorcu. Kljub temu pa ρ ne more biti manjši kot:

$$\rho = \frac{-1}{(B-1)}.$$

Če je ρ negativen, je vzorčni učinek seveda manjši od 1 in vzorčenje v skupinah je v takem primeru natančnejše kot SRS vzorčenje. Kot rečeno, pa so v praksi negativne vrednosti ρ izjemno redke. Običajno ima ρ majhne pozitivne vrednosti, večinoma pod ρ = 0.15, tako da je vzorčni učinek praviloma večji od 1. Poudariti je treba, da je lahko vzorčni učinek $Deff(\bar{y}_c)$ kljub majhnim vrednostim ρ pri velikem številu elementov v skupini izredno velik.

Najvišja možna vrednost za ρ je 1 in se pojavi takrat, ko imajo znotraj vsake skupine vsi elementi enako vrednost, npr. vsi stanovalci v soseski hkrati vidijo določen televizijski program ali pa ga ne vidijo. V takem primeru je seveda $Deff(\bar{y}_c)$ enostavno enak številu elementov v skupini.

Za ponazoritev se vrnimo k našemu primeru vzorčenja dijakov. Predpostavimo, da skupine predstavlja A = 78 razredov, v katerih je po B = 24 dijakov. Z organizacijskega kot tudi finančnega stališča je nedvomno primerno, da v izbranem razredu anketiramo vse dijake. Izberimo vzorec a = 10 razredov, skupno torej 240 dijakov. Spodnji pregled prikazuje deleže dijakov, ki so prebrali določeno knjigo, v vsakem od desetih izbranih razredov:

$$\frac{9}{24},\frac{11}{24},\frac{13}{24},\frac{15}{24},\frac{16}{24},\frac{17}{24},\frac{18}{24},\frac{20}{24},\frac{20}{24},\frac{21}{24}.$$

Skupni delež, izražen v odstotkih, je:

$$p_c = \frac{160}{240} = 66.7\%.$$

Iz izraza (13) lahko izračunamo tudi oceno za vzorčno varianco:

$$v(p_c) = \left(1-\frac{10}{78}\right)\frac{0.02816}{10} = 0.002455,$$

iz česar sledi, da je standardna napaka ocene enaka:

$$se(p_c) = 0.04955$$

oziroma 4.96 %. Zgornja ocena $se(p_c)$ pa ima le $(t-1)$ = 10–1 = 9 prostostnih stopenj (*angl. degrees of freedom*), zato moramo namesto normalne porazdelitve tokrat uporabiti t-porazdelitev. Pri 5-odstotnem tveganju imamo naslednjo vrednost:

$$t_{\alpha/2}(9) = 2.26$$

in na tej osnovi je 95-odstotni interval zaupanja enak:

$$(66.7 \pm 2.26 \times 4.96)\%.$$

Populacijska vrednost se torej pri 5-odstotnem tveganju nahaja med 55.5 % in 77.9 %.

Varianco $v(p_c)$ lahko primerjamo z varianco SRS vzorca enake velikosti (6):

$$v(p_0) = \left(1 - \frac{240}{1872}\right)\frac{0.6667 \times 0.3333}{239} = 0.0008106.$$

Standardna napaka je v tem primeru enaka $se(p_0)$ = 0.02847 oziroma 2.85 %. Od tod lahko ocenimo tudi vrednost vzorčnega učinka $Deff(p_c)$:

$$deff(p_c) = \frac{0.002455}{0.0008106} = 3.029.$$

Z uporabo izraza (14) lahko s statistiko ρ' ocenimo še ρ:

$$\rho' = \frac{deff(p_c) - 1}{B - 1} = 0.088.$$

Očitno je pozitivna vrednost ρ – skupaj z velikostjo skupin B – povzročila, da je vzorčenje v skupinah precej manj natančno kot SRS vzorec enake velikosti. Če zanemarimo učinek faktorja končne populacije *(FPC)*, mora biti v našem primeru vzorec v skupinah približno trikrat večji, da bi bil tako natančen kot SRS vzorec.

Kot kaže izraz (14), je $Deff(\bar{y}_c)$ odvisen od dveh faktorjev: od intraklasne korelacije ρ in od velikosti skupin B. Velika vrednost za $Deff(\bar{y}_c)$ v zgornjem primeru je nastala zaradi obeh razlogov, tako zaradi visoke homogenosti spremenljivke (gledanje televizije) znotraj razredov kot zaradi precejšnje velikosti razredov.

Ob tem velja ponoviti, da lahko tudi pri majhni vrednosti ρ zavzame vzorčni učinek $Deff(\bar{y}_c)$ velike vrednosti, če je faktor $(B - 1)$ velik in je torej v skupinah veliko elementov (B). In obratno, majhne skupine opazno znižujejo vrednost $Deff(\bar{y}_c)$. Če bi npr. imeli v zgornjem primeru velikost skupine B = 8 (prej B = 24), bi se pri enaki intraklasni korelaciji ocenjeni vzorčni učinek $Deff(\bar{y}_c)$ zmanjšal s prejšnjih 3.029 na samo 1.62. Kadar torej pri enaki stopnji homogenosti zmanjšamo velikost skupine B, se skoraj linearno zmanjša tudi $Deff(\bar{y}_c)$.

V splošnem se vrednost ρ nekoliko veča tudi z manjšanjem povprečne velikosti populacijskih skupin – npr. v popisnih okoliših kot skupinah bo ρ manjši kot v krajevnih skupnostih – vendar je tovrstna stopnja naraščanja ρ majhna, tako da ima prevladujoč vpliv na $Deff(\bar{y}_c)$ predvsem neposredna velikost B.

Pri vzorčenju v skupinah je torej vzorčni učinek $Deff(\bar{y}_c)$ pri manjših skupinah v splošnem nižji. Seveda pa morajo biti po drugi strani skupine dovolj velike, da zagotovijo prihranek pri stroških anketiranja. Šele ko je to zagotovljeno, izberemo skupine čim manjšega velikostnega obsega. V našem primeru bi tako lahko dijake združevali tudi po letnikih; ker pa to v postopku anketiranja ne prinaša nobenih stroškovnih prednosti, se raje odločamo za manjše skupine – šolske razrede.

Večstopenjsko vzorčenje

Pri vzorčenju v skupinah največkrat izbiramo med obstoječimi »naravnimi« skupinami, ki so bile oblikovane v neke druge namene. Pogosto pa so take skupine prevelike, da bi jih lahko učinkovito uporabili pri vzorčenju. To največkrat rešujemo tako, da skupine nadalje delimo na podskupine. Pri tem praviloma uporabimo slučajnostno izbiro, s čimer seveda uvajamo dodatno stopnjo vzorčenja. Opraviti imamo torej z večstopenjskim vzorčenjem.

Obravnavajmo najprej dvostopenjski *(angl. two-stage – TS)* vzorec, v katerem izberemo *a* skupin s SRS vzorčenjem iz populacije *A* skupin. V drugem koraku, to je na drugi stopnji vzorčenja, pa v vsaki izbrani skupini s SRS izbiro določimo *b* izmed *B* elementov; skupno torej $n = ab$. Aritmetična sredina $\bar{y}_{ts}$ v dvostopenjskem (*TS*) vzorcu je zato:

$$\bar{y}_{ts} = \sum_{\alpha=1}^{a}\sum_{\beta=1}^{b}\frac{y_{\alpha\beta}}{n} = \sum_{\alpha=1}^{a}\frac{\bar{y}_{\alpha}}{a}$$

in je nepristranska cenilka za populacijsko aritmetično sredino. Seveda pa se pri dvostopenjskem vzorčenju $\bar{y}_{\alpha}$ vzorčna aritmetična sredina za skupino α:

$$\bar{y}_{\alpha} = \sum_{\beta=1}^{b}\frac{y_{\alpha\beta}}{b}$$

razlikuje od populacijske aritmetične sredine $\bar{Y}_{\alpha}$, ki je bila pri vzorčenju v skupinah izračunana na vseh elementih v skupini. Vzorčna varianca ima zaradi dodatnega variiranja $\bar{y}_{\alpha}$ okoli $\bar{Y}_{\alpha}$ v primerjavi z (12) zato dva člena:

$$V(\bar{y}_{ts}) = \left(1-\frac{a}{A}\right)\frac{S_a^2}{a} + \left(1-\frac{b}{B}\right)\frac{S_b^2}{ab},$$

kjer je S_a^2 varianca med skupinami, ki smo jo že predstavili (12a), povprečna varianca znotraj skupin S_b^2 pa je enaka:

$$S_b^2 = \sum_{\alpha=1}^{A}\sum_{\beta=1}^{B}\frac{(Y_{\alpha\beta}-\bar{Y}_{\alpha})^2}{A(B-1)}.$$

Prvi del v izrazu za varianco $V(\bar{y}_{ts})$ torej predstavlja – podobno kot pri vzorčenju v skupinah (12) – učinek skupin, drugi del pa vsebuje komponento druge (dodatne) stopnje zaradi vzorčenja znotraj izbranih skupin. Kadar je $b = B$, je dodatna varianca seveda enaka 0 in zgornji izraz se poenostavi v varianco aritmetične sredine vzorčenja v skupinah (12).

Kadar je $a = A$, so v vzorec vključene vse skupine, s čimer so skupine postale stratumi. Prvi člen v zgornjem izrazu za varianco je zaradi faktorja ($1 - a/A$) v takem primeru enak 0, drugi člen pa preide v varianco proporcionalno stratificiranega vzorca (10), kjer velja:

$$f = \frac{b}{B},\ n = ab,\ S_w^2 = S_b^2 .$$

Teoretični izračuni so pokazali, da je nepristranska cenilka za vzorčno varianco $V(\bar{y}_{ts})$ enaka:

$$v(\bar{y}_{ts}) = \left(1 - \frac{a}{A}\right)\frac{s_a^2}{a} + \frac{a}{A}\left(1 - \frac{b}{B}\right)\frac{s_b^2}{ab},$$

kjer je:

$$s_a^2 = \frac{\sum_{\alpha=1}^{a} (\bar{y}_\alpha - \bar{y}_{ts})^2}{a-1},$$

$$s_b^2 = \frac{\sum_{\alpha=1}^{a}\sum_{\beta=1}^{b} (y_{\alpha\beta} - \bar{y}_\alpha)^2}{a(b-1)}.$$

Zgornji izraz za varianco $v(\bar{y}_{ts})$ je razmeroma zapleten, saj bi morali zaradi s_b^2 računati elementarno varianco v vsaki izbrani skupini v vzorcu. Ker pa je na prvi stopnji vzorčni delež ($f_a = a/A$) običajno zanemarljivo majhen, se prvi člen zaradi:

$$\left(1 - \frac{a}{A}\right) \approx 1$$

poenostavi. Podobno lahko drugi člen $v(\bar{y}_{ts})$ v celoti izpustimo zaradi njegove majhnosti v primerjavi s prvim členom, kar izhaja iz faktorja:

$$\frac{a}{A} \ll 1.$$

V takem primeru dobimo približek:

$$v(\bar{y}_{ts}) \approx \frac{s_a^2}{a}, \qquad [15]$$

ki ga je mogoče tudi nadvse enostavno izračunati. Zgornja poenostavitev v resnici obravnava prvo stopnjo vzorčenja kot SRS vzorčenje s ponavljanjem in ne kot SRS vzorčenje brez ponavljanja. Ker pa je taka poenostavitev v praksi upravičena, se pogosto uporablja pri kompleksnih načrtih in tudi pri računalniških programih za izračunavanje vzorčne variance.

Končne skupine

K dvostopenjskemu vzorčenju lahko pristopimo tudi nekoliko drugače. V tem primeru najprej vsako od *A* skupin v populaciji razdelimo v *B/b* končnih skupin (*angl. ultimate clusters – UC*), to je skupin na zadnji stopnji izbire. Vse končne skupine v populaciji imajo zato enako velikost *b* elementov. Postopek v vsaki od *A* populacijskih skupin z *B* elementi poteka takole: izmed *B* elementov izberemo SRS vzorec velikosti *b*, nato pa iz preostalih (*B-b*) elementov nadaljnji SRS vzorec velikosti *b*, vse dokler v vsaki populacijski skupini ne oblikujemo *B/b* končnih skupin velikosti *b*. Zaradi konceptualne enostavnosti smo v tem primeru seveda predpostavili, da je *B/b* celo število.

Dvostopenjski vzorec v skupinah lahko v tem primeru obravnavamo kot neposredni SRS vzorec, pri katerem med skupno *AB/b* končnimi skupinami izberemo *a* skupin, v katere vključimo v vzorec vseh *b* elementov.

Opisani vzorčni načrt je povsem podoben siceršnjemu postopku dvostopenjskega vzorčenja v skupinah. Razlikuje se le v tem, da pri dvostopenjskih vzorcih izbiramo po eno končno skupino iz vsake vzorčene populacijske skupine, medtem ko pri modelu končnih skupin te omejitve ni in bi lahko v določeni skupini med *B* elementi izbrali dvakrat po *b* elementov. Ker pa je faktor *a/A* običajno majhen, je verjetnost, da bosta v vzorec izbrani dve končni skupini iz iste populacijske skupine, največkrat povsem zanemarljiva. Opisani način je zato dober približek dvostopenjskega vzorčenja v skupinah. Če je torej *a/A* majhen in ne upoštevamo faktorja *FPC*, dobimo siceršnjo poenostavitev za cenilko vzorčne variance in ocenjena varianca $V(\bar{y}_{ts})$ izhaja neposredno iz izraza (12):

$$v(\bar{y}_{ts}) = \left(1 - \frac{ab}{AB}\right)\frac{s_a^2}{a}. \qquad [16]$$

Pri tem je stopnja vzorčenja:

$$f = \frac{a}{\left(\frac{AB}{b}\right)} = \frac{ab}{AB} = \frac{n}{N}$$

vrednost s_a^2 pa se nanaša na varianco aritmetičnih sredin za vse izbrane končne skupine. Kadar je faktor *a/A* majhen, lahko *FPC* v izrazu (16) izpustimo in s tem dobimo varianco (15) pri vzorčenju s ponavljanjem.

Pri modelu končnih skupin izhaja vzorčni učinek $Deff(\bar{y}_{ts})$ za vzorčno aritmetično sredino iz izraza za vzorčni učinek $Deff(\bar{y}_c)$ pri vzorčenju v skupinah (14). Pri tem je ρ intraklasna korelacija v končnih skupinah, *b* pa velikost končnih skupin. Kadar so končne skupine izbrane s pomočjo SRS vzorčenja, lahko privzamemo, da je njihova homogenost v grobem enaka kot v celotni skupini. Zato je približek za $Deff(\bar{y}_{ts})$ pri dvostopenjskem vzorcu enak:

$$Deff(\bar{y}_{ts}) \approx 1 + (b-1)\rho . \qquad [17]$$

Kot kaže gornji izraz, vzorčni učinek $Deff(\bar{y}_{ts})$ pada, če se manjša velikost b končnih skupin, to je velikost izbranih podvzorcev v skupinah. Pri fiksni velikosti vzorca $n = ab$ torej manjšanje velikosti podvzorcev (vrednost b) – in hkratno večanje števila skupin (vrednost a) – izboljšuje natančnost ocen za aritmetično sredino. Seveda pa z večanjem števila skupin naraščajo tudi stroški anketiranja. Kadar pri anketni raziskavi z osebnim anketiranjem razpolagamo z vnaprej določenimi sredstvi, lahko pri večjem številu skupin – zaradi povečanih potnih stroškov anketarjev – uporabimo bistveno manjši vzorec.

Natančnost in stroški

Da bi dosegli optimalno razmerje števila skupin a in števila elementov b, ki smo jih izbrali znotraj populacijske skupine velikosti B, moramo uravnotežiti natančnost in stroške. Za to seveda potrebujemo predhoden model stroškov raziskave. Preprost model strukture stroškov raziskave je lahko naslednji:

$$C = aC_a + nc,$$

kjer so C skupni stroški, c stroški za posamezen element, ki smo ga izbrali v raziskavo (npr. variabilni stroški anketarja za eno anketo), C_a pa stroški na izbrano skupino (npr. šolanje anketarja, potovanje anketarja na mesto, kjer je skupina). Na tej podlagi je optimalna izbira za vrednost b, ki pri danih stroških minimizira varianco vzorčne aritmetične sredine, enaka (Kish, 1965: 8.3B):

$$b_{OPT} = \sqrt{\frac{C_a}{c}\frac{(1-\rho)}{\rho}} . \qquad [18]$$

Ob predpostavki, da so drugi dejavniki nespremenjeni, iz zgornjega izraza izhaja, da je optimalni vzorec bolj razpršen (b_{opt} je manjši in a_{opt} je večji) pri večji homogenosti skupin (večji ρ), pri večjih stroških na element anketiranja (c) ter pri manjših stroških na skupino (C_a). Če imamo, denimo, relativno razmerje stroškov $C_a/c = 17$ in $\rho = 0.07$, sledi $b_{opt} = 15$. Na tej podlagi lahko pri razpoložljivih sredstvih določimo tudi optimalno število skupin a_{opt} in celotno velikost vzorca n.

Stroškovni model v izrazu (18) je lahko v določenih primerih preveč preprost, vendar običajno zadostuje za grobo oceno optimalne velikosti. Seveda lahko uporabljamo tudi bolj kompleksne modele, vendar je vprašljivo, ali je dodatna kompleksnost modela tudi uporabna. Ocenjevanje stroškov se namreč izkaže za izredno težavno celo pri najpreprostejših modelih. Poleg ocene stroškov pa zahteva izračunavanje b_{opt} tudi oceno za ρ, kar lahko dobimo iz prejšnjih raziskav, ki so proučevale enake ali podobne spremenljivke.

Ker pa so anketne raziskave običajno večnamenske, ima lahko vsaka spremenljivka drugačno vrednost b_{opt}. V takem primeru je treba doseči določen operativen kompromis, ki lahko temelji tudi na bolj zapletenih odločitvenih modelih, ki upoštevajo pomembnost posameznih spremenljivk.

Večstopenjsko vzorčenje torej uporabljamo predvsem zaradi stroškovne učinkovitosti, kar je najbolj aktualno pri osebnem anketiranju. Prihranek v stroških je posebej velik pri prostorskih vzorcih, pri katerih izdelujemo sezname stanovanj samo v enotah zadnje stopnje (npr. popisni okoliš, mestna četrt ipd.). Pri zbiranju podatkov z osebnim anketiranjem zato večstopenjsko vzorčenje omogoča izjemne prihranke. Če je populacija velika in razpršena, je namreč enostopenjski vzorec elementov neprijetno razpršen, medtem ko večstopenjski vzorec koncentrira intervjuje v manjše število lokacij. Pri telefonskih in poštnih anketah pa tovrstno vzorčenje v splošnem ne omogoča opaznejših prihrankov pri zbiranju podatkov, čeprav obstajajo določeni posebni primeri, kjer je ugodno vzorčiti skupine elementov.

Denimo, da v nekem mestu izvajamo raziskavo z osebnim anketiranjem. Če je mesto majhno in imamo na voljo seznam hiš ali stanovanj, potem je primerno izbrati enostopenjski stratificirani vzorec. Če pa je mesto majhno in nimamo seznama, je primernejši dvostopenjski vzorec, s čimer prihranimo stroške izdelave celotnega vzorčnega okvira. V takem primeru bi na prvi stopnji izbrali stratificirani vzorec sosesk oziroma popisnih okolišev, nakar bi v izbranih soseskah izbrali stanovanja. V velikem mestu je – tudi če imamo na voljo seznam stanovanj – v vsakem primeru bolje izbrati dvostopenjsko vzorčenje, ki zmanjša stroške potovanja anketarjev in prihrani njihov čas.

Pri večjih populacijah se lahko število stopenj vzorčenja nadalje poveča. Nacionalni vzorci z osebnim anketiranjem v ZDA običajno uporabljajo za raziskave splošne populacije vzorčenje v najmanj treh stopnjah, kar bomo podrobneje obravnavali v poglavju o prostorskih vzorcih.

Do zdaj smo vedno predpostavljali, da so bile skupine in elementi zaradi enostavnosti izbrani s pomočjo SRS vzorčenja. V praksi pa so lahko skupine in tudi elementi izbrani s stratifikacijo ali sistematičnim vzorčenjem. Izkaže se, da je stratifikacija celo pomembnejša pri vzorčenju v skupinah kot pri vzorčenju elementov, saj omogoča večje izboljšave v natančnosti. Pri stratifikaciji enot prve stopnje *(*PSU – *primary sampling units)* namreč pogosto uporabimo toliko stratumov, kolikor je enot prve stopnje. Včasih pa stratificirani vzorec PSU dodatno izboljšujemo še s tehniko nadzorovanega izbora (*angl. controlled selection*), ki je opisana v Goodman et al. (1950) in Hess et al. (1975).

Če iz stratuma izberemo samo po eno vzorčno enoto prve stopnje (PSU), potem variance v takem stratumu seveda ne moremo oceniti. Da bi kljub temu omogočili oceno vzorčnih napak, pogosto kombiniramo pare podobnih stratumov in jih obravnavamo kot en sam stratum. Opisani postopek združevanja stratumov (*angl. collapsed strata*) sicer povzroča določeno precenjevanje vzorčnih napak, ki pa ni kritično.

V posebnem primeru, ko je v vsakem stratumu le po ena enota prve stopnje in smo v vsakem na novo združenem stratumu torej dobili po dve enoti prve stopnje, se število stratumov pri tem seveda prepolovi. Pogosteje pa že v osnovnem načrtu stratifikacije predvidimo po dve enoti v vsakem stratumu, kar imenujemo stratifikacijsko izbiranje parov (*angl. paired selection design).* Pri tem enote prve stopnje (PSU) v združenih stratumih običajno obravnavamo, kot da smo jih vzorčili s ponavljanjem, zato uporabimo preproste cenilke variance, ki smo jih že opisali.

6. SORAZMERNO VZORČENJE

Do zdaj smo pri vzorčenju v skupinah predpostavljali, da so vse skupine enake velikosti, kar pa je v praksi izredno redko. Elementi se namreč pogosto povezujejo v skupine različnih velikosti. Tako npr. šolski razredi ne bodo imeli vedno točno 24 dijakov, ampak verjetneje med 20 in 30, še bolj pa se bodo razlikovale ulice v mestu glede na število stavb ali število prebivalcev. V nadaljevanju bomo zato nekoliko podrobneje obravnavali težave, ki nastajajo zaradi tovrstnega variiranja pri vzorčenju v skupinah.

Za lažji prikaz bomo uporabili primer, ki vključuje EPSEM izbiro stanovanj (elementov) iz 9 stanovanjskih blokov – enot prve stopnje (PSU) – ki imajo skupno 315 stanovanj. Pri tem želimo v vzorec izbrati 21 stanovanj, in sicer na prvi stopnji 3 bloke, nato pa v teh treh blokih naključno izberemo 3×7=21 stanovanj. Vzorčni delež je torej 1/15. Na začetku predpostavimo, da velikosti blokov natančno poznamo. Označimo jih z B_α.

Tabela 4: Populacijska velikost izbranih blokov B_α

Blok	1	2	3	4	5	6	7	8	9	Skupaj
B_α	20	100	50	15	18	43	20	36	13	315

Če vzorčni načrt na prvi stopnji zahteva SRS izbiro treh izmed devetih stanovanjskih blokov, ima vsak blok verjetnost 1/3, da bo izbran v vzorec. Na splošno je pri dvostopenjski izbiri verjetnost, da se pojavi v vzorcu element β iz skupine α, enaka:

$$P(\alpha\beta) = P(\alpha)\times P(\beta|\alpha) \qquad [19]$$

Pri tem je $P(\alpha)$ verjetnost, da bo izbrana skupina α, $P(\beta|\alpha)$ pa verjetnost, da bo na drugi stopnji izbran element β v skupini α. Enačbo seveda lahko razširimo tako, da pokrije več stopenj vzorčenja. Zgornji zapis običajno imenujemo enačba izbora (*angl. selection equation*).

V našem primeru zahtevamo EPSEM vzorčni načrt z vzorčnim deležem:

$$f = P(\alpha\beta) = 1/15\,.$$

Ker imajo izbrane skupine enake verjetnosti $P(\alpha) = 1/3$, iz izraza (19) sledi, da je $P(\beta|\alpha) = 1/5$. Z drugimi besedami, vzorčni delež v vsakem izbranem bloku mora biti enak 1/5. Oglejmo si nekaj vzorcev, ki jih dobimo iz takega načrta. V prvem skrajnem primeru bi lahko izbrali ravno najmanjše bloke (št. 4, 5 in 9), v drugem pa bi lahko izbrali tri največje bloke (št. 2, 3 in 6). V prvem primeru bi nato z vzorčnim deležem 1/5 dobili 9, v drugem pa 39 stanovanj. V povprečju vseh vzorcev bi sicer dobili 21 izbranih stanovanj v treh blokih, vendar bi se pri posameznem vzorcu velikost največkrat bistveno razlikovala od pričakovane vrednosti.

Velika variabilnost v končni velikosti vzorca je deloma posledica majhnega števila izbranih skupin. Vendar pa bi končna velikost vzorca variirala tudi pri večjem številu izbranih skupin, če bi obstajale razlike pri njihovi velikosti. Očitno bi torej potrebovali večji nadzor velikosti končnega vzorca. Čeprav le redko potrebujemo povsem natančno velikost vzorca, jo moramo vseeno obdržati znotraj zmernih meja.

Ena od možnosti za zmanjšanje variabilnosti je stratifikacija skupin po velikosti. V tem primeru bi bloke razdelili v tri stratume po velikosti: prvi stratum bi vključeval bloke številka 2, 3 in 6, drugi bloke 1, 7 in 8, tretji pa bloke 4, 5 in 9. Izbira po enega bloka iz posameznega stratuma bi zmanjšala variabilnost v končni velikosti vzorca od najmanj 15 (bloki 6, 7 in 9) pa do največ 31 elementov (bloki 2, 5 in 8). Stratifikacija po velikosti skupin nam torej lahko prinese bistveno večji nadzor nad končno velikostjo vzorca, vendar pa s tem omejimo uporabo stratifikacije po drugih faktorjih. Zato raje uporabljamo alternativni način nadzora nad velikostjo vzorca, ki ga bomo podrobneje predstavili v nadaljevanju.

Vzorčenje PPS

Navedimo najprej pogoje, ki bi jih radi izpolnili:
(1) vzorec naj bo EPSEM,
(2) vzorec naj bo omejen na 3 bloke,
(3) velikost vzorca naj bo vedno n = 21 ne glede na izbor blokov.

Prvi in tretji pogoj pomenita, da je enačba izbora enaka $f = P(\alpha\beta) = 1/15$, drugi in tretji pogoj pa izpolnimo tako, da v treh izbranih blokih vsakič vključimo v vzorec po 7 stanovanj. Pri takem nastavku je verjetnost izbora na drugi stopnji vzorčenega bloka α enaka:

$$P(\beta \mid \alpha) = \frac{7}{B_\alpha}.$$

Če v enačbo (19) vstavimo $P(\alpha\beta) = 1/15$ in $P(\beta|\alpha) = 7/B_\alpha$, dobimo:

$$\frac{1}{15} = P(\alpha)\frac{7}{B_\alpha},$$

od koder izhaja, da je verjetnost izbire bloka α enaka:

$$P(\alpha) = \frac{B_\alpha}{105}.$$

Če torej izpolnimo vse tri pogoje, bloke izbiramo z verjetnostjo, ki je sorazmerna z njihovo velikostjo B_α (*angl. probability proportional to size – PPS*), kar krajše označujemo kot »PPS izbira«.

Splošna enačba izbora za dvostopenjski vzorec PPS je zato enaka:

$$P(\alpha\beta) = f = \frac{n}{N} = \left(\frac{aB_\alpha}{\sum B_\alpha}\right) \times \left(\frac{b}{B_\alpha}\right), \quad [20]$$

kjer smo a enot prve stopnje (PSU) izbrali s postopkom PPS, b elementov pa smo iz vsake PSU izbrali s SRS izbiro. Pri tem seveda velja:

$$n = ab, \ N = \sum_\alpha B_\alpha .$$

Enačbo lahko razširimo tudi na trostopenjski model, pri katerem a predstavlja število enot prve stopnje, b število enot druge stopnje, c pa število elementov, ki jih izberemo na zadnji stopnji. Enačbo izbora lahko v tem primeru zapišemo kot:

$$P(\alpha\beta\gamma) = f = \frac{n}{N} = \left(\frac{aB_\alpha}{\sum B_\alpha}\right) \times \left(\frac{bB_{\alpha\beta}}{B_\alpha}\right) \times \left(\frac{c}{B_{\alpha\beta}}\right).$$

Tokrat je velikost vzorca seveda produkt $n = abc$, vrednost $B_{\alpha\beta}$ pa pomeni velikost enote druge stopnje (*angl. secondary sampling unit – SSU*) z oznako β znotraj enote prve stopnje (PSU) z oznako α. Seveda velja:

$$N = \sum_\alpha B_\alpha = \sum_\alpha \sum_\beta B_{\alpha\beta} .$$

Izbiro PPS lahko razmeroma enostavno dosežemo s pomočjo kumulativne vsote velikosti blokov.

Tabela 5: Izračun kumulative za PPS izbiro blokov B_α

Blok	1	2	3	4	5	6	7	8	9
B_α	20	100	50	15	18	43	20	36	13
kumulativa	20	120	170	185	203	246	266	302	315

V kumulativni vsoti pripada vsakemu bloku niz števil:
- bloku 1 ustrezajo števila od 001 do 020;
- bloku 2 pripada 100 števil od 021 do 120;
- bloku 3 pa pripada 50 števil, ki segajo od 121 do 170, in tako naprej.

Tako ima vsak blok število elementov, ki ustreza njegovi velikosti B_α. Slučajno izbrano število med 001 in 315 zato določa izbrani blok z metodo PPS. Če bi npr. izbrali število 197, bi v vzorec vključili blok številka 5, saj ga določajo števila med 186 in 203.

Izbrati bi torej morali 3 slučajna števila, da bi s PPS izbiro določili 3 bloke. Pri takem postopku je neugodno le to, da ima vsak blok možnost, da je izbran večkrat, zato za izbiro PPS vzorca brez ponavljanja raje uporabljamo sistematično vzorčenje.

PPS izbiro sistematično izvedemo tako, da skupno velikost vzorca ($n = 315$) razdelimo glede na število enot prve stopnje ($a = 3$). S tem dobimo v našem primeru vzorčni interval velikosti $n/a = 105$. Izberemo slučajno število med 000 in 105, denimo 047, ki določa blok številka 2. Nato dodamo vzorčni interval (105) in dobimo število 152, ki določa blok številka 3. Če še enkrat prištejemo vzorčni interval, dobimo število 257 in s tem blok številka 7.

S PPS vzorčnim načrtom lahko uporabimo tudi metodo »končnih skupin«, ki smo jo že opisali. Predpostavimo, da na drugi stopnji PPS vzorčnega načrta s SRS vzorčenjem izberemo b elementov iz vsake PSU. Potem bi morali – da bi dobili načrt končnih skupin – v vsaki PSU v populaciji določiti končne skupine. Zato – tako kot prej s SRS vzorcem – izberemo b elementov, ki predstavljajo prvo končno skupino. Nato med preostalimi spet izberemo b elementov za drugo končno skupino itd. V tem primeru je vsaka enota prve stopnje α z B_α elementi razdeljena v B_α / b končnih skupin. Tudi tukaj zaradi enostavnosti privzamemo, da je B_α / b celo število. Očitno je torej SRS vzorec končnih skupin enak PPS vzorcu brez ponavljanja. Metoda končnih skupin bi sicer lahko izbrala več kot eno končno skupino iz izbrane PSU, vendar je to s PPS vzorčnim načrtom malo verjetno. Ker je verjetnost izbire dveh končnih skupin iz ene PSU majhna, je razlika med vzorčnima načrtoma povsem zanemarljiva. Podobnost med načrtoma je še posebej opazna, ker je verjetnost za izbor končne skupine v določeno PSU sorazmerna številu končnih skupin v tej PSU, to je B_α / b. Število končnih skupin B_α / b je pri tem sorazmerno z velikostjo PSU.

Ker ima opisani PPS načrt lastnost EPSEM, je velikost končnega vzorca povsem določena in s tem konstanta n. Enostavna vzorčna aritmetična sredina je zato nepristranska cenilka populacijske aritmetične sredine:

$$\bar{y}_p = \sum_\alpha \sum_\beta \frac{y_{\alpha\beta}}{n} = \sum_\alpha \frac{\bar{y}_\alpha}{a}.$$

Uporaba metode končnih skupin daje cenilko variance $V(\bar{y}_p)$ iz izraza (16), pri čemer zanemarimo faktor končne populacije (*FPC*):

$$v(\bar{y}_p) \approx \frac{s_a^2}{a}. \qquad [21]$$

Približek za vzorčni učinek $Deff(\bar{y}_p)$ vzorčne aritmetične sredine pri PPS izbiri na prvi stopnji ter SRS izbiri na drugi stopnji lahko izračunamo po že znani formuli (17):

$$Deff(\bar{y}_p) \approx 1+(b-1)\rho .$$

Vzorčenje PPES

V praksi je opisano PPS vzorčenje redko izvedljivo, saj so velikosti vzorčnih enot B_α običajno neznane. Pogosto pa lahko dobimo določene ocene velikosti, ki jih nato uporabimo namesto pravih velikosti pri PPS izbiri. Pri tem moramo seveda razlikovati med uporabo resničnih velikosti skupin in uporabo ocenjenih velikosti. Tako bomo pojem PPS uporabili le takrat, ko razpolagamo z dejansko velikostjo B_α, dodatno pa bomo uvedli poseben izraz za izbiro, ki je sorazmerna z ocenjeno velikostjo (*angl. probability proportional to estimated size* – PPES). Tovrstno izbiro bomo torej imenovali »PPES izbira«. Kadar pa bomo namesto pravih B_α uporabljali ocenjene velikosti, jih bomo označevali z M_α.

Tako kot v izrazu (20) lahko tudi enačbo izbora za PPES zapišemo kot:

$$P(\alpha\beta) = f = \frac{aM_\alpha}{\sum M_\alpha} \times \frac{b}{M_\alpha} . \qquad [22]$$

Da bi imel vzorec s PPES izbiro lastnost EPSEM, moramo na drugi stopnji zagotoviti ustrezen vzorčni delež, ki je v tem primeru enak b/M_α. V enoti (PSU) z B_α elementi je zato pričakovana velikost vzorca v splošnem enaka:

$$b \times \frac{B_\alpha}{M_\alpha} .$$

Pričakovana velikost vzorca bo variirala med enotami prve stopnje in bo enaka želeni velikosti b, samo kadar bosta prava in ocenjena velikost enaki, $B_\alpha = M_\alpha$.

Denimo, da v našem primeru ne poznamo resničnih velikosti (B_α) devetih blokov, ampak le ocene M_α, ki jih uporabimo v PPES izbiri. Ocene so navedene v spodnji tabeli. Pričakovano velikost vzorca bi dobili tako, da pomnožimo vzorčni delež $7/M_\alpha$ s pravo velikostjo skupin B_α. Pri PPES izbiri zato dobimo pričakovane velikosti, ki se razlikujejo od vrednosti $b = 7$.

Tabela 6: Populacijska (B_α) in ocenjena velikost blokov (M_α)

Blok	1	2	3	4	5	6	7	8	9
B_α	20	100	50	15	18	43	20	36	13
M_α	30	110	50	20	20	50	10	50	20
bB_α/M_α	4.7	6.4	7.0	5.3	6.3	6.0	14.0	5.0	4.6

Pričakovana velikost vzorca variira zaradi nenatančnosti ocen M_α, vendar so odstopanja sprejemljiva. Večje odstopanje opazimo le pri PSU številka 7, kjer smo pravo velikost $B_\alpha = 20$ z oceno $M_\alpha = 10$ močno podcenjevali. Pri določanju ocenjenih velikosti moramo zato paziti, da se takemu neskladju – če je le mogoče – izognemo. Zgodi se namreč lahko, da smo imeli ob zadnjem popisu mestno četrt z 10 stanovanji, ki je narasla na npr. 800 stanovanj.

V zgornji tabeli lahko opazimo, da je v večini PSU pričakovana velikost vzorca manjša od želene vrednosti $b = 7$, saj v skoraj vseh PSU ocenjena vrednost M_α precenjuje B_α. Imamo namreč $\sum M_\alpha = 360$ v primerjavi z $\sum B_\alpha = 315$. Pri načrtovani velikosti vzorca $n = 21$ je vzorčni delež zato enak 21/360, pričakovana velikost vzorca na podlagi ocenjenih vrednosti pa je samo $(21/360)\times 315 = 18.4$, kar dodatno opozarja, da morajo biti ocenjene velikosti čim natančnejše.

Razmernostna cenilka

Posledica PPES vzorčenja je tudi to, da velikost vzorca ni konstanta, ampak slučajna spremenljivka, ki je odvisna od izbire enot prve stopnje. Zato je primerno, da variabilno velikost vzorca označimo z x in ne več z n. Vzorčna aritmetična sredina je zato razmerje:

$$r = \frac{y}{x},$$

kjer je y oznaka za vsoto spremenljivke y v vzorcu, x pa tokrat pomeni velikost vzorca, saj so velikosti skupin v vzorcu označene z x_i. Imamo torej:

$$y = \sum_{i=1}^{n} y_i \text{ in } x = \sum_{i=1}^{n} x_i \,. \qquad [22a]$$

Zgoraj opredeljeno aritmetično sredino imenujemo razmernostna aritmetična sredina (*angl. ratio mean)* ali razmernostna cenilka (*angl. ratio estimator*), ker je izračunana kot razmerje dveh slučajnih spremenljivk. Razmernostna cenilka ni nepristranska cenilka populacijske aritmetične sredine. Njena pristranskost pa je zanemarljiva, če je variabilnost v velikosti vzorca x dovolj majhna. Pristranskosti običajno ni treba upoštevati, če je koeficient variacije x manjši od 0.1, kar pomeni, da za razmerje standardne napake spremenljivke x in pričakovane vrednosti $E(x)$ za velikost vzorca x velja:

$$CV(x) = \frac{SE(x)}{E(x)} < 0.1.$$

Ponovno velja opozoriti, da se oznaka x nanaša na zgoraj definirano velikost vzorca in ne morda na vrednost x_i spremenljivke x na elementih vzorca. Izračun variance za razmernostno cenilko je bolj zapleten, ker je v imenovalcu slučajna spremenljivka. Dober približek je mogoče izdelati le za velike vzorce, in sicer na podlagi Taylorjeve vrste. Približek lahko uporabimo le pod pogojem, da je koeficient variacije x manjši od 0.2 ali – še bolje – manjši od 0.1. Splošna oblika približka za oceno variance razmernostne cenilke r je:

$$v(r) \doteq \frac{v(y) + r^2 v(x) - 2rc(x,y)}{x^2}, \qquad [23]$$

kjer je $c(x,y)$ vzorčna kovarianca slučajnih spremenljivk: vsote za spremenljivko x in za spremenljivko y. Oglejmo si nekoliko podrobneje izraze za $v(y)$, $v(x)$ in $c(x,y)$. Pri tem naj bo vzorčni načrt stratificiran EPSEM vzorec. V primeru brez stratifikacije pa lahko v spodnjih izrazih enostavno izpustimo seštevanje po stratumih. Uporabili bomo naslednje oznake:

$y_{h\alpha}$ vsota spremenljivke y v vzorcu za enoto (PSU) α v stratumu h,
$x_{h\alpha}$ velikost vzorca za enoto (PSU) α v stratumu h,
y_h vsota spremenljivke y v vzorcu za vseh a_h enot (PSU) v stratumu h,
y_h velikost vzorca v stratumu h.

Imamo naslednje ocene:

$$v(y) \doteq \sum_h a_h s_{yh}^2,$$

$$v(x) = \sum_h a_h s_{xh}^2,$$

$$c(x,y) = \sum_h a_h s_{xyh},$$

kjer je:

$$s_{yh}^2 = \frac{1}{(a_h - 1)} \sum_\alpha \left[y_{h\alpha} - \left(\frac{y_h}{a_h} \right) \right]^2,$$

$$s_{xh}^2 = \frac{1}{(a_h - 1)} \sum_\alpha \left[x_{h\alpha} - \left(\frac{x_h}{a_h} \right) \right]^2,$$

$$s_{xyh} = \frac{1}{(a_h - 1)} \sum_\alpha \left[x_{h\alpha} - \left(\frac{x_h}{a_h} \right) \right] \times \left[y_{h\alpha} - \left(\frac{y_h}{a_h} \right) \right].$$

Splošni izraz (23) za varianco razmernostne cenilke $v(r)$ zasluži posebno pozornost. Velja namreč za vsak EPSEM vzorčni načrt ne glede na verjetnosti, ki se uporabljajo za izbiro PSU, in ne glede na to, kako smo izbrali vzorec na drugi stopnji, to je znotraj izbranih PSU. Uporabimo ga seveda lahko tudi pri vzorcih, pri katerih enote prve stopnje (PSU) niso stratificirane (imamo torej samo en stratum), in tudi za vzorce, pri katerih se velikost vzorca ne spreminja in je torej $v(x) = 0$ ter $c(x, y) = 0$. Takrat izraz (23) preide v (15). Edina omejitev za uporabo izraza (23) je – poleg veljavnosti predpostavke o vzorčenju s ponavljanjem – zahteva, da je koeficient variacije spremenljivke x (velikosti vzorca) majhen, npr. pod 0.2. Z določenimi prilagoditvami za M_α in $\sum M_\alpha$ pa lahko izraz (23) uporabimo tudi za verjetnostne vzorce, ki nimajo EPSEM vzorčnega načrta (Kish, 1965: poglavje 6).

Problemi sorazmernih vzorcev

Preden končamo obravnavo PPS in PPES vzorčenja, je treba opozoriti še na nekaj tipičnih težav pri takšnem vzorčenju. Za ponazoritev bomo uporabili prejšnji primer vzorčenja stanovanj v blokih. Tokrat izbiramo tri bloke med desetimi s pomočjo ocenjenih velikosti M_α s skupno vsoto $\sum M_\alpha = 315$.

Tabela 7: Primer prevelike in prazne ocene za velikost blokov (M_α)

Blok	1	2	3	4	5	6	7	8	9	10
M_α	20	120	45	15	18	43	5	0	36	13

Tako kot v prejšnjem primeru imamo pričakovan vzorec velikosti 21 in tudi podvzorec na drugi stopnji naj ima pričakovano velikost $b = 7$. Pri tovrstni izbiri vzorca se pojavita dve tipični težavi.

Prva nevarnost je, da z vzorčnim intervalom velikosti 315/3=105 blok 2 izberemo dvakrat, saj je velikost bloka ($M_2 = 120$) večja od velikosti intervala. Blok 2 se bo zato gotovo pojavil v vzorcu, obstaja pa tudi verjetnost, da se pojavi dvakrat, in sicer v primeru, ko je slučajni začetek med 21 in 35. Ena rešitev je, da v takem primeru v vzorec vključimo dva različna podvzorca iz izbranega bloka. Druga rešitev pa je, da celotni blok 2 označimo za stratum in ga izločimo iz PPES izbire enot prve stopnje. Elemente v tem stratumu vzorčimo z vzorčnim intervalom 1/15, iz preostalih blokov pa izberemo le preostali dve enoti s pomočjo PPES, prav tako s skupnim vzorčnim intervalom 1/15. Seveda tokrat potrebujemo manjše vrednosti b, saj velja:

$$\sum_{\alpha \neq 2} M_\alpha = 195,$$

od koder sledi:

$$\frac{2b}{195} = \frac{1}{15}.$$

Imamo torej vrednost $b = 6.5$. Primeri, ko je določena PSU tako velika, da ima možnost večkratnega izbora, pa so v praksi nadvse pogosti. Največkrat z njimi ravnamo kot z ločenimi stratumi, čeprav se še vedno (nekoliko zavajajoče) obravnavajo kot samostojne PSU (*angl. self-representing* PSU).

Prevelike PSU imajo torej v enačbi izbora verjetnost za vključitev v vzorec večjo od 1. Zapišimo to verjetnost za blok številka 2:

$$\frac{aM_\alpha}{\sum M_\alpha} = \frac{3 \times 120}{315} = \frac{360}{315} > 1.$$

Druga nadvse pogosta nevšečnost PPES izbire so premajhne PSU. Blok številka 7 tako sploh ne vsebuje dovolj stanovanj, da bi jih lahko v vzorec izbrali 6 ali 7 ($b = 6.5$). Preprost način rešitve te težave je združitev s sosednjo enoto, pri čemer oba bloka obravnavamo kot en blok (PSU). Združevanje opravimo pred izborom vzorca, lahko pa tudi kasneje, če uporabimo pravila povezovanja (*angl. linking procedures*), kar je podrobno opisano v Kish (1965: 244–245). Če je takšnih PSU veliko in bi povezovanje povzročalo težave med anketiranjem, jih lahko postavimo v poseben stratum in vzorčimo drugače. Pogosto pa je najmanjša velikost PSU že vnaprej določena, tako da je precej večja od b, s čimer se izognemo vsem težavam zaradi izbora vseh ali skoraj vseh enot v skupini. Pri $b = 6.5$ bi bila primerna minimalna velikost PSU okoli 13, saj le taka velikost zagotavlja, da vzorčni delež ne presega $f = 1/2$.

Na koncu je treba opozoriti še na blok številka 8 z ocenjeno velikostjo $M_\alpha = 0$, saj tak blok sploh nima možnosti, da bi bil izbran v vzorec. Ker je M_α samo ocenjena velikost, je seveda možno, da blok 8 kljub temu vsebuje nekaj stanovanj, če smo pri oceni M_α uporabili zastarele podatke. Torej je tudi v tem primeru priporočljivo povezati blok s sosednjim, saj mu s tem dajemo možnost, da ga izberemo v vzorec. To še posebej velja za prostorske vzorce, pri katerih vsakemu koščku naseljene površine dajemo možnost za izbor v vzorec.

7. DRUGI VERJETNOSTNI NAČRTI

Kombinacija vzorčnih načrtov, o katerih smo govorili v prejšnjih poglavjih, zadošča za obravnavo večine vzorčnih problemov. Obstajajo pa še tri vzorčne tehnike, ki jih pogosto uporabljamo v posebnih primerih: dvofazno vzorčenje (*angl. two-phase sampling*), replicirano vzorčenje (*angl. replicated sampling*) in vzorčenje v panelnih raziskavah (*angl. panel research*).

Dvofazno vzorčenje

Pri dvofaznem vzorčenju v prvi fazi najprej izberemo elemente v začetni vzorec, v drugi fazi pa v podvzorec izberemo le določene elemente iz prve faze. Metodo lahko razširimo v več faz, vendar običajno zadostujeta dve. Tipičen primer je enkratno naročilo večjega vzorca telefonskih številk iz telefonskega imenika, kjer v drugi fazi izbiramo manjše vzorce za posamezne anketne raziskave.

Dvofazno vzorčenje uporabljamo tudi takrat, ko za različne cenilke potrebujemo različne ocene natančnosti in tudi različne velikosti vzorcev. V tem primeru lahko podatke, katerih ocene morajo biti izračunane na večjem vzorcu, dobimo v prvi fazi, preostale ocene pa pridobimo na manjšem (in cenejšem) vzorcu v drugi fazi. Ne le, da s pomočjo dvofaznega postopka zmanjšamo stroške in čas zbiranja podatkov, pomembno lahko zmanjšamo tudi breme anketirancev. Znan primer uporabe dvofaznega vzorca je popis prebivalcev v ZDA, kjer so samo glavne demografske značilnosti zbrane na celotni populaciji, dodatne spremenljivke pa so zbrane v drugi fazi na podvzorcih iz te populacije.

Dvofazni vzorci nastopajo tudi pri uporabi populacijskih podatkov za izdelavo učinkovitega vzorčnega načrta, kadar so stroški zbiranja podatkov na celi populaciji preveliki. V takih primerih je pogosto cenejše, če podatke zberemo v prvi fazi na velikem vzorcu in nato šele v drugi fazi anketiramo ustrezne elemente. Prvo fazo torej uporabimo samo za pridobitev stratifikacijske informacije, za ocenjevanje velikosti enot pri izvedbi PPES vzorčenja ter za oceno stroškov vzorčnega načrta druge faze. Pri ocenjevanju učinkovitosti dvofaznega vzorčnega načrta moramo seveda vnaprej poznati stroške izvedbe obeh faz, pri čemer je vzorec v drugi fazi nujno manjši od vzorca v prvi fazi. Zato je dvofazni vzorčni načrt najučinkovitejši, če so stroški v prvi fazi bistveno manjši od stroškov v drugi fazi. Velike razlike pa lahko nastanejo le, če uporabimo različne načine zbiranja podatkov: v prvi fazi npr. podatke zberemo s poštno oziroma telefonsko anketo – ali pa jih celo pridobimo iz že obstoječih podatkovnih zbirk – v drugi fazi pa uporabimo osebno anketiranje ali celo zahtevnejše tehnike, ki so pogoste v medicinskih raziskavah.

Dvofazno vzorčenje se pogosto uporablja pri vzorčenju redkih populacij, za katere ni ločenega vzorčnega okvira. Take skupine so npr. družine z majhnimi otroki, osebe z določenimi boleznimi, vojni veterani, manjšine ipd. Vzorčni načrt za redke populacije pomeni pri načrtovanju vzorcev nasploh enega največjih izzivov (Kish, 1965: 11.4). Pogosto tako v prvi fazi z nizkimi stroški identificiramo člane redkih skupin, anketiramo pa jih šele v drugi fazi, kar največkrat pomeni dvofazno stratificirano vzorčenje. Elemente vzorca prve faze tako razdelimo v dva ali več stratumov glede na to, ali so člani redke populacije ali ne, stratume pa nato vzorčimo disproporcionalno. Če je v prvi fazi identifikacija redke populacije brez napak, je vzorčni delež med člani redke populacije lahko 1, med nečlani pa 0. Če pa pri identifikaciji obstaja možnost napak, izberemo v stratumu nečlanov vzorčni delež večji od 0, saj le tako dobijo tudi napačno razvrščeni, vendar ustrezni elementi, neničelno verjetnost za izbor v vzorec. Kadar pregledovanje v prvi fazi ni brez napak, je vsekakor bolje imeti nekaj neustreznih elementov, kot pa izgubiti napačno identificirane, vendar ustrezne elemente.

Kot primer za dvofazno vzorčenje redke populacije lahko navedemo študijo otrok z izgubo sluha. V prvi fazi bi otroke identificirali po gospodinjstvih s poceni in enostavnimi preizkusi, da bi v drugo fazo prišli prav vsi otroci s potencialno izgubo sluha. Šele v drugi fazi bi na izbranih otrocih iz prve faze opravili dražje in bolj zapletene preizkuse v laboratorijskem okolju.

Kot poseben primer uporabe dvofaznega vzorčenja lahko navedemo tudi raziskavo političnega razpoloženja med volivci v enem od evropskih mest, kjer je bil v vzorčnem okviru na voljo seznam imen, urejen po abecednem redu. Ker je bilo mesto veliko in so anketo izvajali z osebnim anketiranjem, je bilo zaradi nižanja stroškov treba vzorčiti v skupinah. Skupine bi sicer lahko določili tudi iz obstoječega seznama, vendar bi bilo to drago in zamudno. Zato so izbrali desetkrat večji vzorec in njegove elemente razmestili v skupine podobnih velikosti, ki so temeljile na geografski bližini. V drugi fazi so izbrali približno desetino teh skupin, ki so predstavljale tudi končni vzorec.

Vzorčenje z replikacijami

Pri vzorčenju z replikacijami je celoten vzorec sestavljen iz niza podvzorcev, ki so narejeni na enak način. Vzorec torej obravnavamo kot zaporedje ponovitev povsem enakih podvzorcev. Pri tem je bistveno, da vsak podvzorec generira neodvisno oceno za populacijski parameter. Na tej osnovi izračunamo tudi skupno oceno. Replikacije pogosto uporabljamo tudi za študije nevzorčnih napak, kot je npr. variabilnost rezultatov, dobljenih z različnimi anketarji ali osebami, ki so podatke kodirali. Posebej pogosto pa uporabljamo replikacije tudi za ocenjevanje standardnih napak.

Kot preprost primer replikacij navedimo študijo o vplivu anketarja. Denimo, da potrebujemo SRS vzorec velikosti n = 1,000 elementov in da terensko delo opravlja skupina 20 anketarjev. Pri običajnem vzorcu bi vzorec porazdelili med anketarje glede na splošno usposobljenost anketarja in geografsko bližino. To npr. pomeni, da bi najboljši anketarji opravljali delo v krajih, v katerih je anketiranje težje. Bolj izkušen anketar pa bi se angažiral tudi ob zavrnitvi sodelovanja. Takšna nenaključna razporeditev anketarjev seveda povzroča razlike zaradi njihovih različnih sposobnosti, zaradi različnih težavnosti anketirancev ali zaradi obojega, kar pa je težko ločiti. Če bi hoteli oceniti samo vpliv anketarjev, bi morali izbrati enostaven repliciran vzorčni načrt, npr. 1,000 elementov v obliki 20 neodvisnih ponovitev (replikacij) SRS vzorcev po 50 elementov. Vsak anketar bi moral opraviti 50 anket v eni od takih replikacij. Ker so replikacije primerljive, lahko razlike pripišemo anketarju oziroma njegovim sposobnostim. Seveda je opisan postopek bistveno lažje izvedljiv v telefonskih kot v terenskih anketah.

Za ocenjevanje vpliva anketarjev bomo v nadaljevanju uporabili standardno (npr. Iverson in Norpoth, 1976) enosmerno analizo variance (*angl. one-way analysis of variance*). Pri tem je treba dodati, da so rezultati nekoliko drugačni, če replikacije vsebujejo kompleksne vzorčne načrte.

