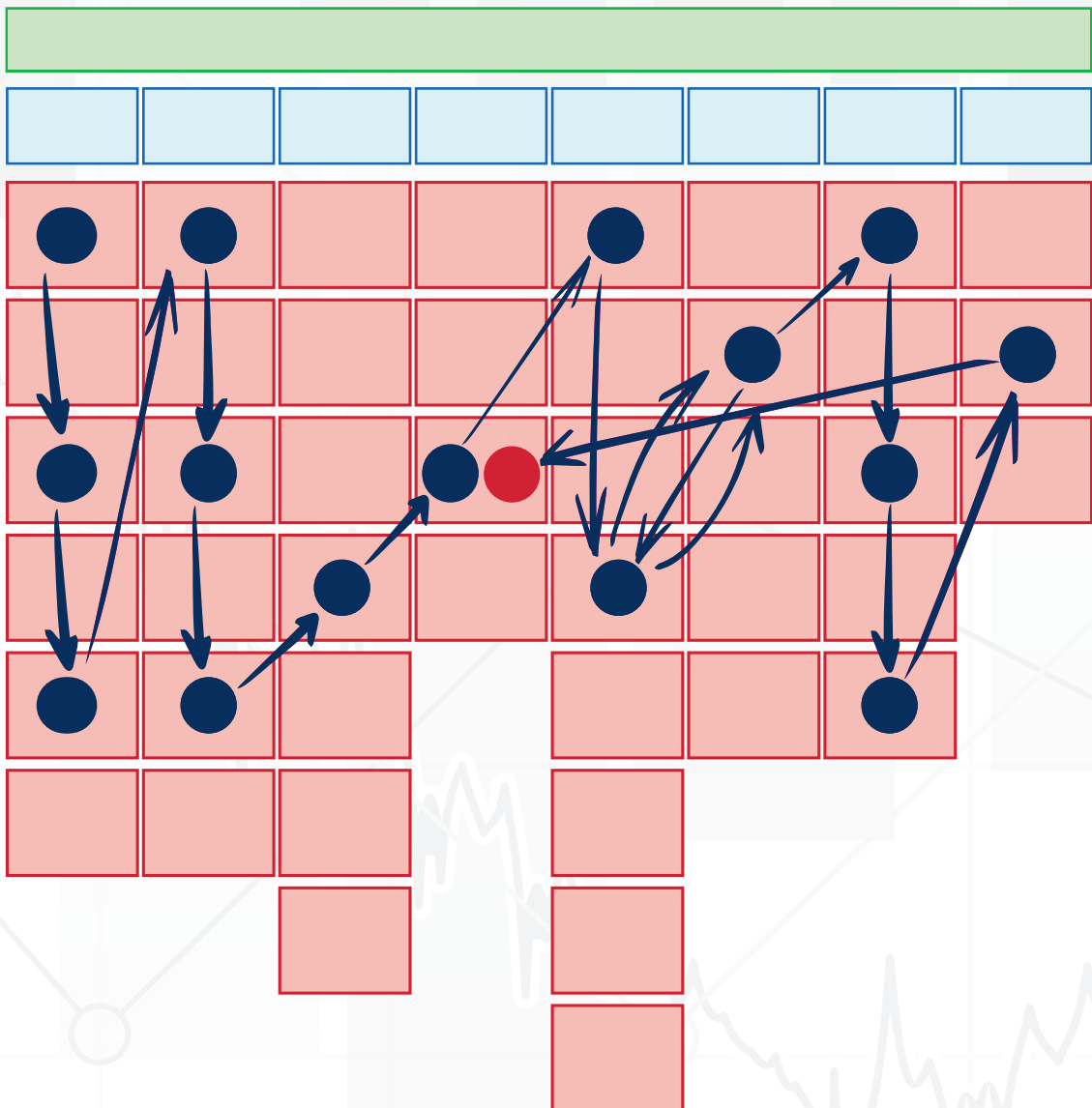


UPUTE ZA KVALITETU PREMA GENERIČKOME MODELU STATISTIČKOGA POSLOVNOG PROCESA (GSBPM-u)

lipanj 2022.



Tiskano u elektroničkom obliku

MOLIMO KORISNIKE DA PRI KORIŠTENJU PODATAKA NAVEDU IZVOR.

Sadržaj

PREGOVOR	5
KRATICE	6
OPĆA NAČELA	8
1. Definiranje potreba za podacima i potrebnih rezultata	13
1.1. Definiranje potreba za informacijama i potrebnih rezultata	13
1.2. Provjera raspoloživih izvora podataka	14
1.3. Utvrđivanje koncepta za proizvodnju rezultata te analiza i testiranje mogućnosti	14
2. Priprema i izrada statističkih metodologija	16
2.1. Definiranje i izrada metodologije za prikupljanje podataka i provedbu istraživanja	16
2.2. Definiranje okvira i metodologije za izbor uzorka	18
2.3. Izrada metodologije za obradu podataka.....	20
3. Izrada potrebnih instrumenata za provedbu	23
3.1. Izrada projektnih zahtjeva	23
3.2. Izrada instrumenata za prikupljanje podataka	25
3.3. Izrada programske podrške	28
3.4. Testiranje alata za prikupljanje i obradu podataka	29
3.5. Konfiguriranje tijeka proizvodnih procesa	31
4. Prikupljanje podataka	32
4.1. Izbor ciljne populacije / uzorka.....	34
4.2. Priprema za prikupljanje podataka	36
4.3. Prikupljanje primarnih podataka	38
4.4. Preuzimanje podataka iz administrativnih i drugih sekundarnih izvora	43
4.5. Unos prikupljenih podataka	45
5. Obrada podataka	50
5.1. Integriranje prikupljenih podataka.....	50
5.2. Kontrola, uređivanje i ispravljanje podataka	52
5.3. Imputiranje i utežavanje	53
5.4. Proizvodnja izvedenih varijabli.....	57
5.5. Izračunavanje agregata	58
5.6. Izrada konačnih datoteka podataka.....	60
5.7. Proizvodnja i ažuriranje statističkih registara i baza podataka	62

6.	Analiza	65
6.1.	Statistička analiza rezultata	65
6.2.	Kontrola kvalitete rezultata	67
6.3.	Detaljna analiza i tumačenja podataka za objavljivanje	68
6.4.	Zaštita povjerljivih podataka	68
7.	Diseminacija.....	71
7.1.	Dizajn i proizvodnja diseminacijskih proizvoda.....	74
7.2.	Upravljanje objavljivanjem diseminacijskih proizvoda	74
7.3.	Promidžba diseminacijskih proizvoda.....	77
7.4.	Upravljanje odnosima s korisnicima	78
8.	Vrednovanje.....	80
8.1.	Prikupljanje informacija za vrednovanje	80
8.2.	Vrednovanje rezultata	81
8.3.	Akcijski plan za poboljšanje	82
	LITERATURA	84

PREDGOVOR

Državni zavod za statistiku glavni je nositelj, diseminator i koordinator sustava službene statistike Republike Hrvatske te glavni predstavnik nacionalnoga statističkog sustava pred europskim i međunarodnim tijelima nadležnima za statistiku.

Statistički podaci koje Državni zavod za statistiku prikupi, obradi, objavi i diseminira rezultat su složenih postupaka i procesa provedbe statističkih istraživanja koji su propisani godišnjim provedbenim planovima statističkih aktivnosti Republike Hrvatske.

Učinkovita i pravodobna priprema i diseminacija kvalitetnih statističkih pokazatelja koji odražavaju ekonomske i društvene pojave i procese te osiguravaju korisnicima pouzdan izvor za analizu postojećeg stanja i donošenje odluka među glavnim su zadacima hrvatskoga statističkog sustava.

Uzimajući u obzir sve odrednice kvalitete u statistici, Državni zavod za statistiku odlučio je primijeniti prilagođeni model GSBPM jer iscrpno opisuje i definira skup poslovnih procesa potrebnih za izradu službene statistike. Navedeni model pruža standardni okvir i jedinstvenu terminologiju kojima se omogućuje modernizacija statističkih proizvodnih procesa te razmjena metoda i komponenata. Također se primjenjuje za integraciju podataka i standardizaciju metapodataka, kao predložak za procesne dokumentacije, usklađivanje statističkih računalnih infrastruktura, osiguranje okvira za procjenu procesa kvalitete te za daljnja poboljšanja.

Na temelju strukture modela GSBPM Državni zavod za statistiku proveo je analizu mogućnosti njegove primjene u hrvatskome statističkom sustavu te se na temelju analiza pripremio prilagođeni model i ovu publikaciju koja na jednome mjestu obuhvaća sve poslovne procese i potprocesne, uključujući primjere iz prakse i stavljajući naglasak na ostvarenje veće kvalitete statističkih proizvoda i usluga.

Ovom je publikacijom u osnovnim crtama dan prikaz statističkih aktivnosti po svim procesima i potprocesima od definiranja potreba za podacima pa do njihova vrednovanja. Čitatelji će moći na jednome mjestu dobiti uvid u sveobuhvatno statističko poslovanje i planove za daljnja unapređenja koristeći se suvremenim tehnologijama.

KRATICE

APIS-IT	Agencija za podršku informacijskim sustavima i informacijskim tehnologijama d. o. o.
BDP	bruto domaći proizvod
CAPI	računalno potpomognuto anketiranje
CATI	računalno potpomognuto telefonsko anketiranje
CAWI	računalno potpomognuto internetsko anketiranje
CD	kompaktni disk
Eurostat	Statistički ured Europske unije
DVD	digitalni višenamjenski disk
DV-PO	Dječji vrtići i druge pravne osobe
DZS	Državni zavod za statistiku
ESA	Europski sustav nacionalnih računa
eSPRI	Statistički poslovni registar
ESS	Europski statistički sustav
ESSC	Odbor Europskoga statističkog sustava
EU	Europska unija
Fina	Financijska agencija
GeoSTAT	portal Državnog zavoda za statistiku za kartografske prikaze
GSBPM	generički model statističkoga poslovnog procesa
PO-31a/Q	Prodaja i otkup poljoprivrednih proizvoda
PO-31b/Q	Prodaja i otkup poljoprivrednih proizvoda
PO-32	Izveštaj o ostvarenim prirodnim ranih usjeva i voća
PP/T-11	Tromjesečni izvještaj o pomorskom i obalnom prijevozu
PR/T-11P	Tromjesečni izvještaj o prekrcaju
PSR	Prostorni statistički registar
RAD-1G	Godišnji izvještaj o zaposlenima i isplaćenju plaći
RB-1	Statistički izvještaj o razvedenom braku
SIT	Sektor informatičkih tehnologija
SOP	Strukovno osposobljavanje u poduzećima
IAKS	Integrirani administrativni i kontrolni sustav
IDU-OK	Investicije u zaštitu okoliša i izdaci za dobra i usluge u okolišu

IKT-DOM	Godišnje istraživanje o korištenju informacijskih i komunikacijskih tehnologija u kućanstvima i kod pojedinaca
IKT-POD	Godišnje istraživanje o korištenju informacijskih i komunikacijskih tehnologija u poduzećima
INOV	Inovacijske aktivnosti poduzeća
INSPIRE	Infrastruktura za prostorne informacije u Europskoj zajednici
INV-P	Godišnji izvještaj o investicijama u dugotrajnu imovinu
KLASUS	alat za sve korisnike klasifikacija koji na jednome mjestu nudi mogućnost pregledavanja i pretraživanja klasifikacija prema nazivu i šifri
MMF	Međunarodni monetarni fond
MB	matični broj
MBS	matični broj subjekta upisa
MMF	Međunarodni monetarni fond
NKD	Nacionalna klasifikacija djelatnosti
NN	Narodne novine
NSK	Nacionalna i sveučilišna knjižnica
OECD	Organizacija za ekonomsku suradnju i razvoj
OIB	osobni identifikacijski broj
PA/M-11	Tromjesečni izvještaj o cestovnome linijskom prijevozu putnika
PA/T-11	Statističko istraživanje o cestovnom prijevozu robe
PC AXIS	besplatni softver koji je razvio nacionalni statistički ured Švedske
PDV	porez na dodanu vrijednost
PO-22/STR	Izvještaj o strukturi poljoprivrednih gospodarstava i metodama poljoprivredne proizvodnje
PRS	Prostorni statistički registar
SPC	Odbor za statistički program
SPR	Statistički poslovni registar
SRPG	Statistički registar poljoprivrednih gospodarstava
STS	kratkoročne poslovne statistike
Š-O/KP	Statistički list za osnovne škole
Š-S/KP	Godišnji izvještaj za srednje škole
TU-11	Mjesečni izvještaj o dolascima i noćenjima turista
TU-14	Izvještaj o putničkim agencijama
UNECE	Ekonomsko povjerenstvo Ujedinjenih naroda za Europu

OPĆA NAČELA

STATISTIČKO ISTRAŽIVANJE

U provedbi statističkih poslovnih procesa središnje mjesto zauzima statističko istraživanje. Statističko istraživanje jest bilo koji sveobuhvatni skup aktivnosti dizajniran za potrebe prikupljanja podataka. Kada govorimo o izravnom prikupljanju podataka na odabranome slučajnom uzorku jedinica promatranja, tada je riječ o klasičnom konceptu, za razliku od širega koncepta koje obuhvaća različite dodatne poslovne procese, osobito u fazi prikupljanja podataka. Statistička istraživanja stoga se mogu podijeliti prema:

VRSTAMA OBUHVATA:

- popis – promatra se cijela populacija
- istraživanje na uzorku – promatra se izabrani uzorak populacije
- izvedene statistike – koriste se otprije poznati statistički rezultati
- postojeći statistički agregati (npr. statistika nacionalnih računa)

NAČINIMA PRIKUPLJANJA PODATAKA:

- intervju (telefonom ili izravno)
- samopopunjavanje (analogni ili elektronički upitnik)
- drugi načini prikupljanja (promatranje)

PERIODIKAMA:

- mjesečna
- tromjesečna
- godišnja
- drugačije periodike (tjedna, višegodišnja)

IZVORIMA PODATAKA:

- statistički (primarni) izvori
- administrativni i drugi sekundarni izvori podataka.

