



SMERNICE ZA ZAGOTAVLJANJE KAKOVOSTI



št. 1 | 2017





SMERNICE ZA ZAGOTAVLJANJE KAKOVOSTI

2. izdaja

Smernice za zagotavljanje kakovosti
Zbrali in uredili: Ana Božič Verbič, Tina Steenvoorden

Publikacija je na voljo na spletnem naslovu: <http://www.stat.si/StatWeb/Catalogue/Index>

Informacije daje Informacijsko središče:

tel. (01) 241 64 04

elektronska pošta info.stat@gov.si

 @StatSlovenija

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

005.336.3:311(035)(0.034.2)

SMERNICE za zagotavljanje kakovosti [Elektronski vir] / [zbrali in uredili Ana Božič Verbič, Tina Steenvoorden]. - 2. izd. - El. knjiga. - Ljubljana : Statistični urad Republike Slovenije, 2017. - (Metodološki priročniki / Statistični urad Republike Slovenije, ISSN 2232-4879 ; 2017, št. 1)

Način dostopa (URL): <http://www.stat.si/StatWeb/Catalogue/Index>

ISBN 978-961-239-367-0 (pdf)

1. Božič Verbič, Ana

290680832

Izdal in založil Statistični urad Republike Slovenije, Ljubljana, Litostrojska cesta 54 – © SURS – Uporaba in objava podatkov dovoljeni le z navedbo vira – ISSN 2232-4879 – ISBN 978-961-239-367-0

UVODNA BESEDA

Statistični urad Republike Slovenije (SURS) je glavni izvajalec ter usklajevalec in povezovalac dela na področju državne statistike. Dejavnost državne statistike izvaja skupaj z drugimi pooblaščenimi izvajalci, ki so opredeljeni v veljavnem srednjeročnem programu statističnih raziskovanj. Vsi skupaj delujejo na podlagi Zakona o državni statistiki (Uradni list RS, št. 45/95 in 9/01) in Uredbe (ES) št. 223/2009 Evropskega parlamenta in Sveta o evropski statistiki.

Naslednje najpomembnejše naloge SURS so: mednarodno sodelovanje, določanje metodoloških in klasifikacijskih standardov, pravočasno zaznavanje potreb po podatkih in pripravljane ustreznih odzivov nanje (zbiranje, obdelava in izkazovanje podatkov) ter skrb za zagotavljanje statistične zaupnosti s pomočjo pravnih, organizacijskih in tehnoloških postopkov.

Pri opravljanju svojih nalog se ravna po splošnih načelih vodenja kakovosti in Kodeksu ravnanja evropske statistike. Ta je skladen s Temeljnimi načeli uradne statistike, ki jih je potrdila Komisija za statistiko pri Združenih narodih, in z nacionalno statistično zakonodajo in z zakonodajo EU. SURS ob tem upošteva (skladno s svojo izjavo o kakovosti) tudi naslednja načela: strokovno neodvisno delovanje, procesno usmerjeno delovanje, skrb za kakovostne izdelke in storitve, usmerjeno načrtovanje izboljšav, spodbudno delovno okolje za zaposlene, prijaznost do dajalcev podatkov ter usmerjenost k uporabnikom.

Prav zaveza SURS, da bo nenehno spremljal in izboljševal kakovost, je bila povod, da je leta 2012 izdal priročnik za sistematičen in usmerjen razvoj upravljanja kakovosti z naslovom Smernice za zagotavljanje kakovosti. V priročniku so bili tako prvič na enem mestu sistematično in podrobno opisani posamezni deli procesa izvedbe statističnega raziskovanja in podane smernice, ki jim morajo izvajalci slediti, da potem lahko pričakujejo tudi kakovostno izvedbo in kar najbolj kakovostne končne rezultate.

Spremembe in razvoj delovnega okolja ter prizadevanja za kar najbolj kakovostne storitve in produkte zahtevajo nenehne spremembe posameznih faz kompleksnega statističnega procesa. Prav zato je bilo treba priročnik, izdan leta 2012, posodobiti.

Tudi nova različica priročnika Smernice za zagotavljanje kakovosti je tako kot prejšnja rezultat dela in znanja številnih oz. skoraj vseh zaposlenih na SURS. Vsak je strokovnjak na svojem področju in ima v procesih, ki so tukaj popisani, svojo vlogo. Priznanje njihovem delu bomo vsi, ki smo kakorkoli povezani z izvedbo statističnih raziskovanj, najlepše izkazali tako, da bomo knjigo intenzivno uporabljali in upoštevali smernice, obenem pa jih tudi izboljševali in izpopolnjevali.

Genovefa Ružič,
generalna direktorica



KAZALO

UVOD	7
1 ANALIZA POTREB IN ZAHTEV	9
1.1 UGOTAVLJANJE POTREBE PO PODATKIH	9
1.2 PREUČITEV VIROV	10
1.3 PREVERJANJE METODOLOGIJE	10
2 NAČRTOVANJE IN PRIPRAVA RAZISKOVANJA	12
2.1 NAČRTOVANJE SREDSTEV IN DOLOČITEV SEZNAMA AKTIVNOSTI Z ROKI	12
2.2 DEFINIRANJE REZULTATOV RAZISKOVANJA	13
2.3 PRIPRAVA METODOLOGIJE ZBIRANJA PODATKOV	14
2.4 PRIPRAVA METODOLOGIJE IZBORA ENOT OPAZOVANJA	16
2.5 PRIPRAVA METODOLOGIJE OBDELAVE PODATKOV	18
2.6 NAČRTOVANJE IN TESTIRANJE IZVEDBE RAZISKOVANJA	19
3 IZBOR ENOT OPAZOVANJA	21
3.1 PRIPRAVA PODATKOVNIH VIROV ZA IZGRADNJO VZORČNEGA OKVIRA	21
3.2 PRIPRAVA VZORČNEGA OKVIRA	23
3.3 IZBOR ENOT OPAZOVANJA	24
3.4 IZDELAVA ADRESARJA	26
4 ZBIRANJE PODATKOV	27
4.1 PRIPRAVA NA ZBIRANJE	27
4.2 PREVZEM ADMINISTRATIVNIH IN DRUGIH ZBIRK PODATKOV	28
4.3 ZBIRANJE PODATKOV IN KOMUNICIRANJE S POROČEVALSKIMI ENOTAMI	29
4.4 ZAJEM PODATKOV	30
5 OBDELAVA PODATKOV	32
5.1 UREJANJE ADMINISTRATIVNIH in drugih ZBIRK PODATKOV	32
5.2 INTEGRACIJA RAZLIČNIH PODATKOVNIH VIROV	33
5.3 UREJANJE PODATKOV NA MIKRORAVNI	34
5.4 VSTAVLJANJE PODATKOV (IMPUTACIJE)	35
5.5 UTEŽEVANJE	36
5.6 IZRAČUN STATISTIČNIH OCEN (AGREGACIJA)	38
5.7 DEFLACIJA	39
5.8 UREJANJE PODATKOV NA MAKRORAVNI	40
5.9 TABELACIJA	41
5.10 STATISTIČNA ZAŠČITA PODATKOV	41

6 ANALIZA PODATKOV	45
6.1 ANALIZA ČASOVNIH VRST	45
6.2 ANALIZA USTREZNOSTI TER POTRDITEV REZULTATOV.....	47
6.3 INTERPRETACIJA REZULTATOV	47
7 IZKAZOVANJE PODATKOV.....	49
7.1 POSODABLJANJE IZHODNIH PODATKOV	49
7.2 PREDSTAVITEV REZULTATOV.....	50
7.3 OBJAVLJANJE	51
7.4 PODPORA UPORABNIKOM.....	51
7.5 HRAMBA STATISTIČNIH MIKROPODATKOV	52
7.6 HRAMBA AGREGIRANIH PODATKOV	53
8 DOKUMENTIRANJE IN EVALVACIJA RAZISKOVANJA	54
8.1 IZDELAVA DOKUMENTACIJE O RAZISKOVANJU	54
8.2 ZBIRANJE INFORMACIJ ZA OCENO KAKOVOSTI.....	55
8.3 IZVEDBA OCENE POSTOPKOV IN PROCESOV	56
KRATICE.....	57
VIRI IN LITERATURA	58

UVOD

Druga, dopolnjena izdaja priročnika **Smernice za zagotavljanje kakovosti** je dopolnitev prilagojene različice splošnega statističnega poslovnega procesnega modela, sprejetega in veljavnega na SURS; gre za prilagojeno različico mednarodno sprejetega splošnega statističnega poslovnega procesnega modela (angl. Generic Statistical Business Process Model - GSBPM), ki je nastal pod okriljem Gospodarske komisije Združenih narodov za Evropo (angl. United Nations Economic Commission for Europe – UNECE). Ta model opisuje ter opredeljuje nabor poslovnih procesov, ki so potrebni za izdelavo uradnih statističnih podatkov. Tako splošni procesni model SURS kot smernice za zagotavljanje kakovosti tvorijo standardni okvir in vsebujejo standardno terminologijo, ki je potrebna za usklajeno delovanje in usposabljanje zaposlenih na SURS in je obenem tudi osnova za sistematično uvajanje izboljšav v posamezne procese in raziskovanja.

Glavne spremembe v novi različici priročnika glede na prejšnjo so naslednje:

- nekateri prej samostojni podprocesi v okviru 2. procesa (Načrtovanje in priprava raziskovanja) so bili odstranjeni, njihova vsebina pa vključena v druge podprocese;
- nekateri podprocesi so bili nekoliko preimenovani, z namenom, da bi bila vsebina posameznega podprocesa razumljivejša;
- opisi in smernice za zagotavljanje kakovosti posameznih procesov in podprocesov so bili posodobljeni v skladu s spremembami in razvojem državne statistike v zadnjih letih (tudi izrazje).

Splošni procesni model SURS obsega 8 procesov; ti se nato delijo na različno število podprocesov. Poglavlja in podpoglavja v dokumentu sledijo strukturi tega procesnega modela. Vsako poglavje sestavlja kratek opis celotnega procesa, vsa podpoglavja pa so nato sestavljena iz dveh sklopov: prvi vsebuje splošen opis podprocesa, drugi pa smernice za zagotavljanje kakovosti, ki naj bi jih izvajalci uporabljali kot nekakšen kontrolni seznam elementov kakovosti, na katere morajo biti pozorni, ko izvajajo podproces. Čeprav procesni model sledi v zaporedju procesov in podprocesov kolikor mogoče dejanski izvedbi raziskovanja, pa so posamezne faze pogosto povezane in prepletene, da si lahko statistični proces predstavljamo kot enostaven linearni model. Zato moramo na model gledati kot na shematičen prikaz teoretičnega zaporedja postopkov in se ob tem zavedati, da se v praksi vsakokrat prilagodi značilnostim izvedbe posameznega statističnega raziskovanja.

Druga, dopolnjena izdaja priročnika Smernice za zagotavljanje kakovosti je namenjena vsem izvajalcem dejavnosti državne statistike, ki so kakor koli vključeni v statistični proces, drugim izvajalcem statističnih raziskovanj in seveda vsem, ki jih poleg končnih izsledkov (statističnih podatkov in informacij) izvajalcev dejavnosti državne statistike zanimata tudi ozadje in celoten postopek, ki na koncu pripelje do uporabne, verodostojne statistične informacije.

Splošni procesni model SURS, 2017

1	ANALIZA POTREB IN ZAHTEV	Ugotavljanje potrebe po podatkih	Preučitev virov	Preverjanje metodologije
2	NAČRTOVANJE IN PRIPRAVA RAZISKOVANJA	Načrtovanje sredstev in določitev seznama aktivnosti z roki	Definiranje rezultatov raziskovanja	Priprava metodologije zbiranja podatkov
				Priprava metodologije izbora enot opazovanja
				Priprava metodologije obdelave podatkov
				Načrtovanje in testiranje izvedbe raziskovanja
3	IZBOR ENOT OPAZOVANJA	Priprava podatkovnih virov za izgradnjo vzorčnega okvira	Priprava vzorčnega okvira	Izbor enot opazovanja
				Izdelava adresarja
4	ZBIRANJE PODATKOV	Priprava na zbiranje	Prevzem administrativnih virov	Zbiranje podatkov in komuniciranje s poročevalskimi enotami
				Zajem podatkov
5	OBDELAVA PODATKOV	Urejanje administrativnih podatkovnih zbirk	Integracija različnih podatkovnih virov	Urejanje podatkov na mikroravni
				Vstavljanje podatkov (imputacije)
				Uteževanje
				Izračun statističnih ocen (agregacija)
				Deflacija
				Urejanje podatkov na makroravni
				Tabelacija
				Statistična zaščita podatkov
6	ANALIZA PODATKOV	Analiza časovnih vrst	Analiza ustreznosti ter potrditev rezultatov	Interpretacija rezultatov
7	IZKAZOVANJE PODATKOV	Posodabljanje izhodnih podatkov	Predstavitvev rezultatov	Objavljanje
				Podpora uporabnikom
				Hramba statističnih mikropodatkov
				Hramba agregiranih podatkov
8	DOKUMENTIRANJE IN EVALVACIJA RAZISKOVANJA	Izdelava dokumentacije o raziskovanju	Zbiranje informacij za oceno kakovosti	Izvedba ocene postopkov in procesov

1 ANALIZA POTREB IN ZAHTEV

Odločitev za zbiranje podatkov se sprejme na podlagi temeljite analize potreb in zahtev širokega kroga uporabnikov po statističnih podatkih (npr. izpolnjevanje pravnih zavez države, zahteve nacionalnih in mednarodnih institucij, potrebe strokovne in splošne zainteresirane javnosti). Njihova pričakovanja in potrebe so zaradi družbenih sprememb, globalnih procesov in novih pobud na mednarodni in nacionalni ravni različni. Pravočasno zaznavanje potreb po podatkih in priprava ustreznih odzivov nanje sta ključna za sistematično in strateško odločanje. Potrebe po podatkih ugotavljamo v stalnem dialogu z uporabniki po različnih komunikacijskih kanalih, obenem pa je treba upoštevati tudi zmožnosti (kadrovske, finančne, IT-opreme) in okvire delovanja državne statistike.

Po analizi potreb po podatkih ugotovimo, ali lahko želene podatke pridobimo iz že obstoječih virov (statističnih, administrativnih in drugih), saj lahko tako bistveno racionaliziramo pridobivanje podatkov. Pri tem moramo upoštevati širok seznam meril, ki nam bodo omogočala ustrezno izbiro primerne vira podatkov (npr. opredelitev enot opazovanja, določitev identifikatorjev, ugotavljanje ustrezne vsebine, kakovosti virov in drugih značilnosti), in tudi načela Kodeksa ravnanja evropske statistike. Pred načrtovanjem statističnega raziskovanja (dalje raziskovanja) preverimo tudi ustreznost zdajšnje, tj. obstoječe metodologije za vse nadaljnje faze predvidenega raziskovanja (to lahko že obstaja, lahko pa se uvaja na novo).

Proces analize potreb in zahtev uporabnikov statističnih rezultatov po podatkih daje osnovo za pripravo vseh nadaljnjih aktivnosti in za sprejemanje odločitev o načrtovanju raziskovanja.

1.1 UGOTAVLJANJE POTREBE PO PODATKIH

Ugotavljanje potrebe po podatkih se sproži, ko podatkov še ni ali ko obstoječi podatki ne zadovoljujejo vseh uporabnikovih potreb po podatkih, lahko pa tudi na podlagi novih zahtev zakonodajalca ali sklenitve partnerskih dogovorov.

Pobude o potrebah po podatkih prihajajo od različnih vrst uporabnikov: izpolnjevanje pravnih zavez države, iz nacionalnih in mednarodnih institucij ter od strokovne in splošne zainteresirane javnosti. V postopku ugotavljanja potreb po podatkih zaznavamo, kaj uporabniki pričakujejo (kateri podatki naj bi se jim po njihovih pričakovanih zagotovili). Za ta namen je treba sodelovati z vsemi zainteresiranimi uporabniki na različne načine (statistični sosveti, statistični svet, delovne skupine, dvostranski dogovori, sodelovanje pri pripravi zakonskih podlag, izvajanje ankete o zadovoljstvu uporabnikov). Seznaniti se je treba s pričakovanji uporabnikov, predvsem ugotoviti, katere podatke uporabniki potrebujejo, kdaj jih potrebujejo, na kakšen način, predvsem pa, za kakšen namen. Zato je ključnega pomena dobro in stalno sodelovanje z uporabniki podatkov, saj to pripomore k razumevanju in sprotnemu spremljanju njihovih potreb. Spremljanje potreb uporabnikov podatkov obsega tako spremljanje potreb po novih podatkih, ukinjanje (opušcanje) zbiranja podatkov kot tudi spremembe v obstoječih podatkih, ki so posledica sprememb v metodologiji.

Po identificiranju pomembnih novih ali spremenjenih potreb po statističnih podatkih je treba določiti nabor podatkov, ki so potrebni za zadovoljitev uporabnikovih potreb, skladno z možnostmi. Pri tem je pomembno, da se v največji možni meri doseže soglasje glede kakovosti podatkov.

Smernice za zagotavljanje kakovosti

- Ugotavljanja potreb po podatkih se je treba lotiti celovito in sistematično z vključitvijo vseh zainteresiranih uporabnikov.
- Pri uporabnikih je treba njihove potrebe po podatkih redno preverjati.
- O njihovih potrebah po podatkih se je treba dobro poučiti, predvsem o tem, katere podatke uporabniki potrebujejo, kdaj, za kakšen namen, na kakšen način in kako podrobno razčlenjene.
- Pri ugotavljanju potreb uporabnikov po podatkih naj se upoštevajo tudi njihove morebitne sorodne potrebe, ki jih morda z malo dodatnega truda lahko enostavno izpolnimo.
- Pri sprejemanju odločitev o zbiranju novih podatkov je treba oceniti, kako pomembni (potrebni) so določeni podatki. Pretehtati je treba stroške in koristi, ki jih bo prineslo zbiranje podatkov.
- Proučijo naj se zahteve po podatkih, ki izhajajo iz uredb, in zahteve uporabnikov; na podlagi teh ugotovitev se potem pripravijo sezname statistik (rezultatov) in za vsako statistiko se določi raven podrobnosti objavljanja.

1.2 PREUČITEV VIROV

Potem ko se na podlagi zaznanih novih potreb uporabnikov po statističnih podatkih sprejme odločitev, da se bodo določeni podatki pridobili, je treba pregledati vsa raziskovanja, ki se že izvajajo, in vsebino administrativnih in drugih zbirk podatkov ter najprej ugotoviti, ali lahko novim potrebam po podatkih zadostimo na podlagi že obstoječih podatkov, ki so na voljo v kakšnem že obstoječem viru (statističnem ali v administrativnih in drugih zbirkah podatkov). Če so taki podatki na voljo, je naslednji korak preučitev skladnosti obstoječih virov s potrebami po novih podatkih oziroma preučitev morebitnih omejitev, ki bi onemogočale uporabo teh virov za zadostitev novih potreb po podatkih (razlike v metodologiji, periodika in namen zbiranja podatkov, stabilnost vira idr.). Preučitev virov se izvede predvsem kot podpora odločitvi o tem, ali bo obstoječi vir mogoče uporabiti kot neposredni vir podatkov in v kolikšni meri. Če želenih podatkov za izpolnitev novih potreb po podatkih v pregledanih in dostopnih virih ni, je treba pridobivanje teh podatkov vključiti v že obstoječa raziskovanja ali uvesti novo raziskovanje.

Skladno s prizadevanji, da se zmanjšajo administrativne ovire in racionalizira poslovanje, se je treba zavzemati, da se za ta namen uporabijo v čim večji meri administrativni viri. Pri tem se je treba zavedati dejstva, da je bil administrativni vir prvotno izdelan za druge, tj. nestatistične namene. Pomembno je, da poznamo razlike v vzorčnem okviru, v definicijah, v pravočasnosti podatkov, v kakovosti vira itd. Dobro moramo poznati tudi zakonske podlage, na podlagi katerih je bil administrativni vir vzpostavljen. Pomembno je, da stalno sodelujemo s skrbnikom administrativnega vira, zlasti pri pripravi dogovora in ob morebitnih spremembah. Če je mogoče, se je treba vključiti že takrat, ko se pripravljata ali uvajata nov administrativni vir in tudi ko se pripravljata zakonske podlage za uvedbo vira.

Smernice za zagotavljanje kakovosti

- Preučiti je treba, ali za pridobitev zelenih podatkov lahko uporabimo že obstoječe statistične in/ali administrativne podatke, ali bi zadostovalo, da se za ta namen dopolni že obstoječe raziskovanje, ali pa je treba načrtovati novo raziskovanje.
- Preučiti je treba kakovost administrativnih in drugih virov (v širšem pomenu). Če so ustrezne kakovosti, ima uporaba teh virov prednost pred neposrednim zbiranjem podatkov.
- Pripraviti je treba merila za določanje hierarhije podatkovnih virov ter s tem tudi prioritete njihovega prevzemanja; pri tem se lahko upoštevajo kakovost podatkov v virih, pogostost posodabljanja (osveževanja) vira, relevantnost vira, metodološka ustreznost vsebin, časovna dostopnost podatkov.
- Redno je treba ugotavljati stabilnost podatkov v administrativnem viru in preverjati njihovo kakovost (npr. s primerjavo posameznih zaporednih nizov podatkov ali s posebej za ta namen pripravljenimi vzorčnimi raziskovanji).
- Pripraviti je treba primerjavo administrativnih in statističnih virov za posamezne vsebine, če so na voljo.
- Preučiti je treba, ali metodologija upravljanja administrativnega vira vpliva na izkazovanje statističnih rezultatov in v kolikšni meri, še zlasti, če se podatki za posamezno spremenljivko v raziskovanju, ki so se prej zbirali neposredno, zdaj v celoti prevzemajo iz administrativnih virov.
- Ob vzpostavitvi novega administrativnega vira ali ob predlaganih spremembah zakonodaje si je treba prizadevati, da se upošteva interes in zagotovi vpliv državne statistike. Proučiti je treba vpliv predlaganih sprememb na metodologijo, vsebino in na statistične procese v posameznem raziskovanju.

1.3 PREVERJANJE METODOLOGIJE

Preverjanje ustreznosti metodologije raziskovanja glede na zaznane potrebe po podatkih vključuje preverjanje teoretičnih vidikov metodologije raziskovanja (npr. ciljna populacija, definicije, metodologija vzorčenja, uporaba obstoječega vira, obdelava podatkov idr.) in preverjanje praktičnih vidikov metodologije raziskovanja (npr. oblika vprašalnika, način zbiranja podatkov, postopek obdelave podatkov, vrste objav idr.).

Potek tega postopka je odvisen tudi od tega, ali se metodologija pripravljata na podlagi evropske zakonodaje ali na podlagi standardov drugih mednarodnih institucij. Če gre za prvo, je treba upoštevati, ali je ta uredba že sprejeta (preverjamo, na kakšen način se lahko uskladimo z zahtevami uredbe) ali postopek sprejemanja oziroma spreminjanja uredbe še poteka (ugotavljamo, ali bi bilo raziskovanje mogoče izvesti tako, kot je predlagano, in na tej osnovi oblikujemo stališča za Slovenijo). Ta stališča je treba potem utemeljevati na različnih ravneh: v delovnih in direktorskih skupinah, na Odboru za evropski statistični sistem (angl. ESSC) in na Svetu Evropske unije (Svet EU).

Smernice za zagotavljanje kakovosti

- Pri preverjanju ustreznosti metodologije dajemo prednost obstoječim metodam in dobrim praksam. V vseh fazah preverjanja ustreznosti metodologije sledimo tehnološkemu razvoju in skušamo prilagoditi obstoječe postopke v skladu z danimi zmožnostmi institucije.
- Pri preverjanju izvedbe posameznih faz raziskovanja je treba vedno upoštevati dane možnosti izvedbe.
- V pripravo metodologije raziskovanja, ki se izvaja na ravni EU, se je treba aktivno vključiti že na začetku, ko se začne tematika obravnavati v delovni skupini. Tako lahko vplivamo na metodologijo (npr. na ciljno populacijo, definicije, klasifikacije, periodiko). Hkrati preučimo tudi tisti del metodologije, ki običajno ni predpisan (npr. način zbiranja podatkov, metodologija statistične obdelave).

2 NAČRTOVANJE IN PRIPRAVA RAZISKOVANJA

Ustrezno načrtovanje in priprava raziskovanja sta ključna za uspešen potek in izvedbo samega raziskovanja. Pri tem izhajamo iz prepoznanih potreb uporabnikov in drugih ugotovitev, do katerih smo prišli s preučevanjem potreb in zahtev po podatkih, in iz načel Kodeksa ravnanja evropske statistike, ki nas zavezujejo h kakovosti z upoštevanjem načel glede dobre metodologije, ustreznih statističnih postopkov, razumne obremenitve dajalcev podatkov in stroškovne učinkovitosti.

Ključni elementi raziskovanja morajo biti navedeni v letnem programu statističnih raziskovanj (LPSR), npr.: namen, vsebina raziskovanja, ciljna populacija, sporočitelj podatkov (kdo sporoči podatke), rok za oddajo podatkov (do kdaj je dolžen to storiti), rok za objavo rezultatov in drugi elementi. LPSR je javno objavljen, in tako lahko uporabniki preverijo, katere statistične podatke lahko pričakujejo, opazovane enote (izbrane osebe, gospodinjstva, kmetijska gospodarstva, podjetja, organizacije itd.) pa se lahko seznanijo, katere podatke bodo morale sporočiti. Pri pripravi raziskovanj, ki jih izvajamo na podlagi evropskih uredb ali dogovorov, upoštevamo zahteve in priporočila, ki so navedena v teh podlagah. Podrobnejši načrt posamezne faze v izvajanju raziskovanja je treba navesti v letnem delovnem načrtu. Pri raziskovanjih, za katera potrebujemo dodatna finančna sredstva (npr. za anketiranje oseb, gospodinjstev in kmetijskih gospodarstev, za popise), dodatno informacijsko tehnološko (IT-) opremo in dodatne človeške vire, je treba vse vire (kadrovske, finančne, IT-opremo) načrtovati že zelo zgodaj. Za takšna raziskovanja se poleg letnega delovnega načrta podrobneje opredeli tudi finančni načrt.

Začetne faze pri načrtovanju raziskovanja so običajno naslednje: opredelitev vsebine, spremenljivk in rezultatov raziskovanja; opredelitev ciljne populacije oziroma opazovane populacije; opredelitev virov podatkov in načina zbiranja podatkov.

Nadaljnje korake določimo glede na predvidene nadaljnje faze raziskovanja:

Če bomo podatke pridobili neposredno pri opazovanih enotah, moramo določiti metodo za izbiro enote (verjetnostni ali neverjetnostni vzorec, popis), določiti način zbiranja podatkov, glede na izbrani način zbiranja podatkov pripraviti ustrezne vprašalnike in seznam spremljajočih gradiv.

Kadar del podatkov ali vse podatke za določeno raziskovanje pridobimo iz že obstoječih virov, najprej preverimo, ali so podatki že dostopni.

Če podatkov še ne prevzemamo, moramo prej, v dogovoru z institucijo, od katere želimo podatke prevzeti, začeti postopek za prevzem in pripraviti dogovor, tehnični protokol in druge potrebne dokumente.

Če bomo del podatkov pridobili neposredno pri opazovanih enotah, del pa iz že obstoječih virov, ustrezno načrtujemo in pripravimo postopke za izvedbo obeh načinov pridobivanja podatkov.

Metodo za izbiro enot opazovanja podrobno opredelimo že v zgodnji fazi načrtovanja in priprave raziskovanja. Tako podrobneje opredelimo tudi posamezne korake izvedbe (metodologijo izbora enot opazovanja, zbiranja podatkov, statistične obdelave podatkov). Določiti je treba programska orodja za izvedbo posameznega koraka, opredeliti tok podatkov in definirati baze podatkov. V čim večji meri je treba uporabiti standardizirane in že izdelane rešitve. Za kakovostno izvedbo raziskovanja je treba posamezne postopke in rešitve pred samo izvedbo testirati. Ob vsem tem je pomembno, da se že ob načrtovanju raziskovanja predvidi sprotna priprava dokumentacije, ki zagotavlja sledljivost postopkov

2.1 NAČRTOVANJE SREDSTEV IN DOLOČITEV SEZNAMA AKTIVNOSTI Z ROKI

Za uspešno izvedbo raziskovanja je treba načrtovati porabo finančnih in materialnih sredstev ter določiti letni delovni načrt za izvedbo nalog. Natančno načrtovanje je ključno za učinkovito izvedbo statističnega procesa.

Vsako raziskovanje je treba vpisati v letni program statističnih raziskovanj (LPSR). Raziskovanje, katerega metodologija se šele postavlja ali ki se bo izvedlo kot pilotno raziskovanje, vključimo v razvojni del, raziskovanje, ki se bo izvajalo redno, pa v redni del programa. V LPSR zapišemo tudi datum prve objave rezultatov. Ta se vključi tudi v Koledar objav, katerega namen je informiranje uporabnikov o objavah podatkov.

Natančen potek dela in razdelitev delovnih faz se določita v letnem delovnem načrtu. Ta vsebuje seznam aktivnosti in imena posameznikov, ki so zadolženi za njihovo izvajanje, zaporedje izvajanja aktivnosti in časovne okvire. Časovni okvir aktivnosti in seznam posameznikov, ki so zadolženi za izvajanje posameznih aktivnosti, je treba uskladiti z vsemi deležniki v procesu.