S simboli $\bar{y}_1, \bar{y}_2, ..., \bar{y}_c$ bomo označili aritmetične sredine, ki smo jih dobili iz c ponovitev oziroma podvzorcev (replikacij) različnih anketarjev. Varianco c aritmetičnih sredin lahko ocenimo:

$$v_1 = \frac{1}{(c-1)}\sum_{r=1}^{c}(\bar{y}_r - \bar{y})^2 ,$$

kjer je

$$\bar{y} = \frac{1}{c}\sum_{r=1}^{c}\bar{y}_r$$

povprečje vzorčnih aritmetičnih sredin. Cenilka v_1 seveda predpostavlja, da med anketiranci različnih skupin ni sistematičnih razlik. V nadaljevanju bomo ob predpostavki SRS vzorčenja preverili ničelno domnevo, da vplivi anketarja v resnici niso prisotni. Če razlike obstajajo, bo ocena za v_1 seveda večja kot v primeru zgolj slučajnega variiranja med anketarji.

Če zanemarimo faktor končne populacije (*FPC*) iz izraza (3), dobimo za podvzorce naslednjo oceno:

$$v(\bar{y}_r) = \frac{s_r^2}{r} ,$$

kjer je s_γ^2 ocenjena elementarna varianca v podvzorcu γ, $r = n/c$ pa velikost podvzorca.

Aritmetična sredina ocen $v(\bar{y}_\gamma)$ vseh podvzorcev c je:

$$v_1 = \frac{\bar{s}^2}{r} = \frac{1}{c}\sum_{\gamma=1}^{c} v(\bar{y}_\gamma) = \frac{1}{c}\sum_{\gamma=1}^{c}\frac{s_\gamma^2}{r},$$

kjer je:

$$\bar{s}^2 = \frac{1}{c}\sum_{\gamma=1}^{c} s_\gamma^2$$

aritmetična sredina ocen elementarne variance v podvzorcih γ. Primerjava variance med anketarji (v_1) in povprečne variance anketarja (v_2) omogoča preverjanje ničelne domneve z uporabo F-testa:

$$F = \frac{v_1}{v_2} = \frac{rv_1}{\bar{s}^2}.$$

Visoka vrednost F pokaže prisotnost vpliva anketarjeve variance. Kritične vrednosti za F, kadar je F večji od 1, dobimo tako, da vzamemo običajni F-test pri:

$$(c-1) \text{ in } c(r-1)=(n-c)$$

prostostnih stopnjah. Dober kazalec variance anketarja je tudi intraklasna korelacija ρ_γ, ki jo lahko ocenimo z naslednjim izrazom (Kish, 1962):

$$\rho_\gamma^* = \frac{F-1}{F-1+r}.$$

Posledice variiranja med anketarji so podobne vplivu, ki ga običajno dobimo pri prostorskih skupinah. Na anketiranca namreč v določenem obsegu vpliva tudi anketar, podobno kot naselje ali soseska določajo odgovore oseb, ki tam prebivajo. Podobno torej kot vzorčni učinek *Deff* pri vzorčenju v skupinah obstaja tudi vpliv anketarjeve variance pri repliciranem SRS vzorčnem načrtu. Če z $Deff(\bar{y}_\gamma)$ in ρ_γ označimo vzorčni učinek oziroma intraklasno korelacijo zaradi vplivov anketarja, se ustrezna vzorčna varianca v primerjavi s SRS vzorcem v izrazu (3) seveda poveča:

$$Var(\bar{y}_\gamma) = Var(\bar{y}_{SRS}) \times Deff(\bar{y}_\gamma) = Var(\bar{y}_{SRS}) \times (1+(r-1)\rho_\gamma).$$

Podobno kot pri vzorčenju v skupinah lahko celo majhne vrednosti ρ_γ pomenijo precejšnje povečanje variance v primerjavi s SRS varianco, kadar je število r anket na posameznega anketarja visoko.

Ponoviti velja, da običajna ocena variance aritmetične sredine na osnovi SRS vzorca seveda ne vključuje vpliva anketarjev. Zato je v primeru, ko podatki to omogočajo, ugodno preveriti izračun vzorčne variance še z metodo replikacij, ki na osnovi variiranja med podvzorci upošteva tudi vpliv anketarja. Pravilna ocena vzorčne variance SRS vzorca, ki upošteva tudi vpliv anketarja, je torej:

$$Var(\bar{y}_r) = \frac{v_1}{c},$$

kar lahko obravnavamo tudi kot posebno obliko cenilke vzorčne variance pri vzorčenju v skupinah (15).

Seveda pa bi zahteva, da anketarju dodelimo slučajno izbrane elemente – ne glede na prostorsko razporeditev ter učinkovitost anketarjev – pri osebnem anketiranju razpršene populacije zaradi visokih stroškov in logističnih težav močno otežila izvedbo take raziskave in s tem tudi izračun anketarjevega vpliva. Zato pa je tak način nadvse primeren pri telefonskih vzorcih.

Druga uporaba repliciranega vzorčenja je ocenjevanje običajne vzorčne variance. Če imamo *c* ocen z_1, z_2,.... z_c parametra *Z*, ki jih dobimo iz *c* neodvisnih replikacij, je varianca cenilke:

$$\bar{z} = \sum_{\gamma=1}^{c} \frac{z_\gamma}{c}$$

podana z izrazom:

$$V(\bar{z}) = \frac{V(z_\gamma)}{c},$$

kjer $V(z_\gamma)$ ocenimo kot:

$$v_1 = \sum_{\gamma=1}^{c} \frac{(z_i - \bar{z})^2}{(c-1)}.$$

Od tod sledi:

$$v(\bar{z}) = \frac{1}{c(c-1)} \sum_{\gamma=1}^{c} (z_i - \bar{z})^2, \qquad [24]$$

kar je splošen izraz za oceno variance na podlagi repliciranega vzorčnega načrta. Uporabimo jo lahko pri vsakem parametru (indeksna števila, korelacije, regresijski koeficienti, aritmetične sredine, odstotki ...), vzorčni načrt podvzorcev pa je lahko tudi kompleksen, npr. stratificirani večstopenjski PPS vzorec.

Pri uporabi izraza (24) nastaja manjša težava, ker izraz ocenjuje varianco aritmetičnih sredin repliciranih vrednosti $\bar{z}$, ki v splošnem ni enaka siceršnji cenilki $\tilde{z}$, ki jo računamo kot enostavno aritmetično sredino vseh elementov v skupnem vzorcu. Vzorčna aritmetična sredina $\tilde{z}$ je zato običajno bolj zaželena cenilka, vendar je v praksi razlika med $\bar{z}$ in $\tilde{z}$ zanemarljivo majhna. Največkrat zato za oceno parametra Z uporabimo količino $\tilde{z}$, pri izračunu variance pa uporabimo v izrazu (24) varianco $v(\bar{z})$ in s tem ocenimo $v(\tilde{z})$.

Več težav je z izbiro parametra *c*, to je števila replikacij. Če izberemo premajhno število replikacij, bo ocena variance $v(\bar{z})$ nenatančna, kar lahko povzroči preširok interval zaupanja za oceno parametra. Pri c replikacijah ima namreč ocena $v(\bar{z})$ samo $(c-1)$ prostostnih stopenj pri izračunu intervala zaupanja s t-porazdelitvijo. Za ponazoritev si oglejmo SRS vzorec velikosti n=1,000. Običajni 95-odstotni interval zaupanja za $\bar{Y}$ namreč dobimo z izrazom:

$$\bar{y} \pm 1.96 \frac{s}{\sqrt{n}},$$

kjer je $z_{\alpha/2}$= 1.96 običajna vrednost, ki pri velikih vzorcih izhaja iz tabele normalne porazdelitve za tveganje α = 0.05. Pri replikacijah z 10 podvzorci po 100 elementov pa je interval zaupanja enak:

$$\bar{y} \pm 2.26\sqrt{\frac{v_1}{10}},$$

kjer smo $t_{\alpha/2}(9)$ = 2.26 dobili iz t-porazdelitve s $(t-1) = 9$ prostostnimi stopnjami (*angl. degrees of freedom).* Pri replikacijah s 4 podvzorci po 250 elementov je 95 % interval zaupanja še bistveno širši:

$$\bar{y} \pm 3.18\sqrt{\frac{v_1}{4}},$$

kjer $t_{\alpha/2}(3)$ = 3.18 izhaja iz t-porazdelitve s tremi prostostnimi stopnjami. Ker je navedena standardna napaka v vseh primerih nepristranska cenilka za vzorčno varianco slučajne spremenljivke $\bar{y}$, je interval zaupanja z replicirano oceno variance pri c = 10 replikacijah v povprečju za 15 % večji, pri 4 replikacijah pa kar za 62 % večji od običajne ocene. Da bi dobili oceno, ki je relativno natančna, mora biti število c torej dovolj veliko; običajno to pomeni vsaj 20 ali 30 replikacij. Po drugi strani pa lahko pri večjem številu replikacij uporabimo manj podrobno stratifikacijo, saj mora vsak podvzorec imeti vsaj en element v stratumu, kar je lahko velika ovira posebej pri večstopenjskih vzorčnih načrtih. Vzemimo, da potrebujemo npr. 60 enot prve stopnje (PSU). Z običajnim vzorčnim načrtom bi lahko enote (PSU) razdelili na 60 stratumov z eno PSU ali na 30 stratumov z dvema PSU. Pri replikacijah pa se že pri c = 10 največje število stratumov, ki jih lahko uporabimo, zmanjša na samo 6.

Če povzamemo, je prednost ocenjevanja variance z metodo enostavnih replikacij predvsem v enostavnosti uporabe, čeprav se pri tem lahko zmanjša natančnost. Če je c majhen, je natančnost ocene variance manjša zaradi manjšega števila prostostnih stopenj. Če pa je c velik, se natančnost zmanjšuje tudi zaradi manj podrobne stratifikacije. Enostavnih replikacij zato v praksi ne uporabljamo posebej pogosto. Namesto tega raje uporabimo tako imenovane psevdoreplikacije, ki omogočajo stratifikacijo in dopuščajo več prostostnih stopenj. Metode psevdoreplikacij bomo podrobneje predstavili v poglavju o analizi raziskave.

Panelne raziskave

Do zdaj smo implicitno predpostavljali, da izdelujemo vzorčne načrte za raziskave, ki se izvajajo samo enkrat. Seveda pa obstajajo tudi raziskave, pri katerih podatke na istih elementih zbiramo po večkrat – panelne raziskave.

Eden od najpomembnejših razlogov za ponovno anketiranje je raziskovanje spremembe v času. Pri tem moramo ločiti med skupnimi spremembami (*angl. gross changes*) in spremembami v razliki (*angl. net changes*). Skupne spremembe namreč opisujejo spremembe na ravni elementov, spremembe v razlikah pa le na ravni agregatov.

Oglejmo si to na primeru raziskave o učinkih prostočasnih dejavnosti na krvni tlak. Če nas zanimajo spremembe na ravni posameznika, moramo seveda v vzorec izbrati osebe, ki so bile že vključene v predhodno raziskavo, kjer elementi niso bili izpostavljeni aktivnostim, katerih vpliv proučujemo. Le v takem primeru namreč lahko analiziramo morebitno povezanost med spremenljivkama in ugotavljamo, da se je po določenih dejavnostih krvni tlak spremenil. Če pa potrebujemo le oceno razlike obravnavane spremenljivke za celo populacijo, npr. zmanjšanje celotnega deleža oseb z zvišanim tlakom v dveh obdobjih, pa panelna raziskava ni potrebna, čeprav je z vidika natančnosti ocen tudi v tem primeru običajno učinkovitejše, če uporabimo panel. Seveda pa v primeru, ko smo izračunali spremembo v razliki na podlagi dveh neodvisnih anket (brez panela), ne moremo oceniti vplivov prostočasnih dejavnosti na to spremembo.

Podoben primer je tudi anketno spremljanje brezposelnosti. Če bi nas zanimala le tekoča in splošna stopnja brezposelnosti, bi zadostovali neodvisni vzorci. V takem primeru (brez panelne raziskave) pa ne bi mogli oceniti skupne spremembe na nivoju elementov. Tako npr. ne bi vedeli, ali v dveh obdobjih ostajajo brezposelne iste osebe, ali pa so se povsem zamenjale. Ker je za proučevanje brezposelnosti taka informacija izredno pomembna, se na področju delovne sile uporabljajo panelne raziskave. Alternativna rešitev – retrospektivno spraševanje anketirancev o zaposlitvenem statusu pred npr. 12 meseci – je namreč zaradi merskih napak preveč nezanesljiva, kar si bomo ogledali v naslednjem primeru.

Drugi razlog za izvedbo panelne raziskave se namreč nanaša na zbiranje informacij, ki so dostopne le v določeni časovni točki, ko so anketiranci o njih sposobni natančno poročati. Tako bi npr. pri raziskavi, ki zahteva pregled dohodkov gospodinjstva v določenem obdobju, morali izvesti več intervjujev. Dobre podatke za določeno časovno obdobje pa lahko dobimo le takrat, ko se jih anketiranci še spomnijo, torej v času obravnavanega obdobja. Podobno velja za raziskavo, ki bi analizirala povezavo med predšolsko vzgojo otrok in kasnejšim uspehom v šoli. Tako anketo bi morali izvesti najprej v obdobju, ko so otroci v predšolski vzgoji, podatke o šolskem uspehu pa bi morali zbrati kasneje. Kakovost podatkov bi bila namreč močno ogrožena, če bi se zanašali na retrospektivna poročila o predšolski vzgoji v času, ko otroci že obiskujejo osnovno šolo.

Paneli ali longitudinalne študije, pri katerih dobivamo podatke o elementih v več časovnih točkah, prinašajo tudi nekaj novih težav. Prva težava je mobilnost elementov v vzorcu. V večini panelnih raziskav se nekateri elementi, najpogosteje so to osebe ali pa gospodinjstva, med trajanjem panelov preselijo. Osebe, ki se odselijo, pa moramo v vzorcu obdržati, če vztrajamo pri začetnem verjetnostnem vzorcu, kar zahteva učinkovite metode sledenja (*angl. tracking*). Osebe oziroma gospodinjstva, ki so se preselila, seveda zapustijo izbrano geografsko zaokroženo enoto (PSU) v večstopenjskem vzorčnem načrtu, kar vsekakor poveča stroške osebnega anketiranja.

Druga pogosta težava pri panelih je spreminjanje populacije v času: nekateri elementi izginejo in drugi se pojavljajo na novo. Vzemimo npr. dolgoročni panel o zdravju v določeni skupnosti. Na začetku izberemo verjetnostni vzorec oseb in jih spremljamo skozi več let. Med tem pa se populacija spreminja; nekateri zapustijo opazovano skupnost, ker se odselijo (npr. v drugo državo), drugi umrejo, spet tretji pa se pojavijo na novo (rojstva, priselitve). Elementi, ki odhajajo, povzročajo zmanjšanje velikosti vzorca oziroma osip (*angl. attrition*). Panel lahko zaradi tega ostane samo verjetnostni vzorec začetne populacije, ki še vedno živi v določenem predelu. Težava z elementi, ki prihajajo, je v tem, da v vzorcu sploh niso zastopani. Zato začetni vzorec ni več verjetnostni vzorec celotne populacije, ki trenutno živi na območju ciljne populacije. Če v populaciji nastaja dovolj velik del novih elementov, torej potrebujemo nov vzorec. Težave se še povečajo, ko izvajamo ankete v družinah ali gospodinjstvih, saj lahko precejšen delež družin oziroma gospodinjstev med panelom spremeni svojo sestavo.

Nadaljnja težava panelov je prilagajanje anketirancev *(angl. conditioning)*. Večkratno anketiranje ima namreč neugodne posledice za anketirance. Nekateri lahko sodelovanje v anketi sprejemajo kot breme in odklanjajo nadaljnje sodelovanje v panelu, s čimer povzročijo nevarnost pristranskosti. Pri tem je posebej kritično vplivanje vsebine panela, zaradi česar lahko anketiranci odgovarjajo drugače. Navedeni učinek se pojavi npr. v potrošniških panelih, kjer anketirance prosimo, da navedejo izdelke (in cene), anketiranci pa se pri tem spomnijo svojih prejšnjih odgovorov in odgovarjajo konsistentno glede na prejšnje valove ankete.

Zaradi naštetih težav se vključenost anketiranca v panelu pogosto časovno omeji s tako imenovanimi rotirajočimi paneli (*angl. rotation panel*). Kot preprost primer vzemimo, da je vsak anketiranec v anketo vključen le za tri ankete, potem pa ga iz panela izključimo. Za vsak val panela tako izpustimo tretjino panela in namesto tega vključimo novo tretjino, ki bo po tem anketirana še dvakrat. Če to pravilo predstavimo s črkami, lahko prvi vzorec označimo z ABC, naslednji z BCD, tretji s CDE, četrti z

DEF itd. Na tak način se vzorca v dveh zaporednih valovih prekrivata dvotretjinsko.

Kot smo že ugotavljali, panelna raziskava lahko pomembno izboljšuje ocenjevanje razlik. Oglejmo si preprosto cenilko ($\bar{y}_2$ - $\bar{y}_1$) za razliko v aritmetični sredini spremenljivke y v času 1 in 2. Varianca razlike slučajnih spremenljivk je splošno podana kot:

$$Var(\bar{y}_2 - \bar{y}_1) = Var(\bar{y}_1) + Var(\bar{y}_2) - 2Cov(\bar{y}_1, \bar{y}_2) =$$

$$= Var(\bar{y}_1) + Var(\bar{y}_2) - 2\bar{R}\sqrt{Var(\bar{y}_1)Var(\bar{y}_2)}, \qquad [25]$$

kjer je $\bar{R}$ korelacijski koeficient med vzorčnima aritmetičnima sredinama $\bar{y}_1$ in $\bar{y}_2$. Pri dveh neodvisnih vzorcih v dveh valovih ankete je seveda $\bar{R}$ = 0. Zadnji člen v izrazu (25) zato prikazuje pridobitev ($\bar{R}>0$) ali izgubo ($\bar{R}<0$) v natančnosti cenilke zaradi panelnega merjenja.

Da bi dobili dodatni vpogled v učinke prekrivanja vzorca, si oglejmo preprost primer statične populacije in SRS vzorčenja z velikostjo vzorca n. Predpostavimo tudi, da sta elementarni varianci v obeh valovih enaki ($S_1^2 = S_2^2 = S^2$), in zanemarimo faktor končne populacije (*FPC*). V takem primeru se izraz (25) poenostavi:

$$Var(\bar{y}_2 - \bar{y}_1) = 2S^2 \frac{1 - P\bar{R}}{n},$$

kjer je $\bar{R}$ že omenjena korelacija v obeh valovih, P pa delež prekrivanja obeh vzorcev. Primer neodvisnih vzorcev (P = 0) in popolno prekrivanje (P = 1) torej predstavljata le poseben primer. Ko je P = 0, imamo dva neodvisna vzorca in varianco razlike izračunamo s posebej preprostim izrazom:

$$\frac{2S^2}{n}.$$

Razmerje med varianco razlike v panelnem vzorčnem načrtu ter v dveh neodvisnih vzorcih je zato ravno:

$$(1 - P\bar{R}). \qquad [25a]$$

Oglejmo si primer. Predpostavimo, da je korelacija opazovane spremenljivke (npr. individualnih političnih mnenj, brezposelnosti ali krvnega tlaka) v dveh valovih ankete enaka $\bar{R}$ = 0.75. Varianca pri popolnem prekrivanju (P = 1) dveh vzorcev je v tem primeru produkt osnovne variance (pri dveh neodvisnih vzorcih) in faktorja:

$$(1 - 0.75) = 0.25.$$

Pri delnem (npr. dvotretjinskem) prekrivanju je faktor seveda manjši:

$$(1 - 2/3 \times 0.75) = 0.50.$$

Če je korelacija visoka, so torej pridobitve panela pri ocenjevanju razlik lahko precejšnje. Pri rotirajočem panelu pa lahko z uporabo kompleksnih cenilk še dodatno povečamo natančnost (Kish, 1965, 463–464).

Opozoriti je treba, da je pri negativnem $\bar{R}$ ocena razlike lahko celo manj natančna. Negativne vrednosti se lahko pojavljajo, kadar v gospodinjstvih opazujemo nakup večje potrošne dobrine, saj je npr. po nakupu avtomobila za določeno časovno obdobje verjetnost za tak nakup bistveno manjša. Pri $\bar{R} = -0.2$ in popolnem prekrivanju ($P = 1$) imamo namreč:

$$(1 - P\bar{R}) = 1.2,$$

zato je varianca pri panelu večja za 20 %.

Na koncu velja dodati, da pridobitve iz izraza (25) niso omejene samo na primere, ko imamo v panelu iste elemente. Vzorčne načrte, pri katerih ohranimo v panelu iste enote prve stopnje (PSU), v njih pa vsakič izberemo druge elemente, lahko na enak način uporabimo za izboljšanje ocenjevanja razlik. Seveda pa je v takem primeru koeficient korelacije $\bar{R}$ v splošnem nižji kot pri primerjavi istih elementov. Tipičen primer je vzorec stanovanj namesto vzorca gospodinjstev, pri katerem sledenje preseljenim gospodinjstvom ni potrebno, saj vsakič anketiramo trenutnega stanovalca.

8. VZORČNI OKVIRI

Problematiko vzorčenja lahko v splošnem razdelimo na dve podpodročji: na eni strani imamo opraviti z načrtovanjem vzorcev, na drugi strani pa z izračunavanjem vzorčne variance in izborom primerne cenilke. Z izborom cenilke se v tem delu posebej ne ukvarjamo, čeprav pogosto obstaja večja izbira možnih cenilk, predvsem pri ocenjevanju populacijskih vsot v obsežnejših anketah.

V praksi večine mnenjskih in marketinških anket nas namreč problematika izbora cenilke posebej ne zanima, saj pogosto zadošča že navadna vzorčna aritmetična sredina. Še več, pogosto nas ne zanima niti ocenjevanje vzorčne variance, saj izvajalci mnenjskih in marketinških raziskav razmeroma redko poročajo o natančnosti ocen. Tovrstno izračunavanje namreč znatno podraži in podaljša izdelavo raziskovalnega poročila, naročniku raziskave pa podrobna informacija o vzorčni varianci pogosto ne bi pomenila veliko.

Z dvigovanjem strokovnih standardov, z naraščajočo konkurenco med izvajalci anketnih raziskav in predvsem z vse zahtevnejšimi in bolj izobraženimi uporabniki pa se veča tudi potreba po navajanju informacij o natančnosti vzorčnih ocen. Raziskovalne in statistične organizacije, ki opravljajo kakovostnejše storitve, zato v vse večji meri zagotavljajo naročnikom in uporabnikom tudi tovrstne informacije.

Zaradi vseh navedenih razlogov se analiza vzorčnih podatkov v veliki večini anketnih raziskav poenostavlja le na siceršnjo statistično analizo (deskriptivna statistika, bivariatna analiza, multivariatna analiza, preverjanje domnev ipd.). Analizo se običajno izvede s standardiziranimi statističnimi paketi, ki predpostavljajo enostavno slučajno vzorčenje in v primeru razmernostnih spremenljivk tudi normalno populacijsko porazdelitev. Le pri vzorcih uradne statistike, v kakovostnih akademskih anketah, v redkih marketinških raziskavah, pri katerih naročnik izrecno zahteva oceno natančnosti in tudi pri nekaterih mednarodnih raziskavah se opravi pravilno ocenjevanje ciljnih spremenljivk oziroma parametrov statističnih modelov (npr. lastne vrednosti, regresijski koeficient ipd.). Pri tem se največkrat uporablja posebne statistične programe (npr. WesVar, Sudaan) ali metode ponovnega vzorčenja.

V Sloveniji je anketnih raziskav, pri katerih se vzorčna varianca pravilno izračunava, razmeroma malo. Sem sodijo nekateri vzorci Statističnega urada, npr. Anketa o delovni sili (Vehovar, 1994a) ter anketne raziskave, pri katerih to zahtevajo mednarodni metodološki standardi, npr. Anketa o rodnosti (Kožuh-Novak et. al., 1998), Anketa o pismenosti (Možina et al., 1999). Nekoliko pogosteje se varianca izračunava le v poenostavljeni obliki (ob večinoma napačni) predpostavki SRS vzorčenja, s čimer se širina intervalov zaupanja seveda podcenjuje za koren vzorčnega učinka *Deff*.

Problematika izbora cenilke ter ocenjevanja vzorčne variance je torej v anketni praksi omejena le na razmeroma redke uradne, marketinške ali mednarodne raziskave, zato se vzorčenje pogosto poenostavlja na problematiko načrtovanja vzorca. Vendar tudi glede določanja velikosti vzorca pogosto velja determinizem – bodisi zaradi omejenih finančnih sredstev bodisi zaradi standardiziranih ali predpisanih shem, ki v raziskovalnih organizacijah in statističnih uradih urejajo velikost vzorca. Načrtovanje vzorcev se zato v anketni praksi pogosto zoži na vprašanje vzorčnega okvira, to je seznama, s katerega bomo izbirali elemente v vzorec.

Izbor vzorčnega okvira je zato v praksi nemalokrat osrednje vprašanje vzorčenja, hkrati pa tudi najpomembnejši dejavnik, ki določa kakovost anketne raziskave. Iz narave vzorčnega okvira namreč hitro razberemo tudi celotno raven in resnost raziskave.

Temeljna vloga vzorčnega okvira je identifikacija vseh elementov ciljne populacije. Seveda pa sama identifikacija ne zadošča; dober vzorčni okvir mora elemente tudi locirati. Tako so npr. lahko na zastarelem seznamu vsi elementi ciljne populacije pravilno identificirani, njihovi naslovi ali telefonske številke pa zaradi neažurnosti ne omogočajo stika. Kakovosten vzorčni okvir pa mora poleg identifikacije in lociranja elementov omogočiti še dodatne informacije, ki jih je moč uporabiti pri stratificiranju in načrtovanju večstopenjskega vzorčenja.

Idealen vzorčni okvir vsak populacijski element uvrsti na seznam le enkrat, in to brez dodatnih seznamov. V praksi pa se to zgodi le redko, saj nastajajo različna odstopanja, ki jih je Kish (1965: 53–59) strnil v naslednje štiri skupine:

- manjkajoči elementi (*angl. missing elements*): populacijski elementi, ki niso vključeni v vzorčni okvir, čeprav sodijo v ciljno populacijo,
- skupine *(angl. clusters):* vzorčni okvir vključuje na enem naslovu skupino elementov in ne le posamezen element,
- neustrezni elementi (*angl. uneligble elements*): vzorčni okvir vključuje prazne ali tuje elemente, ki ne pripadajo ciljni populaciji,
- podvojeni zapisi *(angl. duplicate listings)*: populacijski elementi se v vzorčnem okviru pojavljajo po večkrat.

Manjkajoči elementi

Denimo, da imamo v anketni raziskavi med študenti na voljo določen seznam študentov fakultete. Prvo vprašanje, ki si ga moramo zastaviti, je, ali seznam vključuje vse študente v ciljni populaciji.

Manjkajoči elementi se lahko pojavijo, ker je okvir neprimeren *(angl. inadequate),* kar pomeni, da ni namenjen naši ciljni populaciji, ali pa nepopoln *(angl. incomplete),* kar pomeni, da je ustrezen, vendar ne vključuje vseh elementov. Razlikovanje med neprimernostjo in nepopolnostjo je praktičnega pomena, saj je prvo kategorijo pogosto laže prepoznati. V opisanem primeru bi bil seznam neprimeren, če bi npr. izključeval izredne študente, ki so del ciljne populacije, nepopoln pa bi bil v primeru, če je nekoliko zastarel in ne vključuje nekaterih novih študentov.

Manjkajoči elementi so eno najresnejših vprašanj pri oblikovanju vzorčnih okvirov, saj elementi, ki sodijo v ciljno populacijo, sploh nimajo možnosti, da bi bili vključeni v vzorec. Njihova verjetnost za vključitev je enaka nič, zato ne moremo govoriti o verjetnostnem vzorcu, v katerem imajo vsi elementi znano in pozitivno verjetnost za vključitev.

Včasih lahko vprašanje manjkajočih elementov zaobidemo z redefiniranjem ciljne populacije, tako da ne vključuje manjkajočih elementov. Taka rešitev je sprejemljiva le, kadar ni na voljo nobene druge možnosti, manjkajoča skupina pa predstavlja zanemarljiv delež celotne populacije in njena izključitev ne vpliva na cilje raziskave. Tako npr. pogosto populacijo oseb v določeni državi zožimo na domača (rezidenčna) gospodinjstva in s tem izločimo osebe, ki živijo v ustanovah.

Pri manjkajočih elementih pogosto poiščemo dodatni okvir, v našem primeru seznam izrednih študentov ali seznam novih študentov. Slednje sicer lahko privede do novih težav, to je do podvajanja elementov, saj se lahko nekateri elementi pojavijo na več seznamih. Podvajanje elementov pa je manjša težava kot manjkajoči elementi in jo rešujemo z eno od metod, ki jih bomo obravnavali v nadaljevanju.

Pogosto pa nimamo na voljo nobenega dodatnega okvira za manjkajoče elemente. Takrat lahko v določenih primerih uporabimo postopke povezovanja (*angl. linking procedures*), ki enolično pridružujejo manjkajoče elemente. Ko je s povezovanjem manjkajočih elementov seznam dopolnjen, obravnavamo povezane elemente skupaj z manjkajočimi elementi kot skupino. Povezovanje torej sproži vprašanje skupin elementov, ki je lažje rešljivo kot vprašanje manjkajočih elementov.

Oglejmo si tipičen primer povezovanja elementov. Denimo, da vzorčni okvir vključuje abecedni seznam študentov. Manjkajoče elemente, ki niso v seznamu, lahko povežemo tako, da se vsak študent s seznama poveže v skupino s študenti, ki so po abecednem redu za njim, vendar pred naslednjim študentom s seznama. Da pokrijemo tudi manjkajoče študente z začetka abecede, seznam obravnavamo kot krožen in manjkajoče študente, ki so po abecedi za zadnjim študentom in pred prvim študentom, povežemo z zadnjim študentom s seznama. Seveda lahko opisani postopek uporabimo le, kadar študente izbiramo v vzorec npr. na podlagi prisotnosti v predavalnici ali z obvestili, ki dosežejo celotno populacijo. Takšno povezovanje imenujemo polodprti interval *(angl. half-open interval)* in ga uporabljamo tudi v drugih primerih. Eden pogostejših primerov uporabe takega povezovanja je tudi vzorčenje stavb na podlagi seznamov hišnih številk. Pri tem z uporabo polodprtega intervala vse stavbe, ki jih ni na seznamu, povežemo z zadnjo hišno številko na seznamu, ki je pred manjkajočo stavbo.

Skupine elementov

Kot smo že omenili, lahko nastanejo nove skupine na osnovi povezovanja manjkajočih elementov. Skupine pa se pojavljajo tudi v drugih primerih, npr. ko potrebujemo vzorec oseb, razpolagamo pa z vzorčnim okvirom stanovanj. Osnovni pristop k vprašanju skupin je, da vključimo v vzorec vse elemente v izbrani skupini, saj imajo s tem vsi elementi v skupini enako verjetnost, da se pojavijo v vzorcu, kot bi jo imel izbrani element. Če so elementi z osnovnega seznama vzorčeni EPSEM, velja v takem primeru to tudi za elemente v skupini, ki se na seznamu »skrivajo«. Če so npr. elementi gospodinjstva, skupine pa stanovanja, je takšna rešitev posebej enostavna, saj večina stanovanj vključuje le eno gospodinjstvo, in tudi ko je gospodinjstev več, je njihovo število majhno. Po drugi strani pa tudi anketiranje več kot enega gospodinjstva v izbranem stanovanju le redko pomeni izvedbeno težavo.

Zgoraj omenjena rešitev torej vodi v vzorčenje skupin, kar lahko pripelje k neugodnemu vzorčnemu učinku *Deff*, posebej če je povprečna velikost skupine velika in je intraklasna korelacija visoka. Če obstaja nevarnost, da je *Deff* velik, lahko uporabimo podvzorčenje znotraj skupin. Pri nekaterih skupinah pa je pomemben razlog za uporabo podvzorčenja tudi medsebojno vplivanje elementov znotraj skupine, zaradi česar pogosto izberemo le po en element v skupini. Do tega pogosto prihaja v raziskavah stališč, ko so elementi osebe, skupine pa gospodinjstva (ali stanovanja). Če v gospodinjstvu anketiramo dve ali več oseb, namreč na odgovore druge anketirane osebe vpliva razprava o vsebini ankete s prvim anketirancem. Nadaljnji anketiranci so lahko zato manj pripravljeni na sodelovanje in ga pogosteje zavrnejo (saj so seznanjeni z vsebino in dolžino ankete), če pa sodelujejo, dajejo pristranske odgovore. Poleg tega obstaja velika verjetnost, da se dodatni anketiranci v gospodinjstvu ne razlikujejo od predhodnih, torej z dodatnimi stroški anketiranja ne pridobimo nobene nove informacije.

Kadar se odločimo za izbiro enega samega elementa v skupini, se seveda spremenijo verjetnosti izbora elementov. Če so skupine različno velike in npr. skupina α vsebuje B_α elementov, izberemo pa en sam element, se je verjetnost izbire tega elementa zmanjšala na:

$$\frac{1}{B_\alpha}P(\alpha),$$

kjer je $P(\alpha)$ osnovna verjetnost izbire za skupino. Če so bile skupine vzorčene EPSEM, potem vzorec elementov očitno ni več EPSEM.

Da bi se izognili pristranskosti, je treba izbiro elementov iz skupin izpeljati po strogih verjetnostih mehanizmih. Vzemimo vprašanje izbire anketiranca v gospodinjstvu. V tem primeru je zaželeno, da anketar že v prvem stiku z gospodinjstvom naključno izbere osebo. Pri tem je ena od možnosti, da anketar člane gospodinjstva najprej vpiše na oštevilčen seznam, nato pa izbere enega s pomočjo tabele slučajnih števil. Slabost takega postopka je, da ga težko preverimo, zato obstaja nevarnost, da ga anketarji uporabijo tako, da izberejo osebe, ki so doma in so pripravljene sodelovati.

Rešitev, ki se pogosto uporablja (Kish, 1965) za izbor anketiranca v gospodinjstvih, je Kisheva metoda izbora osebe *(angl. Kish selection grid)*. Podlaga tega objektivnega in preverljivega postopka je, da anketar vpiše člane gospodinjstva po določenem vrstnem redu na oštevilčen seznam in s posebno tabelo določi zaporedno številko izbrane osebe. Nadvse primeren in nedvoumen način za razvrstitev članov gospodinjstva je razvrstitev po starosti in spolu. Ker je za tako razvrstitev treba ugotoviti le relativno starost, običajno ni treba povpraševati po natančni starosti, saj generacijske razlike zadostujejo za določitev vrstnega reda po starosti znotraj vsakega od spolov. Vzemimo raziskavo zaposlenih ob predpostavki, da gospodinjstvo ne vključuje več kot štiri take osebe. Vprašalnik bi, denimo, anketarju dal naslednja navodila za izbiro osebe:

Število zaposlenih v gospodinjstvu:	1	2	3	4
Anketirana je oseba z zaporedno številko:	1	2	2	3

Razvrstitev številk v drugi vrstici za posamezne vprašalnike se opravi v skladu s tabelo 8. Pri tem bi vsak anketar prejel eno od slučajno določenih šestih tabel (A–F). Anketarju, ki smo ga obravnavali v zgornjem primeru, bi očitno pripadala tabela D.

V gospodinjstvu, kjer je zaposlena ena sama oseba, je ta oseba seveda vedno izbrana. V gospodinjstvu, kjer sta zaposleni dve osebi, pa je izbrana prva oseba s seznama, če vprašalnik vsebuje tabele A, B ali C, druga oseba pa je izbrana v primeru D, E ali F. Iz drugega stolpca tabele 8, ki prikazuje delež vprašalnikov z določeno tabelo, namreč razberemo, da ima polovica vprašalnikov tabele A, B ali C, polovica pa tabele D, E ali F. Tako je v gospodinjstvu, v katerem sta dva zaposlena, možnost za izbor v

Tabela 8: Primer Kisheve izbire oseb v gospodinjstvu

		Skupno število zaposlenih oseb			
		1	2	3	4
Tabela	**Delež vprašalnikov**	**Zaporedna številka izbrane osebe**			
A	1/4	1	1	1	1
B	1/12	1	1	1	2
C	1/6	1	1	2	2
D	1/6	1	2	2	3
E	1/12	1	2	3	3
F	1/4	1	2	3	4

vzorec enaka za vsakega od njiju. Prav tako je v gospodinjstvih s tremi zaposlenimi enaka možnost za vsakega člana, da bo izbran v vzorec. Podobno velja tudi za gospodinjstva s štirimi zaposlenimi. Čeprav prihaja pri izbiranju ene osebe seveda do neenakih verjetnosti pri različnem številu zaposlenih v gospodinjstvu, pa daje Kisheva izbira enake možnosti izbora vsem osebam znotraj izbranega gospodinjstva.

Opisani postopek je predpostavljal največ štiri zaposlene na gospodinjstvo, kar lahko tudi povečamo, vendar so za to potrebne dodatne tabele. Pri vzorčenju odraslih v ameriških gospodinjstvih se pogosto uporabljajo tabele z največ šestimi osebami. Kish (1965: 399) v takem primeru predlaga osem tabel, ki omogočajo enake možnosti izbora za vse člane gospodinjstev velikosti 1, 2, 3, 4 in 6. Za člane gospodinjstva velikosti 5 pa možnosti izbora sicer niso popolnoma enake in tudi v gospodinjstvih, ki imajo več kot 6 članov, nekateri od članov niso upoštevani. Običajno pa je ta pomanjkljivost tako majhna, da jo lahko zanemarimo.

Pri uporabi Kisheve izbire se včasih navaja nevarnost, da je anketarjem težko začeti anketo z urejanjem seznama. Zaradi uvodnega spraševanja o spolu in starosti namreč lahko pride do povečane stopnje zavrnitev. Zato sta Troldahl in Carter (1964) razvila metodo, s katero se izognemo občutljivi izdelavi seznamov; anketarji morajo namesto tega zbrati le podatke o številu ustreznih oseb ter o številu moških oziroma žensk v gospodinjstvu. Anketar nato v posebni razpredelnici na podlagi skupnega števila ustreznih oseb na eni osi ter števila ustreznih moških na drugi osi odčita izbranega anketiranca. Tako kot pri Kishevi izbiri obstaja tudi v tem primeru več različnih tabel z različnimi izbirami. Pri Troldahl-Carterjevi metodi tako obstajajo štiri tabele, ki se porazdelijo med vprašalnike z enako frekvenco. Bryant (1975) je predlagal še nadaljnjo izboljšavo tega postopka zaradi težav, ki nastajajo pri pomanjkanju moških v vzorcu. Pri tej različici prepolovimo verjetnost za pojavljanje tabele, kjer so z večjo verjetnostjo izbrane ženske. Čeprav niso povsem nepristranski, pa se tovrstni postopki pogosto uporabljajo v praksi (Czaja, 1982).

Vsekakor je treba ponoviti, da v primeru, ko so skupine vzorčene EPSEM, samo izbira vseh elementov v skupini daje EPSEM vzorec elementov. Kot rečeno, pa je taka rešitev pogosto nesprejemljiva bodisi zaradi prevelikih skupin bodisi zaradi nevarnosti medsebojnega vplivanja. Izbiranje enega slučajnega elementa v skupini – kar smo predstavili v zgornjih odstavkih – se tej nevarnosti sicer izogne, zavedati pa se je treba, da spreminja verjetnosti izbora, zato je treba uporabiti uteževanje.

Obstaja tudi alternativna rešitev, ki uporablja dvofazno vzorčenje. Pri tem izbiramo le en element iz vsake skupine, hkrati pa ohranimo EPSEM vzorec. V prvi fazi tako najprej izberemo vzorec skupin in elemente v teh skupinah uredimo v seznam, v drugi fazi pa izbiramo elemente s seznama. Oglejmo si primer. Denimo, da so skupine gospodinjstva, pri čemer nobeno od gospodinjstev ne šteje več kot šest odraslih oseb. V prvi fazi najprej izberemo večji vzorec gospodinjstev, ki mora zagotoviti seznam najmanj šestkrat toliko oseb, kot jih potrebujemo. V drugi fazi nato s sistematičnim vzorcem oseb vključujemo po eno osebo iz posameznega gospodinjstva.

Neustrezni elementi

Neustrezni elementi vključujejo prazne elemente (*angl. blanks*), ki jih v populaciji sploh ni (več), npr. osebe, ki so umrle oziroma se izselile, ali pa neobstoječe oziroma porušene hiše. Neustrezni elementi se nanašajo tudi na tuje elemente (*angl. foreign elements*), ki se sicer nahajajo v okviru, vendar ne sodijo v ciljno populacijo, npr. brezposelni v raziskavi o podrobnostih dela na delovnem mestu, ali mlajši od 18 let v volilni raziskavi. Zaradi enostavnejše obravnave bomo tako za prazne kot za tuje elemente uporabljali skupni pojem neustreznih elementov (*angl. uneligible elements*).

Pri neustreznih elementih je rešitev nadvse enostavna – takih elementov enostavno ne upoštevamo. Takšen pristop smo že uporabili pri vzorčenju dijakov, ko smo nekatere identifikacijske številke dijakov obravnavali kot prazne elemente zaradi dijakov, ki so se izpisali iz šole. Pomembna posledica neustreznih elementov v vzorčnem okviru seveda je, da se velikost vzorca zmanjša, zato moramo to upoštevati že pri določanju velikosti vzorca. Neustrezne elemente je namreč treba vnaprej predvideti in vsaj v grobem oceniti njihov delež, nato pa ustrezno povečati začetni vzorec. Pogosta napaka, posebej pri sistematičnem vzorčenju, pa je nadomeščanje neustreznih elementov s sosednjim ali z naslednjim na seznamu. Temu se moramo vsekakor izogibati, saj tak postopek poveča verjetnost izbora za »naslednje« elemente, ki so lahko izbrani tako neposredno kot tudi v primeru, ko je izbran predhoden neustrezni element. Pri sistematičnem vzorčenju bi zato morali vzorčni interval izvajati neodvisno od tega, ali smo pri tem naleteli na neustrezne elemente. Če pa smo izbrali neustrezne elemente, jih je treba enostavno izpustiti.

Iz praktičnih razlogov je treba razlikovati med položajem, ko lahko prazne elemente vnaprej odkrijemo iz vzorčnega okvira, in položajem, ko jih vnaprej ni mogoče identificirati. V prvem primeru jih lahko izpustimo že pri vzorčenju, v drugem pa jih moramo najprej poiskati, preden jih lahko izpustimo. Tako je npr. v vzorcu moških, ki jih izbiramo s skupnega seznama moških in žensk, mogoče skoraj vse ženske izločiti že na podlagi njihovih imen, medtem ko bi v raziskavi oseb v starosti 40–64 let morali najprej opraviti pregledni pogovor (*angl. screening*), da bi ugotovili, ali izbrane osebe spadajo v ciljno populacijo.

Poseben primer neustreznih elementov so tudi tako imenovane redke populacije *(angl. rare population*), pri katerih predstavlja ciljna populacija le majhen del vzorčnega okvira. Kadar na podlagi vzorčnega okvira ustreznih elementov ni mogoče identificirati, je treba namreč neustrezne elemente izločati s terenskim pregledovanjem. Eden od načinov take identifikacije je – kot je bilo že omenjeno – uporaba dvofaznega vzorčenja. Pri tem v prvi fazi uporabljamo postopek pregledovanja, s čimer identificiramo elemente, ki pripadajo redki populaciji.

Podvojeni zapisi

Do podvojenih zapisov (*angl. duplicate listings*) prihaja, kadar je vzorčni okvir sestavljen iz več seznamov in se nekateri elementi pojavijo na več kot enem seznamu. Do podvojenih zapisov prihaja tudi, kadar ciljno populacijo oziroma predmet analize predstavljajo določene skupine, npr. gospodinjstva, vzorčni okvir pa sestavljajo zapisi, ki so le del teh skupin, npr. osebe. Pri tem lahko več različnih oseb vodi k istemu gospodinjstvu, ki ima zato seveda večjo verjetnost za izbor v vzorec. Podoben primer lahko nastane, kadar ima gospodinjstvo v imeniku več telefonskih številk.

Vprašanje podvojenih zapisov torej nastaja zaradi spremenjenih verjetnosti izbora, ki so odvisne od (neznanega) števila podvojenih zapisov v vzorčnem okviru. Ena od možnih rešitev je, da odstranimo vse dvojnike iz celotnega vzorčnega okvira, kar pa običajno ni izvedljivo. Druga možnost je uporaba identifikacije (*angl. unique identification*), ki povezuje vsak element ciljne populacije samo z enim od njegovih zapisov, in to na enolično določen način (npr. prvi zapis ali najstarejši zapis), preostale zapise tega elementa pa jemlje kot neustrezne elemente. Tak postopek, se npr. uporablja pri vzorčenju gospodinjstev iz volilnega registra v Veliki Britaniji. V urbanih območjih so namreč volivci oštevilčeni ter urejeni v volilna okrožja glede na ulico in hišni naslov. Vzorec volivcev je pri tem določen s sistematičnim vzorčenjem na podlagi seznama volilnih številk. Če je izbrani volivec prvi na seznamu na določenem naslovu, se izbrani naslov vključi v vzorec, če pa izbrani volivec ni prvi na seznamu na tem naslovu, se celoten naslov obravnava kot neustrezen. V nadaljevanju na izbranih naslovih izberemo v vzorec vsa gospodinjstva.

Seveda pa organizacija vzorčnega okvira pogosto ne omogoča vnaprejšnje enotne identifikacije dvojnikov. V takih primerih lahko enotno identifikacijo izvedemo šele s terenskim pregledovanjem, ko potencialne anketirance povprašamo o potrebnih podatkih, nato pa anketiramo le enega od podvojenih elementov. Na splošno pa tovrstno pregledovanje povzroča precejšnje stroške, zato je taka rešitev stroškovno nadvse neugodna. Alternativna rešitev je, da enostavno sprejmemo vse podvojene elemente, nato pa uporabimo postopke uteževanja, kar podrobneje obravnavamo v poglavju o analizi vzorcev.

Administrativni vzorčni okviri

Pri proučevanju splošne populacije običajno obstajajo določeni administrativni seznami (registri, evidence), ki so namenjeni upravni oziroma poslovni dejavnosti, ob tem pa so primerni tudi kot vzorčni okviri za ankete oseb, gospodinjstev ali stanovanj. Telefonski vzorci imajo v tem pogledu posebno mesto in jih bomo obravnavali v posebnem poglavju. Administrativni vzorčni okviri so pomembni tudi pri vzorčenju poslovnih subjektov, kot so npr. podjetja, trgovine, obrtniki, šole, zdravstveni domovi, kmetije, društva ipd. V Sloveniji je v tem pogledu posebej pomemben Poslovni register Statističnega urada Republike Slovenije, ki vključuje vse pravne osebe. V splošnem pa v anketah poslovnih subjektov *(angl. establishment surveys)* uporabljamo množico najrazličnejših administrativnih in poslovnih seznamov, pri čemer veljajo siceršnja načela za izdelavo vzorca. Vzorčnih okvirov poslovnih subjektov zato v nadaljevanju ne obravnavamo posebej. Namesto tega se osredotočamo predvsem na vzorčne okvire za ankete prebivalstva. Tovrstne ankete so namreč najštevilnejše, poleg tega pa so načela, postopki in problemi vzorčnih okvirov skoraj povsem enaki tudi pri drugih populacijah.

Pri administrativnih vzorčnih okvirih je treba omeniti tudi javne sezname poštnih naslovov, ki so v nekaterih državah – Slovenija v letu 2001 še ni med njimi – izdelani na ravni stanovanjskih enot. Za vzorčenje oseb pa so seveda najzanimivejši uradni podatki registrov in evidenc, kot npr. register prebivalstva, seznami zdravstvenega zavarovanja, pokojninskega zavarovanja ali davčnih zavezancev. Posebej dragoceni so tisti viri, ki so tudi centralizirani. Najkakovostnejši centralizirani administrativni vir pa je vsekakor register prebivalstva oziroma volilni register, s katerim – v taki ali drugačni obliki – razpolagajo v večini razvitih držav.

Primerni administrativni vzorčni okviri so lahko tudi seznami naročnikov električne energije, naročnikov vodovodnih priključkov, odjemalcev komunalnih storitev ipd. Navedenim primerom pa je nemalokrat skupno tudi to, da so decentralizirani, kar pomeni, da se moramo za pridobitev seznama ločeno pogajati z regionalnimi organizacijami. Pogosto imajo tovrstni seznami še dodatne neugodne lastnosti skupin (npr. skupen vodovodni števec za več gospodinjstev ali stanovanj), težave s posebnimi vrstami neažurnosti ter neusklajenost s prostorskimi enotami, za katere razpolagamo z uradnimi statističnimi podatki.

Dodatna težava pri tovrstnih okvirih so tudi predpisi o varovanju osebnih podatkov, ki otežujejo ali celo onemogočajo uporabo administrativnih virov (registrov, evidenc). V tem pogledu je register prebivalstva določena izjema, saj je v Sloveniji – in v mnogih drugih državah – z zakoni predvidena tudi njegova funkcija vzorčnega okvira za anketne raziskave. Pri tem so pogoji tovrstne uporabe urejeni s podzakonskimi predpisi in diskrecijsko pravico odločevalcev, kar seveda vnaša določeno negotovost v uporabo za potrebe vzorčenja. Tako so bili npr. pogoji za pridobitev vzorca iz registra prebivalstva v Sloveniji od konca osemdesetih let precej različni. V nekaterih obdobjih je bila uporaba povsem onemogočena in je bilo potrebno celo za Anketo o delovni sili izdelati prostorski vzorec. V drugih obdobjih je bilo treba dokazovati – s sodelovanjem registriranih raziskovalcev oziroma registriranih raziskovalnih organizacij – raziskovalno naravo vzorčne ankete. Spet tretjič ni bilo nobenih posebnih omejitev. Podobna različnost in nepredvidljivost sta pogosti tudi pri drugih administrativnih seznamih, omejitve pa so običajno še bistveno strožje.

Tako kot v Sloveniji se administrativni seznami uporabljajo za vzorčenje tudi v številnih drugih državah in tudi drugod tovrstno uporabo urejajo spreminjajoči se predpisi. Vprašanja o varovanju podatkov so namreč tesno povezana z odnosom javnosti do zasebnosti, s politizacijo tovrstnih vprašanj, pa tudi z zlorabami in škandali, ki pri tem nastajajo.

Če govorimo o Sloveniji, je treba poudariti, da je register prebivalstva izjemno kakovosten vzorčni okvir. Centralni register prebivalstva (CRP) se je pričel v Sloveniji vzpostavljati že po Popisu 1970, dokončno obliko pa je dobil z zakonsko vzpostavljeno enotno matično številko občana (EMŠO) leta 1980. Od leta 1998 ga upravlja Ministrstvo za notranje zadeve in s tem je register postal referenčna podatkovna baza za e-upravo. Za uporabo registra v statistične namene pa še naprej skrbi tudi Statistični urad Republike Slovenije.

Kadar to dopuščajo časovne in finančne omejitve, je pri osebnem anketiranju uporaba registra prebivalstva eden najpomembnejših korakov za kakovostno anketno raziskavo. Seveda pa se tudi register prebivalstva srečuje z določenimi težavami. Za njihovo razumevanje je treba v pogledu državljanstva razmejevati naslednje kategorije oseb:

- državljani Republike Slovenije,
- tujci z začasnim prebivališčem v Sloveniji,
- tujci s stalnim prebivališčem v Sloveniji.

Državljane Republike Slovenije pa je treba ločevati glede na kraj in način stalnega prebivališča:

- državljani Republika Slovenije s stalnim prebivališčem v Sloveniji ,
- državljani Republike Slovenije s stalnim in tudi z začasnim prebivališčem v Sloveniji,
- državljani Republike Slovenije s stalnim prebivališčem v tujini,
- državljani Republike Slovenije, ki začasno prebivajo v tujini.

Zgornje kategorije so med seboj seveda močno prepletene. Pri izbiri vzorca je zato treba natančno opredeliti ciljno populacijo. Če gre za raziskavo volilnega razpoloženja, običajno upoštevamo le državljane Slovenije s stalnim prebivališčem v Sloveniji. Zavedati pa se je treba, da ima volilno pravico tudi več tisoč slovenskih državljanov, ki so za stalno odseljeni (izseljenci). Ob državnozborskih volitvah leta 2000 so bili v to kategorijo vključeni državljani iz 31 različnih držav. Istočasno je seveda v volilni raziskavi treba izločiti tujce, ki imajo stalno prebivališče v Sloveniji, nimajo pa državljanstva. Kadar pa nas zanimajo vsebine, kot so npr. zaposlenost, potrošniško obnašanje, medijska potrošnja ipd., so lahko ciljna populacija vse osebe s stalnim prebivališčem v Sloveniji ne glede na državljanstvo in status formalne prijave. V nekaterih primerih (npr. raziskave potovalnih značilnosti) pa enostavno upoštevamo kar vse osebe, ki prebivajo v Sloveniji, čeprav nimajo stalnega prebivališča.

Pri izdelavi vzorcev iz registra prebivalstva se največkrat upoštevajo osebe, ki so stalno prijavljene v Sloveniji. S tem seveda – poleg zdomcev in drugih državljanov Slovenije, ki so stalno prijavljeni v tujini – izpuščamo tudi tujce, ki živijo v Sloveniji, nimajo pa stalnega prebivališča. Po drugi strani s tem zajamemo osebe, ki niso državljani Slovenije, imajo pa v Sloveniji stalni naslov. V to kategorijo se uvrščajo – poleg nekaterih tujih poslovnežev in ostalih delovnih imigrantov – tudi državljani bivše Jugoslavije, ki iz raznih razlogov niso uredili državljanstva, čeprav nepretrgoma živijo v Sloveniji.