OSNOVNI PROCESNI MODEL

Osnovna struktura dokumenta pripremljena je prema prilagođenome generičkom modelu statističkih poslovnih procesa (GSBPM-u), koji podrobno opisuje i definira skup poslovnih procesa potrebnih za izradu službene statistike. Navedeni model pruža standardni okvir i jedinstvenu terminologiju kojima se omogućuje modernizacija statističkih proizvodnih procesa te razmjena metoda i komponenata. Isti se model primjenjuje za integraciju podataka i standardizaciju metapodataka, usklađivanje statističkih računalnih infrastruktura, osiguranje okvira za procjenu kvalitete poslovnih procesa te za daljnja poboljšanja.

Na temelju strukture modela GSBPM Državni zavod za statistiku proveo je analizu mogućnosti njegove primjene u hrvatskome statističkom sustavu te je na temelju provedenih analiza izradio odgovarajući nacionalni model koji slijedi preporuke izvornog modela UNECE/Eurostat/OECD¹ Skupine za statističke metapodatke (METIS), a temelji se na modelu poslovnog procesa kojim se koristi statistika Novog Zelanda.

¹ Poveznica <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>

GSBPM je kao model osmišljen neovisno o izvoru podataka te je prilagođen potrebama hrvatskoga statističkog sustava. Može se koristiti za opis i procjenu kvalitete procesa na temelju provedenih anketa, popisa stanovništva, administrativnih evidencija i ostalih nestatističkih ili kombiniranih izvora podataka.

Ovaj model obuhvaća osam procesa koji su podijeljeni u različite potprocese. Naslovi i podnaslovi u dokumentu preuzeti su iz općeg modela. Svako poglavlje sadržava kratak opis cijelog procesa, svi poddijelovi sastoje se od dva dijela: prvi dio sadržava opći opis, dok drugi dio sadržava upute za osiguranje kvalitete koje bi trebale statističarima poslužiti kao svojevrsni popis elemenata kvalitete. Iako model u stvarnoj primjeni prati slijed procesa i potprocesa, faze su često međusobno povezane i isprepletene. Zato ih nije bilo moguće predstaviti jednostavnim linearnim modelom. Uglavnom, nacionalna statistika treba imati uspostavljeni model i potrebno ga je slijediti koliko god je to moguće u praksi.

Slika 1: Generički model statističkog poslovnog procesa – GSBPM

Upravljanje podacima/Upravljanje metapodacima							
1. Definiranje potreba za podacima	2. Priprema i izrada statističkih metodologija	3. Izrada potrebnih instrumenata za provedbu	4. Prikupljanje podataka	5. Obrada podataka	6. Analiza	7. Diseminacija	8. Evaluacija
1.1. Definiranje potreba za informacijama i potrebnih rezultata	2.1. Definiranje i izrada metodologije za prikupljanje podataka i provedbu istraživanja	3.1. Izrada projektnih zahtjeva	4.1. Izbor ciljane populacije/uzorka	5.1. Integriranje prikupljenih podataka	6.1. Statistička analiza rezultata	7.1. Dizajn i proizvodnja proizvoda za diseminaciju	8.1. Prikupljanje informacija za evaluaciju
1.2. Provjera raspoloživih izvora podataka	2.2. Definiranje okvira i metodologije za izbor uzorka	3.2. Izrada instrumenata za prikupljanje podataka	4.2. Priprema prikupljanja podataka	5.2. Kontrola, uređivanje i ispravljanje podataka	6.2. Kontrola kvalitete rezultata	7.2. Upravljanje diseminacijom	8.2. Evaluacija rezultata
1.3. Utvrđivanje koncepta za proizvodnju rezultata te analiza i testiranje mogućnosti	2.3. Izrada metodologije za obradu podataka	3.3. Izrada programske podrške	4.3. Prikupljanje primarnih podataka	5.3. Imputiranje i utežavanje	6.3. Detaljna analiza i tumačenja podataka za publiciranje	7.3. Promidžba diseminacije	8.3. Akcijski plan za poboljšanje
		3.4. Testiranje alata za prikupljanje i obradu podataka	4.4. Preuzimanje podataka iz administrativnih i drugih sekundarnih izvora	5.4. Proizvodnja izvedenih varijabli	6.4. Zaštita povjerljivih podataka	7.4. Upravljanje odnosima s korisnicima	
		3.5. Konfiguriranje proizvodnih procesa	4.5. Unos prikupljenih podataka	5.5. Izračunavanje agregata			
				5.6. Izrada konačnih datoteka podataka			
				5.7. Proizvodnja i ažuriranje registara i baza podataka			

OCJENA KVALITETE STATISTIČKIH PROCESA I PROIZVODA

Posljednjih se nekoliko godina pri proizvodnji službenih statističkih podataka kontinuirano unapređuju modeli i strukture za praćenje i mjerenje kvalitete statističkih procesa i proizvoda.

Koliko je kvaliteta bitna pokazuje zajednička inicijativa nacionalnih statističkih ureda za usvajanjem i usklađivanjem različitih praksi pri objavljivanju podataka o procijenjenoj razini kvalitete. Prethodnih je godina statistička kvaliteta bila usmjerena na preciznost statističkih rezultata, dok je u današnje doba kvaliteta višedimenzionalna i odnosi se na primjerenost statističkih rezultata u smislu primjenjivosti za različite svrhe i za različite skupine korisnika. Time je preciznost postala samo jedan od nekoliko komponenti kvalitete koje su mjerljive i egzaktno definirane.

Osnovnu ulogu u razvoju i jasnu viziju u metodološkome i praktičnom smislu pri uvođenju suvremenih koncepata mjerenja kvalitete imale su međunarodne organizacije, kao što su Eurostat, MMF i OECD. U tom je smislu u okviru ESS-a u posljednjih nekoliko godina razvijen standardni i jedinstveni model za procjenu kvalitete u statistici, koji se temelji na šest dimenzija kvalitete:

Relevantnost je dimenzija kvalitete unutar koje se žele dobiti odgovori na sljedeća pitanja:

- Hoće li rezultati postavljenoga statističkog istraživanja ispuniti očekivanja korisnika i dati cjelovite odgovore na korisnički zahtjev?
- Može li obuhvat istraživanja zadovoljiti korisničke potrebe?
- U okviru ove dimenzije kvalitete važno je redovito pratiti potrebe korisnika upotrebom alata kao što je zadovoljstvo korisnika.

Točnost je dimenzija za koju smo procijenili da se (ne) podudara sa snimljenim stanjem i ne odgovara stvarnom stanju jer su nepoznate vrijednosti populacije. Nedosljednost se također naziva statistička pogreška. Potrebno je procijeniti različite vrste pogrešaka koje se događaju pri provedbi istraživanja i one bi trebale biti jasno objašnjene i dostupne korisnicima.

Pravodobnost i poštovanje rokova objavljivanja podataka – pravodobnost se odnosi na interval između referentnog razdoblja na koji se odnose i datuma objavljivanja, dok je poštovanje rokova objavljivanja interval mjera između stvarnoga i najavljenog datuma objavljivanja podataka koji se određuje u kalendaru objavljivanja statističkih podataka prema pravnoj osnovi statističkog istraživanja.

Usklađenost se u statistici odnosi na sposobnost da se mogu pouzdano kombinirati podaci na različite načine i za različite korisnike. Ta dimenzija kvalitete ima sljedeće poddimenzije:

- usklađenost između privremenih i konačnih rezultata
- dosljednost godišnjih i strukturnih statistika
- usklađenost unutar istoga statističkog područja
- usklađenost s rezultatima iz područja nacionalnih računa.

Usporedivost govori u kojoj su mjeri prikazani podaci usporedivi u vremenu i prostoru.

Sljedeće su posebno važne poddimenzije:

- geografska usporedivost (mogućnosti usporedbe između država, regija i slično)
- vremenska usporedivost (mogućnost usporedbe podataka u različitim razdobljima)
- usporedivost između područja (npr. usporedba podataka iz dviju industrijskih grana, dvaju različitih nastavnih programa, različitih dobnih skupina i slično).

Dostupnost i jasnoća podataka – dostupnost se odnosi na fizičke uvjete u kojima su ti podaci dostupni korisniku (mogućnost naručivanja, raspored radnih mjesta, cjenovna politika itd.). Jasnoća opisuje podatkovno okruženje u kojem korisnik može pristupiti informacijama. Treba opisati posebno jesu li na raspolaganju odgovarajući metapodaci te je li korisniku odobren pristup kvaliteti objavljenih rezultata.

Na temelju opisanih dimenzija kvalitete nacionalni statistički uredi, organizirani u ESS, izradili su za potrebe korisnika standardizirane dokumente za praćenje kvalitete objavljenih statističkih podataka. Te dokumente nazivamo standardnim izvještajima o kvaliteti (poveznica [Državni zavod za statistiku - Izvještaji o kvaliteti prema statističkim područjima \(gov.hr\)](#)).

Kodeks prakse europske statistike bitan je korak u osiguranju kvalitete statističkih rezultata Europskoga statističkog sustava (u nastavku teksta: Kodeks). Kodeks je prvi put usvojio SPC u 2005. i revidirao ga je ESSC u 2011. i 2017. ESSC je posljednju reviziju potvrdio 16. studenoga 2017. Revidirani Kodeks ima 16 načela koja se tiču institucionalnog okružja, statističkih procesa i statističkih rezultata. Cilj Kodeksa jest osigurati da su statistike proizvedene unutar ESS-a relevantne, pravodobne i točne te da su u skladu s načelima profesionalne neovisnosti, nepristranosti i objektivnosti. Skup pokazatelja najbolje prakse i standarda za svako načelo pruža upute i referencije za ocjenu primjene Kodeksa.

OPĆE UPUTE ZA OSIGURANJE KVALITETE

U poglavljima je dan prikaz Uputa za osiguranje kvalitete u svakom od potprocesa tako da su prilagođeni cjelokupnom modelu poslovnih procesa. Predstavljene su opće upute i postupci pri izradi novog istraživanja kao i postupci dorade postojećih istraživanja.

Za učinkovitu uspostavu pojedinoga poslovnog procesa potrebno je pravodobno planirati na godišnjoj i višegodišnjoj razini. Ponajprije je potrebno procijeniti realne mogućnosti provedbe poslovnih procesa s obzirom na raspoloživost ljudskih, financijskih i informatičkih potencijala te u skladu s preduvjetima pripremiti realan raspored aktivnosti.

Postupci unutar poslovnih procesa trebaju u najvećoj mogućoj mjeri omogućiti ponovljivost i dosljednost. Ponovljivost statističkih poslovnih procesa znači da svaku verziju poslovnih procesa kod kojih su nepromijenjeni ulazni podaci trebaju uvijek biti isti konačni rezultati. Dosljednost znači da je u tome poslovnom procesu svaka promjena u podacima zabilježena jasno i nedvosmisleno. To omogućuje da se u svakom procesu izvršenja priprema zapis o svim izmjenama podataka i prati utjecaj tih promjena na konačne rezultate.