Pred začetkom izvajanja raziskovanja je treba pripraviti tudi kadrovske in finančne načrte, ga pravočasno vključiti v pripravo proračuna in načrtovati potrebe po dodatnih materialnih sredstvih. To velja predvsem za prvo izvajanje posameznega raziskovanja, za večje revizije, za večletna raziskovanja in za večja raziskovanja, pri katerih

prihaja stalno do sprememb. Finančni načrt obsega oceno sredstev, potrebnih za posamezne namene, po mesecih, kadrovski načrt pa oceno ur (dni, mesecev), potrebnih za posamezne aktivnosti, po mesecih. Kadrovski načrt mora upoštevati tako redno zaposlene kot tudi pogodbeni sodelavci (anketarje). Če se načrtuje izvedba določene aktivnosti ali storitve v sodelovanju s pogodbenimi sodelavci ali z zunanjimi izvajalci, je treba pravočasno načrtovati vse potrebno za izvedbo postopkov, s katerimi se ti sodelavci pridobijo.

Smernice za zagotavljanje kakovosti

- Potek raziskovanja je treba načrtovati skladno z izkušnjami iz prejšnjih izvedb raziskovanja in glede na razpoložljive vire (kadri, finance, informacijsko tehnologijo).
- Vpis raziskovanja v letni program statističnih raziskovanj (LPSR) z vsemi elementi, ki jih vsebuje, so ključne informacije za poročevalske enote in za uporabnike, zato je treba pri načrtovanju izvedbe raziskovanja izhajati iz tega zapisa oz. obstoječa raziskovanja ustrezno popisati.
- Objavo podatkov v Koledarju objav je treba načrtovati skladno z zahtevami po pravočasnosti objavljanja podatkov, ki izhajajo iz potreb uporabnikov oz. drugih pravnih zahtev (npr. Uredbe EU), pri čemer upoštevamo načelo, da podatke objavimo v najkrajšem možnem roku.
- Pri načrtovanju sredstev in aktivnosti je pomembno sodelovanje vseh vpletenih v statistični proces, s čimer se doseže optimalna porazdelitev sredstev in časa.
- Izkušnje pri poteku izvedbe raziskovanja je treba ustrezno upoštevati pri pripravi rokovnikov dela za naslednja obdobja.
- Pri načrtovanju finančnih sredstev naj se, če se bodo podatki zbirali na terenu, čim natančneje oceni višina posameznih postavk (plačilo odgovora, neodgovora, kilometrine) in upošteva struktura anketarjev glede na vrsto pogodbenega razmerja. Za pripravo teh ocen se lahko uporabijo preglednice za simulacijo stroškov, v katere se vnesejo posamezni parametri ankete (velikost vzorca, cena odgovora, kilometrina ipd.).
- Zaradi lažjega in nemotenega izvajanja raziskovanj ter zaradi pridobivanja potrebnih materialnih sredstev je treba dovolj zgodaj načrtovati časovno porazdelitev raziskovanj, ki se bodo izvajala v naslednjem letu.

2.2 DEFINIRANJE REZULTATOV RAZISKOVANJA

Pomemben korak pri načrtovanju raziskovanja je, da podrobno opredelimo (definiramo) pričakovane rezultate, ki smo jih okvirno določili že med preučevanjem potreb in zahtev po podatkih. Pri tem si moramo v čim večji meri prizadevati, da bodo vsi elementi raziskovanja čim bolj usklajeni z metodami in koncepti iz prejšnjih ali podobnih raziskovanj ter v skladu s standardi. Po potrebi se dodatno posvetujemo z uporabniki podatkov in s strokovnjaki z obravnavanega področja. Koristen vir informacij so tudi morebitna predhodno izvedena raziskovanja in pilotna raziskovanja.

Glede na določeno ciljno populacijo upoštevamo: vrsto enot opazovanja in lastnosti, ki določajo populacijo, geografsko lokacijo enot, ki določajo populacijo, ter časovno (referenčno) obdobje, za katero veljajo lastnosti populacije, ki nas zanimajo. Pripraviti moramo tudi podroben seznam vseh statistik, ki jih želimo z raziskovanjem pridobiti, ter pri vsaki označiti vir podatkov (vprašalnik, administrativni ali drugi), zahtevano referenčno obdobje in razpoložljivost vira.

Pri določanju klasifikacij si pomagamo z že obstoječimi klasifikacijami, ki se nahajajo v klasifikacijskem strežniku KLASJE, ter z obstoječimi klasifikacijami domačih in mednarodnih organizacij (Eurostata, Združenih narodov, Organizacije za ekonomsko sodelovanje in razvoj, Svetovne banke, Mednarodne organizacije dela in drugih). Klasifikacije je treba opredeliti tako, da poročevalskim enotam olajšamo razvrščanje pojava, hkrati pa omogočimo objavo na dovolj podrobni ravni, ne da bi pri tem imeli težave glede zaupnosti in natančnosti podatkov; to lahko dosežemo s tako imenovanimi hierarhičnimi klasifikacijami.

Smernice za zagotavljanje kakovosti

- Pri določitvi (opredelitvi) pričakovanih rezultatov raziskovanja v čim večji meri upoštevamo potrebe EU in potrebe, ki izhajajo iz nacionalne zakonodaje.
- Kjer je le mogoče, uporabimo standardne koncepte, spremenljivke in klasifikacije, in sicer tako pri nazivih kot pri definicijah.
- Vse koncepte, spremenljivke in klasifikacije, ki jih bomo uporabili v raziskovanju, podrobno opredelimo, posebej pa opišimo še vsa morebitna neskladja s standardnimi koncepti.
- Pri uporabi klasifikacij je priporočljivo v čim večji meri uporabljati tako imenovane hierarhične klasifikacije.

2.3 PRIPRAVA METODOLOGIJE ZBIRANJA PODATKOV

Priprava metodologije za zbiranje podatkov vključuje metodologijo in instrumente, potrebne za zbiranje podatkov. Če podatke pridobivamo neposredno pri opazovanih ali njihovih poročevalskih enotah, so elementi te metodologije nekoliko drugačni, kot če jih pridobimo iz že obstoječih podatkovnih baz. Na odločitev o tem, kako bomo podatke zbrali, vplivajo različni dejavniki: značilnost ciljne populacije, cilji zbiranja, razpoložljivost virov, časovne omejitve. Čedalje pogosteje se del podatkov pridobi neposredno, del pa iz obstoječih podatkovnih baz. Z ustrezno pripravo metodologije za zbiranje podatkov si zagotovimo, da se podatki zberejo sistematično. Zavedati se je treba, da je izbira načina zbiranja podatkov pomembna odločitev, saj vpliva na celoten statistični proces in na višino sredstev, potrebnih za izvedbo raziskovanja.

V tabeli je prikazana osnovna razvrstitev načinov zbiranja glede na to, ali podatke zbira anketar in ali je pri zbiranju ali oddajanju podatkov uporabljen računalnik.

Načini zbiranja podatkov glede na vključenost anketarja in uporabo računalnika pri zbiranju podatkov

Vključenost anketarja	Uporaba računalnika	
	Ne. Vprašalniki v tiskani obliki (P&P, angl. Paper and Pencil Surveys).	Da. Računalniško podprto zbiranje (CASIC, angl. Computer-assisted Survey Information Collection).
Da. Vprašalnik izpolnjuje anketar.	<ul style="list-style-type: none"> – Tiskani vprašalnik izpolnjuje anketar, medtem ko anketira respondenta – PAPI (angl. Paper and Pen Interviewing). 	<ul style="list-style-type: none"> – Računalniško podprto osebno anketiranje – CAPI (angl. Computer Assisted Personal Interviewing; face-to-face). – Računalniško podprto telefonsko anketiranje, prek fiksnega ali mobilnega telefona – CATI (angl. Computer Assisted Telephone Interviewing).
Ne. Samoizpolnjevanje. (respondent vprašalnik izpolnjuje sam).	<ul style="list-style-type: none"> – Tiskani vprašalnik prejme respondent po pošti in ga sam izpolni ter vrne po pošti (poštna metoda) – PAP. – Dnevnik (npr. dnevnik izdatkov). 	Elektronski vprašalniki <ul style="list-style-type: none"> – spletni vprašalnik – WEB – po meri narejeni vprašalniki (Excel).

Pri pripravi vprašalnikov v tiskani obliki je treba predvideti kasnejšo digitalizacijo podatkov.

Če bomo podatke zbirali neposredno pri poročevalski enoti, moramo določiti najustreznejši (optimalen) način zbiranja podatkov in tudi način obveščanja izbranih enot o njihovi vključitvi v raziskovanje. Načine zbiranja podatkov lahko tudi kombiniramo. Na primer: najprej lahko anketiramo po telefonu, nato pa še osebno na terenu; lahko uporabimo spletni vprašalnik v kombinaciji s telefonskim in/ali osebnim terenskim anketiranjem. Vsak način zbiranja podatkov ima svoje prednosti in slabosti; v nadaljevanju navajamo najznačilnejše.

Navzočnost anketarja pri zbiranju podatkov (tiskani vprašalnik za anketarja ali računalniško podprto osebno ali telefonsko anketiranje) je zaradi vključitve anketarjev, časa in finančnih sredstev najdražji način zbiranja podatkov. Anketarjeva navzočnost sicer omogoča, da anketar usmerja situacijo in da pojasni postavljena vprašanja, vendar pa lahko (nehote) vpliva tudi na odgovore in zaradi njegove navzočnosti lahko ostanemo tudi brez odgovora na občutljiva vprašanja. Zato je včasih boljše, da anketirana oseba sama odgovori na del vprašanj, če predvidimo, da je tak vpliv možen.

Zbiranje podatkov s telefonskim anketiranjem je precej cenejše od anketiranja na terenu, vendar pa je odzivnost pri tem načinu slabša. Čeprav je opremljenost s fiksnimi in mobilnimi telefoni velika, je težava, ker prebivalci Slovenije svoje telefonske številke v čedalje manjši meri objavljajo v telefonskem imeniku. Vpliv anketarja na odgovore anketirancev je bolje nadzorovan, saj anketarje v telefonskem studiu lahko sproti spremljamo pri njihovem delu.

Samoizpolnjevanje je predvsem stroškovno ugoden način izpolnjevanja vprašalnikov. Poleg tega je najprimernejši za zbiranje občutljivih podatkov. Izbrana enota lahko izpolni vprašalnik takrat, ko to sama želi. Ni primeren za zelo dolge, zapletene vprašalnike, ki zelo obremenijo respondenta (enoto, ki odgovarja na vprašalnik). Pri tej metodi nimamo nadzora nad izpolnjevanjem vprašalnika. Če gre za vprašalnik v tiskani obliki, tak način izpolnjevanja navadno poveča delo pri urejanju podatkov.

Vprašalniki v tiskani obliki so klasičen način zbiranja podatkov: opazovane enote dobijo vprašalnik na svoj poštni naslov, ga izpolnijo in nato pošljejo nazaj pošiljatelju. Tovrstno zbiranje podatkov povzroča poleg stroškov za tisk vprašalnika tudi poštno stroške pošiljatelju in prejemniku in stroške zaradi digitalizacije prejetih podatkov. V današnjih časih je ta način poročanja med podjetji in drugimi poslovnimi subjekti nepriljubljen, saj jim s prepisovanjem podatkov iz lastnih podatkovnih baz na tiskani vprašalnik jemlje čas in povzroča dodatne stroške (pri prepisovanju pa lahko nastanejo tudi napake).

Elektronski vprašalniki za samoizpolnjevanje so lahko spletni vprašalniki ali po meri narejeni elektronski vprašalniki (npr. preglednice v Excelu, v katere enote vpišejo zahtevane podatke in potem podatke v obliki xml oddajo prek spletne aplikacije). Spletni vprašalniki omogočajo uporabo logičnih kontrol in vključitev samodejnega usmerjanja po vprašalniku, dostop do šifrantov ipd. S tem lahko zmanjšamo količino napak ob vnosu in omogočimo hitrejše sestavljanje zbranih podatkov. Pri raziskovanjih, v katerih se statistično opazujejo podjetja in drugi poslovni subjekti, ki so sicer prešli na elektronski način sporočanja podatkov, uporabimo za zbiranje podatkov tiskane vprašalnike običajno kot dodatno možnost za sporočanje podatkov.

Kombinirani način zbiranja podatkov

Najpogostejši način zbiranja podatkov pri podjetjih in drugih poslovnih subjektih, če upoštevamo število enot, ki odgovorijo na vprašalnik, je elektronsko poročanje (spletni vprašalniki ali po meri narejeni elektronski vprašalniki), sledijo tiskani vprašalniki za samoizpolnjevanje in telefonsko anketiranje.

Pri zbiranju podatkov pri osebah in gospodinjstvih je prevladujoči način telefonsko anketiranje, sledi mu osebno terensko anketiranje, oba pa čedalje bolj dopolnjuje spletno izpolnjevanje vprašalnikov. Če uporabimo kombinirani način zbiranja podatkov, običajno začnemo z najcenejšo, končamo pa z najdražjo metodo. Zaradi lažjega obvladovanja procesa se priporoča uporaba le dveh zaporednih načinov. Če kombiniramo več načinov, se s tem zelo poveča delo pri procesih (priprava in testiranje vprašalnikov za različne načine; priprava adresarjev za pošiljanje opomnikov in prilagajanje pisem glede na način, združevanje podatkov; ugotavljanje vplivov uporabe različnih načinov na rezultate raziskovanja).

Način izvedbe zbiranja podatkov je odvisen tudi od vrste raziskovanja. Ločimo dve vrsti raziskovanj: longitudinalna in presečna. V longitudinalnih raziskovanjih spremljamo dinamiko opazovanega pojava; v takem raziskovanju zbiramo podatke za iste enote več kot enkrat v rednih intervalih. V presečnih raziskovanjih pa zberemo podatke samo za določeno časovno točko. Pri longitudinalnih raziskovanjih lahko zbiramo podatke vsakič na enak način, lahko pa so metode zbiranja podatkov pri posameznih izvedbah različne.

Med instrumente, potrebne za zbiranje podatkov, štejemo vprašalnik in druga gradiva, ki jih potrebujemo za izvedbo raziskovanja. Pri pripravi vprašalnika upoštevamo najnovije standarde in usmeritve. Osnutek vprašalnika pripravimo na podlagi pripravljenih spremenljivk in vprašanj, vanj vključimo logične kontrole in preskoke. Vprašanja v vprašalniku razdelimo v smiselne vsebinske sklope. Naš cilj je, da so vprašalniki čim krajši (da vsebujejo čim manj vprašanj), vprašanja pa zlahka (takoj) razumljiva, da čim manj obremenjujejo poročevalske enote. Za vsako vprašanje v vprašalniku mora obstajati utemeljen razlog. Le z ustrezno pripravljenim vprašalnikom lahko pridobimo od enot verodostojne (želene) podatke. Oblikovanje vprašalnika je ponavljajoči se (iterativni) postopek, pri katerem na podlagi izvedenih testiranj izboljšujemo osnovno različico vprašalnika. Pomembno je, da vprašalnik pregledamo tako z vsebinskega kot s tehničnega vidika. V pregledovanje vključimo osebe, zadolžene za posamezen proces/podproces, tj. tiste, na katerih delovni proces bi vplivale spremembe vprašalnika. Pri pripravi vprašalnika sodelujejo tudi druge osebe, npr. področni strokovnjaki (glede na temo raziskovanja), strokovnjaki za pripravo in testiranje vprašalnikov, anketarji, lektorji in predstavniki ciljne populacije. Vsak od njih lahko v veliki meri prispeva k izboljšanju vprašanj. Zaželeno je, da se pred uporabo novega vprašalnika na terenu izvede kognitivni test vprašalnika.

Pomembno je, da pred izvedbo raziskovanja pripravimo načrt komuniciranja z dajalci podatkov. Določimo načine obveščanja opazovanih in poročevalskih enot, rokovnik in vrste gradiv, ki jih bomo pripravili. Gradiva na eni strani vzpostavijo uvodni stik s poročevalsko enoto pred samo izvedbo raziskovanja, na drugi strani pa z njimi ohranjamo stik tudi v času, ko poteka zbiranje podatkov. S temi gradivi dajalce podatkov informiramo in jih skušamo motivirati za sodelovanje v posameznem raziskovanju. Pojasnimo namen, ozadje in vsebino raziskovanja, podamo informacijo o njihovi vključitvi v raziskovanje, informacije glede obveznosti poročanja, informacije o uporabi administrativnih in drugih virov v raziskovanju, o varovanju statistične zaupnosti in varovanju podatkov, informacije glede posredovanja podatkov (kam, komu, kako in do kdaj). Pomembno je, da jih ne pozabimo obvestiti, kje lahko dobijo morebitne dodatne informacije (povezava na spletni strani, podatki za stik).

Če v procesu zbiranja podatkov sodelujemo z zunanjimi izvajalci, moramo načrtovati postopke za njihovo pridobitev (razpis, objava na spletni strani), ter pripravo pogodb, dogovorov in drugih dokumentov, s katerimi se formalno dogovorimo glede organizacijskih in tehničnih zadev. Če podatke zbirajo anketarji, je treba pripraviti načrt za njihovo usposabljanje.

Smernice za zagotavljanje kakovosti

- Izbiro načina izvedbe raziskovanja je treba prilagoditi vsebini. Način zbiranja podatkov naj se prilagaja ciljem, ciljni populaciji in metodologiji raziskovanja, in ne obratno.
- Metodologija zbiranja podatkov, instrumenti za zbiranje podatkov in načrt komuniciranja s poročevalskimi enotami morajo biti skrbno načrtovani.
- Instrumente in metode za zbiranje podatkov je treba načrtovati tako, da bosta sta zagotovljena čim večja kakovost podatkov in nadzor nad merskimi napakami ob čim manjši obremenitvi poročevalskih enot in ob stroškovni učinkovitosti.
- Izogibamo se kombiniranju več kot dveh načinov zbiranja podatkov, da bomo lažje obvladovali procese priprave vprašalnika, zbiranja in obdelovanja podatkov.
- Pri osebnem terenskem anketiranju v čim večji meri uporabimo računalniško podprto anketiranje; ročno zapisujemo podatke na papir le, kadar je to smiselno (bolj racionalno).
- Kadar podatke zberemo s tiskanimi vprašalniki, izvedimo digitalizacijo podatkov v čim večji meri s skeniranjem ali optičnim branjem.
- Vprašanja morajo biti prilagojena načinu zbiranja podatkov in značilnostim poročevalskih enot.
- Za zagotavljanje čim večje primerljivosti rezultatov raziskovanj je pomembno, da so vprašanja oblikovana na enak način. Pri tem moramo upoštevati tudi dejstvo, da so nekatera vprašanja »občutljiva« za to, kaj smo vprašali pred tem (vsebina prejšnjega vprašanja lahko vpliva na odgovor naslednjega).
- Pri vprašalnikih za samoizpolnjevanje morajo biti vprašanja kratka in jasna, vprašalnik ne sme biti obsežen, vsebovati mora jasna navodila in primere.
- Gradiva za komuniciranje s poročevalskimi enotami morajo biti razumljiva, kratka, jasna in brez nepotrebnih informacij.

2.4 PRIPRAVA METODOLOGIJE IZBORA ENOT OPAZOVANJA

Potem ko določimo ciljno populacijo, moramo pridobiti seznam enot te populacije skupaj s čim več njihovimi lastnostmi. Seznam teh enot imenujemo vzorčni okvir. Vzorčni okvir je v praksi približek ciljne populacije. Ta približek je posledica bodisi zavestne odločitve bodisi dejstva, da viri, s katerimi sestavimo vzorčni okvir, niso popolni. Spremenljivke vzorčnega okvira, ki opisujejo lastnosti enot, imenujemo pomožne spremenljivke.

Pri pripravi metodologije za izbiro enot, ki bodo predmet opazovanja, upoštevamo predvideni način zbiranja podatkov. Želene podatke lahko pridobimo iz različnih administrativnih in drugih virov, lahko pa jih izpeljemo ali modeliramo iz obstoječih statističnih virov ali s posrednim ali neposrednim stikom z enotami opazovane populacije.

Ne glede na način zbiranja podatkov lahko raziskovanje izvedemo kot vzorčno raziskovanje ali kot popis. Če se odločimo za vzorčno raziskovanje, potem ga lahko izvedemo na dva načina: z neverjetnostnim vzorčenjem ali z verjetnostnim vzorčenjem. Postopek izbora enot opazovanja pri neverjetnostnem vzorcu je navadno zelo enostaven, poceni in ne zahteva veliko časa, hkrati pa je tak izbor subjektiven, in je posledično treba predpostaviti reprezentativnost vzorca glede proučevanih spremenljivk in zahtevanih statistik, da je mogoče na podlagi zbranih podatkov iz takega vzorca sklepati o lastnostih celotne populacije. Tako se na primer pri raziskovanjih podjetij in drugih poslovnih subjektov pogostokrat odločimo za tako imenovani zajem s pragom. Pri tem postopku izberemo vzorec opazovanih enot (tj. podjetij ali njihovih delov) tako, da enote, ki imajo število zaposlenih ali prihodek pod določenim pragom, izvzamemo iz vzorca, druge pa v vzorec vključimo z gotovostjo. Včasih pa namesto absolutnega števila (zaposlenih ali prihodka) za prag določimo ciljni delež enot glede na izbrano spremenljivko, na primer tako, da vzamemo toliko največjih enot, da kumulativno število zaposlenih preseže določeni ciljni delež vseh zaposlenih.

V nasprotju z neverjetnostnim vzorčenjem je uporaba verjetnostnega vzorčenja zahtevnejša, saj so enote vzorca izbrane z neko neničelno verjetnostjo, ki jo določimo pred izborom, zato sam postopek izbora zahteva več časa. Ker pa za verjetnostnim vzorčnim načrtom stoji celotna teorija vzorčenja, lahko pri takem načinu izbora precej bolj zanesljivo sklepamo o lastnostih celotne populacije in celo izračunamo vzorčne napake, to je ocene tega, koliko se ocena populacijskega parametra razlikuje od populacijskega parametra. Postopek izbora

enot pri verjetnostnem vzorcu je mogoče izvesti na mnogo načinov. Način izbora verjetnostnega vzorca imenujemo vzorčni načrt. Izbira načina je odvisna:

- od pomožnih spremenljivk vzorčnega okvira;
- od »variabilnosti« enot populacije (tu po navadi mislimo na variabilnost vrednosti proučevane spremenljivke);
- od statistik in domen, na katerih statistike objavljamo;
- od stroškov raziskovanja.

Del vzorčnega načrta je tudi določitev velikosti vzorca; velikost vzorca namreč neposredno vpliva na stroške raziskovanja. Cilj je, da pri čim manjši velikosti vzorca pridobimo čim boljše ocene populacijskih parametrov. Če imamo v vzorčnem okviru dovolj podatkov, lahko velikost vzorca izračunamo (navadno glede na zahtevane natančnosti statistik), sicer pa jo določimo na podlagi izkušenj (npr. drugih statističnih raziskovanj).

Najpogosteje uporabljeni vzorčni načrti so:

- enostavno slučajno vzorčenje
- sistematično vzorčenje
- vzorčenje, sorazmerno z velikostjo
- stratificirano vzorčenje
- vzorčenje v skupinah
- dvostopenjsko vzorčenje.

Najpomembnejši viri za sestavo vzorčnega okvira so naslednji: za vzorčenje oseb in gospodinjstev Centralni register prebivalstva (CRP), za vzorčenje podjetij Statistični poslovni register Slovenije (SPRS), za vzorčenje kmetijskih gospodarstev Statistični register kmetijskih gospodarstev (SRKG). Drugi pomembnejši registri ali baze podatkov, ki jih je treba v čim večji meri upoštevati, so: Register nepremičnin (REN), Register prostorskih enot (RPE), Statistični register delovno aktivnega prebivalstva (SRDAP), davčni podatki podjetij, zaključni računi, baza gospodinjstev, dohodninski podatki.

CRP je primeren vir za izbor oseb; za izbor gospodinjstev pa ga uporabljamo posredno, saj nas izbrana oseba (njen naslov) privede do gospodinjstva. S pomočjo terenskih podatkov o številu članov gospodinjstva nato ocenimo verjetnosti izbora posameznega gospodinjstva v vzorec.

Pri vzorčenju podjetij in drugih poslovnih subjektov zajamemo največja podjetja vedno v vzorec z gotovostjo, preostala pa z verjetnostnim vzorcem. Merila za določitev največjih podjetij so število zaposlenih, prihodek ali kombinacija obojega. Meje za določitev največjih podjetij precej variirajo med različnimi raziskovanji.

Smernice za zagotavljanje kakovosti

- Pri določitvi vzorčnega okvira je treba oceniti, kateri viri (registri, podatkovne baze ...) so na voljo in kako jih lahko povežemo med sabo.
- Če je izvedljivo, je treba oceniti kakovost vzorčnega okvira takoj, ko je ta sestavljen. Če ocenimo, da sestava okvira ni zadovoljive kakovosti, je treba pri sestavljanju vprašalnika vključiti še dodatna vprašanja, s katerimi pozneje lažje ocenimo kakovost vzorčnega okvira; te ugotovitve upoštevamo pri sestavi vzorčnega okvira za naslednje referenčno obdobje.
- Če je vzorčni okvir sestavljen na podlagi načina anketiranja (npr. osebe iz telefonskega imenika namesto vseh oseb iz CRP), potem je treba podati oceno vpliva podpokritja na kakovost rezultatov.
- Pri sestavi okvira je treba posebno pozornost nameniti pomožnim spremenljivkam, saj nam dostop do uporabnih pomožnih spremenljivk omogoča izračun velikosti vzorca, izvedbo kompleksnejšega vzorčnega načrta in določitev natančnejših ocen (statistik).
- Če uporabljamo neverjetnostno vzorčenje, je treba še pred izvedbo raziskovanja oceniti pristranskost statistik, ki nastane zaradi tega.
- Prednost je treba dati čim enostavnejšim vzorčnim načrtom.
- Odločitev za kompleksnejše vzorčne načrte moramo dobro pretehtati. Uporaba večstopenjskih vzorčnih načrtov na primer lahko zniža stroške raziskovanja, vendar zmanjša natančnost ocen.
- Pri določitvi velikosti vzorca je treba upoštevati predvideno stopnjo odgovora in tudi rezultate analize neodgovora iz prejšnjega referenčnega obdobja (če so podatki na voljo). Oceniti je treba, kolikšna stopnja odgovora še zagotavlja zadostno kakovost rezultatov.

2.5 PRIPRAVA METODOLOGIJE OBDELAVE PODATKOV

Z obdelavo podatkov tu označujemo vse postopke, ki jih po končani fazi zbiranja ali prevzema podatkov uporabimo za to, da bodo končni statistični rezultati čim točneje odražali lastnosti opazovane populacije. V fazi načrtovanja raziskovanja je predvsem treba v čim večji mogoči meri predvideti, kateri postopki so za izvedbo načrtovanega raziskovanja smiselni, kateri metodološki pristopi naj se v teh postopkih uporabijo in katera programska orodja bi bila za izvedbo najprimernejša.

Različna raziskovanja vsebujejo v svoji izvedbi različne postopke za statistično obdelavo. Sestavni del statističnega raziskovanja so tudi postopki za statistično obdelavo podatkov; ti so različni, kot so različna tudi raziskovanja. Najpogostejši so naslednji:

- **Obravnavanje neodgovora in uteževanje.** Pri izvedbi raziskovanja se – razen pri nekaterih raziskovanjih z malo opazovanimi enotami – neizogibno srečamo s problemom neodgovora, tj. s problemom (delno ali popolnoma) manjkajočih podatkov pri delu opazovanih enot. Postopke za zmanjšanje deleža neodgovora je sicer treba vključiti v aktivnosti, povezane z zbiranjem podatkov, v fazi načrtovanja statistične obdelave pa je treba predvideti, katere postopke bomo po že končanem zbiranju podatkov uporabili za to, da bomo v čim večji meri zmanjšali vpliv neodgovora na končne statistične rezultate (in s tem zmanjšali pristranskost ocen). Problem neodgovora se lahko pojavlja tudi pri administrativnih virih podatkov. Eden najpogostejših načinov prilagajanja zaradi neodgovora je uteževanje podatkov.
- **Povezovanje (integracija) različnih podatkovnih virov.** V raziskovanjih, v katerih se uporablja več različnih podatkovnih virov, je treba te vire povezati. Pri združevanju virov lahko pridobimo vrednosti nekaterih spremenljivk iz več različnih virov; pri tem moramo določiti zaporedje prevzemanja vrednosti iz posameznega vira.
- **Urejanje podatkov.** Izraz urejanje podatkov označuje vse postopke za iskanje in odpravljanje napak v podatkih. Ob načrtovanju teh postopkov je treba izbirati take, ki bodo v čim večji možni meri zmanjševali tako časovno kot stroškovno zahtevnost. »Ročno« urejanje podatkov skušamo nadgraditi z vpeljavo sodobnejših pristopov, kot so selektivno urejanje, polavtomatsko in avtomatsko urejanje podatkov.
- **Agregacija in tabelacija.** V tem delu procesa dobimo iz primerno statistično »obdelanih« podatkov na mikroravni s postopkom, imenovanim agregacija, statistične agregate (statistike), ki so končni izdelek statističnega procesa. V fazi načrtovanja je treba predvsem predvideti, katere statistike se bodo izračunavale, ter natančno in jasno zapisati pravila za njihov izračun – tako v vsebinski kot v formalni obliki (matematične formule). Predvideti je treba tudi raven podrobnosti tabelacije rezultatov. Pri tem je treba upoštevati (morebitne) zahteve iz mednarodnih uredb, zahteve domačih uporabnikov in tudi omejitve, ki jih v zvezi z objavljanjem na malih področjih (domenah) določajo pravila za zaščito podatkov in pravila o objavljanju oz. neobjavljanju rezultatov s premajhno stopnjo natančnosti.