Pri uporabi registra prebivalstva, kjer izbiramo stalno prijavljeno rezidenčno prebivalstvo, je treba upoštevati naslednje praktične omejitve:

- Osebe, ki imajo v Sloveniji tudi začasno prebivališče, bomo težje našli na stalnem naslovu.
- Nekatere osebe iz določenih razlogov (stanovanjska pravica, maloobmejni promet ipd.) ali zaradi enostavnega odlašanja z administrativno prijavo dejansko živijo drugje, kot so prijavljene.
- Pri registraciji vitalnih in migracijskih dogodkov (rojstvo, smrt, ločitev, preselitev) obstajajo določeni administrativni zamiki, ki pa se z naraščanjem kakovosti registra hitro manjšajo. Kljub temu lahko prispevajo k določeni neažurnosti.
- Pogosto se pozablja, da se vzorec izdela bistveno prej kot raziskava, kar z naraščajočim zamikom povzroča dodatno neažurnost.

- Znaten del oseb, ki so v registru navedene s stalnim prebivališčem v Sloveniji, dejansko stalno živi v tujini.
- Osebe, ki dlje časa ali primarno živijo v ustanovah (domovi za ostarele, zapori, vojašnice, bolnišnice ipd.), predstavljajo tako imenovano nerezidenčno (*angl. nonresidential*) oziroma institucionalno populacijo (*angl. institutional population*), ki predstavlja okoli 1 % slovenskega prebivalstva. Vse te osebe so izvzete iz običajnih anket. Le najbolj zahtevne ankete (npr. Ankete o delovni sili) v nekaterih državah vključujejo tudi institucionalno populacijo.

Običajni koncept domače (rezidenčne) populacije, to je populacije, ki prebiva v gospodinjstvih, torej avtomatično izključuje institucionalno populacijo ter populacijo, ki stalno živi v tujini, zato je treba take elemente, če so bili kljub temu izbrani, obravnavati kot neustrezne. Podoben pristop velja tudi v vseh drugih primerih, ko imamo opravka z neustreznimi elementi (npr. osebe brez državljanstva ali osebe pod 18 let v volilni anketi). Kot smo že omenili, je v takih primerih posebej neprimerno uporabljati nadomeščanje neustreznih elementov s sosednjim ali naslednjim elementom na seznamu, saj s tem umetno povečujemo verjetnost izbora za vse osebe, ki so bili v vzorčnem okviru »blizu« neustreznim elementom.

Vzorčenje iz registra prebivalstva torej ne bo zajelo določenih delov prebivalstva, ki sodijo v anketo, upoštevalo pa bo nekatere segmente, ki jih ne bi smelo vključiti. Kljub navedenim pomanjkljivostim pa je register prebivalstva izjemno kakovosten vzorčni okvir, saj so pri alternativnih okvirih možnosti napak bistveno večje. Dosledno izvedena vzorčna anketa, ki vključuje tudi določeno sledenje preseljenim osebam, bo pri vzorčenju iz registra prebivalstva izpustila le nekaj odstotkov rezidenčnega prebivalstva v Sloveniji. Ker je delež tovrstnega nepokritja izredno majhen, je pristranskost odgovarjajočih cenilk za večino spremenljivk zanemarljiva. Nobenih znakov namreč ni, da so manjkajoči segmenti (npr. osebe, ki prebivajo na drugačnem naslovu od formalnega) bistveno drugačni od siceršnje populacije.

Pomembna prednost vzorca na podlagi registra prebivalstva je tudi nadzor anketarjev, saj poznamo ime, priimek in naslov anketiranca, zato lahko pisno, telefonsko ali osebno preverimo anketarjevo delo na celotnem vzorcu ali na podvzorcu.

Register prebivalstva pa ima še eno prednost: vključuje namreč tudi dodatne sociodemografske informacije, kar omogoča, da osebe izbiramo glede na določene lastnosti, npr. glede na starost in spol. Kadar nas zanimajo osebe določene starosti (npr. ankete rodnosti, ankete družin z novorojenci ipd.), je to seveda izredno koristna informacija. Na osnovi enotne identifikacije (EMŠO) pa je za statistične potrebe – vsaj načeloma – mogoče rezultate anketnih raziskav povezovati tudi z drugimi bazami podatkov (dohodnina, zaposlenost ipd.).

9. NEODGOVORI

Pravilno izveden postopek statističnega sklepanja omogoča pri verjetnostnih vzorcih izračunavanje vzorčne napake in oblikovanje intervalov zaupanja. Na tej osnovi lahko – na način, ki je razumljiv tudi statistično neizobraženim uporabnikom – kvantificiramo tveganje pri posploševanju statističnih ugotovitev iz vzorca na populacijo. Ob tem velja spomniti na osnovni predpogoj za verjetnostno vzorčenje: vsakemu elementu v vzorčnem okviru moramo vnaprej dodeliti znano in pozitivno verjetnost za izbor v vzorec. Če se pri tem pojavijo težave z vzorčnimi okviri, jih lahko rešimo s posebnimi postopki za obravnavo tovrstnih problemov. V celoti gledano torej lahko na podlagi ustreznega vzorčnega okvira vedno izdelamo tudi kakovosten verjetnostni vzorec ciljne populacije. Vendar pa vse to za dobro anketno raziskavo ne zadošča. Poleg primernega vzorčnega okvira in dosledne uporabe načel verjetnostnega vzorčenja moramo namreč za vse izbrane elemente zbrati tudi podatke.

Če podatkov od elementov, ki so za raziskavo ustrezni in so bili vključeni v vzorec, ne dobimo, imamo opraviti z neodgovori (*angl. nonresponse*). Neodgovori predstavljajo v anketnem raziskovanju resen in naraščajoč problem, saj so anketiranci vse manj pripravljeni sodelovati v raziskavah (Steh, 1981; Groves et al., 2001). Osnovna težava z neodgovori je namreč v nevarnosti, da se elementi, ki so v anketni raziskavi sodelovali, razlikujejo od ostalih ustreznih elementov, ki v raziskavi niso sodelovali. V takem primeru daje lahko sklepanje, ki temelji zgolj na odgovorih kooperativnih anketirancev, pristranske ocene populacijskih parametrov.

Da bi čimbolj nazorno predstavili potencialno nevarnost napak zaradi neodgovorov, bomo populacijo razdelili na elemente, ki v anketni raziskavi odgovarjajo – »respondente« (*angl. respondents)*, in na elemente, ki v anketni raziskavi ne odgovarjajo – »nerespondente« (*angl. nonrespondents)*. Na tem mestu izraz »respondenti« uporabljamo le v okviru specifične obravnave neodgovorov; v splošnem namreč elemente, ki so bili vključeni v anketno raziskavo, imenujemo »anketiranci«. Problematika neodgovorov je posebej podrobno obravnavana v Vehovar (2001).

Imamo torej dva stratuma, stratum respondentov in stratum nerespondentov. Ker ima v praksi slučajnost pomembno vlogo pri odločanju, ali bodo določeni elementi sodelovali v posamezni anketni raziskavi ali ne, je opisani model seveda poenostavljen, vendar je za ilustracijo problematike povsem ustrezen. Dodatno bomo v nadaljevanju pri obravnavi napak zaradi neodgovorov predpostavljali, da razpolagamo s celotno populacijo in ne samo z vzorcem.

Denimo torej, da želimo izračunati aritmetično sredino $\bar{Y}$. Zapišemo jo lahko na naslednji način:

$$\bar{Y} = W_r \bar{Y}_r + W_m \bar{Y}_m .$$

Pri tem sta $\bar{Y}_r$ oziroma $\bar{Y}_m$ aritmetični sredini v stratumu elementov, ki odgovarjajo, oziroma elementov, ki v anketni raziskavi ne odgovarjajo. Indeks r torej označuje respondente, m pa manjkajoče odgovore oziroma nerespondente. Količini W_r in W_m predstavljata odgovarjajoča populacijska deleža obeh stratumov, pri čemer velja:

$$W_r + W_m = 1.$$

V anketni raziskavi seveda ne razpolagamo z odgovori elementov iz stratuma nerespondentov, ampak le z elementi, ki pripadajo stratumu respondentov. Pristranskost (*angl. bias*) zaradi neodgovorov – to je razliko med izmerjenim $\bar{Y}_r$ in dejanskim populacijskim parametrom $\bar{Y}$ – lahko izrazimo na naslednji način:

$$Bias(\bar{Y}_r) = \bar{Y}_r - \bar{Y} = \bar{Y}_r - (W_r \bar{Y}_r + W_m \bar{Y}_m) = W_m (\bar{Y}_r - \bar{Y}_m). \qquad [26]$$

Pristranskost zaradi sklepanja na osnovi respondentov je torej odvisna od dveh faktorjev:

- delež nerespondentov v populaciji (W_m),
- razlika med aritmetičnima sredinama respondentov in nerespondentov ($\bar{Y}_r - \bar{Y}_m$).

Če sta stratuma respondentov in nerespondentov oblikovana povsem slučajno, sta aritmetični sredini respondentov in nerespondentov seveda enaki. V takem primeru pristranskost zaradi neodgovorov sploh ne obstaja. Vendar pa je v praksi nevarno predpostavljati, da elementi z manjkajočimi odgovori nastajajo povsem slučajno. Večinoma je namreč ravno nasprotno, za kar obstaja vrsta razlogov. Tako npr. premožnejše osebe težje odgovarjajo na vprašanja o dohodku, zaradi česar je razlika med ocenjenimi dohodki respondentov in nerespondentov lahko precejšnja. Najboljši način, da zmanjšamo pristranskost zaradi neodgovorov, je zato dodatni napor za skrčenje stratuma nerespondentov, s čimer bo prvi faktor W_m kar najbolj zmanjšal skupni člen $W_m (\bar{Y}_r - \bar{Y}_m)$. Na drugi faktor v navedenem členu, ki izraža populacijske razlike med respondenti in nerespondenti, namreč ne moremo vplivati.

Pri obravnavi neodgovorov v anketnih raziskavah moramo razlikovati med dvema skupinama neodgovorov. To sta:

- neodgovor elementa (*angl. unit nonresponse*), do katerega pride, kadar za izbrani element ne dobimo nobene informacije,
- neodgovor spremenljivke (*angl. item nonresponse*), kjer v fazi anketira-

nja sicer zberemo določene informacije o elementu, vendar za nekatere spremenljivke nimamo odgovorov.

Včasih namesto izraza »neodgovor elementa« uporabljamo tudi oznako »neodgovor enote«. Ker pa smo že uvodoma pojem »element« uporabili za označevanje posameznega anketiranca, pojem »enota« pa za (prostorske) združbe elementov, bomo v nadaljevanju govorili o neodgovoru elementa. Pogosto pa neodgovor elementa oziroma enote imenujemo kar neodgovor in samo pri neodgovoru spremenljivke izrecno navedemo celoten izraz.

Neodgovor elementa

V anketnih raziskavah, ki temeljijo na osebnem anketiranju, lahko neodgovor elementa delimo v naslednje kategorije:

- zavrnitev sodelovanja (*angl. refusal*), npr. pomanjkanje časa, načelna zavrnitev,
- nezmožnost sodelovanja, npr. bolezen, gluhost, nepoznavanje jezika,
- odsotnost anketiranca,
- nezmožnost locirati ustreznega anketiranca,
- nedostopnost anketiranca,
- izgubljeni ali poškodovani vprašalniki, npr. pri prevozu ali obdelavi.

Neodgovor je torej lahko posledica različnih razlogov, čeprav prevladujeta predvsem odsotnost ustreznega anketiranca ter njegova zavestna odločitev za nesodelovanje.

Med vzroki za neodgovor elementa v osebnem anketiranju prevladujejo predvsem zavrnitve in odsotnost anketiranca, ostali razlogi pa so – vsaj pri večini raziskav splošne populacije – manj pomembni. Pri verjetnostnih vzorcih z osebnim anketiranjem vsaka od obeh kategorij pogosto predstavlja približno polovico vseh neodgovorov.

V raziskavah po pošti je natančna ocena vzrokov za neodgovore bolj težavna. Nekateri anketiranci sicer sporočijo, da zavračajo sodelovanje in v določenih primerih sosedje ali sorodniki obvestijo, da je izbrana oseba npr. preveč bolna, da bi lahko odgovarjala. Podobno se lahko nekateri vprašalniki vrnejo pošiljatelju, ker naslovnik ne obstaja ali pa pošiljke ni bilo mogoče dostaviti. Pri veliki večini neodgovorov v poštnih anketah pa običajno vemo le to, da je bil vprašalnik odposlan, ne pa tudi vrnjen. Razmeroma nejasno torej ostaja, ali je to zaradi zavračanja oziroma začasne odsotnosti (neodgovori) ali zaradi napačnega naslova (neustrezni elementi). Pri obravnavi neodgovorov je namreč nadvse pomembno, da neustrezne elemente v celoti izločimo iz vseh kategorij neodgovorov.

Podobna nejasnost obstaja tudi pri telefonskem anketiranju, kjer lahko pri določenem elementu v vseh poskusih kontaktiranja naletimo na zvonenje v prazno (*angl. ring-no-answer*). V takem primeru zato ni jasno, ali gre za osebo, ki je v času anketiranja začasno odsotna (neodgovor), ali pa gre za napačno osebo oziroma prazno stanovanje (neustrezni element).

Očitno je natančno opredeljevanje neodgovorov – predvsem zaradi razmejevanja z neustreznimi elementi – v nekaterih primerih precej težavno. Ob tem velja še enkrat poudariti, da se neodgovori vedno nanašajo samo na ustrezne elemente. Če bi kot neodgovor obravnavali tudi neustrezne elemente, bi namreč porušili načelo verjetnostnega vzorčenja, saj bi ustrezni elementi, ki so v vzorčnem okviru blizu neustreznih, s tem imeli preveliko verjetnost za vključitev v vzorec.

Kot smo navedli že v poglavju o vzorčnih okvirih, je pravilna obravnava neustreznih elementov nadvse enostavna – neustrezne elemente preskočimo in jih ne upoštevamo. Postopek je torej bistveno drugačen od obravnave neodgovorov, kjer namesto manjkajočega elementa – eksplicitno ali implicitno (to je z uteževanjem) – vstavimo določeno nadomestno vrednost.

Razmejevanje med neustreznimi elementi in neodgovorom elementa je posebej kritično pri vzorčenju iz registra prebivalstva. Za osebe, ki so se preselile v drug kraj, je namreč lahko nejasno, ali so to le neodgovori ustreznih oseb – navsezadnje so to državljani s stalnim prebivališčem v Sloveniji, ki v celoti ustrezajo pogojem raziskave – ali pa so take osebe za anketno raziskavo neustrezne (prazni elementi), saj na navedenem naslovu v resnici ne prebivajo. V tovrstnih dilemah se pogosto zatečemo k arbitrarnim rešitvam, ki so primerne za določeno anketno raziskavo. Tako npr. lahko za neustrezne razglasimo samo tiste elemente, za katere anketar dobi izrecno informacijo, da ne sodijo v ciljno populacijo (npr. umrli, stalno odseljeni iz države).

Zavrnitev sodelovanja

Zavračanje sodelovanja v anketni raziskavi lahko izvira iz načelnega odpora do anketiranja, lahko pa le iz začasne preobremenjenosti oziroma pomanjkanja časa. Razlogi se seveda od ankete do ankete razlikujejo, zato je njihovo poznavanje pomembno, kadar izbiramo strategijo za zmanjšanje neodgovorov. Obseg zavrnitev je namreč v anketnih raziskavah različen in je v splošnem odvisen od vrste okoliščin, predvsem od vsebine raziskave, dolžine vprašalnika ter od spretnosti in naporov anketarjev oziroma raziskovalcev.

V splošnem je najučinkovitejši način za reševanje problema neodgovorov njihovo preprečevanje, kar še posebej velja za zavrnitve sodelovanja. Tako obstaja danes množica podrobno razdelanih postopkov za zmanjšanje števila zavrnitev. Seveda pa se tovrstne strategije razlikujejo glede na način anketiranja. Pri raziskavah z osebnim anketiranjem se zato že v fazi usposabljanja anketarji natančno poučijo o pristopih za preprečevanje zavrnitev. Tako ima lahko anketar navodilo, da se v primeru zavrnitve ponovno vrne in poskusi opraviti intervju v primernejšem času. Nemalokrat pa opravi ponovni poskus anketiranja drug anketar, ki je bolj izkušen ali celo specializiran za obravnavo zavrnitev (*angl. refusal conversion*).

Običajni pristopi k preprečevanju zavrnitev poskušajo prepričati anketirance tudi s poudarjanjem pomena raziskave, kar se največkrat utemeljuje z navedbo uglednega naročnika (npr. Univerza, Statistični urad). To je posebej pomembno pri anketnih raziskavah po pošti, kjer potencialnemu anketirancu argumentov ni mogoče posredovati interaktivno. Za pridobitev sodelovanja je ugodno tudi navajanje širših družbenih koristi, ki nastajajo s tem, ko anketiranec sodeluje v raziskavi. Po drugi strani pa se – posebej v marketinških raziskavah – sodelovanje dosega s spretnimi psihološkimi prijemi, ki poudarjajo respondentovo pomembnost. Vse pogosteje pa se uporabljajo tudi različnima darila in nagrade (*angl. incentives*).

Praviloma se anketirancem izrecno zagotovi tudi anonimnost in zaupnost podatkov, s čimer se odpravijo morebitni pomisleki in strahovi v zvezi z nadaljnjo uporabo podatkov.

Sodelovanje respondentov se zagotavlja tudi s primerno sestavljenim vprašalnikom, ki se prične z enostavnimi in privlačnimi vprašanji. S tem se izognemo tveganju, da bi anketiranec prekinil s sodelovanjem, ker bi se moral že takoj na začetku soočiti z zahtevnim vprašanjem ali takšnim, ki ga spravlja v zadrego.

Nekontakti

Poleg zavračanja sodelovanja so nekontaktirani anketiranci najobsežnejša kategorija neodgovorov. Pri tem govorimo o ustreznih elementih, ki so bili v času vseh poskusov kontaktiranja odsotni.

V primeru odsotnosti elementov v raziskavah z osebnim anketiranjem se običajno opravijo poskusi ponovnega kontaktiranja. Pri anketah, ki temeljijo na verjetnostnih vzorcih, se pogosto opravi vsaj štiri ali pet poskusov, praviloma ob različnih dnevih in različnih urah, vključno z nekaj obiski ob večeru. Anketarje se tudi vzpodbuja, da dodatno opravijo kontakt vsakič, ko se znajdejo v soseščini potencialnih anketirancev. Za povečanje možnosti, da izbrano osebo uspešno kontaktiramo, so posebej koristni tudi vnaprej dogovorjeni termini za izvedbo intervjuja.

Pri telefonskem anketiranju je nekoliko enostavneje izvajati ponovne kontakte (*angl. call-back*), zato pri verjetnostnih telefonskih vzorcih pogosto vztrajamo pri najmanj desetih poskusih kontaktiranja.

V raziskavah po pošti se postopek nadaljnjega kontaktiranja nanaša na pošiljanje dodatnih dopisov (*angl. follow-up*). Dodatni dopisi pa niso usmerjeni le proti začasno odsotnim elementom (nekontakti), ampak nasploh večajo in vzpodbujajo odziv potencialnih anketirancev. Tako elementom, ki se niso odzvali na prvi dopis, po določenem času običajno pošljemo pismo – opomnik (*angl. reminder letter).* Vsem, ki se niso odzvali tudi po dveh dopisih, pa lahko pošljemo še en opomnik, tokrat s ponovno priloženim vprašalnikom.

V tem okviru so posebej znana tako imenovana priporočila TDM (Total Design Method), kjer pošljemo prvi opomnik že po tednu dni. Pri tem se hkrati zahvalimo respondentom in vzpodbudimo nerespondente. Drugi opomnik (torej tretji dopis) pošljemo čez dva tedna (tri tedne po začetku raziskave) preostalim nerespondentom skupaj s ponovno priloženim vprašalnikom. Po potrebi izvedemo še četrti dopis, za katerega je priporočljivo, da ga opravimo z drugim načinom anketiranja (osebno, telefonsko) ali s priporočeno pošto. Podrobnosti o tem navaja Dillman (2000).

Stopnje odgovorov

Zaradi različnih kategorij neodgovorov ter nejasnosti pri razmejevanju neodgovorov od neustreznih elementov so primerjave stopenj odgovorov za različne ankete razmeroma težavne. Nemalokrat prihaja zaradi različnih opredelitev celo do večjih nesporazumov glede obsega neodgovorov. Za izražanje stopenj neodgovorov velja zato upoštevati mednarodno standardizirane kazalce. Pri tem lahko izpostavimo *Ameriško združenje javnomnenjskih raziskav* – AAPOR (http://www.aapor.org), ki je izdelalo podrobno klasifikacijo stopenj neodgovorov. Na tej osnovi podajamo v nadaljevanju pregled najpomembnejših indikatorjev na področju neodgovorov v sodobni anketni praksi.

Stopnja odgovorov (*angl. response rate*) je razmerje med številom izpolnjenih vprašalnikov (vsi elementi z odgovori – respondenti) in številom vseh ustreznih elementov, ki smo jih vključili v vzorec. Seveda se lahko izražamo tudi s komplementarnim kazalcem; tako je **stopnja neodgovorov** *(angl. nonresponse rate)* delež nerespondentov med vsemi ustreznimi elementi.

Čeprav je zgoraj navedena definicija videti enostavna, prihaja v praksi do precejšnjih težav, posebej pri obravnavi neustreznih (npr. praznih ali tujih) elementov. V skladu z zgornjo opredelitvijo bi namreč morali vse neustrezne elemente izključiti tako iz imenovalca kot iz števca. Pogosto pa organizacije, ki izvajajo anketiranje, obeh kategorij ne ločujejo pravilno. Nemalokrat je tako ločevanje tudi izredno težavno, saj ni mogoče določiti, ali je vzorčeni element neustrezen (npr. prazen) ali pa gre za neodgovor (npr. odsotnost). V naključnem telefonskem vzorcu je tako lahko številka, na kateri tudi po več klicih ni odgovora, nedelujoča oziroma poslovna številka (neustrezen element), lahko pa je to samo gospodinjstvo, kjer v vseh poskusih ni bilo nikogar doma oziroma se na klic niso odzvali (neodgovor).

Podobno je lahko npr. v raziskavi, kjer proučujemo mlajšo populacijo v starosti od 18 do 24 let, izbrano gospodinjstvo, ki ga v procesu anketiranja nismo uspeli kontaktirati. Za tako gospodinjstvo torej ne vemo, ali je neustrezno (ne vsebuje nobene osebe v starosti 18–24 let), ali pa gre za neodgovor in je to gospodinjstvo – čeprav v času anketiranja odsotno – za raziskavo ustrezno, saj vključuje npr. enega ali več elementov ciljne populacije. V praksi je v tovrstnih primerih priporočljivo sporne elemente razdeliti med neodgovore in neustrezne elemente s pomočjo ocenjenega razmerja med obema kategorijama. Pri tem lahko ocene pridobimo iz prejšnjih anket, dodatnih analiz ali na osnovi ekspertne presoje. Vsekakor pa je treba stopnje neodgovorov vsakič kritično preveriti in ugotoviti, kako so bile izračunane.

Stopnja ustreznosti (*angl. eligibility rate*) ima v števcu skupno število ustreznih elementov, v imenovalcu pa število vseh elementov vključenih v vzorec. Stopnja ustreznosti torej govori o naravi vzorčnega okvira, nikakor pa ne o uspešnosti pri pridobivanju sodelovanja.

Stopnja anketiranja (*angl. completion rate*) ima v števcu vse elemente z odgovori (respondente), v imenovalcu pa vse elemente, ki smo jih vključili v vzorec. Izračunavanje stopnje anketiranja je zato razmeroma enostavno. Seveda pa je taka stopnja tudi manj informativna. Nizka stopnja anketiranja je namreč lahko posledica vzorčnega okvira z veliko praznimi elementi in torej ne izraža zgolj (ne)uspešnosti pri pridobivanju sodelovanja. Stopnja ustreznosti in stopnja odgovorov predstavljata torej dve komponenti stopnje anketiranja, ki je v matematičnem smislu tudi produkt obeh stopenj.

Stopnja nekontaktiranja (*angl. noncontact rate*) izraža delež ustreznih elementov, ki jih nismo uspeli kontaktirati, **stopnja kontaktiranja** (*angl. contact rate*) pa je temu komplementarna mera.
Stopnja zavračanja (*angl. refusal rate*) se nanaša na delež ustreznih elementov, ki so izrecno zavrnili sodelovanje. Na podoben način lahko izračunamo stopnje tudi za ostale kategorije neodgovorov (npr. nezmožnost sodelovanja). Vsota vseh stopenj za posamezne kategorije neodgovorov pa daje – zaradi enakega imenovalca (število ustreznih elementov) – skupno stopnjo neodgovorov.

Včasih se stopnja zavračanja računa le med kontaktiranimi elementi, kar pa otežuje primerjave, zato je bolj primerno, da jo imenujemo **stopnja nesodelovanja** *(angl. noncooperation rate)*. V takem primeru je njen komplementarni kazalec **stopnja sodelovanja** *(angl. cooperation rate),* torej delež respondentov med vsemi kontaktiranimi elementi. V obeh primerih seveda upoštevamo le tiste kontaktirane elemente, ki so bili za raziskavo tudi ustrezni. Produkt stopnje kontaktiranja in stopnje sodelovanja daje stopnjo odgovorov, s čimer analiziramo obe ključni komponenti, na kateri lahko vpliva anketna organizacija.

Tabela 9: Stopnje neodgovorov

STOPNJA	ANGLEŠKI IZRAZ	ŠTEVEC	IMENOVALEC
Stopnja odgovorov	*response rate*	elementi z odgovori	ustrezni elementi
Stopnja neodgovorov	*nonresponse rate*	neodgovor elementa	ustrezni elementi
Stopnja ustreznosti	*eligibility rate*	ustrezni elementi	vsi elementi v vzorcu
Stopnja anketiranja	*completion rate*	elementi z odgovori	elementi v vzorcu
Stopnja nesodelovanja	*noncooperation rate*	zavrnitve	kontaktirani elementi
Stopnja sodelovanja	*cooperation rate*	elementi z odgovori	kontaktirani elementi
Stopnja kontaktiranja	*contact rate*	kontaktirani elementi	ustrezni elementi
Stopnja nekontaktiranja	*noncontact rate*	nekontaktirani elementi	ustrezni elementi
Stopnja zavrnitve	*refusal rate*	zavrnitve	ustrezni elementi

In kakšne so stopnje neodgovorov v praksi? Stopnje neodgovorov so predvsem izredno raznolike in odvisne od vrste dejavnikov, posebej izrazito pa se razlikujejo glede na način anketiranja.

Stopnje odgovorov pri nezapletenih anketnih raziskavah z osebnim anketiranjem le redko presežejo 70 % ali 80 %. Samo najpomembnejše – in tudi najdražje – uradne in akademske raziskave lahko z izjemnimi napori presežejo stopnjo 90 %. V marketinških raziskavah so stopnje odgovorov bistveno nižje, večinoma pod 50 %, pogosto pa jih sploh ni mogoče izračunati (npr. pri kvotnih vzorcih). Nemalokrat se take ankete zadovoljijo – posebej pri proučevanju organizacij oziroma podjetij – celo s samo nekaj odstotki odgovorov. Seveda v takih primerih težko govorimo o verjetnostnih vzorcih.

V telefonskih raziskavah so stopnje odgovorov še nekoliko nižje kot pri osebnem anketiranju, glavni vzrok za to pa je – če le opravimo dovolj poskusov kontaktiranja – predvsem večja stopnja zavrnitev sodelovanja. Pri telefonskih raziskavah prihaja v večjem obsegu tudi do prekinjenih intervjujev, ko anketiranec prekine pogovor, še preden odgovori na vsa vprašanja. V takem primeru govorimo **o delnih odgovorih** (*angl. partial response*). O delnih odgovorih seveda govorimo le med elementi z odgovori oziroma respondenti. V primeru delnega odgovora se zato včasih pojavi vprašanje, kdaj sploh še lahko govorimo o respondentih in kdaj imamo opraviti že z neodgovorom elementa oziroma z nerespondenti. V praksi običajno razvrstimo med respondente tiste elemente, ki imajo dovolj izpolnjenih vprašanj, da jih uvrstimo v vsebinsko analizo. Običajno pomeni to odgovore vsaj na demografski sklop vprašanj. Elemente, kjer je anketiranec odgovoril na tako malo vprašanj, da je celoten element neuporaben za vsebinske analize, pa uvrstimo med nerespondente (in ne med delne odgovore).

Stopnja odgovorov v raziskavah po pošti se od raziskave do raziskave močno razlikuje, od skromnih 10 % (v nekaterih marketinških raziskavah so odstotki seveda še bistveno nižji) do več kot 90 % v primerih, ko z relevantnim vprašalnikom anketiramo specifične (motivirane) populacije. Pri anketiranju splošne populacije pa dajejo ankete po pošti večinoma nižje stopnje odgovorov kot telefonske ankete. Za ankete po pošti še posebej velja, da je stopnja odgovorov odvisna od naporov, ki so bili vloženi v raziskavo (predvsem v ponovljene dopise), od ugleda sponzorjev ter od aktualnosti vsebine ankete za ciljno populacijo.

Posebej kritične so stopnje odgovorov v spletnih anketah (Vehovar et. al., 2001). Seveda pa je treba pri spletnih anketah jasno ločiti ankete na osnovi verjetnostnih vzorcev od ostalih spletnih anket. Med spletnimi anketami namreč prevladujejo priložnostne spletne ankete (npr. posamezno anketno vprašanje na spletnih straneh on-line dnevnikov) ter spletne ankete, kjer se respondenti vključijo na osnovi javnega oglaševanja, torej brez osebnega vabila (*angl. unsolicited Web surveys*), kot npr. v spletni anketi RIS (2001). Pri takih anketah so stopnje sodelovanja seveda izredno nizke. Pogosto odgovori nanje manj kot odstotek oseb, ki so bile izpostavljene takemu vabilu. V takih primerih je zato stopnje odgovorov v praksi skoraj nemogoče izračunati. Pri spletnih anketah, ki temeljijo na verjetnostnih vzorcih, pa se stopnje odgovorov lahko približajo stopnjam v poštnih anketah.

Obravnava neodgovorov

Tveganja, da pride zaradi neodgovorov do pristranskosti, ne smemo podcenjevati. Neodgovori so namreč pogosto porazdeljeni neenakomerno v določenih podskupinah oziroma razredih ciljne populacije. Tako so npr. neodgovori pri osebnem anketiranju običajno mnogo višji v mestnih središčih kot drugje. Posledica tega je, da se struktura dobljenega vzorca razlikuje od tiste v ciljni populaciji. Zelo verjetno je tudi – ni pa nujno – da bo zaradi tega prišlo do pristranskosti zaradi neodgovorov. Kadar pa se vrednosti ciljne spremenljivke v teh razredih razlikujejo, potem lahko skoraj gotovo pričakujemo tudi določeno pristranskost zaradi neodgovorov. Tako npr. bomo v anketi o potovalnih značilnostih podcenili obseg potovanj, če osebe v mestnih središčih – kjer je stopnja neodgovorov večja – potujejo bistveno več kot ostale osebe. Seveda pa lahko prihaja do pristranskosti zaradi neodgovorov tudi, kadar v sociodemografski strukturi našega vzorca ni nobenih odstopanj od populacijskih vrednosti. Nerespondenti se namreč lahko od respondentov razlikujejo samo v ciljni spremenljivki.

Kako se borimo proti pristranskosti zaradi neodgovorov? Osnovna strategija je vsekakor njihovo preprečevanje. Po tem, ko nekateri neodgovori manjkajo, namreč pravega nadomestka za manjkajočo informacijo ne moremo najti. Glavnino pozornosti za obravnavo problema neodgovorov je zato potrebno usmeriti v njihovo preprečevanje. V tem okviru obstaja vrsta podrobno izdelanih postopkov za obravnavo posameznih vrst neodgovorov pri določenem načinu anketiranja. Nekatere najosnovnejše pristope smo omenili že pri obravnavi zavrnitev in nekontaktiranja.

Določene rešitve seveda obstajajo tudi po tem, ko so podatki že zbrani. Tako lahko uporabimo postopek uteževanja (*angl. weighting adjustment*), posebej kadar je mogoče identificirati jasne podskupine oziroma razrede z različnimi stopnjami neodgovorov. Pri tem gre največkrat za sociodemografsko opredeljene množice elementov. Na osnovi tako opredeljenih razredov lahko, ob upoštevanju razkoraka med strukturo vzorca in strukturo ciljne populacije, respondentom priredimo določene uteži. Treba je poudariti, da takšni posegi popravijo predvsem sociodemografsko strukturo vzorca, kar lahko vpliva tudi na zmanjšanje pristranskosti nekaterih spremenljivk. Jasno pa se moramo zavedati, da ni nobenega zagotovila, da bo uteževanje odpravilo ali zmanjšalo pristranskost zaradi neodgovorov pri vseh spremenljivkah. Uteževanje je namreč uspešno le tedaj, ko so nerespondenti glede na ciljno spremenljivko slučajna podmnožica v razredu, kjer želimo opraviti uteževanje. To pa je v praksi razmeroma redko. Bolj običajno je, da uteževanje delno odpravi pristranskost le pri nekaterih spremenljivkah, kar seveda ne rešuje celotnega problema neodgovorov.

Pri uteževanju tudi ne gre pozabiti, da dodatno povečuje vzorčno varianco. Uspešnost postopka uteževanja se zato od primera do primera močno razlikuje in ne more biti nadomestilo za prizadevanja, s katerimi moramo v fazi zbiranja podatkov doseči čim višje stopnje odgovorov.

Podobne učinke kot uteževanje imajo tudi postopki vstavljanja (*angl. imputation)* nadomestnih elementov. V takem primeru namesto manjkajočih elementov (neodgovorov) vstavljamo celoten nadomestni element, ki ga v praksi največkrat izberemo kar med respondenti iste ankete. Seveda pri tem običajno izvedemo določen izbor nadomestnih elementov znotraj določenih razredov. Tako npr. namesto neodgovora v določeni regiji vstavimo slučajno izbranega respondenta iz iste regije. Učinki tovrstnega vstavljanja so podobni kot pri uteževanju, pomembna primerjalna prednost vstavljanja pa je, da razpolagamo s popolnjeno matriko podatkov. Ob tem se razvijajo tudi mnoge nadvse obetavne metode modeliranja neodgovorov (Rubin, 1988), posebej metoda večkratnega vstavljanja (*angl. multiple imputation*).

V nekaterih okoljih se v primeru neodgovorov opravi vstavljanje že v fazi zbiranja podatkov, in sicer z anketiranjem nadomestnih elementov oziroma »rezervnih« elementov (*angl. field substitutions*). Tak postopek skriva nemajhne nevarnosti (Vehovar, 1999), zato pri verjetnostnem vzorčenju – z nekaj redkimi izjemami – ni priporočljiv. Težave z nadzorom nad anketarji, podaljševanje faze zbiranja podatkov, dodatna pristranskost v cenilkah in številne druge slabosti namreč ne odtehtajo odgovarjajočih prednosti. Pri tem velja posebej poudariti, da so prednosti takega postopka običajno zgolj navidezne (Vehovar, 1995a). Izjema so lahko nekateri telefonski vzorci (Brick in Waksberg, 1992) in modelsko nadomeščanje elementov na osnovi večkratnega vstavljanja (Rubin in Zanutti, 2001).

Neodgovor spremenljivke

Neodgovor spremenljivke se kaže kot manjkajoči podatek v zapisu določenega odgovora (spremenljivke) pri elementu, ki je sicer sodeloval v anketni raziskavi. **Stopnjo odgovora spremenljivke** (*angl. item response rate*) izražamo z razmerjem med številom elementov, za katere pri obravnavani spremenljivki razpolagamo z odgovorom, ter številom vseh ustreznih elementov, ki so v raziskavi na to vprašanje odgovarjali. **Stopnja neodgovora spremenljivke** (*angl. item nonresponse rate*) pa je komplementarna mera, ki izraža delež elementov z manjkajočo vrednostjo pri obravnavani spremenljivki med vsemi za to spremenljivko ustreznimi elementi.

Vzroki za neodgovor spremenljivke so lahko različni:

- anketiranci morda ne poznajo odgovorov na nekatera vprašanja,
- anketiranci posredno zavrnejo odgovor na določeno vprašanje, ker so preveč občutljiva, jih spravljajo v zadrego ali pa menijo, da niso relevantna za cilje raziskave,
- anketar lahko zaradi naglice, nespretnosti in drugih vzrokov vprašanje preskoči ali pozabi zapisati odgovor,
- odgovor je sicer lahko vpisan v vprašalnik, vendar je izločen, ker ni bil konsistenten z drugimi odgovori,
- nekateri odgovori se lahko v procesu obdelave podatkov izgubijo.

Obseg neodgovora spremenljivke je različen in variira glede na naravo spremenljivke ter glede na način zbiranja podatkov. Pogosto imajo preprosta demografska vprašanja zanemarljivo majhno stopnjo neodgovora spremenljivke, pri vprašanjih o dohodkih pa lahko manjka 10–30 % ali celo več odgovorov. Tudi pri izjemno občutljivih ali pa težkih vprašanjih je lahko stopnja neodgovora spremenljivke razmeroma visoka.

Eden od običajnih pristopov k reševanju neodgovora spremenljivke je omejitev analize samo na elemente z vsemi odgovori. Pri univariatnih analizah moramo v takem primeru upoštevati vse omejitve, ki smo jih uvodoma navedli pri obravnavi neodgovorov, vključno z uporabo izraza (26) za pristranskost. Seveda velja upoštevati, da se v primeru neodgovora spremenljivke oznaka W_m nanaša le na delež elementov, ki niso odgovorili na določeno vprašanje, in ne na število vseh nerespondentov. Težave zaradi neodgovora spremenljivke so v tem primeru torej povsem enake kot pri neodgovoru celotnega elementa.

Kadar imamo manjkajoče podatke pri večjem številu vprašanj (spremenljivk), je zoževanje analize samo na popolne elemente – torej tiste, ki so odgovorili na vsa vprašanja – nadvse vprašljivo. Z izločanjem vsakega elementa, ki ima pri vsaj enem vprašanju manjkajoči podatek (neodgovor spremenljivke), namreč lahko odstranimo prevelik del respondentov. V takih primerih so zato primernejše drugačne rešitve.

V splošnem neodgovor spremenljivke najpogosteje obravnavamo z enostavnimi postopki vstavljanja (*angl. imputation*), ki smo jih omenili že pri obravnavi neodgovorov celotnega elementa. Problem neodgovorov spremenljivke je mogoče reševati tudi z metodami neposredne analize, kot je npr. EM algoritem (Rubin, Stern in Vehovar, 1995). Pri obravnavi neodgovora spremenljivke v praksi pa gre največkrat predvsem za enostavne postopke, ki vstavijo vrednosti za manjkajoče odgovore s pomočjo odgovorov na ostala vprašanja in tudi na osnovi drugih pomožnih informacij (*angl. auxilary information*).

Tako npr. pogosto razdelimo vzorec v razrede na podlagi odgovorov na druga relevantna vprašanja (običajno so to sociodemografske spremenljivke), nato pa pri manjkajočih vrednostih (neodgovorih spremenljivke) vstavimo aritmetično sredino, ki smo jo izračunali pri respondentih tega razreda. Tak pristop odstrani razlike v stopnji neodgovorov v različnih razredih in je, kar zadeva ocenjevanje populacijske aritmetične sredine, enakovreden uteževanju. Slabost takega postopka pa je v tem, da izkrivlja porazdelitev spremenljivke, saj ustvarja gostitve okoli aritmetičnih sredin znotraj posameznih razredov. Porazdelitev spremenljivke je zato umetno spremenjena, zmanjšana pa je tudi elementarna varianca. Ena od rešitev je v tem, da se vsaki vstavljeni vrednosti (aritmetični sredini določenega razreda) prišteje še določena slučajna napaka.

Posebej pogost in nadvse praktičen postopek vstavljanja so razvili na Ameriškem uradu za popise (U.S. Bureau of the Census). Gre za tako imenovane »hot deck« postopke vstavljanja. Pri tej metodi znotraj razredov namesto manjkajočih vrednosti vstavljamo vrednosti respondentov, ki smo jih zabeležili v isti raziskavi. Elemente znotraj izbranih razredov zato razvrstimo (sortiramo) v določeno zaporedje in namesto neodgovora spremenljivke vsakič vstavimo zadnjo vrednost respondenta, ki je na to vprašanje veljavno odgovoril. Pred tem seveda premišljeno opredelimo razrede, znotraj katerih izvajamo takšno vstavljanje. Opisan pristop je posebej privlačen zaradi enostavne računalniške obdelave, njegova slabost pa je – poleg podcenjevanja variance – tudi v tem, da lahko določeno vrednost pripišemo večim elementom. Navedene slabosti rešujejo bolj zapleteni postopki vstavljanja, kot npr. vstavljanje v anketi o dohodkih (Welniak in Coder, 1980) ter metoda večkratnega vstavljanja (Rubin, 1988).

Manjkajoče vrednosti, ki nastajajo zaradi neodgovora spremenljivke, lahko nadomestimo tudi z ocenami na osnovi modelov. Tako lahko uporabimo npr. regresijsko enačbo. Za določitev regresijskih koeficientov uporabimo odgovore respondentov na druga vprašanja (spremenljivke), namesto manjkajočih vrednosti pa nato vstavimo ocene, ki jih pri elementih z neodgovorom spremenljivke dajejo rezultati regresije. Tudi v tem primeru pa nastaja slabost zmanjševanje elementarne variance in gostitev vstavljenih vrednosti okoli regresijske krivulje. Podobno kot pri vstavljanju aritmetične sredine znotraj razredov se lahko proti temu borimo z dodajanjem naključnih ostankov (*angl. residuals*), ki izhajajo iz regresijske enačbe.

Ponoviti velja, da je ena od najpomembnejših prednosti postopkov vstavljanja dejstvo, da na tej osnovi razpolagamo s celotno in popolnjeno matriko podatkov, brez vseh manjkajočih vrednosti. Pri mnogih analizah je to izredno praktično, posebej pri multivariatnih metodah. Seveda pa se moramo pri analizi podatkov vsakič jasno zavedati, da smo uporabili postopek vstavljanja. Vstavljene vrednosti je treba tudi nedvoumno označiti, tako da jih lahko v toku analize razlikujemo od dejanskih odgovorov. Podatkov, ki vsebujejo tudi vstavljene vrednosti, zato ne smemo nekritično analizirati in jih obravnavati enako kot popolne podatke.

Ob koncu velja dodati, da v statističnem smislu vsaka uporaba vstavljanja vpliva tudi na povečanje vzorčne variance in s tem povzroča širitev intervalov zaupanja. Postopki vstavljanja namreč povečujejo vzorčno varianco podobno – nemalokrat pa še bolj – kot postopki uteževanja. Za razliko od uteževanja pa dodatno povečanje v vzorčni varianci v primeru vstavljanja težje ocenimo, predvsem pa se ga bistveno manj zavedamo. Še več, enostavne ocene vzorčne variance iz popolnjenih podatkov celo kažejo, da se vzorčna varianca po vstavljanju zmanjšuje, kar je zgolj posledica umetnega povečanja velikosti vzorca. Seveda takih ocen ne smemo uporabljati brez ustreznih popravkov.

Poleg podcenjene vzorčne variance, se je treba sprijazniti, da določena pristranskost zaradi neodgovorov, kljub vstavljanju, vsekakor ostaja. Izkušnje namreč kažejo, da metode za obravnavo neodgovorov – vključno z vstavljanjem – le redko uspejo odstraniti več kot polovico pristranskosti zaradi neodgovorov. Po drugi strani pa lahko vstavljanje, posebej če je narejeno površno, močno popači povezave med spremenljivkami in s tem škodljivo vpliva na korelacije ter multivariatne analize (Kalton in Kasprzyk, 1982).

10. ANALIZA RAZISKAVE

Za analizo podatkov iz vzorčnih raziskav so na voljno številne statistične metode. Pri tem standardni paketi za statistično analizo največkrat predpostavljajo, da so bili podatki zbrani s SRS vzorcem. Seveda pa taka predpostavka pogosto pomeni povsem nerealno poenostavitev, saj večina anketnih raziskav temelji na kompleksnih vzorcih. Dejstvo, da imamo kompleksni vzorčni načrt in ne SRS vzorca, ima namreč nemajhne posledice za celotno statistično analizo. V nadaljevanju seveda ne moremo obravnavati vseh vplivov kompleksnega vzorčnega načrta na uporabo statističnih metod. Dotaknili se bomo le nekaterih bistvenih posebnosti, ki jih je treba upoštevati pri statistični analizi kompleksnih vzorcev. Osredotočili se bomo predvsem na uporabo uteži in na izračunavanje vzorčnih napak.

Uteževanje

Uteži dajejo nekaterim elementom v vzorcu večji relativni pomen kot drugim. Največkrat je to potrebno zato, ker so bili elementi izbrani z različnimi verjetnostmi, pa tudi zaradi drugih razlogov (npr. poststratifikacija, neodgovori ipd).

Neenake verjetnosti izbora

Začnimo z analizo uporabe uteži pri ne-EPSEM vzorčnem načrtu. Za primer uteževanja vzemimo vzorec desetih elementov v raziskavi o številu učbenikov med študenti. Denimo, da je edini dostopni vzorčni okvir spisek študentov za vsako od študijskih smeri. Če npr. študent obiskuje dve študijski smeri, se njegovo ime v končnem seznamu seveda pojavlja dvakrat. S skupnega seznama izberemo s sistematičnim vzorčenjem vzorec zapisov in nato vključimo ustreznega študenta. Denimo, da je bilo vseh zapisov ravno N = 970, zato daje, ob izbranem naključnem začetku, izbor vsakega 97. zapisa ravno vzorec desetih zapisov, n = 10. Imamo torej:

$$N = 970 \quad \text{in} \quad f = \frac{10}{970} = \frac{1}{97}.$$

Ker večina študentov obiskuje več kot eno študijsko smer, zgoraj opisani EPSEM vzorec zapisov seveda ne določa tudi EPSEM vzorca študentov.

Pri večjem številu smeri, ki jih obiskuje določen študent, ima namreč tak študent tudi večjo verjetnost za izbor v vzorec.

Vzemimo, da želimo oceniti aritmetično sredino za število kupljenih učbenikov. Število učbenikov, ki jih je kupilo 10 študentov, in število študijskih smeri, ki so jih ti študentje vpisali, sta razvidni iz tabele 10.

Tabela 10: Primer uteževanja študentov glede na število študijskih smeri

Številka študenta	Število učbenikov (y_i)	Število študijskih smeri (r_i)	Utež: $w_i = \frac{12}{r_i}$	$u_i = w_i y_i$
1	2	1	12	24
2	5	2	6	30
3	6	3	4	24
4	8	3	4	32
5	3	2	6	18
6	7	4	3	21
7	6	4	3	18
8	3	2	6	18
9	5	3	4	20
10	2	2	6	12
Vsota	47	-	54	217

Enostavna vzorčna aritmetična sredina za število kupljenih učbenikov je:

$$\bar{y}_b = \sum \frac{y_i}{n} = \frac{47}{10} = 4.70.$$

Zgornja ocena za populacijsko aritmetično sredino števila učbenikov je seveda pristranska. Verjetnosti izbora so namreč neenake, obstaja pa tudi povezava med verjetnostjo izbora in številom kupljenih knjig. Če podrobneje pogledamo tabelo 10, ugotovimo, da več študijskih smeri, ki jih obiskuje določen študent, pomeni tudi več kupljenih učbenikov. Zato enostavna ocena aritmetične sredine precenjuje populacijsko aritmetično sredino. Da bi vzorčni načrt popravili, potrebujemo uteži, ki so obratno sorazmerne z verjetnostjo izbora. Če je verjetnost izbora elementa p_i, mora biti utež enaka:

$$\frac{k}{p_i},$$

pri čemer je k konstanta, ki jo določimo naknadno.

Če vzamemo za izhodiščno izbiro $k = 1$, je utež:

$$w_i = \frac{1}{p_i}.$$

Ker so bili zapisi vzorčeni v razmerju 1 proti 97, je torej verjetnost p_i za izbor študenta sorazmerna $r_i/97$, pri čemer je r_i število študijskih smeri, ki jih obiskuje *i*-ti študent. Uteži so v našem primeru enake:

$$w_i = \frac{1}{p_i} = \frac{97}{r_i}.$$

Študent 1 bi imel tako utež 97, študent 2 utež 48.5, študent 3 utež 32.3 itd.

Izbira $k = 1$ je uporabna, če potrebujemo cenilke za populacijske vsote, kot npr. celotno število učbenikov, ki so jih kupili študentje fakultete. Pri $k = 1$ namreč populacijsko vsoto *(angl. population total)* ocenimo z uteženo vsoto v vzorcu:

$$\hat{Y} = \sum_{i=1}^{n} w_i y_i .$$

Pogosto pa so verjetnosti za izbor elementa majhna števila, s katerimi je težko operirati. V tem primeru je možna alternativna izbira vrednosti *k*, s katero poenostavimo uteži, kar si bomo za tabelo 10 ogledali v nadaljevanju. Kadar pa izberemo vrednost $k \neq 1$, moramo tudi uteženo vsoto spremenljivke *y* v izrazu $\sum_i w_i y_i$ deliti s *k*, da dobimo pravilno oceno populacijske vsote. Tovrstni popravki pa niso potrebni za aritmetične sredine, odstotke in druge statistike, ki temeljijo na aritmetičnih sredinah oziroma razmerjih vzorčnih vrednosti.

Drugi način določanja uteži je enostavno enačenje uteži z obratnimi vrednostmi števila smeri, ki jih študent obiskuje, to je $1/r_i$, saj ravno število študijskih smeri povzroča neenako verjetnost izbora. Uteži so v tem primeru enake 1.00 za prvega študenta, 0.50 za drugega, 0.33 za tretjega itd. Opisana shema torej implicitno uporablja vrednost $k = 1/97$ in je zato povsem ustrezna, vendar zahteva nekaj zaokroževanja, npr. pri 1/3 se moramo zadovoljiti s približkom 0.33. Da bi se izognili zaokroževanju, so uteži v tabeli 10 določene z vrednostjo $12/r_i$, kar pomeni, da smo implicitno privzeli $k = 12/97$.

Kadar uporabljamo uteži, je vzorčna aritmetična sredina v splošnem opredeljena kot:

$$\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}. \qquad [26a]$$

V našem primeru imamo zato:

$$\bar{y}_w = \frac{217}{54} = 4.02,$$

kar se precej razlikuje od pristranske ocene na osnovi enostavne aritmetične sredine SRS vzorca $\bar{y}_0 = 4.70$, kjer nismo upoštevali neenakih verjetnosti za vključitev v vzorec.

V nadaljevanju bomo – podobno kot v (22a) – vsoto uteži v vzorcu označili:

$$w = \sum_{i=1}^{n} w_i \, . \qquad [26b]$$

Zgornji izraz seveda ni konstanta, ampak med vzorci variira, zato je utežena vzorčna aritmetična sredina $\bar{y}_w$ v resnici razmernostna cenilka. Kot smo že omenili (23), je razmernostna cenilka pristranska cenilka populacijske aritmetične sredine, vendar je pristranskost zanemarljiva, če je koeficient variacije imenovalca manjši od 0.1 oziroma 0.2.

Varianco vsote uteži lahko ocenimo kot vsoto n neodvisnih varianc za posamezne elemente v vzorcu:

$$v(w) = v\left(\sum_{i=1}^{n} w_i\right) = n s_w^2 = n \sum_{i=1}^{n} \frac{(w_i - \bar{w})^2}{n-1} = \frac{624}{9} = 69.33 \, .$$

Pri tem je s_w^2 seveda ocena za S_w^2 (elementarna varianca uteži):

$$s_w^2 = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \bar{w})^2 ,$$

aritmetična sredina uteži v vzorcu pa je enaka:

$$\bar{w} = \sum \frac{w_i}{n} = \frac{w}{n} = \frac{54}{10} = 5.4 \, .$$

Ocenjeni koeficient variacije za velikost vzorca (oziroma za vsoto uteži) je zato:

$$cv(w) = \frac{\sqrt{v(w)}}{w} = \frac{8.327}{54} = 0.15 \, . \qquad [26c]$$

Koeficient variacije uteži bomo v nadaljevanju še podrobneje obravnavali, npr. (30).

Na tem mestu velja izpostaviti neugodno podvajanje označevanja, ki pogosto nastaja v statistični literaturi, npr. Kish (1965). Oznaka »w« namreč pomeni spremenljivko uteži, hkrati pa tudi njihovo vsoto (26b), kar smo uvedli že pri obravnavi razmernostne spremenljivke (22a). Tokrat bomo – zaradi ločevanja od koeficient variacije za vsoto spremenljivke w (26c) – elementarni koeficient variacije spremenljivke w, ki ga v splošnem sicer označujemo z $cv(w)$, dodatno označili:

$$cv_e(w) = \frac{\sqrt{s_w^2}}{\bar{w}} = 0.48 \, . \qquad [26d]$$

Elementarni koeficient variacije $cv_e(w)$ in koeficient variacije $cv(w)$ za vsoto w sta v naslednjem odnosu:

$$cv_e(w) = \frac{\sqrt{s_w^2}}{\bar{w}} = \frac{\sqrt{v(w)/n}}{w/n} = \frac{\sqrt{v(w)}}{w} \times \frac{n}{\sqrt{n}} = \sqrt{n} \times cv(w). \qquad [26e]$$

Koeficient variacije za velikost vzorca $cv(w)$ pa je enak tudi koeficientu variacije za cenilko aritmetične sredine uteži:

$$cv(\bar{w}) = \frac{se(\bar{w})}{\bar{w}} = \frac{se(\bar{w})}{\bar{w}} \times \frac{n}{n} = \frac{\sqrt{n^2 v(\bar{w})}}{n\bar{w}} = \frac{\sqrt{v(n\bar{w})}}{n\bar{w}} = cv(w).$$

Čeprav je v našem primeru koeficient variacije za velikost vzorca $cv(w)$ večji kot 0.1, je še vedno dovolj majhen, da zagotavlja sprejemljivo pristranskost razmernostne cenilke aritmetične sredine. Ker se koeficient variacije $cv(w)$ nadalje zmanjšuje z večanjem velikosti vzorca, je namreč to pri nekoliko večjem vzorcu v praksi vsekakor mogoče zanemariti.