Svakako se preporučuje koristiti se dobrim primjerima iz prakse. Zaposlenici trebaju biti redovito obaviješteni o formiranju i provedbi statističkih poslovnih procesa te o postojanju standarda. Uzimanje dobre prakse u ostalim područjima može se potaknuti poboljšanjem protoka informacija između različitih statističkih područja. To se može najlakše postići organiziranjem radionica, seminara i drugih oblika internog usavršavanja.

1. Definiranje potreba za podacima i potrebnih rezultata

Pri analiziranju zahtjeva i potreba za podacima potrebno je usredotočiti se na vanjske korisnike statističkih podataka, tj. na podatke koji su prikupljeni prema statističkim načelima odnosno na one podatke koji su nužni za donošenje odluka te za provođenje javnih politika.

Analiza potreba i zahtjeva pružaju osnovu na kojoj će se graditi buduće aktivnosti i donositi odluke o smjernicama za dizajniranje istraživanja. Očekivanja i potrebe korisnika za statističkim rezultatima sve su veća i usko specijalizirana zbog promjena u društvu, globalnih procesa i nove inicijative na međunarodnoj i nacionalnoj razini.

Pravodobno prepoznavanje potreba za podacima i priprema odgovarajućih podataka ključni su za sustavno i strateško odlučivanje. Potrebe za podacima definiraju se izravno s korisnicima putem različitih komunikacijskih kanala, vodeći pritom računa o sposobnostima i okvirima nacionalnoga statističkog sustava.

Pri proučavanju podataka treba provjeriti i statističke i administrativne, već postojeće izvore podataka jer to može znatno utjecati na racionalno korištenje podataka. Svima je bitno da se omogući odabir pravog izvora informacija koji će sadržavati definiciju jedinice promatranja, imati mogućnost određivanja identifikatora, relevantnih sadržaja, kvalitetnih resursa i druga obilježja.

Pri odlučivanju između izravnog prikupljanja podataka na terenu i eventualne mogućnosti korištenja administrativnih izvora podataka potrebno je uzeti u obzir da je cijena korištenja administrativnih izvora podataka jeftinija. Važnu ulogu ima odluka o korištenju administrativnog izvora podataka, zatim kvaliteta resursa, njegova relevantnost i metodološka prikladnost te održivost.

Prije planiranja provedbe statističkog istraživanja potrebno je provesti provjeru metodologije, odnosno svih faza istraživanja. Provjera metodologije može se odnositi na postojeće istraživanje ili na metodologiju za uvođenje novoga statističkog istraživanja.

1.1. Definiranje potreba za informacijama i potrebnih rezultata

Opis

Utvrđivanje potreba za podacima treba pokrenuti kada podaci ne postoje ili kada postojeći podaci ne ispunjavaju sve potrebe korisnika podataka.

Potreba za podacima utvrđuje se od raznih korisnika: ministarstava, HNB-a, nacionalnih i međunarodnih institucija, istraživača, znanstvenika i opće zainteresirane javnosti. U procesu potreba za podacima može se točno odrediti što korisnik može očekivati od nacionalnoga statističkog sustava. Za tu svrhu u ovaj bi proces konzultacija trebali biti uključeni zainteresirani korisnici u različitim oblicima suradnje od statističkih savjetodavnih odbora, statističkih vijeća pa do zahtjeva nacionalnih i međunarodnih institucija. Državni zavod za statistiku upoznat je s korisničkim očekivanjima, osobito s vrstama informacijama koje korisnik treba, kada ih treba i za koju svrhu. Stoga je potrebno uspostaviti kontinuiranu suradnju s korisnicima jer ako se prethodno utvrđene potrebe korisnika mijenjaju, potrebno je promijeniti i postupke.

Nakon toga utvrđuju se poslovni procesi koji su potrebni za isporuku skupova podataka potrebnih za ispunjavanje potreba korisnika, a u skladu s mogućnostima koje nudi Državni zavod za statistiku. U ovoj fazi potrebno je postići međusobni dogovor u vezi s podacima kvalitete statističkih podataka koji se temelje na njezinih šest dimenzija, a radi izbjegavanja nesporazuma kod isporuke konačnog skupa podataka.

Upute za osiguranje kvalitete

Pri utvrđivanju potreba za podacima potreban je sveobuhvatan i sustavan pristup koji uključuje sve zainteresirane korisnike. Kako bi se Državni zavod za statistiku upoznao s potrebama korisnika, potrebno je redovito razmjenjivati informacije s korisnicima te upoznati potrebe korisnika, a posebno je potrebno saznati koje podatke korisnik treba, kada i za koju svrhu i na koji način. Analitičkim pristupom treba uzeti u obzir i sve povezane potrebe korisnika koje se uz malo dodatnog napora mogu jednostavno ispuniti i specifične potrebe za podacima za koje je potrebno analizirati troškove i koristi koje može donijeti prikupljanje podataka.

1.2. Provjera raspoloživih izvora podataka

Opis

Nakon provedenih analiza u vezi s novim zahtjevima korisnika potrebno je provjeriti sva statistička istraživanja koja se koriste administrativnim izvorima podataka, a zatim provjeriti dostupnost podataka u nekim drugim izvorima. Ako su dostupni drugi izvori, potrebno je provjeriti pod kojim uvjetima, zatim je potrebno utvrditi metodološke razlike u odnosu na statističku metodologiju te periodiku u vezi s prikupljanjem podataka i drugo. Provjerom raspoloživih administrativnih izvora podataka utvrđuje se koliko se administrativni podaci mogu koristiti kao izravni izvor podataka. Ako podaci iz korisničkog zahtjeva uopće nisu raspoloživi, potrebno je organizirati prikupljanje tih skupova podataka i uključiti ih ili u jedno od uhodanih statističkih istraživanja ili pripremiti novo statističko istraživanje.

Upute za osiguranje kvalitete

Kvalitetu administrativnih izvora podataka potrebno je uvijek provjeriti, a u slučaju odgovarajuće kvalitete, upotreba administrativnih izvora podataka u svakom slučaju ima prednost u odnosu na izravno prikupljanje podataka od ispitanika.

Što se tiče kvalitativnog aspekta, potrebno je voditi računa o kvaliteti administrativnih izvora podataka, učestalosti ažuriranja, relevantnosti izvora podataka, metodološkoj usporedivosti sadržaja izvora podataka te o vremenskoj dostupnosti podataka. Razinu stabilnosti podataka kod administrativnih izvora podataka treba redovito pratiti i procjenjivati. Kvalitetu administrativnih podataka potrebno je redovito provjeravati uspoređivanjem pojedinačnih uzastopnih skupova podataka ili metodom uzorkovanja na anketnom istraživanju koje je posebno dizajnirano za tu svrhu.

Korisnicima podataka treba posvetiti posebnu pozornost, osobito u slučajevima kada su podaci potpuno preuzeti iz administrativnih izvora za statističke svrhe.

1.3. Utvrđivanje koncepta za proizvodnju rezultata te analiza i testiranje mogućnosti

Treba razmotriti može li se postojeće istraživanje nadopuniti kako bi se dobili željeni podaci, može li se upotrijebiti administrativni izvor podataka ili treba planirati novo istraživanje. Nadalje je potrebno pripremiti analizu kako bi se utvrdila hijerarhija izvora podataka.

Osim toga, potrebno je u sadržajnom smislu pripremiti usporedbu administrativnih i statističkih podataka za pojedine dostupne varijable. Treba provjeriti poštivanje metodologije kao i upravljanje administrativnim izvorima podataka za svrhu prezentacije statističkih podataka.

U skladu sa zakonskim odredbama potrebno je osigurati poštivanje statističke povjerljivosti te fizičku i informacijsku zaštitu podataka koji su preuzeti iz administrativnih izvora tako da se ograniči pristup podacima te organizira praćenje pristupa podacima.

Upute za osiguranje kvalitete

S posjednikom administrativnih izvora podataka potrebno je sklopiti odgovarajući sporazum o razmjeni podataka i ovisno o vrsti podataka potrebno je definirati katalog podataka. Dobar praktični primjer jest Katalog međusobnih potraživanja statističkih podataka iz evidencija Porezne uprave i Državnog zavoda za statistiku. Katalog, primjerice, sadržava sljedeće podatke: naziv poreza/obrasca, slog podataka, namjenu, periodiku dostave, rok za dostavljanje prvih rezultata podataka, rok za dostavljanje konačnih rezultata podataka, oblik i način dostave podataka, prijamni odjel u Državnom zavodu za statistiku, korisnički odjel u Državnom zavodu za statistiku, odgovornu osobu za korištenje datoteke u Državnom zavodu za statistiku, odgovornu osobu u Poreznoj upravi, osobu zaduženu za prijam u Poreznoj upravi, odgovornu osobu u APIS-IT te osobu odgovornu za formiranje podataka – APIS-IT. Državni zavod za statistiku obvezan je u ovom slučaju o razini kvalitete podataka redovito izvještavati posjednike administrativnih izvora podataka. Pri utvrđivanju izvora i predlaganju promjena u obzir se mora uzeti interes statistike.

Preuzimanje podataka i njihovu kvalitetu treba pratiti tijekom cijele godine. Ako postoje nelogičnosti, potrebno ih je prijaviti vlasniku podataka.

2. Priprema i izrada statističkih metodologija

Analizom potreba korisnika i drugih opažanja utvrđene su potrebe za statističkim istraživanjima koja se periodično planiraju. Pri pripremi statističkih istraživanja u okviru ESS-a vodi se računa o tome da se provode na temelju europskih propisa ili sporazuma i da su u skladu s propisanim standardima. Svrha, sadržaj istraživanja, ciljna populacija te ostali elementi moraju biti u skladu s potrebama i navedeni u godišnjemu provedbenom planu. Detaljan plan za svaku fazu provedbe statističkog istraživanja mora biti naveden u planu rada. Kod anketa koja se provode na uzorku i za koje je potrebno dodatno financiranje (npr. za intervjuiranje osoba, kućanstava i poljoprivrednih gospodarstava), dodatna oprema i dodatno osoblje, sve to zajedno (ljudske resurse, financije, IT opremu) potrebno je nekoliko godina unaprijed planirati. Za ankete je, osim plana rada, potrebno izraditi i detaljan financijski plan.

Početne faze planiranja statističkih istraživanja i anketa obično su sljedeće: definiranje sadržaja varijabli i rezultata istraživanja; definiranje ciljne populacije i populacije promatranja; utvrđivanje izvora podataka i metoda prikupljanja podataka.

U nastavku se, ovisno o gore navedenim definicijama, određuju daljnji koraci. Ako se preuzima informacija izravno od izvještajnih jedinica, potrebno je odrediti metodu odabira jedinica (vjerojatnosna metoda uzorkovanja), pripremiti nacrt upitnika i popis prateće dokumentacije. Kada se dio podataka ili informacija za određeno statističko istraživanje preuzima iz postojećih izvora, treba provjeriti jesu li podaci već dostupni u Državnom zavodu za statistiku. Ako podaci o traženim varijablama u Državnom zavodu za statistiku nisu dostupni, potrebno je u dogovoru s institucijom od koje želimo preuzeti podatke pokrenuti postupak u vezi sa sklapanjem sporazuma te, ovisno o situaciji, katalog za razmjenu podataka i ostale potrebne dokumente.

U ranoj fazi planiranja i pripreme statističkog istraživanja potrebno je detaljno definirati metodologiju od odabira jedinica promatranja pa do statističke obrade podataka. Za kvalitetnu provedbu istraživanja bitno je da se već u osmišljavanju ovog istraživanja planira provesti eksperimentalno statističko istraživanje s pomoću kojeg se testiraju sve faze ili barem one faze u provedbi obrade statističkog istraživanja koje još prethodno nisu bile testirane.

Pri planiranju i pripremi statističkog istraživanja Državni zavod za statistiku vodi računa o tome da opterećenje ispitanika bude što manje i da preuzimanje podataka od izvještajnih jedinica bude dogovoreno na obostrano zadovoljstvo i uz što je moguće niže troškove provođenja istraživanja te uz odgovarajuću kvalitetu koja će dati statističke pokazatelje o području koje se promatra istraživanjem.

2.1. Definiranje i izrada metodologije za prikupljanje podataka i provedbu istraživanja

Opis

Definiranje i izrada metodologije za prikupljanje podataka i provedbu istraživanja uključuje izradu svih potrebnih metodologija (metode, instrumenti za prikupljanje, varijable, definicije, opisi, upute, sporazumi i ugovori s davateljima podataka, sadržaj upitnika, plan diseminacije itd.).

Priprema opisa metapodataka o prikupljenim i izvedenim varijablama i klasifikacijama ključan je preduvjet za sljedeće faze. Utvrđuju se najprikladnije metode i instrumenti za prikupljanje podataka. Aktivnosti ovise o metodama za prikupljanje podataka (CAPI, PAPI, CATI i CAWI), uključujući testiranje instrumenata. Izrađuju se svi formalni ugovori o dostavljanju podataka, kao što su memorandum i potvrda pravne osnove za prikupljanje podataka.