Včasih je treba v postopku statistične obdelave podatkov načrtovati še naslednje korake, če to zahtevajo narava raziskovanja ali pa potrebe uporabnikov:

- Če gre za vzorčno raziskovanje, je treba predvideti postopke za oceno vzorčne napake. Predvsem se je treba odločiti, katerega od pristopov in katero programsko orodje bomo pri tem uporabili.
- Predvsem pri finančnih podatkih je običajno treba uporabniku posredovati tudi tako imenovane deflacirane statistične rezultate. Deflacija je postopek, s katerim predvsem iz rezultatov, ki prikazujejo spreminjanje kakega pojavnosti (npr. indeks povprečne plače), odstranimo vpliv spreminjanja cen. V fazi načrtovanja je treba predvideti, katerega od dveh osnovnih pristopov (deflacioniranje na mikro- ali na makroravni) bomo uporabili, kakšna bo metodologija izračuna deflatorjev, katere osnovne vire bomo pri tem potrebovali in kako bomo postopek umestili v celoten statističen proces.
- Če gre za periodično raziskovanje in rezultate raziskovanja lahko obravnavamo kot časovne vrste, je treba predvideti tudi to, ali bo treba podatke desezonirati. Uporaba teh postopkov je sicer odvisna od tega, ali sezonski vpliv in vpliv števila delovnih dni na rezultate obstajata; to je sicer mogoče analitično ugotoviti šele po določenih ponovitvah izvedbe, običajno pa je tak vpliv mogoče predvideti že na podlagi poznavanja vsebine.

Smernice za zagotavljanje kakovosti

- Pri načrtovanju statističnega procesa je treba v največji mogoči meri uporabiti sodobne statistične metode ter metodološke rešitve, ki so bile kot dobre prakse vpeljane v drugih statističnih organizacijah ali ki jih na podlagi sodobnih teoretskih dognanj vpeljemo in uveljavimo (implementiramo) sami.
- Pri načrtovanju orodij za izvedbo statistične obdelave je treba ugotoviti, ali že obstajajo splošna programska orodja, ki bi jih lahko uporabili, ali pa bo treba razviti rešitve »po meri« raziskovanja. Če je treba vpeljati nove

programske rešitve, je treba izbrati najprimernejše programsko orodje. Pri tem naj se upoštevajo tako funkcionalne zmožnosti posameznih programskih orodij kot tudi standardna praksa na uradu.

- Pri obravnavi neodgovora enote se je treba odločiti, ali bomo uporabili postopek uteževanja podatkov ali postopek vstavljanja podatkov. Pri postopkih urejanja podatkov je treba v čim večji mogoči meri uporabiti metode selektivnega, polavtomatskega in avtomatskega urejanja podatkov, saj te lahko bistveno zmanjšajo stroške in obremenitev poročevalskih enot zaradi ponovnega vzpostavljanja stika z njimi.
- Pri določitvi podrobnosti ravni objavljanja podatkov se, če je le mogoče (in če ni predpisano drugače), izogibajmo prepodrobni ravni objavljanja, da se v čim večji meri izognemo prevelikemu številu praznih, zakritih ali zaščiteneh celic.
- Predvideti moramo izračun kazalnikov kakovosti, s katerimi bomo ocenjevali kakovost rezultatov, pridobljenih v raziskovanju, in sicer tako, da se ti v kar največji meri izračunajo med procesom in da so na voljo za vpogled.

2.6 NAČRTOVANJE IN TESTIRANJE IZVEDBE RAZISKOVANJA

Pred začetkom izvajanja raziskovanja je treba določiti proces izvedbe in tok podatkov od zbiranja podatkov do objave rezultatov. Obravnavati je treba vse podprocese in njihove posamezne korake. Zagotoviti je treba, da je tok podatkov med posameznimi koraki čim bolj tekoč, brez manjkajočih korakov ali podvajanja. Splošno pravilo je, da se v celotnem procesu izvedbe raziskovanja v čim večji meri uporabljajo standardizirani postopki, metode, programska orodja in že izdelane programske rešitve (storitve, sistemi, podatkovne baze). Obstoječe rešitve je treba proučiti, da se ugotovi, ali ustrezajo značilnostim obravnavanega raziskovanja oz. v kolikšni meri, in da se prepoznajo morebitne posebnosti raziskovanja, za katere programska orodja ali metode morda še niso razviti. V takih primerih je treba te dodatno razviti.

Pomemben element pri načrtovanju in testiranju izvedbe raziskovanja je jasna določitev oseb, izvajalcev posameznih vlog v izvedbi posameznih korakov in rokov, do katerih morajo biti orodja in metode, potrebni za izvedbo raziskovanja, razviti.

Pred začetkom izvajanja novega raziskovanja (še posebej, če gre za novo in doslej še ne raziskano vsebino ali ciljno populacijo) ali pri večji tehnični ali metodološki reviziji obstoječega raziskovanja je treba načrtovane postopke ustrezno testirati. Testiramo lahko izvedbo vseh posameznih korakov od zbiranja do analize podatkov. Priporočljivo je izvesti pilotno raziskovanje, s pomočjo katerega lahko dokončno določimo nekatere metodološke vidike raziskovanja (npr. potrebno velikost vzorca). S pilotnim raziskovanjem tudi testiramo postavljeni tok podatkov. Testirati je treba orodja za zbiranje, obdelavo in analizo podatkov in s tem zagotoviti, da je proces čim bolj optimalen. Na podlagi izsledkov testiranja ali izvedbe pilotnega raziskovanja se potrdi proces izvedbe raziskovanja.

Vse opisane postopke, ki jih je treba opraviti v fazi načrtovanja in testiranja izvedbe raziskovanja, izvedemo, tudi če podatke v celoti pridobimo iz administrativnih virov.

Smernice za zagotavljanje kakovosti

- Pred izvedbo raziskovanja je treba določiti tok podatkov (načrt izvedbe raziskovanja) od zbiranja podatkov do objave rezultatov. Pomembno je, da pri tem sodelujejo vsi izvajalci posameznih vlog, ki bodo vključeni v izvedbo posameznih korakov.
- Da se lahko izdelajo in testirajo vsi potrebni postopki, je treba načrt izvedbe načrtovati dovolj zgodaj (pravočasno).
- V čim večji meri je treba uporabiti metode in postopke, ki so že razviti in vpeljani. Pri tem moramo biti pozorni na to, ali ustrezajo značilnostim konkretnega raziskovanja.
- Pomembno je, da pred izvedbo raziskovanja načrtovani proces izvedbe analiziramo z vidika manjkajočih korakov in morebitnega podvajanja. Če je treba izdelati novo programsko rešitev, je treba izbrati najprimernejše orodje. Pri tem upoštevamo funkcionalne zmožnosti posameznih orodij in možnost njihove združitve z drugimi rešitvami.
- Pomembno je, da so vloge izvajalcev jasno opredeljene. Vsi, ki bodo v procesu izvedbe konkretnega raziskovanja sodelovali, morajo pravočasno vedeti, kaj morajo opraviti in do kdaj.
- Nestandardnim rešitvam se izogibajmo v največji možni meri. Uvajanje takih rešitev v proces namreč pomeni dodatno porabo človeških in finančnih virov ter težave pri tehničnem in kadrovskem vzdrževanju takšnih rešitev.
- Cilje pilotnega raziskovanja je treba jasno opredeliti.

- V pilotnem raziskovanju lahko testiramo različne metodološke vidike: način zbiranja podatkov, vpliv komunikacijskih gradiv na respondenta (tj. na enoto, ki odgovarja na vprašalnik), na stopnjo odgovora enote in/ali spremenljivke, na formulacijo vprašanj itd.
- Ciljem pilotnega raziskovanja je treba prilagoditi vzorčni načrt, tako da lahko testiramo razlike v odgovorih pri posameznih podskupinah, učinek različnih komunikacijskih gradiv, različne oblike vprašanj itd.
- Pilotno raziskovanje je namenjeno tudi testiranju vprašalnika, vendar morata biti tehnično testiranje vprašalnika in kognitivni testi izvedena že pred pilotnim raziskovanjem.
- Pred izvedbo raziskovanja se testirajo postopki za zbiranje, obdelavo in analizo podatkov.
- Izsledke pilotnega raziskovanja je treba uporabiti pri pripravi glavnega raziskovanja.

3 IZBOR ENOT OPAZOVANJA

Ker je izvedba raziskovanja na celotni populaciji draga, zahteva veliko časa in je zelo obremenjujoča tako za tiste, ki raziskovanje izvajajo, kot za tiste, ki na vprašalnike odgovarjajo, večino raziskovanj izvajamo na podlagi vzorcev izbranih enot populacije. To pomeni, da na podlagi pridobljenih podatkov le o delu populacije sklepamo o lastnostih celotne populacije, ki nas pri posameznem raziskovanju zanima. Še preden se odločimo za vzorčni načrt, s katerim bomo določili način izbora vzorca, moramo teoretično opredeljeno populacijo in čim več osnovnih lastnosti o njej dejansko pridobiti. Dejanski zapis enot populacije skupaj z nekaterimi lastnostmi, ki so pomembne tako za izbor vzorca kot za izračun ocen populacijskih parametrov, imenujemo vzorčni okvir. Pred pripravo vzorčnega okvira je treba najprej pripraviti vse podatkovne vire (registre, podatke iz drugih raziskovanj in podobno), s pomočjo katerih bomo sestavili vzorčni okvir. V praksi je vzorčni okvir bolj ali manj natančen približek ciljne populacije (in nekaterih njenih lastnosti), njegova kakovost pa v veliki meri vpliva na kakovost končnih rezultatov raziskovanja. Zato se je treba pri pripravi podatkovnih virov, potrebnih za sestavo vzorčnega okvira, kar najbolj potruditi, da dosežemo najvišjo kakovost vzorčnega okvira, ki je mogoča. Ko je vzorčni okvir določen, se določi vzorčni načrt, s katerim izberemo vzorec enot oz. vzorčne enote, na katerih bomo dano raziskovanje izvedli. Vzorec enot je lahko verjetnosten, neverjetnosten ali kombinacija obeh. Ko je vzorec enot določen, se sestavi adresar enot, izbranih v vzorec. Ta vsebuje seznam naslovov enot iz vzorca (in drugih za vzpostavitev stika potrebnih podatkov) in naslove poročevalskih enot (tudi enot, ki sporočijo podatke za enoto (v imenu enote) v vzorcu) ter morebitni seznam ključnih poročevalskih enot, to je tistih, pri katerih si še posebej prizadevamo, da nam sporočijo zahtevane oz. zelene podatke.

3.1 PRIPRAVA PODATKOVNIH VIROV ZA IZGRADNJO VZORČNEGA OKVIRA

Vzorčni okvir je seznam enot, iz katerega izbiramo enote, ki bodo sestavljale vzorec. Glede na opredeljeno ciljno populacijo in izbrano referenčno obdobje opazovanja je treba določiti vse podatkovne vire, ki jih bomo uporabili pri pripravi vzorčnega okvira. Cilj je uporabiti vse razpoložljive vire, ki so nam na voljo, da je vzorčni okvir čim bolj skladen s teoretično opredeljeno ciljno populacijo. Pred uporabo virov je treba preveriti njihovo kakovost. Če pri določenih virih ugotovimo pomanjkljivosti, ki bi lahko znatno vplivale na izvedbo raziskovanja in kakovost rezultatov, je treba izvesti aktivnosti, s katerimi se te odpravijo. Če teh pomanjkljivosti ni mogoče odpraviti, teh virov pri sestavi vzorčnega okvira ne upoštevamo.

Pomembno je, da razumemo razliko med ciljno populacijo in vzorčnim okvirom. Ciljna populacija raziskovanja je množica enot, katerih določene lastnosti opazujemo in na katero se nanašajo vsi rezultati tega raziskovanja. Vzorčni okvir je dejanski seznam enot ciljne populacije, ki ga imamo v času priprave raziskovanja na voljo in ga uporabimo za izbor enot ciljne populacije, ne glede na to, ali je raziskovanje vzorčno ali ne.

Preden določimo vzorčni okvir, moramo natančno definirati ciljno populacijo (npr. osebe, gospodinjstva, stanovanja, kmetijska gospodarstva, podjetja):

- glede na lastnosti, ki nas zanimajo, določimo, kdo ali kaj je ciljna populacija in katerim pogojem mora enota zadoščati, da je del populacije;
- določiti moramo referenčno obdobje (čas), v katerem nas zanimajo lastnosti ciljne populacije.

Primer: Naša ciljna populacija so osebe, ki živijo v zasebnih gospodinjstvih. Na določeni datum so bile prebivalci Slovenije in so bile na ta datum stare vsaj 18 let. Pri tem moramo upoštevati definicijo prebivalstva Slovenije; definirati je treba pogoje, ki določajo, kdaj je oseba član zasebnega gospodinjstva (kaj je zasebno gospodinjstvo; kako ravnati, če oseba živi nekaj časa v enem, nekaj časa v drugem gospodinjstvu; h kateremu gospodinjstvu sodi taka oseba). S temi pogoji definiramo ciljno populacijo in posledično tudi pogoje, ki jih mora posamezna enota izpolnjevati, da je ustrezna za raziskovanje.

V raziskovanju razlikujemo:

- vzorčno enoto (enota iz vzorčnega okvira, ki je bila izbrana v vzorec),
- enoto opazovanja oz. opazovano enoto (enota, o kateri zbiramo podatke in je torej del ciljne populacije) in
- poročevalsko enoto (enota, ki nam sporoča podatke, ki jih zbiramo, oz. enoto, ki je sporočitelj podatkov).

V nekaterih raziskovanjih so vse tri lastnosti združene v eni enoti, pogosto pa to ne velja.

Primer: Vzorčna enota je izbrana oseba iz Centralnega registra prebivalstva (CRP) z natančno določenim naslovom; ta naslov nam določa enoto opazovanja, to je stanovanje, ki je del naše ciljne populacije. Ni nujno, da izbrana oseba sploh živi v tem stanovanju. Podatke pa nam sporoči oseba, ki živi v tem stanovanju. Ta oseba je poročevalska enota (sporočitelj podatkov).

Vse to moramo upoštevati pri sestavi vzorčnega okvira; to torej pomeni, da vzorčnega okvira pogosto ne sestavljajo neposredno enote, ki jih opazujemo, ampak enote, ki nas pripeljejo k enoti opazovanja.

Ko vemo, kaj je naša ciljna populacija in kako lahko dostopamo do enot te populacije, se lotimo določitve vzorčnega okvira. Glede na ciljno populacijo uporabimo različne vire, in sicer v časovnih točkah ali iz obdobj, ki so čim bližje referenčnemu obdobju raziskovanja ali času zbiranja podatkov.

Če uporabljeni podatkovni vir pokriva referenčno obdobje raziskovanja, potem populacijo v referenčnem obdobju zajamemo točneje, kot če takega vira nimamo na razpolago. Vendar imamo zato v vzorčnem okviru več neustreznih enot (to so enote, ki v času zbiranja podatkov niso ustrezne, v referenčnem obdobju raziskovanja pa so še bile) ali pa so se nekatere lastnosti enot spremenile (če npr. podjetje spremeni dejavnost). Če pa je podatkovni vir novejši glede na čas zbiranja podatkov, potem po eni strani zmanjšamo število neustreznih enot do najmanjše stopnje, po drugi strani pa imajo enote v okviru nekatere lastnosti, ki veljajo zdaj, ne veljajo pa za referenčno obdobje opazovanja, zato jih lahko napačno združujemo v skupine ali celo stratumе.

Pri zbiranju virov moramo poskrbeti tudi za klasifikacijske spremenljivke, ki določajo stratumе in ki jih potrebujemo pri določanju vzorčnega načrta in/ali pri objavi rezultatov. Na primer: če je poraba energije v gospodinjstvih, ki živijo v blokih, drugačna od porabe v tistih, ki živijo v hišah, jih moramo s stratumi razdeliti glede na to, v katerem tipu stavbe prebivajo. Pri raziskovanjih podjetij in drugih poslovnih subjektov so klasifikacijske spremenljivke velikostni razredi podjetij in njihova dejavnost.

Običajno izberemo en glavni podatkovni vir za vzorčni okvir, potem pa izboljšamo kakovost okvira z dodatnimi viri. Za glavni vir si izberemo vir, ki nam omogoča, da z njim čim bolje zajamemo ciljno populacijo. Z dodatnimi podatkovnimi viri lahko dopolnimo glavni vir z dodatnimi podatki o enotah opazovanja.

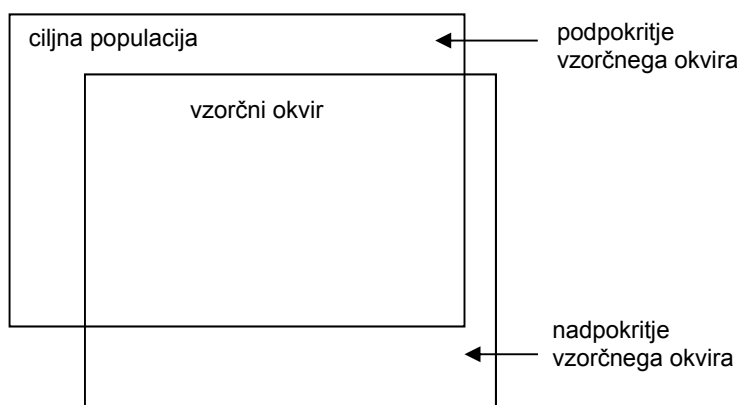
Pri raziskovanjih podjetij in drugih poslovnih subjektov je praviloma glavni podatkovni vir Poslovni register Slovenije (PRS); v njem dobimo seznam vseh podjetij. Dodatni podatkovni viri pa so lahko: zaključni računi, davčni podatki podjetij, Statistični register delovno aktivnega prebivalstva (SRDAP); ti viri nam na primer omogočajo določitev velikostnih razredov podjetij.

Pri raziskovanjih, v katerih zbiramo podatke o osebah in gospodinjstvih, je za terenske ankete glavni podatkovni vir CRP; ta vir dopolnimo z Registrom prostorskih enot (RPE), da dopolnimo geografsko lokacijo opazovanih enot, lahko tudi z REN, če nas zanimajo stanovanja gospodinjstev; za telefonske ankete je ravno tako glavni podatkovni vir CRP, ki ga dopolnimo z RPE, kot nujen podatkovni vir pa dodamo še telefonski imenik.

Včasih en sam podatkovni vir ne zadošča, zato moramo zgraditi vzorčni okvir iz več različnih virov (administrativni viri, popis, druga raziskovanja). To se zgodi zlasti takrat, kadar različni podatkovni viri pokrivajo različna obdobja in različne dele populacije. Tu je zelo pomembno, da imamo v vseh podatkovnih virih enolično identifikacijo enot. Zato so za nas na splošno ustrezni le podatkovni viri, v katerih lahko to identifikacijo določimo. Tudi v tem primeru pa se en podatkovni vir določi kot osnovni vir.

Ciljna populacija in vzorčni okvir se le redko ujemata v vseh enotah. Obstajajo enote, ki so del ciljne populacije, a jih ni v seznamu enot vzorčnega okvira, prav tako pa so v vzorčnem okviru enote, ki v resnici niso del ciljne populacije. Oba primera povzročata napake zaradi neustreznega pokritja, ki v veliki meri vplivajo na kakovost končnih rezultatov raziskovanja.

Prikaz podpokritja in nadpokritja vzorčnega okvira



Smernice za zagotavljanje kakovosti

- Referenčna obdobja, na katera se nanašajo podatki v virih, ki jih bomo uporabili pri gradnji vzorčnega okvira, naj bodo čim bližje referenčnemu obdobju raziskovanja. Če so referenčna obdobja v različnih virih različna, je treba razlike dokumentirati.
- Treba je preveriti, ali so spremenljivke, ki nastopajo v več različnih virih, skladne glede na metodološke opredelitve. Če se ugotovi, da se po teh opredelitvah razlikujejo oz. ne ujemajo, je treba te razlike dokumentirati.
- Vzorčni okvir mora vsebovati naslednje podatke: enolični identifikator enote; kontaktne podatke; klasifikacijske spremenljivke. Koristno je, da vzorčni okvir vsebuje tudi datum stanja okvira in povezovalne spremenljivke za povezovanje z drugimi podatkovnimi viri ter pomožne spremenljivke, ki jih kasneje uporabimo pri kalibraciji, če so v času priprave vzorčnega okvira že na voljo.
- Poskrbeti je treba, da so vsi podatkovni viri pred združevanjem prečiščeni z vidika formalnih kontrol (npr. veljavnost šifrantov, ustrezen obseg vrednosti spremenljivk, problem podvojenih enot).
- Podatkovne vire, ki smo jih uporabili za gradnjo okvira, je treba ustrezno dokumentirati.
- Zapise, ki jih ne bomo mogli uporabiti za gradnjo okvira, je treba ustrezno dokumentirati.
- Ker želimo, da bi bile napake zaradi neustreznega pokritja čim manjše, je treba pri pripravi vzorčnega okvira uporabiti prav vse podatkovne vire, ki so na voljo, saj vsak prispeva k temu, da je vzorčni okvir čim bolj skladen s ciljno populacijo. Morebitna neskladja med vzorčnim okvirom in ciljno populacijo je treba ustrezno dokumentirati.

3.2 PRIPRAVA VZORČNEGA OKVIRA

Vzorčni okvir za konkretno raziskovanje se izdelava na podlagi opredelitve ciljne populacije, razpoložljivih podatkovnih virov ter obstoječe programske opreme. Naš končni cilj je pripraviti enotno tabelo podatkov, ki bo zajemala seznam enot, po lastnostih čim bližje teoretsko opredeljeni ciljni populaciji, in v kateri bodo vsaki enoti določene vrednosti tistih spremenljivk, ki jih potrebujemo v poznejšem postopku izbora enot opazovanja. Ključni korak pri pripravi vzorčnega okvira je določitev postopka za izbor enot, vključenih v vzorčni okvir. Izhodišče vzorčnega okvira je praviloma eden izmed osnovnih registrov v izbrani časovni točki. Seznam enot, ki je določen z izbrano časovno točko registra, lahko nato dopolnimo z uporabo dodatnih administrativnih in statističnih virov.

Za pripravo kakovostnega vzorčnega okvirja je velikokrat treba uporabiti več podatkovnih virov. Potem ko se ti viri identificirajo in pripravijo v primerni računalniški obliki, jih je treba z ustrezno metodologijo združiti ter pripraviti spremenljivke, ki določajo ključne lastnosti enot v vzorčnem okviru.

Ključni korak pri pripravi vzorčnega okvira je določitev postopka za izbor enot, ki so vključene v vzorčni okvir. Izhodišče vzorčnega okvira je praviloma eden izmed osnovnih registrov v izbrani časovni točki (le izjemoma je drugače). Seznam enot, ki je določen z izbrano časovno točko registra, nato skušamo »izboljšati« z uporabo dodatnih administrativnih in statističnih virov. Prvenstveni cilj je zaznati in nato iz seznama izločiti enote, ki sicer so v registru, vendar v resnici niso del ciljne populacije. Razlog za obstoj takih neustreznih enot v registru so lahko pomanjkljivosti v postopkih osveževanja registra ali pa administrativna narava registra. Včasih pa vključimo v vzorčni okvir tudi enote, ki jih v registru ni. Predvsem pri raziskovanjih, v katerih se opazujejo

podjetja in drugi poslovni subjekti so to t. i. fiktivne enote, ki jih zaznamo z uporabo dodatnih podatkovnih virov. Razlog za tako »nepokritost« običajno izvira že iz opredelitve populacije, na katero se uporabljeni register nanaša, saj ta ne zadošča popolnoma specifičnim potrebam raziskovanja.

Če želimo, da bo vzorčni okvir služil svojemu namenu, je treba zagotoviti, da bo ta vseboval za vsako svojo enoto vrednosti spremenljivk, ki jih bomo v poznejših postopkih potrebovali. Pri vzorčnem raziskovanju so to predvsem stratifikacijske spremenljivke, spremenljivke, ki določajo skupino prve stopnje (pri dvostopenjskem vzorčnem načrtu), ali druge spremenljivke, ki jih potrebujemo pri uporabljenem vzorčnem načrtu ali ki definirajo domene, na katerih želimo objaviti rezultate. Tudi pri uporabi neverjetnostnega vzorčnega načrta (npr. pri zajemu s pragom) običajno potrebujemo vrednosti nekaterih spremenljivk (npr. prihodek podjetja ali število zaposlenih), ki so ključne za izvedbo izbora. Zagotoviti moramo tudi, da je za vsako enoto določen enolični identifikator, ki se uporablja pozneje v postopku izvajanja raziskovanja.

Če verjetnostni vzorec izbiramo z metodo stalnih (enakomerno porazdeljenih) naključnih števil, ki se uporablja predvsem pri periodično ponavljajočih se vzorčnih raziskovanjih (s tem postopkom razpršimo obremenitev poročevalskih enot), potem prenesemo vsa naključna števila iz predhodnega okvira v trenutni vzorčni okvir, novim enotam v okviru pa naključno število (med 0 in 1) dodamo z ustrezno računalniško proceduro. Metoda stalnih naključnih števil je zelo uporabna pri koordiniranem vzorčenju, pri katerem izbiramo vzorce za več raziskovanj glede na že izbrane vzorce. S koordiniranim vzorčenjem lahko nadzorujemo obremenitev poročevalskih enot pri več raziskovanjih hkrati.

Smernice za zagotavljanje kakovosti

- Če se različna raziskovanja nanašajo na isto osnovno ciljno populacijo in isto referenčno obdobje, je treba za določitev vzorčnega okvira uporabiti enake postopke, saj s tem bistveno pripomoremo k večji skladnosti statističnih rezultatov.
- Kar najbolj se moramo potruditi (tako v postopku identificiranja podatkovnih virov kot v postopku izdelave), da izločimo neustrezne in podvojene enote, saj s tem dosežemo boljše kakovost statističnih rezultatov.
- Če vrednosti spremenljivk vzorčnega okvira določamo iz več različnih podatkovnih virov, je treba skrbno preučiti postopke za določitev prednosti (prioritet) posameznih podatkovnih virov.
- Kakovost postopkov gradnje vzorčnega okvira in tudi kakovost uporabljenih podatkovnih virov je treba na podlagi povratnih informacij, zbranih v statističnem postopku, redno in sistematično ocenjevati. Če zaznamo večje odmike od še sprejemljivih standardov, je treba izvesti postopke za izboljšanje kakovosti.
- Pri pripravi vzorčnega okvira je treba osebe ali podjetja in druge poslovne subjekte, ki so pred kratkim sodelovali v raziskovanju, izločiti iz vzorčnega okvira in uporabiti koordinirano vzorčenje. Tako zagotavljamo enakomerno porazdelitev bremen med poročevalskimi enotami.

3.3 IZBOR ENOT OPAZOVANJA

Na podlagi izdelanega vzorčnega okvira in določenega vzorčnega načrta se izdelava vzorec enot opazovanja, torej seznam enot, ki so dejansko vključene v raziskovanje in o katerih bomo skušali v poznejših fazah pridobiti želene podatke. V grobem ločimo tri osnovne načine izbora vzorca: izbor na podlagi slučajnih mehanizmov (verjetnostni vzorci), izbor na podlagi determinističnih pravil (neverjetnostni vzorec) in kombinacijo prvih dveh načinov (verjetnostni vzorec se izbere iz vzorčnega okvira, določenega s pragom). Rezultat procesa izbora vzorca so tudi vzorčne uteži (če gre za verjetnostni vzorec) ter seznam ključnih enot. Izjema so raziskovanja, v katera so vključena gospodinjstva; v teh raziskovanjih namreč verjetnost izbora lahko izračunamo, ko pridobimo povratne informacije s terena. Vzorčne uteži so števila, ki jih skupaj s podatki, pridobljenimi od enot v vzorcu, uporabimo pri izračunu ocen populacijskih parametrov. Ključne enote so enote, ki so za naše raziskovanje glede na njihov predvideni vpliv na končni rezultat pomembnejše kot preostale enote in jih zato tako v fazi zbiranja podatkov kot v fazi urejanja podatkov obravnavamo drugače.

Če velikost vzorca ni določena pri načrtovanju raziskovanja, jo izračunamo oz. ustrezno ocenimo na podlagi pomožnih spremenljivk, ki so v dobri korelaciji s proučevanimi spremenljivkami, na podlagi podatkov proučevanih spremenljivk iz prejšnjega obdobja ter s pomočjo razmerja vzorčne variance uporabljenega vzorčnega načrta in vzorčne variance, če bi uporabili metodo enostavnega slučajnega vzorčenja brez ponavljanja. To razmerje se imenuje vzorčni učinek. Izračunano velikost vzorca delimo še s pričakovano stopnjo neodgovora in tako dobimo končno velikost vzorca.