Oglejmo si še oceno vzorčne variance. Ker ima utežena aritmetična sredina slučajni spremenljivki v števcu in v imenovalcu, moramo uporabiti izraz za varianco razmernostne cenilke (23). Zato najprej zapišemo:

$$\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i} = \frac{\sum u_i}{\sum w_i} = \frac{u}{w},$$

pri čemer je spremenljivka u_i definirana kot $u_i = w_i y_i$. Če je koeficient variacije uteži manjši od 0.2, je približna cenilka variance za $\bar{y}_w$ enaka:

$$v(\bar{y}_w) \approx \frac{v(u) + \bar{y}_w^2 v(w) - 2\bar{y}_w c(u,w)}{w^2},$$

kar izhaja iz izraza (23). Če uporabimo podatke iz tabele 10 in upoštevamo $u = 217/10 = 21.7$, dobimo:

$$v(u) = n\sum \frac{(u_i - \bar{u})^2}{n-1} = 360.11,$$

$$c(u,w) = n\sum \frac{(u_i - \bar{u})(w_i - \bar{w})}{n-1} = 5.78.$$

Imamo torej:

$$v(\bar{y}_w) = \frac{1}{54^2}\left[\frac{3241}{9} + \left(\frac{217}{54}\right)^2 \times \frac{624}{9} - 2\times\left(\frac{217}{54}\right)\times\left(\frac{52}{9}\right)\right] = 0.4915$$

in $se(\bar{y}_w) = 0.70$.

Koristno je primerjati natančnost obravnavanega ne-EPSEM vzorca z enostavnim slučajnim vzorcem (SRS) enake velikosti. V ta namen potrebujemo cenilko elementarne variance za število kupljenih učbenikov. Ocena te količine je zaradi neenakih verjetnosti izbora nekoliko drugačna kot v navadnem SRS vzorcu (1a):

$$s_w^2 = \frac{n}{n-1}\frac{\sum w_i (y_i - \bar{y}_w)^2}{\sum w_i}. \qquad [27]$$

Pri SRS vzorcu velikosti n = 10 je torej vzorčna varianca aritmetične sredine ob neupoštevanju faktorja končne populacije (*FPC*) ocenjena z izrazom:

$$v(\bar{y}_0) = \frac{s_w^2}{n} = 0.4382.$$

Ocenjeni vzorčni učinek v ne-EPSEM vzorcu je zato:

$$deff(\bar{y}_w) = \frac{v(\bar{y}_w)}{v(\bar{y}_0)} = \frac{0.4915}{0.4382} = 1.12, \qquad [27a]$$

kar nakazuje povečanje variance za 12 % zaradi neenakih verjetnosti izbora. Pri velikih razlikah v verjetnostih za izbiro elementov v vzorec se lahko natančnost zelo zmanjša, zato se moramo velikemu variiranju v utežeh izogibati.

Skupine elementov

Uteževanje potrebujemo tudi pri odpravljanju težav z vzorčnimi okviri, posebej v primeru skupin elementov. Vzemimo, da je EPSEM vzorec a stanovanj izbran med A stanovanji v nekem mestu, pri čemer iz vsakega izbranega stanovanja z uporabo Kisheve sheme (Kish, 1965) določimo eno odraslo osebo. Verjetnost, da bo v vzorec izbrana oseba β v stanovanju α, je dana z enačbo izbora:

$$P(\alpha\beta) = P(\alpha) \times P(\beta \mid \alpha) = \frac{a}{A} \times \frac{1}{B_\alpha},$$

kjer je B_α število oseb v stanovanju α. Da bi upoštevali neenake verjetnosti izbora, moramo uporabiti uteži, ki so sorazmerne z:

$$w \propto \frac{1}{P(\alpha\beta)} = \frac{AB_\alpha}{a}.$$

Logična izbira v tem primeru je torej utež, ki je sorazmerna s številom oseb v stanovanju (B_α). Čeprav so te uteži teoretično potrebne, pa se v praksi ne uporabljajo pogosto, ker je variiranje števila odraslih oseb na stanovanje razmeroma majhno, povezanost s ciljnimi spremenljivkami pa je majhna. Zato imajo največkrat le zanemarljiv vpliv na rezultate raziskave (Kish, 1965: 400).

Disproporcionalna stratifikacija

Oglejmo si še en primer, pri katerem potrebujemo uteževanje – disproporcionalno stratifikacijo. V poglavju o stratifikaciji (6b) smo pokazali, da lahko cenilko populacijske aritmetične sredine izračunamo s pomočjo vzorčne aritmetične sredine v vsakem stratumu, nakar ocene kombiniramo v uteženo aritmetično sredino:

$$\bar{y}_{st} = \sum W_h \bar{y}_h \, .$$

Drug način za dosego tega rezultata pa je, da vsakemu izbranemu elementu pripišemo ustrezno utež. Utež je v tem primeru enaka za vse elemente v istem stratumu, vendar različna med stratumi. Na tako uteženem vzorcu uporabimo običajno oceno (26a) za uteženo aritmetično sredino $\bar{y}_w$. Uteži so tudi tokrat določene obratno sorazmerno z verjetnostjo izbora elementov v stratumu, to je:

$$w_{hi} \propto \frac{N_h}{n_h} = k\frac{N_h}{n_h},$$

za vse elemente v stratumu *h*.

Sledi:

$$\bar{y}_w = \frac{\sum_h \sum_i w_{hi} y_{hi}}{\sum_h \sum_i w_{hi}} = \frac{\sum_h k N_h \bar{y}_h}{\sum_h k N_h} = \sum_h W_h \bar{y}_h \, .$$

Očitno je, da sta oceni $\bar{y}_w$ in $\bar{y}_{st}$ enakovredni. Poudariti pa je treba, da $\bar{y}_w$ tokrat ni razmernostna cenilka, ker je imenovalec konstanten. Prednost uporabe $\bar{y}_w$ namesto $\bar{y}_{st}$ je predvsem enostavnost pri računanju. Ko so uteži določene, namreč lahko za statistično analizo uporabimo standardne računalniške programe.

Poststratifikacijske uteži

Podobno se lahko uteževanje uporablja tudi pri tehniki stratifikacije po opravljenem zbiranju podatkov – poststratifikacija (*angl. poststratification*). Pri tej tehniki lahko znanje o populacijski porazdelitvi zunanjih spremenljivk uporabimo za izboljšanje natančnosti cenilk. Če je npr. porazdelitev starosti za osebe v ciljni populaciji znana iz popisa prebivalcev, lahko razdelimo vzorec v starostne skupine, nato pa izračunamo aritmetično sredino $\bar{y}_h$ ciljne spremenljivke za vsako starostno skupino posebej. Dobljene aritmetične sredine kombiniramo v splošno poststratificirano oceno:

$$\bar{y}_{pst} = \sum_h W_h \bar{y}_h ,$$

kjer je W_h delež populacije v skupini h. Kot pri disproporcionalni stratifikaciji (6b) je lahko tudi pri poststratifikaciji aritmetična sredina izražena kot utežena aritmetična sredina, kjer je vsakemu elementu pripisana utež sorazmerno z izrazom:

$$\frac{N_h}{n_h} .$$

Če za trenutek pozabimo na morebitne težave zaradi neodgovorov in nepokritja, ugotovimo, da poststratifikacija v celoti prilagodi vzorčno porazdelitev – ki je pri vsakem vzorcu podvržena naključnim odstopanjem – dejanski populacijski porazdelitvi glede na stratume. S tem smo v pogledu obravnavane spremenljivke vzorec povsem uskladili z znano populacijsko porazdelitvijo.

Seveda pa tovrstna korekcija rezultatov dodatno povečuje vzorčno varianco. Vendar je v primeru, ko je v poststratifikacijskih celicah vsaj deset ali več elementov, varianca poststratificirane vzorčne aritmetične sredine približno enaka varianci aritmetične sredine (10) v proporcionalno stratificiranem vzorcu (Cochran 1977, 5A.9).

Poststratifikacija je uporabna predvsem takrat, kadar poznamo populacijske deleže W_h, vendar elementom ne moremo vnaprej določiti stratuma, ki mu pripadajo. V tem primeru zato ne moremo opraviti predhodne stratifikacije, vendar pa lahko od elementov, ki so bili vključeni v raziskavo, zberemo informacije, ki omogočajo uvrstitev v stratume in s tem uporabo poststratifikacije. Poststratifikacijo lahko uporabimo tudi za izkoriščanje dodatnih stratifikacijskih faktorjev poleg tistih, ki smo jih uporabili pri načrtovanju vzorca. Tako kot pri proporcionalni stratifikaciji se tudi pri poststratifikaciji natančnost povečuje, če obstaja glede na ciljne spremenljivke heterogenost med stratumi in homogenost znotraj stratumov.

Uteževanje vzorca v skladu z znano populacijsko porazdelitvijo pa ne odpravlja le slučajnih odstopanj v strukturi vzorca, ampak pripomore tudi k reševanju težav zaradi neodgovorov in nepokritja. Če je npr. stopnja neodgovorov višja med mladimi ali pa mladi niso bili v zadostnem obsegu vključeni v vzorčni okvir, lahko z uteževanjem vzorca in uskladitvijo z znano populacijsko porazdelitvijo uravnotežimo te dejavnike. Seveda pa s tem vplivamo le na razmerja med starostnimi skupinami ter na tiste spremenljivke, ki so odvisne od starosti. Ker pa pogosto obstajajo – kot smo omenili že pri obravnavi neodgovorov – nadaljnje razlike med respondenti in nerespondenti tudi znotraj vsake starostne skupine, bo določena pristranskost ostala kljub uteževanju.

V primeru, ko želimo opraviti poststratifikacijo z več kontrolnimi spremenljivkami hkrati, pa se pogosto srečujemo z dejstvom, da nimamo ustreznih populacijskih podatkov. Tako npr. lahko razpolagamo s populacijsko strukturo za starost in tudi za izobrazbo, nimamo pa podatkov za posamezne starostno-izobrazbene razrede. V primeru, ko pa razpolagamo z bolj podrobnimi podatki, se lahko zgodi, da v vzorcu v ustreznih razredih oziroma celicah nimamo dovolj elementov. V določenem obsegu lahko navedeni težavi rešujemo s tako imenovanim postopkom »raking«, kjer vzorec iterativno prilagajamo posameznim spremenljivkam.

Uteževanje zaradi neodgovorov in nepokritja

Podobno kot poststratifikacija zahtevajo tudi popravki zaradi neodgovorov oziroma nepokritja poznavanje populacijske porazdelitve kontrolnih spremenljivk. Le v posebnih primerih lahko opravimo uteževanje tudi samo na podlagi informacij v vzorcu, kadar so dostopne tako za respondente kot za nerespondente. Pri takem načinu uporabljamo predvsem informacije o stratumih ali vzorčnih enotah prve stopnje (PSU), v katerih so elementi.

Denimo, da je vzorec razdeljen na geografska območja, ki so nadalje razdeljena v razrede glede na tip naselja (mesto, primestje, vas). Z EPSEM vzorcem lahko popravke za neodgovore izvedemo z utežmi:

$$w_h = \frac{n_h}{r_h},$$

ki jih pripišemo vsem respondentom v razredu h, kjer je n_h celotna velikost vzorca na območju h, odgovarjajoče število odgovorov v vzorcu pa je označeno z r_h. Tovrstno uteževanje prilagodi porazdelitev respondentov dejanski porazdelitvi celotnega vzorca znotraj razredov, tako da respondenti v vsakem razredu predstavljajo tudi nerespondente. Seveda pa lahko tako uteževanje uporabljamo le pri neodgovorih, ne pa pri nepokritju, saj elementi, ki manjkajo v vzorčnem okviru, ne morejo biti vključeni v števec n_h.

Dodajmo, da se zgoraj opisano uteževanje lahko uporablja tudi pri vzorčenju v skupinah, kjer v prostorski enoti zadnje stopnje (npr. v popisnem okolišu) uspemo anketirati le r_α od načrtovanih n_α elementov.

V praksi lahko postane uteževanje precej zapleteno, saj je pogosto potrebna kombinacija različnih tipov uteži. Uteži namreč uporabljamo za popravke neenakih verjetnosti izbora, za neodgovore in za prilagoditev vzorčne porazdelitve znani populacijski porazdelitvi. Pri oblikovanju uteži je zato potrebna velika pozornost, saj zlahka pride do resnih napak.

Izračunavanje vzorčnih napak

V predstavitvah različnih vzorčnih načrtov smo videli, da je velikost vzorčne napake odvisna od velikosti in tudi od kompleksnosti konkretnega vzorca. Kljub temu pa v praksi pri izrazih za izračun standardne napake pogosto predpostavljamo SRS vzorce s ponavljanjem, čeprav imamo opraviti s kompleksnimi vzorci. Seveda izrazov za SRS vzorce s ponavljanjem ne bi smeli nekritično uporabljati pri drugih vzorčnih načrtih, saj s tem lahko povzročimo precenjevanje ali – kar je bolj neugodno in bolj pogosto – podcenjevanje vzorčne napake.

Predpostavka o SRS vzorcu
Pri enostavnem slučajnem vzorčenju brez ponavljanja (SRS) je varianca vzorčne aritmetične sredine (2) manjša od variance aritmetične sredine vzorca v enostavnem slučajnem vzorcu s ponavljanjem za faktor končne populacije (*FPC*):

$$FPC = 1 - f = 1 - \frac{n}{N}.$$

Pri velikih populacijah je vzorčni delež f običajno majhen, zato je *FPC* praktično enak 1. V tem primeru lahko uporabimo izraz za standardno napako na podlagi SRS vzorca s ponavljanjem tudi pri SRS vzorcu brez ponavljanja.

Proporcionalno stratificiran vzorčni načrt s SRS vzorčenjem znotraj stratumov daje cenilke, ki so vsaj tako natančne kot pri SRS načrtu. Cenilke so natančnejše, kadar je stratum notranje homogen glede na ciljne spremenljivke. Uporaba izraza za standardno napako ob predpostavki SRS vzorčenja bo zato pri proporcionalni stratificikaciji povzročila določeno precenjevanje vzorčne napake. Če pa zanemarimo še *FPC*, se precenjevanje načeloma še poveča. Res pa je, da so izboljšave v natančnosti zaradi proporcionalne stratifikacije običajno majhne (posebej v družboslovju, kjer ocenjujemo predvsem odstotke), zato lahko v praksi uporabimo mnogo enostavnejši izraz za standardno napako SRS vzorčenja s ponavljanjem. Seveda je treba pred tem natančno preveriti, ali je v resnici mogoče povsem zanemariti pridobitve, ki jih prinese proporcionalna stratifikacija. V vsakem primeru pa je uporaba izrazov za SRS vzorčenje s ponavljanjem pri proporcionalni stratifikaciji konservativna in je torej varna, saj vzorčno varianco kvečjemu precenjuje, s čimer se preprečujejo prenagljeni sklepi.

Položaj pri disproporcionalnem stratificiranem vzorčenju je bolj zapleten. Ker je disproporcionalna stratifikacija ne-EPSEM vzorčni načrt, zahteva vsako ocenjevanje parametrov predhodno uteževanje. Pri uteževanju, ki nevtralizira ne-EPSEM vzorčni načrt, pa moramo elementarno populacijsko varianco oceniti z uteženo oceno:

$$s_w^2 = \frac{n}{n-1} \frac{\sum w_i (y_i - \bar{y}_w)^2}{\sum w_i}$$

iz izraza (27). Zato je ocena za standardno napako vzorčne aritmetične sredine SRS vzorca s ponavljanjem enaka:

$$se(\bar{y}_w) = \frac{s_w}{\sqrt{n}}.$$

Seveda pa je vprašanje, ali jo smemo uporabiti. Učinki disproporcionalne stratifikacije na natančnost cenilk so namreč bistveno manj jasni kot pri proporcionalni stratifikaciji. Disproporcionalno stratificirano vzorčenje lahko zato daje cenilke, ki so bolj ali pa tudi manj natančne kot pri SRS vzorcu s ponavljanjem enake velikosti.

Po eni strani – če so stroški na element enaki za vse stratume – daje optimalna (Neymanova) stratifikacija pri ocenjevanju populacijske aritmetične sredine vsaj tako natančno oceno kot proporcionalna stratifikacija. Večjo natančnost kot proporcionalna stratifikacija pa daje optimalna stratifikacija šele v primerih, ko elementarna varianca izraziteje variira po stratumih. Izraz za standardno napako SRS vzorca s ponavljanjem bo zato v takem primeru precenjeval dejansko vzorčno napako.

Po drugi strani pa lahko disproporcionalna stratifikacija, ki je prilagojena eni spremenljivki, povzroči podcenjevanje standardne napake pri ocenjevanju preostalih spremenljivk.

Disproporcionalno stratifikacijo pogosto uporabimo, kadar potrebujemo ločene ocene za določene domene. Posebej stratumi, ki predstavljajo manjša področja, so pogosto vzorčeni z višjim vzorčnim deležem, s čimer zagotovimo zadostno velikost vzorca za vsako domeno. Uporaba disproporcionalne stratifikacije zato nemalokrat vodi v zmanjšanje natančnosti ocene na celotnem vzorcu v primerjavi s SRS vzorcem enake velikosti. Zmanjšanje je lahko precejšnje, če so nekatera področja (stratumi) vzorčena s posebej visokimi vzorčnimi deleži.

Kot primer vzemimo dva stratuma, vsak od njiju naj predstavlja domeno, za katero je potrebna ločena ocena. Pri tem eden od stratumov obsega 90 %, drugi pa 10 % populacije. Vzemimo, da imata oba stratuma enako aritmetično sredino in varianco. Če iz vsakega stratuma izberemo vzorce enake velikosti, so za oblikovanje celotne ocene potrebne uteži v razmerju 9 : 1. Če zanemarimo učinek *FPC*, bo taka disproporcionalna stratifikacija oziroma ustrezno uteževanje povečalo varianco aritmetične sredine celotnega vzorca približno za faktor 1.64 v primerjavi s SRS vzorcem s ponavljanjem enake velikosti. Če imamo torej velik razpon oziroma variabilnost v utežeh, se natančnost ocen lahko zelo zmanjša. Uporaba izrazov za standardno napako SRS vzorca s ponavljanjem zato v takih primerih resno podcenjuje vzorčne napake.

Vzorčenje v skupinah privede do zmanjšanja natančnosti v primerjavi s SRS vzorcem enaki velikosti, če je koeficient intraklasne korelacije ρ pozitiven, kar pa je skoraj vedno. Zmanjšanje natančnosti je odvisno tako od velikosti ρ kot tudi od povprečne velikosti podvzorcev v skupini, kar smo podrobno obravnavali že v predhodnih poglavjih (14), kjer smo velikost skupin v vzorcu označevali z b. Če je povprečna velikost skupin velika, se lahko močno zmanjša tudi natančnost, kljub relativno nizki vrednosti ρ. Izrazi za standardno napako SRS vzorca brez ponavljanja zato praviloma podcenjujejo – pogosto celo izrazito – vzorčne napake pri večstopenjskem vzorčenju v skupinah.

V praksi se pogosto srečujemo s kompleksnimi vzorčnimi načrti, ki na vsaki stopnji vzorčenja vključujejo večstopenjsko vzorčenje in stratifikacijo. Pri tem posebej pogosto uporabljamo proporcionalno stratifikacijo in EPSEM vzorčne načrte. Splošno pravilo pri izračunu vzorčnih napak v takšnih načrtih je naslednje: zmanjšanje natančnosti zaradi vzorčenja v skupinah običajno presega pridobitve proporcionalne stratifikacije. Kompleksni vzorčni načrt zato v resnici daje precej manj natančne cenilke kot SRS vzorec s ponavljanjem enake velikosti. Z drugimi besedami, vzorčni učinek *Deff* je običajno večji od 1. Velikost vzorčnega učinka je seveda odvisna od množice dejavnikov, med drugim od narave skupin, povprečne velikosti podvzorca v skupinah, narave uporabljene stratifikacije, ciljnih spremenljivk in tipa cenilke. Pri običajnih nacionalnih prostorskih verjetnostnih vzorcih so vzorčni učinki *Deff* za aritmetične sredine in deleže osnovnih demografskih spremenljivk, kot sta starost in spol, povsem blizu 1, kar pomeni, da je v skupinah notranja homogenost pri teh spremenljivkah izredno majhna. Vzorčni učinki za ciljne spremenljivke pa so v splošnem večji od 1. Pri običajnih nacionalnih večstopenjskih vzorcih oseb se vrednosti *Deff* pogosto nahajajo med *Deff* = 1.5 in *Deff* = 2.5. Pri tem so vzorčni učinki za aritmetične sredine ali deleže v podrazredih (npr. ocene za ciljne spremenljivke med mladimi), ki so enakomerno razporejeni v vzorcu *(angl. crossclass)*, praviloma manjši kot na celotnem vzorcu.

Vzorčni učinki za razlike med aritmetičnima sredinama dveh podrazredov so praviloma manjši od vzorčnih učinkov za aritmetične sredine znotraj samih podrazredov. Vzorčni učinki za regresijske koeficiente so pogosto podobni vzorčnim učinkom za razlike med aritmetičnimi sredinami.

Ne glede na to, kakšne so cenilke, pa so vzorčni učinki kompleksnih vzorčnih načrtov skoraj vedno večji od 1, včasih le malo, včasih pa občutno. Uporaba izrazov za standardno napako SRS vzorca s ponavljanjem zato običajno precenjuje natančnost in podcenjuje vzorčno varianco.

V zadnjem času je bilo razvitih več računalniških programov za izračun vzorčnih napak pri cenilkah, ki temeljijo na kompleksnih vzorčnih načrtih. Večina programov predpostavlja, da smo enote prve stopnje (PSU) izbirali s ponavljanjem, čeprav je v praksi izbiranje pogosto izvedeno brez ponavljanja. Tovrstna predpostavka zato povzroča precenjenost variance, toda obseg precenjevanja je majhen, če je vzorčni delež na prvi stopnji majhen. Največji prednosti predpostavke o uporabi vzorčenja s ponavljanjem sta bistveno zmanjšanje računanja in večja splošnost uporabe. Kot smo videli v poglavju o večstopenjskem vzorčenju (15) oziroma (23), lahko standardno napako aritmetične sredine – če je vzorčni delež na prvi stopnji dovolj majhen – ocenimo samo iz variiranja aritmetičnih sredin v vzorčenih enotah prve stopnje (PSU). Ocene variiranja znotraj vzorčnih enot prve stopnje (in tudi znotraj nadaljnjih stopenj) zato sploh niso potrebne, kar prihrani mnogo dodatnega računanja. Še pomembnejša prednost opisanega pristopa je splošnost izračuna pri vzorčenju s ponavljanjem, saj je za določeno cenilko potreben samo en izraz za standardno napako, ne glede na način podvzorčenja znotraj enot prve stopnje PSU. Tako je npr. isti izraz uporaben, če so znotraj izbranih PSU elementi vzorčeni s SRS izbiro, sistematično, stratificirano ali pa z nadaljnjimi stopnjami. Edini pogoj je, da ima vzorec lastnost EPSEM oziroma je bilo pred tem izvedeno ustrezno uteževanje.

Opisana splošnost je privlačna ne le zato, ker en sam program zadošča za izračun standardne napake za katero koli obliko podvzorca, ampak tudi zato, ker uporabniku ni treba navajati podrobnosti vzorčnega načrta za podvzorce. Uporaba programa zato zahteva samo to, da vsak element vsebuje identifikacijo ustrezne enote prve stopnje in stratuma.

Taylorjeva vrsta

Obstaja več načinov ocenjevanja vzorčnih napak za splošne cenilke (linearne in nelinearne) v kompleksnih vzorčnih načrtih. Pri tem s pojmom linearne cenilke razumemo vsako cenilko, ki se izraža kot linearna kombinacija vrednosti spremenljivk v vzorcu, npr. aritmetična sredina. Kot primer nelinearnih cenilk lahko navedemo razmerja, regresijski koeficient, korelacijski koeficient ipd. Eden najpogostejših načinov je razvoj cenilke v Taylorjevo vrsto (*angl. Taylor series expansion*), kar včasih imenujemo tudi linearizacija oziroma metoda delta. Navedeni postopek smo omenili že pri obravnavi razmernostne cenilke (23). Osnovna ideja takega načina je v zapisu cenilke v poenostavljeno linearno obliko, nakar lahko uporabimo postopke za ocenjevanje variance pri (bolj enostavnih) linearnih cenilkah.

Uporaba Taylorjeve vrste je izvedbeno razmeroma enostavna, zato se pogosto uporablja, posebej za enostavnejše cenilke. Uporablja se tudi v številnih programih (npr. SUDAAN), posebej za izračun variance vzorčnih aritmetičnih sredin, deležev, deležev v podrazredih, razlik med aritmetičnimi sredinami ter razlik med deleži v kompleksnih vzorčnih načrtih.

Kot rečeno, uporaba Taylorjeve metode omogoča tudi ocenjevanje variance za cenilko razmerja. Razmerje je namreč eden najpogostejših primerov nelinearnih statistik. Cenilko vzorčne variance za razmerje smo obravnavali v poglavju o sorazmernem vzorčenju (23). Pri tem smo navedli, da mora biti koeficient variacije imenovalca manjši od 0.1 oziroma od 0.2. Večina programov za ocenjevanje variance navaja vrednosti tega koeficienta v svojih izpisih, saj je treba – posebej pri analizi podrazredov – te vrednosti rutinsko preverjati.

Metoda BRR

Alternativni način za ocenjevanje standardne napake je oblikovanje posebne množice vzorcev, ki omogočajo enostaven izračun ocen za standardne napake katerokoli cenilke. Kot smo že omenili (24), lahko to dosežemo z metodami repliciranega vzorčenja, kjer vzorec razbijemo v niz ponovitev neodvisnih podvzorcev oziroma replikacij z enakim vzorčnim načrtom. Variiranje med ocenami posameznih replikacij nato omogoča enostavno oceno standardne napake za cenilko v celotnem vzorcu. Pri tem ni pomembno, kako zapletena (npr. nelinearna) je cenilka ali kako kompleksen je vzorčni načrt.

Pri obravnavi enostavnega repliciranega vzorčenja smo omenili, da je pomembna omejitev nasprotje med potrebo po dovolj velikem številu replikacij, ki omogočajo natančne ocene, ter željo po podrobni uporabi stratifikacije (znotraj vsake replikacije), kar ravno tako povečuje natančnost ocen. Zaradi opisanega nasprotja se enostavno replicirano vzorčenje redko uporablja. Namesto tega uporabljamo posebne tehnike psevdoreplikacij, ki vključujejo vse prednosti enostavnega repliciranega vzorca, hkrati pa zagotavljajo natančne cenilke za standardno napako in preprečujejo omejitve pri stratifikaciji. V nadaljevanju bomo najprej na kratko opisali metodo uravnoteženih ponovljenih replikacij (*angl. Balanced Repeated Replications – BRR*), ki jo imenujemo tudi replikacije polovičnega vzorca oziroma replikacije polvzorca (*angl. half-sample replication*). Podrobnosti o tem najdemo v obsežni literaturi (Kish in Frankel, 1970, 1974; Frankel, 1976; McCarthy, 1966).

Metodo BRR uporabljamo pri vzorčnem načrtu z dvema enotama prve stopnje v vsakem stratumu (*angl. paired selection design)*, kjer v vsakem stratumu izberemo natanko dve PSU. Če so PSU stratificirane tako, da je bila v vsakem stratumu izbrana samo po ena PSU, je za oceno variance treba pred tem uporabiti tehniko združenih stratumov (*angl. collapsed strata*). Združevanje parov stratumov s tem preide v vzorčni načrt z dvema enotama (PSU) v stratumu. Podobno lahko preoblikujemo tudi vzorec, kjer so bile enote prve stopnje sicer izbrane brez stratifikacije, vendar s sistematično izbiro iz predhodno geografsko urejenega seznama (implicitna stratifikacija); v takem primeru v stratume vključujemo pare sosednjih PSU.

Pri metodi BRR obe izbrani PSU iz vsakega stratuma obravnavamo, kot da sta vzorčeni povsem neodvisno. S stališča repliciranega vzorčenja zato tak vzorec obravnavamo v dveh polvzorcih oziroma v dveh replikacijah – prva vsebuje naključno izbrano PSU iz vsakega stratuma, druga pa preostale (komplementarne) PSU. Če je z' vzorčna ocena za parameter

Z (npr. regresijski koeficient), ki temelji na prvi replikacij oziroma polvzorcu, z'' pa cenilka, ki temelji na drugi polovici vzorca, je vzorčna aritmetična sredina:

$$\bar{z} = \frac{z' + z''}{2}.$$

Varianca zgornje cenilke izhaja iz izraza (24), kjer upoštevamo $c = 2$:

$$v(\bar{z}) = \frac{(z' - \bar{z})^2 + (z'' - \bar{z})^2}{2}. \qquad [28]$$

V praksi za oceno populacijskega parametra Z raje uporabimo enostavno oceno aritmetične sredine $\tilde{z}$ na celotnem vzorcu, ki ga sestavljata oba polvzorca. Ker pa sta ocena $\bar{z}$ – to je povprečje aritmetičnih sredin polvzorcev – in ocena $\tilde{z}$ običajno blizu, lahko zaradi poenostavljenega izračunavanja namesto ocene $\bar{z}$ uporabimo kar oceno $\tilde{z}$. Brez večjih posledic bi torej lahko v izrazu (28) namesto $\bar{z}$ uporabili oceno $\tilde{z}$.

Tudi v zgornjem primeru pa ostaja omejitev opisane cenilke v nestabilni oceni, saj temelji na le eni prostostni stopnji in je zato praktično neuporabna. Pri metodi BRR je rešitev tega vprašanja v ponavljajočem generiranju slučajno izbranih polvzorcev iz osnovnega vzorca ter vsakokratnem izračunu ocene. Na tej podlagi lahko generiramo veliko replikacij in s tem tudi ocen, iz katerih nato izračunamo varianco. Če je z'_t vzorčna cenilka za Z, ki temelji na t-tem polvzorcu, z''_t pa vzorčna cenilka, ki temelji na njegovem komplementarnem polvzorcu, je ocena variance za $\tilde{z}$:

$$v(\bar{z}) = \frac{\sum((z'_t - \bar{z})^2 + (z''_t - \bar{z})^2)}{2T},$$

kjer je T skupno število polvzorcev oziroma njihovih komplementov, količino $\bar{z}$ iz izraza (28) pa tokrat nadomestimo z $\tilde{z}$.

Iz zgornjega opisa je razvidna temeljna značilnost ponovljenih replikacij (*angl. repeated replications)*. Lastnost uravnoteženosti (*angl. balanced*) se pri tem nanaša na način, kako so izbrani polvzorci. Skupno število T polvzorcev namreč ni izbrano povsem slučajno in neodvisno, ampak na poseben in uravnotežen način, ki daje najučinkovitejšo cenilko variance. Optimalna razporeditev polvzorcev je razvidna v posebnih tabelah za BRR vzorčenje. Da dobimo čim boljši rezultat, mora biti število izbranih polvzorcev T večje ali enako številu stratumov, poleg tega pa tudi mnogokratnik števila štiri. Tako npr. pri 22 stratumih (to je 44 PSU v vzorčnem načrtu z dvema enotama v stratumu) potrebujemo za uravnoteženo izbiro $T = 24$ polvzorcev, pri 47 stratumih pa $T = 48$ polvzorcev. Če je izračun vsake ocene z_t zahteven in je število stratumov veliko, lahko postane računanje preobsežno. V takem primeru so na voljo dodatne tehnike za dosego delnega ravnotežja z manjšim številom polvzorcev. Ob vse zmogljivejših računalnikih pa postajajo tovrstne računske omejitve vse manj pomembne.

Metoda Jack-knife

Pomembna tehnika za ocenjevanje variance pri kompleksnih vzorčnih načrtih (Frankel, 1971; Kish in Frankel, 1974) je tudi Jack-knife ponovljena replikacija (*angl. JRR – Jack-knife Repeated Replications*). Tako kot pri BRR gre tudi pri JRR za ponavljanje določenih replikacij. Vendar tokrat replikacije generiramo z opustitvijo samo ene PSU, preostale PSU v stratumu pa utežimo, da obdržimo osnovno porazdelitev vzorca po stratumih. Če nimamo stratifikacije, izvajamo replikacije s postopnim izpuščanjem vsake od PSU v vzorcu. Replikacije lahko generiramo tudi tako, da izpustimo po več PSU hkrati. Če imamo npr. 120 PSU, jih torej lahko s sistematičnim izborom razdelimo v 30 replikacij po 4 PSU in v vsaki od 30 replikacij izpustimo 4 PSU. Kadar je število enot PSU majhno (npr. *a*), *PSU* v replikacijah seveda ne bomo združevali, ampak bomo izpuščali le posamezne enote in s tem dobili *a* replikacij. Seveda pa je pri vseh tovrstnih postopkih nujno, da iz vsakega stratuma izpustimo vsaj eno PSU. Če tega ne naredimo, varianca teh stratumov ne bo zastopana v celotni oceni variance.

Naj z_{ht} označuje oceno Z, ki temelji na t-ti replikaciji, oblikovani iz stratuma h. Potem je JRR cenilka za varianco $\tilde{z}$:

$$v(\tilde{z}) = \sum_{h=1}^{H} \sum_{t=1}^{t_h} \frac{(a_h - 1)(z_{ht} - \tilde{z})^2}{t_h},$$

kjer je a_h število vzorčenih PSU v stratumu h, t_h pa število replikacij, pridobljenih z izpustitvijo določenih skupin enot (PSU) iz stratuma h. Če izpustimo vse PSU, je seveda $t_h = a_h$. Kot kaže zgornji izraz, je prednost JRR pred BRR v tem, da brez težav obravnava tudi vzorčne načrte, ki v stratumih niso izbirali po dve PSU, to je $a_h = 2$.

Zgoraj opisane metode ocenjevanja variance so samo približne, vendar simulacije kažejo, da dajejo dovolj dobre rezultate. Kadar se odločamo za eno od opisanih metod, upoštevamo prednosti zmanjšanega obsega računanja, razpoložljivost programov ter praktično primernost za ocenjevanje v konkretnem vzorčnem načrtu. Taylorjevo metodo največkrat uporabljajo pri relativno enostavnih cenilkah, BRR in JRR pa je moč brez težav uporabljati tudi za najbolj zapletene ocene. Metoda BRR je sicer omejena na stratificirani vzorčni načrt z dvema enotama na stratum, kar pa v praksi – z nekoliko prilagoditve – ustreza večini vzorcev. Pri vzorcih, ki niso stratificirani, pa seveda lahko uporabimo JRR.

Pri izračunavanju vzorčne variance se moramo zavedati, da vzorčne raziskave običajno vključujejo več spremenljivk, zato nas v analizah zanimajo številni izvedeni rezultati o spremenljivkah in njihovih razmerjih. Čeprav so na voljo programi za izračunavanje standardne napake, pa je le redko mogoče izračunati standardne napake prav za vse ocene v raziskavi. Tudi če bi vse standardne napake izračunali, bi poročilo postalo preobsežno. Zato običajno izračunamo samo standardne napake najpomembnejših spremenljivk in posebej aktualnih ocen. Na tej podlagi lahko izdelamo posplošene modele, iz katerih izračunamo standardne napake tudi za ostale spremenljivke, ki bi nas zanimale (Kish, 1965: 574–582).

Nekatere podrobnosti o praktičnem ocenjevanju vzorčnih napak in uporabi posplošenih modelov opisuje Kalton (1977). Kish in Frankl (1974) pa govorita o splošnih metodah za ocenjevanje napak pri vzorčenju in predstavljata rezultate simulacij za primerjavo metod Taylorjeve vrste, BRR in JRR.

Uteži in vzorčna varianca

Danes skoraj ni anketne raziskave, ki bi se lahko izognila uteževanju. Za uteževanje namreč obstajajo – kot smo že omenili – mnogi razlogi, številne in kompleksne pa so tudi posledice, ki jih ima uteževanje za statistično analizo. Ena od pomembnih in nadvse neugodnih posledic je tudi povečanje vzorčne variance, zato si bomo v nadaljevanju ogledali enostaven in praktičen način za oceno tega povečanja. Opisani pristop se pogosto uporablja v raziskovalni praksi, razmeroma redko in samo delno pa je predstavljen tudi v literaturi, npr. Kish (1965, poglavje 11.7).

Porast vzorčne variance zaradi uteževanja

Ocena povečanja vzorčne variance, ki nastaja zaradi uporabe uteževanja, izhaja iz elementarnega koeficienta variacije $CV(w)$ za spremenljivko uteži w. Že uvodoma velja poudariti, da se omejujemo na uteževanje na osnovi disproporcionalne stratifikacije, kjer so uteži fiksne in se pri ponavljajočem izbiranju vzorca ne spreminjajo. Fiksna je tudi vsota uteži:

$$w = \sum_{i=1}^{n} w_i .$$

V nadaljevanju torej primerjamo varianco uteženega vzorca na osnovi disproporcionalne stratifikacije s proporcionalnim stratificiranim in neuteženim vzorcem enake velikost, torej z vzorcem, kjer uteži niso bile potrebne. Z drugimi besedami, zanima nas porast vzorčne variance zaradi odstopanja verjetnosti izbora od EPSEM vzorčnega načrta, kjer imajo vsi elementi enake verjetnosti. Obravnavano povečanje v vzorčni varianci pa je enako tudi v primeru, ko elementom v obstoječem SRS vzorcu (torej v EPSEM vzorčnem načrtu) pripišemo določene (arbitrarne) uteži.

V nadaljevanju ocenjujemo uteženo aritmetično sredino, ki jo na osnovi zgornjega izraza za vsoto uteži zapišemo podobno kot v (26a):

$$\bar{y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} = \frac{1}{w}\sum_{i=1}^{n} w_i y_i \ .$$

Zaradi nadaljnjih izpeljav zapišimo varianco zgornje cenilke v stratifikacijski obliki:

$$Var(\bar{y}_w) = Var\left(\frac{1}{w}\sum_{i=1}^{n} w_i y_i\right) = Var\left(\frac{1}{w}\sum_{h=1}^{H}\sum_{j=1}^{n_h}(w_{hj} y_{hj})\right).$$

Pri tem i-ti element v celotnem vzorcu označuje j-ti element, $j = 1...n_h$, v stratumu h, $h = 1...H$, zato imamo identiteto $y_{hj} \equiv y_i$.

Če upoštevamo elementarno zakonitost za izračunavanje variance vsote slučajnih spremenljivk x in y, kjer sta a in b konstanti:

$$Var(ax + by) = a^2 Var(x) + b^2 Var(y) \ , \qquad [29]$$

preide zgornja cenilka v naslednjo obliko:

$$Var(\bar{y}_w) = Var\left(\frac{1}{w}\sum_{h=1}^{H}\sum_{j=1}^{n_h}(w_{hj} y_{hj})\right) = \frac{1}{(n\bar{w})^2} Var\left(\sum_{h=1}^{H}\sum_{j=1}^{n_h}(w_{hj} y_{hj})\right), \qquad [29a]$$

kjer upoštevamo fiksno velikost vzorca, fiksno vsoto uteži in tudi fiksno povprečno utež:

$$\bar{w} = \frac{w}{n} \ .$$

V naslednjem koraku upoštevamo neodvisnost vzorčenja v stratumih:

$$Var(\bar{y}_w) = \frac{1}{(n\bar{w})^2} Var\left(\sum_{h=1}^{H}\sum_{j=1}^{n_h}(w_{hj} y_{hj})\right) = \frac{1}{(n\bar{w})^2}\left(\sum_{h=1}^{H} Var\left(\sum_{j=1}^{n_h}(w_{hj} y_{hj})\right)\right).$$

Nadaljujemo s podobnimi postopkom še za uteži w_{hj}, ki so v stratumih konstantne in s tem neodvisne od spremenljivke y, zato lahko upoštevamo $w_{hj} = w_h$ (čeprav tega nismo eksplicitno zapisali):

$$Var(\bar{y}_w) = \frac{1}{(n\bar{w})^2}\left(\sum_{h=1}^{H} w_{hj}^2 Var\left(\sum_{j=1}^{n_h} y_{hj}\right)\right).$$

Če pri vsakem elementu predpostavimo enako in neodvisno porazdelitev s fiksno elementarno varianco $S^2 = Var(y_i) = Var(y_{hj})$, sledi:

$$Var(\bar{y}_w) = \frac{1}{(n\bar{w})^2}\left(\sum_{h=1}^{H} w_{hj}^2 \sum_{j=1}^{n_h} Var(y_{hj})\right) = \frac{1}{(n\bar{w})^2}\left(\sum_{h=1}^{H} w_{hj}^2 \sum_{j=1}^{n_h} S^2\right) = \frac{1}{(n\bar{w})^2}\left(\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj}^2 S^2\right) =$$

$$= \frac{1}{(n\bar{w})^2}\left(\sum_{i=1}^{n} w_i^2 S^2\right) = \frac{S^2}{n\bar{w}^2}\left(\frac{1}{n}\sum_{i=1}^{n} w_i^2\right). \qquad [29b]$$

Pri tem smo v zadnji vrstici celoten izraz zopet predstavili v obliki števca *i* na celotnem vzorcu in ne več v obliki dveh števcev, med stratumi in znotraj stratumov tako kot v (29a).

Če s pomočjo znanega izraza za elementarno varianco spremenljivke *w* v SRS vzorcu:

$$Var(w) = V(w_o) = \frac{1}{n}\sum_{i=1}^{n}(w_i - \bar{w})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} w_i^2\right) - \bar{w}^2 \qquad [29c]$$

zapišemo vsoto kvadratov uteži $\frac{1}{n}\sum_{i=1}^{n} w_i^2$ v izrazu (29b), dobimo poenostavljen zapis:

$$Var(\bar{y}_w) = \frac{S^2}{n\bar{w}^2}\left(\bar{w}^2 + Var(w)\right) = \frac{S^2}{n}\left(1 + \frac{Var(w)}{\bar{w}^2}\right) = \frac{S^2}{n}\left(1 + CV^2(w)\right). \qquad [30]$$

V zgornji izpeljavi smo izraz (29c) za varianco v SRS vzorcu uporabili zgolj tehnično oziroma algebraično, ne pa tudi v vsebinskem smislu. Stratumske uteži namreč v primeru celovitega SRS vzorca niso več fiksne, tako kot so znotraj posameznih stratumov, ampak v skupnem SRS vzorcu variirajo zaradi razlik v stratumskih utežeh. Na celotnem vzorcu pa so lahko uteži tudi korelirane s spremenljivko *y*, kar znotraj stratumov ni bilo mogoče.

Člen $CV^2(w)$ je kvadrat elementarnega koeficienta variacije uteži in izraža razmerje med elementarno varianco uteži in kvadratom aritmetične sredine, kar včasih označujemo tudi z izrazom relativna varianca (*angl. relvariance*). Pri obravnavi uteževanja ga vključujemo v poseben faktor povečanja variance (*angl. Variance Inflation Factor – VIF*):

$$VIF = 1 + CV^2(w), \qquad [30a]$$

ki izraža povečanje vzorčne variance uteženega vzorca v primerjavi z vzorčno varianco vzorca enake velikost, kjer nismo potrebovali uteži. Tipično se to nanaša na varianco v primeru disproporcialno stratificiranega vzorca, kjer moramo uteževanje uporabiti zaradi neenakih verjetnosti za vključitev elementov v vzorec. V takem primeru *VIF* izraža primerjavo z vzorčno varianco pri predpostavki proporcionalno stratificiranega vzorca enake velikosti, kjer tovrstno uteževanje ni bilo potrebno. Povečanje v vzorčni varianci zaradi uteževanja se zato izraža:

$$Var(\bar{y}_w) = Var(\bar{y}) \times VIF. \qquad [30b]$$

Izraz *VIF* ocenjujemo s cenilko *vif*, ki izhaja iz ocene za elementarni koeficient variacije spremenljivke *w*, kar nadvse enostavno izračunamo iz najosnovnejših opisnih statistik. Če pa je povprečje uteži enako $\bar{w} = 1$ – in tako je v veliki večini običajnih vzorcev – potem ocena za *VIF* enostavno izhaja iz elementarne variance za spremenljivko uteži:

$$vif = 1 + cv(w) = 1 + \frac{sd^2(w)}{\bar{w}^2} = 1 + sd^2(w) .$$

Seveda na tem mestu oznako *w* uporabljamo v smislu spremenljivke in ne kot vsoto uteži. Zato je obravnavani $cv(w)$ elementarni koeficient variacije uteži, kar smo v (26e) izjemoma označevali tudi s $cv_e(w)$.

Povečanje v vzorčni varianci zaradi uteževanja (30) lahko zapišemo v obliki, kjer je eksplicitno razvidna primerjava z varianco SRS vzorca:

$$Var(\bar{y}_w) = \frac{S^2}{n}\left(1 + CV^2(w)\right) = Var_{SRS}(\bar{y}) \times VIF .$$

Izkaže se, da je zgornji izraz robusten, torej neobčutljiv na odstopanja, in ga je mogoče uporabiti tudi v kompleksnih vzorčnih načrtih.

Korelacija med utežjo in spremenljivko

Uporabnost ocene za porast vzorčne variance lahko ogroža korelacija med utežmi in vrednostmi spremenljivke. Oglejmo si to v primeru, ki smo ga predstavili v tabeli 10. Razberemo lahko, da študentje, ki obiskujejo več študijskih smeri – in so naj tej osnovi prejeli tudi manjšo utež – kupujejo večje število knjig. Korelacija je bila visoka, saj jo lahko ocenimo na $r(w,y) = -0.73$. Posredno seveda navedena korelacija vpliva tudi na odgovarjajoče izračune za vzorčno varianco razmernostne cenilke. Pri tem je korelacija med utežjo in spremenljivko *u* (produkt *w* in *y*) enaka $r(w,u) = 0.04$, drugačna pa je tudi korelacija med številom študijskih smeri in številom učbenikov, $r(r,y) = 0.81$. Seveda so zaradi majhnosti vzorca vse navedene ocene za populacijski parameter korelacije *R* nadvse nenatančne. Vsekakor je navedena korelacija vplivala na določeno zmanjšanje vzorčne variance razmernostne cenilke (26a). Določeno podcenjevanje pa je nastalo tudi zaradi neupoštevanja implicitne disproporcionalne stratifikacije v oceni (26a). Na osnovi primerjave z vzorčno varianco SRS vzorca – zaradi česar mora *VIF* sovpadati z *Deff* – smo v izrazu (27a) izračunali:

$$vif^* = deff(\bar{y}_w) = 1.12 .$$

Zgornja ocena je nekoliko drugačna – zato smo jo označili z *vif** – od neposredno ocenjenega faktorja *VIF* na osnovi ocene za elementarni koeficient variacije uteži $cv(w) = 0.48$, kar smo v izrazu (26d) označili tudi kot $cv_e(w)$. Na osnovi ocene za izraz (30a) imamo namreč:

$$vif = 1 + cv^2(w) = 1.23\,.$$

Ker gre za dve različni in – zaradi majhnosti vzorca – precej nenatančni oceni za faktor *VIF*, je očitno nastala določena razlika. K razliki vsekakor prispeva tudi omenjena negativna korelacija med spremenljivko in utežjo, $r(w,y) = -0.73$. Ocena *vif** namreč vključuje vse dejavnike povečanja vzorčne variance zaradi uteževanja, ocena *vif* pa le eno komponento tega povečanja – porast zaradi disproporcionalne alokacije fiksnih uteži.

V praksi sicer prevladuje prepričanje, da je vpliv obravnavane korelacije majhen, saj običajno ni posebnih razlogov za izrazitejšo povezanost med utežmi in vrednostmi spremenljivk, pa tudi vpliv korelacije ni neposreden, ampak je precej kompleksen. Kljub temu je primerno, da se vsakič preveri potencialno nevarnost za obstoj tovrstne korelacije.

Po drugi strani velja opozoriti, da je korelacija med utežmi in spremenljivko – na nivoju celotnega vzorca – predpogoj za nastanek in tudi za zmanjšanje pristranskosti neutežene ocene. Zapišimo razliko med uteženo in neuteženo oceno:

$$bias(\bar{y}_w) = \bar{y}_w - \bar{y} = \frac{1}{\bar{w}}\sum_{i=1}^{n} w_i y_i - \bar{y} = \frac{1}{\bar{w}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} w_i y_i\right) - \bar{w}\times\bar{y}\right).$$

Razberemo lahko, da je izraz na desni ravno enak izrazu za kovarianco, zato zapišemo:

$$bias(\bar{y}_w) = \frac{1}{\bar{w}}cov(w,y) = \frac{1}{\bar{w}}r(w,y)s_y s_w.$$

Če vpeljemo še izraze za koeficient variacije, lahko oceno za relativno pristranskost utežene ocene zapišemo v odvisnosti od ocen za korelacijo in za oba koeficienta variacija:

$$\frac{bias(\bar{y}_w)}{\bar{y}} = r(w,y)cv(y)cv(w)\,.$$

Variabilne uteži

Kot rečeno se *VIF* nanaša na fiksne uteži, ki izhajajo iz neenakih verjetnosti izbora. Samo na tej osnovi namreč lahko prejme določen element enake uteži tudi v ponavljajočih vzorcih. Kot fiksne lahko obravnavamo predvsem uteži, ki nastajajo pri disproporcionalni stratifikaciji in v drugih primerih, ko verjetnosti za izbor v vzorec niso enake (ne-EPSEM). Tipičen primer za to je stratum, katerega elementi so bili v vzorec vključeni s polovico manjšo verjetnostjo kot ostali. Vsi elementi v takem stratumu zato prejmejo utež, ki je sorazmerna faktorju 2, in to celo neodvisno od tega, kateri elementi iz populacije tega stratuma so bili vključeni v konkretni vzorec. Podoben primer, vendar brez eksplicitno razvidne stratifikacije, je tudi npr. gospodinjstvo z dvema telefonskima priključkoma v vzorcih na osnovi telefonskega imenika. Zaradi povečane verjetnosti izbora namreč prejme tak element utež, ki je sorazmerna $^1/_2$, in to vsakič, ko se pojavi v vzorcu. Na tej osnovi ne more tak element prejeti nobene druge uteži; utež je torej fiksna.

Obstajajo pa tudi variabilne uteži, kjer zgoraj opisano ne velja. Take so npr. uteži, ki jih zaradi neodgovorov dodeljujemo respondentom na nivoju skupin v večstopenjskih vzorcih. Tako v primeru, ko v določeni skupini odgovori samo 8 od 10 predvidenih elementov, preostali respondenti prejmejo utež 10/8. Ker pa bi elementi v ponavljajočih vzorcih prejemali vsakič drugačne uteži – odvisno od slučajnosti v variiranju neodgovorov (npr. 10/7 ali 10/9) – veljajo v takih primerih drugačne zakonitosti (Vehovar, 1999), ki se močno razlikujejo od faktorja *VIF*.

Podobne razmere imamo tudi pri populacijskem uteževanju (npr. poststratifikacija, raking). Tako je npr. delež žensk – v primeru enakih in ponavljajočih vzorcev – vsakič nekoliko drugačen in zato so vsakič nekoliko drugačne tudi odgovarjajoče uteži, ki spolno strukturo v vzorcu prilagodijo populacijski. Določen element torej lahko v različnih vzorcih prejme različne uteži, ki so razmeroma neodvisne od njegove siceršnje verjetnosti za izbor v vzorec.

Za razliko od problema korelacije, kjer lahko z dodatnimi napori kljub vsemu izračunamo pravilen obseg povečanja vzorčne variance zaradi uteževanja, pa je to pri variabilnih utežeh bistveno težje. Ne obstaja namreč enostaven način, ki bi zaobjel vpliv variabilnih uteži na povečanje vzorčne variance. Delno lahko vpliv variabilnih uteži – posebej v primeru populacijskih korekcij (npr. raking, poststratifikacija) – zajamemo le v zahtevnejših metodah za izračunavanje variance, ki temeljijo na replikacijah oziroma ponovnem vzorčenju. Seveda pa moramo pred tem v vsaki replikaciji ponoviti celoten postopek uteževanja. Predhodno opisani postopek (30b), ki izhaja iz ocene *VIF* in elementarnega koeficienta variacije uteži, namreč vključuje samo eno komponento povečanja variance, druge pa zaradi predpostavke o fiksnih utežeh v celoti izpušča.

Še težje je pravilno ovrednotiti učinke variabilnih uteži, ki nastajajo zaradi neodgovorov pri vzorčenju v skupinah. Na srečo je v takih primerih porast v vzorčni varianci zaradi uteževanja manjši kot to izhaja iz *VIF*. Ocena osnovi *VIF* je zato konservativna, to je varna (Vehovar, 1995a).

Ocena, ki temelji na *VIF,* torej vključuje le tisti del porasta v vzorčni varianci, ki izhaja iz fiksnih stratumskih uteži (implicitnih ali eksplicitnih). Kadar v konstrukcijo končne uteži vstopajo še druge komponente – zaradi variabilnih uteži ali zaradi korelacije uteži s spremenljivko – je ocena na osnovi *VIF* lahko pomanjkljiva.