Provjera metodologije statističkog istraživanja obuhvaća provjeru teorijskog dijela metodologije (npr. ciljnu populaciju, definicije, metodologiju uzorkovanja, upotrebu postojećeg izvora podataka i obradu podataka) i provjeru praktičnog dijela metodologije statističkog istraživanja (npr. format upitnika, metoda prikupljanja podataka, proces obrade podataka, vrste publikacija i slično).

Postupak provjere metodologije odnosi se na primjerenost metodologije u odnosu na uočene potrebe. Tijek ovog postupka također ovisi o tome je li metodologija pripremljena na temelju propisa EU-a ili ne. Ako je pripremljen na temelju propisa EU-a, mora se uzeti u obzir je li ovaj propis već usvojen ili je u postupku usvajanja ili izmjene. Kod slučajeva kada je potrebno primijeniti odredbe propisa, tada se ispituje mogućnost usvajanja zahtjeva iz propisa, a u drugom slučaju utvrđuje se može li se istraživanje provesti onako kako je predloženo u uredbi ili u izmjeni i dopuni uredbe koja je u postupku izrade i usvajanja. U slučaju kada se priprema očitovanje za Republiku Hrvatsku, tada se stajališta moraju usvojiti na različitim razinama, od radnih, direktorskih, stručnih radnih skupina u ESSC-u i u Vijeću EU-a.

Provedbu i primjenu metodologije provjerava nositelj istraživanja, odnosno osoba koja je odgovorna za istraživanje te statističari koji su zaduženi za određeno područje i sudjeluju u procesu. Što se tiče izvedivosti i primjerenosti metodologije, nositelj istraživanja može predložiti sazivanje stručne radne skupine na kojoj će se raspraviti o predloženom rješenju i dati svoje mišljenje o metodološkim prijedlozima. Prije donošenja konačne odluke o metodologiji prijedlog bi trebao razmotriti stručni kolegij.

Upute za osiguranje kvalitete

Kada se provjerava metodologija, prednost imaju postojeće metode i dobre prakse, dok se u fazi pripreme mora voditi računa o tome da definicije budu napisane na razumljiv način, da se jasno definira što je obuhvaćeno, a što ne.

Kod prikupljanja podataka s pomoću upitnika važno je da su pitanja razumljiva i upute za popunjavanje precizne. Međutim, pri preuzimanju podataka iz postojećih administrativnih izvora bitno je sklopiti odgovarajući sporazum o načinu preuzimanja podataka sa svakim određenim tijelom koje je posjednik administrativnih izvora podataka potrebnih za provedbu statističkih istraživanja utvrđenih godišnjim provedbenim planom statističkih aktivnosti Republike Hrvatske te ujedno definirati katalog podataka o vrsti traženih podataka, načinu i rokovima dostave podataka i o odgovornim osobama za pripremu i preuzimanje podataka.

Pri provjeri provedbe pojedinih faza istraživanja uvijek je potrebno uzeti u obzir postavljene mogućnosti provedbe ili potražiti druga rješenja.

U slučajevima kada se statističko istraživanje provodi na razini EU-a, Državni zavod za statistiku trebao bi od početka biti aktivno uključen, odnosno od trenutka kada se o toj temi raspravlja u radnoj skupini koja djeluje na razini EU-a za svako pojedino statističko istraživanje. Prema tome, Državni zavod za statistiku može utjecati na metodologiju (ciljnu populaciju, definicije, klasifikacije, periodičnost objavljivanja i slično). Osim toga, provjerava se dio metodologije koji nije propisan (npr. način prikupljanja podataka, metodologija statističke obrade).

Za uspješnu provedbu statističkog istraživanja potrebno je pripremiti plan potrebnih financijskih, informatičkih i ljudskih resursa. Detaljan raspored aktivnosti ključan je za učinkovitu provedbu statističkog istraživanja.

Svako statističko istraživanje potrebno je navesti u godišnjemu provedbenom planu statističkih aktivnosti Republike Hrvatske. Statistička istraživanja su ovisno o načinu provedbe u godišnjemu provedbenom planu statističkih aktivnosti Republike Hrvatske strukturirana na sljedeći način:

- I. statističko istraživanje na temelju neposrednog prikupljanja podataka (tromjesečno, polugodišnje, jednogodišnje, višegodišnje)
- II statističko istraživanje na temelju administrativnih izvora ili metodom promatranja i praćenja
- III. razvojne aktivnosti, popisi i druga opsežnija statistička istraživanja.

U suradnji s ustrojstvenim jedinicama nositelj istraživanja u planu rada definira popis aktivnosti, statističare zadužene za njihovu provedbu, slijed aktivnosti i rokove.

Prije početka statističkog istraživanja potrebno je pripremiti plan izvora podataka i financijski plan. Aktivnosti je potrebno pravodobno uključiti u proračun te planirati potrebe za dodatnim resursima (primjerice prijenosnim računalima). Ovo je uobičajena procedura kod provedbe pojedinačnoga statističkog istraživanja, zatim kod većih revizija podataka, višegodišnjih statističkih istraživanja, kod statističkih istraživanja koja imaju kontinuirano zahtjeve preko tzv. ad hoc modula i kod istraživanja koja zahtijevaju sudjelovanje anketnog studija.

Pri planiranju izvora podataka i aktivnosti važno je osigurati suradnju nositelja statističkih istraživanja s ustrojstvenom jedinicom za podršku statističkim istraživanjima i postići optimalni omjer i raspodjelu poslova s obzirom na izvor podataka i raspoloživo radno vrijeme.

Provedbe aktivnosti treba pažljivo pratiti tijekom cijele godine. Utvrđene nelogičnosti potrebno je prijaviti odgovornim osobama u skladu s aktualnim organigramom i zaduženjima po pojedinim potprocesima modela GSBPM. Pri pripremi plana rada za sljedeća proračunska razdoblja koju obavljaju nositelji istraživanja u obzir se moraju uzeti sve utvrđene nepravilnosti i otežavajuće okolnosti u provođenju statističkog istraživanja. Kod planiranja financijskih sredstava potrebno je odrediti visine pojedinih proračunskih stavki za vanjske anketare koji su uključeni u statističko istraživanje na temelju sklopljenih ugovora o djelu. Visina proračunske stavke za izračun financijskih stavki dobiva se na temelju veličine uzorka i predviđene stope odgovora i neodgovora za svako pojedino statističko istraživanje. U strukturi anketara prema načinu plaćanja za ovu svrhu potrebno je koristiti se pregledom simulacija troškova za svaki pojedinačni segment ankete. Kod određenih istraživanja postoji mogućnost za planiranjem određenih dodatnih financijskih stavki (npr. naknada kućanstvima za pravilno ispunjen dnevnik potrošnje i slično).

2.2. Definiranje okvira i metodologije za izbor uzorka

Opis

Definiranje okvira i metodologije za izbor uzorka utvrđuje i određuje ciljnu populaciju, definira okvir uzorka (i, gdje je potrebno, registar iz kojega se uzorak uzima) te određuje najprikladnije kriterije i metodologiju uzorkovanja (što može uključivati i cijelu ciljnu populaciju). Najčešći izvori jesu geoprostorni, administrativni i statistički registri te popisi stanovništva. Ti se izvori mogu i kombinirati. Treba provesti analizu da bi se provjerilo obuhvaća li okvir uzorka ciljnu populaciju. Treba izraditi i plan uzorkovanja.

Nakon utvrđivanja ciljne populacije potrebno je dobiti popis jedinica populacije s većim brojem njihovih aproksimacija jer izvori podataka s pomoću kojih se izrađuje okvir za izbor uzorka nisu idealne kvalitete. Varijable okvira za izbor uzorka koje opisuju svojstva jedinica nazivaju se pomoćnim varijablama².

² U literaturi se mogu naći slobodnije definicije pomoćne varijable. Primjerice, varijabla čija je vrijednost populacije poznata prije provođenja istraživanja, a podaci odabranih jedinica uzorka dobiveni su provođenjem ankete, naziva se i pomoćnom varijablom.

Pri pripremi metodologije odabira jedinica u obzir se uzima način prikupljanja podataka. Traženi podaci mogu se dobiti iz različitih administrativnih izvora podataka, izvedeni ili modelirani iz postojećih statističkih izvora podataka, inovativnim pristupima, uključujući znanstveno utemeljene i dobro dokumentirane metode, kao što su imputacija, procjena i modeliranje ili izravnim/neizravnim kontaktiranjem s jedinicama promatrane populacije.

Bez obzira na način na koji se prikupljaju podaci, statističko istraživanje može se provesti ili kao ogledno istraživanje ili kao popis. Ako se organizira istraživanje na bazi uzorka, može se učiniti na dva načina: nevjerovatnosnim ili vjerovatnosnim postupkom uzorkovanja. Odabir jedinica promatranja za nevjerovatnosni uzorak obično je vrlo jednostavan, jeftin i ne zahtijeva puno vremena, ali istodobno je takav odabir subjektivan. Metodolog treba odrediti reprezentativnost uzorka za predmetno istraživanje, ovisno o obilježjima cijele populacije.

Za razliku od nevjerovatnosnog uzorkovanja, upotreba vjerovatnosnog uzorkovanja³ zahtjevnija je jer se jedinice uzorka odabiru uz poznate, pozitivne vjerovatnosti, koje se utvrđuju prije odabira, pa odabir zahtijeva više vremena. Međutim, budući da se plan vjerovatnosnog uzorkovanja temelji na cijeloj teoriji uzorkovanja, mogu se donijeti puno pouzdaniji zaključci o obilježjima cijele populacije i izračunati uzoračke pogreške. Odabir jedinica za vjerovatnosni uzorak može se provesti na više načina. Metoda odabira vjerovatnosnog uzorka naziva se planom uzorka. Izbor načina ovisi o:

- pomoćnim varijablama iz okvira uzorka
- varijabilnosti populacijskih jedinica (ovdje obično mislimo na varijabilnost vrijednosti proučavane varijable)
- statistici i područjima na kojima se statistički podaci objavljuju
- troškovima istraživanja.

Dio plana uzorkovanja jest i određivanje veličine uzorka koja izravno utječe na cijenu ankete. Cilj je dobiti najbolje moguće procjene parametara populacije uz najmanju moguću veličinu uzorka. Ako u okviru uzorkovanja imamo dovoljno podataka, veličina uzorka može se izračunati (obično prema potrebnoj točnosti statističkih podataka). Inače se određuje i na temelju iskustva (npr. druga statistička istraživanja).

Najčešći oblici uzorkovanja koji se koriste u Državnom zavodu za statistiku jesu:

- jednostavni slučajni uzorak
- sistematski uzorak
- stratificirani uzorak
- uzorak skupina
- višestapno uzorkovanje.

Upute za osiguranje kvalitete

Pri određivanju okvira modela potrebno je procijeniti koji su resursi (registri, baze podataka) dostupni i kako se mogu međusobno povezati.

Ako je moguće, kvalitetu postavljenog okvira uzorka treba procijeniti čim je pripremljen. Ako se ocijeni da je sadržaj okvira nezadovoljavajuće kvalitete, potrebno je pri dizajniranju upitnika uključiti dodatna pitanja koja će za sljedeće referentno razdoblje omogućiti pouzdanu procjenu kvalitete uzorka.

Ako je okvir za izbor uzorka sastavljen na temelju anketne metode (npr. osobe iz telefonskog imenika umjesto sve osobe iz baze popisa stanovništva, kućanstava i stanova), tada treba napraviti procjenu utjecaja na kvalitetu rezultata.

³ Kolokvijalno se primjenjuje izraz "slučajno uzorkovanje" umjesto "vjerovatnosnog uzorkovanja", premda se "slučajno uzorkovanje" u literaturi spominje kao poseban slučaj vjerovatnosnog uzorkovanja u kojem svi uzorci određene veličine imaju jednaku vjerovatnost odabira postupkom uzorkovanja.

Najvažniji izvori za dizajn okvira za uzorak jesu sljedeći registri i baze podataka: za uzorkovanje osoba i kućanstava to je baza podataka popisa stanovništva, kućanstava i stanova; za uzorkovanje trgovačkih društava to je SPR; za uzorkovanje obiteljskih poljoprivrednih gospodarstava to je SRPG, baze podataka na temelju članka 117. Zakona o poljoprivredi (NN, br. 118/18.) te IAKS-a.

Pri dizajnu okvira za izbor uzorka treba obratiti pozornost na pomoćne varijable jer pristup pomoćnim varijablama omogućuje izračun veličine uzorka, provedbu složenijeg postupka uzorkovanja i određivanje preciznijih procjena statističkih podataka.