Ko je vzorčni okvir pripravljen in ko je velikost vzorca določena, začnemo postopek izbora vzorčnih enot, ki je bil določen že v fazi načrtovanja raziskovanja. Za to potrebujemo ustrezno programsko opremo in posebne algoritme.

Običajno so vzorčni načrti stratificirani. Že v vzorčnem okviru imamo populacijo razdeljeno na podpopulacije, ki jih imenujemo stratumi. Najprej moramo izračunati stratumsko alokacijo (porazdelitev) vzorca. To pomeni, da moramo za vsak stratum določiti število enot, ki jih bomo izbrali v vzorec. Najpogosteje uporabljamo naslednje tri alokacije:

- enakomerna alokacija: v vsakem stratumu izberemo enako število enot;
- proporcionalna alokacija: v vsakem stratumu izberemo toliko enot, da je delež vsakega stratuma glede na velikost vzorčnega okvira enak deležu izbranih enot iz tega stratuma glede na velikost vzorca (to pomeni, da v večjih stratumih izberemo več enot v vzorec kot v manjših);
- disproporcionalna alokacija: stratumi imajo različne vzorčne deleže, zato se struktura stratumov v vzorcu ne ujema s strukturo stratumov v populaciji;
- optimalna alokacija: glede na podatke za pomožno spremenljivko ali glede na podatke iz prejšnjega referenčnega obdobja izračunamo ustrezne velikosti vzorca v vsakem stratumu, pri čemer poleg velikosti stratumov upoštevamo še variabilnost pomožne oz. proučevane spremenljivke prejšnjega obdobja (v stratumih z večjo variabilnostjo določimo večjo velikost vzorca).

Pri vseh omenjenih alokacijah pazimo, da število izbranih enot ni v nobenem stratumu premajhno. Če je po alokaciji velikost vzorca v kakem stratumu premajhna, določimo neko minimalno fiksno število za velikost (navadno okoli deset enot) ali pa v vzorec izberemo cel stratum, če je njegovo število enot manjše od tega fiksnega števila.

Tudi izbor neverjetnostnega vzorca ali izbor na podlagi determinističnih pravil (npr. zajem s pragom, popis) navadno izvedemo s posebnim algoritmom.

Deterministične metode se razlikujejo od metod verjetnostnega izbora po tem, da se pri determinističnih metodah uporabljajo za izbor vzorca natančno določena pravila, v katera ni vključen noben slučajnostni mehanizem. V praksi se za izbor vzorca najpogosteje uporabljata predvsem dve deterministični metodi: popis ter zajem s pragom. Pri popisu je postopek v resnici zaključen že z določitvijo vzorčnega okvira, saj s tem pojmom označujemo tiste redke primere, pri katerih si lahko privoščimo, da v raziskovanje vključimo vse enote vzorčnega okvira. Pri zajemu s pragom skrbimo seznam enot vzorčnega okvira (podobno kot pri verjetnostnem vzorcu), in sicer tako, da na podlagi vnaprej postavljenih meril nekatere enote vzorčnega okvira izločimo. Merila za izločitev navadno določimo na podlagi vrednosti kake spremenljivke; tiste enote, pri katerih je vrednost spremenljivke nad določeno mejo, vključimo v izbor, drugih pa ne.

Tudi pri zajemu s pragom obstajata dva osnovna pristopa. Prvi pristop je pristop fiksne praga. Kriterij za izbor enot v vzorec se pri tej metodi opredeli s točno določeno vrednostjo kriterijske spremenljivke. Najenostavnejši primer take izbire enot v vzorec je postopek, pri katerem, na primer pri raziskovanjih podjetij in drugih poslovnih subjektov, v raziskovanje vključimo samo enote (npr. podjetja), ki zaposlujejo večje število oseb od števila, ki je določeno (npr. 20). Druga možnost je uporaba variabilnega praga. Pri tej metodi populacijo najprej razdelimo (podobno kot pri slučajnem stratificiranem vzorcu) na podskupine in nato v vsaki podskupini določimo toliko največjih enot (glede na kriterijsko spremenljivko), da bodo izbrane enote glede na vrednosti kriterijske spremenljivke presegale neki vnaprej določeni delež (npr. 75 %). V praksi se pri zajemu s pragom pogosto uporablja kombinacija obeh zgoraj opisanih pristopov.

Smernice za zagotavljanje kakovosti

- Pri verjetnostnem vzorcu naj se (če le ni tehtnih razlogov, ki bi temu nasprotovali) uporabi za izbor enot stratifikacija, saj s tem bistveno pripomoremo k reprezentativnosti vzorca in obenem k večji natančnosti statističnih rezultatov. Izbrane stratifikacijske spremenljivke naj bodo ali visoko korelirane s proučevanimi spremenljivkami ali pa take, ki določajo področja (domene) objavljanja. Glede na teorijo vzorčenja je najučinkovitejša taka stratifikacija, pri kateri so si enote v stratumu čim bolj podobne (glede na neko ključno spremenljivko), enote iz različnih stratumov pa čim bolj različne.
- Pri izrazito asimetričnih porazdelitvah ciljnih spremenljivk v populaciji (ko ima majhno število enot tako visoko vrednost spremenljivke, da je njihov delež v skupni vsoti prevladujoč) je treba z vzorčnim načrtom nekatere stratume vključiti v vzorec z gotovostjo. To je priporočljivo predvsem pri raziskovanjih, v katera so vključena podjetja in drugi poslovni subjekti; pri teh vključimo v vzorec z gotovostjo velika podjetja.

- Zajem s pragom naj se uporabi namesto verjetnostnega vzorčenja samo pri izrazito heterogeni populaciji (glede na proučevano spremenljivko), in sicer takrat, kadar ima populacija pod pragom opazovanja zanemarljiv vpliv na načrtovane statistične rezultate. Če se uvede zajem s pragom, je treba pred tem izvesti študijo, katere rezultati bodo upravičili uvedbo take metode izbora.
- V izračunu alokacije naj se upošteva tudi predvidena stopnja neodgovora v posameznih podskupinah. V podskupine, v katerih je predvidena višja stopnja neodgovora, naj se alocira več enot kot v tistih, v katerih je predvidena nižja stopnja neodgovora. Pri zajemu s pragom naj se merila za določitev praga za podskupine s predvideno visoko stopnjo neodgovora ustrezno prilagodijo.
- Pri verjetnostnem izboru vzorca je pomembno, da se uporabi računalniško generirani slučajni mehanizem. Pri standardnih vzorčnih načrtih se priporoča uporaba že pripravljenih procedur.
- Pri periodičnih raziskovanjih je treba stalno preverjati, ali natančnost rezultatov, dobljenih na podlagi verjetnostnega vzorca, zadošča zahtevanim merilom in standardom. Če ni tako, je treba vzorčni načrt ustrezno spremeniti. V okviru načrtovanih sprememb je treba najprej raziskati možnosti za uporabo učinkovitejšega vzorčnega načrta. Šele če se ugotovi, da s postopki vzorčenja ne bi bilo mogoče izboljšati natančnosti, je treba razmisliti o možnostih za povečanje velikosti vzorca.
- Pri periodičnih raziskovanjih in pri zajemu s pragom je treba stalno preverjati, ali »izpuščeni« del populacije ne povzroča prevelike pristranskosti statističnih rezultatov. To lahko opravimo z občasnim vzorčnim raziskovanjem ali celo s popisom. Če je ocenjena pristranskost glede na vnaprej določena merila in standarde previsoka, je treba ustrezno spremeniti postopke za določitev praga opazovanja.
- Priporočljivo je, da ključnih poročevalskih enot ni preveč (največ tretjina celotnega vzorca), če želimo z določitvijo in uporabo teh enot doseči želeni učinek.

3.4 IZDELAVA ADRESARJA

Za vsako raziskovanje, za katero se podatki zbirajo primarno, je treba izdelati tudi adresar. Z izrazom adresar poimenujemo seznam opazovanih in poročevalskih enot z naslovi in drugimi informacijami, potrebnimi za zbiranje podatkov. Tak seznam se uporablja pri razpošiljanju vprašalnikov, za preverjanje prispelosti vprašalnikov, za komuniciranje s poročevalskimi enotami in pri obdelavi podatkov. Vir za pripravo adresarja za posamezno raziskovanje je ustrezen register (glede na vrsto raziskovanja): Poslovni register Slovenije (PRS; raziskovanja podjetij in drugih poslovnih subjektov), Centralni register prebivalstva (CRP; raziskovanja oseb in gospodinjstev), Statistični register kmetijskih gospodarstev (SRKG; kmetijska raziskovanja), Statistični poslovni register Slovenije (SPRS; raziskovanja podjetij in drugih poslovnih subjektov skupaj vključno z bazo fiktivnih enot). Adresar se pripravi na podlagi seznama iz predhodnega podprocesa priprave in izbora enot opazovanja. Izbor enot opazovanja in poročevalskih enot je lahko določen s postopki vzorčenja, lahko je določen z neposrednim izborom iz ustreznega registra na podlagi natančno postavljenih pravil, lahko pa se oblikuje tudi na osnovi adresarja za isto raziskovanje v predhodnem obdobju z dodajanjem oziroma brisanjem enot. V adresarju je pri raziskovanjih, v katera so vključena podjetja in drugi poslovni subjekti, na podlagi seznama spremenljivk v vzorčnem okviru, označen seznam ključnih enot, to je seznam enot, ki so za raziskovanja najpomembnejše z vidika vpliva na končni rezultat in jih predvsem v fazi zbiranja in urejanja podatkov obravnavamo z višjo prioriteto.

Pri periodičnih raziskovanjih je treba adresar redno vzdrževati. Pri tem je treba upoštevati aktualne spremembe v registru in povratne informacije, zbrane med izvajanjem raziskovanja na terenu. V postopku osveževanja (ažuriranja) adresarja se po potrebi dodajo nove enote, spremljajo se demografske spremembe. Zaznane spremembe se ne upoštevajo samodejno (avtomatsko), temveč se sprememba vnese na podlagi odločitve za vsak posamezni primer posebej. Pri pripravi in vzdrževanju adresarjev je treba nenehno skrbeti tudi za povezave med enoto opazovanja in poročevalsko enoto, ki je sporočitelj podatkov, tj. enoto, ki za enoto opazovanja oz. v njenem imenu sporoča podatke. Spremljati je treba vrnjeno pošto, tj. pošto, ki se iz kakršnih koli razlogov vrne s terena (ker je npr. naslovnik neznan, preseljen ali zaradi pomanjkljivega naslova), in tudi določanje načina zbiranja podatkov za posamezne enote, katerih podatki se pridobivajo na več načinov: z elektronskim poročanjem, s tiskanimi vprašalniki (poštna metoda), po telefonu (telefonsko anketiranje).

Smernice za zagotavljanje kakovosti

- Pravočasno je treba pripraviti in posredovati vso ustrezno izvedbeno dokumentacijo za izdelavo adresarja.
- Adresar mora vsebovati enolično identifikacijo opazovane in poročevalske enote in po potrebi naslov.
- Da bo komuniciranje z opazovanimi in poročevalskimi enotami učinkovito, je treba navesti v adresar čim več kontaktnih podatkov in pri tem upoštevati način zbiranja podatkov (elektronski naslovi, telefonske številke ipd.)
- Opazovanih enot ne brišemo iz adresarjev. Če posamezna enota ni več aktivna, se ji določi ustrezen status aktivnosti (npr. posluje/ne posluje).

4 ZBIRANJE PODATKOV

Zbiranje podatkov je postopek, ki ga izvajamo na začetku izvedbe raziskovanja, vendar močno vpliva na končni rezultat. Podatke lahko zberemo na različne načine. Najpogosteje pridobivamo podatke neposredno pri dajalcih podatkov. Čedalje bolj pa se uveljavljata tudi pridobivanje podatkov iz administrativnih in drugih virov ter kombinirano zbiranje podatkov (tj. kombinacija neposredno zbranih podatkov in podatkov iz administrativnih in drugih virov).

Zbiranje podatkov v grobem sestoji iz treh zaporednih faz: priprave na zbiranje, zbiranja in zajema podatkov. Priprava na zbiranje zajema pripravo vseh potrebnih gradiv in programov, potrebnih za izvedbo zbiranja podatkov, usposabljanje vseh sodelujočih pri zbiranju podatkov, zagotovitev ustreznih pravnih podlag za pridobitev administrativnih in drugih virov podatkov ter ureditev medsebojnega sodelovanja z zunanjimi izvajalci. Zbiranje podatkov zajema vso komunikacijo z dajalci podatkov, uporabo različnih instrumentov za pridobivanje podatkov, upravljanje celotnega procesa zbiranja podatkov. Zajem podatkov vključuje pretvorbo zbranih podatkov v ustrezno elektronsko obliko, potrebno za nadaljnjo obdelavo.

Vprašalniki in metode zbiranja podatkov morajo slediti načelom Kodeksa ravnanja evropske statistike, ki nas zavezujejo k čim večji kakovosti podatkov, zbranih z ustreznimi statističnimi postopki, ob razumni obremenitvi poročevalskih enot in stroškovni učinkovitosti.

Glavni težavi, s katerima se spopadamo pri zbiranju podatkov in ki lahko pomembno vplivata na rezultate raziskovanja, sta neodgovor in merske napake (namerno ali nenamerno napačno sporočanje podatkov). Neodgovor zmanjšuje efektivno velikost vzorca; posledica tega je večja vzorčna napaka in s tem manjša natančnost ocen. V številnih primerih se respondenti (enote, ki odgovarjajo na vprašalnik) razlikujejo od nerespondentov (enot, ki ne odgovarjajo). Pristranskost rezultatov zaradi neodgovorov se popravlja z uteževanjem zaradi neodgovorov. Merske napake se učinkovito zmanjšujejo s sistematičnim testiranjem vprašalnikov pri predstavniki ciljne populacije, s skrbnimi pripravami na zbiranje podatkov, s kakovostno izvedbo in z nadzorom zbiranja podatkov. Na stopnjo odgovorov vplivajo številni dejavniki: splošno ozračje v državi, ugled ustanove, ki zbira podatke, ustrezna usposobljenost anketarjev, strategija komuniciranja in drugi dejavniki. Stopnjo odgovorov lahko učinkovito izboljšamo z ustrezno strategijo komuniciranja z opazovanimi enotami in njihovimi poročevalskimi enotami ter s sistematičnim delom z anketarji. Pri tem je treba upoštevati vrsto omejitev: finančna sredstva, roke za posamezne aktivnosti, omejene človeške vire idr.

4.1 PRIPRAVA NA ZBIRANJE

K pripravi na zbiranje podatkov spada uskladitev časovnega načrta za izvedbo zbiranja podatkov, izdelava vseh gradiv za komuniciranje z dajalci podatkov, izdelava in testiranje (kognitivno in tehnično) vprašalnika, izdelava in testiranje programov za zajem podatkov ter izvedba usposabljanja vseh sodelujočih pri zbiranju podatkov.

Načini zbiranja podatkov so različni in imajo določene posebnosti; razlikujejo se po videzu (oblikovanosti) in vsebini vprašalnika, po načinu izpolnjevanja vprašalnika, po uporabljenih postopkih in po zahtevah do opazovane oz. poročevalske enote. Tudi potek in dolžina priprav na zbiranje podatkov sta odvisna od več dejavnikov, med katerimi sta najpomembnejša periodika izvajanja raziskovanja in način zbiranja podatkov (uporaba le enega načina ali kombinacije različnih načinov).

Priprave na zbiranje podatkov vključujejo pripravo vseh gradiv za komuniciranje z dajalci podatkov in orodij za zbiranje podatkov. Razlikujejo pa se glede na način izpolnjevanja vprašalnika.

Pri zbiranju podatkov z vprašalniki v tiskani obliki je treba pripraviti ustrezno število izvodov gradiv (vprašalnikov, obvestilnih pisem, metodoloških gradiv in drugega gradiva). Gradiva pošljemo naslovnikom po pošti. Hkrati je treba pripraviti tudi vse potrebno za evidentiranje odgovorov in programe za vnos podatkov ali/in optično branje. Gradiva, ki jih bodo opazovane ali poročevalske enote vrnila, morajo biti pravilno naslovljena in opremljena z ustrezno identifikacijo, ki omogoča spremljanje odgovorov. Pri tiskanih gradivih se priporoča uporaba črtne kode z identifikacijo na vsaki strani gradiva.

Pri zbiranju podatkov z elektronskim vprašalnikom poteka obveščanje opazovanih in/ali poročevalskih enot pri raziskovanih podjetjih in drugih poslovnih subjektov praviloma po elektronski pošti (po navadni pošti le prvič pri pošiljanju uporabniškega imena in gesla ali izjemoma), pri raziskovanih, v katera so vključene osebe in gospodinjstva, pa po navadni pošti (po elektronski pošti le izjemoma).

Zbiranje podatkov z vključitvijo anketarja je treba še posebej natančno načrtovati zato, ker je povezano z večjo porabo finančnih virov. Poleg priprave gradiv za komuniciranje je treba pripraviti še vse potrebno za izvedbo: gradiva za anketarje, seznam izbranih oseb oz. gospodinjestev (opazovanih enot) in program za spremljanje in nadzor dela anketarjev. Anketarji lahko zelo vplivajo na kakovost podatkov, zato je pomemben del priprave na zbiranje podatkov tudi njihovo usposabljanje. Anketarji pridobijo pri usposabljanju informacije o varovanju in zaupnosti podatkov, o načinih pristopa do izbranih oseb oz. gospodinjestev, o organizaciji anketiranja, o načinu rokovanja s programom za anketiranje (le pri računalniško podprtem anketiranju) in o vsebini raziskovanja, za katero bodo zbirali podatke. Del usposabljanja anketarjev so tudi praktične vaje v izpolnjevanju vprašalnikov.

Smernice za zagotavljanje kakovosti

- Glede na izbrani način izvedbe raziskovanja je treba z udeleženi v procesu uskladiti časovni načrt izvajanja zbiranja podatkov.
- Obdobje zbiranja podatkov se določi glede na velikost vzorca in trajanje izvedbe ankete/raziskovanja.
- Pred začetkom zbiranja podatkov je treba testirati vsa uporabljena orodja za zbiranje podatkov. Pri elektronskem načinu za podjetja in druge poslovne subjekte se testira vprašalnik.
- Pri tiskanih vprašalnikih morajo biti identifikacijski podatki opazovane enote jasno razvidni na vseh gradivih, ki bodo vrnjena z vpisanimi podatki o opazovanih enotah (zaradi morebitne kasnejše komunikacije ali preverjanja podatkov).
- Pri elektronskih vprašalnikih (raziskovanja, v katera so vključena podjetja in drugi poslovni subjekti) je treba pravočasno poskrbeti za obveščanje opazovanih enot, da izberejo ustrezno osebo v podjetju ali drugo podjetje (npr. računovodski servis, nadrejeno enoto), da pripravijo opremo, ki ustreza tehničnim zahtevam za sporočanje oz. oddajo podatkov in da poskrbijo za registracijo v spletno aplikacijo, ki je osnovni pogoj za ta način poročanja.
- Tehnična oprema (npr. prenosni računalniki) za anketarje morajo biti prilagojeni glede na sprejete standarde. Pri tem je treba upoštevati tudi zaščito podatkov pred morebitno krajo.
- Na vseh gradivih za zbiranje podatkov (vprašalniki, obvestilna pisma, opomniki, zgibanke ipd.) je treba uporabljati standardne vsebinske sklope in celostno grafično podobo.

4.2 PREVZEM ADMINISTRATIVNIH IN DRUGIH ZBIRK PODATKOV

Pri zbiranju podatkov se trudimo v čim večji meri zmanjšati obremenitev poročevalskih enot, kar lahko dosežemo z uporabo administrativnih in drugih zbirk podatkov. V raziskovanjih, v katerih pridobimo podatke na podlagi administrativnih in drugih virov, je problem neodgovora običajno manjši, lahko pa se pojavijo drugačne težave, predvsem take, ki izvirajo iz metodoloških neskladij s statističnimi koncepti. Skrbnik administrativne zbirke podatkov namreč v svojih evidencah vodi podatke, ki so bili zbrani za vodenje upravnih ali drugih postopkov, in ne za statistični namen. Prednost uporabe administrativnih zbirk podatkov je učinkovitejša izkoriščenost teh virov. Hkrati pa to pomeni odvisnost od aktivnosti skrbnika administrativne zbirke podatkov, tako metodološko kot tehnično. Mesto in oblika hranjenja podatkov sta odvisna od skrbnika administrativne zbirke, zato se običajno mora sekundarni uporabnik teh podatkov pri prevzemu podatkov in pri nadaljnjih postopkih obdelave podatkov ravnati po njem.

Katalog administrativnih zbirk podatkov se vodi na enem mestu, zato da lahko ob morebitnih dodatnih potrebah najprej preverimo že obstoječe vire. Katalog je treba redno in sprotno urejati. Vodi se tudi sprotne evidencne prevzemov podatkov.

Z upravljavci administrativnega vira je treba skleniti ustrezen dogovor in tehnični protokol o posredovanju podatkov, v katerem se določijo vsebina posredovanih podatkov, periodika posredovanja in način posredovanja podatkov. Sklenjene dogovore je treba redno preverjati in po potrebi posodabljeni.

Pred prvim prevzemom je treba pripraviti in preveriti navodila, postopke in programe za prevzem podatkov in za izvedbo formalnih kontrol (berljivost podatkov, ustreznost zapisov in njihovo število, preverjanje ustreznosti znakov). Običajna pot prevzemanja podatkov je, da po prejemu obvestila, da so podatki pripravljene, sledi prevzem. Vse administrativne zbirke podatkov, se sprejemajo na informacijsko varovani vhodni točki. Sledi tehnično preverjanje podatkov ter odložitvev podatkov na dogovorjeno mesto. Pri prenosu s fizičnim nosilcem se pripravi potrdilo o prevzemu. Fizični nosilec se ustrezno označi, vpiše in po prenosu podatkov na dogovorjeno mesto arhivira v arhivu administrativnih zbirk podatkov.

Načini prenosa podatkov so odvisni od občutljivosti oziroma podrobnosti podatkov. Običajni načini prenosa podatkov so:

- na različnih podatkovnih nosilcih (npr. CD, DVD, prenosni diski, USB-ključi ...)
- prek protokola FTP (angl. file transfer protocol) ali protokola SFTP (angl. safe file transfer protocol)
- po elektronski pošti
- s pomočjo spletnih storitev (angl. web services)
- prek replikacije na ravni podatkovnih strežnikov (popolna replikacija ali prevzem podatkov na podlagi vnaprejšnjega dogovora)

O morebitnih nepravilnostih pri prevzemu ali napakah v strukturi podatkov obvestimo tehničnega skrbnika administrativne zbirke podatkov (poskrbimo tudi za povratno informacijo).

Smernice za zagotavljanje kakovosti

- Pred začetkom prevzemanja podatkov morata biti pripravljena in podpisana dogovor in tehnični protokol.
- Imenovati je treba vsebinskega skrbnika podatkov in njegovega namestnika ter osebe, odgovorne za prevzem podatkov. Vsi udeleženi v procesu morajo biti pred prvim prevzemom seznanjeni z vsebino podatkov.
- Zunanji upravljavec administrativne zbirke podatkov obvešča prejemnika podatkov o času (datumu, roku), ko so podatki pripravljene (na razpolago) za prevzem skladno z dogovorjenimi roki. Oseba, odgovorna za prevzem podatkov, spremlja dogovorjene roke za posredovanje podatkov iz administrativnih zbirk podatkov in opozarja nanje.
- Preizkusiti je treba delovanje dogovorjenega tehničnega prenosa zbirke podatkov. Skrbeti je treba, da deluje sistem varnega prenosa podatkov in sprotne obveščanja o spremembah pri delovanju sistema prenosa.
- O prevzemih administrativnih zbirk podatkov se sproti/ažurno vodi ustrezna evidenca prevzemov.
- V skladu z varnostno politiko se vodi tudi evidenca fizičnih nosilcev podatkov.
- Ob spremembah zakonodaje z delovnega področja je treba slediti vplivom sprememb na administrativno zbirko podatkov, ki jo prevzemamo, in spremembe pravočasno uskladiti z zunanjim upravljavcem administrativne zbirke podatkov.
- Redno in pravočasno je treba poskrbeti za obveščanje uporabnikov o spremembah, morebitnih zamudah in o drugih izrednih dogodkih.
- Če so v administrativnih zbirkah podatkov osebni podatki, moramo z njimi ravnati v skladu z Zakonom o varstvu osebnih podatkov (ZVOP-1).

4.3 ZBIRANJE PODATKOV IN KOMUNICIRANJE S POROČEVALSKIMI ENOTAMI

Zbiranje podatkov zajema vso komunikacijo z dajalci podatkov, uporabo različnih instrumentov za pridobivanje podatkov in upravljanje celotnega procesa zbiranja podatkov.

Visoko stopnjo odgovorov poročevalskih enot lahko dosežemo z jasnimi cilji, z ustreznim komuniciranjem in nudenjem podpore dajalcem podatkov in izbiro najprimernejšega načina in časa zbiranja podatkov. Tudi v raziskovanjih, za katera so opazovane enote po zakonu zavezane k sporočanju podatkov, se spopadamo z neodgovorom (manjkajočim odgovorom oz. z odsotnostjo odgovora).

Neodgovor lahko razdelimo v dve glavni kategoriji: neodgovor spremenljivke in neodgovor enote. O neodgovoru spremenljivke govorimo takrat, kadar odgovori na nekatera vprašanja v vprašalniku, ki je sicer izpolnjen, manjkajo. To se lahko zgodi iz različnih razlogov: ker npr. ni bilo mogoče odgovoriti, zaradi pomanjkanja znanja, zaradi zavračanja odgovora na občutljivo vprašanje, lahko gre za protislovne odgovore, ki niso v skladu s pravili za kontrolo podatkov. O neodgovoru enote pa govorimo takrat, kadar podatki manjkajo pri vseh ključnih spremenljivkah. Pri različnih načinih zbiranja podatkov dosegamo različne stopnje neodgovorov. Pri raziskovanjih, pri katerih je sporočanje podatkov prostovoljno, dosegamo najnižjo stopnjo odgovorov pri vprašalnikih za samoizpolnjevanje (tiskanih in spletnih), sledi računalniško podprto telefonsko anketiranje. Najboljši odziv anketirancev še vedno dosežemo s t. i. osebnim anketiranjem na terenu, vendar pa je to hkrati tudi najdražji način zbiranja podatkov.

Pri samoizpolnjevanju respondent (enota, ki odgovarja na vprašalnik) sam prebere vprašanje in navodila in odgovori tako, kot to sam razume, zato podatki pogosto vsebujejo precej več napak kot pri drugih načinih zbiranja podatkov; to pozneje pri postopkih urejanja terja več dela. Izpolnjeni vprašalniki ali komentarji respondentov lahko prispejo po pošti, kot opomba k spletnemu vprašalniku, po telefaksu, na elektronski naslov ali po telefonu. Evidenco odgovorov redno spremljamo. Poslovnim subjektom, ki se niso odzvali, pošljemo največ dva opomina v elektronski obliki. Pri nekaterih raziskovanjih, v katerih statistično opazujemo podjetja in

druge poslovne subjekte, ter pri tistih za osebe in gospodinjstva, so predvideni tudi opomniki, ki jih pošljemo enoti nekaj dni pred potekom roka za sporočanje podatkov ali na dan, ko poteče rok za sporočanje podatkov. Po potrebi enote, posebej ključne opominjamo tudi telefonsko.

Pri zbiranju podatkov na terenu vprašalnike za opazovane enote izpolnjuje anketar. Ta tudi natančno ve, kaj sprašuje in kako naj bodo odgovori vneseni v vprašalnik, in zagotovi, da dobi na vsa vprašanja sprejemljiv odgovor. Morebitni neodgovor in ustreznost izvajanja zbiranja podatkov se sproti nadzorujeta s kontrolo podatkov, ki jih pošilja anketar.

Pri telefonskem anketiranju anketarje nadzira kontrolor, ki je zadolžen za spremljanje njihovega dela. Spremlja, kako anketiranje napreduje, nadzoruje kakovost, anketarje opozarja na napake in morebitno nezadostno kakovost ter skrbi za odmore med delom. Usposabljanje anketarjev, nadzor in podajanje dodatnih navodil je tu enostavnejše kot pri drugih načinih zbiranja podatkov. Zaželeno je, da je na začetku izvedbe raziskovanja navzoča pri anketiranju tudi oseba, ki dobro pozna vsebino raziskovanja, da odgovarja na metodološka vprašanja anketarjev; ti jo seznanjajo tudi z odzivom anketirancev.

Opazovanim enotam (pri raziskovanjih za podjetij in druge poslovne subjekte) določimo glede na njihovo odzivnost pri sporočanju podatkov oz. njihovo ekonomsko aktivnost v gospodarstvu ustrezen status enote. Statusi opredeljujejo način aktivnosti enote (poslujejo/ne poslujejo, demografske spremembe pri podjetjih, odgovor/neodgovor; vrste neodgovorov ipd.), ki v nadaljnjih postopkih omogočajo izbor načina komuniciranja z enoto in obravnave podatkov poročevalskih enot v naslednjih fazah procesa.