Kot rečeno pa v metodološki praksi kljub temu prevladuje (razmeroma nepreverjeno) prepričanje, da *VIF* zajema glavnino povečanja vzorčne variance, ki je nastalo zaradi uteževanja. Ob tem je osnovni problem variabilnih uteži ravno v tem, da – za razliko od problema korelacij – ne obstaja noben enostaven znak, ki bi kazal, da je ocena na osnovi *VIF* napačna, saj pravilne variance zaradi zapletenosti njenega izračunavanja največkrat ne poznamo. Še težavnejše pa je v kompleksnih vzorcih spekuliranje o tem, kakšna bi bila vzorčna varianca, če ne bi potrebovali uteževanja. Za oceno vpliva uteži namreč potrebujemo – samo tako lahko pravilno ocenimo porast vzorčne variance zaradi uteževanja – tudi oceno za vzorčno varianco v primeru, ko uteži ne bi bile potrebne. Ta ocena namreč pri kompleksnih vzorcih v splošnem ni enaka poenostavljeni oceni za vzorčno varianco, ki jo izračunamo kar na surovih in neuteženih podatkih brez upoštevanja verjetnosti za izbor elementov v vzorec.

Srednja kvadratna napaka

Za hip pozabimo na omejitve pri izračunavanju faktorja *VIF* in si oglejmo njegov širši okvir. Povečanje vzorčne variance namreč lahko razumemo kot ceno, ki jo moramo plačati zaradi uporabe uteži. Po drugi strani pa z utežmi seveda zmanjšujemo pristranskost. Koliko je uteževanje v resnici koristno, lahko ocenimo šele z izračunom srednje kvadratne napake (*angl. Mean Squared Error – MSE*), ki je vsota kvadrata pristranskosti (*angl. bias)* in variance:

$$MSE(\bar{y}) = Bias^2(\bar{y}) + Var(\bar{y}). \qquad [31]$$

Pri tem je pristranskost opredeljena kot razlika med pričakovano vrednostjo na osnovi ponavljajočih vzorcev ter pravo populacijsko vrednostjo parametra:

$$Bias(\bar{y}) = E(\bar{y}) - \bar{Y}. \qquad [32]$$

V praksi lahko pristranskost običajno ocenimo le v obsegu, v katerem jo lahko izboljšuje uteževanje. Primerjamo lahko torej le uteženo oceno z oceno brez uteževanja. Prave vrednosti $\bar{Y}$ namreč brez dodatnih zunanjih informacij – razen nekaterih redkih izjem (npr. dejanski volilni rezultat) – ne poznamo, zato se v nadaljevanju omejujemo le na pristranskost, ki nastaja zaradi neuporabe uteži:

$$Bias_w(\bar{y}) = E(\bar{y}) - E(\bar{y}_w).$$

Uteževanje je na osnovi kriterija srednje kvadratne napake upravičeno le v primeru, ko se je srednja kvadratna napaka – zaradi učinkovitega zmanjševanja pristranskosti – z uteževanjem zmanjšala. Navedeno lahko ugotovimo s primerjavo med srednjo kvadratno napako za uteženo cenilko $\bar{y}_w$, kjer se pojavi dodatni porast v vzorčni varianci zaradi uteževanja, ter srednjo kvadratno napako za neuteženo cenilko $\bar{y}$, kjer ostaja celotna pristranskost, ki smo jo z uteževanjem sicer uspeli odstraniti, povsem nespremenjena:

$$MSE(\bar{y}) = Bias_w^2(\bar{y}) + Var(\bar{y}),$$

$$MSE(\bar{y}_w) = Var(\bar{y}_w) = Var(\bar{y}) \times VIF.$$

V praksi se omejujemo na naslednjo oceno za pristranskost:

$$bias_w(\bar{y}) = \bar{y} - \bar{y}_w$$

in jo skupaj z *var* ($\bar{y}$) in *vif* uporabimo za izračun ocene *mse* za srednjo kvadratno napako. S primerjavo srednje kvadratne napake za uteženo in neuteženo cenilko torej lahko presodimo o učinkovitosti opravljenega uteževanja. Možno seveda je, da se – zaradi nenatančnosti ocen oziroma zaradi slučajnih in neugodno izbranih vzorcev – tudi motimo, vendar sodi to v siceršnje tveganje, ki ga je mogoče kvantificirati pri vsaki statistični analizi na osnovi verjetnostnih vzorcev.

Če se je z uteževanjem srednja kvadratna napaka povečala, potem uporaba uteži verjetno ni bila smiselna. Slednje se v praksi pogosto dogaja. Obsežno in kompleksno uteževanje namreč nemalokrat znatno poveča vzorčno varianco, ne daje pa izboljšav v pristranskosti. Običajno pa iz preventivnih razlogov kljub temu vztrajamo pri analizi na uteženih podatkih, saj temelji na pravilnih verjetnostih za vključitev elementov v vzorec.

Zgornji izraz je torej nadvse praktičen pripomoček za ocenjevanje učinkov – in tudi smiselnosti – uteževanja. Pri tem se je seveda treba zavedati, da v zgornjo primerjavo srednjih kvadratnih napak nismo vključili vseh komponent, ki sodijo vanjo. Tako npr. ni bila vključena preostala pristranskost, ki je uteževanje ne odpravlja. Poleg tega pa srednja kvadratna napaka seveda ne obsega vseh napak, ki nastajajo v anketnem raziskovanju.

Ob tej priložnosti velja omeniti, da lastnost cenilke z majhno srednjo kvadratno napako označujemo kot točnost (*angl. accuracy*). Točna cenilka torej poleg komponente natančnosti (*angl. precision*), ki se nanaša na majhnost vzorčne variance, vključuje še komponento majhne pristranskosti. V primeru večje pristranskosti so zato cenilke lahko natančne (npr. zaradi velikega vzorca je vzorčna varianca zelo majhna), niso pa tudi točne, saj je srednja kvadratna napaka zaradi pristranskosti še vedno lahko zelo velika.

Pojmov točnosti in natančnosti ne gre zamenjevati z lastnostmi anketnega merskega instrumenta, kot sta veljavnost (*angl. validity*) in zanesljivost (*angl. reliability*). Veljavnost pomeni, da (z anketnim vprašalnikom) dejansko merimo tisto, kar naj bi merili oziroma smo nameravali meriti. Zanesljivost pa se nanaša na lastnost meritve, da v enakih ponovitvah daje kar najbolj enake rezultate.

11. VELIKOST VZORCA

Eno od prvih vprašanj pri načrtovanju vzorcev se vsekakor nanaša na njegovo velikost. Razpravo o tem vprašanju smo do zdaj puščali ob strani, saj je odvisna od vrste dejavnikov, ki smo jih postopno obravnavali v prejšnjih poglavjih. V nadaljevanju pa bomo problematiko določanja velikosti vzorca predstavili podrobneje. Ogledali si bomo običajni postopek določanja velikosti vzorca na podlagi standardne napake in na podlagi koeficienta variacije, posebej pa bomo obravnavali tudi problematiko majhnih vzorcev.

Velikost vzorca in standardna napaka

Za ponazoritev temeljnih načel pri določanju velikosti vzorca si oglejmo primer raziskave z osebnim anketiranjem, kjer želimo v manjšem mestu oceniti zanimanje za obiskovanje knjižnice med odraslimi prebivalci. Mesto šteje 15,000 odraslih prebivalcev. Da bi določili velikost vzorca, moramo najprej opredeliti, kakšno natančnost potrebujemo. Opredelitev natančnosti ni lahka naloga in hitro se lahko zgodi, da jo precenimo, kar ima lahko neugodne finančne posledice. Vzemimo najprej cenilko, ki bo vključevala populacijsko vrednost v intervalu, širokem v vsako smer po 0.02 oziroma 2 %, kar pomeni dve odstotni točki. Pri tem naj tveganje, da navedeni interval kljub vsemu ne vključuje populacijske vrednosti, znaša 5 %. Z drugimi besedami, 95-odstotni interval zaupanja naj obsega odstopanja od vzorčne ocene v vsako smer za dve odstotni točki. Interval zaupanja ima torej v vsako smer širino:

$$1.96 \times SE(p) = 0.02 = 2\%,$$

zato imamo:

$$P = p \pm 0.02,$$

pri čemer je p cenilka za populacijski delež P. Pri enostavnem slučajnem vzorcu (SRS) – in ob neupoštevanju faktorja *FPC* – izhajamo iz izraza:

$$SE(p) = \sqrt{\frac{PQ}{n}},$$

kjer $Q = (1 - P)$, n' pa začetna ocena potrebne velikosti vzorca. Iz zgornjega izraza sledi:

$$n' = \frac{PQ}{SE(p)^2}, \tag{33}$$

kjer namesto izraza $SE(p)$ v našem primeru lahko uporabimo:

$$SE(p) = \frac{0.02}{1.96} \doteq 0.010204.$$

Da bi določili n', potrebujemo tudi oceno za vrednost P. Točna vrednost za P je seveda neznana, saj je to ravno količina, ki jo ocenjujemo. Kljub temu lahko iz izkušenj, poznavanja problema in rezultatov iz prejšnjih anket v grobem ocenimo dopustne vrednosti za populacijski delež P. Če pa vrednosti P ne moremo oceniti, je treba upoštevati, da je izraz PQ največji ravno pri $P = Q = 50\,\%$. Konservativna, to je varna, je torej odločitev, da P izberemo tako, da bo čim bližje $P = 50\,\%$. V mnogih raziskavah zato opravimo izračun le za vrednost $P = 50\,\%$ in se s tem avtomatično zavarujemo za vse preostale deleže. Če pa razpolagamo z natančnejšo informacijo ali predpostavko o možnih vrednostih za P, jo lahko vključimo v izračune in s tem izboljšamo našo oceno. Denimo, da vnaprej vemo, da leži delež P med 15 % in 35 %. Potem je konservativna odločitev $P = 35\,\%$ in v tem primeru iz zgornjega izraza sledi ocena $n' = 2{,}185$.

Če je začetna velikost vzorca v primerjavi z velikostjo populacije majhna, lahko zanemarimo faktor *FPC* in s tem bi že izračun n' dal ustrezno velikost vzorca. V našem primeru pa tega faktorja ne moremo povsem zanemariti. Popravljena ocena velikosti vzorca, ki zagotavlja popravek za faktor končne populacije (*FPC*), mora upoštevati:

$$n = \frac{Nn'}{N + n'}. \tag{33a}$$

V našem primeru pri $N = 15{,}000$ in $n' = 2{,}185$ dobimo $n = 1{,}907$.

V zgornjih izračunih smo predpostavljali enostavno slučajno vzorčenje (SRS), za druge oblike vzorčnih načrtov pa so ocene bolj zapletene. Če obstaja določen vzorčni učinek *Deff*, se namreč zgornja ocena za potrebno velikost SRS vzorca poveča še za faktor *Deff*, ki izraža razmerje, za katero je kompleksni vzorec manj učinkovit od SRS vzorca. V nekaterih primerih (npr. stratifikacija) pa je lahko *Deff* tudi manjši od 1, kar v primerjavi s SRS vzorcem zmanjšuje zahtevano velikost vzorca.

Če bi imeli v zgornjem primeru na voljo seznam vseh odraslih oseb v mestu, bi lahko uporabili proporcionalno stratificiran vzorec in zadostoval bi lahko že nekoliko manjši vzorec. Vendar, kot smo že omenili, s proporcionalno stratifikacijo pri ocenjevanju deležev običajno ne pridobimo veliko, zato bo zmanjšanje potrebne velikosti vzorca največkrat skromno. Denimo, da smo s proporcionalno stratifikacijo dosegli vzorčni učinek 0.97. Potrebna velikost vzorca s proporcionalno stratifikacijo, ki daje 95-odstotni interval zaupanja širine 2 %, je zato:

$$0.97 \times 1{,}907 = 1{,}850.$$

Če pa v zgornjem primeru nimamo niti seznama odraslih oseb niti seznama stanovanj, je treba izpeljati prostorsko vzorčenje *(angl. area sampling)*, pri katerem najprej opravimo vzorčenje prostorskih enot (npr. ulic) in stanovanjskih objektov, nato izvedemo popis stanovanj, zatem opravimo vzorčenje stanovanj in končno še slučajni izbor ene (ali več) odraslih oseb iz vsakega izbranega stanovanja. V takem primeru bi skoraj zagotovo uporabili tudi stratifikacijo in izbiro PPS. Denimo, da je *Deff* pri takem stratificiranem večstopenjskem načrtu, v katerem je iz vsake PSU izbrano v povprečju po deset odraslih oseb, ocenjen na 1.3. V takem primeru bi imel ustrezen vzorec naslednjo velikost:

$$1.3 \times 1{,}907 = 2{,}479.$$

Pri določanju velikosti vzorca moramo upoštevati še en dejavnik, in sicer stopnjo anketiranja, ki izraža delež respondentov med vsemi elementi, ki jih vključimo v vzorec. Vzemimo, da je predvidena stopnja anketiranja 75 %. Velikost vzorca, ki bo zagotavljala ciljno velikost 2,479 odraslih oseb v zgornjem opisanem primeru, je zato naslednja:

$$\frac{2{,}479}{0.75} = 3{,}305.$$

Z navedenim popravkom sicer dobimo želeno velikost vzorca, zavedati pa se moramo, da s tem nikakor ne moremo rešiti vprašanja pristranskosti ocen zaradi manjkajočih elementov.

Ko pri načrtovanju velikosti vzorca pridemo do končne ocene, nas pogosto zanima, ali bi zahtevano izhodiščno natančnost lahko znižali. Denimo, da bi bil namesto 2 % sprejemljiv že interval zaupanja, ki je v vsako smer širok po 3 %. V tem primeru lahko velikost vzorca občutno zmanjšamo na n = 1,581, saj daje ponovljeni izračun potrebne standardne napake tokrat večjo vrednost ($SE(p)$ = 0.015) kot v primeru zahteve po 2 % širini intervala zaupanja ($SE(p)$ = 0.01).

V praksi je natančnost, ki jo potrebujemo za cenilko, redko določena povsem dokončno. Nemalokrat pa so tovrstni izračuni celo povsem v ozadju in se velikost vzorca določa predvsem na podlagi grobe ocene stroškov in nadvse približnega razmerja z natančnostjo.

Določanje velikosti vzorca na podlagi zgornjih izrazov je bilo v našem primeru odvisno od vrste parametrov: delež populacije, ki bo uporabljal knjižnico, vzorčni učinek in stopnja anketiranja. Napake pri ocenjevanju katere koli od teh vrednosti seveda povzročajo določena razhajanja med načrtovano in dejansko natančnost cenilke. Kljub temu pa napake pri tovrstnih približkih običajno niso usodne za izračun velikosti vzorca in v večini primerov je taka cenilka – če ni drugih napak – v grobem ustrezna.

Potem ko smo določili velikost vzorca, je običajno treba določiti še vzorčni delež (*angl. sampling fraction*) in vzorčni korak oziroma vzorčni interval (*angl. sampling interval*), s katerim bomo s populacijskega seznama izbirali elemente v vzorec.

Če bo vzorec izbran s seznama 15,000 odraslih oseb, je treba upoštevati še možnost neustreznih elementov – to so prazni elementi (umrli, odseljeni) in tuji elementi (začasni prebivalci) – pa tudi posledice morebitnih postopkov povezovanja (*angl. linking)*, ki se pri prostorskih vzorcih uporabljajo za reševanje vprašanja manjkajočih elementov in majhnih prostorskih enot. Če imamo npr. 4 % neustreznih elementov in se izračun zaradi povezovanja elementov dodatno ne zapleta, lahko vzorčni delež določimo z izrazom:

$$\frac{2{,}479}{0.96 \times 15{,}000} = 0.172,$$

kar pomeni, da potrebujemo vzorčni interval 1:5.81 (1/0.172), da bi dobili vzorec velikosti 2,479. V praksi lahko vzorčni interval zaradi praktičnosti zaokrožimo, npr. na 1 : 5.8 ali celo 1 : 6, tako da dobimo vzorec 2,483 ali 2,400 odraslih.

Pri večstopenjskem prostorskem vzorcu seveda najprej potrebujemo vzorec stanovanj in šele nato izbiramo po eno odraslo osebo na stanovanje. Denimo, da je bilo pri zadnjem popisu v mestu skupno 6,500 naseljenih stanovanj. To številko moramo najprej osvežiti s popravki in spremembami, do katerih je prišlo po popisu, ter jo ustrezno prilagoditi, če se npr. opredelitev mestnih meja v popisu in raziskavi razlikuje zaradi novozgrajenih sosesk. Vzemimo, da je po vseh teh prilagoditvah število naseljenih stanovanj v mestu ocenjeno na 6,750. Poleg opisanih popravkov moramo upoštevati tudi dejstvo, da vzorčenje v anketni raziskavi verjetno ne bo doseglo take pokritosti stanovanj kot popis. Stopnja pokritosti (*angl. coverage rate*) pri anketni raziskavi lahko doseže npr. 95 % stopnje pokritosti pri popisu in v tem primeru potrebujemo za izbor stanovanj naslednji vzorčni delež:

$$\frac{2{,}479}{0.95 \times 6{,}750} = 0.3866,$$

kar predstavlja korak 1 : 2.59. Navedeni vzorčni korak oziroma interval nam torej zagotavlja želeni vzorec 2,479 stanovanj oziroma odraslih, če v izbranih stanovanjih vključujemo po eno odraslo osebo. Tudi v tem primeru bi zaradi praktičnosti lahko korak zaokrožili na 1 : 2.6, s čimer pristanemo na nekoliko manjšo pričakovano velikost vzorca (n = 2,466).

Zgornji primer zgolj ponazarja najpogostejša vprašanja pri določanju velikosti vzorca, zato je treba opozoriti na številne poenostavitve. V praksi so namreč raziskave večnamenske (*angl. multipurpose design*), tako da moramo upoštevati različne ciljne spremenljivke in ne samo ene kot v zgornjem primeru. Poleg tega pogosto potrebujemo ocene ne le za celoten vzorec, ampak za celo vrsto podskupin (*angl. subclass*), denimo za posamezne pokrajine, za različne starostne skupine, različne stopnje izobrazbe ipd. Eden od pomembnih dejavnikov pri določanju velikosti vzorcev je tudi zagotovitev ustrezne natančnosti za ocenjevanje razlik med podskupinami. Posebej večji vzorci pa dajejo tudi več možnosti za podrobno analizo manjših podskupin.

Določitev velikosti vzorca je zato v praksi kompleksen proces, kjer izvedemo vrsto kompromisov in približnih rešitev. Seveda pa je velikost vzorca v največji meri odvisna od ocene razmerja med stroški ter prednostmi, ki jih omogoča podrobnejša analiza večjega, pa tudi dražjega vzorca.

Velikost vzorca in koeficient variacije

Zgoraj predstavljen pristop k določanju velikosti vzorca zasledimo v večini statističnih učbenikov. Podobno velja tudi za učbenike s področja marketinškega raziskovanja. Na opisan način se pogosto postopa tudi v praksi, čeprav je velikost vzorca primerneje obravnavati na druge načine, kar si bomo v nadaljevanju podrobneje ogledali.

Elementarni koeficient variacije

Ponovimo najprej opredelitev elementarnega koeficienta variacije (*CV*), ki je relativna mera variabilnosti za izbrano spremenljivko:

$$CV(y) = \frac{\sigma_y}{\bar{Y}}. \qquad [33b]$$

Oznako »elementarni« uporabljamo v podobnem smislu kot pri elementarni varianci (1). V obeh primerih gre za populacijski parameter spremenljivke y, katere porazdelitev opazujemo v populaciji N elementov. Poleg osnovne spremenljivke y namreč obravnavamo tudi druge spremenljivke (cenilke), za katere ravno tako izračunavamo varianco in koeficient variacije. Spomnimo se, da so cenilke pravzaprav slučajne spremenljivke, s katerimi ocenjujemo populacijske parametre, tako npr. $\bar{y}$ ocenjuje $\bar{Y}$. Seveda pa cenilke zavzemajo vrednosti (ocene) v populaciji vseh vzorcev in ne v populaciji elementov tako kot osnovne spremenljivke.

Elementarni koeficient variacije predstavlja koristen nadzor nad relativno razpršenostjo osnovne spremenljivke. V družboslovju je namreč običajno, da imamo pri intervalnih in razmernostnih merskih lestvicah opravka z elementarnim koeficientom variacije med 0.2 in 0.5. Na nekaterih drugih področjih, npr. v biologiji, so koeficienti variacije bistveno manjši. Slednje velja že npr. za višino odraslih moških ali žensk, pa tudi za inteligenčni kvocient.

Poseben primer spremenljivke y je binomska porazdelitev, ki zavzame na vsakem elementu samo dve vrednosti (npr. 0 in 1, DA/NE, moški/ženski ipd). Take spremenljivke imenujemo tudi dihotomne in sodijo med spremenljivke z nominalno mersko lestvico. Ker jih srečujemo izredno pogosto, bomo zanje uvedli posebno oznako y^p . Spremenljivka zavzema na elementih v vzorcu vrednosti $y^p = 0$ in $y^p = 1$, v populaciji pa $Y^p = 0$ in $Y^p = 1$. Binomska porazdelitev je tudi osnova za porazdelitev deležev, saj delež P pomeni delež elementov z opazovano lastnostjo ($Y^p = 1$) v celotni populaciji (N). Zaradi posebne oblike elementarne variance (Cochran 1977, 51) ima elementarni koeficient variacije naslednjo obliko:

$$CV(y^p) = \frac{\sigma_p}{P} = \frac{1}{P}\left(\sqrt{PQ\frac{N}{N-1}}\right) \approx \sqrt{\frac{Q}{P}}, \qquad [33c]$$

kjer je $Q = 1 - P$.

Če so v družboslovju elementarni koeficienti variacije za intervalne spremenljivke običajno pod 0.5, pa to ne velja za binomsko porazdelitev. Na podlagi zgornjega izraza tako razberemo, da imamo celo v najugodnejšem primeru, to je pri $P = 0.5$, vrednost:

$$CV(y^p) \approx \sqrt{\frac{Q}{P}} = 1.$$

Pri manjših deležih je koeficient variacije še bistveno večji. Tako zavzame elementarni koeficient variacije pri deležu $P = 0.10$ vrednost 3, pri deležu $P = 0.01$ pa že 9.95. Posledica zgornje zakonitosti je dejstvo, da potrebujemo za ocenjevanje deležev bistveno večje vzorce kot pri ocenjevanju parametrov intervalnih spremenljivk.

V primeru normalne porazdelitve je interpretacija elementarnega koeficienta variacije posebej enostavna in tudi nadvse uporabna. Če je elementarni koeficient variacije (CV) za normalno porazdeljen pojav npr. $CV(y) = 0.10$, potem se dve tretjini populacije nahajata znotraj intervala, ki ga določata večkratnika aritmetične sredine $0.90 \times \bar{Y}$ in $1.10 \times \bar{Y}$. Približno 95 % elementov v populaciji pa je v takem primeru znotraj 0.80 in 1.20 vrednosti aritmetične sredine. Elementarni koeficient variacije torej omogoča hitro oceno, kje se nahaja glavnina populacije. Po drugi strani pa elementarni koeficient variacije omogoča tudi enostavno primerjavo relativnih razlik, npr. v plačah med državami, v deležu strank med regijami ipd., s čimer dobimo učinkovit pregled nad tem, kje so pojavi bolj oziroma manj razpršeni.

Elementarni koeficient variacije je koristen parameter tudi pri pregledovanju kakovosti podatkov, saj hitro razkrije napako v podatkih ali obstoj izstopajočih elementov (*angl. outlier*). V takih primerih namreč koeficient variacije pogosto zavzame nenavadno visoke vrednosti.

Koeficient variacije za cenilke

Ko proučujemo vzorčno porazdelitev – in ta nas v pričujoči obravnavi pravzaprav zanima – ima koeficient variacije nekoliko drugačno vsebino. V primerjavi z elementarnim koeficientom variacije, ki govori o relativni razpršenosti vrednosti spremenljivke na elementih v osnovni populaciji, Y_i, $i = 1...N$, pri cenilki opazujemo npr. razpršenost aritmetičnih sredin, to je vzorčnih ocen $\bar{y}_k$, $k = 1...K$, v populaciji vseh vzorcev:

$$CV(\bar{y}) = \frac{\sqrt{Var(\bar{y})}}{\bar{Y}} = \frac{\sigma_{\bar{y}}}{\bar{Y}} = \frac{SE(\bar{y})}{\bar{Y}}. \qquad [34]$$

Zveza z elementarnim koeficientom variacije je v SRS vzorcu s ponavljanjem nadvse enostavna:

$$CV(\bar{y}) = \frac{\sqrt{Var(\bar{y})}}{\bar{Y}} = \frac{\sqrt{\sigma_y^2 / n}}{\bar{Y}} = \frac{\sqrt{\sigma_y^2}}{\bar{Y}\sqrt{n}} = \frac{\sigma_y}{\bar{Y}\sqrt{n}} = \frac{SD(y)}{\bar{Y}\sqrt{n}} = \frac{CV(y)}{\sqrt{n}}. \qquad [34a]$$

Pri tem seveda ločujemo standardno napako

$$SE(\bar{y}) = \sigma_{\bar{y}} \doteq \sqrt{Var(\bar{y})}$$

in standardni odklon:

$$SD(y) = \sigma_y = \sqrt{Var(y)}.$$

Oglejmo si primer. Denimo, da imamo v vzorcu $n = 100$ elementov spremenljivko porazdeljeno na intervalni lestvici 1–5, pri čemer znaša aritmetična sredina $\bar{Y} = 3$, elementarni standardni odklon pa $SD(y) = \sigma_y = 1$. Navedeno razmerje med aritmetično sredino in standardnim odklonom je za normalno porazdeljen pojav nadvse pogosto. Elementarni koeficient variacije $CV(y)$ znaša v tem primeru:

$$CV(y) = \frac{\sigma_y}{\bar{Y}} = 0.333,$$

koeficient variacije za cenilko pa je enak:

$$CV(\bar{y}) = \frac{SE(\bar{y})}{\bar{Y}} = \frac{SD(y)/\sqrt{n}}{\bar{Y}} = 0.033.$$

Pri deležih, kjer aritmetična sredina določa tudi elementarno varianco (33c), imamo na osnovi (34a) naslednjo poenostavitev:

$$CV(p) = \frac{CV(y^*)}{\sqrt{n}} = \sqrt{\frac{Q}{Pn}}\,. \qquad [34b]$$

Zakonitost obnašanja *SE(p)* in *CV(p)* v odvisnosti od velikosti vzorca in od deleža *P* si lahko ogledamo v spodnji tabeli. Tako npr. znaša pri vzorcu velikosti n = 500 in pri ocenjevanju deleža P = 10 % standardna napaka $SE(p)$ = 0.013 oziroma 1.3 %, koeficient variacij pa $CV(p)$ = 0.134 oziroma 13.4 %. V zadnji vrstici je za primerjavo naveden še elementarni koeficient variacije, in sicer najprej v običajni obliki, nato pa še v obliki odstotkov, s katerimi v tabeli izražamo koeficient variacije cenilke.

Tabela 11: *SE(p)* in *CV(p)* v odvisnosti od velikosti vzorca *n* in populacijskega deleža *P*

Velikost vzorca *n*	*SE(p)%* *CV(p)%*	Delež P v populaciji					
		1%	2%	5%	10%	20%	50%
10	*SE(p)%*	3.146	4.427	6.892	9.487	12.649	15.811
	CV(p)%	314.600	221.350	137.840	94.870	63.245	31.622
20	*SE(p)%*	2.225	3.131	4.873	6.708	8.944	11.180
	CV(p)%	222.500	156.550	97.460	67.080	44.720	22.360
50	*SE(p)%*	1.407	1.980	3.082	4.243	5.657	7.071
	CV(p)%	140.700	99.000	61.640	42.430	28.285	4.142
100	*SE(p)%*	0.995	1.400	2.179	3.000	4.000	5.000
	CV(p)%	99.500	70.000	43.580	30.000	20.000	10.000
300	*SE(p)%*	0.574	0.808	1.258	1.732	2.309	2.887
	CV(p)%	57.446	40.415	25.166	17.321	11.547	5.774
500	*SE(p)%*	0.445	0.626	0.975	1.342	1.789	2.236
	CV(p)%	44.500	31.300	19.500	13.420	8.945	4.472
1,000	*SE(p)%*	0.315	0.443	0.689	0.949	1.265	1.581
	CV(p)%	31.500	22.150	13.780	9.490	6.325	3.162
10,000	*SE(p)%*	0.099	0.140	0.218	0.300	0.400	0.500
	CV(p)%	9.900	7.000	4.360	3.000	2.000	1.000
100,000	*SE(p)%*	0.031	0.044	0.069	0.095	0.126	0.158
	CV(p) %	3.100	2.200	1.380	0.950	0.630	0.316
Elementarni CV	*CV(y)*	9.9	7	4.9	3	2	1
	CV(y)%	994.987	700.000	489.897	300.000	200.000	100.000

Majhni deleži

Oglejmo si še zanimiv in v praksi nadvse pogost primer. Kadar je P dovolj majhen, tako da je:

$$(1-P) \approx 1\,,$$

(npr. P < 5 %), nastanejo pri izračunu koeficienta variacije cenilke p znatne poenostavitve, ki jih bomo v nadaljevanju podrobno predstavili.

Ker je v primeru analize deležev osnovna spremenljivka y^p porazdeljena binomsko in torej zavzema na elementih vrednosti 0 in 1, lahko zapišemo:

$$\sum_{i=1}^{N} Y_i^p = N_a, \sum_{i=1}^{n} y_i^p = n_a,$$

$$P = \frac{N_a}{N}, p = \frac{n_a}{n}.$$

Količina N_a se torej nanaša na število elementov z opazovano lastnost, ki predstavljajo v populaciji delež P. Produkt $n_a = np$ pa pomeni – podobno tudi produkt nP – velikost podskupine oziroma število elementov v vzorcu z obravnavano lastnostjo. Na tej podlagi preide izraz (34b) v obliko:

$$CV(p) = \sqrt{\frac{Q}{Pn}} = \sqrt{\frac{1-P}{Pn}} = \frac{\sqrt{1-P}}{\sqrt{Pn}} \approx \frac{1}{\sqrt{n_a}}, \qquad [35]$$

kar omogoča pri majhnih deležih nadvse enostavno izračunavanje koeficienta variacije. Če imamo v vzorcu npr. 10 oseb, ki berejo določeno revijo v izbrani regiji, potem cenilka p ocenjuje populacijski delež P oseb, ki pripadajo izbrani regiji in hkrati berejo analizirano revijo. Odgovarjajoči koeficient variacije lahko ocenimo:

$$cv(p) \approx \frac{1}{\sqrt{10}} = 0.33.$$

Koeficient variacije je pri majhnih deležih neodvisen od n, N in tudi od P. Seveda je pri dani velikosti podskupine n_a delež p odvisen od velikosti vzorca, kar pa ne velja za odgovarjajoči koeficient variacije. V vzorcu n = 1,000 zato $n_a = 10$ pomeni delež $p = 1$ %, v vzorcu 5,000 pa pomeni $n_a = 10$ delež $p = 0.2$ %. Kljub temu pa je natančnost ocene izražena s koeficientom variacije v obeh primerih enaka. Pri tem velikost deleža ni pomembna, če je le dovolj majhna.

Zgornji rezultat lahko dobimo tudi drugače. V primeru, ko v vzorcih velikosti n opazujemo elemente z določeno lastnostjo oziroma njihov delež, imamo namreč opraviti z običajno Bernoullijevo porazdelitvijo, ki določa verjetnost, da se v n ponovitvah dogodek z verjetnostjo P zgodi natanko k-krat:

$$B_n(k) = \binom{n}{k} P^k (1-P)^{n-k}. \qquad [35a]$$

Dogodek se seveda nanaša na dejstvo, da ima element določeno lastnost in se v določenem vzorcu zgodi tolikokrat, kolikor je elementov z obravnavano lastnostjo *($k = n_a = np$)*. Pri tem je pričakovana vrednost za n_a enaka nP.

Iz statistične teorije (Jamnik, 1980) pa je znano, da zgornja porazdelitev pri majhnih verjetnostih P in pri velikem številu poskusov (velikost vzorca n) preide v Poissonovo porazdelitev:

$$\frac{\lambda^k e^{-\lambda}}{k!},$$

ki ima en sam parameter λ, saj je elementarna varianca te porazdelitve enaka pričakovani vrednosti:

$$E(n_a) = \lambda = nP = \sigma^2_{Poisson} .$$

Elementarni koeficient variacije, to je razmerje med elementarnim standardnim odklonom in pričakovano vrednostjo, je zato:

$$\frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}},$$

kar se ujema z izrazom (35), če namesto λ vstavimo n_a. Slučajna spremenljivka v Poissonovi porazdelitvi in slučajna spremenljivka p v izrazu (35) se sicer razlikujeta za konstanto n, kar pa na koeficiente variacije ne vpliva.

Poissonova porazdelitev je posebej uporabna v primerih, ko velikosti vzorca ne poznamo. Tako se redki dogodki, kot so npr. prometne nesreče s smrtnim izidom, dogajajo z neznano (vendar izredno majhno) verjetnostjo ob velikem številu poskusov, to je ob npr. skupnem številu voženj v določenem tednu. Oboje je seveda neznano, znano pa je končno število tedenskih nesreč n_a, ki predstavlja oceno za pričakovano vrednost in tudi za elementarno varianco. Poissonova porazdelitev se nasploh pogosto uporablja kot teoretični model porazdelitve pri analizi nominalnih spremenljivk.

Koeficient variacije in interval zaupanja

Podobno kot daje elementarni koeficient variacije informacijo, kje se nahaja glavnina elementov v populaciji, omogoča koeficient variacije za cenilko vpogled, kje je glavnina vzorčnih ocen.

Ker je vzorčna porazdelitev za cenilke pri velikih vzorcih vedno normalna, lahko iz koeficienta variacije podrobno sklepamo o njeni porazdelitvi, razpršenosti in s tem tudi o kakovosti cenilke.

Tako se npr. pri vrednosti $CV(\bar{y}) = 0.10$ nahaja 95 % vseh ocen $\bar{y}$ iz vzorcev velikosti n v verjetnostnem intervalu, ki ga določata večkratnika 0.8 in 1.2 vrednosti populacijske aritmetične sredine. Dve tretjini vseh ocen pa se nahajata v razmiku 0.9 in 1.1 aritmetične sredine. Verjetnostni interval, v katerem se giblje 95 % aritmetičnih sredin vseh vzorcev, namreč lahko na osnovi (4b) zapišemo tudi v naslednji obliki:

$$P\left[\bar{Y} - 1.96 \times SE(\bar{y}) < \bar{y} < \bar{Y} + 1.96 \times SE(\bar{y})\right] = 0.95 .$$

Z uporabo koeficienta variacije pa zapišemo:

$$P\left[\bar{Y}\left(1-1.96\times CV(\bar{y})\right)<\bar{y}<\bar{Y}\left(1+1.96\times CV(\bar{y})\right)\right]=0.95\,.$$

Podobno velja tudi za oblikovanje običajnega intervala zaupanja (4a), ki ga lahko zapišemo v naslednji obliki:

$$\bar{Y}=\bar{y}\pm 1.96\times se(\bar{y})=\bar{y}\left(1\pm 1.96\times\frac{se(\bar{y})}{\bar{y}}\right)\approx\bar{y}\left(1\pm 2\times cv(\bar{y})\right),$$

kjer male črke za *se* ($\bar{y}$) in *cv* ($\bar{y}$) tokrat označujejo vrednosti slučajne spremenljivke iz vzorca in ne fiksne vrednosti iz populacije, kot je bilo to pri verjetnostnem intervalu. Seveda lahko v zgornjem izrazu uporabimo tudi odgovarjajoče populacijske vrednosti, če jih poznamo.

Iz zgornjega izraza izhaja, da lahko npr. pri ocenjevanju deleža $P = 0.1$ in pri koeficientu variacije $CV(p) = 0.2$, pričakujemo intervale zaupanja široke 40 % vrednosti deleža, torej ±0.04 oziroma ±4 %:

$$P=p\left(1\pm 2\times cv(p)\right)=p\left(1\pm 0.4\right)=0.1\pm 0.04.$$

Določanje velikosti vzorca

Oglejmo si torej vlogo koeficienta variacije pri določanju velikosti vzorca. Ko smo govorili o določanju velikosti vzorca s pomočjo standardne napake, je bilo eno ključnih meril tveganje, na katero pristajamo. Pri tem največkrat uporabljamo tveganje $\alpha = 5\,\%$ ($z_{\alpha/2} = 1.96$), ki v raziskovalni praksi močno prevladuje. Le pri manj zahtevnih sklepanjih uporabimo $\alpha = 10\,\%$ ($z_{\alpha/2} = 1.65$), pri posebej občutljivih vsebinah pa $\alpha = 1\,\%$ ($z_{\alpha/2} = 2.56$).

Drugi parameter, ki ga moramo opredeliti pri izbiri vzorca, je želena širina intervala zaupanja oziroma standardna napaka, kar lahko izrazimo s koeficientom variacije, če zapišemo:

$$e=1.96\times SE(\bar{y})=1.96\times CV(\bar{y})\times\bar{Y}\,.$$

Uporaba koeficienta variacije *CV* ima v primerjavi s kriterijem absolutne širine intervala zaupanja *e* pomembno prednost, da *CV* ni odvisen od aritmetične sredine. Zahtevana širina intervala zaupanja v obsegu npr. dveh odstotnih točk ($e = 2\,\%$, $SE(p) = 1\,\%$) ima namreč povsem drugačen pomen pri ocenjevanju deleža $P = 50\,\%$ kot pa pri deležu $P = 2\,\%$. Če se izrazimo v odstotkih, imamo v navedenih primerih naslednja intervala zaupanja:

$$P_1=(50\pm 2)\%\quad\text{oziroma}\quad P_2=(2\pm 2)\%.$$

Čeprav gre v drugem primeru za bistveno slabšo natančnost, to iz širine intervala zaupanja ni razvidno. Po drugi strani pa lahko velike razlike v natančnosti zlahka razberemo iz primerjave koeficientov variacije:

$$CV(p_1) = \frac{1}{50} = 0.02 \quad \text{oziroma} \quad CV(p_2) = \frac{1}{2}$$

V praksi je zato pri načrtovanju velikosti vzorcev ugodno izhajati iz koeficienta variacije in ne iz absolutne širine intervala zaupanja oziroma iz standardne napake za cenilko. V takem primeru lahko ocenimo potrebno velikost vzorca na podlagi (34a) definicije CV:

$$n' = \frac{SD^2(y)}{CV^2(\bar{y})\bar{Y}^2}.$$

Pri manjših populacijah in vzorcih brez ponavljanja pa moramo dodatno upoštevati:

$$n = \frac{Nn'}{(N + n')}.$$

Kadar ocenjujemo majhne deleže, se določanje velikosti vzorca še posebej poenostavi. Izhajamo namreč iz izraza (35) in dobimo:

$$n_a = \frac{1}{CV^2(p)}.$$

Tabela 12: Potrebna velikost vzorca oziroma podskupin glede na želeni koeficient variacije $CV(p)$ pri ocenjevanju majhnih deležev

$CV(p)$	n_a
0.01	10,000
0.02	2500
0.03	1111
0.04	625
0.05	400
0.06	277
0.10	100
0.15	44
0.20	25
0.33	9

Iz zgornje tabele razberemo, da potrebujemo npr. za populacijsko podskupino, ki ima v populaciji majhen delež, okoli 100 elementov v vzorcu, če želimo doseči $CV(p)$ = 0.1. Iz koeficienta variacije lahko torej v takih primerih neposredno določamo potrebno velikost podskupin in tudi velikost celotnega vzorca. Dodati velja, da v primeru majhnih podskupin (npr. $n_a < 30$) ocene iz zgornje tabele niso povsem pravilne, saj vzorčna porazdelitev ni več normalna, kar bomo podrobno obravnavali v naslednjem poglavju.

Kritične vrednosti koeficienta variacije

Koeficient variacije za cenilko torej avtomatično opredeljuje relativno širino intervala zaupanja glede na aritmetično sredino. S tem imamo tudi možnost, da opredelimo standarde za mejne oziroma kritične vrednosti. Tako so npr. merila kakovosti vzorčnih ocen lahko naslednja:

- $CV(\bar{y}) < 0.01$ – ključne spremenljivke v pomembnih anketah,
- $CV(\bar{y}) < 0.05$ – običajno meja sprejemljive relativne natančnosti,
- $CV(\bar{y}) < 0.10$ – dopustna meja za mnenjske spremenljivke,
- $CV(\bar{y}) < 0.20$ – ocena, ki jo še objavimo brez posebnega opozorila,
- $CV(\bar{y}) < 0.33$ – ocena, ki jo objavimo le informativno, z opozorilom,
- $CV(\bar{y}) > 0.33$ – ocena, ki se načeloma NE objavlja.

Za ponazoritev lahko navedemo, da Evropski statistični urad (Eurostat) za ankete o delovni sili (Labour Force Surveys) zahteva pri ocenjevanju stopnje brezposelnosti znotraj evropskih regij natančnost, ki je izražena s koeficientom variacije $CV(p) = 5\,\%$.

Dodati je seveda treba, da so zgoraj postavljene meje nekoliko arbitrarne. Različni statistični uradi in druge organizacije, ki objavljajo informacije na podlagi vzorčnih ocen, imajo namreč posebne strategije pri objavljanju podatkov. Tako npr. kanadski statistični urad ne dovoljuje objave za ocene, kjer koeficient variacije presega $CV(\bar{y}) > 0.20$ (v določenih primerih tudi $CV(\bar{y}) > 0.30$) .

Statistični urad Republike Slovenije je v preteklih letih pogosto postavljal naslednje omejitve pri objavljanju rezultatov:

- ocene $CV(\bar{y}) < 0.1$ se objavljajo brez omejitve,
- ocene $0.1 < CV(\bar{y}) < 0.2$ se navajajo v oklepaju, npr. (36),
- ocene $0.2 < CV(\bar{y}) < 0.33$ se navajajo v dveh oklepajih, npr. ((16)),
- ocene $CV(\bar{y}) > 0.33$ se nadomestijo s piko (torej ».«), kar pomeni, da je vrednost sicer neničelna, vendar preveč nenatančna za objavo; seveda pa je tako oznako potrebno jasno ločevati od ostalih kategorij, npr. »ni pojava«, »manjkajoči podatek« ipd.

Izpuščanje ocene in njeno nadomeščanje s piko je po drugi strani seveda določena izguba zbrane informacije. Kljub temu je zaradi nevarnosti napačne interpretacije – posebej s strani statistično neizobraženih uporabnikov – pogosto primerno, da se tako nenatančne ocene enostavno odstrani. Zavedati se namreč moramo, da iz vrednosti $CV(\bar{y}) = 0.33$ izhaja izredno širok interval zaupanja:

$$\bar{Y} = \bar{y}(1 \pm 1.96 \times CV(\bar{y})) \approx \bar{y}(1 \pm 0.66),$$

kar npr. pri ocenjevanju deleža $P = 50\,\%$ pomeni interval $(50 \pm 33)\,\%$. Po drugi strani pa obstajajo primeri, ko je tudi taka informacija lahko koristna. Tako nam npr. vzorčna ocena za populacijski delež P:

$$P = (5 \pm 3)\%$$

kljub nenatančnosti pove, da populacijski delež nikakor ne dosega 10 %, kar je lahko informacija, ki uporabniku povsem zadošča.

Problem uporabe tovrstnih (nenatančnih) informacij na podlagi majhnih vzorcev bomo podrobneje obravnavali v posebnem poglavju.

Intervalne spremenljivke

V družboslovju se pogosto srečujemo z anketami, kjer je ciljna spremenljivka porazdeljena na ordinalni lestvici (1–5). Pri tem vrednost 1 običajno označuje skrajno nestrinjanje s kako trditvijo, vrednost 5 pa pomeni izrazito strinjanje, npr. »povsem se strinjam«, »povsem zadovoljen« ipd. Tovrstne lestvice v statistični analizi pogosto obravnavamo kot intervalne in normalno porazdeljene spremenljivke. V določenih primerih je to seveda vprašljivo, zato je treba tako predpostavko vsakič preveriti.

Ker so navedene spremenljivke v družboslovnem raziskovanju pogoste, se z njimi srečujemo tudi pri določanju velikosti vzorca. V nadaljevanju bomo obravnavali najenostavnejši in tudi najpogostejši primer, ko zavzame aritmetična sredina obravnavane spremenljivke na lestvici (1–5) vrednost $\bar{Y} = 3$. Ponovimo še enkrat, da gre pri tem seveda za ordinalno lestvico, ki jo za praktične potrebe obravnavamo kot intervalno. Predpostavljamo tudi normalno porazdelitev, kar je dober približek za večino tovrstnih spremenljivke. Standardni odklon naj zavzame tretjino vrednosti aritmetične sredine, kar je v praksi nadvse tipična vrednost. Imamo torej:

$$SD(y) = \sigma_y = 1.$$

Opisano normalno porazdelitev označujemo z N(3,1). Hitro lahko razberemo (34a) – faktor končne populacije pri tem zanemarimo – naslednjo zakonitost:

$$n = \frac{SD^2(y)}{CV^2(\bar{y})\bar{Y}^2}.$$

Pri npr. koeficientu variacije $CV(\bar{y}) = 0.05$ in porazdelitvi N(3,1) zato potrebujemo velikost vzorca $n = 44$. Izračun za nekatere tipične vrednosti CV je v spodnji tabeli 13. Tudi tokrat velja dodati, da v primeru majhnih vzorcev izračuni niso povsem pravilni, saj nastopijo določene posebnosti, kar bomo obravnavali v naslednjem poglavju.

Tabela 13: Potrebna velikost vzorca (zaokrožene vrednosti) za normalno porazdeljeno spremenljivke N(3,1) glede na želeno velikost koeficienta variacije $CV(\bar{y})$

$CV(\bar{y})$	n
0.01	1,111
0.02	277
0.03	123
0.04	69
0.05	44
0.06	30
0.10	11
0.15	4
0.20	2
0.33	1

Razberemo lahko, da so vzorci, ki jih potrebujemo za ocenjevanje aritmetične sredine normalno porazdeljene spremenljivke izredno majhni v primerjavi z ustreznimi vrednostmi pri deležih (tabela 11, tabela 12). Seveda pa v raziskovalni praksi poleg koeficienta variacije pri določanju velikosti vzorca uporabljamo še druga merila, kot je npr. najmanjša velikost analiziranih podskupin v vzorcu. Običajno zahtevamo, da je v celici oziroma v podskupini, za katero izvajamo analize, v vzorcu vsaj 30 elementov.

Vsekakor pa iz zakonitosti v zgornji tabeli razberemo, da že v primeru z n = 30 elementi dobimo sprejemljivo natančnost, saj je takrat $CV(\bar{y})$ = 0.06, interval zaupanja pa je enak:

$$\bar{Y} = \bar{y} \pm 1.96 \times SE(\bar{y}) = 3 \pm 1.96 \times 0.18 = 3 \pm 0.358\,.$$

kjer smo upoštevali:

$$SE(\bar{y}) = \frac{SD(\bar{y})}{\sqrt{n}} = \frac{1}{\sqrt{30}} = 0.18.$$

Dodajmo, da v intervalu zaupanja tokrat izjemoma uporabljamo populacijsko vrednost $SE(\bar{y})$, ki jo po predpostavki poznamo. V splošnem pa na tem mestu seveda nastopa ocena $se(\bar{y})$.

Pogosto nas zanimajo tudi primerjave med skupinami in ne le intervalna ocena določenega parametra. Denimo, da nas zanima opazovanje spremembe – to je razlike v aritmetičnih sredinah – za normalno porazdeljeno spremenljivko. Stopnja tveganja naj bo α = 5 %. Predpostavimo, da želimo z dvema enako velikima in neodvisnima vzorcema, ki ju izvedemo zaporedno v določenem časovnem razmaku, na lestvici 1–5 odkriti povečanje povprečne ocene všečnosti neke TV oddaje za razliko:

$$D = \bar{Y}_2 - \bar{Y}_1 = 0.2.$$

Odkriti torej želimo povečanje npr. od $\bar{Y}_1$ = 2.9 na $\bar{Y}_2$ = 3.1. Navedeno povečanje ni majhno, saj predstavlja znaten del vrednosti same ocene. Relativna sprememba namreč znaša:

$$\frac{D}{\bar{Y}} = \frac{0.2}{3} = 0.07$$

oziroma 7 % aritmetične sredine. Na tem mestu je primerno ponoviti, da v nadaljevanju – če izrecno ne opredelimo drugače – vedno govorimo o enostavnih slučajnih vzorcih, pri čemer zanemarjamo tudi faktor končne populacije.

Spomnimo se najprej splošnega izraza (25) za standardno napako razlike dveh slučajnih spremenljivk:

$$Var(\bar{y}_2 - \bar{y}_1) = Var(\bar{y}_2) + Var(\bar{y}_1) - 2 \times Cov(\bar{y}_2, \bar{y}_1).$$

Utemeljeno lahko predpostavimo, da se elementarna varianca spremenljivke kljub morebitni spremembi v aritmetični sredini, ne spreminja. Kadar imamo dva neodvisna (R = 0) in enako velika vzorca, zato izhajamo iz izraza, ki smo ga obravnavali že pri panelnih vzorcih:

$$SE(\bar{y}_2 - \bar{y}_1) = \sqrt{\frac{2}{n}} \times SD(y) = \sqrt{2} \times SE(\bar{y}) \approx 1.41 \times SE(\bar{y}).$$

Izračun pokaže, da pri dveh enako velikih in neodvisnih vzorcih za proučevanje razlike aritmetičnih sredin zadošča že velikost, ki je za faktor 1.4 večja od velikosti, ki jo potrebujemo pri siceršnjem ocenjevanju aritmetične sredine. Interval zaupanja za razliko aritmetičnih sredin zato v takem primeru zavzema naslednjo obliko:

$$(\bar{Y}_2 - \bar{Y}_1) = (\bar{y}_2 - \bar{y}_1) \pm 1.41 \times 1.96 \times se(\bar{y}) = (\bar{y}_2 - \bar{y}_1) \pm 2.8 \times se(\bar{y}).$$

Vrednost 2.8 × SE ($\bar{y}$) je torej mejna vrednost razlike oziroma spremembe, ki jo s 5-odstotnim tveganjem še lahko prepoznamo kot statistično značilno spremembo. V spodnji tabeli si lahko v predzadnjem stolpcu ogledamo najmanjše razlike, ki jih pri določeni velikosti vzorca še lahko zaznamo ob 5-odstotnem tveganju. V zadnjem stolpcu pa je razlika prikazana kot odgovarjajoča relativna sprememba oziroma odstotek od celotne aritmetične sredine ($\bar{Y}$= 3).

Že z dvema neodvisnima vzorcema velikosti $n_1 = n_2$ = 200 bi torej lahko zaznali 6.333-odstotno relativno spremembo v aritmetični sredini normalno porazdeljene spremenljivke N(3,1). Seveda pa je v primeru normalno porazdeljene spremenljivke pri n = 200 tudi siceršnja natančnost ocene nadvse ugodna, $CV(\bar{y})$ = 2.36 %. Če pa bi želeli zaznati 5-odstotno spremembo, lahko v zadnjem stolpcu ugotovimo, da je vrednost 5 % med peto in šesto vrstico, iz česar sledi, da potrebujemo velikost vzorca $n_1 = n_2 \approx$ 350.

Tabela 14: Statistično značilna razlika (α = 0.05) aritmetičnih sredin pri normalno porazdeljeni spremenljivki N(3,1) in dveh neodvisnih vzorcih $n_1 = n_2 = n$

n	$SE(\bar{y})$	$1.96{\times}SE(\bar{y})$	$CV(\bar{y})\%$	$D{=}2.8{\times}SE(\bar{y})$	$D/\bar{Y}\times 100$
30	0.183	0.359	6.100	0.512	17.173
100	0.100	0.196	3.333	0.280	9.333
200	0.071	0.139	2,367	0.199	6.633
300	0.058	0.114	1.933	0.162	5.400
400	0.050	0.098	1.667	0.140	4.667
500	0.045	0.088	1.500	0.126	4.200
600	0,041	0.080	1.367	0.115	3.833
700	0,038	0.074	1.267	0.106	3.533
800	0,035	0.069	1.167	0.098	3.267
1000	0,032	0.063	1.067	0.090	3.000

Če uporabimo panelno raziskavo s popolnim prekrivanjem – v takem primeru iste anketirance sprašujemo v dveh časovnih točkah (odvisni vzorec) – se natančnost (vzorčna varianca) nadalje izboljša (25a) še za faktor (1-R), interval zaupanja pa se zoži za koren navedenega izraza.

Kakšna je v praksi korelacija R, je v splošnem težko reči. Za primerjavo: medletni status stopnje brezposelnosti (p = 0.08) je koreliran z vrednostjo 0.39, delovna aktivnost (p = 0.6) pa z 0.85 (Vehovar in Zaletel, 1994). Ker v mnenjskih anketah npr. zadovoljni anketiranci običajno ostajajo zadovoljni in obratno, je pogosto povsem realno pričakovati vrednosti $R > 0.5$ in s tem zožitev intervala za najmanj 30 %. Z uporabo panela bi torej dosegli enako natančnost z vzorcem, ki je bistveno manjši.

Rezultati so še ugodnejši, če se spremenljivka porazdeljuje bolj koničasto oziroma bolj ozko. V primeru $CV(\bar{y}) = \sigma = 0.66$, torej N(3,0.66), bi z vzorcem n = 200 lahko namesto D = 0.199 (tabela 14) zaznali že razliko:

$$D = 1.41 \times 1.96 \times SE(\bar{y}) = 0.131,$$

npr. spremembo od 3.00 na 3.13. Pri siceršnjem standardnem odklonu $SD(\bar{y}) = 1$ pa bi za to potrebovali vzorec okoli n = 500.

V primeru, da nas zanima tudi analiza podskupin, potrebna velikost vzorca hitro naraste, saj potrebujemo zgornje velikosti za vsako podskupino, v kateri bi želeli dosegati predpisano natančnost.