Baza podataka popisa stanovništva, kućanstava i stanova prikladan je izvor za odabir fizičkih osoba u uzorak, s tim da se neizravno za odabir uzorka koriste kućanstva jer odabrana osoba prema adresi stanovanja vodi do pripadajućega kućanstva. S pomoću terenskih podataka o broju članova kućanstva procjenjuje se vjerojatnost odabira kućanstava na uzorku.

Prednost treba dati najjednostavnijemu mogućem dizajnu uzorka. Odluku za planiranje složenijih postupaka uzorkovanja treba dobro razmotriti. Na primjer, upotreba metode uzorkovanja u više etapa može smanjiti troškove istraživanja, ali i točnost procjena.

Pri određivanju veličine uzorka treba uzeti u obzir očekivanu stopu odgovora i rezultate analize neodgovora iz prethodnoga referentnog razdoblja. Dakle, potrebno je procijeniti koja razina odgovora pruža zadovoljavajuću kvalitetu statističkih rezultata.

2.3. Izrada metodologije za obradu podataka

Opis

Statistička obrada podataka odnosi se na sve postupke koji se koriste nakon završetka prikupljanja podataka. Izrada metodologije za obradu podataka uključuje određivanje postupaka šifriranja, uređivanja, imputacije, procjene, integracije, povezivanja geoprostornih i statističkih informacija, ovjeravanja i finaliziranja skupova podataka s rokovima i nositeljima. Osnovni je cilj da konačni statistički rezultati što točnije odražavaju obilježja promatrane populacije za dobivanje informacija, odnosno pokazatelja kojima se omogućuju praćenje i usporedivost za nacionalne i međunarodne potrebe. Navedeni pokazatelji također služe u kreiranju politika, u različitim znanstvenim analizama, te, općenito, za informiranje šire javnosti.

U fazi planiranja istraživanja potrebno je predvidjeti koji postupci imaju smisla za provedbu planiranoga statističkog istraživanja, koji se metodološki pristupi trebaju primijeniti u tim postupcima i koji bi softverski alati bili najprikladniji za provedbu.

Različita statistička istraživanja u svojoj provedbi sadržavaju sljedeće različite postupke statističke obrade.

Rješavanje neodgovora. Tijekom provođenja ankete često se događa da je stopa odgovora niska kod određenog skupa podataka statističkog istraživanja ili da nema nikakvih podataka. Postupci za smanjenje stope neodgovora trebali bi biti uključeni u aktivnosti prikupljanja podataka. Međutim, u fazi planiranja statističke obrade treba predvidjeti koji će se postupci koristiti nakon završetka prikupljanja podataka kako bi se minimizirao utjecaj neodaziva statističke rezultate i smanjila razina pristranosti. Obično se rezultati istraživanja ponderiraju kako bi se prilagodili dizajnu uzorka, a za jedinice neodgovora proizvedu se valjani rezultati za ciljnu populaciju. Jedinica neodgovora obračunava se ponovnim ponderiranjem. Problem neodgovora može se pojaviti i kod administrativnih izvora podataka.

Postupci uređivanja podataka. Postupak uređivanja podataka odnosi se na sve postupke pronalaženja i uklanjanja pogrešaka u podacima. Kada se planiraju postupci uređivanja podataka, potrebno je izabrati postupke koji će što je moguće više smanjiti utrošeno vrijeme i strukturu troškova. Stoga bi trebalo uložiti velike napore i naći rješenje kojim će se tzv. postupak "ručnog" uređivanja podataka zamijeniti uvođenjem modernijih pristupa, poput selektivnoga i automatskog uređivanja podataka.

Uređivanje podataka može se primijeniti i u fazi prikupljanja podataka primjenom logičko-računskih kontrola unutar upitnika i tako smanjiti broj pogrešaka u obradi.

Postupci agregiranja i tabeliranja. Ovaj dio procesa jest "konstanta" statističkih istraživanja, s pomoću kojih se iz obrađenih podataka za statističke svrhe na mikrorazini izračunavaju statistički agregati (tzv. konačni proizvod statističkog procesa). U fazi planiranja potrebno je predvidjeti koje će se vrste statistika izračunavati te jasno i nedvosmisleno opisati pravila za njihov izračun i popratni tekst potkrijepiti matematičkim formulama. Također je potrebno iz perspektive korisnika i proizvođača podataka osigurati zadovoljavajuću razinu detalja u tablici rezultata.

Isto tako, nužno je kod postupka statističke obrade, ovisno o vrsti statističkog istraživanja ili potreba korisnika, planirati sljedeće.

Kod anketnih istraživanja na uzorku moraju se osigurati postupci za procjenu pogreške uzorkovanja. Potrebno je odlučiti o korištenju triju osnovnih pristupa (analitički pristup, modelni pristup i pristup ponovnog postupka uzorkovanja) i softverskih alata.

Osobito treba biti oprezan kod financijskih podataka, kada se korisniku dostavljaju deflacionirani statistički rezultati. Deflacioniranje je postupak odstranjivanja utjecaja promjena cijena na vrijednosno izražene pojave. Provodi se dijeljenjem nominalnih vrijednosti s odgovarajućim indeksom cijena. U fazi planiranja potrebno je predvidjeti koji će se pristup primijeniti, primjerice, hoće li se primijeniti deflacija na mikrorazini ili makrorazini, zatim koja će se metodologija primijeniti za izračun deflatora te koji će osnovni resursi biti potrebni i kako će se taj proces uklopiti u cjelokupni statistički proces.

Najvažniji ekonomski pokazatelji često su pod utjecajem sezonskih i kalendarskih varijacija koje mogu prikriti relevantna kratkoročna i dugoročna kretanja te otežati razumijevanje ekonomskih pojava (BDP, promet industrije, obujam građevinskih radova, promet u trgovini na malo i obujam industrijske proizvodnje). Ako je riječ o povremenome statističkom istraživanju i ako su rezultati istraživanja prikazani u vremenskim serijama, potrebno je predvidjeti hoće li se podaci morati sezonski prilagođavati. Primjena ovih postupaka ovisi o tome postoji li sezonski učinak i učinak broja radnih dana na statističke rezultate. O čemu je zapravo riječ može se analitički utvrditi nakon određenih broja ponavljanja aktivnosti. Isto tako, navedeni utjecaj može se predvidjeti kada se detaljno upozna sadržaj statističkog istraživanja.

Upute za osiguranje kvalitete

Pri planiranju statističkih aktivnosti potrebno je slijediti suvremene statističke metode i metodološka rješenja kao i dobre prakse drugih nacionalnih statističkih ureda. Kod dizajna alata za statističku obradu potrebno je utvrditi postoje li softverski alati koji bi se mogli koristiti ili je potrebno razviti rješenja po mjeri. Ako je potrebno primijeniti nova softverska rješenja, nužno je odabrati najprikladniji softverski alat. Treba uzeti u obzir funkcionalne mogućnosti pojedinih softverskih alata i dobru praksu nacionalnih statističkih ureda.

U slučajevima neodgovora jedinice na uzorku potrebno je odlučiti hoće li se za sve varijable koristiti utežavanje podataka ili umetanje. Za donošenje konačne odluke treba uzeti u obzir sljedeće parametre:

- broj varijabli koje treba umetnuti (metode umetanja obično se koriste kao rješenje problema neodgovora jedinice na uzorku samo kad je u anketi malo varijabli)
- je li riječ o periodičnom istraživanju u kojem se povijesni podaci mogu koristiti za umetanje podataka (podaci pojedine ispitane jedinice na uzorku)
- postoje li vanjski (npr. registri) administrativni izvori podataka na mikrorazini koji bi se mogli koristiti za umetanje podataka.

Metodama selektivnoga i automatskog uređivanja podataka trebalo bi se koristiti što je više moguće u postupcima uređivanja.

Pri određivanju detalja o razini objavljivanja podataka potrebno je izbjegavati predetaljnu razinu objavljivanja podataka koja bi mogla rezultirati velikim brojem praznih, zatamnjenih ili zaštićenih ćelija. Treba voditi računa o statističkoj povjerljivosti, koja podrazumijeva zaštitu povjerljivih podataka, a mogu se odrediti izravnom identifikacijom (naziv/ime i prezime, adresa, javno dostupan identifikacijski broj) i neizravnom identifikacijom (svaki drugi način osim izravne identifikacije). Ta preporuka ne vrijedi u slučajevima kada je drugačije propisano.

3. Izrada potrebnih instrumenata za provedbu

Pri određivanju ciljne populacije za pojedino statističko istraživanje potrebno je odlučiti hoće li se provesti na cijeloj populaciji ili samo na dijelu populacije. Budući da je provođenje ankete na cijeloj populaciji skupo, dugotrajno i zahtjevno i za one koji provode anketu i za one koji odgovaraju na upitnike, većina anketa provodi se na temelju izbora uzorka. To znači da se na temelju dobivenih podataka ispituju samo dijelovi populacije sa specifičnim obilježjima važnima za provedbu statističkog istraživanja.

Čak i prije nego što se donese odluka u vezi s planiranjem postupka uzorkovanja, kojim će se odrediti način odabira uzorka, potrebno je teoretski definirati populaciju i pripremiti što više osnovnih obilježja o njoj. Stvarni zapis jedinica populacije, zajedno s nekim obilježjima koja su bitna i za odabir uzorka i za izračunavanje procjena parametara populacije, naziva se okvir za izbor uzorka. Prije pripreme okvira za izbor uzorka potrebno je pripremiti sve izvore podataka (registre, podatke iz drugih istraživanja) s pomoću kojih će se sastaviti okvir za izbor uzorka. U praksi je okvir za izbor uzorka više ili manje točna aproksimacija ciljne populacije (i nekih njegovih obilježja). Kvaliteta uzorka odražava se na kvalitetu konačnih rezultata istraživanja.

U pripremu izvora podataka koji su potrebni za sastavljanje okvira za izbor uzorka treba uložiti sve napore kako bi se postigla najviša moguća kvaliteta okvira za izbor uzorka. Nakon što se utvrdi okvir za izbor uzorka, utvrđuje se plan postupka uzorkovanja kojim se obavlja odabir jedinica u uzorak za provedbu statističkog istraživanja. Uzorak jedinica može biti vjerojatnosni, nevjerojatnosni ili njihova kombinacija. Jedanput kada se utvrdi uzorak jedinica promatranja, sastavlja se adresar jedinica koje su izabrane na uzorku. Inače, adresar sadržava popis adresa (i ostale detalje za kontakt) izvještajnih jedinica.

3.1. Izrada projektnih zahtjeva

Opis

Izrada projektnih zahtjeva uključuje zahtjeve za prikupljanje i obradu podataka s rokovima i nositeljima u skladu s uputama za izradu projektnog zahtjeva SIT-a, bez obzira na to je li riječ o istraživanju na uzorku ili istraživanju na punom obuhvatu (popis).

Bitno je razumjeti razliku između ciljne populacije i okvira za izbor uzorka. Ciljna populacija statističkog istraživanja jest skup jedinica čija se specifična obilježja promatraju i na koja se odnose svi rezultati određenog istraživanja. Okvir za izbor uzorka stvarni je popis jedinica ciljne populacije koji je dostupan u vrijeme pripreme istraživanja, a koristi se za odabir jedinica ciljne populacije u uzorak.

Prije određivanja okvira za izbor uzorka mora se precizno definirati ciljna populacija (osobe, kućanstva, stanovi, obiteljska poljoprivredna gospodarstva, poduzeća itd.). Ovisno o obilježjima, određuje se ciljna populacija i uvjeti koje jedinica mora ispuniti kako bi bila dio populacije. Isto tako, potrebno je odrediti zemljopisni položaj jedinica te referentno razdoblje (vrijeme) s obilježjima populacije važnima za provedbu istraživanja.

Primjer: Naša ciljna populacija jesu osobe koji žive u privatnim kućanstvima, koje su određenog datuma bile stanovnici Republike Hrvatske i na taj su datum imale najmanje 18 godina. Pritom je potrebno uzeti u obzir definiciju "stanovnik Republike Hrvatske", potrebno je definirati uvjete kada je osoba član privatnoga kućanstva (što je privatno kućanstvo; kako postupiti ako osoba neko vrijeme živi u jednom e, dok drugo vrijeme živi u nekome drugom kućanstvu; kojem kućanstvu pripada takva osoba). Ti uvjeti definiraju ciljnu populaciju i, prema tome, uvjete koje jedinica mora ispunjavati kako bi bila ili ne bi bila prikladna za istraživanje.