Smernice za zagotavljanje kakovosti

- Pri zbiranju in komuniciranju z opazovanimi oz. poročevalskimi enotami se upoštevajo navodila, napisana v izvedbeni dokumentaciji.
- Nenehno si je treba prizadevati, da ugotovimo, kaj je z enotami, s katerimi nismo vzpostavili stika. Uporabljati je treba tudi ustrezne postopke za zmanjšanje zavrnitev. Identificirati je treba segmente populacije z višjo stopnjo neodgovorov in povečati intenzivnost vzpostavljanja stikov z njimi. Treba je zagotoviti odgovor ključnih enot.
- Določiti je treba minimalen nabor podatkov, ki jih morajo enote sporočiti, da jih še lahko upoštevamo kot odgovor.
- Zagotoviti je treba, da izbrane enote zaradi sporočanja podatkov nimajo stroškov ali da so ti najmanjši možni.
- V obdobju zbiranja podatkov je treba redno spremljati število prispelih odgovorov in ustreznost izvajanja postopkov, povezanih z zbiranjem podatkov.
- Pri pripravi ustreznih postopkov je treba upoštevati tudi obdobje zbiranja podatkov, trajanje zbiranja podatkov ter predvideno obremenitev za vse udeležene (način zbiranja, dolžina vprašalnika, težavnost vsebine, periodika raziskovanja).
- Določiti je treba največje še primerno število stikov z enoto. Običajno opazovanim oz. poročevalskim enotam (pri raziskovanjih za podjetja in druge poslovne subjekte) pošljemo največ dva pisna opomina.
- Odzivnost enot s časom praviloma hitro pada. Zato moramo pri opominjanju doseči primerno ravnotežje med časom, porabljenim za opominjanje, in učinkom opominjanja.
- K uspešnemu izvajanju komunikacijskega načrta prispeva stalno posodabljanje kontaktnih podatkov poročevalcev. S stalnim spremljanjem odzivnosti enot se lahko sproti prilagaja načrt komuniciranja z enotami oziroma se odzivnost enot upošteva pri pripravi komunikacijskega načrta za naslednje obdobje.

4.4 ZAJEM PODATKOV

Zbiranje podatkov poteka v različnih oblikah. Če jih prejmemo v elektronski obliki, so že ustrezni za združevanje v vhodni bazi; podatke na tiskanih vprašalnikih pa je treba za poznejše obdelave najprej pretvoriti v elektronsko obliko. Pretvorba v elektronsko obliko lahko poteka na dva načina:

- z ročnim vnosom (rezultat tega postopka so podatki v elektronski obliki) ali
- vnos z optičnim branjem s pomočjo optičnega čitalnika (rezultata tega postopka sta vedno dva: slike vprašalnikov in podatki v elektronski obliki).

Ustrezen način vnosa podatkov se določi glede na količino in možno obliko vnosnih polj na vprašalniku, glede na število opazovanih enot oz. vprašalnikov in glede na vsebinsko zahtevnost vprašalnika.

Izjemoma se kontrola vprašalnikov izvede pred vnosom, in sicer takrat, kadar gre za tiskan vprašalnik in kadar ni treba originalnih sporočenih podatkov shranjevati v elektronski obliki. Izbira načina vnosa podatkov je odvisna tudi od načina izpolnjevanja vprašalnikov. Če za vnos podatkov uporabimo optično branje, je treba upoštevati tehnične pogoje, ki jih optično branje zahteva (npr. vpis podatkov v točno določena polja, izpolnjevanje z modrim ali črnim kemičnim svinčnikom, šifriranje v za to predvidena polja ipd.). Pri ročnem vnosu podatkov te omejitve ne veljajo in je zaradi kakovosti vnosa podatkov zaželeno le čitljivo izpolnjevanje vprašalnika.

Kadar podatke pridobimo s kombiniranimi načini poročanja (npr. tiskani, elektronski vprašalniki, telefonski vprašalniki idr.) ali kadar se kombinirajo načini za vnos podatkov (npr. optično branje in ročni vnos) ali kadar smo del podatkov prevzeli iz administrativne zbirke podatkov, je za nadaljnji statistični proces treba zajete podatke združiti v enotno bazo.

Smernice za zagotavljanje kakovosti

- Pripraviti je treba jasna in natančna navodila za preverjanje podatkov pred vnosom, če je to potrebno.
- Pri zajemu podatkov v tiskani obliki je stroškovno in časovno primernejše optično branje. Pri tem moramo z navodili ali ustreznim oblikovanjem vprašalnika poskrbeti, da bodo podatki zapisani jasno in čitljivo.
- Če uporabljamo hitri vnos brez ponavljanja, je treba vprašalnik oblikovati tako, da je mogoče izdelati notranje kontrole (npr. posebno polje za seštevke). Pred zajemom podatkov s tiskanih vprašalnikov se je treba izogibati kontroli vprašalnikov pred vnosom. Šifriranje podatkov se lahko izvaja, če je to potrebno.

5 OBDELAVA PODATKOV

Obdelava podatkov označuje vse postopke, s katerimi podatke, ki smo jih zbrali v raziskovanju, z uporabo ustreznih statističnih metod obdelamo ter uredimo v obliko, primerno za analizo in objavo. Proces zaobjema tako obdelavo podatkov na mikroravni (npr. odprava napak ter manjkajočih vrednosti na ravni posamezne enote) in postopke preračuna na populacijo (uteževanje, agregacija) kot tudi obdelavo že agregiranih podatkov (npr. desezoniranje, zaščita podatkov). V procesu obdelave podatkov uporabljamo različne statistične metode, katerih namen je predvsem omogočiti izračun čim točnejših in čim bolj nepristranskih statističnih ocen. Podatke, ki so v izvedbi raziskovanja osnova za izračun statističnih ocen, lahko pridobimo na različne načine oziroma iz različnih virov. Vse razpoložljive vire podatkov je treba ustrezno povezati. Opredeliti je treba morebitno uporabo statističnih modelov ter določiti postopek za izračun populacijskih ocen.

Obdelava podatkov se je v celotni izvedbi raziskovanja zaradi razvoja statističnih metod in informacijske tehnologije v zadnjem obdobju najbolj spremenila, predvsem je postala hitrejša in učinkovitejša. Vse več je računalniško podprtih, avtomatiziranih postopkov, ki omogočajo hitro in učinkovito izvajanje ter ponovljivost postopkov. Kljub nedvomnim prednostim takega razvoja pa obstaja tudi pomanjkljivost pri avtomatski obdelavi podatkov, to pomeni, da so nekateri postopki izvedeni na avtomatiziran način, brez vpliva izvajalca. Da bi omilili negativne vplive takega izvajanja, je treba proces statistične obdelave podatkov skrbno načrtovati, jasno in podrobno dokumentirati, predvsem pa omogočiti pripravo potrebnih procesnih metapodatkov, takih, ki nudijo izvajalcu vpogled v izvajanje avtomatiziranih postopkov. V zadnjem času je precej aktualen tudi tako imenovani metapodatkovno vodeni (ang. metadata driven) pristop, pri katerem se nastavitve oziroma parametrizacije posameznih delov procesa lahko določijo v nekem zunanjem okolju (običajno v podatkovni bazi, ki je ločena od osnovne računalniške aplikacije) in jih lahko določijo oziroma spreminjajo sami izvajalci raziskovanja.

5.1 UREJANJE ADMINISTRATIVNIH in drugih ZBIRK PODATKOV

Tudi administrativne in druge zbirke podatkov, ki jih uporabljamo bodisi kot neposreden vir podatkov bodisi kot pomožne spremenljivke, vsebujejo različne vrste napak. Zato je treba tudi te zbirke podatkov pred uporabo v čim večji meri urediti oziroma morebitne napake v njih v čim večji meri odkriti in odpraviti. Urejanje tovrstnih zbirk podatkov je pomembno predvsem takrat, kadar povezujemo več podatkovnih zbirk.

V tem podprocesu gre za urejanje administrativnih in drugih zbirk podatkov, ki ga izvajamo pred postopkom povezovanja različnih virov podatkov. Najpogostejše vrste kontrol pri takem urejanju so: preverjanje vrednosti, ki so vezane na šifrante; preverjanje obsega vrednosti pri številskih spremenljivkah (npr. negativne vrednosti); preverjanje dvojnikov (po enoličnih identifikatorjih in/ali po vseh spremenljivkah); preverjanje okvirno pričakovanega števila zapisov; preverjanje ustreznosti referenčnih obdobj, na katera se nanašajo podatki, ki smo jih prejeli. Na tem mestu se lahko kontrolira pravilnost povezav med več spremenljivkami v istem viru. Večina kontrol doslednosti se sicer izvaja pozneje v procesu statističnega urejanja; takrat se preverja tudi skladnost med spremenljivkami iz različnih virov, tako administrativnih kot statističnih.

Smernice za zagotavljanje kakovosti

- Vse vire podatkov je treba ob prevzemu jasno in natančno dokumentirati, saj bomo s tem olajšali delo pri poznejši uporabi administrativnih in drugih zbirk podatkov v raziskovanjih.
- Za vsako spremenljivko, ki je vezana na šifrant, je treba natančno ugotoviti, katera verzija šifranta je glede na veljavnost podatkov ustrezna. Ta verzija šifranta se nato uporabi kot podlaga za preverjanje vrednosti.
- Za vse številске spremenljivke je treba skrbno preučiti njihovo naravo in nato določiti nabor sprejemljivih vrednosti. V večini primerov se je treba predvsem odločiti, ali bodo dovoljene ničelne ali celo negativne vrednosti. Če ni vsebinskih in/ali tehničnih omejitev, je treba omogočiti sledljivost vseh sprememb podatkov. Izvorni podatki naj se torej v fazi urejanja ne prekrijejo s popravki, ampak naj se, če je bil podatek popravljen, vedno tvori nova verzija podatka. Vsak popravljeni podatek naj bo tudi »opremljen« z ustreznim metapodatkom, ki podaja informacijo, kje v procesu in zakaj je bil podatek popravljen.
- Vse kontrole smiselnosti za posamezne spremenljivke je treba natančno dokumentirati. Če je vrednost pri določenih spremenljivkah omejena, je treba navesti razlog za tako »omejitev«.
- Povratna informacija o zaznanih napakah naj se zunanjemu upravljavcu administrativne in druge zbirke podatkov posreduje le v splošni obliki, in ne na ravni posameznih enot, vendar le, če napaka ni bila ugotovljena na podlagi uporabe dodatnih informacij iz drugih podatkovnih virov.
- Z administrativnimi in drugimi zbirkami podatkov, ki vsebujejo osebne podatke, je treba ravnati v skladu z Zakonom o varstvu osebnih podatkov (ZVOP-1).

- Postopke, po katerih se je treba ravnati, če se zaznajo napake, je treba opredeliti smiselno in racionalno, da bi se izognili preveliki časovni zahtevnosti postopkov.
- Tehnična rešitev naj bo čim splošnejša, enostavna in dokumentirana.

5.2 INTEGRACIJA RAZLIČNIH PODATKOVNIH VIROV

V raziskovanjih, v katerih se uporablja več različnih podatkovnih virov, je treba te različne vire povezati, združiti in jih pripraviti v obliko, ki omogoča nadaljnjo obdelavo. Če v virih obstajajo enolični identifikatorji, poteka združevanje neposredno prek teh, drugače pa je treba definirati in pripraviti postopek za združevanje zapisov prek drugih parametrov (posredno povezovanje). Za lažji prikaz postopkov predpostavimo, da pri postopku združevanja vedno obstaja referenčni nabor enot opazovanja (referenčni vir); ta je pri postopkih povezovanja izhodiščni vir. Združeni nabor podatkov, ki je rezultat postopka integracije, naj bi vseboval vse enote iz referenčnega vira.

Neposredno povezovanje zapisov iz različnih virov poteka prek enoličnega identifikatorja, ki obstaja v obeh virih. Zapise iz pridruženega vira, ki nam jih prek enoličnega identifikatorja ni uspelo povezati z referenčnim virom, (po potrebi) zapišemo v poseben seznam.

Če v referenčnem in pridruženem viru ni skupnega enoličnega identifikatorja, ju povežemo s postopki posrednega povezovanja. V tem primeru si moramo pomagati z drugimi izbranimi skupnimi spremenljivkami, ki obstajajo v obeh virih. Enote referenčnega vira v tem primeru »razpadejo« na dva dela. V prvem delu so enote, ki smo jih prek posrednih povezovalnikov enolično povezali s pridruženim virom. Pri teh enotah se vrednosti posrednih povezovalnikov v obeh virih povsem ujemajo in v pridruženem viru obstaja samo en tak zapis. V drugem delu so enote, ki nam jih ni uspelo enolično povezati s pridruženim virom; to pomeni, da ima vsak zapis iz referenčnega vira »pridruženih« več zapisov iz pridruženega vira. Pridruženi zapisi so v tem primeru tisti, pri katerih so vrednosti posrednih povezovalnikov v obeh virih najbolj »podobni«. Podobnost mora biti seveda povsem nedvoumno definirana, običajno pa se za ta namen uporabi ena izmed že obstoječih funkcij v različnih programskih okoljih za numerično vrednotenje ujemanja nizov.

Za nadaljnjo statistično obdelavo je treba tudi v drugem naboru podatkov izbrati samo en zapis. Ali ta izbor poteka ročno ali programsko na podlagi dogovorjenega algoritma, se je treba odločiti pri vsakem posameznem raziskovanju posebej.

Pri združevanju virov pridobimo vrednosti kake spremenljivke pogosto iz več različnih virov. V takem primeru je treba natančno določiti prioriteto prevzema virov in za to spremenljivko dodati novo spremenljivko s kazalnikom, ki za vsako enoto pove, kateri vir je bil pri določitvi vrednosti spremenljivke določen kot prednost.

Smernice za zagotavljanje kakovosti

- Pred določitvijo postopka združevanja podatkov iz različnih virov je treba upoštevati njihovo razpoložljivost in njihovo uporabnost (to pomeni: ali spremenljivke v določenem viru zadoščajo zahtevanim statističnim opredelitvam in ali bodo podatki iz administrativnih virov na voljo pravočasno).
- Če se za isto spremenljivko uporabi več različnih virov, je treba ugotoviti, ali so podatki iz teh različnih virov med seboj skladni.
- Če se uporabljajo postopki posrednega povezovanja, jih je treba postaviti tako, da zadoščajo tako zahtevam po čim večji pokritosti kot tudi zahtevam po časovni racionalnosti. Predvsem se je treba izogibati preveliki količini »ročnih« postopkov, ki bi lahko v veliki meri zmanjšali pravočasnost zagotovitve statističnih rezultatov.
- Zagotoviti je treba, da ima ob koncu postopka povezovanja vsaka enota opazovanja v bazi ustrezen status enote, vsaka spremenljivka pa ustrezen status spremenljivke. Če so vrednosti določene spremenljivke pridobljene iz več različnih virov, je treba zagotoviti, da je za vsako vrednost nedvoumno razvidno, iz katerega vira je bila pridobljena.
- Za vsako spremenljivko, ki je vezana na šifrant, je treba natančno ugotoviti, katera različica šifranta je glede na veljavnost podatkov ustrezna. Ta različica šifranta se nato uporabi za kontrolo vrednosti spremenljivke.
- Za vse številске spremenljivke je treba skrbno preučiti njihovo naravo in nato določiti nabor sprejemljivih vrednosti. V večini primerov se je predvsem treba odločiti, ali bodo dovoljene ničelne ali celo negativne vrednosti.
- Vse kontrole, ki jih uporabimo v tem delu procesa, je treba natančno dokumentirati. Če je morda vrednost pri določenih spremenljivkah »omejena«, je treba navesti razlog za tako »omejitev«.

- Postopke za ravnanje ob morebitnih zaznanih napakah je treba definirati smiselno in racionalno, da se izognemo preveliki časovni zahtevnosti postopkov.
- Tehnična rešitev naj bo čim splošnejša in ustrezno dokumentirana.

5.3 UREJANJE PODATKOV NA MIKRORAVNI

Postopke urejanja podatkov lahko glede na uporabljeno raven podatkov v grobem razdelimo na urejanje podatkov na mikroravni in na urejanje podatkov na makroravni. Pri urejanju podatkov na mikroravni izvajamo postopke na ravni podatkov posameznih enot. Tudi pri urejanju podatkov na mikroravni ločimo več postopkov, in sicer glede na to, kako so bili podatki pridobljeni:

- Pri računalniško podprtih vprašalnikih se celoten ali delni nabor logičnih kontrol izvaja že med anketiranjem. Po končanem zbiranju se podatki združijo – skupaj z morebitnimi sekundarnimi viri – v skupno datoteko; v tej datoteki se izvedejo še dodatne kontrole, morebitne zaznane napake pa popravijo, in to z ročnimi ali avtomatskimi postopki popravkov.
- Pri izpolnjevanju tiskanega vprašalnika se logične kontrole izvajajo po končanem vnosu. Kontrole celotnega nabora podatkov se izvajajo hkrati (paketno), in sicer z avtomatskimi popravki in z generiranjem statistike napak. Zaznane napake se tudi v tem primeru popravijo ali z ročnimi ali avtomatskimi postopki popravkov.
- Če v raziskovanju uporabljamo kombinacijo več različnih virov podatkov (npr. statističnih in administrativnih), je treba najprej urediti vsak vir posebej, nato pa izvesti še kontrolo v združenem naboru podatkov. Pri tem je treba poudariti, da je treba napake, ki se zaznajo v združeni datoteki, v čim večji meri odpraviti v vsakem posameznem viru, in ne v združeni datoteki, saj se le tako lahko zagotovi ponovljivost postopkov.

Ker je proces urejanja podatkov tako s časovnega kot s stroškovnega vidika zelo zahteven, je treba postopke procesa urejanja podatkov čim bolj avtomatizirati. Računalniški programi so izdelani tako, da napake hkrati odkrivajo in odpravljajo; to pomeni, da so popravljeni podatki določeni z eno od ustaljenih metod vstavljanja podatkov.

Predvsem pri raziskovanjih podjetij in drugih poslovnih subjektov se pogosto uporablja selektivno urejanje podatkov. To je postopek, s katerim določimo enotam različne prioritete za urejanje podatkov. Treba je torej opredeliti postopek, s katerim v fazi urejanja podatkov določimo enote, ki so za naše rezultate res pomembne, in jim je v fazi urejanja podatkov treba nameniti več pozornosti. Enote, ki jih v fazi selekcije označimo z višjo prioriteto, običajno urejamo »ročno« (z morebitno ponovno vzpostavitvijo stika s poročevalskimi enotami), druge enote pa s postopki avtomatskega urejanja.

Tudi avtomatske popravke lahko v grobem razdelimo v dve skupini: na deterministične in na verjetnostne popravke. Prve popravke določimo z enostavnimi determinističnimi pravili, ki jih lahko zapišemo v obliki logično-aritmetičnih izrazov (npr. $\text{if } x > 10 \text{ and } y < 1 \text{ then } y = 1$). Druge popravke določimo na podlagi verjetnostnih postopkov, običajno na podlagi minimalne spremembe sporočenih podatkov (Fellegi-Holtov pristop).

Smernice za zagotavljanje kakovosti

- Logične kontrole morajo biti definirane konsistentno in neprotislovno. Logične kontrole je treba testirati v testnem programu. Pri računalniško podprtem zbiranju podatkov je treba opredeliti logične kontrole, ki bodo vključene že v sam vprašalnik.
- Izgibati se je treba praksi, da bi v želji, da bi z raziskovanjem zbrane podatke čim bolj »prečistili«, postavili preveč kontrol, s prestrogo postavljenimi merili. Taka praksa namreč vodi do tako imenovanega pretiranega urejanja (ang. *overediting*), in posledica takega urejanja je, da tako dobimo preveč enot, katerih vrednosti bi morali znova preveriti.
- Raziskati je treba, ali je mogoče uvesti postopke za avtomatsko urejanje podatkov, saj bi lahko tako bistveno zmanjšali stroške, skrajšali čas in povečali učinkovitost urejanja. Preden bi se taki postopki uvedli v redni proces, bi bilo treba podrobno preučiti, ali so smiselni in izvedljivi (tako z vsebinskega kot s tehničnega vidika).
- Vse postopke urejanja podatkov je treba podrobno in urejeno dokumentirati.
- Pri periodičnih raziskovanjih je treba zagotoviti, da rezultate urejanja uporabimo za izboljšanje kakovosti podatkov v naslednjih izvedbah raziskovanj.
- Proces urejanja podatkov je treba postaviti tako, da bo omogočal ponovljivost postopkov; to pomeni, da moramo omogočiti, da bodo postopki urejanja podatkov, pri katerih bodo uporabljeni enaki vhodni podatki in enake metode, vedno dali enak končni rezultat.

- Omogočiti moramo sledljivost vseh sprememb podatkov. Izvorni podatki naj se torej v fazi urejanja ne prekrijejo s popravki, ampak naj se, če je bil podatek popravljen, vedno tvori nova verzija podatka. Vsak popravljen podatek naj bo tudi »opremljen« z ustreznim metapodatkom, ki podaja informacijo, kje v procesu in zakaj je bil podatek popravljen.

5.4 VSTAVLJANJE PODATKOV (IMPUTACIJE)

Izraz vstavljanje podatkov (imputacije) označuje vse postopke, pri katerih manjkajoče ali v procesu urejanja podatkov zaznane napačne vrednosti nadomestimo s statističnimi ocenami. Pri tem moramo razločevati postopke, pri katerih vrednosti nadomeščamo s statističnimi ocenami, od postopkov, pri katerih popravljene vrednosti pridobimo s ponovno vzpostavitvijo stika s poročevalskimi enotami ali z uporabo administrativnega in drugega vira podatkov. O vstavljanju podatkov govorimo samo v prvem primeru.

Postopki vstavljanja podatkov naj bi pomagali izboljšati statistične ocene na ravni agregatov. Da bi ta namen dosegli, moramo metode vstavljanja podatkov izbrati skrbno in pretehtano. Te metode morajo biti izbrane tako, da bodo vstavljene vrednosti čim boljši približki pravih (toda neznanih) vrednosti, da bodo zagotovila konsistentne podatke (glede na podan nabor logičnih kontrol) in bodo v čim večji meri ohranjala osnovno porazdelitev pridobljenih podatkov.

Obstaja veliko metod vstavljanja podatkov. Tukaj bomo na kratko opisali le nekatere najpogosteje uporabljene:

- **Metoda logičnih imputacij.** Pri tej metodi vstavimo vrednost, ki logično sledi iz podatkov, ki nam jih je za obravnavano enoto uspelo pridobiti. Če npr. za posamezno osebo poznamo njen datum rojstva, ne pa njene starosti, lahko manjkajočo starost (za določeno časovno točko) izračunamo.
- **Metoda povprečne vrednosti.** Pri tej metodi nadomestimo manjkajočo vrednost s povprečno vrednostjo podatkov tistih enot, katerih podatke imamo. Povprečne vrednosti običajno ne izračunavamo iz celotnega nabora podatkov, ampak iz podatkov za domeno, v kateri je enota, za katero vstavljamo podatek (npr. iz podatkov za občino).
- **Metoda notranjega darovalca.** Pri tej metodi vstavimo vrednost, ki jo privzamemo od kake druge enote (darovalca), za katero imamo podatek. Pri tem je darovalec lahko točno določen (na podlagi funkcije pomožnih in drugih proučevanih spremenljivk) ali izbran slučajno. Od izvedbe metode je odvisno, ali vrednost darovalca preprosto prepisemo ali pa jo ustrezno prilagodimo (npr. za razmerje vrednosti pomožne spremenljivke enote, ki ji vstavljamo podatek, in darovalca).
- **Metoda zunanjega darovalca.** Pri tej metodi vstavimo vrednost, ki jo privzamemo iz kakega zunanjega vira, npr. iz istega raziskovanja za prejšnje referenčno obdobje (pri periodičnih raziskovanjih), iz drugega raziskovanja ali iz administrativnih podatkovnih virov. Kot pri metodi notranjega darovalca lahko tudi tu vrednost preprosto prepisemo ali pa jo ustrezno prilagodimo (npr. za koeficient rasti pomožne spremenljivke).
- **Metoda deležev oz. strukture podatkov.** To metodo uporabimo, kadar je navedena vsota in moramo vstaviti posamezne komponente te vsote. Vrednost posamezne komponente vstavimo tako, da vsoto pomnožimo z nekim deležem. Ta delež lahko izračunamo na več načinov (odvisno od izvedbe metode), pomembno pa je, da je vsota deležev vseh komponent vsote enaka 1.
- **Metode regresije.** Pri tej metodi vstavimo vrednost, ki jo izračunamo iz ustreznega regresijskega modela. Najenostavnejši regresijski model je model linearne regresije. Če npr. za izračun vrednosti spremenljivke Y uporabimo dve pomožni spremenljivki (X , Z), je matematična enačba takega modela naslednja: $Y = \alpha \cdot X + \beta \cdot Z + \varepsilon$. Pri tem sta α in β parametra, ki ju ocenimo na podlagi enot, za katere poznamo vrednosti vseh treh spremenljivk (Y , X , Z), ε pa je slučajni ostanek (ang. residual).

Metode vstavljanja podatkov se delijo tudi na:

- deterministične metode (ocenjena vrednost se izračuna z analitičnim postopkom, pri katerem se uporablja ustrežna deterministična funkcija)
- stohastične metode (ocenjena vrednost se izračuna s postopkom, pri katerem se uporablja verjetnostni mehanizem).

Da bi izvajalcem raziskovanja omogočili boljši vpogled v to, kakšen vpliv imajo uporabljeni postopki vstavljanja na statistične rezultate, je treba v procesu omogočiti izračunavanje ustreznih kazalnikov kakovosti. Dva ključna kazalnika kakovosti za proces vstavljanja podatkov sta delež vstavljenih podatkov in vpliv vstavljanja podatkov na statistike:

- kazalnik delež vstavljenih podatkov se izračunava vsaj za ključne spremenljivke, in sicer kot razmerje med številom enot, za katere smo podatek za obravnavano spremenljivko vstavili, in med številom vseh enot, ki bi morale imeti podatek za to spremenljivko;
- kazalnik vpliv vstavljanja podatkov na statistike se izračunava vsaj za ključne statistike, in sicer kot (relativna) razlika med statistiko, izračunano pred postopkom vstavljanja podatkov, in statistiko, izračunano po tem postopku.

Smernice za zagotavljanje kakovosti

- Manjkajoče podatke, ki so posledica neodgovora spremenljivke, je treba vedno nadomestiti z ustrežno statistično oceno. Ohranjanje manjkajočih vrednosti v končnem naboru podatkov lahko po eni strani povzroči nezanesljivo pristranskost rezultatov ali pa oteži izvajanje zahtevnejših statističnih analiz.
- Pri manjkajočih podatkih, ki so posledica neodgovora enote, naj se metode vstavljanja uporabijo samo pri majhnem številu spremenljivk ali če obstajajo tehtni razlogi za domnevo, da lahko uporaba teh metod bistveno zmanjša pristranskost. Sicer naj se pri manjkajočih podatkih, ki so posledica neodgovora enote, uporabi izračun uteži neodgovora (glej 5.5).
- Pred izbiro metod vstavljanja podatkov je treba izvesti natančne in dovolj obsežne simulacijske študije, ki pokažejo, katera metoda vstavljanja podatkov bi bila najustreznejša. Z izvedbo simulacijskih študij je treba opraviti predvsem naslednje:
 - določiti mehanizem ustvarjanja manjkajočih podatkov, ki bo čim bližje predpostavljeni realnosti izvedbe raziskovanja;
 - pri testiranju metod uporabiti vse pomožne podatke, ki so na voljo med izvajanjem statistične obdelave podatkov;
 - pri testiranju metod predvsem ocenjevati, katera metoda najbolje ohranja statistične lastnosti rezultatov na makroravni, in ne tega, katera metoda najbolje napoveduje vrednosti na mikroravni.
- Vsak vstavljeni podatek moramo v statističnem procesu označiti z uporabo ustreznega statusa spremenljivke, iz katerega je razvidno, katero metodo vstavljanja smo uporabili. V vsaki fazi procesa moramo biti sposobni ločiti sporočene podatke od podatkov, ki so rezultati statističnih ocen.
- Pri pripravi programske opreme za izvedbo postopkov vstavljanja je treba v čim večji meri uporabiti že obstoječe splošne programske rešitve.
- Zagotoviti je treba izračunavanje kazalnikov kakovosti, ki ponujajo vpogled v avtomatizirane postopke vstavljanja. Izračunavajo naj se tako kazalniki na mikroravni kot tudi kazalniki na makroravni. Posebno skrbno je treba spremljati delež vstavljenih podatkov in vpliv postopkov vstavljanja na končne rezultate.
- Ustreznost uporabljenih metod vstavljanja je treba stalno preverjati in analizirati. Če analize pokažejo, da uporabljene metode niso (več) ustrezne, je treba postopek ustrezno prilagoditi.

5.5 UTEŽEVANJE

Pri večini raziskovanj pridobimo podatke o želenih lastnostih populacije le od dela te populacije (tudi če gre za popis oz. popolni zajem, nam prav vse enote navadno ne odgovorijo), kar pomeni, da moramo postopke za izračun statističnih ocen prilagoditi, saj nimamo podatkov za proučevane spremenljivke za celotno populacijo. Uteži dajejo nekaterim elementom v vzorcu večji pomen kot drugim. Razlogi so različni – neenaka verjetnost izbora, neodgovor, prilagoditev na populacijske vrednosti. Postopek uteževanja določata predvsem vzorčni načrt ter pomožne populacijske spremenljivke, ki jih imamo na voljo.