Zgornji izračuni seveda predpostavljajo zanesljive in veljavne merske lestvice in tudi nadzor nad problemom neodgovorov, posebej pri majhnih končnih populacijah. V naši obravnavi smo tudi predpostavljali, da se spremenljivke z ordinalno lestvico porazdeljujejo normalno na intervalni lestvici, kar je pogumen privzetek, ki ga je treba vsakič preveriti. Ker pa moramo tovrstne izračune nemalokrat opraviti še preden imamo podatke, lahko uporabimo porazdelitve podobnih spremenljivk v anketah, ki so bile že opravljene. Če gre za spremenljivko, o kateri nimamo predhodnih informacij, je priporočljivo izvesti poskusno (pilotno) anketo, saj najosnovnejšo informacijo o porazdelitvi razberemo že iz vzorca 30 ali celo 10 elementov. Če tudi za poskusno anketo ni časa ali sredstev, preostane le ekspertna ocena.

Končna velikost vzorca je v največji meri odvisna od najpomembnejše ciljne spremenljivke. Če je takih spremenljivk več, si izberemo bodisi spremenljivko, ki zahteva največji vzorec, bodisi vrednost mediane ciljnih spremenljivk (Blejec, 1970).

Nominalne spremenljivke

Vsi izračuni koeficienta variacije za aritmetično sredino veljajo tudi za deleže, ki so le poseben primer aritmetičnih sredin. Ker pa v družboslovju srečujemo deleže posebej pogosto, je primerno navesti poenostavitve, ki nastanejo v primeru nominalnih spremenljivk.

V spodnji tabeli so povzeti in prirejeni rezultati iz tabele 11, kar daje enostaven pregled nad velikostmi vzorcev pri ocenjevanju deležev. Tako npr. potrebujemo pri koeficientu variacije $CV(p)$ = 0.05 za ocenjevanje deleža P = 0.10 velikost vzorca n = 3,600. V primeru deleža P = 0.5, za katerega so potrebni najmanjši vzorci, pa potrebujemo n = 400.

Tabela 15: Velikost vzorca glede na koeficient variacije in velikost ciljnega deleža P

$CV(p)$	P = 1%	P = 5%	P = 10%	P = 50%
0.01	990,000	190,000	90,000	10,000
0.02	247,500	47,500	22,500	2,500
0.03	110,000	21,111	10,000	1,111
0.04	61,875	11,875	5,625	625
0.05	39,600	7,600	3,600	400
0.06	27,500	5,277	2,500	278
0.10	9,900	1,900	900	100
0.15	4,400	844	400	44
0.20	2,475	475	225	25
0.33	909	174	83	9

Rezultate lahko primerjamo s tabelo 13 za intervalno spremenljivko. Še enkrat lahko potrdimo, da potrebujemo pri ocenjevanju deležev bistveno večje vzorce kot pri normalno porazdeljenih spremenljivkah, saj v primeru ocenjevanja aritmetične sredina normalne porazdelitve N(3,1) pri $CV(p)$ = 0.05 zadošča že vzorec n = 44.

Podobno velja tudi za ocenjevanje razlike, kar je za delež P = 0.5 prikazano v tabeli 16. Tako npr. lahko pri 5-odstotnem tveganju zaznamo v dveh neodvisnih vzorcih spremembo za okoli D = 0.10 (npr. od 50 % na 60 % ali od 45 % na 55 %) šele pri velikosti dveh neodvisnih vzorcev n = 200. Dodajmo, da predstavlja navedena sprememba okoli 20 % vrednosti deleža P. Da bi zaznali spremembo D = 0.05 (npr. od 47 % na 52 %), ki predstavlja okoli 10 % deleža, pa bi potrebovali dva neodvisna vzorca velikosti n = 800.

Pri ocenjevanju razlike aritmetične sredine normalno porazdeljene spremenljivke (tabela 14) potrebujemo bistveno manjše vzorce. Tako npr. so za zaznavanje relativne spremembe v obsegu petine vrednosti (natančneje 17.17 %) potrebni vzorci okoli n = 30, za relativne spremembe v obsegu 10 % vrednosti aritmetične sredina pa vzorci okoli n = 100, kar je osemkrat manj kot pri deležu P = 0.5 v tabeli 16.

Tabela 16: Statistično značilna (α = 0.05) razlika deležev pri dveh neodvisnih vzorcih, P = 0.5

n	$SE(p)$	$1.96 \times SE(p)$	$CV(p)=SE(p)/P$	$D=2.8 \times SE(p)$	$D/P \times 100$
30	0.091	0.178	0.182	0.255	51.000
100	0.050	0.098	0.100	0.140	28.000
200	0.035	0.069	0.070	0.098	19.600
300	0.029	0.057	0.058	0.081	16.200
400	0.025	0.049	0.050	0.070	14.000
500	0.022	0.043	0.044	0.062	12.400
600	0.020	0.039	0.040	0.056	11.200
700	0.019	0.037	0.038	0.053	10.600
800	0.018	0.035	0.036	0.050	10.000
1,000	0.016	0.031	0.032	0.045	9.000

Pri manjših deležih se zahteve nadalje poslabšajo. Kot vidimo v četrtem stolpcu tabele 17, bi že za samo ocenjevanje deleža P = 0.1 potrebovali vzorce okoli n = 1,000, saj koeficient variacije CV šele takrat zdrsne pod 0.1. Glede odkrivanja statistično značilne spremembe pa bi lahko, kot kaže predzadnji stolpec tabele 17, zaznali razliko D = 0.05 pri n = 300. Seveda pa tokrat D = 0.05 pomeni relativno spremembo za 50 %, to je že polovica vrednosti samega deleža. Za zaznavanje relativne spremembo v obsegu 20 % bi potrebovali vzorec okoli n = 2,000, za odkrivanje relativne spremembe za 10 % pa več kot n = 4,000.

Tabela 17: Statistično značilna razlika (α = 0.05) deležev pri dveh neodvisnih vzorcih, P = 0.1

n	$SE(p)$	$1.96 \times SE(p)$	$CV(p)=SE(p)/P$	$D=2.8 \times SE(p)$	$D/P \times 100$
30	0.055	0.107	0.548	0.152	151.821
100	0.030	0.059	0.300	0.083	83.156
200	0.021	0.042	0.212	0.059	58.800
300	0.017	0.034	0.173	0.048	48.010
400	0.015	0.029	0.150	0.042	41.578
500	0.013	0.026	0.134	0.037	37.188
600	0.012	0.024	0.122	0.034	33.948
700	0.011	0.022	0.113	0.031	31.430
800	0.011	0.021	0.106	0.029	29.400
1,000	0.009	0.019	0.095	0.026	26.296

Če torej analiziramo npr. delež določene politične stranke P = 0.1, potem pri dveh neodvisnih vzorcih potrebujemo velikost n = 1,000, da bi ob običajnem tveganju zaznali spremembo – porast ali upad podpore – za tretjino (natančneje za 26 %) obstoječe vrednosti (npr. povečanje od 9 % na 12 % ali zmanjšanje od 11 % na 8 %). Za odkrivanje manjših sprememb bi potrebovali še večje vzorce. Seveda je povsem mogoče, da uporabnikom zadošča že manj natančno odkrivanje razlik, kar pa vključuje večje tveganje, da je tako sklepanje napačno.

Iz zgornjih primerjav lahko povzamemo, da je v anketah bistveno bolje oblikovati vprašanja, ki temeljijo na ordinalni oziroma intervalni lestvici kot pa na nominalni oziroma dihotomni. Če je le mogoče, torej vprašanje oblikujemo v smislu:

»Kako se na lestvici od 1 do 5 strinjate z naslednjo trditvijo?«

in ne:

»Ali se strinjate z naslednjo trditvijo ali ne?«

Tudi pri uporabi intervalne lestvice pa velja podobno zakonitost kot pri deležih, da lahko večje vrednosti (npr. $\bar{Y}$ = 3) natančneje ocenimo kot majhne (npr. $\bar{Y}$ = 1). Poleg tega je ugodno zagotoviti normalno obliko osnovne porazdelitve. Zato je pri lestvicah 1–5 priporočljivo postaviti vprašanje tako, da se bo srednja vrednost gibala okoli $\bar{Y}$ = 3 in se s tem tudi najbolj približala normalni porazdelitvi. Če bi se torej anketiranci na lestvici 1–5 strinjali z neko trditvijo le v povprečni vrednosti $\bar{Y}$ = 2, je tako trditev priporočljivo preoblikovati v blažjo obliko, kar bo povečalo stopnjo strinjanja in s tem aritmetično sredino približalo vrednosti $\bar{Y}$ = 3.

12. KRITIČNA VELIKOST PODSKUPIN

Ko smo obravnavali vrednosti koeficientov variacije, smo se seznanili tudi z osnovnimi omejitvami pri objavljanju vzorčnih ocen. Tako so npr. ocene deležev, ki se nanašajo na podskupine z manj kot $n_a = 10$ elementi, povsem neprimerne za javno objavo. Pri tem je nepomembno, kako velika je populacija in kako velik je vzorec, če je le delež relativno majhen. Zahteva po najmanj $n_a = 10$ elementih v analizirani podskupini pa je le skrajni minimum. Večina raziskovalnih organizacij in statističnih uradov namreč uporablja strožje pogoje. Že iz tabele 12 npr. razberemo, da lahko brez večjih zadržkov, torej pri pogoju $CV(p) = 0.2$, objavimo ocene šele pri $n_a = 25$. Natančnejše ocene majhnih deležev, npr. $CV(p) < 0.1$, pa so mogoče šele s podskupinami, ki imajo v vzorcu vsaj $n_a = 100$ elementov.

Pri ocenjevanju aritmetičnih sredin so razmere nekoliko ugodnejše. Kljub temu pri natančnosti $CV(\bar{y}) = 0.1$ v primeru predpostavke o normalni vzorčni porazdelitvi potrebujemo najmanj $n_a = 11$ elementov (tabela 13). Seveda pa moramo v tem primeru namesto standardizirane normalne z-porazdelitve uporabiti t–porazdelitev, ki daje pri $n_a = 11$ elementih za 15 % širše intervale.

Pri majhnih vzorcih in pri t-porazdelitvi namreč koeficient variacije izgubi svojo siceršnjo vlogo za opredeljevanje natančnosti cenilk in tudi za določanje intervalov zaupanja, saj vzorčna porazdelitev ni več normalna. Tako bi pri t-porazdelitvi za cenilko aritmetične sredine normalno porazdeljene spremenljivko N(3,1) širina običajnega intervala zaupanja dosegla 20 % vrednosti aritmetične sredine že pri velikosti $n_a = 13$ in ne šele pri $n_a = 11$ kot pri z–porazdelitvi. Relativna širina intervala zaupanja (20 % vrednosti ocene) se namreč ujema z mejno vrednostjo koeficienta $CV(\bar{y}) = 0.10$ samo, kadar je vzorčna porazdelitev normalna.

Opredelitev kritične velikosti

Kadar so vzorci premajhni, lahko statistična analiza vodi k zavajanju naročnikov oziroma uporabnikov. Da bi to preprečili, je statistična stroka sprejela ustrezne etične standarde. Skoraj vsi kodeksi statističnih združenj se zato jasno opredeljujejo do kakovosti informacij, ki jih statistiki posredujejo naročnikom oziroma javnosti. Oglejmo si, kaj o tem pravi kodeks (http://www.sigov.si/zrs/sds/etika.htm) Slovenskega statističnega društva:

»Da bi ohranili zaupanje javnosti, morajo statistiki pri svojem delu upoštevati stroga poklicna merila (vključno z znanstveno in poklicno etiko) glede metod in postopkov za zbiranje, obdelavo, shranjevanje in objavljanje statističnih podatkov. Statistiki ne smejo pretiravati glede točnosti in analitične vrednosti podatkov oziroma informacij.«

»Statistiki ne smejo pristati (javno ali na skrivaj) na izbiro takšnih metod, ki bi dajale zavajajoče rezultate, in tudi ne na nepravilno interpretiranje statističnih rezultatov.«

»Statistiki morajo zagotoviti naročniku celovito oceno o prednostih in pomanjkljivostih posameznih pristopov. Pri prevzemanju nalog morajo ravnati strokovno, odgovorno in kompetentno.«

Poleg splošnih etičnih načel so nam v pomoč tudi konkretna merila, ki jih določene organizacije uporabljajo pri majhnih vzorcih oziroma podskupinah. Merila Statističnega urada Republike Slovenije smo že navedli, v nadaljevanju pa si oglejmo, kako vprašanje kritičnih podskupin rešujejo nekatere druge organizacije:

- Pogosto se poleg minimalnega koeficienta variacije kot kriterij kakovosti za vzorčno oceno vključuje še zahteva, da mora ocenjevana podskupina v vzorcu dosegati predpisano število elementov. Tako npr. kanadski statistični urad v nekaterih raziskavah zahteva kot merilo za objavo – poleg koeficienta variacije $CV(\bar{y}) < 0.33$ – še dodatni pogoj, da se ocena nanaša na vsaj $n_a = 30$ elementov. Vzorčne ocene so tako razdeljene v tri kategorije:
  - *sprejemljiva natančnost*: koeficient variacije je pod $CV(\bar{y}) = 0.165$, velikost podskupine v vzorcu pa je nad $n_a = 30$ elementov – ocene se v tem primeru lahko objavljajo brez opozorila;
  - *mejna natančnost*: koeficient variacije je nad $CV(\bar{y}) = 0.165$, vendar pod $CV(\bar{y}) = 0.33$, hkrati pa je velikost podskupine nad $n_a = 30$ elementov – takim ocenam je treba dodati opozorilni komentar;
  - *nesprejemljiva natančnost*: bodisi je koeficient variacije večji od $CV(\bar{y}) = 0.33$ bodisi je število elementov manjše od $n_a = 30$ – ocene v takem primeru niso sprejemljive za objavo. Če pa se takšne ocene kljub temu navajajo, mora biti dodana opomba, da so nenatančne in ne ustrezajo merilom kanadskega statističnega urada za javno objavo.
- Meja $n_a = 30$ elementov – včasih se navaja tudi $n_a = 20$ elementov – je pogosto navedena kot najmanjša velikost podskupine, ki sploh lahko nastopa kot samostojna enota v statističnih analizah (Little in Rubin, 1987). Pri tem ne gre le za analizirane podskupine v vzorcu, ampak tudi za celice, v katerih opravimo uteževanje ali druge korekcijske metode.
- Podobna merila imajo tudi večje marketinške organizacije, npr. Media Matrix (http://www.mediametrix.com/landing.jsp) ki pri ocenjevanju obiskanosti spletnih predstavitev objavlja rezultate le, če so zbrali vsaj $n_a = 40$ elementov (oseb, ki so obiskale določeno spletno predstavitev).

- Nekatere marketinške organizacije včasih uberejo enostavnejšo pot in enostavno objavijo vse rezultate v vseh celicah, hkrati pa navedejo splošno opozorilo, da so vse ocene, ki temeljijo na podskupinah z manj kot n_a = 120 elementi, lahko nenatančne (http://www.nielsen-netratings.com).

Problematika majhnih vzorcev

Želja po analizi majhnih podskupin je posebej kritična v majhnih državah. Ker sta vzorčna porazdelitev in standardna napaka neodvisni od velikosti populacije, nacionalne anketne raziskave (npr. telemetrija, anketa o delovni sili, volilne ankete itd) zahtevajo velikosti vzorcev, ki so načeloma neodvisne od velikosti države. Seveda pa je v manjši državi bistveno težje zagotoviti sredstva za ustrezno velikost nacionalnih vzorcev, saj je – med drugim – manjše tudi potencialno število uporabnikov takih raziskav. Kadar pa velikost vzorca zaostaja za predvidenim obsegom, se pri statistični analizi majhnih populacijskih skupin soočamo s perečim problemom, ko imamo v odgovarjajočih podskupinah v vzorcu premajhno število elementov.

Majhni vzorci predstavljajo posebno poglavje tudi v statistični teoriji. Tako so v statističnih učbenikih poglavja o majhnih vzorcih običajno ločena od statističnega sklepanja pri velikih vzorcih (npr. Wonnacott in Wonnacott, 1990). Pri velikih vzorcih namreč velja centralni limitni izrek, zato je vzorčna porazdelitev normalna in neodvisna od porazdelitve obravnavane spremenljivke. Normalnost vzorčne porazdelitve omogoča tudi enostaven in standardiziran način izračunavanja intervalov zaupanja ter poenostavljene postopke preverjanja domnev. Po drugi strani pa pri majhnih vzorcih centralni limitni izrek ne velja in vzorčne porazdelitve zato večinoma niso normalne, ampak so vsakič drugačne, odvisne od konkretne porazdelitve spremenljivke v osnovni populaciji in od velikosti obravnavanega vzorca. Statistično sklepanje na osnovi majhnih vzorcev je zato bistveno bolj zapleteno.

Poudariti velja, da ostra in enotna meja, ki bi ločevala velike in majhne vzorce, ne obstaja. Pri ocenjevanju aritmetične sredine oziroma deležev smo sicer večkrat omenjali vrednost n = 30, kar pa ne velja za ocenjevanje drugih parametrov. Posebej pri nelinearnih parametrih (npr. elementarna varianca) so za prehod vzorčne porazdelitve v normalno obliko potrebni nekoliko večji vzorci. Podobno velja tudi za nekatere nenavadne porazdelitve spremenljivk (npr. izrazita asimetrija, izstopajoči elementi ipd).

Običajno je, da za ocenjevanje aritmetične sredine normalno porazdeljene spremenljivke pri manj kot 30 elementih namesto z-porazdelitve uporabimo t-porazdelitev, čeprav so pri 10–30 elementih razlike med porazdelitvama še vedno razmeroma majhne. Pri 5-odstotnem tveganju in velikosti vzorca $n = 30$ so intervali zaupanja na osnovi t-porazdelitve (v primerjavi s standardizirano normalno z-porazdelitvijo) širši za 4 %, pri velikosti $n = 20$ za 7 %, pri 10-tih elementih za 14 %, pri 5-tih elementih za 31 % in pri treh elementih za 62 %.

Podobna meja za majhne vzorce velja tudi pri ocenjevanju deležev. Za praktične potrebe se pogosto privzema normalnost vzorčne porazdelitve že pri manjšem številu elementov; včasih se omenja celo število 15. Opozoriti velja, da v primeru skrajnih deležev (npr. 0.001 ali 0.999) to ne velja, saj se pojavi večja asimetrija vzorčne porazdelitve (Cochran, 1977: 54).

Na tem mestu – razen ocenjevanja deležev in aritmetične sredine za normalno porazdeljeno spremenljivko – ne obravnavamo drugih porazdelitev oziroma parametrov. Poudariti velja, da je ocenjevanje parametrov pri majhnih vzorcih v splošnem precej bolj težavno, kot je bilo to v obravnavanih primerih deležev in aritmetičnih sredin normalno porazdeljenih spremenljivk. Že zgornje ilustracije pa jasno kažejo, da se v primeru majhnih vzorcev oziroma podskupin srečujemo z znatnimi strokovnimi in tudi z etičnimi dilemami.

Bayesov pristop k analizi majhnih vzorcev

Osnovni problem majhnih vzorcev izhaja iz dejstva, da ne obstaja – tako kot pri velikih vzorcih – vnaprej znana in enaka vzorčna porazdelitev, ampak imamo pri vsakem ocenjevanem parametru in pri vsaki porazdelitvi spremenljivke drugačno oziroma specifično vzorčno porazdelitev. Ker klasično statistično sklepanje izhaja iz vnaprejšnjega poznavanja vzorčne porazdelitve, je sklepanje pri majhnih vzorcih razmeroma težavno. K statističnemu sklepanju pa je mogoče pristopiti tudi drugače. V nadaljevanju si bomo ogledali alternativen pristop k analizi majhnih vzorcev, kjer statistično sklepanje ne izhaja iz vzorčne porazdelitve.

Začnimo s primerom. Denimo, da v večjem vzorcu analiziramo spolno strukturo za podskupino bralcev določene revije v izbrani regiji. Pri tem sta 2 od 8 elementov (bralcev revije v tej regiji) moška. Ob tem vsekakor obstaja intuitivni občutek, da bi to lahko posplošili tudi na populacijo, v smislu, da je večina bralcev v izbrani regiji ženskega spola. Po drugi strani se pri taki trditvi seveda soočamo s tveganjem, da je zaradi majhnosti vzorca taka ocena napačna, s čimer bi uporabnike (npr. medijske načrtovalce) lahko zavedli v zgrešeno alokacijo oglaševalskih sredstev. Ker je $n_a < 10$, je taka ocena neprimerna tudi z vidika vseh priporočenih omejitev za objavljanje vzorčnih ocen. Vendar – ali smo zato, ker nimamo niti 10 elementov, zavezani k strokovni molčečnosti? Ali pa podatki kljub temu skrivajo določeno informacijo, ki bi uporabniku lahko koristila?

Če bi problem poenostavili in kljub majhnosti podskupine predpostavili normalno vzorčno porazdelitev – kar je seveda napačno – je interval zaupanja za delež elementov, ki imajo opazovano lastnost, naslednji:

$$p \pm 1.96 \times \sqrt{\frac{pq}{n}} = 0.25 \pm 1.96 \times \sqrt{\frac{0.25 \times 0.75}{8}} = 0.25 \pm 0.30 .$$

Če pa bi – kar le nekoliko manj napačno – pri tem upoštevali *t*-porazdelitev, je interval zaupanja širši:

$$p \pm t(8-1) \times \sqrt{\frac{pq}{n}} = 0.25 \pm 2.365 \times \sqrt{\frac{0.25 \times 0.75}{8}} = 0.25 \pm 0.36 .$$

Ob navedenih poenostavitvah dobimo vtis, da večina elementov – kljub precejšnji nenatančnosti – verjetno nima obravnavane lastnosti, kar je lahko koristna informacija, ki morda v celoti zadostuje potrebam naročnika. Seveda pa je bila zgornja predpostavka o normalni vzorčni porazdelitvi zaradi premajhnega vzorca napačna, saj centralni limitni izrek ne velja. Vzorčni deleži se pri velikosti n = 8 namreč ne porazdeljujejo normalno niti se ne porazdeljujejo s *t*-porazdelitvijo. Delno se to razkriva tudi v nerealni negativni spodnji meji obeh intervalov zaupanja.

Primer si bomo ogledali s pomočjo Bayesove teorije statističnega sklepanja, ki – za razliko od klasičnega sklepanja – ne izhaja iz posploševanja na osnovi vzorčne porazdelitve, ampak upošteva dejansko porazdelitev spremenljivke. Na tej osnovi je mogoče razmeroma enostavno posploševanje pri poljubnem parametru, poljubni porazdelitvi in tudi pri poljubni velikosti vzorca.

V pričujoči obravnavi se podrobneje v načela Bayesovega statističnega sklepanja seveda ne spuščamo, omenimo le, da gre za pristop, ki ga je prvi omenil angleški statistik Thomas Bayes (Essay towards solving a problem in the doctrine of chances, 1764), ponovno pa so ga oživili in radikalizirali številni statistiki v drugi polovici 20. stoletja. Bayesov pristop je osnova za konceptualno povsem drugačen statističen pristop, zato se statistiki pogosto delijo na klasične statistike in »Bayesijance«.

Osnovna ideja Bayesovega sklepanja je v obravnavi vzorčne ocene kot fiksne vrednosti. Ocenjevani populacijski parameter je pri tem neznana variabilna količina (slučajna spremenljivka), ki se porazdeljuje okoli fiksne vrednosti iz dobljenega vzorca. To je seveda ravno obratno kot pri klasičnem statističnem sklepanju, kjer predpostavljamo, da je parameter fiksna (čeprav neznana) vrednost, vzorčne ocene pa spremenljiva količina, ki v ponavljajočih vzorcih oblikuje vzorčno porazdelitev. V primeru velikih vzorcev dajeta oba pristopa skoraj povsem enake rezultate, obstajajo pa številne okoliščine, kjer je Bayesov pristop primernejši. Pri Bayesovem pristopu se namreč porazdelitev ocenjevanega parametra računa kot produkt predhodne porazdelitve – ki izhaja iz morebitnega vnaprejšnjega poznavanja pojava – ter iz porazdelitve na osnovi dobljenega vzorca. S tem je tovrstno sklepanje bolj prilagodljivo glede vključevanja informacij o vnaprejšnjem poznavanju obravnavanega pojava. Posebej primerno je za majhne vzorce in tudi za potrebe statističnega modeliranja.

Predpostavimo, da o deležu, ki ga ocenjujemo, vnaprej ne vemo ničesar. Imamo torej enakomerno (*angl. flat*) predhodno porazdelitev (*angl. prior distribution*) populacijskega deleža. Včasih jo imenujemo tudi neinformativna (*angl. non-informative*) predhodna porazdelitev. To pomeni, da pred izvedbo anketne raziskave z enako verjetnostjo dopuščamo prav vse populacijske deleže: $P = 0.0$, $P = 0.1$, $P = 0.2, \ldots$, $P = 0.9$, $P = 1.0$.

V naslednjem koraku moramo opredeliti porazdelitev analizirane spremenljivke. Izhajamo iz nominalne porazdelitve, saj gre za posedovanja oziroma ne-posedovanja opazovane lastnosti (moški/ženska), zato elemente, ki so bili izbrani v vzorec, obravnavamo kot Bernoullijevo zaporedje neodvisnih poskusov z verjetnostjo P. Na tej osnovi bomo izračunali tako imenovano posteriorno porazdelitev *(angl. posterior distribution)*, ki poleg predhodne porazdelitve upošteva tudi empirične rezultate.

Bernoullijevo porazdelitev bi lahko zamenjali s hipergeometrijsko porazdelitvijo, če bi omenjenih 8 elementov izbirali iz končne populacije, čeprav že pri N = 100 elementih med porazdelitvama ni večjih razlik. Podobno bi pri majhnih verjetnostih P lahko uporabili Poissonovo porazdelitev.

Pri Bernoullijevi porazdelitvi je verjetnost, da se v n neodvisnih poskusih dogodek z verjetnostjo P zgodi ravno k-krat, enaka:

$$B_n(k) = \binom{n}{k} P^k (1-P)^{n-k}.$$

V našem primeru imamo v osmih poskusih (od tega sta bila dva ugodna) naslednji izraz:

$$B_8(2) = \frac{8 \times 7}{2 \times 1} P^2 (1-P)^{8-2};$$

ki opredeljuje porazdelitev parametra P. V drugem stolpcu $B_n(k)$ v spodnji tabeli so opravljeni izračuni za verjetnosti $B_8(2)$ pri različnih vrednostih P. Vrednost v drugi vrstici ($P = 0.1$):

$$B_n(k) = 0.1488034800$$

tako npr. pomeni verjetnost, da se v 8 poskusih opazovani dogodek (v našem primeru izbor elementa z opazovano lastnostjo), ki ima vsakič verjetnost P = 0.1, zgodi ravno dvakrat. V stolpcu $B^*_n(k)$ smo ugodne izide Bernoullijeve porazdelitve normalizirali, tako da je vsota enaka 1. Pri vrednosti $P = 0.1$ imamo sedaj:

$$B^*_n(k) = 0.1340894141,$$

ki pomeni verjetnost, da je dobljena ocena v vzorcu (p = 2/8 = 0.25) nastala pri populacijski vrednosti P = 0.1.

Zaradi predpostavke o enakomerni predhodni porazdelitvi se je nadaljnji izračun povsem poenostavil. Če pa bi zaradi vnaprejšnjega poznavanja problema ocenili, da nekatere vrednosti v populaciji niso možne (npr. deleži manjši od P = 0.2 niso mogoči), bi se zgornji izračuni dodatno izostrili. Pri izračunavanju $B^*_n(k)$ bi namreč vrednosti $B_n(k)$ predhodno pomnožili še z vnaprejšnjimi verjetnostmi za deleže P. V primeru, ko vemo, da npr. deleži pod 0.2 niso mogoči, bi vrednosti $B_n(k)$ v prvih treh vrsticah (torej vrednosti za P = 0.0, P = 0.1, P = 0.2) pomnožili s faktorjem nič. S tem bi posteriorno porazdelitev v tretji koloni ocenili bistveno natančneje.

Tabela 18: Posteriorna porazdelitev pri $B_8(2)$

P	$B_n(k)$	$B^*_n(k)$	Kumulativa $B^*_n(k)$
0.0	0.0000000000	0.0000000000	0.0000000000
0.1	0.1488034800	0.1340894141	0.1340894141
0.2	0.2936012800	0.2645692401	0.3986586542
0.3	0.2964754800	0.2671592319	0.6658178861
0.4	0.2090188800	0.1883505625	0.8541684486
0.5	0.1093750000	0.0985597223	0.9527281710
0.6	0.0412876800	0.0372050494	0.9899332203
0.7	0.0100018800	0.0090128687	0.9989460891
0.8	0.0011468800	0.0010334736	0.9999795627
0.9	0.0000226800	0.0000204373	1.0000000000
1.0	0.0000000000	0.0000000000	1.0000000000
	1.10973324000	1.0000000000	1.0000000000

Iz zgornje tabele lahko izračunamo posteriorne verjetnosti za populacijski parameter P pri dobljenem izidu (p = 0.25). Sestavljamo lahko različne trditve:

a) Verjetnost, da je bila dejanska populacijska vrednost pri dobljeni oceni p = 0.25 ravno enaka P = 0.2, razberemo v tretji vrstici tretjega stolpca:

$$B^*_8(2) = 0.2645692401.$$

Pri tem z nekaj poenostavitve upoštevamo, da sta spodnja in zgornja meja za npr. kategorijo P = 0.2 v prvem stolpcu enaki P = 0.15 oziroma P = 0.25. Ob danem rezultatu iz vzorca (p = 0.25) se torej populacijski delež P z verjetnostjo 0.265 nahaja med 15 % in 25 %. Povedano drugače: tvegamo 74 % (1 – 0.26 = 0.74), če trdimo, da je dobljeni izid nastal na osnovi parametra v intervalu $P \in (0.15, 0.25)$. Za natančnejše izračune bi sicer potrebovali podrobneje razdeljene vrednosti P v prvi koloni, npr. na 100 namesto na 10 vrednosti, vendar se rezultat ne bi bistveno spremenil.

b) Iz vsote ustreznih vrednosti v tretjem stolpcu izhaja tudi izračun verjetnosti:

$$0.134 + 0.265 + 0.267 + 0.188 + 0.099 = 0.953,$$

da je populacijska vrednost parametra P v intervalu:

$$P \in (-0.05,\ 0.55),$$

ki smo ga pred tem (napačno) izračunali kot interval zaupanja pri 5-odstotnem tveganju ob poenostavljeni predpostavki normalne porazdelitve. Enako vrednost seveda razberemo tudi iz kumulative v četrtem stolpcu, pri čemer smo ponovno predpostavili, da je zgornja meja pri kategoriji $P = 0.5$ enaka 0.55.

c) Verjetnost, da je populacijska vrednost manjša od 50 %, je ravno vsota $B^*_n(k)$ za vse verjetnosti pod $P = 0.50$. Natančnejši izračun potrdi, da je točna vrednost 0.9150124867. S skoraj 92-odstotno verjetnostjo torej lahko trdimo, da je populacijski delež $P < 0.5$. V našem primeru torej pri trditvi, da je večina bralcev ženskega spola, obstaja le 8 % tveganja, da se motimo.

d) Podobno lahko izračunamo tudi verjetnostni interval, za katerega bi s 5-odstotnim tveganjem trdili, da je v njem populacijski parameter P. Kvantilni rang z vrednostjo 0.975 namreč pripada kvantilu $P = 0.585$, kvantilni rang 0.025 pa vrednosti $P = 0.07$. Interval, v katerem je s 95-odstotno verjetnostjo populacijska vrednost P, je torej enak:

$$P \in (0.07\ ,\ 0.585),$$

kar lahko primerjamo z napačnima intervaloma (0.25 ± 0.30) in (0.25 ± 0.36), ki smo ju dobili s poenostavljenim računanjem ob zgrešeni predpostavki normalne oziroma t-porazdelitve. Očitno gre v obravnavanem primeru za določeno asimetrijo vzorčne porazdelitve.

Na osnovi zgoraj opisanih izračunov bi torej uporabniku lahko posredovali koristno informacijo o deležu žensk v obravnavani podskupini, posebej ugotovitvi (c) in (d), kjer je pravilno opredeljeno tudi tveganje za napako. Uporabnik pa bi nato, na osnovi siceršnjega tveganja pri sprejemanju poslovnih odločitvah, bodisi sprejel poslovno odločitev (npr. o oglaševalski akciji) bodisi bi naročil raziskavo na večjem vzorcu.

Zgoraj opisani pristop lahko uporabimo tudi za druge porazdelitve oziroma parametre in s tem vsakič kvantificiramo tveganje pri statističnem sklepanju v primeru majhnih vzorcev. Seveda pa bi se pri večjih vzorcih (npr. $n > 30$) verjetnostni intervali povsem približali intervalom zaupanja, ki jih poznamo iz klasičnega statističnega sklepanja.

Kritične podskupine in velikost vzorca

Zgornji rezultati omogočajo dodatno poenostavitev pri načrtovanju vzorcev. Do zdaj smo namreč velikost vzorca določali na podlagi koeficienta variacije ali s pomočjo izračunov, ki temeljijo na različnih izpeljankah absolutne mere variabilnosti, kot so:

- absolutna napaka (*angl. margin*), to je odstopanje znotraj intervala zaupanja v eno smer (npr. $1.96 \times SE$),
- standardna napaka (SE),
- skupna širina intervala zaupanja pri določenem tveganju (npr. $2 \times 1.96 \times SE$).

Velikost vzorca lahko določimo tudi enostavneje, in sicer samo na podlagi kritične velikosti podskupin v vzorcu. Pri tem v prvem koraku izračunamo potrebno velikost vzorca za podskupino oziroma za vse podskupine, v katerih želimo dosegati predvideno natančnost, v drugem koraku pa zahtevane velikosti podskupin povežemo v skupni vzorec.

V praksi lahko postopek poenostavimo in obravnavamo kritično število elementov – tipično gre za 10, 20, 25, 30 ali 100 elementov – samo v najmanjši podskupini. Če nas npr. zanimajo podskupine, določene s spolom, starostjo (3 kategorije) in petimi regijami, imamo torej $2 \times 3 \times 5 = 30$ podskupin, v katerih moramo dosegati kritično velikost, ki jo označimo z n°. Če so vse populacijske podskupine enako velike in smo se odločili za najmanjšo kritično velikost $n^\circ = 10$, potem potrebujemo skupni vzorec velikosti $n = 300$ elementov.

Običajno pa podskupine niso enako velike. Denimo, da iz populacijskih podatkov izhaja, da je najmanjša podskupina – ki bo v našem vzorcu nastopala tudi kot najmanjša celica (*angl. cell*), za katero bomo izdelali oceno – podskupina moških v starosti nad 50 let v določeni regiji. Če znaša populacijski delež navedene podskupine $P = 2\,\%$ ciljne populacije, imajo druge opazovane podskupine večji delež in nam pri vzorčenju z enakimi verjetnostmi (EPSEM) zanje ni treba skrbeti. Če smo se odločili za kritično velikost $n^\circ = 10$, moramo torej zagotoviti 10 elementov samo v tej najmanjši celici in s tem smo avtomatično zagotovili minimalno število elementov tudi v preostalih celicah. Potrebovali bi torej vzorec velikosti:

$$n = \frac{n^\circ}{P} = \frac{10}{0.02} = 500 \,.$$

Če pa bi v vsaki celici zahtevali najmanjšo kritično velikost n° = 100 elementov – kar bi omogočilo običajno natančnost ciljne spremenljivke v vsaki podskupini – bi seveda potrebovali vzorec 5,000 elementov.

Včasih želimo doseči predpisano natančnost ocen samo za vsako posamezno spremenljivko – torej posebej za spol, posebej za starost in posebej za regijo, ne da bi nas zanimala njihova presečna podskupina. V takem primeru je za določitev velikosti vzorca kritična samo najmanjša kategorija ene od spremenljivk; v obravnavanem primeru bi jo verjetno določala celica z najmanjšo regijo.

Doslej smo predpostavljali, da ocenjujemo delež izbrane podskupine ali aritmetično sredino določene spremenljivke samo na celotni podskupini v vzorcu, to je v izbrani celici. Če pa znotraj celice dodatno analiziramo deleže določene spremenljivke ali izvajamo nadaljnje analize in drobljenje elementov, se izračuni dodatno zapletejo. Če bi nas zanimal delež moških v starosti nad 50 let znotraj izbrane regije, ki gledajo določeno TV oddajo, se potrebni vzorec nadalje poveča, čeprav načela določanja velikosti vzorca ostajajo enaka. Izračuni se zapletejo tudi, če bi želeli na podlagi dejstva, da vse skupine niso enako velike, prihraniti pri velikosti vzorca, saj bi to vodilo v disproporcionalno stratifikacijo oziroma v vzorec, ki je neuravnotežen. Kljub temu pa nam opisani pristop daje nadvse praktičen napotek za hitro in enostavno določanje velikosti vzorca.

Nadaljnji korak pri poenostavljanju procesa določanja velikosti vzorca je upoštevanje tipičnih velikosti, ki so pri anketnih raziskavah običajne. Tako npr. Sudman (1976) navaja naslednjo tabelo za velikosti vzorcev:

Tabela 19: Običajne velikosti vzorcev (Vir: Sudman, 1976)

Število analiziranih podskupin	Osebe ali gospodinjstva		Organizacije ali institucije	
	Nacionalni vzorec	Regionalno, specifično	Nacionalni vzorec	Regionalno, specifično
Malo ali nič	1,000–1,500	200–500	200–500	50–200
Običajno	1,500–2,500	500–1,000	500–1,000	200–500
Nadpovprečno	2,500 +	1,000 +	1,000 +	500 +

Pri določanju velikosti vzorca je primerno opozoriti tudi na vprašanje ocenjevanja znotraj geografsko zaokroženih podpopulacij. Tako želimo ciljno populacijo pogosto analizirati po določenih regijah, npr. vzorce v Evropi analiziramo in primerjamo po državah, enako je v ZDA. Znotraj držav pa so pogoste analize po pokrajinah, tako npr. v Sloveniji brezposelnost prikazujemo znotraj uradno določenih zaposlitvenih območij.

V primerjavi s podskupinami, ki so razpršene (*angl. subclass*) po celotnem vzorcu (npr. spolno-starostne skupine), so domene (*angl. domain*) geografsko ali kako drugače strnjene. Pogosto nastopajo celo kot samostojni stratumi, v katerih izvedemo neodvisno vzorčenje. Če želimo torej dosegati želeno natančnost tudi na domenah oziroma na podpopulacijah, potrebujemo predpisano velikost vzorca prav v vsaki od navedenih podpopulacij. Velikosti vzorca se zato enostavno seštevajo. Če npr. za običajno nacionalno mnenjsko anketo potrebujemo vzorec n = 1,000 elementov, potem za 15 držav potrebujemo vzorec 15,000 oseb.

Opisana logika je tudi razlog, zakaj v večjih državah v nekaterih anketah kljub vsemu potrebujemo večje vzorce. Pogosto namreč izdelujemo ocene za standardizirane regije, ki obsegajo nekaj milijonov prebivalcev. Tako npr. evropska regijska razdelitev Slovenijo obravnava na drugi regijski ravni (NUTS II) kot eno regijo, večje evropske države pa imajo seveda po več takih regij. Če se torej na navedeni ravni (NUTS II) zahteva ocena brezposelnosti s koeficientom variacije $CV(p)$ = 0.05, bodo države z npr. petimi takimi regijami v grobem potrebovale petkrat večji vzorec kot Slovenija. Seveda pa pri mnogih nacionalnih anketah – npr. s področja rodnosti, pismenosti, javnega mnenja – tovrstne podpopulacije niso pomembne, zato velikost populacije ne vpliva na velikost vzorca.

13. PROSTORSKI VZORCI

Kadar pri izdelavi verjetnostnega vzorca nimamo seznama elementov ciljne populacije, pogosto uporabimo tako imenovane prostorske vzorce (*angl. area sample*). Prostorski vzorci se uporabljajo predvsem za vzorčenje oseb, gospodinjstev, kmetij in stanovanj, včasih pa tudi za vzorčenje poslovnih subjektov. Z nekoliko poenostavitve lahko rečemo, da je prostorski vzorec pravzaprav nadomestek za seznam stanovanj. Izdelava takega seznama je načeloma enostavna, saj sledimo zaporednemu verjetnostnemu izbiranju prostorskih enot (npr. krajevnih skupnosti, popisnih okolišev ipd.), na zadnji prostorski enoti pa enostavno popišemo vsa stanovanja in izberemo v vzorec ustrezne elemente.

Tovrstni vzorci so zgodovinsko izšli iz prostorskega načrtovanja vzorcev v ZDA, pa tudi v Indiji. V obeh primerih imamo opraviti z velikima državama in razpršenimi populacijami, kjer je bilo – ob odsotnosti administrativnih seznamov – izbiranje prostorskih enot edini način, ki je zagotavljal verjetnostno izbiro. Tudi pri mednarodnih raziskavah – npr. raziskave pismenosti, rodnosti ali zdravja, ki jih na standardiziran način vodijo strokovnjaki agencij Združenih narodov – pogosto srečamo uporabo prostorskih vzorcev.

Kadar nimamo administrativnih seznamov, so prostorski vzorci edini način verjetnostne izbire elementov tudi v majhnih državah. Res pa je, da v manjših državah ali v manjših regijah lažje najdemo sprejemljiv administrativni vzorčni okvir. Posebej v informatiziranih družbah se srečujemo s kakovostnimi administrativnimi seznami, ki so pogosto na voljo tudi v elektronski obliki.

Prostorski vzorci v ZDA

V ZDA so prostorski vzorci osnovni način verjetnostnega vzorčenja za raziskave, ki se izvajajo z osebnim anketiranjem. V nadaljevanju bomo predstavili vzorec za anketne raziskave dveh vodilnih akademskih ustanov na področju družboslovnega raziskovanja, *Survey Research Center, Institute for Social Research* (SRC) na Univerzi v Michiganu in *National Opinion Center* (NORC) v Chicagu. Vsako desetletje, po opravljenem popisu prebivalstva, organizaciji obnovita podlago za skupen vzorčni okvir s svežimi popisnimi podatki. Zahtevnost pri izdelavi tovrstnega vzorca je namreč tako velika, da sta SRC in NORC zaradi praktičnosti vzpostavili sodelovanje, čeprav sta na področju empiričnega družboslovnega raziskovanja med seboj konkurenčni. V nadaljevanju opisujemo primer prostorskega vzorca, ki sta ga organizaciji izvedli po popisu leta 1980.

Pri SRC/NORC vzorcu gre za stratificiran večstopenjski prostorski vzorec, ki uporablja izbiro z verjetnostmi, ki so v sorazmerju z ocenjeno velikostjo (*Probability Proportional to Estimated Size* – PPES). Upoštevati je treba, da je zaradi razpršenosti populacije in s tem povezanimi stroški v ZDA prisotna težnja k čim manj enotam prve stopnje, kar so potrdili tudi optimizacijski izračuni.

Enote prve stopnje (PSU) za nacionalni NORC/SRC vzorčni načrt so naslednje prostorske enote:

- standardna velemestna območja (*Standard Metropolitan Statistical Areas* – SMSA),
- okraji (*angl. counties*) ali skupine okrajev z najmanj 4,000 prebivalci ob zadnjem popisu prebivalstva,
- šestnajst največjih SMSA (med njimi New York, Chicago, San Francisco, Boston, St. Louis in Atlanta) je bilo avtomatično vključenih v vzorec. To so tako imenovane »samovključene enote prve stopnje« (*angl. Selfrepresented* PSU–SPSU), ki se obravnavajo kot stratumi. Enote druge stopnje, ki nastopajo v takih SPSU, so zato v resnici enote prve stopnje.

Poleg 16 enot SPSU je bilo s PPES vzorčenjem izbrano v vzorec 68 PSU, pri čemer je bila mera velikosti (*angl. measure of size*) število rezidenčnih gospodinjstev na podlagi zadnjega popisa. V postopku izbire so bile PSU najprej razdeljene v 68 stratumov približno enake velikosti. Stratume so oblikovali tako, da so PSU najprej razdelili na štiri pokrajine (severo-centralna, severo-vzhodna, južna in zahodna), znotraj pokrajin pa na SMSA in ne-SMSA. SMSA so bile nadalje stratificirane po zemljepisni lokaciji in velikosti največjih mest, ne-SMSA pa po zemljepisni lokaciji ter velikosti. Nato je bila s PPS izborom v vsakem od 68 stratumov izbrana po ena PSU, pri čemer so bili izvedeni še nekateri drugi postopki, ki zagotavljajo dobro zastopanost glede na razpoložljive kontrolne spremenljivke, kot so npr. delež afriških Američanov na podeželskem jugu ali delež populacije latinskega izvora na zahodu (Goodman in Kish, 1950; Hess 1975).

Naslednji korak je bil izbor enot druge stopnje (*angl. secondary sampling units* – SSU v 16 SPSU in v 68 PSU. V ta namen so bile izbrane skupine oziroma bloki v urbanih območjih, za katera so v popisnih podatkih na voljo potrebne informacije, drugod pa popisni okoliši (*angl. enumeration areas*) oziroma popisna okrožja (*angl. enumeration districts*). V vsaki taki enoti/skupini je moralo biti najmanj 48 gospodinjstev. Kjer to ni bilo doseženo, je bilo izvedeno povezovanje s sosednjimi enotami/skupinami.

Iz vsake PSU so v vzorec NORC/SRC izbrali po šest enot druge stopnje (SSU). V enotah SPSU pa so v vsaki od osmih najmanjših SMSA za vsako od organizacij (SRC in NORC) izbrali po šest enot druge stopnje (SSU) – ki pa imajo zaradi stratumske narave SPSU v resnici vlogo PSU –

medtem ko so večje SPSU zagotovile večje število enot. V newyorški SPSU so tako za glavni vzorec vsake od organizacij izbrali po 24 enot, v losangeleški SPSU pa po 18.

Izbor elementov (gospodinjstev) znotraj enot druge stopnje (SSU) je bil izveden s PPES vzorčenjem. Merilo velikosti je bilo tudi tokrat število gospodinjstev ob zadnjem popisu prebivalstva. Uporabljeno je bilo sistematično vzorčenje s predhodno urejenega seznama, kar je zagotovilo dodatno implicitno stratificiranje za izbrane spremenljivke. V SMSA in tudi v vzorčenih PSU so bile SSU predhodno urejene po okrajih, upravnih enotah, popisnem območju, okrajnih številkah ali številkah blokov. Okraji so bili urejeni glede na velikost in zemljepisno lego. V 20 državah, za katere so bili znani podatki o velikosti in mediani družinskega dohodka v manjših upravnih enotah (lokalne upravne enote), so upoštevali tudi razvrstitev po velikosti in mediani dohodka.

Enote druge stopnje (SSU) so bile različnih velikosti, od 50 gospodinjstev pa do 700 in več. Pri večjih SSU so zato izvedli še eno stopnjo vzorčenja, da bi SSU zmanjšali na obvladljivo velikost. V ta namen je bilo treba SSU razdeliti v segmente in vsakemu segmentu določiti ocenjeno velikost. Postopek je temeljil na ogledu, ki so ga opravili terenski delavci SRC/NORC. Pri tem so izvedli razdelitev gospodinjstev na segmente, nato pa s PPES izbiro iz takih SSU izbrali po en segment.

Na zadnji stopnji izbora je terensko osebje popisalo oziroma oštevilčilo (enumeriralo) stanovanja v vseh izbranih segmentih oziroma enotah druge stopnje (SSU). Tako dobljeni seznami stanovanj oziroma gospodinjstev so predstavljali vzorčni okvir za vse ankete obeh raziskovalnih organizacij.

V opisanem primeru imamo opraviti s tako imenovanim krovnim vzorcem (*angl. master sample*), s katerim se pogosto srečujemo v prostorsko velikih državah. Kadar je izdelava prostorskih vzorcev zaradi razprostranjenosti ozemlja in drugih razlogov izredno draga ali zapletena, je namreč ugodno na daljši rok (npr. vsakih 10 let) izdelati en sam večji krovni vzorec. Ko je tak krovni vzorec izdelan, se za vsako posamezno raziskavo izbirajo elementi brez dodatnega terenskega dela za potrebe vzorčenja. Pri tem se lahko način izbiranja elementov iz krovnih vzorcev od raziskave do raziskave razlikuje. Izbor se lahko razlikuje tudi znotraj enot druge stopnje oziroma znotraj segmentov. V vsakem primeru pa opisani postopek omogoča izdelavo seznama stanovanj, v katerem za vsak element natančno poznamo verjetnost, s katero je bil vključen v krovni vzorec. Na tej podlagi lahko opravimo verjetnostno izbiro elementov tudi za vsak posamezen vzorec.

Seveda je treba upoštevati, da seznami sčasoma zastarajo in jih je treba obnoviti, kar lahko izvedemo z dodatnimi navodili v fazi terenskega anketiranja, v določenih primerih pa je treba v izbranih prostorskih enotah izvesti tudi dopolnilno ali celo povsem novo popisovanje oziroma oštevilčevanje (enumeriranje) stanovanj.

Ko sta SRC ali NORC potrebovala vzorce za terenske anketne raziskave, sta torej deset let uporabljala iste enote prve stopnje. Število in izbor enot druge stopnje pa sta bila vsakič odvisna od konkretnih potreb določenega vzorca, predvsem od velikosti, ki se tipično nahaja v intervalu od 1,000 do 10,000 elementov. Pri manjših vzorcih so npr. potrebovali le po eno SSU znotraj vsake PSU, pri večjih vzorcih pa po več. Seveda pa je ena SSU lahko služila večjemu številu anket, saj npr. 50 gospodinjstev na zadnji stopnji vzorčenja načeloma omogoča uporabo desetih vzorčnih anket, če vsaka izbere le po pet elementov.

Prostorski vzorec AKP 1991–1992

V Sloveniji zaradi dobrih administrativnih virov ne potrebujemo krovnih prostorskih vzorcev. Prostorski vzorci so zato redki, saj je tak vzorec v primerjavi z vzorčenjem iz registra prebivalstva bistveno bolj zapleten. Izkušnja s pravilno izvedenim verjetnostnim prostorskim vzorcem to izrazito potrjuje. Gre za vzorec za *Anketo o kadrovskem potencialu* (AKP) 1991–1992. Ker je pred tem in po tem vzorčenje za navedeno raziskavo potekalo na podlagi registra prebivalstva, so mogoče tudi neposredne primerjave obeh postopkov (Vehovar, 1994a). Prostorski vzorec AKP je sledil standardnim navodilom za izbiro tovrstnih vzorcev:

1. Enote prve stopnje (PSU) so bile krajevne skupnosti in z načrtom PPES je bilo v prvem koraku izbranih 150 krajevnih skupnosti. Ocenjena velikost prostorskih enot je temeljila na popisnih podatkih o številu stanovanj. Ker nas je bolj kot visoka natančnost ocen za stopnjo brezposelnosti zanimalo doseganje kritične velikosti za podskupine, je vzorec z razmeroma nizkim številom enot prve stopnje dopuščal visoke vrednosti vzorčnih učinkov (*Deff*). Sredstva, ki so bila na voljo, so bila namreč namesto večje razpršenosti (in večje natančnosti) usmerjena v večjo velikost vzorca.
2. V izbranih krajevnih skupnostih sta bila v drugem koraku na enak način (PPES) izbrana po dva popisna okoliša. Zaradi spreminjanja meja krajevnih skupnosti in popisnih okolišev, predvsem pa zaradi premajhnih popisnih okolišev je bilo treba – če smo želeli zagotoviti dosledno uporabo načel verjetnostnega vzorčenja – pri tem opraviti veliko zamudnega administrativnega dela.
3. V tretjem koraku so bile v izbranih popisnih okoliših (iz registra hišnih številk) izpisane vse stavbe, ki so jih popisovalci (enumeratorji) obiskali in izdelali seznam stanovanj.
4. S seznama stanovanj so bila z metodo slučajnih števil v vzorec izbrana stanovanja, ki so jih nato obiskali anketarji in v njih anketirali vsa gospodinjstva. V povprečju je bilo v vsakem popisnem okolišu (SSU) izbranih okoli 10 stanovanj, skupno torej $150 \times 2 \times 10 = 3{,}000$ stanovanj.

Zaradi mnogih odstopanj in težav pri izdelavi vzorca je bilo potrebno tudi obsežno izračunavanje uteži. Ker je šlo pri navedenem vzorcu za panelno raziskavo, so pri tem nastali dodatni zapleti. Vsako leto je namreč v raziskavo vstopilo novih 3,000 gospodinjstev, ki so ostala v panelu po tri leta. Stroški, zamudnost in obseg strokovnega dela, ki je bil potreben za izdelavo opisanega vzorca, so bistveno presegli primerljiva prizadevanja pri vzorčenju iz registra prebivalstva. Groba ocena je, da je prostorski vzorec dražji za najmanj 30 %.

Prostorski vzorec je po drugi strani imel tudi določene prednosti. V primerjavi z registrom prebivalstva je namreč zajel vse osebe, ki prebivajo v Sloveniji, ne glede na državljanstvo in formalni/administrativni status njihove prijave.

Iz zgornjega opisa je razvidno, da je prostorsko vzorčenje drago in zamudno. Celo v primerih, ko registrskih podatkov ni mogoče dobiti, je zaradi časovnih oziroma finančnih omejitev, izdelava prostorskega vzorca tako težavna, da se pogosto odločamo za številne poenostavitve. Posebej tipična je opustitev oštevilčevanja stanovanj. Seveda pa ob takih odstopanjih v večini primerov ne moremo več govoriti o verjetnostnih vzorcih. Odstopanja sicer omogočajo minimalno ravnotežje med statistično stroko na eni strani ter zahtevami naročnikov in konkurenčnim okoljem izvajalcev anketnih raziskav na drugi strani. Kljub temu se je treba jasno zavedati, da je za statistične ocene, pri katerih želimo kvantificirati tveganje za napako, potrebno dosledno upoštevati načela verjetnostnega vzorčenja.