U statističkom istraživanju razlikujemo:

- jedinice za uzorkovanje (jedinice iz okvira za izbor uzorka koji su izabrane uzorkovanjem)
- jedinicu promatranja (jedinica o kojoj prikupljamo podatke i stoga je dio ciljne populacije)
- izvještajnu jedinicu (jedinica koja dostavlja podatke koji se prikupljaju).

U nekim su studijama sva tri obilježja kombinirana u jednoj cjelini, ali to često nije slučaj.

Primjer: Jedinica za uzorkovanje jest odabrana osoba iz baze podataka popisa stanovništva, kućanstava i stanova koja živi u stanu na određenoj adresi i dostavila je informacije propisane istraživanjem. Ta je osoba izvještajna jedinica.

Sva obilježja potrebno je uzeti u obzir pri sastavljanju okvira za izbor uzorka. To znači da se okvir za izbor uzorka često ne sastoji od jedinica koje su predmet promatranja, nego od jedinica koje statističare vode do jedinica promatranja.

Kada je poznata ciljna populacija i ako se može pristupiti jedinicama te populacije, pristupa se postupku kojim se definira okvir za izbor uzorka. Ovisno o ciljnoj populaciji, koriste se različiti izvori podataka, bilo u vremenskim intervalima, bilo u razdobljima koja su bliža referentnom razdoblju istraživanja ili vremenu prikupljanja podataka.

Ako korišteni izvor podataka pokriva referentno razdoblje istraživanja, tada se populacija preciznije bilježi u referentnom razdoblju. Može biti riječ o jedinicama koje nisu bile relevantne za vrijeme prikupljanja podataka. Međutim, još su uvijek bile prisutne u referentnom razdoblju istraživanja ili su se neka obilježja jedinica promijenila (npr. tvrtka je promijenila djelatnost). Ako je izvor podataka novijeg datuma u odnosu na vrijeme prikupljanja podataka, broj jedinica koje ne zadovoljavaju postavljene kriterije sveden je na minimum. Također se može dogoditi i da jedinice imaju neka obilježja koje vrijede sada, ali ne i za referentno razdoblje te se mogu pogreškom grupirati u skupine ili klastere.

Obično se bira jedan osnovni izvor podataka kao okvir za izbor uzorka. Nakon toga poboljšava se kvaliteta okvira za izbor uzorka dodatnim izvorima. Za osnovni izvor podataka bira se izvor koji najbolje omogućuje prikupljanje podataka o ciljnoj populaciji. Dodatni izvori podataka nadopunjuju osnovni izvor podataka.

U pravilu su osnovni izvor podataka poslovna statistička istraživanja odnosno SPR Državnog zavoda za statistiku, u kojem je sadržan popis svih tvrtki. Dodatni izvori podataka mogu biti godišnji financijski izvještaj Fine i podaci o PDV-u. Navedeni izvori podataka omogućuju određivanje veličine raspona poduzeća.

U statističkim istraživanjima u kojima se prikupljaju podaci o osobama i kućanstvima, za anketne upitnike koji se koriste za terensko istraživanje, glavni izvor podataka jest baza popisa stanovništva, kućanstava i stanova. Popis stanovništva, kućanstava i stanova provodi se svakih deset godina. Baza popisa stanovništva, kućanstava i stanova na godišnjoj se razini ažurira podacima iz matica rođenih i umrlih te migracijskim saldom; za anketne upitnike koji se provode telefonski glavni izvor podataka može biti telefonski imenik.

Katkad jedan izvor podataka nije dovoljan, stoga je potrebno izraditi okvir za izbor uzorka iz nekoliko različitih izvora podataka (administrativni izvori podataka, baza popisa stanovništva, kućanstava i stanova te druga statistička istraživanja). To je slučaj kada različiti izvori podataka pokrivaju različita razdoblja i različite dijelove populacije. U ovom slučaju bitno je da se koristi jedinstvena identifikacija za sve populacijske jedinice kod svih vrsta izvora podataka. Stoga su relevantni samo izvori podataka u kojima je moguće utvrditi zajedničku identifikaciju. U ovom slučaju jedan izvor podataka treba biti definiran kao primarni izvor podataka.

Ciljna populacija i okvir za izbor uzorka rijetko se podudaraju u svim jedinicama. Postoje jedinice koje su dio ciljne populacije, ali ih nema na popisu jedinica okvira za izbor uzorka. Postoje i jedinice u okviru za izbor uzorka koje zapravo nisu dio ciljne populacije. Oba slučaja mogu prouzročiti pogreške u obuhvatu koje uvelike određuju kvalitetu konačnih rezultata statističkog istraživanja.

Upute za osiguranje kvalitete

Referentna razdoblja na koja se podaci odnose u izvorima podataka koji će se koristiti u izradi okvira za izbor uzorka trebaju biti što bliža referentnom razdoblju statističkog istraživanja. Ako su referentna razdoblja različita u različitim izvorima podataka, razlike bi se trebale evidentirati i dokumentirati.

Potrebno je provjeriti jesu li varijable koje se pojavljuju u nekoliko različitih izvora podataka u skladu s metodološkim definicijama. Ako se utvrdi da se prema tim definicijama razlikuju ili se ne podudaraju, te razlike moraju biti evidentirane i dokumentirane.

Rubrika mora sadržavati sljedeće podatke: jedinstveni identifikator jedinice promatranja, kontaktne podatke i klasifikacijsku oznaku varijable. Korisno bi bilo kada bi okvir za izbor uzorka sadržavao i datum stanja okvira za izbor uzorka i varijable veze.

Potrebno je osigurati konsolidaciju svih izvora podataka sa stajališta formalnih kontrola prije agregiranja, kao što su valjanost popisa šifri, odgovarajući raspon vrijednosti varijabli i problem dvostrukih jedinica.

Izvori podataka koji se koriste za izradu okvira za izbor uzorka trebali bi biti pravilno dokumentirani s kratkim opisom tih izvora kao i zapisi koji se neće moći koristiti za izradu okvira.

Svi bi se dostupni izvori podataka trebali koristiti u pripremi okvira za izbor uzorka jer svaki od njih doprinosi maksimalnom usklađivanju okvira za izbor uzorka s ciljnom populacijom. Trebalo bi dokumentirati sve razlike između okvira za izbor uzorka i ciljne populacije.

3.2. Izrada instrumenata za prikupljanje podataka

Opis

Podrazumijeva izradu instrumenata koji će se koristiti u fazi prikupljanja podataka. Instrument za prikupljanje podataka izrađuje se na temelju specifikacija dizajna izrađenih u fazi pripreme.

Prije provođenja statističkog istraživanja uz upitnik je potrebno izraditi popratnu dokumentaciju koja se koristi za komunikaciju s izvještajnim jedinicama. U najavnom pismu izvještajna jedinica upoznaje se s pojedinim statističkim istraživanjem, glavnim ciljevima istraživanja, metodološkom osnovom provedbe istraživanja, na koga je istraživanje usmjereno, obvezom statističkog izvještavanja, načinom prikupljanja i rokovima, povjerljivošću podataka te osobama za kontakt u slučaju pitanja ili nejasnoća. Namjera je potaknuti izvještajne jedinice na sudjelovanje. Stoga se široj javnosti predstavljaju najzanimljiviji rezultati istraživanja. Rezultatima se može naglasiti smisao provođenja istraživanja i privući izvještajne jedinice na sudjelovanje. Obično se dostavljaju dva podsjetnika, samo iznimno i tri. Ako je riječ o manjem broju izvještajnih jedinica, može ih se i telefonski "požuriti" i nije nužno slati drugi podsjetnik. U slučajevima kada je prekinuta suradnja zbog prestanka poslovanja ili je Državni zavod za statistiku preusmjerio svoje aktivnosti na druge izvore podataka, više nije potrebno sudjelovanje svake izvještajne jedinice kao i u anketama s povremenim rotacijama promatranih jedinica kada pojedina izvještajna jedinica više nije uključena u prikupljanje podataka.

Sva se dokumentacija priprema za tekuću godinu na temelju standardiziranih predložaka. Dokumentaciju je potrebno lektorirati prije distribucije na teren.

Priprema upitnika važan je dio statističkog postupka jer kada je dobro i kvalitetno pripremljen, onda će i preuzimanje pouzdanih podataka od izvještajnih jedinica biti na zadovoljavajućoj razini. Zahvaljujući dugogodišnjoj suradnji s izvještajnim jedinicama, Državni zavod za statistiku upoznat je s tim u kojem obliku izvještajne jedinice imaju potrebne statističke podatke.

Prije izrade radne verzije upitnika potrebno je konzultirati se s korisnicima i posjednicima podataka, prikupiti i pregledati upitnike svih prethodno provedenih statističkih istraživanja i pripremiti popis svih varijabli za predmetno istraživanje. Za svaku varijablu unutar statističke aktivnosti koja se prikuplja godišnjim provedbenim planom statističkih aktivnosti Republike Hrvatske potrebno je navesti izvor podataka odnosno je li izvor podataka statističkog istraživanja upitnik, administrativni izvor ili neki drugi sekundarni izvor podataka. U slučaju da je riječ o podacima iz registra, potrebno je navesti referentno razdoblje i informacije o dostupnosti izvora podataka. Odluka o načinu prikupljanja podataka mora biti donesena prije nego što se počne s pripremom radne verzije upitnika i treba uzeti u obzir najnovije standarde koji vrijede za statistički sustav Republike Hrvatske.

Formulacija pitanja ovisi o raznim čimbenicima, kao što su: način prikupljanja podataka (mrežni upitnici, Excel-upitnici, telefonska anketa, terensko prikupljanje podataka pomoću prijenosnih računala, tiskani upitnici za samopopunjavanje), obilježja davatelja podataka (ispitanika), izbjegavajući prekomjerno opterećenje jedinica te složenost podataka koji su prikupljeni s visokom razinom povjerljivosti podataka i osjetljivosti informacija. Isto tako, potrebno je razmotriti usporedivost s rezultatima drugih istraživanja te voditi računa o smislenosti. Postavljena pitanja trebaju imati isto značenje za sve davatelje podataka (ispitanike) koji su uključeni u istraživanje kao i ostali kriteriji koji trebaju biti zadovoljeni kako slijedi:

- dostupnost potrebnih podataka
- spremnost ispitanika da dostavi kvalitetne podatke
- vjerojatnost neodgovora
- formulacija u semantičkom smislu svakog pitanja
- izgled upitnika
- izvor pogrešaka kod iskazivanja vrijednosti u mjernim jedinicama i pogrešaka u odgovorima
- unos podataka (unos ankete, unos brzine popunjavanja).

Priprema radne verzije upitnika

Nakon pripreme pitanja i varijabli priprema se verzija upitnika s logičkim kontrolama i preskakanjima. Pri pripremi elektroničkog upitnika provodi se savjetovanje s programerima oko tehničkih mogućnosti i standarda. Nakon toga priprema se radna verzija upitnika. Pitanja u upitniku podijeljena su ovisno o temama. Ako je riječ o upitniku koji trebamo preuzeti u skladu s harmoniziranim upitnikom (radi usporedivosti između zemalja EU-a), treba posvetiti pozornost prijevodu s engleskoga na hrvatski jezik jer je često riječ o izrazima koji nisu toliko poznati ili nisu uvriježeni u našem jeziku.

Provjera i revizija upitnika

Prije testiranja upitnika na ciljnoj populaciji bitno je provjeriti sve stavke upitnika. Upitnik se šalje na lekturu i tehničko uređenje. Važno je služiti se izrazima za koje smo sigurni da su stručno dobro prevedeni. Internom bi se kontrolom sadržaja upitnika trebala provjeriti gramatička ispravnost postavljenih pitanja te poboljšanja kod nespretno postavljenih i nerazumljivih pitanja. Interna kontrola podrazumijeva da tekst upitnika testiraju osobe koje nisu izravno uključene u određeno statističko istraživanje. Osobe koje testiraju upitnik mogu biti stručnjaci iz drugih statističkih istraživanja, anketari ili predstavnici ciljne populacije. Svi navedeni elementi mogu pridonijeti poboljšanju konačne verzije upitnika.

Tehničko testiranje upitnika

Ako se podaci prikupljaju elektroničkim upitnikom (angažiranjem anketara ili primjenom mrežnog upitnika za samopopunjavanje), tada odgovorni statističar(i) tehnički provjerava(ju) sadržaj upitnika. Takva provjera obuhvaća testiranje skokova i funkcioniranje logičkih kontrola. Kada u upitniku sve tehnički dobro funkcionira, onda se provodi i provjera sadržaja.