Omejimo se na verjetnostne vzorce, na podlagi katerih lahko uporabimo statistično sklepanje in znamo oceniti natančnost naših ocen. Vsaka enota vzorčnega okvira ima vnaprej znano neničelno verjetnost, da bo izbrana v vzorec. Obratno vrednost verjetnosti izbora imenujemo vzorčna utež. Vzorčna utež torej pove, koliko enot v okviru predstavlja (zastopa) izbrana enota vzorca.

Če gre za verjetnostni vzorec imenujemo formule za izračun statističnih ocen statistike. Eden najpogostejših načinov prilagajanja zaradi neodgovora je uteževanje. Pri uteževanju vsak odgovor pomnožimo z nekim pozitivnim številom, ki je večje ali enako 1. To število imenujemo utež zaradi neodgovora. Če so vse enote vzorca odgovorile, potem so uteži zaradi neodgovora enake 1. Pred samim postopkom uteževanja moramo najprej opraviti analizo neodgovora, s katero skušamo ugotoviti, ali se enote, ki niso odgovorile, razlikujejo od enot, ki so odgovorile, in od neustreznih enot (npr. ali nam neki poseben del populacije ni odgovoril na vprašalnik). Pri raziskovanih podjetjih in drugih poslovnih subjektov, oseb, gospodinjstev in kmetij največkrat predpostavimo, da se neodgovori ne razlikujejo od odgovorov in neustreznih enot; z drugimi besedami: da je struktura ustreznih in neustreznih enot v populaciji (in med neodgovori) enaka strukturi enot, o katerih podatke imamo (smo jih pridobili).

Predpostavka o dvofaznem vzorčnem načrtu

Pri raziskovanjih podjetij in drugih poslovnih subjektov ter raziskovanjih kmetij izračun uteži zaradi neodgovora temelji na predpostavki o dvofaznem vzorčnem načrtu. V prvi fazi iz okvira izberemo vzorec velikosti n , v drugi fazi pa iz tega vzorca izberemo vzorec odgovorov in neustreznih enot. V prvi fazi ima vsaka enota iz okvira verjetnost, da bo izbrana v vzorec $\pi = \frac{n}{N}$ in vzorčno utež $w = \frac{N}{n}$. V drugi fazi pa iz tega vzorca izberemo vzorec odgovorov in neustreznih enot. Za vsako enoto, izbrano v tej drugi fazi, obstaja verjetnost, da bo izbrana.

$$\frac{\text{število odgovorov} + \text{število neodgovorov}}{n}$$

in utež zaradi neodgovora

$$\frac{n}{\text{število odgovorov} + \text{število neustreznih}}.$$

Pri raziskovanjih oseb in gospodinjstev pa utež zaradi neodgovora izračunamo po formuli

$$\frac{\text{število odgovorov} + \text{število neodgovorov}}{\text{število odgovorov}}.$$

V praksi želimo ocene izboljšati (in s tem povečati njihovo natančnost oz. zmanjšati njeno vzorčno varianco v primerjavi z vzorčno varianco pri uporabi vzorčne uteži). Izboljšamo jih tako, da uporabimo tako imenovani faktor: vzorčne uteži pomnožimo z nekim številom, ki ga imenujemo kalibracijski popravek. Izračunamo ga na podlagi informacij o populaciji iz drugih virov. S kalibracijo želimo vzorčne uteži popraviti tako, da če jih uporabimo za uteževanje podatkov za neko pomožno spremenljivko iz okvira, potem dobimo natančno populacijsko vrednost te pomožne spremenljivke.

Najpogosteje uporabljene kalibracijske metode so:

- **Razmernostni popravek**
Če je pomožna spremenljivka (označimo jo z x) v dovolj dobri korelaciji (vsaj 0,5) s proučevano spremenljivko, razmernostni popravek izračunamo tako, da populacijsko vsoto pomožne spremenljivke x delimo z uteženo vsoto vrednosti pomožne spremenljivke za enote vzorca, pri čemer za uteži uporabimo vzorčne uteži w_i . Razmernostni popravek izračunamo po formuli

$$w_{\text{razm}} = \frac{\sum_{k=1}^N x_k}{\sum_{k=1}^n \left(\frac{1}{\pi_k}\right) x_k}, \text{ kjer je } N \text{ velikost okvira in } n \text{ velikost vzorca.}$$

Razmernostni popravek se najpogosteje uporablja pri raziskovanjih podjetij in drugih poslovnih subjektov, saj imamo pri teh raziskovanjih zelo veliko pomožnih spremenljivk iz administrativnih in drugih virov (npr. število zaposlenih, prihodek podjetij, plače, investicije ipd.).

- **Raking**
Velikokrat želimo izvesti kalibracijo na več pomožnih spremenljivkah (npr. spol in starostni razredi). Ta postopek se imenuje raking. Raking je iterativni postopek, pri katerem postopoma izvajamo kalibracijo najprej po eni in nato po drugi pomožni spremenljivki (in nato po tretji itd.) in postopek ponavljamo toliko časa, da so razlike zanemarljive.
- **Poststratifikacija**
Pri tem postopku poznamo populacijske vrednosti pomožnih spremenljivk za kake podpopulacije, ki jih nismo načrtovali z vzorčnim načrtom (na primer zato, ker nismo imeli dovolj informacij v vzorčnem okviru). Pri poststratifikaciji preračunamo uteži na vsaki taki podpopulaciji (oz. poststratumu) posebej, in sicer tako, da so utežene ocene vsote pomožne spremenljivke enake njeni pravi vrednosti na vseh podpopulacijah.

V praksi je uteževanje lahko precej zapleteno, saj pogosto uporabljamo kombinacijo različnih tipov uteži (verjetnost izbora, neodgovor, prilagajanje na populacijske vrednosti).

Pri določenih raziskovanjih se za ocenjevanje lastnosti celotne populacije uporabljajo uteži, ki jih izračunamo iz vrednostnih podatkov drugih raziskovanj (raziskovanje o porabi v gospodinjstvih, vrednosti končne potrošnje prebivalstva, vrednost uvoženega in izvoženega blaga, vrednost domače proizvodnje, itd.). Uteži so pomemben element pri izračunu agregiranih indeksov, za izračun katerih se uporablja Laspeyresov indeks fiksnega tipa, kar

pomeni, da je izračunan z utežmi iz istega baznega ponderacijskega obdobja in z enakim zajemom opazovanih enot.

Smernice za zagotavljanje kakovosti

- Če je le mogoče, raziskovanje izvedemo na verjetnostnem vzorcu, saj lahko le na podlagi verjetnostnega vzorčenja uporabimo statistično sklepanje in posledično izračunamo natančnost statističnih ocen.
- Pri uteževanju je treba posebno pozornost nameniti osamelcem, to je enotam, ki imajo zelo veliko (ali zelo majhno) vrednost proučevane spremenljivke glede na druge enote v vzorcu. Če utemeljeno domnevamo, da je taka enota osamelec tudi na ravni celotne populacije (ali pripadajočega stratuma), potem ji utež popravimo na vrednost 1.
- Vsota končnih uteži mora biti vsaj približno enaka številu enot v populaciji (ali na domenah populacije).
- Pri uporabi kalibracijskih metod se priporoča, da se popravljene uteži ne razlikujejo bistveno od vzorčnih uteži.
- Če je stopnja neodgovora zelo visoka, potem je priporočljivo izvesti dodatno raziskovanje na vzorcu enot, ki določajo neodgovore, in na podlagi analize tega vzorca preračunati vzorčne uteži.
- Če nimamo na voljo nobene pomožne informacije, potem končne uteži, pri katerih upoštevamo neodgovor, izračunamo s pomočjo predpostavke o dvofaznem vzorčnem načrtu.
- Če imamo pomožne informacije, ki so v dobri korelaciji s proučevano spremenljivko, potem lahko neodgovor analiziramo. Rezultati analize nam pomagajo določiti skupine, v katerih se enote obnašajo podobno, in predpostavko dvofaznega vzorčnega načrta izvedemo na teh skupinah ali pa uporabimo metodo kalibracije, s katero vpliv neodgovora odpravimo.
- Pri raziskovanjih imamo seznam statusov za vse enote vzorca; ta je v grobem ločen na tri skupine: statusi odgovora, statusi neodgovora in statusi neustreznosti.
- Če je kakšna enota dodana v adresar po fazi izbora enot, mora biti to zabeleženo, prav tako razlogi za to.
- Če enota sporoči, da pripada skupini, ki ni del stratuma, v katerem je bila izbrana, a je ta skupina še vedno del opazovane populacije (npr. drugega stratuma), potem običajno izračunavamo uteži na ravni stratuma, v katerem je bila enota izbrana, statistike pa objavljamo na ravni skupine, o kateri je taka enota poročala.

5.6 IZRAČUN STATISTIČNIH OCEN (AGREGACIJA)

S pojmom izračun statističnih ocen ali agregacija označujemo tisti del procesa, pri katerem iz končnih mikropodatkov izračunamo populacijske ocene, ki jih na kratko imenujemo tudi statistike. Ker v večini raziskovanj ne opazujemo celotne populacije, ampak le njen del (verjetnostni vzorec), pri izračunu statistik običajno upoštevamo tudi preračun na populacijske vrednosti ali izračun populacijskih ocen. Postopek ali funkcijo, s katerima preračunamo podatke iz vzorca na populacijsko oceno, imenujemo cenilka.

V splošnem obstaja mnogo različnih vrst agregatov (npr. vsota, povprečje, razmerje) in cenilk. Nekaj prevladujočih je predstavljenih v nadaljevanju. Pri tem predpostavljamo, da ima vsaka enota vzorca že izračunano ustrezno populacijsko utež (w_i) in da je bil postopek izračuna uteži izveden tako, da je bil

upoštevani uporabljeni vzorčni načrt. Predpostavimo, da so $\{y_i\}_{i=1}^n$ na vzorcu izmerjene vrednosti proučevane spremenljivke Y . Populacijsko oceno označimo z \hat{Y} .

- **Populacijska vsota.** Oceno populacijske vsote spremenljivke Y izračunamo po formuli $\hat{Y} = \sum_{i=1}^n w_i \cdot y_i$.
- **Populacijsko povprečje.** Oceno populacijskega povprečja spremenljivke Y izračunamo po formuli

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

- **Število enot z določeno lastnostjo.** Naj bo D_i spremenljivka, ki za vsako opazovano enoto zavzame le dve vrednosti: vrednost 1 pripišemo, če enota ima opazovano lastnost, vrednost 0 pa, če enota te lastnosti nima. Število enot z določeno lastnostjo izračunamo po formuli $\hat{Y} = \sum_{i=1}^n w_i \cdot D_i$.

- **Delež enot z določeno lastnostjo.** Populacijski delež enot z opazovano lastnostjo ocenimo po formuli

$$\hat{p} = \frac{\sum_{i=1}^n w_i \cdot D_i}{\sum_{i=1}^n w_i}, \text{ pri čemer je spremenljivka } D_i \text{ definirana kot v prejšnji točki.}$$

- **Razmerje populacijskih vsot.** Razmerje populacijskih vsot dveh spremenljivk (X,Y) ocenimo po formuli:

$$\hat{R} = \frac{\sum_{i=1}^n w_i \cdot y_i}{\sum_{i=1}^n w_i \cdot x_i}.$$

Če oceno populacijske vrednosti izbrane statistike izračunamo na podlagi podatkov, pridobljenih z vzorcem, takšna ocena vsebuje tudi vzorčno napako. Pri verjetnostnem vzorcu je treba to napako oceniti in jo na primeren način prikazati pri objavi rezultatov. Postopki za izračun vzorčne napake so lahko precej kompleksni, določata pa jih predvsem vrsta cenilke ter uporabljeni vzorčni načrt. V grobem lahko postopke za oceno vzorčne napake razdelimo v tri skupine:

- uporaba direktnih formul za oceno standardnih napak linearnih cenilk ter uporaba »približnih formul« Taylorjeve linearizacije za nelinearne cenilke (analitični pristop);
- ocena standardne napake na podlagi izbire podvzorcev in nato uporaba ustreznih formul za tako ravnanje (ponovno vzorčenje);
- uporaba modela za oceno standardnih napak ocen na podskupinah obravnavane populacije (model).

Obstaja tudi več načinov, kako ocenjene vzorčne napake predstavimo v objavah. Uporabljata se dva načina: vzorčna napaka se eksplicitno (vrednostno) objavi skupaj z ocenjeno statistiko ali pa se manj natančne ocene označijo s posebnimi oznakami.

Smernice za zagotavljanje kakovosti

- Pri izračunu populacijskih ocen na podlagi vzorca je treba izbrati cenilke, ki so (čim bolj) nepristranske.
- Prizadevati si je treba, da se izračun ocen ter izračun vzorčnih napak teh ocen izvedeta v istem procesu. Tako skupaj z ocenjeno statistiko dobimo na razpolago tudi ocenjeno vzorčno napako ter posledično oceno tega, kako zanesljiva je ocenjena statistika.
- Pri pripravi postopka za izračun vzorčne napake je treba vedno upoštevati uporabljeni vzorčni načrt.
- Pri pripravi navodil za prikaz vzorčne napake je treba upoštevati splošno sprejete standarde.
- Vzorčne napake je treba redno analizirati, predvsem zaradi zaznavanja ocen s premajhno natančnostjo (s preveliko vzorčno napako). Če analize pokažejo, da so izračunane ocene premalo natančne, je treba uvesti ukrepe za zmanjšanje vzorčnih napak. Taki ukrepi gredo lahko v smer povečanja vzorca (če razpoložljiva sredstva to dopuščajo), spremembe alokacije pri izboru vzorca ali pa v smer definiranja učinkovitejše cenilke (predvsem z uporabo pomožnih spremenljivk).

5.7 DEFLACIJA

Deflacija v statističnem procesu označuje postopek, pri katerem iz vrednostnih podatkov izločimo vpliv spremembe cen v določenem obdobju. Na teoretski ravni si vedno lahko predstavljamo, da je vrednostni podatek sestavljen iz produkta dveh komponent: iz količine (oz. obsega) in iz cene. Ko računamo razmerje vrednostnega podatka v dveh časovnih točkah, sta torej v izračunano razmerje zajeti sprememba količine in sprememba cene. Če hočemo na podlagi izmerjene spremembe vrednosti priti do čim natančnejše ocene spremembe količine, je iz vrednostnega razmerja (indeksa) treba odstraniti vpliv spremembe cen. Indeks, ki se izračuna iz »originalnih« individualnih podatkov, imenujemo nominalni indeks, indeks, ki se izračuna iz podatkov, iz katerih smo izločili vpliv spremembe cen, pa realni indeks.

Pri izvedbi postopka deflacije lahko izberemo dva osnovna pristopa. Prvi je postopek deflacije na mikro-, drugi pa postopek deflacije na makroravni. Pri deflaciji na mikroravni najprej vsak podatek na individualni ravni delimo z ustreznim deflatorjem in tako (v smislu spremembe cen) vse podatke preračunamo na isto (fiktivno) časovno točko. Če na primer kot deflator uporabimo indeks cen življenjskih potrebščin na fiksno bazo leto (povprečje leta

2005), potem vsak preračunan individualni vrednostni podatek predstavlja vrednost, ki bi (s takrat veljavnimi cenami) veljala v fiktivni časovni točki povprečje leta 2005.

Pri deflaciji na makroravni izvedemo postopke »odstranitve« cen na že izračunanih agregatih oziroma indeksih. Od izračuna na mikroravni se razlikuje predvsem v tem, da tu potrebujemo deflator na isto obdobje, kot je indeks, ki ga deflacijemo. Če deflacijemo indeks tekoči mesec glede na pretekli mesec, potrebujemo tudi deflator (torej ustrezen indeks cen) tekoči mesec glede na pretekli mesec, če pa na primer deflacijemo indeks na fiksno bazno leto, potrebujemo tudi deflator na fiksno bazno leto (npr. povprečje leta 2005). Pri deflaciji na mikroravni tako potrebujemo le eno časovno vrsto deflatorjev, pri deflaciji na makroravni pa toliko časovnih vrst deflatorjev, kolikor je časovnih vrst, ki jih želimo deflacijirati. To je tudi največja prednost deflacije na mikroravni. Prednost deflacije na makroravni pa je v večji »robustnosti« postopka, saj je deflacija na makroravni zaradi manjšega števila podatkov, ki jih obdelujemo, za obdelavo precej manj zahtevna.

Smernice za zagotavljanje kakovosti

- Pri odločitvi o tem, ali se uporabijo postopki deflacije na mikroravni ali na makroravni, je treba pretehtati vse prednosti in slabosti obeh pristopov. Izbrati je treba predvsem tak pristop, ki bo omogočal enostavnejšo in racionalnejšo umestitev postopka v celoten statistični proces in ki vsebinsko najbolj ustreza pojavu.
- Pri določitvi postopkov za pripravo deflatorjev je v prvem koraku treba preveriti, kateri indeksi cen so najprimernejši kot osnovni podatek za pripravo deflatorjev. V drugem koraku je treba preveriti, ali je sam indeks cen že primeren deflator ali pa ga je treba kombinirati z drugimi, strukturnimi podatki. Cilj je, da bi se izračunani deflator čim bolj približal dejanskemu gibanju cen pojava, ki ga merimo na podlagi vrednostnih podatkov.
- Zagotoviti je treba čim enostavnejši in po možnosti avtomatiziran prevzem indeksov cen. Pri tem je treba zagotoviti tudi omejen dostop, saj so indeksi cen v nekaterih primerih (če jih npr. prevzemamo pred uradno objavo ali če prevzemamo nižje ravni od objavljenih) zelo občutljiv podatek.
- Ustreznost uporabljenega deflatorja je treba stalno preverjati.

5.8 UREJANJE PODATKOV NA MAKORAVNI

Z urejanjem podatkov na makroravni iščemo potencialno napačne podatke s postopki, ki vključujejo podatke nabora več enot. V grobem lahko te postopke razdelimo v dve skupini: kontrola in analiza agregiranih podatkov ter iskanje osamelcev. Običajno se postopki urejanja na makroravni v procesu izvajanja raziskovanja umeščajo za fazo urejanja podatkov na mikroravni. Na makroravni praviloma poteka samo zaznavanje napak, popravljanje napak pa naj bi se vedno izvajalo na mikroravni.

V obeh skupinah postopkov urejanja na makroravni obstaja veliko različnih metod. V nadaljevanju podajamo kratek opis nekaj teh metod, ki jih v praksi uradne statistike največkrat uporabljamo:

- **Metoda kontrole agregatov.** Osnovna zamisel metode je v tem, da z logičnimi kontrolami, prilagojenimi za makroraven, najprej preverjamo že agregirane podatke na določeni podskupini (npr. na 4-mestni ravni SKD), nato pa opravimo »klasično« logično kontrolo na mikroravni za vse enote samo iz tistih skupin, ki so bile na makroravni zaznane kot vprašljive. Na tak način lahko bistveno zmanjšamo količino urejanja na mikroravni.
- **Metoda postopnega izločanja.** Osnovna zamisel, na kateri temelji ta metoda, je v tem, da s pomočjo ustrezne računalniške aplikacije omejimo kontrolo in urejanje podatkov le na tiste enote, katerih vpliv na končno oceno v določeni podskupini ni zanemarljiv. Postopek nam prek ustrezne aplikacije izpiše enote, katerih vpliv na agregate ni zanemarljiv, in v naslednjih korakih nato kontroliramo samo vrednosti teh enot.
- **Grafične metode.** Namen različnih izvedb grafičnih metod je grafično prikazati porazdelitev uteženih vrednosti podatka, ki ga želimo kontrolirati. Grafična predstavitev porazdelitve ter ključnih parametrov porazdelitve (npr. kvartili) nam tako lahko omogoča, da meje ekstremnih vrednosti lažje določimo. Grafične metode uporabljamo predvsem kot »dopolnilno« orodje pri izvajanju drugih metod urejanja podatkov.
- **Računsko iskanje osamelcev.** Osamelce določimo kot vrednosti, ki so ali manjše od $Q1 - C * Q$ (levi osamelec) ali večje od $Q3 + C * Q$ (desni osamelec), kjer sta $Q1$, $Q3$ prvi in tretji kvartil, Q pa kvartilni razmik. Parameter C določa »strogost« merila za določitev osamelcev. Standardni vrednosti parametra C sta 1,5 za osamelce in 3 za ekstremne osamelce. Metoda je predvsem primerna za vsaj približno simetrične porazdelitve.

Smernice za zagotavljanje kakovosti

- Kontrole agregatov je treba premišljeno opredeliti in jih tudi formalno zapisati v podobni formalni obliki kot kontrole na mikroravni.

- Treba se je izogibati situaciji, da bi zaradi prestrogo postavljenih kontrol na makroravni povzročili problem pretiranega in neučinkovitega urejanja podatkov na mikroravni.
- Priporoča se uporaba grafičnih metod, saj je vizualna predstavitev običajno najučinkovitejša pot za zaznavanje odstopajočih vrednosti.
- Urejanje podatkov na makroravni naj se uporablja le za zaznavanje napačnih oziroma vprašljivih vrednosti. Vsi popravki podatkov naj se opravijo na ravni individualnega zapisa (mikroraven).
- Vse postopke je treba čim bolj avtomatizirati in s tem izboljšati učinkovitost celotnega statističnega procesa.

5.9 TABELACIJA

V procesu tabelacije se pripravijo strukturirane agregirane informacije za različne vrste objav: za elektronsko objavo, za podatkovno bazo, za tiskane publikacije, za tabele, ki ustrezajo zahtevam standardnih uporabnikov, za tabele, namenjene za mednarodno poročanje. Pripravljene tabele se lahko uporabijo tudi v okviru drugih podprocesov kot kontrolne tabele: pri urejanju na makroravni, pri statistični zaščiti podatkov, pri analizi ustreznosti in pri potrditvi rezultatov. Pomembno je, da so predvsem tabele za javno objavo rezultat izvedbe vnaprej pripravljenih računalniških postopkov, ki se lahko izvedejo avtomatsko ali tako, da jih je mogoče nadzorovati. S takim ravnanjem se povečajo možnosti za organizacijo in hranjenje vseh izračunanih statistik, zmanjša se tveganje, da bi bila zaradi časovne stiske ogrožena pravočasnost objave rezultatov, in veliki meri se izloči vpliv človeškega faktorja na pravilnost rezultatov.

Tabele morajo biti oblikovane optimalno glede na vsebino, glede na tehnologijo priprave podatkov, glede na tehnične možnosti orodij in glede na sprejete standarde objavljavanja podatkov. Postopek priprave tabel lahko poteka samostojno, večinoma pa ga je smiselno združiti z nekaterimi drugimi procesi, predvsem s statistično zaščito podatkov in z izračunom statističnih ocen.

Smernice za zagotavljanje kakovosti

- Za izvedbo tabelacij je treba uporabljati standardna navodila.
- Če se pri izdelavi tabel uporabljajo šifranti, je treba vedno uporabljati šifrante; ti se nahajajo na klasifikacijskem strežniku KLASJE.
- Tabele za javno objavo naj se oblikujejo na podlagi vnaprej pripravljenih postopkov.
- Za pripravo internih priložnostnih (ad hoc) in kontrolnih tabel naj se uporabljajo standardna interaktivna analitična orodja.
- Ker ta postopek lahko združuje različne procese in ker so vanj tako vključeni različni oddelki, je treba delovne postopke skrbno usklajevati, standardizirati in ustrezno dokumentirati.
- Pri izbiri strukture tabel je treba zagotoviti, da imajo tabele čim manj zaščitenih in praznih celic.

5.10 STATISTIČNA ZAŠČITA PODATKOV

Doseganje strateških ciljev in izpolnjevanje nalog državne statistike sta med drugim odvisni od zaupanja opazovanih enot oz. dajalcev podatkov: oseb, gospodinjestev, podjetij, kmetijskih gospodarstev in drugih organizacij. To pomeni, da nam bodo svoje podatke zaupali, če bodo prepričani, da ravnamo tako, da v celoti zagotavljamo njihovo zaupnost (tj. varujemo njihovo identiteto) in zaupnost danih informacij, in da bodo pridobljeni podatki uporabljeni samo za statistični namen. S tem namenom jih tudi nenehno obveščamo o tem, jim razlagamo uporabljene postopke, in temu je namenjeno tudi to metodološko pojasnilo.

Del zaščite (varovanja) zaupnosti podatkov temelji na tem, da izkazani podatki ne omogočajo neposredne identifikacije (s pomočjo direktnih identifikatorjev) ali posredne identifikacije (na kateri koli drug način). K temu nas zavezujeta tako slovenska kot tudi evropska zakonodaja. Glede na to, da je poslanstvo državne statistike posredovati in objavljati statistične rezultate v čim širšem obsegu, vendar hkrati tako, da je tveganje razkritja informacij o enotah minimalno, so potrebne metode statistične zaščite, s katerimi dosežemo, da se spoštuje zakonodaja in obenem ohranja zadovoljiva oz. čim večja obvestilna (informacijska) vrednost posredovanih oziroma objavljenih statističnih rezultatov.

Ločimo dve vrsti metod za statistično zaščito podatkov:

- **metode, uporabljene na osnovnih mikropodatkih:**
 - metode, ki spreminjajo podatke (podatki naj bi se z uporabo take metode spremenili čim manj, vendar v tolikšni meri, da je tveganje razkritja najmanjše možno):
 - statistična zaščita mikropodatkov: metoda PRAM, dodajanje šuma, zaokroževanje, menjava podatkov;

- statistična zaščita tabel: zaokroževanje, metoda CTA;
- metode, ki ne spreminjajo podatkov (podatki imajo manjšo informacijsko vrednost, vendar so še vedno točni):
 - statistična zaščita mikropodatkov: metoda manjkajočih vrednosti, združevanje v razrede, kodiranje vrha, kodiranje dna, mikroagregacija, vzorčenje;
 - statistična zaščita tabel: metoda manjkajočih vrednosti, združevanje v razrede;
- **metode za generiranje sintetičnih podatkov** (podatke dobimo z upoštevanjem statističnega modela).

Statistična zaščita tabel

Vsako posredovano statistiko si lahko predstavljamo kot tabelo dimenzije 1 x 1, saj individualnih podatkov nikoli ne posredujemo. V nadaljevanju se bomo osredotočili na tabele, kakršne poznamo, saj je večina agregiranih podatkov uporabnikom na voljo v tabelarni obliki.

Tabela (preglednica) je podatkovna struktura, ki realizira matematično strukturo matrike in razvrsti podatke v celice po vrsticah in stolpcih. V njej so prikazani agregirani podatki (različne statistike), ki opisujejo združene podatke več enot. Spremenljivke, ki napenjajo tabelo, se imenujejo pojasnjevalne spremenljivke. Število pojasnjevalnih spremenljivk določa število dimenzij tabele.

Pri postopku statistične zaščite tabel določimo pravila za določanje primarno občutljivih celic oz. vrednosti, ki so prikazane v celicah tabele. Od tipa tabele je odvisno, katera pravila bomo uporabili. Ločimo tri tipe tabel:

- **frekvenčne tabele** (v vsaki celici tabele je vpisano število enot, ki pripadajo tej celici);
- **vrednostne tabele** (v vsaki celici je vpisana vsota vrednosti določene spremenljivke enot, ki pripadajo tej celici);
- **druge tabele** (v celicah so vpisane vrednosti drugih tipov statistik – deležev, razmerij, indeksov in drugih), ki so običajno povezane z vrednostnimi in frekvenčnimi tabelami.

Pravila za določanje primarno občutljivih celic so:

- **prag** (če je število enot manjše od praga, potem je celica primarno občutljiva) – minimalna vrednost za prag naj bo 3;
- **pravilo dominantnosti** (če n največjih prispevkov prispeva več kot $k\%$ vrednosti celice, potem je celica primarno občutljiva);
- **p%-pravilo** (če lahko enota z drugim največjim prispevkom oceni največji prispevek bolje kot na $p\%$ natančno, potem je celica primarno občutljiva).

Pri frekvenčnih tabelah upoštevamo le prag, medtem ko pri vrednostnih in drugih tabelah upoštevamo prag in p%-pravilo ali pa prag in pravilo dominantnosti. Pravila za določanje primarno občutljivih celic so zaupna in se ne smejo posredovati, saj bi z razkritjem parametrov zmanjšali zaščito enot.

Če se določene enote strinjajo (pridobimo njihovo pisno dovoljenje), da se njihovi individualni podatki objavijo, ali če je del podatkov pridobljen iz javnega vira, se lahko uporabi t. i. pravilo zahteve. To pomeni, da tem enotam dodelimo poseben status in tako zmanjšamo izgubo informacije v tabelah, v katerih se pojavljajo.

Z metodo za statistično zaščito tabel primarno občutljive celice vedno zaščitimo. Zaradi povezav znotraj tabele pa najpogosteje zaščitimo še dodatne celice (pravimo jim sekundarno zaščitene celice); če povezav ni, to ni potrebno. Pri metodi manjkajočih vrednosti zakrijemo primarno in sekundarno občutljive vrednosti tako, da jih ne objavimo, tj. jih ne vpišemo v celico, temveč namesto vrednosti vpišemo črko „z“. Izgubo informacije določa delež zakritih celic.