Vidiki prostorskih vzorcev v Sloveniji

Oglejmo si nekatere pomembnejše vidike izdelave prostorskega vzorca v Sloveniji:

- *Izbor prostorskih enot.* V Sloveniji je prek 10,000 popisnih okolišev, več kot 1,000 krajevnih skupnosti, okoli 3,000 volišč, okoli 200 občin, več kot 60 upravnih enot, okoli 10 regij ter približno 7,000 naselij. Pri izdelavi vzorcev imamo zato na vsaki stopnji veliko možnosti izbora prostorskih enot. Pogosto pa se omejimo le na dve stopnji in zato potrebujemo le eno stopnjo izbora prostorskih enot, saj na drugi stopnji vzorčimo stanovanja.
- *Število stopenj.* Tako kot pri drugih večstopenjskih vzorcih je treba upoštevati, da v Sloveniji zaradi majhnosti pogosto zadošča neposredna izbira končne prostorske enote, npr. popisnega okoliša ali volišča. Redkeje se odločamo za dva koraka izbire prostorskih enot (kar pomeni tri stopnje vzorčenja), npr. krajevna skupnost – popisni okoliši. Poseben primer trostopenjskega vzorca je tudi vzorec za anketo *Slo-*

vensko javno mnenje (1965–2000). Praktično nepoznani pa so primeri treh prostorskih korakov, npr. občina – krajevna skupnost – popisni okoliš. Zaradi majhne prostorske razpršenosti populacije so namreč potovalni stroški nizki, zato optimiziranje variance in stroškov običajno kaže k večjemu številu enot prve stopnje s po manj elementi.

- *Enote prve stopnje* (*PSU)*. Kot smo omenili že pri obravnavi razmernostne cenilke, enota izbora na prvi stopnji ne sme imeti prevelike variabilnosti v velikosti. Če je koeficient variacije za ocenjeno velikost enote prve stopnje previsok (npr. večji od 0.2), take enote kot PSU niso priporočljive. Navedena omejitev velja npr. za naselja, katerih velikost se razteza od nekaj oseb pa do več sto tisoč prebivalcev. Poleg tega za enote prve stopnje izbiramo tiste prostorske skupine, za katere obstajajo tudi urejeni uradni statistični podatki. V tem pogledu npr. skupine gospodinjstev na osnovi električnih razdelilnikov (»trafo postaje«) niso primerne, saj ne sovpadajo z uradnimi prostorskimi enotami.
- *Število enot prve stopnje*. Ker število enot prve stopnje neposredno vstopa v izraze za ocenjevanje variance (15), se je v praksi – ne glede na postopke za izračun optimalne velikosti vzorca – izoblikovalo praktično vodilo o okoli 100 enotah prve stopnje. Navedeno priporočilo glede števila enot prve stopnje – kot tudi omejitve pri variiranju v velikosti teh enot – velja za vse večstopenjske vzorce in ne samo za prostorske vzorce.
- *Število PSU v populaciji in v vzorcu*. Zaželeno je, da so enote prve stopnje v populaciji številčne, tako da je delež enot, ki jih izberemo v vzorec, majhen. Na tej osnovi so namreč pri izračunavanju variance mogoče velike poenostavitve. Zato v Sloveniji pogosto izberemo popisne okoliše kot neposredne enote prve stopnje. Razen naselij in popisnih okolišev tudi nimamo drugih za vzorčenje primernih prostorskih enot, kjer bi npr. 100 v vzorec izbranih enot predstavljalo majhen delež celotnega števila enot v populaciji. Pri večjih vzorcih in pri varčevanju s sredstvi se včasih odločamo tudi za krajevne skupnosti kot enote prve stopnje in v njih z dodatno stopnjo izberemo popisne okoliše. Naselja pa so, kot rečeno, zaradi različnih velikosti razmeroma neprimerne enote v večstopenjskem vzorcu, razen če jih pred tem stratificiramo. Podobno velja za občine in upravne enote, ki jih je poleg tega v populaciji razmeroma malo. Če namreč izbiramo enote prve stopnje med prostorskimi enotami, ki niso številčne, je tovrstni izbor treba opraviti posebej pazljivo, saj v takem primeru nastopijo posebni izrazi za ocenjevanje variance (Cochran, 1977, 9A.6).
- *Prostorske enote zadnje stopnje*. Povsod na svetu so na zadnji stopnji izbire prostorskih enot najprimernejši popisni okoliš (*angl. enumeration area*). Na tej ravni namreč razpolagamo z uradnimi popisnimi podatki o številu elementov (oseb, stanovanj, gospodinjstev). Poleg tega so popisni okoliši glede na njihov namen določeni tako, da so geografsko zaokroženi in vključujejo nekaj sto oseb. Popisni okoliši so namreč

oblikovani za potrebe desetletnega popisa prebivalstva, pri katerem morajo imeti popisovalci glede obsega terenskega dela približno enake pogoje. Popisni okoliš je tudi razmeroma stabilna enota, na katero ne vplivajo administrativno-politične spremembe (tako kot npr. na občine, volišča, volilne okraje). Zato je popisni okoliš tudi osnovni element, iz katerega so sestavljene druge prostorske enote, ki običajno ne sekajo meja popisnih okolišev. Ker pa so popisni okoliši pogosto premajhni, so na Statističnem uradu izvedli poseben postopek njihovega povezovanja. Na tej podlagi so nastale posebne *vzorčne enote* (Zaletel in Vehovar, 1998). Njihovo število je sicer bistveno manjše od števila popisnih okolišev, zato pa zagotavljajo, da je najmanjše število oseb v vsaki vzorčni enoti večje od 20 (Arnež et al., 1996).

- *Število elementov na zadnji stopnji*. Skoraj vse ankete, ki temeljijo na verjetnostnih vzorcih v Sloveniji, na zadnji stopnji izbire vključijo od 5 do 30 elementov. Analize stroškov in variance namreč skoraj nikoli ne pokažejo optimalne velikosti, ki bi presegla 20 elementov.
- *Ocenjene velikosti prostorskih enot*. Pri obravnavi večstopenjskih vzorcev smo videli, da za izdelavo takega vzorca ne potrebujemo točne velikosti enot. Zadošča že ocenjena velikost, ki je podlaga za PPES izbiro. Točna velikost enot v večstopenjskem vzorčenju sicer zagotavlja znano (fiksno) število izbranih elementov na zadnji stopnji vzorčenja, kar je seveda ugodno, ni pa nujno. Odstopanja ocenjenih velikosti prostorskih enot od dejanskih velikosti so namreč običajna, saj prostorski vzorci pogosto temeljijo na popisnih podatkih, ki so stari do 10 let. Pravzaprav je glavna prednost prostorskih vzorcev ravno v tem, da ne zahtevajo točnih podatkov o velikosti prostorskih enot. Teoretično bi zadoščala celo poljubna ocenjena velikost, določena povsem arbitrarno. Seveda pa bodo večja razhajanja med ocenjeno in dejansko velikostjo prostorskih enot povzročila na zadnji prostorski stopnji večje razlike v številu izbranih elementov, zaradi česar lahko nastane vrsto praktičnih težav. Ob tem so podatki iz popisa odlično izhodišče. Enako velja tudi za ocene iz registra prebivalstva, posebej v primeru, ko vzorčimo osebe. V takem primeru se lahko povsem približamo PPS (in ne samo PPES) izbiri.
- *Viri podatkov o prostorskih enotah*. V večini razvitih držav, in tudi v Sloveniji, so za potrebe prostorskih vzorcev na voljo številni javni podatki, predvsem statistični in geodetski viri. V Sloveniji sta posebej pomembna register prostorskih enot in register hišnih številk. Zanju ne veljajo omejitve varovanja podatkov, ki pogosto otežujejo uporabo administrativnih seznamov oseb. Seznam hišnih številk v izbranih popisnih okoliših je zato posebej dragocena pomoč pri izdelavi vzorcev, saj anketarju izročimo celoten seznam hišnih številk, s čimer na enostaven način povečamo kakovost vzorca. Tako kot drugod po svetu pa so tudi v Sloveniji osnovno izhodišče za izdelavo prostorskih vzorcev popisni podatki o številu oseb, gospodinjstev, stanovanj ipd. na

ravni osnovnih prostorskih enot. V ustreznih publikacijah Statističnega urada (npr. Rezultati raziskovanj) tako najdemo navedene podatke, zbrane na ravni naselij, na ravni upravnih enot ter na ravni krajevnih skupnosti, na voljo pa so tudi izpisi po popisnih okoliših. Pri zahtevnejših prostorskih vzorcih lahko naročimo najnovejše podatke o številu oseb v prostorskih enotah tudi iz registra prebivalstva. Pomagamo pa si lahko tudi z digitaliziranimi podatki Geodetskega zavoda, ki so posebej dragoceni pri združevanju prostorskih enot, saj so digitalizirane vse razmejitve med popisnimi okoliši in njihovimi združbami.

Poenostavitve prostorskih vzorcev

Pri poenostavitvah prostorskih vzorcev moramo ločevati poenostavitve na zadnji stopnji vzorčenja od poenostavitev na prvi stopnji vzorčenja.
1) V prvem primeru je verjetnostno vzorčenje pravilno izvedeno vse do zadnje prostorske enote (npr. popisnega okoliša), kjer se opušča verjetnostna izbira stanovanj. Posebej tipično je opuščanje pravilnega oštevilčevanja (enumeriranja) stanovanj. Namesto tega se uporabljajo drugačne sheme izbora elementov, od enostavnega izbiranja vsakega *k*-tega stanovanja s priložnostno začetno točko pa do kvotne izbire. Na tej točki torej obstajajo številne možnosti za poenostavitev prostorskih vzorcev. Čeprav nobena poenostavitev ne zagotavlja verjetnostnega vzorca, vse rešitve niso enako slabe.
2) Nevarnejše odstopanje je opustitev verjetnostne izbire prostorskih enot pri prvih stopnjah vzorčenja. Sem sodi npr. opustitev PPS izbire na prvi stopnji ob hkratnem vztrajanju pri fiksnem številu izbranih elementov v prostorskih enotah na zadnji stopnji. Tak postopek – ko prostorske enote izbiramo kar z enako verjetnostjo – je v praksi presenetljivo pogost, čeprav vodi do prevelikega števila redkeje naseljenih enot. V navedenem primeru namreč izberemo preveč majhnih naselij, preveč majhnih ulic, preveč majhnih krajevnih skupnosti ipd., skratka enot, v katerih nato izbiramo fiksno število elementov. Resno in nepredvidljivo napako lahko povzroča tudi izbor enot, za katere nimamo podatkov o verjetnosti izbora. Posebej pogosto so to ulice, za katere v publikacijah Statističnega urada nimamo podatka o velikosti in zato tudi ne moremo nadzorovati verjetnosti njihovega izbora.

Metoda slučajne poti

Odstopanja od načel verjetnostnega vzorčenja se pri prostorskih vzorcih pogosto formalizirajo v tako imenovano vzorčenje slučajne poti (*angl. ran-*

dom-route sampling). Pri tovrstnih postopkih je treba najprej določiti slučajne začetne točke, v katerih se anketar prične gibati po točno določenih pravilih slučajne poti in pri tem izbira elemente v vzorec.

Problematičnost metode se kaže v dveh vidikih. Prvi, najbolj kritičen vidik, je izbira slučajnih začetnih točk (*angl. starting points*) vzorčenja. Začetne točke namreč določajo tudi prostorske enote in ne le začetek anketarjeve slučajne poti, na podlagi katere izbira v vzorec elemente – tipično 5, 10 ali 15 elementov. Pravilni izbor začetnih točk je tako pomemben, da je treba v nekaterih državah (npr. v Nemčiji) tovrstni seznam kupiti po visoki ceni pri specializiranih organizacijah, saj je za pravilno določanje slučajnih začetnih točk treba upoštevati podatke, ki javno niso dostopni.

V Sloveniji je najboljše izhodišče za izbor začetnih točk register prebivalstva. V ta namen je treba izpisati naslove npr. 100 oseb, ki smo jih s sistematičnim izborom določili v kumulativi ciljne populacije. S tem je seznam začetnih točk povsem določen. V primeru varovanja osebnih podatkov točnega imena in priimka niti ne potrebujemo, saj zadošča že izpis ulice in hišne številke izbranih oseb.

Slučajno točko lahko izberemo tudi s postopkom prostorskega vzorčenja. Tako lahko z verjetnostno izbiro določimo npr. popisne okoliše, v katerih bi morali za pravilno izbiro začetne točke opraviti še oštevilčevanje (enumeriranje) stanovanj in tudi verjetnostno izbiro začetnega stanovanja.

V državah, kjer je telefonsko pokritje visoko, je mogoče začetne točke določiti tudi z uporabo telefonskega imenika. Tako bi lahko npr. s korakom 6,000 v imeniku 600,000 rezidenčnih telefonskih številk določili 100 rezidenčnih številk, ki predstavljajo naslove začetnih točk. Pri večjem telefonskem nepokritju je tak pristop seveda neugoden, enako tudi v primeru, ko ne razpolagamo z ažurnim telefonskim imenikom v elektronski obliki. Izbira iz tiskanega telefonskega imenika je namreč – kot bomo videli v poglavju o telefonskem vzorčenju – nekoliko bolj tvegana.

Telefonski imenik se lahko uporabi ne samo za izbor slučajnih začetkov, ampak tudi za samo izbiro elementov (gospodinjstev). Z določenim korakom namreč lahko za slučajno izbranim začetnim elementom iz telefonskega imenika vključujemo še nadaljnje elemente in s tem poenostavimo izbiro vzorca. Seveda pa je vzorec v takem primeru manj kakovosten kot v primeru, ko iz telefonskega imenika izberemo samo začetne točke, ostale elemente pa izberemo z verjetnostno izbiro na podlagi terenskega dela. Izbira vseh elementov neposredno iz telefonskega imenika namreč povečuje pomanjkljivosti telefonskega vzorčnega okvira, predvsem neažurnost, nepokritje in tajne številke.

Najpogostejša napaka pri izbiri slučajnih točk je uporaba metod, ki temeljijo izključno na prostorskih informacijah, brez upoštevanja podatkov o številu oseb. Oglejmo si primer. Denimo, da smo v vzorec z verjetnostno izbiro vključili določen popisni okoliš, naselje, krajevno skupnost, volišče ali ulico. Izbrati moramo le še slučajni začetek. Če nimamo seznama oseb, stanovanj ali gospodinjstev, bi morali za zagotovitev verjetnostnega vzorca seveda najprej opraviti oštevilčevanje (enumeriranje) stanovanj. Pri povsem slučajnem izboru stavbe (hišne številke), ki ga opravimo iz seznama stavb bomo namreč prepogosto izbrali redkeje naseljene stavbe in redkeje naseljene soseske. Stolpnice in individualne hiše imajo namreč pri taki strategiji enako verjetnost za izbor, kar daje osebam, ki prebivajo v individualnih hišah, preveliko verjetnost izbire. Podobna nevarnost velja tudi za tako imenovano »slepo« izbiranje začetne točke iz zemljevida oziroma mestnega načrta. Tudi v tem primeru imajo osebe v večjih stavbah premajhno verjetnost izbora. Ker imata pri takem postopku npr. enostanovanjska hiša s 5 osebami in stolpnica s 500 osebami enako verjetnost izbora, bodo osebe v stolpnici izbrane s 100-krat premajhno verjetnostjo. Problem lahko omilimo s povečano verjetnostjo izbora večstanovanjskih stavb, s povečanim številom izbranih stanovanj v večstanovanjskih stavbah in z uteževanjem. Seveda pa vse to ne more v celoti nadomestiti pravilnega postopka verjetnostne izbire.

V drugem koraku preide metoda slučajne poti v pravila, ki določajo anketarjevo gibanje na terenu in izbiro stanovanj. Pravila so na tej točki lahko zelo stroga in obsežna, saj morajo predvideti izredno pestrost dogajanja na terenu.

Anketarji se morajo v izbrani začetni točki (*angl. starting point*) običajno postaviti v določeno nebesno smer, nadaljevati gibanje v točno določeni smeri (npr. gibanje v smeri urinega kazalca) in po predpisanem ključu izbirati v vzorec stanovanja, na katera naletijo na svoji poti. Pri večstanovanjskih hišah veljajo dodatna in razmeroma zapletena pravila. Ker so stavbe, stanovanja, predvsem pa ulice in ceste urejene na brezštevilne – in včasih povsem nelogične – načine, pravila slučajnostnih poti niti teoretično niso enolična. Toliko manj so enolična v praksi, kjer se soočajo s specifičnim razumevanjem anketarja in tudi z nevarnostjo zavestnega prilagajanja pravil konkretnim razmeram. Zato tudi v primeru, ko smo izbrali popolne začetne vzorčne točke, metoda slučajne poti ne zagotavlja vsem elementom (to je stanovanjem) vnaprej znane verjetnosti za izbor (Hoffmeyer-Zlotnik in Krebs, 1996). Kljub temu pa lahko dobro zasnovana pravila, izšolani anketarji in izdelane kontrole pripomorejo, da tovrstna izbira povsem zadošča za večino marketinških in mnenjskih raziskav.

Dodati velja, da je v Sloveniji pri določanju slučajne poti odličen pripomoček seznam hišnih številk, kar v veliki meri poenostavlja predpisovanje slučajne poti in tudi nadzor anketarjevega gibanja.

14. TELEFONSKI VZORCI

Telefonsko anketiranje predstavlja zaradi vrste razlogov najpogostejši način anketiranja. Med pomembne prednosti telefonskega anketiranja, ki so pripomogle k njegovi razširjenosti, sodijo tudi postopki vzorčenja, ki omogočajo razmeroma enostavno izdelavo verjetnostnih vzorcev.

Razvoj telefonskega vzorčenja

Telefonsko anketiranje se je začelo uveljavljati v 60-ih letih, čeprav se je uporabljale že mnogo prej. Razvoj je najhitreje potekal v ZDA, kjer je telefonsko pokritje že pred 2. svetovno vojno preseglo 30 %, v petdesetih letih se je povzpelo na 60 %, že v začetku sedemdesetih let pa je preseglo 90 %. Ovira pa ni bilo samo telefonsko pokritje, ampak tudi tehnične težave, saj so vse do šestdesetih let zapleteni sistemi prevezovanja ter dragi medkrajevni pogovori oteževali normalno pogovorno uporabo. Razmeroma zgodaj se je pojavilo tudi vprašanje vzorčnih okvirov, saj gospodinjstva zaradi agresivne telefonske prodaje ne želijo imeti telefonskih številk v javnem imeniku.

Tehtni pomisleki so v začetku izhajali tudi iz metodoloških razlogov. Šele v sedemdesetih letih so v ZDA opravili poglobljene študije, ki so preverile kakovost tovrstnega anketiranja (Groves in Kahn, 1979). Izkazalo se je, da sam način anketiranja (*angl. survey mode*) ni sporen in tudi nepokritost (*angl. noncoverage*) ni prinašala večjih napak. Z uporabo posebnega dvostopenjskega postopka telefonskega vzorčenja W-M (Waksberg, 1978), kar bomo v nadaljevanju podrobno opisali, je bila v ZDA odpravljena še zadnja ovira za splošno uporabo telefonskih anket. V osemdesetih letih se je telefonsko anketiranje v razvitem svetu izredno razmahnilo in postalo tako imenovana nadomestna tehnologija (*angl. replacement technology*) za osebno anketiranje. Posebej marketinška industrija je cenovno prednost v celoti izkoristila, saj predstavljajo v geografsko velikih državah stroški telefonskega anketiranja pogosto le tretjino stroškov osebnega anketiranja. Na telefonsko anketiranje so hitro prešli tudi v akademskih in uradnih anketnih raziskavah.

V Evropi je bila širitev telefonskega anketiranja kasnejša in bolj specifična. V mnogih državah je zaradi zapletenega sistema telefonskih številk tovrstno anketiranje precej oteženo. Tako so šele v zadnjih letih izdelali približne postopke za izdelavo verjetnostnih vzorcev gospodinjstev s telefonom v Veliki Britaniji (Nicolaas et al., 2000) in Nemčiji (Gabler and Häder, 2000). Po drugi strani se ponekod v Evropi še vedno borijo z nizkim pokritjem (npr. Vzhodna Evropa), v komunikacijsko najrazvitejših državah pa se že soočajo s problemom mobilne telefonije (npr. Skandinavija), ki dramatično spreminja načine telefonskega vzorčenja. Tako je npr. na Finskem že v letu 1999 petina prebivalstva uporabljala samo mobilni telefon in v gospodinjstvu ni imela fiksnega telefona.

Dodatni impulz k razmahu telefonskega anketiranja je prinesla tudi tehnologija računalniško podprtega anketiranja, ki je bila v prvi fazi namenjena predvsem za podporo telefonskemu anketiranju in se je šele kasneje selila v druge načine anketiranja: osebno anketiranje, samoanketiranje (*ang. self-administered surveys*) ter spletno anketiranje (*angl. Web survey*). Računalniško podprto telefonsko anketiranje je tako doseglo tehnološko zrelost že konec osemdesetih let.

V začetku novega tisočletja tudi na tem področju občutimo hiter razvoj telekomunikacij. Veliki anketni centri s cenitvijo mednarodnih impulzov namreč lahko izvajajo globalne telefonske ankete iz multietičnih lokacij (npr. London, Los Angeles, New York). Selitev upravljanja celotnega procesa telefonskega anketiranja na svetovni splet (WWW) pa omogoča izvedbo distribuiranega anketiranje na večih lokacijah, predvsem pa kombinacijo s spletnim anketiranjem, kar prinaša novo razsežnost v razvoju anketne industrije.

Določene spremembe na področju telefonskega anketiranja nastajajo tudi zaradi tehnologije prepoznavanja govora (*angl. voice recognition*). Nekatere enostavne in standardizirane ankete pa je mogoče opraviti brez anketarja tudi tako, da anketiranec odgovore (na vnaprej posneta vprašanja) enostavno vtipka v telefonsko številčnico. Tovrstne pocenitve in poenostavitve sicer nimajo neposrednih posledic za telefonsko vzorčenje, res pa je, da postaja z njimi vzorčenje manj pomembna komponenta celotnega procesa.

Pomembno spremembo v telefonskem vzorčenju so povzročili tudi sistemi vnaprejšnje izbire (*angl. predictive dialing)*, kjer računalnik izbere telefonsko številko, počaka, da klicani dvigne slušalko, in ga šele nato preveže k anketarju. Navedeni postopek lahko pri krajših anketah poveča izkoristek anketarjevega časa od običajnih 30–50 % na več kot 80 %, kar je izjemen prihranek. Pri tem pa nastaja neprijeten tehnični problem značilnega zvoka, ki čakajočega anketiranca spominja na telefonsko prodajo, kjer se tovrstni sistemi posebej intenzivno uporabljajo. Navedena omejitev je pomembno zavrla širitev sistemov vnaprejšnje izbire.

Sistem vnaprejšnje izbire vsekakor poenostavljajo celotno problematiko vzorčenja. Načeloma je namreč mogoče povsem slučajnostno (npr. SRS) generirati in preklicati vse 9- ali 10-mestne številke, odstraniti vse javne poslovne številke, nato pa številke, kjer se anketiranec odzove, prevezati na anketarja. Izbor telefonskega vzorca je v tem primeru nadvse enostaven, saj prazne in napačne telefonske številke ter številke, kjer se nihče ne javlja, ne pomenijo časovnega ali finančnega bremena.

Čeprav se postopek vnaprejšnje izbire pri anketnem raziskovanju (vsaj v začetku novega tisočletja) ni posebej razširil, je njegovo temeljno načelo nadvse privlačno. To posebej velja za velike anketne centre z več kot 100 delovnimi postajami, kjer je mogoče brez težav upravljati proces tako, da imamo vedno na voljo prostega anketarja, na katerega prevežemo novega anketiranca.

V državah, kjer telefonski operaterji to omogočajo, se pogosto uporablja samo ena komponenta postopka vnaprejšnje izbire, in sicer vnaprejšnje preverjanje aktivnosti telefonskih številk. V tem primeru se po izboru telefonskih številk še pred izvedbo anketiranja na vse izbrane številke odpošljejo posebni impulzi, ki preverijo – ne da bi bili potencialni anketiranci moteni – aktivnost vseh izbranih telefonskih številk. Opisani postopek preverjanja se običajno izvede ponoči, njegov rezultat pa je vzorec, ki vsebuje le aktivne telefonske številke, saj odkrije in izloči prazne oziroma nezasedene številke.

Telefonski vzorci v ZDA

Ker skoraj vsa ameriška gospodinjstva posedujejo telefon, so telefonske ankete že več desetletij privlačen način zbiranja anketnih podatkov. Kljub temu je treba upoštevati, da je v ZDA največ gospodinjstev brez telefona med prebivalstvom z nizkim dohodkom, med ne-belci in med mlajšimi od 35 let. Tako je npr. stopnja brezposelnosti med gospodinjstvi brez telefona nekaj desetkrat večja kot sicer, kar lahko vpliva na pristranskost celo pri 95-odstotnem telefonskem pokritju. V primerih, ko raziskava zahteva zastopanost teh podskupin, je telefonskemu vzorcu primerno dodati še prostorski vzorec (Groves in Lepkowski, 1982), kar vodi k podvojenim okvirom (*angl. dual frames*). V primeru panelnih anket je zato primerno, da je prvi kontakt opravljen osebno. Tako se npr. v mesečni Anketi o delovni sili (*Current Population Survey* – CPS) prvi stik vedno izvede z osebnim anketiranjem.

Telefonski imeniki so v ZDA razmeroma neustrezen vzorčni okvir, saj mnoge telefonske številke niso v imeniku; ponekod manjka celo več kot polovica gospodinjstev. Pomemben razlog za to so tudi preselitve – v ZDA se letno preseli 20 % prebivalcev (v Sloveniji 2 %).

Zaradi težav z vzorčnimi okviri so raziskovalci predlagali vrsto alternativnih rešitev na osnovi slučajnega generiranja številk (*angl. Random Digit Dialing* – RDD), od prištevanja konstante vsaki rezidenčni številki v imeniku do variiranja zadnjih dveh cifer (Frankel in Frankel, 1977). Običajno pa take metode ne zagotavljajo verjetnostne izbire.

Pri verjetnostnem RDD vzorčenju je treba upoštevati, da so v ZDA telefonske številke sestavljene iz desetih cifer v treh delih, npr. 301-555-

1212. Prvi del je območna koda (*angl. area code*), drugi je lokalna koda (*angl. central office code*), zadnje štiri cifre pa predstavljajo lokalni dodatek (*angl. suffix*). V ZDA je več kot 100 območnih kod, več kot 30,000 lokalnih kod, zadnje štiri cifre pa znotraj vsake območno-lokalne kode generirajo nadaljnjih 10,000 telefonskih številk. Seveda je večina teh številk nedelujočih ali pa gre za poslovne številke. V najenostavnejšem primeru generiranja RDD telefonskega vzorca je zato treba najprej slučajno izbrati prva dva dela, to je prvih šest cifer območnih in lokalnih kod, nato pa še naključno štirimestno številko od 0000 do 9999. Opisani postopek sicer pokrije vse telefonske številke, vendar je delež neustreznih – gre za prazne (nedelujoče številke) in tuje elemente (poslovne številke) – v takem primeru izredno visok, običajno nad 60 %. Neustrezne elemente seveda zavržemo, vendar je tak postopek drag zaradi številnih neuspešnih klicev.

Alternativni postopek RDD vzorčenja je opisal Waksberg (1978) v tako imenovanem W-M (Waksberg-Mitofski) postopku. Pri tem postopku izhajamo iz skupin – imenujemo jih bloki *(angl. banks)* – telefonskih številk po 100 elementov, ki so določeni z variiranjem zadnjih dveh cifer. Posamezen blok torej določimo na podlagi kombinacije območne kode, lokalne kode ter prvih dveh cifer lokalnega dodatka, torej na prvih osmih cifrah telefonskih številk. Tako izbrane bloke nato uporabimo kot enote prve stopnje (PSU) v dvostopenjskem načrtu. Izbiramo jih z enako verjetnostjo, nato pa v izbranih blokih slučajno izberemo eno telefonsko številko. Če izbrana številka ni rezidenčna telefonska številka, zavržemo celoten blok. Pri rezidenčni številki izvedemo anketo, nato pa v takem bloku naključno izberemo še druge številke, dokler ne dosežemo potrebnega, vnaprej določenega števila b rezidenčnih gospodinjstev, ki jih moramo anketirati v vseh vključenih blokih.

Verjetnost, da bo določen blok α izbran in tudi sprejet v vzorec – pogoj za to je seveda dejstvo, da je bila prva slučajno izbrana številka rezidenčna – je torej sorazmerna z deležem rezidenčnih telefonskih številk v tem bloku:

$$P_\alpha = \frac{B_\alpha}{100},$$

kjer je B_α skupno število rezidenčnih številk v bloku. Verjetnost, da bo določena rezidenčna telefonska številka izbrana, če smo izbrali blok α, pa je enaka:

$$\frac{b}{B_\alpha},$$

kjer je b fiksno število rezidenčnih številk, ki jih izberemo v blokih, v katerih je bila prva številka rezidenčna. Tako je verjetnost za izbor rezidenčne telefonske številke β v bloku α sorazmerna:

$$P(\alpha\beta) = \frac{B_\alpha}{100} \times \frac{b}{B_\alpha} = \frac{b}{100}.$$

Opisana shema daje vsem rezidenčnim številkam enako verjetnost izbora (EPSEM). Bloki (PSU) so namreč izbrani sorazmerno z velikostjo (PPS), pri čemer je velikost bloka ravno B_α, to je število rezidenčnih telefonskih številk v bloku. Vrednosti B_α seveda ne poznamo, vendar je elegantnost opisanega postopka ravno v tem, da vrednosti B_α ni treba poznati, saj se v zgornjem izrazu okrajšajo.

Seveda pa vsaka uporaba dvostopenjskega vzorčenja pomeni tudi manj natančne ocene, saj je odgovarjajoči vzorčni učinek *Deff* skoraj vedno večji od 1. Dvostopenjsko vzorčenje je torej upravičeno le, če z njegovo uporabo prihranimo glede na stroške in natančnost. Običajno pa je vzorčni učinek majhen v primerjavi s prihranki zaradi manjšega števila neuspešnih klicev. Ker se rezidenčne telefonske številke gospodinjstev gostijo v skupine, sta namreč pri W-M postopku približno dve od treh izbranih številk tudi rezidenčni telefonski številki, pri enostavni SRS shemi slučajnega generiranja pa je npr. med petimi le ena taka številka. Prihranek pri stroških zato v opisanem W-M postopku običajno presega minimalno izgubo v natančnosti zaradi vzorčnega učinka.

Dodati je treba, da je v devetdesetih letih v ZDA začelo prevladovati vzorčenje s pomočjo seznamov (*angl. list-assisted telephone sampling*). V tem primeru je izhodišče vzorčnega okvira seznam vseh blokov, v katerih je vsaj ena javna rezidenčna številka, to je številka, ki jo je bilo mogoče zaslediti v javnem telefonskem imeniku. S tem se seveda zavestno odpovemo gospodinjstvom, ki so v blokih, v katerih ni nobene javne rezidenčne številke, vendar so analize pokazale, da zaradi tega nastaja le majhna pristranskost. Še več, razmeroma majhno pristranskost tvegamo celo, kadar izpuščamo tudi bloke, v katerih je samo ena javna rezidenčna številka.

Na tej poenostavitvi je mogoče zgraditi vrsto različnih postopkov. Najenostavnejša možnost je enostavna slučajna (SRS) izbira blokov, nakar v drugem koraku sledi izbira ene same telefonske številke iz vsakega bloka. Za vse izbrane številke se nato preizkusi – z avtomatiziranim preverjanjem, ki smo ga omenili pri obravnavi postopkov vnaprejšnje izbire – njihova dejanska aktivnost in odstranijo se vse javne poslovne številke. Opisani postopek zagotavlja EPSEM vzorec, saj gre za običajni dvostopenjski PPS vzorec, pri katerem na drugi stopnji izbiramo le en element (Connor in Heeringa, 1992).

Glavna prednost opisanega postopka pred klasično W-M metodo je dejstvo, da ne potrebujemo sekvenčnega (zaporednega) preverjanja rezidenčnosti. Pri W-M metodi je namreč treba v vsakem izbranem bloku postopno dodajati številke, dokler ne dosežemo želenega fiksnega števila rezidenčnih številk *b*. Za vključevanje dodatne številke pa je treba predhodno razčistiti status vseh številk, ki so bile vključene v določenem bloku. Anketna raziskava na podlagi W-M metode zato traja bistveno dlje.

Navedeni pomanjkljivosti W-M postopka se sicer lahko izognemo s fiksno izbiro številk v vsakem bloku. Tako npr. lahko v vsakem izbranem bloku pokličemo 15 telefonskih številk. Pri tem seveda končamo s precej različnim številom rezidenčnih številk v posameznih blokih, kar nadomestimo z uteževanjem (Brick in Waksberg, 1992). Seveda se moramo v takem primeru – zaradi izrazitega variiranja v številu rezidenčnih stanovanj v izbranih blokih – sprijazniti z znatnim povečanjem vzorčne variance zaradi uteževanja. Ker pa gre za variabilne uteži, je tovrstno povečanje vzorčne variance manjše kot je to običajno pri fiksnih utežeh (Vehovar, 1999).

Zgoraj opisani postopki telefonskega vzorčenja se uporabljajo predvsem pri vzorcih v zahtevnejših uradnih in akademskih anketah. Za večino marketinških anket pa največkrat zadoščajo že narejeni vzorci, ki jih prodajajo specializirane komercialne organizacije.

Glede stratifikacije so pri telefonskem vzorčenju možnosti razmeroma omejene. Največkrat namreč razpolagamo le s podatkom o pripadnosti telefonske številke določeni območni oziroma lokalni kodi in na tej podlagi lahko vzorec stratificiramo samo s pomočjo geodemografskih podatkov (Groves in Kahn, 1979).

Telefonski vzorci v Sloveniji

V Sloveniji se je telefonsko anketiranje izraziteje pojavilo konec osemdesetih let, prve računalniško podprte telefonske ankete pa šele v devetdesetih letih. Tudi telefonsko pokritje *(angl. telephone coverage)* je šele v začetku devetdesetih let preseglo 60 %. Sledila je hitra rast telefonskega pokritja, konec devetdesetih let pa se je dokončno ustalil in profesionaliziral tudi trg ponudnikov telefonskega anketiranja.

V začetku novega tisočletja je telefonsko anketiranje v Sloveniji verjetno doseglo vrhunec svojega razvoja. Eden pomembnejših razlogov za to je visoko telefonsko pokritje, ki je preseglo 90 %, kar dopušča le še majhno nevarnost pristranskosti. Iz izrazov za neodgovore (26), ki jih lahko uporabimo tudi pri nepokritju, namreč izhaja, da znaša npr. pri ocenjevanju deležev pristranskost kvečjemu desetinko (odstotek nepokritja) razlike med deležem spremenljivke v »telefonski« in v »ne-telefonski« populaciji:

$$Bias(p_t) = 0.1 \times (P_t - P_n).$$

V primeru, da določeno stališče podpira npr. P_t = 15 % telefonske populacije in P_n = 20 % ne-telefonske, bomo pri ocenjevanju na podlagi telefonske populacije podcenili pravo populacijsko vrednost (P = 15.5 %) za:

$$Bias(p_t) = 0.1 \times (0.15 - 0.20) = -0.005 = -0.5\%,$$

to je za samo pol odstotne točke, kar seveda izhaja tudi neposredno:

$$P - P_\pi = 15.5\% - 15\% = -0.5\%.$$

V praksi je taka pristranskost pogosto skoraj povsem zanemarljiva, ker so razlike med telefonsko in ne-telefonsko populacijo običajno majhne in skoraj nikoli ne dosežejo obsega kot v navedenem primeru. Kljub temu moramo biti pozorni na spremenljivke, ki so občutljive na posedovanje telefona. Tipična taka spremenljivka je npr. modemski dostop do interneta od doma, saj gospodinjstva brez telefona nimajo modema. Raziskave pa kažejo, da so s posedovanjem telefona v gospodinjstvu poleg nekaterih sociodemografskih značilnosti močneje povezane tudi spremenljivke s področja religije oziroma vernosti (Vehovar, 1990).

K uspešnosti telefonskega anketiranja v Sloveniji prispevajo tudi razmeroma visoke stopnje sodelovanja. Pri zbiranju podatkov v okviru mednarodne študije (Hox in deLeeuw, 2001) so bile v slovenskih raziskovalnih organizacijah ugotovljene razmeroma visoke stopnje odgovorov, čeprav jih zaradi posebnosti posameznih raziskav (število klicev, vsebina ankete, zahtev naročnikov, dolžina ankete ipd.) med seboj ne moremo primerjati. Podatki so bili namreč zbrani (Rutar, 2000) za različne ankete *Statističnega urada Republike Slovenije, Fakultete za družbene vede (Center za proučevanje javnega mnenja in množičnih komunikacij)* in marketinških raziskovalnih podjetij (*Cati center, IRM-Mediana, Gral-Iteo*).

Verjetno najugodnejše sodelovanje v telefonskih anketah v Sloveniji pa je bilo doseženo v anketi *Viktimizacija* leta 1992. V tej raziskavi se je uporabilo predhodno pisno obvestilo vsem sodelujočim, postopek 20 telefonskih klicev ter spreobračanje zavrnitev *(angl. refusal conversion)*. Anketa je imela tudi izrazito akademsko naravo, saj sta jo izvajali *Pravna fakulteta, Inštitut za kriminologijo* in *Fakulteta za družbene vede, Center za metodologijo in informatiko*. Tako visoko stopnjo odgovorov je danes vsekakor težko ponoviti.

Tabela 20: Stopnje odgovorov v anketi *Viktimizacija*, 1992

Stopnja ustreznosti	93%
Stopnja odgovorov	90%
Stopnja zavrnitev	3%
Stopnja anketiranja	85%

K ugodnim pogojem telefonskega anketiranja prispevata tudi razmeroma skromna in neagresivna telefonska prodaja *(angl. telemarketing)* ter siceršnja nezasičenost s telefonskimi anketami. Še v letu 1996 je namreč polovica aktivne populacije izjavljala, da v preteklih 12 mesecih ni bila vključena v nobeno anketo (Vehovar et al., 1996).

K razširjenosti telefonskega anketiranja v Sloveniji nenazadnje prispeva tudi ažuren in javno dostopen telefonski imenik v elektronski obliki. Pri tem je bil odstotek tajnih telefonskih številk, ki niso objavljene v javnem telefonskem imeniku, konec devetdesetih še vedno razmeroma nizek – pod 5 %.

V prihodnje lahko pričakujemo slabšanje trendov. Delež tajnih številk bo naraščal, upadala bo pripravljenost anketirancev za sodelovanje, tehnologija pa bo omogočila vse večjo zaščito pred nezaželenimi klici. Povsem mogoče so tudi zakonske omejitve pri poseganju v zasebnost.

Večalo se bo tudi število gospodinjstev, ki imajo mobilni telefon, ne pa tudi fiksnega telefona. Že v letu 2000 je bilo po podatkih ankete *Slovensko javno mnenje* v Sloveniji 3 % takih gospodinjstev. Mobilna telefonija vsekakor predstavlja naraščajoč problem telefonskega anketiranja in odpira povsem novo poglavje na področju izdelave telefonskih vzorcev.

Kot rečeno so pomembna ovira za telefonsko anketiranje tudi tehnične omejitve, ki jih omogoča sodobna tehnologija. Gre za telefonske tajnice, kontrole za vhodne (klicoče) številke in za inteligentne vmesnike, ki se na klic odzovejo z ustreznim obnašanjem. Tako se lahko po klicu izbrane številke vsakič sproži avtomatski poziv, da vtipkamo nadaljnjo izbiro, npr. 1 za osebni klic, 2 za poslovni klic, 3 za telefonsko prodajo, 4 za telefonsko anketiranje. Pri tem se – glede na programirano prioriteto, ki jo uporabnik nameni tovrstnim klicem – bodisi dovoli takojšen osebni kontakt bodisi se klic zavrne ali preusmeri v govorni predal, na elektronski naslov, spletno stran ipd. Z razvojem tehnologije prepoznavanja govora pa lahko pričakujemo še dodatne omejitve.

Vzorci iz tiskanih imenikov

Pri izbiri telefonskih številk iz tiskanega imenika je treba – poleg težav s tajnimi številkami – upoštevati predvsem problem neažurnosti. Ob večjih spremembah telefonskih številk so namreč tiskani imeniki seveda povsem neuporabni.

Ker govorimo predvsem o telefonskih vzorcih gospodinjstev, se moramo soočiti tudi s problemom poslovnih in drugih nerezidenčnih telefonskih številk. Pri ločevanju rezidenčnih (gospodinjstva) od ostali (poslovnih) telefonskih številk namreč lahko nastajajo znatne težave. Pravilen pristop je vsekakor vnaprejšnja ocena deleža poslovnih številk, nakar za ustrezni odstotek povečamo izhodiščni vzorec. Vsekakor pa je treba nerezidenčne številke, ko naletimo nanje, izpustiti tako kot vse neustrezne elemente. V primeru, ko bi namesto poslovne številke izbrali najbližjo (naslednjo) rezidenčno številko, namreč končamo s prevelikim številom priimkov blizu nazivom poslovnih številk, ki smo jih preskočili. Tako ima npr. priimek »Zavodnik« kot prvi priimek, ki sledi množici neustreznih (nerezidenčnih) elementov z nazivi »Zavod za ...«, preveliko verjetnost, da se pojavi v vzorcu.

Težavo lahko povzročijo tudi nesorazmerno razporejeni bloki oglasov. Pri izbiranju telefonskih številk je zato povsem pravilen le sistematičen izbor številk z vrvico, ki omogoča, da je interval med dvema številkama določen s fiksno dolžino. Dober približek je lahko v določenih primerih tudi zaporedna izbira strani z določenim korakom, npr. vsaka deseta stran, pri čemer na izbranih straneh opravimo nadaljnjo stopnjo vzorčenja, npr. sedmo in dvajseto vrstico. Ker pa je – vsaj v primerjavi z običajno velikostjo vzorca – strani v tiskanem imeniku razmeroma malo, moramo na izbranih straneh vključiti v vzorec več kot eno telefonsko številko, zaradi česar se pojavijo neugodni učinki vzorčenja v skupinah.

Podobno kot pri prostorskih vzorcih obstajajo v praksi številne poenostavitve tudi pri izboru iz tiskanega telefonskega imenika. Vsem tovrstnim poenostavitvam pa je skupno, da prinašajo večja ali manjša odstopanja od verjetnostnega vzorčenja in s tem tudi bolj ali manj neznano nevarnost pristranskosti ocen.

Vzorci iz elektronskih imenikov

Telekom Slovenije omogoča enostavno izdelavo vzorcev iz telefonskega imenika na podlagi standardiziranih programov za izdelavo slučajnih seznamov rezidenčnih številk. Poslovne številke so posebej označene, zato njihova izločitev ne povzroča nobenih težav. Edina večja skrb torej je, da izberemo v vsaki regiji delež, ki ustreza številu prebivalstva. Posebej kritična je ljubljanska regija, v kateri je delež rezidenčnih številk bistveno večji od deleža prebivalstva.

Zaradi ažurnosti, majhnega števila tajnih številk in dostopnosti izdelave vzorcev iz telefonskega imenika alternativne metode telefonskega vzorčenja v Sloveniji niso posebej aktualne. Postopki slučajnega generiranja telefonskih številk (RDD), kjer se telefonske številke generirajo računalniško, dajejo zaradi velikega števila praznih številk izjemno nizek izkoristek. Nekatere izkušnje kažejo na izkoristek pod 20 %, kar pomeni, da je med slučajno generiranimi telefonskimi številkami le vsaka peta številka rezidenčna številka gospodinjstva. Nekoliko boljši izkoristek dajejo metode variiranja zadnjega števila. V takih primerih lahko iz osnovnega vzorca telefonskih številk generiramo devet novih vzorcev z enostavnim spreminjanjem zadnje cifre. Tako npr. telefonska številka 1-234-5678 določa devet novih telefonskih številk: 1-234-5679, 1-234-5670, 1-234-5671 itd. Osnovna ideja takega postopka je seveda v predpostavki, da se rezidenčne številke gostijo v skupine.

Z naraščanjem deleža tajnih številk in z morebitno uporabo tehnik vnaprejšnje izbire pa lahko tudi v Sloveniji pričakujemo razmah alternativnih metod RDD vzorčenja telefonskih številk.

Vidiki telefonskega vzorčenja

V nadaljevanju se bomo na kratko dotaknili nekaterih pomembnejših vidikov telefonskih vzorcev, ki so posredno povezani s problematiko vzorčenja. Sicer pa je telefonsko vzorčenje izčrpno obdelano v mnogih odličnih učbenikih (Lavrakas 1993, Groves et al., 1988).

- *Velikost gospodinjstev*. Če govorimo o vzorcih z enakimi verjetnostmi (EPSEM), je treba upoštevati, da dajejo telefonski vzorci v primerjavi z vzorci oseb (npr. *Slovensko javno mnenje*) pravilne (enake) verjetnosti za vključitev gospodinjstev in napačne (neenake) verjetnosti za vključitev oseb. Oseba v 10-članskem gospodinjstvu ima tako 10-krat manjšo verjetnost, da se pojavi v telefonskem vzorcu, kot pa oseba v enočlanskem gospodinjstvu. Pri vzorčenju oseb namreč potrebujemo 6 % oseb iz enočlanskih gospodinjstev, kolikor je v slovenski populaciji leta 2000 oseb, ki živijo v enočlanskih gospodinjstvih. Po drugi strani pa v običajni telefonski vzorec načeloma vključimo 18 % takih oseb. Takšen je namreč delež enočlanskih gospodinjstev v slovenski populaciji gospodinjstev. Če torej na podlagi telefonskega vzorca analiziramo populacijo oseb, je potrebno dodatno uteževanje, ki je obratno sorazmerno s številom ciljnih oseb v gospodinjstvu. Tovrstno uteževanje pa se v praksi običajno ne opravi, in to brez večjih posledic. Razloga za to sta dva. Prvič, visoke stopnje neodgovorov med enočlanskimi gospodinjstvi avtomatično poskrbijo za manjše število oseb iz takih gospodinjstev. Drugič, mnenjske spremenljivke so razmeroma neobčutljive za tovrstna odstopanja. Kljub temu obstajajo spremenljivke, npr. število osebnih avtomobilov, skupni dohodki gospodinjstva, osebni računalnik ipd., ki so močneje povezane z velikostjo gospodinjstva.
- *Število telefonskih priključkov v gospodinjstvu*. Gospodinjstva imajo lahko več telefonskih priključkov in s tem tudi večjo verjetnost za izbor v vzorec. Gospodinjstvo z npr. dvema telefonoma bi torej moralo dobiti dodatno utež $W = 1/2$.
- *Izbor oseb v gospodinjstvu*. Kakovostna telefonska anketa mora poskrbeti za nadzorovano izbiro oseb v izbranih gospodinjstvih. Načeloma je primerno uporabiti Kishev postopek, ki je sicer nekoliko zamuden, vendar pa je razumljiv, vzbuja občutek resnosti in daje preverjeno dobre rezultate. Metode zadnjega in naslednjega rojstnega dne, ki so izvedbeno najhitrejše, imajo namreč nemalokrat priokus neresnosti, anketiranci jih ne razumejo, predvsem pa so preveč svobodni pri nava-

janju izbrane osebe. Podobno kot pri kvotnih vzorcih obstaja tudi pri izbiri anketiranca znotraj gospodinjstva v marketinški industriji cel spekter postopkov, ki so sicer praktični, ne zagotavljajo pa verjetnostne izbire. Tako lahko izberemo kar prvo osebo, ki je dvignila telefonsko slušalko. Izberemo lahko tudi osebo, ki je navedena v imeniku, kar omogoča personaliziran uvodni nagovor, izravna pa tudi prevelik delež žensk, ki pogosteje dvigujejo telefon. Anketirance lahko izbiramo tudi samo med prisotnimi osebami, pogosto pa se uporabljajo tudi kvote. Pri uporabi kvot je neugodno, da v določenem gospodinjstvu včasih ne anketiramo nikogar – ko so prisotne le osebe, ki pripadajo skupinam, katerih kvote smo že zapolnili. V takih primerih je bil torej telefonski klic opravljen zaman, kar je vsekakor nepotreben strošek. Bolj primerno je zagotoviti želeno strukturo z ustreznim vnaprej prilagojenim fiksnim uvodnim nagovorom. Tako lahko npr. izbiramo moške z nekoliko večjo verjetnostjo. Morebitna preostala razhajanja pa uredimo z uteževanjem. Nadaljnja slabost kvot je tudi dejstvo, da onemogočajo izračunavanje pravilnih stopenj odgovorov.

- *Število klicev*. Če govorimo o verjetnostnih vzorcih, moramo zagotoviti tudi sprejemljive stopnje odgovorov. Pri tem je število klicev eden najpomembnejših mehanizmov za zagotavljanje visokih stopenj odgovorov. Analiza srednje kvadratne napake in števila stikov pogosto kaže (Kveder in Vehovar, 1999), da je optimalno število klicev okoli 5. V telefonski raziskavi projekta RIS med podjetji se je celo izkazalo, da je optimalno število klicev pravzaprav 3 (Vehovar et. al., 2000a). Z nadaljnjimi klici namreč hitro upada tudi njihov prispevek k zbranim informacijam. Po petih klicih se zato izkoristek bistveno zmanjša, nadaljnji intervjuji se podražijo in zato ti klici glede na stroške niso optimalni. Po drugi strani pa se tako imenovani zgodnji respondenti, ki jih lociramo že po nekaj klicih, lahko razlikujejo od poznih respondentov. V kakovostnih anketah zato raje ne tvegamo in zagotovimo vsaj 10 ali 20 poskusov kontaktiranja.
- *Število anketarjev*. Vsak anketar vpliva na anketiranca s svojo osebnostjo. Tako npr. nekateri anketarji izraziteje izvabljajo pozitivne odgovore oziroma izjave strinjanja (Vehovar in Belak, 1996). V tem pogledu ima anketar podoben učinek kot prostorska enota anketiranja. Zato ga lahko obravnavamo kot stopnjo vzorčenja, kar smo si ogledali že pri repliciranem vzorčenju. Elementi, ki pripadajo določenemu anketarju, so si namreč zaradi njegovega vpliva podobni, tako kot so si podobni elementi, ki pripadajo določeni prostorski enoti (npr. soseski). Dodatna stopnja vzorčenja pa seveda prinaša tudi dodatni vzorčni učinek. Analiza pri anketi s 14-timi anketarji je pokazala, da je ta učinek v splošnem okoli *Deff* = 1.5, pri nekaterih mnenjskih spremenljivkah pa je celo dvakrat večji (Belak in Vehovar, 1995). Majhno število anketarjev je lahko zato – posebej pri mnenjskih anketah – velik vir nenatančnosti.

- *Neodgovori*. Omenili smo že razmeroma ugodno razpoloženje glede sodelovanja v telefonskih anketah v Sloveniji. K temu lahko pripomorejo dober uvodni nagovor, zagotovilo, da ne gre za telefonsko prodajo, kakovosten sponzor oziroma naročnik, predvsem pa razdelan nadzor nad anketarji. Učinkovito sredstvo za doseganje visokega sodelovanja postajajo tudi nagrade oziroma darila (*angl. incentives*), ki so postale v razvitih državah izredno pogoste. Za zmanjševanje neodgovorov je lahko nadvse učinkovito tudi ponovno anketiranje oseb, ki so zavrnile sodelovanje, z bolj izkušenim anketarjem (*angl. refusion conversion*). Izkušnje kažejo, da je v tem pogledu uspešnih okoli 40 % poskusov.
- *Številke brez odziva*. Pri telefonskem anketiranju uporabljamo običajno terminologijo za označevanje kategorij neodgovorov. Večja težava je le pri kategoriji telefonskih številk, kjer se ni nihče javil (*angl. ring-no-answer*). Njihova uvrstitev v kategorijo odsotnih anketirancev (neodgovori) bi namreč bistveno večala stopnje neodgovorov, uvrstitev v kategorijo praznih številk (neustrezni elementi) pa bi zmanjšala stopnjo ustreznosti in dajala vtis o ugodnih stopnjah sodelovanja. Težavo lahko rešimo z arbitrarno razdelitvijo med obe kategoriji, npr. 50 : 50. Seveda pa je ugodno, da razmerje prilagodimo konkretnim značilnostim telefonskega anketiranja, predvsem številu ponovnih klicev.
- *Uteževanje rezultatov*. Uteževanje ima v telefonskih anketah dve nalogi. Po eni strani zagotovi videz (*angl. cosmetic weighting*), da je vzorec »reprezentativen«, po drugi strani pa dejansko popravi ocene, ki so zaradi porušene sociodemografske strukture v telefonskih vzorcih nekoliko pristranske. Pri popravljanju vzorca iz telefonskih raziskav je zato treba upoštevati kar največ zunanjih informacij o populaciji. Telefonski vzorci običajno izraziteje odstopajo glede deleža žensk, visoko izobraženih ter prebivalcev kmečkih okolij (Vehovar, 1990), čeprav se z izboljševanjem telefonskega pokritja odstopanja manjšajo. V telefonskih anketah se podcenjuje tudi delež manj premožnih slojev in delež vernih oseb. Vse navedeno je primerno – skupaj z drugimi kontrolnimi spremenljivkami – rutinsko popraviti v postopkih poststratifikacije ali rakinga. Izkušnje pa kažejo, da celo v primeru, ko imamo na voljo vse razpoložljive kontrolne spremenljivke – posebej kritične so lahko dohodek in veroizpoved – z uteževanjem ne moremo odstraniti več kot polovico pristranskosti (Vehovar, 1992a). Zato pa nam uteževanje daje nadvse koristno informacijo o smeri popravka.