Testiranje i revizija upitnika

U ovaj su postupak uključene sljedeće aktivnosti: neformalno testiranje upitnika, kognitivne metode, fokus skupine, izvještaji anketara, šifriranje ponašanja anketara ili ispitanika, razni eksperimenti (npr. testiranje podijeljenog uzorka) i probna testiranja upitnika. Poželjno je izvršiti test kognitivnog upitnika prije provođenja probnih istraživanja na terenu.

Popunjavanje upitnika

S obzirom na to da je dizajn upitnika iterativni postupak, bilo bi dobro s vremena na vrijeme testirati upitnik.

Upute za osiguranje kvalitete

Pri pripremi pitanja bitno je da izvještajna jedinica razumije postavljena pitanja, da ima informacije za pripremu odgovora na postavljena pitanja i da je voljna odgovoriti na pitanja. Način postavljanja pitanja potrebno je prilagoditi načinu prikupljanja podataka.

Uvijek je potrebno prvo provjeriti koji su podaci dostupni u administrativnim izvorima podataka, za koje referentno razdoblje i kada su dostupni. Ovisno o raspoloživosti podataka kod administrativnih izvora pristupa se izradi upitnika koji bi sadržajno trebao biti što kraći. Nadalje je potrebno provjeriti jesu li u istraživanje uključene sve osnovne varijable koje su propisane propisom ili metodologijom.

Prije provedbe probnih istraživanja ili redovitog istraživanja upitnik se kontrolira na detaljnoj razini. Ako je riječ o elektroničkom upitniku, najprije se interno provjerava cijeli upitnik, zatim slijedi tehničko testiranje (funkcioniranje logičkih kontrola i preskakanja) i na kraju se provjerava sadržaj upitnika na ciljnoj populaciji odnosno razumijevanje postavljenih pitanja.

Pri pripremi pitanja uvijek treba voditi računa o obilježjima ispitanika, tj. o davateljima podataka te njima prilagoditi razinu složenosti pojmova i strukturu rečenica koje se koriste u pitanjima. Postavljena pitanja široj javnosti odnosno ciljnoj populaciji trebaju svim skupinama biti jasna i razumljiva.

Kod samopopunjavanja upitnika postavljena pitanja trebaju biti kratka i jasna. Upitnik ne smije biti opsežan. Treba sadržavati jasne upute i primjere. Korisno je uz postavljena pitanja dati praktične primjere.

Kod takvih vrsta upitnika ne postoji posrednik, primjerice anketar koji bi ispitaniku mogao objasniti sadržaj pitanja. Odluka o načinu prikupljanja podataka i načinu unosa podataka ovisi o raspoloživim financijskim sredstvima, vrijednosti stope neodgovora, vremenskog okvira u kojem su se podaci trebali prikupiti, ciljnoj populaciji i informacijama koje su dostupne u okviru za izbor uzorka. Iznimno, tiskani upitnici mogu se koristiti i za prikupljanje podataka na terenu ako se procijeni da će prikupljanje podataka s pomoću prijenosnih računala vremenski trajati dulje od popunjavanja na papiru.

Metodolozi se trebaju potruditi da broj pitanja u upitniku bude minimalan. Za svako postavljeno pitanje treba postojati razlog zašto je pitanje uključeno u upitnik. Stoga je nužna provjera u administrativnim izvorima podataka i evidencijama.

Kako bi se ispitaniku olakšalo popunjavanje upitnika, postoji mogućnost da se upitnik automatski popuni iz nekoga drugog izvora podataka i da ispitanik u tom slučaju može odgovor na pitanje ispraviti ako nije točan.

Iz praktičnih razloga katkad je korisno dodati u upitnik pitanje koje je potrebno u nekome drugome statističkom području. Ovakav način rada je za Državni zavod za statistiku zahtjevniji, dok za izvještajnu jedinicu predstavlja manje opterećenje.

Posebnu pozornost treba posvetiti postavljenim pitanjima s pomoću kojih se želi preuzeti složenije podatke. U ovakvim slučajevima u uputama treba dodati pitanja kojima se obuhvaćaju složene teme. Upute pomažu anketarima i ispitanicima kako bi lakše pripremili odgovore na postavljena pitanja.

Povjerljivost i osjetljivost podataka utječu na dizajn upitnika. U upitniku i na obavijesti u najavnom pismu treba biti navedeno da su prikupljeni statistički podaci o fizičkim ili pravnim osobama, ako se oni izravno ili neizravno mogu dovesti u vezu s fizičkom ili pravnom osobom, statistički povjerljivi i službena su tajna. Povjerljivi statistički podaci iskazuju se u obliku u kojem statističku jedinicu nije moguće identificirati ni izravno ni neizravno i ne mogu biti podloga za utvrđivanje bilo kakvih prava i obveza izvještajnih jedinica, kao što su upravne, pravne ili porezne svrhe, ili za provjeru izvještajnih jedinica.

Ako su predmet anketiranja osjetljive teme odnosno teme o kojima je neugodno razgovarati, u ovom slučaju mogu se koristiti tehnike pristupačnijeg načina postavljanja pitanja.

Isto tako, bitno je da se stranim ispitanicima pitanja postavljaju na govornom jeziku ciljne populacije. Ako je upitnik preveden sa stranog jezika, nužno je pri prevođenju uzeti u obzir kulturološke razlike i navike. Primjerice, u popisima stanovništva, kućanstava i stanova upitnici se moraju prevesti na jezike nacionalnih manjina.

Kako bi se postigla usporedivost rezultata istraživanja, bitno je da su pitanja osmišljena na isti način.

3.3. Izrada programske podrške

Opis

Opisuje radnje koje treba poduzeti za izgradnju novih i unapređenje postojećih softverskih komponenata potrebnih za poslovni proces, dizajniranih u fazi pripreme. Komponente mogu uključivati kontrolne tablice i izvještaje, baze podataka, tablice rezultata, alate za transformaciju podataka te alate za upravljanje podacima, geoprostornim podacima i metapodacima.

Sve faze poslovnih procesa trebale bi biti jasno i detaljno dokumentirane. Na primjer, za provođenje obrade statističkog istraživanja s pomoću Generatorsa potrebno je pripremiti "Projektni zahtjev", koji je ujedno i radni nalog za sve poslove vezane za obradu istraživanja koje statističke ustrojstvene jedinice dostavljaju Sektoru informatičkih tehnologija (SIT-u).

Razlikuju se sljedeće vrste poslova:

1. zahtjev za novom obradom
2. dopuna zahtjeva koji je već predan SIT-u
3. izmjena u zahtjevu koji je već predan SIT-u.

U zahtjevu se nalazi niz obrazaca standardnog oblika i sadržaja i služi za dokumentaciju obrade istraživanja. Predaje se SIT-u za sve nove obrade, kao i za dopune i promjene u već postojećim obradama. Komplet priloga mora biti takav da se iz njega mogu uočiti i provesti sve potrebne intervencije. Na zahtjevu je potrebno označiti o kojem je tipu posla riječ.

Dokumentacija treba sadržavati i materijalno-tehničke opise poslova. Dokumentacija treba u najvećoj mogućoj mjeri sadržavati ocjene po različitim fazama poslovnih procesa.

Upute za osiguranje kvalitete

Središnja baza metapodataka CROMETA jest jezgra sustava za obradu statističkih istraživanja tzv. Integriranoga statističkoga informacijskog sustava. Nazvana je CROMETA prema modelu koji je nastao u suradnji Državnog zavoda za statistiku i Švedske statistike (SCB-a) te uz sponzorstvo Švedske agencije za međunarodni razvoj (SIDA-e).

Prva verzija modela napravljena je 2005., a sve do danas kontinuirano se razvija prema potrebama. Model baze metapodataka CROMETA oblikovan je prema tzv. referentnome modelu koji je nastao u Eurostatovu projektu METANET (2000. – 2003.). Referentni model objedinjuje različite vrste metapodataka koji opisuju statističke podatke i procese.

Unutar modela CROMETA postoje različite vrste metapodataka prema svrsi. Neki metapodaci služe za deklarativni opis sadržaja statističkih istraživanja i podataka, dok su drugi procesno orijentirani te opisuju načine prikupljanja i obrade podataka. Uz njih postoje globalni koncepti i metodologije koje se koriste neovisno o statističkom istraživanju, kao i općeniti metapodaci koji opisuju organizacijsku strukturu, autorizaciju i prava pristupa.

Aplikacija Metadata Manager služi za upravljanje metapodacima u centralnoj bazi metapodataka CROMETA. Aplikacija je dostupna samo autoriziranim korisnicima koji imaju dopuštenja za rad s metapodacima, ovisno o skupini kojoj pripadaju.

Warehouse Browser jest aplikacija namijenjena za pregled podataka spremljenih u statističkom skladištu podataka i opisanih u središnjoj bazi metapodataka te se s pomoću nje mogu generirati izlazne tablice i spremati rezultati u obliku agregiranih podataka (kocke) ili pripremljenih tablica.

3.4. Testiranje alata za prikupljanje i obradu podataka

Opis

Uključuje tehničko testiranje i odobravanje novih programa i postupaka. Uključuje i testiranje interakcije između komponenti te se brine da proizvodni sustav funkcionira kao usklađen skup komponenti. Uključuje prikupljanje podataka za eksperimentalna istraživanja radi testiranja instrumenata za prikupljanje podataka.

Prije pokretanja istraživanja osobito ako je riječ o novom istraživanju ili o situaciji kada nije istražen ni sadržaj ni ciljna populacija, treba provesti probno istraživanje s pomoću kojeg je moguće utvrditi i pripremiti metodologiju istraživanja. Ciljevi probnog istraživanja moraju biti jasno definirani jer se njemu treba prilagoditi plan i svrha istraživanja na uzorku. U probnom istraživanju može se testirati cijeli postupak ili samo pojedini dijelovi poslovnog procesa. Provedbom probnog istraživanja najčešće se provode sljedeće aktivnosti, kao što su: testiranje pitanja, potrebna veličina uzorka, plan postupka uzorkovanja i način prikupljanja podataka. Veličina uzorka obično je mnogo manja nego kod organiziranja i provedbe redovitoga statističkog istraživanja. Nakon provedbe probnog istraživanja rezultati se detaljno analiziraju i na temelju dobivenih rezultata utvrđuju se daljnji postupci u provedbi istraživanja.

Probno istraživanje može se provesti i slučajevima kada se podaci potpuno preuzimaju iz administrativnih izvora. Bez obzira na to što je riječ o preuzimanju podataka iz administrativnog izvora, potrebno je definirati ciljeve probnog istraživanja.

Upute za osiguranje kvalitete

Državni zavod za statistiku u skladu sa Zakonom o službenoj statistici (NN, br. 25/20.) te programskim dokumentima prikuplja podatke s pomoću statističkih obrazaca koji su osnovni alat za provedbu statističkih istraživanja na terenu.

Statistički obrasci koji se koriste u tom postupku različitih su informatičko-metodoloških oblika i sadržaja. Državni zavod za statistiku u skladu s raspoloživim resursima postupno razvija novi elektronički sustav prikupljanja podataka. Elektronički obrasci omogućuju izvještajnim jedinicama alternativni način dostavljanja podataka koji se od njih traže, za razliku od sustava koji se temelji na ispunjavanju klasičnih obrazaca u tiskanom obliku koji se uobičajeno dostavljaju poštom. Elektroničkim sustavom prikupljanja podataka obuhvaćena su sljedeća statistička istraživanja:

- Statistička istraživanja industrije
- Statistička istraživanja o primjeni informacijskih i komunikacijskih tehnologija (IKT-POD, IKT-DOM)
- Statistička istraživanja za energiju
- Statistička istraživanja za građevinarstvo
- Mjesečni izvještaj trgovine na malo (TRG-1)
- statistička istraživanja o cijenama usluga (SPPI)
- Mjesečni izvještaj distributivne trgovine i ostalih usluga (USL-M)
- Statističko istraživanje o cestovnom prijevozu robe (PA/T-11)
- Statistička istraživanja transporta (PA/M-11, PR/T-11P, PP/T-11)
- Statistička istraživanja turizma (TU-14, TU-11i)
- Anketa o putovanjima
- Izvještaj u sektoru poljoprivrede, ribarstva i zaštite okoliša (PO-22/STR, PO-32)
- Prodaja i otkup poljoprivrednih proizvoda (PO-31a/Q, PO-31b/Q)
- Statistička istraživanja kaznenog pravosuđa
- Investicije u zaštitu okoliša i izdaci za dobra i usluge u okolišu (IDU-OK)
- Godišnji izvještaj o investicijama u dugotrajnu imovinu (INV-P)
- Godišnji izvještaj o zaposlenima i plaći (RAD-1G)
- Strukovno osposobljavanje u poduzećima u 2020. (SOP)
- Inovacijske aktivnosti poduzeća u razdoblju 2018. – 2020. (INOV)
- Godišnji izvještaj za osnovne škole (Š-O/KP)
- Godišnji izvještaj za srednje škole (Š-S/KP)
- Dječji vrtići i druge pravne osobe (DV-PO)
- Statistički izvještaj o razvedenom braku (RB-1).