Pri statistični zaščiti tabel je treba paziti na hierarhije pojasnjevalnih spremenljivk, število dimenzij tabel in povezave med tabelami.

Statistična zaščita mikropodatkov

Mikropodatke (individualne podatke) je treba zaščititi zaradi velikega števila spremenljivk, ki so v datoteki mikropodatkov, saj so nekatere kombinacije spremenljivk redke in zato je tveganje, da se razkrije identiteta enote, veliko.

V procesu statistične zaščite mikropodatkov moramo vedeti, kateremu tipu uporabnikov bodo zaščitene datoteke mikropodatkov namenjene. Če je taka datoteka namenjena raziskovalcem, potem bo izguba

informacije manjša, saj enote zaščitimo le pred nenamernim razkritjem. V takem primeru raziskovalec podpiše tudi pogodbo, s katero se zaveže k varovanju podatkov, ki jih bo prejel. Če je datoteka namenjena javnosti (v to skupino štejemo tudi študente), potem bo izguba informacije večja, saj enote zaščitimo tudi pred namernim razkritjem. V takem primeru je datoteka na voljo na spletni strani ponudnika podatkov, ne da bi bilo treba podpisati pogodbo.

Pri procesu statistične zaščite mikropodatkov moramo najprej izbrati statistike, katerih vrednosti bomo skušali čim manj spremeniti. V naslednjem koraku je treba določiti 3 nabore spremenljivk:

- **identifikacijske spremenljivke:** spremenljivke, ki omogočajo razkritje enote:
 - direktni identifikatorji: spremenljivke, ki enolično določajo enote (npr. davčna številka, matična številka podjetja), omogočajo neposredno identifikacijo;
 - indirektni identifikatorji: spremenljivke, katerih kombinacije lahko privedejo do razkritja enote (npr. ime, naslov, spol, starost, regija), omogočajo posredno identifikacijo;
- **zaupne izhodne spremenljivke:** spremenljivke, ki nosijo občutljivo informacijo o enoti (npr. prihodek, vera, politično prepričanje, zdravstveno stanje ...);
- **nezaupne izhodne spremenljivke:** druge spremenljivke.

Redke kombinacije preostalih identifikacijskih spremenljivk so tiste, ki jih je treba zaščititi, saj obstaja velika verjetnost, da bo enota, na katero se ta redka kombinacija nanaša, lahko prepoznana. Nato določimo sprejemljivo tveganje razkritja (verjetnost prepoznave pri točno določenem scenariju). Za to imamo na voljo več metod. Omenili bomo dve največkrat uporabljeni:

- na podlagi števila enot, ki posamezni kombinaciji identifikacijskih spremenljivk pripadajo (ta metoda je uporabna, če imamo v datoteki celotno populacijo);
- z zahtevnejšo matematično metodo, pri kateri določimo statistični model za populacijo na podlagi vzorčnih podatkov, ki so v datoteki.

Določimo število enot, ki jih uporabnik lahko potencialno še vedno prepozna. Zavedati se moramo, da kljub statistični zaščiti tveganje razkritja ni nikoli ničelno.

V naslednjem koraku določimo metode statistične zaščite in jih uporabimo na podatkih. Direktne identifikatorje in določene indirektno identifikatorje, tj. take, katerih kombinacija z veliko verjetnostjo enolično določa enoto (npr. ime, priimek, naslov), moramo odstraniti. Preostali indirektni identifikatorji so spremenljivke, ki jih v podatkih obdržimo, saj so pomembne za statistično analizo (npr. starost, spol, statistična regija). Te spremenljivke imenujemo ključne spremenljivke. Izberemo metode, ki jih bomo uporabili na ključnih spremenljivkah, da bo verjetnost posredne identifikacije sprejemljiva. Metode se izberejo glede na namen podatkov (npr. datoteka za javnost, datoteka za raziskovalce). Zaščitimo lahko tudi zaupne izhodne spremenljivke. Nekatere metode so uporabne le za zvezne spremenljivke (npr. dodajanje šuma), druge le za znakovne spremenljivke (npr. metoda manjkajočih vrednosti); obstajajo pa tudi metode, ki so primerne za obe vrsti spremenljivk (npr. združevanje v razrede). Metode uporabimo v taki meri, da zmanjšamo tveganje razkritja enot v zaščiteni datoteki do sprejemljive (dopustne) stopnje. Uporabniku zaščitenih podatkov pojasnimo, katere metode smo uporabili na podatkih, saj se mora zavedati, da so določeni podatki spremenjeni (zmoteni) in da se vrednosti v zaščiteni datoteki razlikujejo od vrednosti v originalni datoteki. Ne smemo pa razkriti parametrov metod (npr. nabora indirektnih identifikatorjev, praga pri metodi manjkajočih vrednosti), saj bi z razkritjem parametrov zmanjšali zaščito enot.

Smernice za zagotavljanje kakovosti

Statistična zaščita tabel

- Povezave med tabelami (pojasnjevalnimi spremenljivkami in statistikami), ki jih ščitimo, je treba podrobno preučiti. Upoštevati je treba vse objavljene in posredovane tabele, ki so ali bodo vsebovale iste podatke (npr. tabele za nacionalno objavo in tabele, posredovane na Eurostat).
- Število dimenzij tabele naj bo manjše od 5.
- Razredi pojasnjevalne spremenljivke naj se ne definirajo podrobneje, kot je to potrebno, saj se s tem lahko močno vpliva na izgubo informacije.
- Če se odločamo med p%-pravilom in pravilom dominantnosti (2,k), je prvo primernejše zaradi naravnejše definicije primarne občutljivosti.
- Pri vrednostnih tabelah moramo obvezno uporabiti pravilo dominantnosti ali p%-pravilo. Če uporabimo to pravilo (pravilo dominantnosti), moramo obvezno gledati tudi dominantnost prvih dveh največjih enot.

- Minimalni prag je 3 (če bi bila namreč 1 sama enota, bi razkrili njen podatek, če bi bili 2, bi lahko razkrili druga drugo).
- Pri vzorčnem raziskovanju moramo upoštevati uteži.
- Javnih podatkov ne ščitimo.

Statistična zaščita mikropodatkov

- Vrsta uporabnikov zaščitene datoteke mikropodatkov mora biti točno določena (javnost, študenti, raziskovalci).
- Treba je določiti, katere statistike so najpomembnejše, in jih skušamo čim manj spremeniti.
- Število spremenljivk v datoteki mikropodatkov naj ne bo večje, kot je to potrebno za načrtovane analize.
- Povezave med spremenljivkami v datoteki morajo biti točno specificirane.
- Pazljivo je treba določiti ključne spremenljivke, da katere ne zgrešimo, in da s tem ne povečamo možnosti razkritja enot.
- Javnih podatkov ne ščitimo.

6 ANALIZA PODATKOV

Statistična analiza podatkov je postopek, pri katerem podatke analiziramo z različnimi orodji in tehnikami, in sicer z namenom, da bi preučili in pojasnili stanja in dogajanja, izluščili določene zakonitosti, ki so značilne za opazovani pojav, ter povzeli in interpretirali rezultate. Običajno najprej analiziramo podatke na makroravni, po potrebi pa tudi na mikroravni.

Podatke analiziramo zato, da z analizo potrdimo njihovo ustreznost oziroma odkrijemo morebitne pomanjkljivosti in jih nato odpravimo ter s tem izboljšamo kakovost podatkov. Če analiza pokaže sistemske pomanjkljivosti, izsledke uporabimo za izboljšanje kakovosti procesov ali za dopolnitev ali celo spremembo metodologije. Podatke analiziramo zato, da lahko korektno interpretiramo in prikazujemo stanja in dogajanja na področjih, ki jih opazujemo, ali celo ugotavljamo določene zakonitosti, ki se v okviru opazovanega področja pojavljajo.

Na makroravni lahko analiziramo podatke, vezane na določeno časovno točko (različne strukture), ali pa časovne vrste podatkov (običajno indeksi). Če pri analizi podatkov na makroravni ugotovimo bistvene odmike od pričakovanih vrednosti, a jih na tej ravni ne moremo pojasniti, izvedemo še analizo na mikroravni in na tej ravni podrobneje analiziramo tiste mikropodatke, iz katerih so bili izvedeni rezultati na makroravni, pri katerih smo ugotovili največje odmike oziroma pomanjkljivosti.

Analiza podatkov obsega tudi analizo vpliva različnih postopkov, ki jih izvajamo pri statistični obdelavi podatkov, na podatke same. Na primer: analiziramo vpliv neodgovora in vstavljanja podatkov na končne rezultate, vpliv primernosti izbranega postopka za vstavljanje podatkov, razlike med začasnimi in končnimi podatki in podobno. Vsi ti izsledki so osnova za opis kakovosti podatkov in tudi osnova za dopolnjevanje in izpopolnjevanje metodologije raziskovanja.

Pri analizi podatkov uporabimo tudi druge vire podatkov in informacij, ki se neposredno ali posredno nanašajo na področje, ki ga obravnavamo, tako da še na podlagi zunanjega vira preverimo kakovost podatkov.

Preden začnemo statistične pojave pojasnjevati, razlagati, moramo z analizo podatkov ugotoviti, kaj podatki kažejo. Podatke pretvorimo v uporabne informacije, ki jih kasneje praviloma objavimo. V tej fazi procesa se moramo spraševati, kaj podatki kažejo, ter z analizo in interpretacijo podatkov odgovarjati na ta vprašanja.

6.1 ANALIZA ČASOVNIH VRST

Časovna vrsta je časovno urejeno zaporedje podatkov, v našem primeru statističnih podatkov (npr. indeks industrijske proizvodnje, indeks stroškov dela). Poznamo različne časovne vrste: glede na periodiko npr. ločimo mesečne, četrletne, letne časovne vrste; nekatere časovne vrste so trenutne (podatek je vrednost v danem trenutku), nekatere pa intervalne (podatek je agregirana vrednost v danem časovnem intervalu). Področje statistike, ki proučuje časovne vrste, se imenuje analiza časovnih vrst.

Časovno vrsto analiziramo zato, da opišemo, razložimo, napovemo ali nadzorujemo neki pojav. Pri analizi časovne vrste nas najprej zanima ozadje časovne vrste (to so vse informacije o časovni vrsti, ki jih lahko dobimo). Nato narišemo graf podatkov v odvisnosti od časa. Na njem hitro opazimo dolžino časovne vrste, ali kak podatek manjka, izstopajoče vrednosti (osamelce, probleme in napake v podatkih), ali podatki 'v povprečju' naraščajo, padajo, ostajajo enaki (trend), kako so podatki odvisni od obdobja v letu (sezonska nihanja). Potem podatke uredimo, če je potrebno: odkrijemo in popravimo očitne napake, vstavimo (imputiramo) manjkajoče podatke, spremenimo osamelce idr. Vprašamo se tudi, ali so bile merjene najprimernejše spremenljivke in ali so bile merjene primerno natančno.

Pri modeliranju je pomembno, da statistično teorijo združimo z zdravo pametjo in z znanjem o pojavu. Statistiki, kot so povprečje, mediana, modus, standardni odklon ipd., navadno ne računamo za časovno vrsto, saj so lahko zelo zavajajoče. Lahko pa izračunamo npr. časovno vrsto rasti. Za intervalne časovne vrste je smiselno izračunati časovno vrsto kumulativ in časovno vrsto vsot. Večinoma pa želimo časovno vrsto razčleniti na komponente (npr. pri desezoniranju). Pri modeliranju se odločimo, za katere vplive gre v časovni vrsti, pri čemer si pomagamo z ozadjem. Od namena analize časovne vrste pa je odvisno, kako vplive modeliramo ali jih morda odstranimo. Na časovno vrsto lahko vpliva več dejavnikov: zelo pogosto opazimo vplive sezone (ljudje smo podvrženi različnim naravnim in socialnim ritmom), pogosto pa tudi vplive koledarja (med te spadajo vpliv števila delovnih dni, vpliv prestopnega leta, vpliv praznikov, vpliv velike noči). Da lahko uporabimo teorijo časovnih vrst, moramo včasih časovno vrsto transformirati, da npr. stabiliziramo varianco, dobimo stacionarno časovno vrsto. Uradna statistika se ukvarja v glavnem z desezoniranjem mesečnih in četrletnih časovnih vrst. Časovno vrsto desezoniramo tako, da ji postavimo model v programu za desezoniranje (metoda TRAMO/SEATS). Lahko

izberemo logaritemsko transformacijo, določimo osamelce, vplive koledarja, ARIMA-model ... Glede na izbrani model program razčleni časovno vrsto na komponente:

- komponenta trend-cikel (sestavljajo jo trend (dolgoročno gibanje povprečja časovne vrste) in ciklična nihanja s periodo, daljšo od enega leta);
- sezonska komponenta (sestavljajo jo vplivi sezone in vplivi koledarja);
- iregularna komponenta (sestavljajo jo slučajna nihanja in nekateri osamelci).

Če smo izbrali logaritemsko transformacijo, potem je časovna vrsta enaka produktu svojih komponent (multiplikativni model), sicer pa vsoti (aditivni model).

Osamelci so večje spremembe vrednosti, ki v časovni vrsti še posebej izstopajo. Poznamo več tipov osamelcev, najpogostejši pa so:

- aditivni osamelec (časovna vrsta nenadoma precej opazno naraste ali pade, potem pa se takoj vrne na prejšnjo vrednost, tako da se vpliv pozna samo pri enem podatku);
- prehodna sprememba (vrednosti v časovni vrsti nenadoma narastejo ali padejo, potem pa se vrsta počasi vrne v stanje pred motnjo);
- sprememba ravni (pri povprečni vrednosti časovne vrste pride v neki točki do preloma – spremeni se raven časovne vrste, in ta sprememba ostane).

Rezultati analize so odvisni od programske opreme, metode in izbranega modela.

Če iz časovne vrste odstranimo sezonsko komponento, dobimo desezonirano časovno vrsto. Ta postopek imenujemo desezoniranje. Z desezoniranjem podatke poenostavimo, da jih ustrezneje interpretiramo, saj sezonska nihanja lahko zameglijo druga pomembna gibanja. Večino časovnih vrst desezoniramo neposredno (direktno; to pomeni, da imajo v programu za desezoniranje postavljen model). Za časovno vrsto, ki jo desezoniramo posredno (indirektno), pa rezultate desezoniranja preračunamo iz rezultatov desezoniranja drugih časovnih vrst.

Ko spremenimo originalne podatke (dodamo nov podatek, popravimo stare podatke), časovno vrsto potem na novo desezoniramo. S temi podatki znova ocenimo parametre modela. Temu načinu osveževanja rečemo delna sprotna prilagoditev. Sprememba v originalnih podatkih povzroči, da se nekoliko spremenijo tudi ocene parametrov modela in zato tudi rezultati desezoniranja vzdolž celotne časovne vrste.

Rezultati desezoniranja na koncu časovne vrste so manj zanesljivi kot drugi rezultati desezoniranja in se lahko z dodajanjem podatkov še precej spremenijo. Temu problemu rečemo problem končnih točk.

Uporabljamo in objavljamo lahko npr. naslednje podatke in rezultate:

- originalni podatki;
- desezonirani podatki so podatki z izločenimi vplivi sezone in koledarja (to so originalni podatki, iz katerih so odstranjeni vplivi sezone in koledarja);
- podatki z izločenimi vplivi koledarja (to so originalni podatki, iz katerih so odstranjeni vplivi koledarja);
- komponenta trend-cikel.

Zelo zanimivo je napovedovanje prihodnjih vrednosti. Metode za napovedovanje delimo na:

- subjektivne (pri njih uporabimo svojo presojo ali kako pomembno informacijo);
- univariatne (napovedi temeljijo na modelu, ki ga določimo znani časovni vrsti);
- multivariatne (napovedi vsaj delno temeljijo na vrednostih ene ali več napovedovalnih spremenljivk).

Izbira metode je odvisna od uporabe napovedi in značilnosti časovne vrste, od podatkov, ki so na voljo, naših veščin, izkušenj itd.

Napoved je lahko izražena kot točkovna vrednost (skupaj z njo nas pogosto zanima tudi napaka v napovedi), lahko pa tudi kot interval, na katerem prihodnja vrednost leži z neko verjetnostjo. Na splošno lahko bolje ocenimo manjkajoče vrednosti znotraj časovne vrste kot pa njene prihodnje vrednosti. Napovedi originalnih podatkov in napake v napovedih dobimo samodejno za časovne vrste, ki imajo postavljen model v programu za desezoniranje.

Smernice za zagotavljanje kakovosti

- Poznati moramo vse spremembe, povezane s časovnimi vrstami (npr. spremembe v metodologiji).
- Časovne vrste moramo skrbno pregledati, podatke urediti in jih analizirati, da bi odkril morebitne napake.
- Če želimo časovno vrsto desezonirati, mora biti dolga vsaj 3 leta (mesečna časovna vrsta) oz. 4 leta (četrtna časovna vrsta).
- Razmisliti moramo o ustreznosti modelov (npr. o značilnosti vplivov sezone in koledarja, vzrokih in tipih osamelcev, skrajšanju časovnih vrst) in rezultatov (npr. kaj nam povedo, ali imajo ustrezen predznak).
- Vse datoteke moramo ustrezno shraniti, da kasneje lahko analizo ponovimo.
- Pri desezoniranju je dobro, da približno enkrat letno podrobno pregledamo vse modele časovnih vrst, pri čemer upoštevamo časovno obdobje, ki nam je takrat na voljo. Takrat lahko izvedemo tudi večje spremembe. V obdobju med podrobnima pregledoma pa si prizadevamo, da so spremembe čim manjše. Do njih pride večinoma zaradi začasnega osamelca (to je osamelec, ki se pojavi na koncu časovne vrste in še ni dokončno vključen v model).
- Če primerjamo dva podatka v isti časovni vrsti, potem pri istem mesecu (četrtnetih) dveh različnih let (npr. januar 2009 in januar 2008) primerjamo podatke z izločenimi vplivi koledarja, sicer (npr. januar 2009 in december 2008) pa primerjamo desezonirane podatke. Desezonirane podatke uporabimo tudi pri primerjavah različnih časovnih vrst.

6.2 ANALIZA USTREZNOSTI TER POTRDITEV REZULTATOV

Analiza ustreznosti rezultatov je proces, v katerem se preverja smiselnost rezultatov, njihova notranja skladnost, skladnost in primerljivost v času in prostoru ter skladnost z obstoječimi notranjimi in zunanjimi referenčnimi viri podatkov.

Analiza ustreznosti oz. potrditev rezultatov se opravlja po procesu urejanja rezultatov na makroravni in lahko zajema naslednje postopke:

- Preverjanje notranje skladnosti rezultatov (če ta ni že vgrajena v urejanje na makroravni), npr. preverjanje rezultatov na podlagi znanih oziroma pričakovanih razmerij med rezultati (ali je npr. vrednost proizvodnje večja od dodane vrednosti).
- Preverjanje skladnosti rezultatov glede na rezultate iz preteklih referenčnih obdobj (to velja predvsem za tista raziskovanja, katerih osnovni namen ni merjenje sprememb v času).
- Preverjanje skladnosti rezultatov s sorodnimi ali s povezanimi rezultati iz drugih raziskovanj (tudi z rezultati raziskovanj, ki jih izvajajo druge institucije).
- Občasno preverjanje »ustreznosti in smiselnosti« rezultatov z zunanjimi strokovnjaki.

Smernice za zagotavljanje kakovosti

- Pred končno potrditvijo rezultatov je treba še enkrat analizirati vse procesne podatke (podatke, ki so na voljo v zvezi z izvajanjem procesa obdelave podatkov), predvsem rezultate urejanja na makroravni.
- Pri izboru metod in meril za ocenjevanje skladnosti in primerljivosti rezultatov je treba v čim večji meri upoštevati značilnosti osnovnih podatkov in končnosti rezultatov (začasni, končni).
- Pri izboru metod in meril za ocenjevanje primerljivosti in skladnosti rezultatov je treba opredeliti, kateri postopki se opravljajo pred vsako objavo rezultatov, katere pa izvajamo le občasno – skupaj s periodiko njihovega izvajanja.
- Postopke preverjanja je treba prilagoditi periodiki raziskovanja, ciljni populaciji, načinu zbiranja podatkov ter vrsti podatkovnega vira.
- Izsledke analize je treba uporabiti za izboljšanje kakovosti pri naslednjih izvedbah raziskovanja.

6.3 INTERPRETACIJA REZULTATOV

V fazi interpretacije podatke pretvorimo v informacije. Statistične pojave uporabniku pojasnimo in razložimo na jasn in razumljiv način.

Podatki morajo biti relevantni in uporabni, vsi zaključki morajo biti podprti s podatki, ki smo jih pridobili v statističnem procesu. Pri interpretaciji podatkov moramo upoštevati način zbiranja podatkov in druge informacije, povezane z zbiranjem podatkov. Vse to je treba uporabnikom posredovati v ustreznih metapodatkih (metodološka pojasnila in metapodatki, npr. opombe, navedba vira, navedba statističnih znamenj itd.), ki se nanašajo na komentar, tabelarično ali grafično predstavitev. Še posebej je treba opisati vse pomanjkljivosti

podatkov, kot so na primer odkloni med ciljno populacijo in populacijo, ki smo jo dejansko opazovali. Pri interpretaciji podatkov je treba upoštevati tudi načelo zaupnosti podatkov.

Pri interpretaciji podatkov moramo presoditi, katere pojave bomo predstavili uporabniku. Pri tem upoštevamo najbolj relevantne pojave, aktualnost tematike in izstopajoče podatke. Pojasnjevalna vrednost predstavljenih podatkov se poveča s primerjavo podatkov v času, prostoru in z uporabo klasifikacij in nomenklatur. Tako lahko uporabnik primerja predstavljene podatke ter jih ovrednoti glede na izbrane primerljive referenčne podatke.

Ker so uporabniške skupine različne (strokovna in splošna javnost), mora biti tudi interpretacija podatkov temu prilagojena. Razmišljati moramo o tem, kaj želi uporabnik vedeti in na kakšen način mu bomo to predstavili, da bo razumljivo, zanimivo in uporabno.

Smernice za zagotavljanje kakovosti

- Interpretacija rezultatov mora biti prilagojena ciljni populaciji in mediju, v katerem bodo podatki pozneje objavljeni.
- Interpretacija rezultatov mora biti nepristranska, objektivna, točna, jasna in razumljiva.
- Interpretacija kratkoročnih statistik je praviloma drugačna kot interpretacija strukturnih statistik (pri prvih se bolj osredotočamo na gibanje pojava, pri drugih pa na strukturo pojava).
- Uporaba primerljivih referenčnih podatkov ima močan vpliv na interpretacijo podatkov. Zaradi tega moramo, zlasti pri časovnih primerjavah, izbrati referenčne točke, ki podatke prikazujejo objektivno.
- Pri interpretaciji rezultatov v obliki indeksov in drugih relativnih števil je treba izbrati smiselno obdobje primerjave, ki bo uporabniku podatkov omogočalo učinkovito interpretacijo gibanja kakega pojava.
- Pri predstavitvi rezultatov v obliki indeksov in drugih relativnih števil je treba biti pozoren na pravilno interpretacijo sprememb v pojavu, kadar so te izražene v odstotkih (%) ali v odstotnih točkah.

7 IZKAZOVANJE PODATKOV

Izkazovanje statističnih podatkov in informacij je sporočanje namensko, sistematično zbranih podatkov javnosti, uporabnikom, in sicer prek različnih sredstev javnega obveščanja (medijev) in z različnimi vrstami objav. Vključuje objave v elektronskih in tiskanih publikacijah, objave na podatkovnem portalu, interaktivne predstavitve podatkov na spletni strani, objave metapodatkov (vprašalniki, metodološka pojasnila, poročila o kakovosti), posredovanje mikropodatkov in podatkov posameznikom, organiziranje novinarskih konferenc, objave člankov, interaktivno izmenjavo informacij preko družbenih medijev, televizijske in radijske intervjuje.

Rezultati in izsledki raziskovanj morajo biti kakovostni, tj. nepristranski in objektivni, ustrezni, natančni in zanesljivi, pravočasni in točni, skladni in primerljivi ter dostopni in jasni.

Z različnimi načini objavljanja podatkov se približamo različnim skupinam uporabnikov. Zato objavljamo podatke na različne načine in prek različnih kanalov. Najpomembnejši kanal (informacijska pot) za objavljanje podatkov uradne statistike je spletna stran. Podatki morajo biti objavljeni tako, da so dostopni vsem uporabnikom hkrati in na enak način. Uporabnikom so na razpolago podatki in metapodatki, vključno z opisi metodologije in informacijami o kakovosti zbranih podatkov. Pomemben vidik izkazovanja podatkov je tudi obveščanje uporabnikov o objavah podatkov ter o revizijah in popravkih napačnih podatkov.

Izkazovanje podatkov vključuje tudi posredovanje podatkov uporabnikom na podlagi posameznih zahtevkov ter redno sporočanje podatkov notranjim in mednarodnim organizacijam.

7.1 POSODABLJANJE IZHODNIH PODATKOV

Objavljanje statističnih podatkov in informacij poteka po standardiziranih postopkih v različnih tehnologijah. Standardizirani postopki temeljijo na vnaprej predpisanih strukturah, formatih in metapodatkih, in ti se upoštevajo že pri pripravi tabel (v fazi statistične obdelave podatkov). Za vse vsebine veljajo enaki, standardizirani postopki dela v skladu z načelom transparentnosti procesov in ob upoštevanju pravočasnosti. Tako pri pripravi novih vsebin kot pri rednem posodabljanju izhodnih podatkov so pomembni urejena dokumentacija, interna izmenjava znanj, standardne komunikacijske poti ter arhiviranje gradiv in postopkov dela.

Za vsako obliko objave statističnih podatkov in informacij je treba najprej posodobiti izhodne podatke; to so lahko podatki v obliki baz podatkov ali pa v obliki tabel s končnimi agregiranimi podatki, pripravljenimi za objavo.

Za raziskovanja, pri katerih se podatki revidirajo za zagotavljanje višje kakovosti, je treba v skladu z navodili o revidiranju statističnih podatkov v fazi izvedbe revizije posodobiti vse izhodne podatke ter revidirane podatke objaviti.

Način posodabljanja izhodnih podatkov se razlikuje glede na vrsto objave:

- dodajanje nove časovne točke datotekam v podatkovni bazi
- manipulativna opravila na podatkovni bazi
- priprava podatkov za posredovanje na mednarodne organizacije
- priprava podatkov za interaktivna orodja
- revidiranje podatkov.

Smernice za zagotavljanje kakovosti

- Pri posodabljanju izhodnih podatkov je treba upoštevati standardizirane postopke in jih izvajati transparentno (pregledno, jasno). Procese, ki se ponavljajo periodično (v enakih časovnih presledkih), je smiselno avtomatizirati.
- Dokumentacija mora biti urejena, objavljena, dostopna in sproti posodobljena.
- Ker je pri tem postopku udeleženih več oseb z različnih področij, mora biti interna koordinacija dokumentirana.
- Vsebine, ki so bile umaknjene zaradi spremembe metodologije, morajo biti ustrezno arhivirane.
- Upoštevati je treba pravila revidiranja podatkov.

7.2 PREDSTAVITEV REZULTATOV

Podatke objavljamo v tiskanih in elektronskih publikacijah, interaktivnih spletnih orodjih ter v podatkovnih bazah. Vsebinske elemente objave je treba pripraviti glede na potrebe in razumevanje ciljne publike in glede na namen objave. Podatke je treba ustrezno predstaviti in objaviti. Predstavitev rezultatov naj bo učinkovita, razumljiva, jasna in zanimiva in naj vsebuje ustrezne primerjave.

Pri predstavitvi rezultatov je treba upoštevati načelo dvojezičnosti, zato praviloma objavimo vse dvojezično (slovenska in angleška različica).

Pri predstavitvi podatkov se je treba zavedati, da so uporabniki različno statistično pismeni (sposobnost razumevanja in razlage statističnih podatkov in zmožnost uporabe) in zaradi tega različno razumejo in uporabljajo statistične podatke in informacije. Strokovna javnost pretežno uporablja podrobneje razčlenjene podatke za potrebe izvajanja nadaljnjih analiz podatkov, zato je tovrstne podatke smiselno objaviti v podatkovnih bazah v formatih, ki omogočajo nadaljnje obdelave (elektronska oblika). Splošno javnost zanimajo predvsem najpomembnejši ali najzanimivejši splošni statistični podatki in informacije, prikazani na jasen in razumljiv način (poljudno). Avtor predstavitve podatkov mora razmišljati o tem, kaj želi uporabnik vedeti in na kakšen način bo to predstavil, da bo to razumljivo, zanimivo in uporabno. Podatki, ki so objavljeni v elektronski prvi objavi in publikacijah, naj bodo predstavljeni v statistični zgodbi, kar pomeni, da podatke umestimo v širši kontekst, jih povežemo tudi z drugimi sorodnimi podatki, če je to možno, in naredimo celoto iz komentarja in vizualizacije. Pri predstavitvi podatkov je treba upoštevati standarde za objavo.

Načela predstavitve rezultatov se razlikujejo glede na vrsto in medij objave, a kljub temu veljajo neke splošne zakonitosti.

Komentar mora biti učinkovit, tj. razumljiv, jasen, zanimiv in preprost ter obenem nevtralen oz. nepristranski. Vsebuje naj primerjave (npr. časovne, krajevne, med kategorijami, dejavnostmi, področji); če so vzroki znani, tudi navajanje teh.