Zgornja obravnava je bila namenjena predvsem tistim vidikom telefonskega anketiranja, ki so vsaj posredno povezani z vzorčenjem. Pri tem ne gre pozabiti na izredno pomembno vlogo, ki jo imajo izbira, šolanje, nadzor in motivacija anketarjev. Ustrezen nadzor (*angl. monitoring*) anketarjev je pri tem še posebej pomemben. V raziskavi organizacij, ki izvajajo telefonsko anketiranje, se je sicer izkazalo, da postopki izbire, nadzora in šolanja pri nas še niso posebej razviti (Kocjanc, 1998), čeprav postajajo vse bolj pomembni. V razvitejših okoljih se namreč s strukturiranim vprašalnikom nadzoruje tudi do 20 % anketarjevih pogovorov, rezultati pa se nato analizirajo in obravnavajo v skupinskih pogovorih. Posebej koristne so lahko tudi poglobljene analize obnašanja anketarjev, od načina uvodnega nagovora do ravnanja ob zavrnitvi, kar vse lahko vodi k izboljšavam celotnega procesa (Hox in deLeeuw, 2001).

Zavedati se je torej treba, da je sodobno telefonsko anketiranje – bolj kot metodološko vprašanje – predvsem kompleksen poslovni proces. Celovito upravljanje tega procesa je zato prvi pogoj za njegovo uspešnost. K temu lahko pomembno prispeva tudi učinkovita računalniška podpora. Izbira in analiza vzorca sta v tem procesu le eden od množice potrebnih pogojev za izvedbo kakovostne telefonske ankete.

15. NEVERJETNOSTNO VZORČENJE

Doslej smo se ukvarjali predvsem z verjetnostnim vzorčenjem, vendar bi bilo neustrezno, če bi se povsem izognili metodam neverjetnostnega vzorčenja. Neverjetnostno vzorčenje je namreč v praksi nadvse razširjeno, saj so taki vzorci mnogo cenejši in izvedbeno hitrejši, ocene pa se nemalokrat ne razlikujejo od ocen iz verjetnostnih vzorcev. Poleg tega pri mnogih raziskavah zahteva po kvantificiranju natančnosti – vključno z izračunavanjem intervalov zaupanja – ni tako zelo pomembna. Zato se bomo v nadaljevanju na kratko dotaknili neverjetnostnega vzorčenja in tudi razširjene tehnike kvotnih vzorcev.

Ponovimo najprej, da je glavna prednost verjetnostnega vzorčenja v mehanizmu verjetnostnega izbora, ki je predpogoj za uporabo statistične teorije. Tako lahko vnaprej izberemo cenilke z majhno pristranskostjo ali celo brez pristranskosti. Statistična teorija omogoča pri verjetnostnih vzorcih tudi izračunavanje natančnosti, oblikovanje intervalov zaupanja in preverjanje domnev.

Ker za neverjetnostne vzorce teorija statističnega sklepanja ne velja, pri teh vzorcih ne moremo izračunati intervalov zaupanja niti ne moremo preverjati domnev. Ocenjevanje kakovosti neverjetnostnih vzorcev je zato subjektivne narave, saj porazdelitev vzorčnih ocen ne temelji na statističnih zakonitostih. Zato tudi v primeru, kadar izkušnje kažejo, da je bila neverjetnostna metoda uspešna, to še ne zagotavlja, da bo takšna ostala še naprej. Kljub navedenim slabostim pa se neverjetnostni vzorci v praksi pogosto uporabljajo – kot rečeno – predvsem zaradi stroškov in enostavnosti izvedbe.

Priložnostno vzorčenje

Ena od najenostavnejših oblik neverjetnostnega vzorčenja je priložnostno vzorčenje (*angl. accidental sampling*). Navedimo nekaj primerov:

- pacienti nekega zdravnika,
- otroci v neki šoli,
- ankete, opravljene na ulici, tržnici ipd.,
- osebe, ki so odgovorile na anketo, priloženo v določeni reviji,
- osebe, ki se po telefonu odzivajo na televizijsko vprašanje (npr. »televoting«),
- osebe, ki izpolnijo kratek vprašalnik na svetovnem spletu.

Glede na precejšnjo verjetnost, da pride pri takšnih vzorcih do pristranskosti, je rezultate tovrstnih raziskav tvegano uporabljati za sklepanje o celotni populaciji.

Ekspertna izbira

Druga oblika neverjetnostnega vzorčenja je tako imenovani strokovni izbor oziroma ekspertna izbira (*angl. expert selection*). V tem primeru vzorec, ki naj bi bil »reprezentativen«, izbere strokovnjak, ki pozna vsebino raziskovanega področja. Tako lahko raziskovalci s področja izobraževanja opravijo izbor manjšega števila šol, ki predstavljajo tipičen presek šol v nekem mestu. Podobno lahko demografi izberejo tipično naselje, ki bo nastopalo v mednarodni študiji, v katero je iz vsake države vključeno le po eno naselje ali dve. Posebej pogosto se uporablja ekspertna izbira na področju statistike cen, kjer poznavalci na prvi stopnji izberejo v vzorec večje oziroma tipične kraje, na drugi stopnji pa še trgovine, v katerih se mesečno spremlja gibanje drobnoprodajnih cen. Kot primer ekspertne izbire na področju volilnih študij velja omeniti tudi izbiranje kraja, ki ga v nekaterih državah (npr. ZDA, Avstrija) raziskovalci določijo na osnovi »reprezentativnega« obnašanja pri prejšnjih volitvah.

Seveda pa bi se v praksi več strokovnjakov le redko strinjalo glede tega, kaj spada v »reprezentativni« vzorec. Poleg tega obstaja pri takem izboru tveganje o pristranskosti neznanega obsega, saj o obsegu odstopanja vzorčnih aritmetičnih sredin od populacijskih vrednostih nimamo nobenih objektivnih meril.

Kljub temu pa je v primeru, ko potrebujemo npr. le en element ali dva, tovrstni izbor primernejši kot verjetnostna izbira. Če primerjamo verjetnostni in neverjetnostni vzorec enake velikosti, se namreč nevarnost netočnih cenilk pri neverjetnostnih vzorcih z velikostjo vzorca ne zmanjšuje. Če je vzorec zelo majhen, bo varianca cenilke v verjetnostnem vzorcu izredno velika, tako da je, primerjalno gledano, pristranskost cenilke vzorca na podlagi ekspertne izbire v primerjavi s tem lahko zanemarljiva. Ko pa se velikost vzorca povečuje, se varianca cenilke verjetnostnega vzorca zmanjšuje, medtem ko se netočnost cenilke v drugem primeru največkrat ne spreminja. Zaradi opisanih razlogov je ekspertna izbira lahko upravičena, kadar je vzorec majhen, za večje vzorce pa je priporočljivo verjetnostno vzorčenje. Če npr. raziskovalec lahko opravi raziskavo le v enem ali dveh mestih, je verjetno bolje izbrati mesta na podlagi ekspertne izbire kot pa tvegati veliko variabilnost verjetnostnega izbora, ki lahko privede do čudnega vzorca. Če pa vzorec povečamo na nekaj deset elementov, npr. 50 mest, je stratificiran verjetnostni vzorec skoraj zagotovo ustreznejša možnost.

Kvotno vzorčenje

Tretja oblika neverjetnostnega vzorčenja, ki jo v praksi srečujemo izredno pogosto, je kvotno vzorčenje (*angl. quota sampling*). Kvotno vzorčenje ima številne različice, najpogosteje pa se uporablja pri marketinških raziskavah, saj je cenejše, lažje izvedljivo in ga je mogoče izpeljati bistveno hitreje kot primerljive verjetnostne vzorce. Bistvo kvotnega vzorčenja so kvote – gre za fiksno število oseb z določenimi značilnostmi, ki jih morajo anketarji vključiti v vzorec. Tako lahko npr. anketar dobi naslednje kvote:

- 6 moških, starih 35 let ali manj,
- 5 moških nad 35 let,
- 5 zaposlenih žensk,
- 8 delovno neaktivnih žensk.

Cilj uporabe kvot je izogibanje oziroma vsaj nadzorovanje pristranskosti pri anketarjevem izboru oseb v vzorec. Anketar ima namreč proste roke pri izbiri oseb samo znotraj predpisanih kvot in torej le ni povsem svoboden tako kot pri priložnostnih vzorcih.

Kvote so lahko medsebojno povezane, kot je to v zgornjem primeru, ali pa so neodvisne. Neodvisne kvote bi v zgornjem primeru pomenile ločene kvote za spol (npr. 10 moških, 13 žensk) in za starost (11 oseb mlajših od 35 let, 12 oseb, starih 35 let in več).

Pomembna izboljšava kvotnega vzorčenja je njegovo kombiniranje z verjetnostnim vzorčenjem (*angl. probability sampling with quotas – PSQ*), ki se običajno izvede pri izboru enot prve stopnje. Tako se prvih nekaj stopenj vzorčenja pri nacionalnih kvotnih vzorcih pogosto opravi po enakih verjetnostnih metodah kot pri verjetnostnih vzorcih, kjer se na podlagi popisnih ali registrskih podatkov verjetnostno izbirajo naselja, krajevne skupnosti, popisni okoliši ipd. Do razlike pride šele v zadnji fazi, pri izbiri oseb. Pri verjetnostnem vzorcu namreč anketiramo elemente, ki so bili izbrani na podlagi verjetnostnega mehanizma. Pri kvotnih vzorcih pa anketarji zgolj izpolnjujejo svoje kvote, pri čemer se včasih dodaja še druge omejitve, posebej glede anketarjevega gibanja na terenu (npr. metoda slučajne poti) in dnevnega časa, ko se izbira elemente v vzorec (npr. samo popoldne).

Tipično navodilo pri kakovostnih kvotnih vzorcih zato opredeljuje postopek iskanja ustreznih anketirancev v verjetnostno izbranih blokih (ali ulicah) na podlagi enolično določene začetne točke in nadaljnjega točno predpisanega gibanja slučajne poti. Pri tem anketarji – upoštevajoč kvotne omejitve – izbirajo v vzorec po eno osebo v vsakem izbranem stanovanju.

Kadar v vzorec najprej izbiramo stanovanja – mnogo kvotnih vzorcev sicer poteka kar na ulici ali v trgovini – je nadvse pomemben tudi izbor anketiranca v stanovanju. Omejitev na izbor ene same osebe v vsakem stanovanju je seveda ugodna, saj razprši vzorec in prepreči vrsto težav pri opravljanju več anket v enem stanovanju. Po drugi strani pa lahko taka omejitev privede do premajhnega števila oseb iz stanovanj z bolj številnimi stanovalci (Stephenson, 1979). Seveda se s tako težavo srečujemo tudi pri verjetnostnem vzorčenju, kadar izbiramo samo eno osebo iz vsakega stanovanja, vendar lahko pri verjetnostnih vzorcih porušeno razmerje popravimo z uteževanjem, ki je obratno sorazmerno z verjetnostjo izbora. Osebe iz stanovanj, kjer je manjše število oseb, tako dobijo manjšo utež. Uteževanje v kvotnih vzorcih, kjer verjetnosti izbora ne poznamo, pa ima razmeroma nejasne posledice.

Kvotne skupine so na videz podobne stratumom, saj imamo v obeh primerih populacijske skupine elementov, iz katerih izbiramo elemente v vzorec. Seveda pa ne smemo spregledati pomembne razlike. Pri stratumih so elementi izbrani z verjetnostnim mehanizmom, pri kvotnih skupinah pa ne, kar vodi v različna merila za oblikovanje stratumov in kvotnih skupin. Ker se z verjetnostnim vzorčenjem povsem izognemo pristranskosti izbora, je pri določanju stratumov treba poskrbeti predvsem za večanje natančnosti cenilk. Kot smo videli, dosegamo to z oblikovanjem stratumov, ki so notranje homogeni glede na ciljne spremenljivke. Stratume torej izbiramo tako, da je variabilnost znotraj stratumov kar najmanjša, razlike med stratumi pa so čim večje.

Pri kvotnem vzorčenju pa je pozornost usmerjena v skupine, ki zagotavljajo, da je pristranskost pri izboru – kar pri stratifikaciji sploh ni problem – čim manjša. Pri tem je sicer koristno oblikovanje skupin, ki so notranje homogene glede ciljne spremenljivke, vendar je pomembno oblikovati predvsem skupine, ki so notranje homogene glede neodgovorov in dosegljivosti za anketo. Drugače povedano, kvotne skupine se morajo med seboj razlikovati predvsem glede dosegljivosti in pripravljenosti na sodelovanje v anketi. Zadnja zahteva je Nacionalni center za mnenjske raziskave na Univerzi v Chicagu (NORC) privedla do uporabe uvodoma omenjenih štirih kvotnih skupin za raziskave, ki temeljijo na verjetnostnem vzorčenju s kvotami (PSQ). Že v šestdesetih in sedemdesetih letih pa se pri PSQ vzorčenju tipično uporabljajo naslednje kvote (Sudman, 1966; Stephenson, 1979): mlajši moški, starejši moški, zaposlene ženske in delovno neaktivne ženske.

Navedene kvote so izdelane predvsem zato, da zagotovijo ustrezno zastopanost mladih moških – zaradi slabšega sodelovanja oziroma dosegljivosti je skupina mlajših moških v anketah najbolj kritična – in delovno neaktivnih žensk, ki so v anketah najlažje dosegljiva in so v vzorcih nadreprezentirane. Ker je kvotna izbira pri PSQ vzorčenju vključena šele znotraj verjetnostno izbranih stanovanj, lahko strog nadzor v prvih stopnjah vzorčenja zagotovi kakovost vzorca tudi glede drugih sociodemografskih struktur (regija, status, rasa), saj je izbira stavb oziroma stanovanj opravljena z verjetnostnim vzorčenjem.

Nadaljnja poenostavitev pri zagotavljanju ustreznega deleža zaposlenih žensk in mlajših moških pa je lahko kvotno pravilo, da se v izbranem stanovanju vedno izbere v vzorec najmlajši prisotni moški. Kadar med prisotnimi ni moških, pa se izbere najstarejšo zaposleno žensko. Šele po tem pride na vrsto morebitno popolnjevanje ostalih kvot.

Ko se v anketni raziskavi odločimo za kvote, je treba zanje najprej izračunati populacijske deleže in jih nato porazdeliti med anketarje. Populacijska razmerja največkrat izračunamo na podlagi popisnih podatkov, tako izračunane kvote pa so za vse anketarje običajno bolj ali manj enake. Variirajo lahko kvečjemu glede na razlike v sociodemografski strukturi ciljne populacije na posameznih območjih, na katerih anketarji opravljajo svoje delo.

Opozoriti je treba tudi na nevarnost, da so podatki, na podlagi katerih so določene kvote, netočni. Tipičen primer je npr. zastarela popisna ocena o izobrazbeni strukturi. V takem primeru porazdelitev kvotnega vzorca seveda ne ustreza dejanski porazdelitvi populacije. V nasprotju s tem pri verjetnostnih vzorcih nimamo teh težav, saj se tovrstne napake pri verjetnostni izbiri avtomatično odpravijo.

Omeniti velja še eno težavo pri kvotnem vzorčenju. Včasih nastaja vtis, da se s kvotnim vzorčenjem izognemo problemu neodgovorov. V resnici pri kvotnem vzorcu pride zgolj do zamenjave odsotnega ali na sodelovanje nepripravljenega anketiranca z anketirancem, ki je pripravljen sodelovati. Zaradi tega imamo v kvotnih vzorcih seveda še vedno premalo oseb, s katerimi je težko priti v stik oziroma ne želijo sodelovati, čeprav kvotni vzorec zagotovi zahtevano velikost in sociodemografsko strukturo. Pri kvotnem vzorčenju prihaja do premajhne zastopanosti takšnih oseb celo bolj izrazito kot pri verjetnostnem vzorcu, kjer morajo anketarji vložiti bistveno več napora, da opravijo anketo s točno določenimi osebami.

Oglejmo si še dva dodatna razloga, zakaj je – kljub teoretskim pomanjkljivostim – uporaba kvotnega vzorčenja tako razširjena. Prvi razlog je v poenostavitvah, ki izhajajo iz tega, da za izbor anketirancev ne potrebujemo vzorčnega okvira ciljnih oseb. Drugi razlog pa je v zmanjšanem številu potrebnih stikov. Anketarjem se namreč v primeru odsotnega anketiranca ni treba vračati in ponovno poskusiti z anketiranjem. Pri kvotnih vzorcih se opravi samo en stik, torej en sam poskus anketiranja, nakar anketar preprosto poskusi z naslednjo osebo ali z naslednjim stanovanjem. Taki postopki omogočajo bistveno enostavnejšo in veliko hitrejšo izvedbo raziskave kot pri verjetnostnem vzorčenju. Ugodna posledica navedenih poenostavitev je tudi nižja cena kvotnega vzorčenja. Cena kvotnega vzorčenja je zato v veliki meri odvisna prav od obsega poenostavitev pri nadzoru nad postopki izbire v vzorec: manjši kot je nadzor, nižji bodo stroški. Hkrati s tem pa seveda narašča tudi tveganje za resnejšo pristranskost dobljenih ocen.

Kvotno vzorčenje v praksi

Oglejmo si še pronicljiv pogled na vlogo kvotnega vzorčenja v anketnem raziskovanju, kot ga je zapisal Leslie Kish pol stoletja po uporabi verjetnostnega vzorčenja in skoraj 35 let po izidu svojega klasičnega učbenika Survey Sampling (1965). Slednji je še danes – v povsem nespremenjeni obliki – eno najpogosteje uporabljanih besedil pri praktičnem vzorčenju. Omenjeni učbenik med drugim obravnava tudi kvotne vzorce. Ko so največje marketinške agencije v ZDA profesorja Kisha prosile, da ponovno osvetli uporabnost tovrstnih vzorcev, je Leslie Kish v odgovor navedel kratek zapis (1998), ki navajamo v celoti:

> »Podpoglavje 13.7 v Survey Sampling (Kish, 1965) povzema bistveno o tovrstnem vzorčenju. Danes ni temu veliko dodati, kajti kvotno vzorčenje se ni spremenilo. Novo je le to, da ga izvaja vse več organizacij, vse pogosteje in tudi po telefonu. Nasprotno pa je verjetnostno vzorčenje po letu 1965 močno napredovalo.
>
> O kvotnem vzorčenju je nasploh težko razpravljati, ker to sploh ni znanstvena metoda z jasnimi definicijami. To je bolj nekakšna veščina (*angl. art*), ki jo v najrazličnejših okoliščinah uporabljajo ljudje z različnim spektrom strokovnega znanja in tudi z različno uspešnostjo. Značilno je tudi, da ni nobenega učbenika o kvotnem vzorčenju, na katerega bi se lahko sklicevali. Že samo to dejstvo je dovolj veliko svarilo, ki govori proti uporabi kvotnega vzorčenja.
>
> Kadar je kvotno vzorčenje poceni in hitro, potem je navadno tudi slabo. Kadar pa je bolje opravljeno, ni veliko cenejše od verjetnostnega vzorčenja. Boljše kvotne metode pa se uporabljajo vselej, ko je rezultate mogoče preveriti z zunanjimi viri, kot je to npr. pri volilnih napovedih.
>
> Kvotno vzorčenje je v veliki meri povezano z domiselnostjo raziskovalcev in je zato tudi cenovno učinkovito, vendar se moramo zavedati velike nevarnosti nepredvidene pristranskosti. Prav v kvotnih vzorcih pa pristranskosti pogosto sploh ni mogoče preveriti, saj največkrat proučujemo pojave, o katerih nimamo kontrolnih populacijskih podatkov, kot je npr. odnos do določenega izdelka.
>
> Kakovostni kvotni vzorci izhajajo iz bolj podrobnih kvotnih zahtev (npr. preverjanje zaposlitvenega statusa, gmotnega položaja ipd.) in ne le običajnih starostno-spolnih kvot. Seveda pa so tovrstne kvote tudi bistveno dražje, saj je npr. v nekem mestu težko najti diplomanta visoke šole, ki pripada določeni etnični skupini in ima več kot 65 let. Določene kvotne lastnosti je tudi težko opaziti ali po njih spraševati potencialne anketirance (npr. dohodek, nacionalnost). Največ izboljšav, pa tudi dodatnih stroškov pri kakovostnih kvotnih vzorcih, pravzaprav izhaja iz razporejanja anketarjev na

lokacije, ki so bile izbrane po načelih verjetnostnega vzorčenja in v tem se torej kvotni vzorci od verjetnostnih ne razlikujejo.

Po drugi strani so običajne (starost/spol) kvote teoretično in tudi praktično skoraj brez vrednosti. Spol in starost namreč pogosto pojasnjujeta samo nekaj odstotkov variabilnosti pri določeni spremenljivki. Po statistični teoriji bi lahko dobili podobne rezultate tudi, če bi anketarjem prepustili, da povsem poljubno izberejo katerih koli 1,000 (ali n) oseb in bi nato rezultate utežili v skladu s starostno-spolnimi kvotami.

In zakaj so kvotni vzorci cenejši od verjetnostnih vzorcev?

- Najprej, splošno znana in razširjena je domneva, da kvotno vzorčenje stane x-krat manj kot verjetnostno vzorčenje, npr. polovico manj. Dejansko pa gre le za faktor, ki se nanaša na nekatere stroške anketiranja (gre torej za faktor f in ne za faktor x), medtem ko mnogi drugi vidiki (intervju, kodiranje, poročanje) ostajajo enaki. Razlika največkrat izvira le iz stroškov, ki nastajajo pri iskanju anketiranca, kar v večini primerov ne presega 30 minut. Seveda pa je tak dodatni strošek nadvse pomemben za nizkoproračunske ankete, ki imajo kratke intervjuje. V splošnem pa je tovrstna primerjava pogosto precej nejasna zaradi vrste drugih dejavnikov, ki otežujejo primerjanje in razbijanje stroškov anketnih raziskav.
- Administrativni stroški kvotnega vzorčenja so manjši, saj ni treba plačati postopkov in osebja za verjetnostno vzorčenje. Seveda pa to tudi pomeni, da nimamo nobene dokumentacije o vzorcu niti ustreznega nadzora nad potekom izbora elementov v vzorec.
- Kvotno vzorčenje je praviloma opravljeno brez ponovnih stikov z izbrano osebo, kar seveda zmanjšuje stroške, čeprav najbrž ne bistveno. Anketiranje se torej največkrat opravi pri anketirancu, ki se prvi oglasi (pri vratih, na telefonski klic). To ima za posledico prostovoljne anketirance, ki v mnogih anketah vodijo k pristranskim ocenam. Na tem mestu bi lahko navedli vrsto neuspešnih kvotnih vzorcev volilnih anket in tudi pristranske rezultate v mnogih anketah potrošnikovega obnašanja. Celo v verjetnostnih vzorcih imajo namreč lažje dosegljivi in na sodelovanje pripravljeni anketiranci pogosto drugačne lastnosti kot ostali elementi v vzorcu.

Cenovne prednosti kvotnega vzorčenja se vsekakor maščujejo v dodatni pristranskosti in dodatni nenatančnosti vzorčnih ocen. Izkušnje sicer kažejo, da so ocene iz kvotnih vzorcev za večino spremenljivk lahko dovolj dobre, zato pa so pri določenih spremenljivkah nepredvidljivo napačne.

Postavlja pa se tudi vprašanje o vzorčni varianci. Nadvse značilno namreč je, da pri rezultatih, ki temeljijo na kvotnih vzorcih, nikoli ne najdemo ocene vzorčne variance niti intervalov zaupanja. To je tudi razumljivo, saj variance načeloma ni mogoče izračunati, če vnaprej ne poznamo verjetnosti za vključitev elementov v vzorec. Vsekakor pa ne moremo sprejeti npr. poenostavitve za izračun vzorčne variance deležev:

$$Var(p) = Deff \times \frac{PQ}{n},$$

kjer je *Deff* nekakšen splošni vzorčni učinek. Prvič, enotni *Deff* ne obstaja, saj ima vsaka spremenljivka drugačne lastnosti. Drugič, pravilna ocena za *Deff* bi razkrila bistveno večje faktorje.

Tako je npr. pravilna cenitev, ki sta jo opravila Moser in Strent (Kish, 1965) pokazala bistveno večje vzorčne učinke v kvotnih kot pa v verjetnostnih vzorcih. Posamezna razhajanja so bila celo tako velika, da so povzročila resne dvome o smiselnosti uporabe kvotnih vzorcev. Seveda pa bi morali kot merilo učinkovitosti upoštevati stroške glede na doseženo natančnost (varianco) oziroma točnost (MSE), ne pa samo stroške na anketarja.

Takoj velja dodati, da je ocenjevanje variance v kvotnih vzorcih, ki bi omogočilo primerjavo z verjetnostnimi vzorci, mogoče le pri PSQ vzorcih in še tam zgolj približno, saj bi morali paroma primerjati skupine anketarjev, ki so bili naključno izbrani znotraj iste enote prve stopnje. Šele rezultati na podlagi takih poskusov omogočajo ocenjevanje razlike med vzorčnima variancama. Primerov takih analiz, z izjemo prej omenjene študije Moser in Strent (Kish, 1965), skoraj ne srečamo.

Tudi če zanemarimo dejstvo, da vzorčne variance za neverjetnostne vzorce načeloma ni mogoče izračunati in obstajajo le grobi približki, pa se pojavlja še naslednje vprašanje: Ali je vzorčno varianco sploh mogoče oceniti, če pa so ocene pristranske? To bi namreč pomenilo, da ne upoštevamo pomembne komponente srednje kvadratne napake (MSE), ki je sestavljena iz variance in kvadrata pristranskosti. Pri ocenjevanju deležev, kjer imamo pristranskost, moramo namesto vzorčne variance namreč upoštevati:

$$MSE(p) = Var(p) + Bias^2(p).$$

Če že poskušamo oceniti vzorčno varianco v kvotnih vzorcih, moramo torej vključiti tudi dodatno pristranskost, ki nastaja zaradi uporabe kvotnih vzorcev.«

Na koncu velja dodati, da kvotnim vzorcem – kljub vsem navedenim ugovorom – ni mogoče odreči določenega potenciala. Posebej v primeru majhnih sredstev so informacije, ki jih lahko pridobimo s tovrstnimi vzorci, lahko nadvse koristne. Osnovna dilema pri uporabi kvotnega vzorčenja je zato predvsem presoja, kdaj in v kakšnem obsegu je primerno uporabiti načela verjetnostnega vzorčenja. Primerjava stroškov in točnosti (MSE) za kvotne in verjetnostne vzorce namreč lahko pokaže, da je prihranek zaradi manjših stroškov kvotnega vzorca povzročil neprimerno večjo napako pri kakovosti vzorčne ocene.

Pri enakih finančnih sredstvih lahko zato verjetnostni vzorec daje – kljub manjši velikosti – bistveno bolj točno oceno kot pa kvotni vzorec. Slednji je pri enakih sredstvih zaradi nižjih terenskih stroškov sicer lahko večji, vendar to ne pomeni avtomatično tudi večje točnosti (MSE) oziroma kakovosti ocen. Pri presoji o morebitni uporabi kvotnih vzorcev moramo zato pazljivo pretehtati vse bistvene vidike kvotnih in verjetnostnih vzorcev: čas, stroške in obseg odgovarjajočih napak.

16. SKLEPNE MISLI

Vzorčenje v anketah je visoko specializirano področje anketnega raziskovanja. Statistiki so pri tem razvili širok izbor vzorčnih tehnik, vendar obstaja pri njihovi praktični izvedbi vrsta nevarnosti. Brez dobrega poznavanja temeljnih načel vzorčenja je zato pri načrtovanju in analizi verjetnostnih vzorcev potrebna kar največja previdnost. Napake pri vzorčenju namreč zlahka zmanjšajo – ali celo povsem izničijo – uporabnost rezultatov anketne raziskave. Za raziskovalca, ki vzorčenja ne pozna, je zato primerno, da se pri zahtevnejših načrtih za nasvet obrne na statistike, ki se s tem ukvarjajo.

Literatura o vzorčenju v anketah je nadvse obsežna tako na teoretičnem kot na praktičnem področju. V pričujočem besedilu je seveda obravnavan le uvodni pregled osnovnih pojmov, najpomembnejših tehnik in elementarnih napotkov za praktično uporabo.

Bralci, ki želijo podrobnejšo obravnavo, jo lahko najdejo v obsežni literaturi. Za praktična vprašanja vzorčenja je posebej priporočljiv učbenik Survey Sampling (Kish, 1965). Mnoge opredelitve v navedenem učbeniku so bile izhodišče tudi za obravnavo v pričujočem delu. V okviru praktičnih del velja omeniti še Hansen et al. (1953), Yates (1981), pa tudi Deming (1960), kjer so poleg repliciranega vzorčenja predstavljeni še številni praktični nasveti.

Med teoretsko zahtevnejšimi deli velja izpostaviti predvsem klasični učbenik Cochran (1977), pa tudi Sukhatme (1970) in Murthy (1967).

Za uvodno obravnavo vzorčenja so primerni tudi Ray (1972), Levy in Lemeshow (1980), Sudman (1976) ter Stuart (1976), ki je vzorčenje predstavil povsem nematematično, pa tudi poglavja o vzorčenju v Moser in Kalton (1971).

V zadnjih desetletjih je izšlo še več kakovostnih del s področja vzorčenja, od specializirane obravnave ocenjevanja vzorčne variance (Wolter, 1985), modelskega pristopa k verjetnostnemu vzorčenju (Särdnal et al., 1992), do modernih učbenikov kot sta Scheaffer et al. (1996) in Lohr (1999).

LITERATURA

ARNEŽ, M., NOVAK, T., VEHOVAR, V. in ZALETEL, M. (1996). Geoinformacijski sistem za načrtovanje vzorcev. V I. Tršinar in I. Ograjenšek (ur.). *Ekonomija, Slovenija, Evropska unija*, 268–275. Statistični urad Republike Slovenije, Statistično društvo Slovenije, Ljubljana.

BELAK, E. in VEHOVAR, V. (1995). Interviewers' effects in telephone surveys: the case of international victim surveys. V A. Ferligoj in A. Kramberger (ur.). *Contributions to Methodology and Statistics (Metodološki zvezki 10)*, 85–98. Fakulteta za družbene vede, Ljubljana.

BLALOCK, H. M. (1972). *Social Statistics.* McGraw-Hill, New York.

BLEJEC, M. (1970). *Slovensko javno mnenje 1969*. Višja šola za sociologijo, politične vede in novinarstvo, Ljubljana.

BOX, G. E. P. in TIAO, G. C. (1992). *Bayesian Inference in Statistical Analysis*. Wiley, New York.

BRICK, M. in WAKSBERG, J. (1992). Avoiding Sequential Sampling with Random Digit Dialing. *Survey Methodology*, 17, 31–47.

BREHM, J. (1993). *The Phantom Respondents*. The University of Michigan Press, Ann Arbor.

BRYANT, B. E. (1975). Respondent selection in a time of changing household composition. *Journal of Marketing Research*, 12, 129–135.

COCHRAN, W. G. (1977): *Sampling Techniques*. Wiley, New York.

CONNOR, J. in HEERINGA, S. (1992). Evaluation of Two Cost Efficient RDD Designs. *SMP Working Paper Series*. The University of Michigan, ISR, SRC, Ann Arbor.

CZAJA, R., BLAIR, J. in SEBESTIK, J. P. (1982). Respondent selection in a telephone survey: a comparison of three techniques. *Journal of Marketing Research*, 19, 381–385.

DEMING, W. E. (1960). *Sample Design in Business Research.* John Wiley, New York.

DILLMAN, D. A. (1978). *Mail and Telephone Surveys. The Total Design Method.* John Wiley & Sons, New York.

DILLMAN, D. A. (2000). *Mail and Internet Surveys. The Tailored Design Method*. John Wiley & Sons, New York.

FRANKEL, M. R. (1971). *Inference from Survey Samples.* Institute for Social Research, Ann Arbor.

FRANKEL, M. R. in FRANKEL, L. R. (1977). Some recent developments in sample survey design. *Journal of Marketing Research,* 14, 280–293.

FREY, J. in OISHI, S.M. (1995). *How to Conduct Interviews by Telephone and in Person.* Sage, Thousand Oaks.

GABLER, S. in HÄDER, S. (2000). Telephone Sampling in Germany. *Paper presented at International Conference on Social Science Methodology, Cologne.*

GOODMAN, R. in KISH, L. (1950). Controlled selection – a technique in probability sampling. *Journal of the American Statistical Association,* 45, 350–372.

GOYDER, J. (1988). *The Silent Minority.* Polity Press, New York.

GROVES, R. (1989). *Survey Errors and Survey Costs.* Wiley, New York.

GROVES, R. in KAHN, R. L. (1979). *Surveys by telephone: A National Comparison with Personal Interviews*. Academic Press, New York.

GROVES, R. in LEPKOWSKI, J. M. (1982). Alternative dual frame mixed mode survey designs. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 154–159.

GROVES, R., LYBERG, L. in MASSEY, J., ur. (1988). *Telephone survey methodology.* Wiley, New York.

GROVES, R. DILLMAN, D, ELTINGE, J. in R. LITTLE (ur.) (2001). *Nonresponse in Survey.* Wiley, New York.

HANSEN, M. H., HURWITZ, W. N. in MADOW, W. G. (1953). *Sample Survey Methods and Theory, Vols. 1 and 2.* John Wiley, New York.

HESS, I., RIEDEL, D. C. in FITZPATRICK, T. B. (1975). *Probability Sampling of Hospitals and Patients.* Health Administration Press, Ann Arbor.

HOFFMEYER-ZLOTNIK, J. in KREBS, D. (1996). Different methods of survey sampling in Germany. V A. Ferligoj in A. Kramberger (ur.). *Developments in data analysis: proceedings of the International Conference on Statistical Data Analysis and Data Collection, Bled, Slovenia, September 19 – 21, 1994 (Metodološki zvezki 12)*, 75–97. Fakulteta za družbene vede, Ljubljana.

HOX, J. in LEEUW, E., de (2001). The Influence of Interviewers' Attitude And Behavior On Household Survey Nonresponse – An International Comparison. V R. Groves, D. Dillman, J. Eltinge in R. Little (ur.). *Nonresponse in Surveys.* Wiley, New York.

IVERSEN, G. R. in NORPOTH, H. (1976). *Analysis of Variance. Sage University Paper series on Quantitative Applications in the Social Sciences 07–001.* Sage, Beverly Hills.

JAMNIK, R. (1980). *Matematična statistika*. DZS, Ljubljana.

KALTON, G. (1977). Practical methods for estimating survey sampling errors. *Bulletin of the International Statistical Institute,* 47 (3), 495–514.

KALTON, G. (1990). *Introduction to Survey Sampling (Series: Quantitative Applications in the Social Sciences, Sage University Paper 35), Ninth printing.* Sage Publications, Newbury Park, London, New Delhi.

KALTON, G. in KASPRZYK, D. (1982). Imputing for missing survey responses. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 22–31.

KAPLAN, B. in FRANCIS, I. (1979). A comparison of methods and programs for computing variances of estimators from complex sample surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 97–100.

KENDALL, M. G. in SMITH, B. B. (1939). *Tables of Random Sampling Numbers.* Cambridge University Press, Cambridge.

KISH, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association,* 57, 92–115.

KISH, L. (1965). *Survey Sampling.* John Wiley, New York.

KISH, L. (1970). Balanced repeated replications for standard errors. *Journal of the American Statistical Association*, 65, 1071–1094.

KISH, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics*, 8 (2), 183–201.

KISH, L. (1998). *Quota Sampling: Old plus New Thought*. Delovni material, http://www.ris.org/vasja/kish.doc.

KISH, L. in FRANKEL, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society B,* 36, 1–37.

KOŽUH-NOVAK, M., OBERSNEL-KVEDER, D., AERNIČ-ISTENIČ, M., ŠIRCELJ, V. in VEHOVAR, V. (1998). *Rodnostno vedenje Slovencev: nacionalno poročilo*. Znanstvenoraziskovalni center SAZU, Založba ZRC, Ljubljana.

KOCJANC, D. (1998): *Anketiranje po telefonu in anketar: diplomsko delo*. Fakulteta za družbene vede, Ljubljana.

KVEDER, A. in VEHOVAR, V (1999). An Elaborate Calling Strategy – Does It Make Enough Difference? Paper presented at the *International Conference on Nonresponse in Surveys, Portland.*

LAVRAKAS, P.J. (1993). *Telephone Survey Methods: Sampling, Selection, and Supervision (Applied social research methods series, Vol. 7).* Sage, Newbury Park, London, New Delhi.

LEVY, P. S. in LEMESHOW, S. (1980). *Sampling for Health Professionals.* Lifetime Learning Publications, Belmont.

LITTLE, R. in RUBIN, D. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

LOHR, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press, Brooks/ Cole.

LOZAR, K. in VEHOVAR, V. (2000). Number of contacts and mean squared error in surveys. V A. Ferligoj in A. Mrvar (ur.). *Developments in Survey Methodology (Metodološki zvezki 15),* str. 30–31. Fakulteta za družbene vede, Ljubljana.

McCARTHY, P. J. (1966). Replication – An Approach to the Analysis of Data from Coplex Surveys. *Vital and Health Statistics Series 2, No. 14.* Government Printing Office, Washington, DC.

MOSER, C. A. in KALTON, G. (1971). *Survey Methods in Social Investigation.* Heinemann, London.

MOŽINA, E., EMERŠIČ, B., KNAFLIČ, L., RADOVAN, M. in VEHOVAR, V. (1999). *Nacionalna raziskava Pismenost odraslih in udeležba v izo-*

braževanju: *tehnično poročilo (Mednarodna raziskava pismenosti odraslih).* Andragoški center Republike Slovenije, Ljubljana.
MURTHY, M. N. (1967). *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta.
NICOLAAS, G., LYNN, P. in LOUND, C. (2000). Random Digit Dialing in the UK: Viability of the Sampling Method Revisited. *Paper presented at International Conference on Social Science Methodology, Cologne.*
O'MUIRCHEARTAIGH, C. in WONG, S. T. (1981). The impact of sampling theory on survey practice: a review. *Bulletin of the International Statistical Institute*, 49(1), 465–493.
RAJ, D. (1972). *The Design of Sample Surveys*. McGraw-Hill, New York.
RIS, Raba Interneta v Sloveniji (2001). Http://www.ris.org.
RUBIN, D. (1988). *Multiple Imputations for Nonresponse in Surveys.* Wiley, New York
RUBIN, D., STERN, H. in VEHOVAR, V. (1995). Handling "Don't know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association*, 90 (431), 822–828.
RUBIN, D. in ZANUTTI, E. (2001). Multiple Imputations for Substitute Units in Sample Surveys. V R. Groves, D. Dillman, J. Eltinge in R. Little (ur.). *Nonresponse in Surveys.* Wiley, New York.
RUTAR, K. (2000). *Vplivi raziskovalnega okolja na procese anketiranja – iskanje vzrokov za nesodelovanje v anketah: diplomsko delo.* Fakulteta za družbene vede, Ljubljana.
SÄRDNAL, C., SWENSSON, B. in WRETMAN, J. (1992). *Model assisted Survey Sampling*. Springer-Verlag, New York.
SCHEAFFER, R.L., MENDENHALL, W. in OTT, R.L. (1996). *Elementary Survey Sampling, 5th ed*. Duxbury, Massachusetts.
STEEH, C. G. (1981). Trends in nonresponse rates, 1952–1979. *Public Opinion Quarterly*, 45, 40–57.
STEPHENSON, C. B. (1979). Probability sampling with quotas: an experiment. *Public Opinion Quarterly*, 43, 477–495.
STUART, A. (1976). *Basic Ideas of Scientific Sampling*. Griffin, London.
SUDMAN, S. (1966). Probability sampling with quotas. *Journal of the American Statistical Association*, 61, 749–771.
SUDMAN, S. (1976). *Applied Sampling.* Academic Press, New York.
SUKHATME, P. V. in SUKHATME, B. V. (1970). *Sampling Theory of Surveys with Applications.* Asia Publishing House, London.
ŠTEBE, J. (1995). Nonresponse in the Slovene Public Opinion Survey. V A. Ferligoj in A. Kramberger (ur.). *Contributions to Methodology and Statistics (Metodološki zvezki 10)*, 21–37. Fakulteta za družbene vede, Ljubljana.
THORNBERRY, O. T. in MASSEY, J. T. (1978). Correcting for under-coverage bias in random digit dialed national health surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 224–229.

TROLDAHL, V. C. in CARTER, R. E. (1964). Random selection of respondents within households in phone surveys. *Journal of Marketing Research*, 1, 71–76.

U.S. BUREAU OF THE CENSUS (1978). *The Current Population Survey: Design and Methodology. Technical Paper 40.* Government Printing Office, Washington, DC.

VEHOVAR, V. (1990). Pristranskost telefonskih vzorcev v Sloveniji. V G. Ajduković, J. Jug in A. Kramberger (ur.). *Blejsko metodološko srečanje '90: zbornik referatov VIII. strokovnega sestanka Sekcije za metodologijo in statistiko Jugoslovanskega združenja za sociologijo (Metodološki zvezki 7)*, 147–160. Fakulteta za sociologijo, politične vede in novinarstvo, Raziskovalni inštitut, Ljubljana.

VEHOVAR, V., LOZAR-MANFREDA, K, BATAGELJ, Z. in ZALETEL, M. (1990a). Costs and errors of Web surveys in establishment surveys. V: *Survey methods for businesses, farms, and institutions; invited papers: proceedings. Alexandria (Virginia): American Statistical Association*, 449–458.

VEHOVAR, V. (1992). *The imprecision of the estimates from telephone survey.* Postgraduated Diploma in Social Science Methodology and Data Analysis. University of Essex, Colchester.

VEHOVAR, V. (1992a). Plebiscit 1990: Zakaj so anketne napovedi zgrešile? V A. Ferligoj in J. Jug (ur.). *Blejsko metodološko srečanje '91 (Metodološki zvezki 7).* Jugoslovansko združenje za sociologijo, Sekcija za metodologijo in statistiko, Ljubljana.

VEHOVAR, V. (1993).The Field Substitution in the Slovene Labour Force Survey. *Bulletin of the ISI, Contributed Paper, 49th ISI Session, Firenze*, 519–520.

VEHOVAR, V. (1994). *Field substitutions in sample surveys – the case of the "Slovenian Public Opinion" survey: MA Thesis*. University of Essex, Colchester.

VEHOVAR, V. (1994a). Anketa o kadrovskem potencialu – pet let kasneje. V: Sonja Pirher in Ivan Svetlik (ur.). *Zaposlovanje: približevanje Evropi (Teorija in praksa),* 205–223. Fakulteta za družbene vede, Ljubljana.

VEHOVAR, V. (1995). The Field Substitution in the Slovene Public Opinion Survey. V A. Ferligoj in A. Kramberger (ur.). *Contributions to Methodology and Statistics (Metodološki zvezki 10)*, 38–66. Fakulteta za družbene vede, Ljubljana.

VEHOVAR, V. (1995a). *Nadomestne enote v anketnem raziskovanju: doktorska disertacija*. Ekonomska fakulteta, Univerza v Ljubljani, Ljubljana.

VEHOVAR, V., ZALETEL, M., BELAK, E. in ŠTEBE, J. (1996) Does the confidentiality concern increase the nonresponse rates? *Statistical confidentiality: collection of papers. Third International Statistical Seminar on Statistical Confidentiality,* 2–4 October, Bled, 53–63. Statistični urad RS: Eurostat, Ljubljana.

VEHOVAR, V. (1999). Field substitution and unit nonresponse. *Journal of official statistics*, 15 (2), 335–350.

VEHOVAR, V., LOZAR-MANFREDA, K., ZALETEL, M. in BATAGELJ, Z. (2001). Nonresponse in Web surveys. V R. Groves, D. Dillman, J. Eltinge in R. Little (ur.). *Nonresponse in Surveys.* Wiley, New York.

VEHOVAR, V. (2001). *Nepopolnost podatkov v anketnem raziskovanju (Metodološki zvezki 17).* Fakulteta za družbene vede, Univerza v Ljubljani, Center za metodologijo in informatiko, Ljubljana.

WAKSBERG, J. (1978). Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, 73, 40–46.

WARWICK, D. P. in LININGER, C. A. (1975). *The Sample Survey: Theory and Practice.* McGraw-Hill, New York.

WELNIAK, E. J. in CODER, J. F. (1980). A measure of the bias in the March CPS earnings imputation system. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 421–425.

WOLTER, K.M. (1985). *Introduction to variance estimation.* Springer-Verlag, New York.

WONNACOTT, H.W. in WONNACOTT, J.W. (1990). *Introductory statistics*. Wiley, New York.

YATES, F. (1981). *Sampling Methods for Censuses and Surveys.* Macmillan, New York.

ZALETEL, M. in VEHOVAR, V. (1998). The advantages of geo-demographic stratification. *New directions in surveys and censuses: proceedings*, 135–139. Minister of Industry, Ottawa.

SLOVAR ANGLEŠKIH IZRAZOV

accidental sampling – priložnostno vzorčenje
accuracy – točnost
area code – območna koda
area sample – prostorski vzorec
area sampling – prostorsko vzorčenje
art – veščina, spretnost, umetelnost
attrition – osip oziroma zmanjšanje velikosti vzorca
auxilary information – pomožna zunanja informacija
balanced – uravnotežen
Balanced Repeated Replications (*BRR*) – uravnotežene ponovljene replikacije
bank – blok 100 številk v telefonskih vzorcih ZDA
bias – pristranskost
blank element – prazen element
call-back – ponovno kontaktiranje pri telefonskem anketiranju
cell – celica v vzorcu, definirana z vrednostmi ene ali več spremenljivk
census – popis, anketa, ki vključi vse elemente v populaciji
central office code – lokalna trimestna koda v telefonskih vzorcih v ZDA
cluster – skupina
cluster sampling – vzorčenje v skupinah
collapsed strata – združevanje stratumov
completion rate – stopnja anketiranja
complex sample design – kompleksni vzorčni načrt
conditioning – prilagajanje anketirancev na anketna vprašanja v panelnih anketah
confidence interval – interval zaupanja
contact rate – stopnja kontaktiranja
controlled selection – nadzorovan izbor elementov v vzorec
cooperation rate – stopnja sodelovanja
cosmetic weighting – uteževanje, ki zagotovi videz reprezentativnosti
county – okraj, upravna enota v ZDA
coverage rate – stopnja pokritosti
crossclass – podskupine, ki so v populaciji razpršene po vseh segmentih (npr. spolno –starostne podskupine)
degree of freedom – prostostna stopnja
design effect – vzorčni učinek
disproportional stratification – disproporcionalna stratifikacija
domain – domena

domain of study – domena proučevanja
dual frame – podvojeni okvir
duplicate listing – podvojen zapis
eligibility rate – stopnja ustreznosti
enumeration area – popisni okoliš
enumeration district – popisno okrožje
equal probability selection method (EPSEM) – vzorec, kjer ima vsak element v populaciji enako verjetnost za izbor
establishment survey – anketa poslovnih subjektov
estimate – ocena
estimator – cenilka
expected value – pričakovana vrednost, matematično upanje
expert selection – ekspertna izbira
explicit stratification – eksplicitna stratifikacija
face-to-face interviews – terensko osebno anketiranje
field substitution – nadomestni elementi (rezervne enote), ki v terenski fazi anketiranja nadomestijo izpadle elemente zaradi neodgovorov
finite population correction (FPC) – popravek končne populacije
flat distribution – enakomerna porazdelitev
follow-up – ponovno kontaktiranje (pošiljanje dopisov) v raziskavah po pošti
foreign element – tuj element
gross changes – skupne spremembe
half-open interval – polodprti interval
half-sample replication – replikacija polvzorca
implicit stratification – implicitna stratifikacija
imputation – vstavljanje
inadequate – neprimeren
incentive – nagrada oziroma darilo za sodelovanje v anketi
incomplete – nepopoln
institutional population – institucionalna populacija
intraclass correlation – intraklasna korelacija
item nonresponse – neodgovor spremenljivke
item nonresponse rate – stopnja neodgovora spremenljivke
item response rate – stopnja odgovora spremenljivke
jackknife repeated replications (JRR) – Jackknife ponovljena replikacija
Kish selection grid – Kisheva metoda izbora osebe
linking procedure – postopek povezovanja
list-assisted telephone sampling – vzorčenje s pomočjo telefonskih imenikov
margin – absolutna napaka, širina intervala zaupanja v eni smeri
master sample – krovni vzorec
mean squared error (MSE) – srednja kvadratna napaka
measure of size – mera velikosti
missing element – manjkajoč element

monitoring – nadzor na procesom (npr. nadzor na delom anketarjev)
multiple imputation – metoda večkratnega vstavljanja
multipurpose design – večnamenska raziskava
multistage sampling – večstopenjsko vzorčenje
net changes – spremembe v razliki
Neyman allocation – Neymanova optimalna razmestitev elementov
noncontact rate – stopnja nekontaktiranja
noncooperation rate – stopnja nesodelovanja
noncoverage – nepokritost
non-informative prior distribution – neinformativna predhodna porazdelitev
nonprobability sample – neverjetnostni vzorec
nonresidential population – nerezidenčna populacija
nonrespondent – nerespondent
nonresponse – neodgovor
nonresponse rate – stopnja neodgovorov
one-way analysis of variance – enosmerna analiza variance
optimal allocation – optimalna razmestitev
outlier – izstopajoči element
paired selection design – vzorčni načrt z dvema enotama prve stopnje v vsakem stratumu, stratifikacijsko izbiranje parov
panel research – panelne raziskave
partial response – delni odgovori
population – populacija
population element variance – populacijska elementarna varianca
population total – populacijska vsota
posterior distribution – posteriorna porazdelitev
poststratification – poststratifikacija
precision – natančnost
predictive dialing – vnaprejšnja izbira
primary sampling unit (PSU) – enota prve stopnje
prior distribution – predhodna porazdelitev
probability proportional to estimated size (PPES) – verjetnost, ki je sorazmerna z ocenjeno velikostjo enot
probability proportional to size (PPS) – verjetnost, ki je sorazmerna z velikostjo enot
probability sample – verjetnostni vzorec
probability sampling with quotas (PSQ) – kombiniranje kvotnega in verjetnostnega vzorčenja
proportional stratification – proporcionalna stratifikacija
quota sampling – kvotno vzorčenje
random digit dialing (RDD) – slučajno generiranje telefonskih številk
random variable – slučajna spremenljivka
random-route sampling – vzorčenje slučajne poti
rare population – redka populacija

ratio estimator – razmernostna cenilka
ratio mean – razmerje, razmernostna aritmetična sredina
refusal – zavrnitev sodelovanja
refusal rate – stopnja zavračanja
refusion conversion – ponovno anketiranje oseb, ki so zavrnile sodelovanje, z bolj izkušenim anketarjem
reliability – zanesljivost
relvariance – relativna varianca
reminder letter – pismo opomnik
repeated replication – ponovljena replikacija
replacement technology – nadomestna tehnologija
replicated sampling – replicirano vzorčenje
residual – naključni ostanek
respondent – respondent, kooperativni anketiranec
response rate – stopnja odgovorov
ring-no-answer – telefonske številke, kjer se nihče ne javi
rotation panel – rotirajoči panel
sample – vzorec
sample element variance – elementarna varianca v vzorcu
sampling distribution – vzorčna porazdelitev
sampling fraction – vzorčni delež
sampling frame – vzorčni okvir
sampling interval – vzorčni interval
sampling variance – vzorčna varianca
sampling without restrictions – vzorčenje brez omejitev ali vzorčenje s ponavljanjem
scientific sample – znanstveni vzorec, verjetnostni vzorec
screening – krajši pregledni pogovor oziroma predhodni intervju
secondary sampling unit (SSU) – enota druge stopnje
selection equation – enačba izbora
Selfrepresenting PSU (SPSU) – samovključena enota prve stopnje
self-administered surveys – samoanketiranje, anketiranje brez anketarja
self-weighted samples – samouteženi vzorci, ki ne potrebujejo uteževanja
simple random sampling *(SRS)* – enostavno slučajno vzorčenje (brez ponavljanja)
simple random sampling (SRS) with replacement – enostavno slučajno vzorčenje s ponavljanjem
standard deviation – standardni odklon
standard error – standardna napaka
starting point – začetna točka pri metodi slučajne poti
statistical inference – statistično sklepanje
statistics – statistika
stratification – stratifikacija

subclass – podskupina, ki je lahko (za razliko od *cross-class*) tudi bolj segregirana
suffix – lokalni dodatek na osnovi zadnjih štirih cifer pri telefonskem vzorčenju v ZDA
survey – anketna raziskava
survey design – raziskovalni načrt
survey mode – način anketiranja
survey population – anketirana populacija
systematic sampling – sistematično vzorčenje
target population – ciljna populacija
Taylor series expansion – Taylorjeva vrsta
telemarketing – telefonska prodaja/trženje
telephone coverage – telefonsko pokritje
tracking – sledenje (preseljenim anketirancem)
truncation – zaokroževanje navzdol
two stage cluster sampling – dvostopenjsko vzorčenje v skupinah
two stage sampling – dvostopenjsko vzorčenje
two-phase sampling – dvofazno vzorčenje
two-stage sample – dvostopenjski vzorec
ultimate cluster – končna skupina
unbiased estimator – nepristranska cenilka
uneligible element – neustrezen element
unique identification – enotna identifikacija
unit nonresponse – neodgovor elementa
unsolicited Web surveys – spletne ankete, kjer se respondenti vključijo na osnovi javnega oglaševanja, torej brez osebnega vabila
validity – veljavnost
variance inflation factor (VIF) – faktor povečanja variance zaradi uteževanja
voice recognition – prepoznavanje govora
Web survey – spletno anketiranje
weighting adjustment – postopek uteževanja
within stratum variance – varianca znotraj stratumov