Tabele prikazujejo pregledno urejene podatke in predstavljajo najpogostejši element statističnih objav, saj so primerne za prikaz večjega števila podatkov. Z njimi lahko predstavimo kateri koli izbrani pojav, ki ga s podrobnim prikazom v tabeli lahko razčlenimo in natančno analiziramo, s strnjenim pa posplošimo in poenostavimo. Primerne so za prikaz absolutnih podatkov in relativnih števil (koeficienti, gostote, indeksi, povprečja, strukture). Za objavo v tiskanih publikacijah in elektronskih publikacijah na spletni strani so primerne tabele s strnjenim prikazom, podrobne podatke pa objavljamo v tabelah v podatkovni bazi. V tabelah v podatkovni bazi morajo biti objavljeni vsi osnovni rezultati (absolutni in relativni na višjih ravneh agregacije) in tudi podrobni podatki.

Grafikoni so grafične predstavitve statističnih podatkov (pojavov) v obliki stolpcev, linij, krogov ipd. Glede na njihovo obliko ločimo različne vrste grafikonov (stolpčni, linijski, tortni itd.) Značilnosti pojava in podatkov narekujejo, katera vrsta grafikona je za predstavitev podatkov najustreznejša. Grafikoni lahko v primerjavi s komentarjem ali tabelami informacije predstavijo učinkoviteje in popolneje ter lahko nazorneje prikažejo večje število podatkov. Primerni so za prikaz absolutnih podatkov in relativnih števil, in sicer takrat, kadar podatkov ne nameravamo predstaviti zelo natančno.

Karte so najustreznejše orodje za vizualizacijo vzorcev razporeditve pojavov v prostoru. Omogočajo primerjavo različno velikih prostorskih enot (prikaz kazalnikov) in na nazoren način predstavijo veliko količino podatkov. Na statističnih kartah so običajno prikazana relativna števila, zlasti kazalniki na 1.000 prebivalcev in preračuni na površino prostorske enote.

Infografike so predstavitev podatkov in informacij z različnimi grafičnimi elementi. To je izraz za smiselno urejen skupek ilustracij in grafikonov, ki ponazarjajo uporabnikom, kar bi bilo drugače težko ali nemogoče prikazati samo z besedilom. Infografike omogočajo, da z grafičnimi elementi naredimo razumljive in zanimive za širši krog uporabnikov tudi težje razumljive, lahko tudi suhoparne podatke.

Smernice za zagotavljanje kakovosti

- Poskrbeti je treba, da je dokumentacija o pravilih, navodilih in priporočilih za predstavitev podatkov urejena, objavljena in dostopna.
- Pri načinu predstavitve statističnih podatkov (pisanju učinkovitih komentarjev ter izdelavi tabel, grafikonov, kart in infografik) v objavah za javnost je treba upoštevati standarde za objavo.
- Pri pripravi tabel za objavo v podatkovni bazi je treba matrike oblikovati čim bolj optimalno glede na vsebino, tehnologijo priprave podatkov, tehnične možnosti orodij in sprejete standarde objavljanja.
- Pri predstavitvi podatkov je treba upoštevati načela nepristranskosti in objektivnosti, ustreznosti, jasnosti in razumljivosti ter tudi statistične zaupnosti.
- Podatki morajo biti predstavljeni skupaj z metapodatki, kar uporabniku zagotavlja pravilno razlago, uporabo in smiselno primerjavo.
- Pri učinkovitem komentiranju sprememb pojava je treba obvezno zaradi ustreznega in smiselnega podajanja primerjav navesti obdobje opazovanja in obdobje primerjave.
- Priporočeno in zaželeno je sodelovanje z lektorjem in uporaba lektorskih storitev.

7.3 OBJAVLJANJE

Vsi podatki in informacije ter metapodatki se objavljajo in so uporabnikom dostopni prek različnih kanalov, med katerimi je najpomembnejša spletna stran. Podatki na spletni strani so na voljo brezplačno, dostopni so tudi v angleščini, za čim večji obseg uporabe pa se jih aktivno promovira.

Vsaka objava statističnih podatkov mora biti napovedana in potrjena v skladu z načrtom publiciranja v Koledarju objav. Koledar objav se pripravi konec leta za eno leto vnaprej. Pomembno je, da so ob načrtovanju določeni ustrezni datumi objav, ustrezni naslovi za vse objave podatkov ter predvidena načrtovana, redna revidiranja podatkov, saj je od tega odvisna končnost rezultatov (začasni ali končni); določeno mora biti tudi, ali bodo podatki hkrati objavljeni v podatkovni bazi. Med aktivnosti objavljanja poleg načrtovanja, napovedovanja in potrjevanja objav sodi tudi objavljanje na napovedani dan ob napovedani uri, vodenje evidence o naročnikih na statistične objave, obveščanje uporabnikov o objavah, zagotavljanje enakega dostopa za vse uporabnike in odpošiljanje objav naročnikom. Postopki objavljanja podatkov so za različne vrste objav različni (objava novic, podatkovni portal, posredovanje podatkov drugim organizacijam idr.).

Aktivnost, povezana z objavljanjem statističnih podatkov, je tudi odpravljanje napak v objavljenih statističnih podatkih in informacijah, katere namen je oskrbovanje uporabnikov s točnimi in kakovostnimi statističnimi podatki in informacijami. Sistem odpravljanja napak mora biti urejen tako, da je pregleden in jasen za uporabnike in za avtorje. Natančni postopki popravljanje napak v objavljenih statističnih podatkih so predpisani z navodili in dostopni uporabnikom na spletni strani.

Smernice za zagotavljanje kakovosti

- Podatki morajo biti objavljeni točno v skladu z napovedmi v Koledarju objav; ta mora biti dostopen vsem uporabnikom.
- Celoten postopek raziskovanja mora biti načrtovan tako, da so podatki objavljeni pravočasno.
- Podatki morajo biti dostopni vsem uporabnikom na enak način in čim lažje. Prav tako mora biti uporabnikom na voljo čim več enostavnih možnosti dodatnega povpraševanja ter naročanja podatkov in storitev. Pri tem imata pomembno vlogo informacijsko bogata, navigacijsko enostavna in vizualno privlačna spletna stran in informacijsko središče.
- Objavljanje podatkov mora biti čim bolj pregledno. Skupaj s podatki morajo biti tako na voljo tudi vsi ustrezni metapodatki, kot so metodološka pojasnila, osnovni kazalniki kakovosti, vprašalniki itd.
- Omogočena mora biti tudi preglednost in jasnost postopkov popravljanja napak in postopkov revidiranja podatkov.

7.4 PODPORA UPORABNIKOM

Uporabnikom je treba omogočiti dostop do podatkov, ki so rezultat raziskovanj, pri čemer mora biti zagotovljena zaščita statistične zaupnosti. Podpora uporabnikom se zagotavlja predvsem po elektronski pošti in po telefonu, lahko tudi z osebnim stikom. Načini podpore uporabnikom so lahko različni; odvisni so od tega, za katero vrsto/obliko povpraševanja gre.

Podpora uporabnikom obsega posredovanje statističnih podatkov in informacij, svetovanje o dostopu do podatkov in svetovanje o uporabi orodij za pripravo podatkov ter pomoč pri iskanju in pripravi statističnih podatkov in informacij.

Statistični podatki in informacije, ki so objavljeni, in podatki, katerih priprava ne zahteva dodatne obdelave, so brezplačni. Priprava podatkov po želji uporabnika pa se obračuna po veljavnem ceniku. Posredovanje podatkov, ki sodi v okvir poročanja mednarodnim organizacijam, je brezplačno.

Evidenci pisnih in telefonskih povpraševanj po podatkih omogočata analize uporabnikov, na podlagi katere ugotavljamo sestavo uporabnikov, načine komuniciranja z njimi in katera so najbolj iskana statistična področja.

Uporabnikom je omogočen tudi dostop do podatkov za znanstvenoraziskovalni in analitični namen, vendar pod posebnimi pogoji in s posebno pogodbo. Te podatke imajo pravico pridobiti registrirane raziskovalne institucije, registrirani raziskovalci in raziskovalci organov državne uprave.

V sodelovanju z mednarodnimi organizacijami posebej obravnavamo povpraševanja po mednarodno primerljivih podatkih. Posredovanje podatkov mednarodnim organizacijam se zagotavlja z izpolnjevanjem vprašalnikov. V okvir mednarodnega poročanja sodi tudi posredovanje podatkov drugim ustanovam v Sloveniji, ki poročajo mednarodnim organizacijam, in reševanje zahtevkov mednarodnih organizacij, ki nimajo oblike vprašalnikov.

Smernice za zagotavljanje kakovosti

- Podatki in informacije, posredovani uporabnikom, morajo biti točni, ustrezni in primerno predstavljeni.
- Uporabniki morajo podatke prejeti pravočasno. O morebitnih zamudah je treba uporabnika nemudoma obvestiti.
- Uporabnikom posredujemo celovite odgovore (če povpraševanje vključuje več podvprašanj, uporabniku odgovorimo na vsa hkrati).
- Uporabnikom zagotavljamo brezplačne podatke.
- Informacije o tem, kako do statističnih podatkov, so objavljene na več mestih (spletna stran, publikacije).
- Podpora uporabnikom se nudi v času poslovnega časa organa.
- Osnovna pojasnila in navodila v zvezi z dostopom do podatkov za znanstvenoraziskovalni in analitični namen morajo biti objavljena na spletni strani.

7.5 HRAMBA STATISTIČNIH MIKROPODATKOV

Statistični mikropodatki za nadaljnjo uporabo in za uporabo za raziskovalno-analitičen namen raziskovalnih organizacij, vladnih služb in samostojnih raziskovalcev se hranijo v elektronski obliki. Hramba statističnih mikropodatkov v elektronski obliki poteka v sklopu izvajanja politike varnostnega kopiranja, ki se izvaja v skladu z navodilom za izdelavo in hrambo varnostnih kopij.

Vsi podatkovni in datotečni strežniki, ki vsebujejo podatke iz različnih raziskovanj in metapodatke, ki so trajne vrednosti, so podvrženi enotnemu sistemu varnostnega kopiranja in se izvaja na podlagi navodil za izdelavo in hrambo varnostnih kopij. Varnostno kopiranje se izvaja z namenom, da se preprečijo izgube podatkov, do česar lahko pride, če se zgodi incident, kot je na primer tehnična okvara dela informacijskega sistema ali izbris podatkov. Poleg tega se kopije uporabljajo tudi za reševanje logičnih napak izvajalcev/uporabnikov podatkovnih baz z obnovitvijo stanja podatkovne zbirke na določen dan.

Za hrambo končnih mikropodatkov so vzpostavljeni ustrezni postopki in aplikacija. To so podatki in metapodatki, iz katerih so izračunani agregirani objavljeni podatki. Varnostno kopiranje teh podatkov poteka po standardnih postopkih hrambe statističnih mikropodatkov.

Smernice za zagotavljanje kakovosti

- Varnostno kopiranje in hrambo je treba izvajati v skladu z navodili za izdelavo in hrambo varnostnih kopij.
- Hrambo končnih mikropodatkov je treba izvajati v skladu s standardnim postopkom.

7.6 HRAMBA AGREGIRANIH PODATKOV

Statistični agregirani podatki za nadaljnjo uporabo se hranijo v elektronski in tiskani obliki. Hramba agregiranih podatkov v elektronski obliki poteka v sklopu izvajanja politike varnostnega kopiranja in hrambe elektronskih podatkov in v obliki hrambe spletnega portala.

Agregirani statistični podatki v tiskani obliki so lahko objavljeni tako v izdanih kot v prejetih publikacijah z drugih institucij. V knjižničnem arhivu se hranijo publikacije obeh vrst. V obeh primerih gre tako za serijske kot za monografske publikacije. Tiskani izvodi lastnih publikacij se pošiljajo v arhiv Narodne in univerzitetne knjižnice (NUK). Gradiva pregleda komisija in določi seznam gradiv za odpis.

Smernice za zagotavljanje kakovosti

- Redno je treba spremljati strokovne zahteve na področju varnostnega kopiranja in hrambe.
- Redno je treba pošiljati izvode v arhiv NUK-a.

8 DOKUMENTIRANJE IN EVALVACIJA RAZISKOVANJA

Raziskovanja se praviloma izvajajo periodično in tako se celoten statistični proces ponavlja. Pomembno je, da statistični proces vključuje tudi povratno zanko, ki omogoča uvajanje sprememb in izboljšav. Ob zaključku posamezne izvedbe raziskovanja je treba urediti informacije, ki so nastale ob izvajanju statističnega procesa; pripraviti celovito oceno uspešnosti poteka raziskovanja in predvideti morebitne izboljšave za naslednjo izvedbo.

Izdelava dokumentacije raziskovanj je pomemben del statističnega procesa, ki izboljša kakovost podatkov, statistik in metapodatkov, hkrati pa je nujno potrebna tudi za ponovljivost izvedbe raziskovanja.

Zbiranje informacij o kakovosti statističnih podatkov poteka med celotnim statističnim procesom, pri čemer te informacije običajno zbiramo v obliki kazalnikov kakovosti. Kazalniki kakovosti so številske vrednosti, s katerimi izkazujemo doseženo kakovost v okviru posameznih komponent. Sistematično dokumentirane vrednosti kazalnikov kakovosti so pomemben del informacij o poteku raziskovanja in pomagajo pri odkrivanju morebitnih sistematičnih napak v procesu. S pomočjo kazalnikov lahko ocenimo kakovost statističnih podatkov in kritično ovrednotimo pridobljene rezultate, pomembni pa so tudi za uporabnike, saj tako ti pridobijo dodaten vpogled v potek statističnega procesa.

Objavljanje informacij o kakovosti podatkov je transparenten način obveščanja uporabnikov o različnih vidikih statističnih podatkov, ki jih uporabljajo. Na podlagi zbranih informacij se, kadar je to potrebno, pripravi načrt za izboljšave in v skladu s tem načrtom se v proces uvajajo spremembe.

8.1 IZDELAVA DOKUMENTACIJE O RAZISKOVANJU

Postopek priprave dokumentacije o raziskovanju obsega podroben opis statistične aktivnosti (npr. opis pojmov, definicij, uporabljenih metod, procesa obdelave, uporabljenega informacijskega sistema, delovnih navodil). Kakovost dokumentacije o raziskovanju je pomemben kazalnik kakovosti raziskovanja. Je tudi pomembno orodje za komunikacijo med različnimi izvajalci raziskovanj ter med proizvajalci in uporabniki statističnih podatkov. Dokumentacija o raziskovanju je del metapodatkov in zagotavlja ponovljivost izvedbe raziskovanja.

Dokumentacijo raziskovanj delimo na dokumentacijo za uporabnike podatkov raziskovanj (metodološka dokumentacija) in na dokumentacijo za izvajalce raziskovanj (izvedbena in tehnična dokumentacija); druga dokumentacija je interna.

Metodološka dokumentacija opisuje in dokumentira statistične rezultate in vsebuje: opredelitev ciljne populacije, opredelitev vsebine vhodnih podatkov, ključne statistike, raven objave, metodologijo izvedbe raziskovanja ter opis revizij raziskovanja. Namenjena je uporabnikom rezultatov raziskovanja in je javno objavljena.

Izvedbena dokumentacija opisuje statistične postopke in procese, ki smo jih uporabili v celotnem statističnem procesu. Zajema navodila za izvajanje posameznih faz izvedbe raziskovanja (npr. navodila za anketarje, navodila za urejanje podatkov, navodila za tisk, odpremo in sprejem gradiv). Tehnična dokumentacija nastaja pri izdelavi programske opreme in drugih tehničnih pripomočkov (npr. opis izdelave programske opreme za vnos in urejanje podatkov, opis izvajanja letnega pregleda časovnih vrst). Izvedbena in tehnična dokumentacija sta namenjeni izvajalcem raziskovanj za interno uporabo.

Namen dokumentacije za uporabnike je, da pomaga uporabnikom bolje razumeti, kaj podatki in uporabljene statistične metode merijo, da podatke razumejo, da jih lažje poiščejo in da jih po potrebi lahko tudi sami še naprej obdelajo. Primeri vsebin dokumentacije za uporabnike so: vprašalniki za raziskovanja, predstavitevne zgibanke, metodološka pojasnila, uporabniško usmerjena poročila o kakovosti raziskovanj.

Namen dokumentacije za izvajalce raziskovanj je zagotavljanje učinkovite in zanesljive produkcije podatkov; zagotavljanje jasnosti, transparentnosti, sledljivosti in ponovljivosti uporabljenih postopkov; izboljšanje kakovosti statistik; vzdrževanje in razvoj delovnih metod, produkcijskih procesov in informacijskega sistema; predstavitev korakov raziskovanja novim zaposlenim. V dokumentaciji za izvajalce raziskovanj morajo biti predvsem podrobno opisani posamezni koraki, ki smo jih v izvedbi raziskovanja uporabili (npr. določitev ciljne populacije, izbira enot v vzorec (sestava vzorca), oblikovanje vprašalnika, urejanje podatkov, objava podatkov itd.). Opisi naj vsebujejo tudi poročilo o tem, kaj je bilo v posameznem koraku opravljeno, in pojasnilo o tem, zakaj je bil izbran ravno določen način izvedbe. Vse to so namreč koristne informacije za morebiten razvoj procesov raziskovanja v prihodnosti ali za načrtovanje novih raziskovanj. Včasih so nekatere vsebine iz take dokumentacije zanimive tudi za zunanje uporabnike podatkov, in jih je zato koristno javno objaviti.

Smernice za zagotavljanje kakovosti

- Dokumentacija o raziskovanju mora biti natančna, izčrpna in razumljiva uporabniški skupini, ki ji je namenjena.
- Vsebina in stopnja podrobnosti dokumentacije naj bosta prilagojeni uporabniški skupini.
- Pomembno je dokumentirati časovno zgodovino sprememb v procesu.
- Pri pripravi dokumentacije naj se uporabljajo standardne oblike in predloge, kjer obstajajo.
- Dokumentacija za uporabnike in za izvajalce naj bo enostavno dostopna.
- Dokumentacija naj bo zapisana tako, da bo v čim večji meri zagotavljala transparentnost statističnih postopkov in produktov.
- V dokumentaciji za uporabnike naj bo pojasnjeno vse, kar bi morda pri uporabnikih lahko povzročilo nejasnosti ali nerazumevanje.
- V dokumentaciji za izvajalce naj bo dokumentirano vse, kar bo zagotavljalo jasnost in ponovljivost uporabljenih postopkov in procesov, še posebej morebitni odmiki od standardno uveljavljenih postopkov.
- Dokumentacija za izvajalce naj vsebuje tudi navedbo/opis predvidenih postopkov, po katerih se je treba ravnati, če pride do napak, in imena oseb ali oddelkov, s katerimi je treba vzpostaviti stik, če pride do težav ali nejasnosti.

8.2 ZBIRANJE INFORMACIJ ZA OCENO KAKOVOSTI

Proces izvedbe raziskovanja obsega več različnih korakov, ti pa vsebujejo mnogo dejavnikov, ki lahko vplivajo na kakovost rezultatov. Da bi bili na koncu tega procesa v čim večji mogoči meri sposobni oceniti kakovost izkazanih rezultatov, je treba v vsakem koraku statističnega procesa zbrati čim več informacij, ki bi lahko služile za oceno kakovosti predvsem izkazanih rezultatov, posledično pa tudi za oceno statističnega procesa ter institucionalnega okolja, v katerem proces poteka.

Informacije, ki jih zbiramo zaradi ocenjevanja kakovosti, lahko razdelimo v dve skupini. V prvo spadajo informacije, ki izhajajo neposredno iz procesa, v drugo pa tiste, ki jih pridobimo od uporabnikov rezultatov raziskovanja (predvsem zunanjih). Sem spadajo predvsem informacije o zadovoljstvu uporabnikov, informacije o novih ali spremenjenih potrebah po informacijah ali o neustreznosti izkazanih rezultatov za potrebe uporabnikov.

Posebej pri informacijah, ki izhajajo neposredno iz procesa, je treba posebno pozornost nameniti kvantitativnim informacijam, tj. tistim, ki jih običajno označujemo z nazivom kazalniki kakovosti. Dva tipična primera takih kazalnikov sta vzorčna napaka in stopnja neodgovora. Kazalniki so pomembni predvsem zato, ker omogočajo empirično-analitski pristop k spremljanju in zagotavljanju kakovosti produktov in procesov. Izračun teh kazalnikov je treba v čim večji meri vgraditi v sam proces. Pomembno je, da se vrednosti kazalnikov redno analizirajo in uporabijo za pripravo izboljšav, prav tako pa se vključujejo v metodološka gradiva za uporabnike (npr. poročila o kakovosti).

Smernice za zagotavljanje kakovosti

- Skrbno je treba analizirati statistični proces in ugotoviti vse mogoče vire informacij, ki bi bili lahko koristni za oceno kakovosti.
- Pri določitvi kazalnikov kakovosti, ki se bodo izračunavali in pozneje vključevali v poročila o kakovosti, naj se upošteva seznam standardnih kazalnikov kakovosti. Če gre za res pomembne kazalnike, specifične za neko področje, se lahko po potrebi v seznam dodajajo tudi dodatni kazalniki.
- Ob pripravi navodil za izračunavanje standardnih kazalnikov kakovosti naj se dosledno upošteva ustrezna metodologija izračuna.
- Za pridobitev informacij od uporabnikov je treba uporabiti vse razpoložljive kanale.
- Zbrane ali v različnih oblikah sporočene informacije je treba urediti in preoblikovati tako, da bodo uporabne za nadaljnje analize in da jih bo mogoče vključiti tudi v poročilo o kakovosti.
- Poskrbeti je treba, da so vse informacije v zvezi s kakovostjo na voljo v čim krajšem času, saj jih bo le tako mogoče koristno in učinkovito uporabiti.

8.3 IZVEDBA OCENE POSTOPKOV IN PROCESOV

Informacije, zbrane za oceno kakovosti, je treba smiselno urediti in analizirati. Ob koncu vsake izvedbe raziskovanja (tako redne kot pri reviziji) se na podlagi zbranih in pripravljenih informacij pripravi in javno objavi poročilo o kakovosti; to obsega opise vseh dimenzij kakovosti, skupaj z vrednostmi kazalnikov kakovosti.

Če je bila ob zadnji izvedbi raziskovanja opravljena večja revizija (vpeljane pomembne metodološke spremembe, spremembe vprašalnika, prehod na nov način zbiranja podatkov), je treba napisati podrobno poročilo, ki opisuje potek in uspešnost izvedene revizije. Osredotočiti se je treba predvsem na zaznane najbolj kritične točke v procesu oziroma v širšem kontekstu izvajanja raziskovanja.

Pri periodičnih raziskovanjih je treba posebej pozorno analizirati vrednosti kazalnikov kakovosti v primerjavi s prejšnjimi obdobji. Če se zaznajo večje razlike glede na vrednosti iz prejšnjih obdobji, je treba analizirati vzroke za take razlike in pripraviti načrt sprememb, ki bodo v prihodnosti izboljšale področje, katerega kakovost meri »problematični« kazalnik. Oцени se potek statističnega procesa: opiše se, kakšni so bili problemi in zakaj je do njih prišlo; kje so nastajale zamude v procesu in razlogi zanje; identificirajo se kritične točke procesa in proučijo se možnosti za boljši pretok procesa; preučijo se povratne informacije, zbrane med procesom, in možnosti za izboljšave. Na podlagi analize procesa je treba podati predloge za vpeljavo izboljšav za prihodnje izvedbe raziskovanja. Zbrane informacije o kakovosti ter o poteku procesa raziskovanja so koristne za pripravo načrta za izboljšave.

Smernice za zagotavljanje kakovosti

- Rezultati analiz kakovosti naj bodo predstavljeni jasno in enostavno, v obliki, ki bo razumljiva tudi nekomu, ki se neposredno ne ukvarja z obravnavanim raziskovanjem.
- Zbrane informacije je treba analizirati natančno, strokovno in nepristransko. V analizi je treba posebej opozoriti na zaznane šibke točke.
- V analizi kazalnikov kakovosti naj se posebna pozornost nameni primerjavi s kazalniki kakovosti iz preteklih obdobji oziroma iz drugih (sorodnih) raziskovanj.
- Informacije o kakovosti morajo biti javno objavljene in podane na uporabniku razumljiv način. Poročilo o kakovosti se pripravi in objavi čim prej po objavi statističnih podatkov.
- Vsi odzivi uporabnikov naj se kritično pregledajo in analizirajo. Pomembno je, da se v fazi analize poudarijo predvsem tiste zahteve, ki jih lahko vključimo v načrte za prihodnje izboljšave.
- Načrt za izboljšave naj obsega celovit pregled izboljšav, ki jih je mogoče vpeljati v proces raziskovanja, z analizo pa je treba načrt čim bolj približati realnim možnostim.
- Na podlagi opravljenih analiz je treba izboljšave vpeljati v proces, potem pa jih redno spremljati s kazalniki kakovosti in tako ugotavljati učinek izboljšav.

KRATICE

CRP	Centralni register prebivalstva
ES	Evropska skupnost
EU	Evropska unija
IT	informacijska tehnologija
LPSR	Letni program statističnih raziskovanj
NUK	Narodna in univerzitetna knjižnica
PRS	Poslovni register Slovenije
REN	Register nepremičnin
RPE	Register prostorskih enot
SKD	Standardna klasifikacija dejavnosti
SPRS	Statistični poslovni register Slovenije
SRDAP	Statistični register delovno aktivnega prebivalstva
SRKG	Statistični register kmetijskih gospodarstev
SURS	Statistični urad Republike Slovenije

VIRI IN LITERATURA

- Cochran, W.G. (1977): *Sampling Techniques*. London: John Wiley & Sons, Inc.
- Cox, B. G. et al. (1995). *Business Survey Methods*. New York: Wiley.
- Design your questions right. How to develop, test, evaluate and improve questionnaires*. (2004). Stockholm: Statistics Sweden. Pridobljeno 19.6.2012 s spletne strani:
http://www.scb.se/statistik/_publikationer/OV9999_2004A01_BR_X97OP0402.pdf
- Dillman, D. A. (2003). *Mail and Internet Surveys: The Tailored Design Method*. London: John Wiley and Sons, Inc.
- Fellegi, I.P., Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*. 71(353), 17-35.
- Generic Statistical Business Process Model (2016)*. UNECE. Pridobljeno 19.10.2016 s spletne strani:
<http://www1.unece.org/stat/platform/display/GSBPM/Generic+Statistical+Business+Process+Model>
- Granquist, L. (1991). Macro Editing – A Review of Some Methods for Rationalizing the Editing of Survey Data. *Statistical Journal*, 8, 137-145.
- Hidiroglou, M.A., J.M. Berthelot, (1986). Statistical Editing and Imputation for Periodic Business Surveys. *Survey Methodology*, 12, 73-83.
- Levy, P. S. Lemeshow, S. (2004). *Sampling of Populations, Methods and Applications*. New York: Wiley.
- Presser, S., Rothgeb, J. M., Couper, M., Lesser, J., Martin, E., Martin, J., Singer, E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
- Quality Guidelines for Official Statistics*. (2007). Helsinki: Statistics Finland. Pridobljeno 19.6.2012 s spletne strani: http://www.stat.fi/org/periaatteet/qg_2ed_en.pdf
- Quality Guidelines*. (2003). Ottawa: Statistics Canada.
- Sarndal, C.E., Svensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M., Tourangeau, R. (1999). *Cognition and Survey Research*. London: John Wiley & Sons, Inc.
- Snijders, G. (2002). *Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality*. Deen Haag: Statistics Netherlands.
- Survey Methods and Practices*. (2003). Ottawa: Statistics Canada.
- Tehnično priporočilo za varno elektronsko arhiviranje*. (2003). GZS. Pridobljeno 19.6.2012 s spletne strani:
[http://www.gzs.si/e-poslovanje/dokumentacija/eSLOG-Elektronski_arhiv_0.99\(v_pripravi\).pdf](http://www.gzs.si/e-poslovanje/dokumentacija/eSLOG-Elektronski_arhiv_0.99(v_pripravi).pdf)
- Thompson, S. K. (1992). *Sampling*. London: John Wiley & Sons, Inc.
- Tourangeau, R., Rips, L. J., Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Wallgren A., Wallgren B. (2007). *Register-based Statistics; Administrative Data for Statistical Purposes*. London: John Wiley & sons.

KAKO DO STATISTIČNIH PODATKOV IN INFORMACIJ?

◆ **na spletnih straneh Statističnega urada**

www.stat.si

◆ **po pošti, telefonu ali elektronsko**

naslov: Statistični urad Republike Slovenije,
Litostrojska cesta 54, 1000 Ljubljana, Slovenija

telefon: (01) 241 64 04

e-naslov: info.stat@gov.si

 [@StatSlovenija](https://twitter.com/StatSlovenija)

◆ **z naročilom statističnih publikacij**

naslov: Statistični urad Republike Slovenije,
Litostrojska cesta 54, 1000 Ljubljana, Slovenija

telefon: (01) 241 52 85

e-naslov: gp.surs@gov.si

◆ **z obiskom v Informacijskem središču**

poslovni čas: od ponedeljka do četrтка od 9.00 do 15.30
petek od 9.00 do 14.30