Kod klasičnog prikupljanja podataka izvještajnim jedinicama koje nisu odgovorile treba dostaviti podsjetnik ili zatražiti ponovno sudjelovanje u statističkom procesu prikupljanja podataka. Uobičajena je procedura da se dostavljaju dva podsjetnika.

U statističkim istraživanjima za osobe i kućanstva, u kojima sudjeluje anketar, odabranu se izvještajnu jedinicu o dolasku anketara i anketi obavještava najavnom pismom. Uz najavno pismo obično se nalazi letak s odgovorima u vezi s pitanjima koja se najčešće postavljaju ispitanicima ili promatranim jedinicama.

Ako ciljna populacija uključuje i maloljetnike, u najavnom pismu mole se roditelji ili skrbnici da svojoj djeci omoguće sudjelovanje u provedbi istraživanja.

U slučaju da se za statističko istraživanje koristi kombinirana metoda anketiranja, npr. kombinacija telefonskog anketiranja i terenskih anketa, praksa je pokušati anketirati izvještajne jedinice koje nisu bili obuhvaćene telefonskim anketama jer se time želi smanjiti pristranost zbog stope neodgovora. U ovakvom slučaju tim se jedinicama mora poslati dodatno pismo. Sadržaj pisma prilagođuje se ovisno o konačnom razlogu neodgovora kod metode prvog anketiranja.

U poštanskim se anketama dostavlja podsjetnik ili ponovna prijava za sudjelovanje. Pritom se mora paziti da se dostavi omotnica za koju je poštarina već unaprijed plaćena i na kojoj je napisana adresa Državnog zavoda za statistiku, kao i dodatni upitnik, zajedno s ponovljenim zahtjevom za sudjelovanje (podsjetnik).

Najavna pisma trebaju sadržavati osnovne informacije o istraživanju te o načinu rada anketara.

Ciljevi probnog istraživanja trebaju biti jasno definirani. U njemu se trebaju ispitati različiti metodološki aspekti: formulacija pitanja, pojedinačne stope neodgovora, učinak obavijesti itd. Rezultati probnog istraživanja trebaju se koristiti u pripremi glavnog istraživanja.

U probnom istraživanju dizajn uzorka treba prilagoditi tako da se mogu testirati razlike u odgovorima pojedine podskupine; može se testirati učinak obavijesnih pisama, različitih oblika pitanja itd.

Probno istraživanje također je namijenjeno za provjeru upitnika. Međutim, prije provedbe probnog istraživanja moraju se provesti tehnička ispitivanja upitnika i kognitivnih obilježja. Pri testiranju upitnika potrebno je ostvariti osobni kontakt s izvještajnom jedinicom. Katkad je u probnom istraživanju potrebno testirati postupke statističke obrade i tabličnog izračuna.

3.5. Konfiguriranje tijeka proizvodnih procesa

Opis

Konfiguriranje tijeka proizvodnih procesa odnosi se na prikupljanje podataka pa sve do arhiviranja konačnih statističkih rezultata. Uključuje aktivnosti kojima se proces stavlja u proizvodnju, spreman da ga koriste poslovna područja.

Sve statističke aktivnosti koja se provode u Republici Hrvatskoj, a nositelji su im Državni zavod za statistiku i druge upravne organizacije i institucije, opisane su u GPP-u. U bazu CROMETA su na temelju GPP-a uneseni metapodaci o statističkim istraživanjima i dopunjeni kontaktnim informacijama te još nekim metapodacima. GPP sadržava pregled statističkih istraživanja i aktivnosti koje su grupirane u pet osnovnih poglavlja: Demografske i društvene statistike; Ekonomske statistike; Sektorske statistike; Statistika okoliša i statistika za više područja te Metodologija prikupljanja, obrade, diseminacije i analize podataka.

Upute za osiguranje kvalitete

Pravni okvir jest polazna točka za organiziranje i provedbu statističkih aktivnosti u Republici Hrvatskoj, odnosno Zakon o službenoj statistici (NN, br. 25/20.).

Struktura svakoga statističkog istraživanja ili aktivnosti propisana je čl. 36. Zakona o službenoj statistici. Godišnjim provedbenim planom statističkih aktivnosti Republike Hrvatske uređuje se provedba sljedećih aktivnosti službene statistike:

- statističkih istraživanja na temelju neposrednog prikupljanja podataka
- aktivnosti službene statistike čiji se podaci dobivaju iz administrativnih izvora ili metodom promatranja i praćenja
- razvojnih aktivnosti, popisa i drugih opsežnijih statističkih istraživanja

Godišnjim provedbenim planom statističkih aktivnosti Republike Hrvatske utvrđeni su nazivi aktivnosti službene statistike, odgovorni nositelji službene statistike, izvještajne jedinice, periodičnost istraživanja, rokovi za prijenos ili prikupljanje podataka, okvirni rokovi objavljivanja rezultata, razina objavljivanja rezultata, relevantni standardi te prema potrebi i druga obilježja aktivnosti službene statistike.

Kontrolni mehanizam za praćenje statističkih aktivnosti jest statistički priručnik Eurostata pod nazivom Statistical Requirements Compendium, u kojem se točno navode aktualni korisnički priručnici, standardi, tehnička dokumentacija i treninzi za osposobljavanje, npr. u metadatkovnim standardima ESS-a.

4. Prikupljanje podataka

Prikupljanje podataka jest postupak koji se provodi na početku statističkog istraživanja i znatno utječe na konačni rezultat. Poznato je nekoliko načina prikupljanja podataka za statističke svrhe. Najpoznatiji način prikupljanja podataka jesu statistička istraživanja s pomoću kojih se prikupljaju podaci isključivo za statističke svrhe. U posljednje se vrijeme kao drugi način prikupljanja koriste administrativni izvori u kombinaciji s terenskim prikupljanjem podataka za statističke svrhe.

Kod terenskog rada prva faza obuhvaća kontaktiranje s izvještajnim jedinicama, dok druga faza obuhvaća prikupljanje podataka s pomoću upitnika u obliku zatvorenih ili otvorenih pitanja. Pitanja u upitnicima za statistička istraživanja koja provodi Državni zavod za statistiku uglavnom su kvantitativna, a rjeđe kvalitativna. Upitnici i metode prikupljanja podataka moraju omogućiti najvišu moguću kvalitetu podataka, kontrolu pogreške mjerenja, najmanje moguće opterećenje za izvještajne jedinice i što niže troškove.

Na odluku o načinu prikupljanja podataka utječe nekoliko čimbenika: obilježja ciljne populacije, ciljevi prikupljanja podataka, raspoloživi resursi i vremenska ograničenja. Osim toga, potrebno je poznavati obilježja statističkih istraživanja kako bi se mogla izabrati odgovarajuća metoda i odgovarajuća tehnologija prikupljanja podataka.

Razlikuju se dvije vrste statističkih istraživanja: longitudinalna i unakrižna. Longitudinalna istraživanja koriste se za praćenje dinamike promatrane pojave. Kod takvih statističkih istraživanja prikupljaju se podaci za iste jedinice više puta u redovitim intervalima, dok u unakrižnim statističkim istraživanjima podatke prikupljamo samo jedanput ili povremeno. Panel-istraživanje jest istraživanje u koje su uključene iste jedinice u nekoliko uzastopnih faza istraživanja.

Najčešći načini prikupljanja podataka jesu samopopunjavanje (poštanska metoda i elektroničko izvještavanje) i intervjuiranje (telefonsko i osobno anketiranje). Prvi se pristup obično koristi kod poslovnih anketa, dok se drugi koristi u anketama u kojima se promatraju pojedinci i kućanstva. Razlikuju se dvije vrste tehnologija za prikupljanje podataka: snimanje odgovora na papir i snimanje odgovora na računalu. Kada se koriste tiskani upitnici, podaci se moraju digitalizirati; digitalizacija podataka može se izvršiti ručnim unosom ili skeniranjem. Kod oba načina digitalizacije problem su pogreške pri unosu jer i čovjek i skener mogu pogriješiti. Stoga je nakon digitalizacije potrebna dodatna kontrola unesenih podataka.

Također se mogu koristiti kombinirane metode prikupljanja podataka, primjerice telefonska anketa, a zatim terenski razgovori. Svaka metoda prikupljanja podataka ima svoje prednosti i nedostatke, a sljedeće su najtipičnije.

Samopopunjavanje je prije svega isplativ način popunjavanja upitnika, tj. unošenje potrebnih podataka u upitnik. Osim toga, najprikladniji je za prikupljanje osjetljivih podataka. Izvještajna jedinica može ispuniti upitnik kada to želi. Istom primatelju može se istodobno dostaviti nekoliko upitnika. Skup izvještajnih jedinica može biti vrlo velik. Budući da anketar nema utjecaja na odgovore, pouzdanost odgovora može biti lošija. Često je potrebno požurivati one izvještajne jedinice koje ne dostave podatke u roku. Nije prikladan za otvorena pitanja ili za duge, složene upitnike jer opterećuju izvještajnu jedinicu. Primjenom ove metode ne može se postići kontrola nad interakcijom između izvještajne jedinice i drugih osoba, te se ovom metodom obično povećava posao koji se odnosi na uređivanje podataka.

Elektronički upitnici mogu biti mrežni upitnici. Stoga se ispitaniku može poslati poruka/obavijest s poveznicom ili informacijom na kojoj se mrežnoj adresi može popuniti upitnik. Ako elektronički upitnik radi lokalno na nekom računalu onda se koristi kao upitnici CATI ili CAPI. Glavni nedostaci ovog načina popunjavanja upitnika jesu neodgovaranje i premali obuhvat jer dostupnost tehnologije ovisi o različitim čimbenicima (za osobe, na primjer, o prihod ili dobi, a za tvrtke o vrsti djelatnosti ili veličini tvrtke). Elektronički upitnici prikladniji su za ankete u kojima su promatrane jedinice poslovni subjekti. Elektronički upitnici omogućuju upotrebu različitih složenijih oblika, preskakanje pitanja, pristup registrima, popisima šifara itd. Osim toga, može se uključiti šifriranje odgovora, npr. zanimanja. Time se može smanjiti količina ulaznih pogrešaka i omogućiti brža kompilacija prikupljenih podataka jer su logičke kontrole već dio upitnika.

Terenski intervjui (tiskani upitnik za anketara i anketna osoba s pomoću računala) najskuplji su način prikupljanja podataka zbog korištenja ljudskih izvora podataka, vremena i financijskih ograničenja. Prisutnost anketara omogućuje anketaru da usmjeri situaciju i objasni postavljena pitanja, međutim anketar može utjecati na odgovore. Zbog njegove prisutnosti može se ostati i bez odgovora na osjetljiva pitanja, pa je katkad za ispitanika bolje da sam odgovori na neka pitanja, ako postoji mogućnost da se predvidi takav utjecaj. U ovom načinu rada opterećenje izvještajne jedinice je manje, ali neki ne žele da anketar uđe u njihov dom. S pomoću ove metode prikupljaju se podaci u trenutku kada je osoba otkrila događaj. Prikladni su za preuzimanje podataka na mjestima gdje nema telefona, a koriste se i kada izvještajna jedinica nije poznata. Za vrlo kratke upitnike također treba uzeti u obzir vrijeme potrebno za prikupljanje podataka. Katkad je moguće brže ispuniti tiskani upitnik nego uključiti računalo, pokrenuti program i ispuniti elektronički upitnik. Elektronički upitnici nisu prikladni za istraživanja na otvorenome (na primjer na graničnim prijelazima).

Metodom telefonskog anketiranja možemo postići vrlo visoku stopu odgovora ako je upitnik kratak i jasan. Ovome ide u prilog i točnost telefonskih brojeva te spretnost anketara koji mogu pridobiti ispitanika da odgovara na pitanja iz ankete. Od terenskih istraživanja razlikuje se po tome što prostorna raspodjela jedinica ne utječe na troškove, pa su telefonske ankete mnogo jeftinije od terenskih. S druge strane, postoji visok rizik od nedovoljne pokrivenosti (kućanstva bez telefona ili s neobjavljenim brojem). Prikupljanje podataka putem telefonskih anketa može biti vrlo brzo. Zbog računalne podrške u anketi, tijekom ankete može biti složen i može sadržavati otvorena pitanja i mnoštvo preskakanja, ali ne na isti način kao kod terenskih anketiranja. Utjecaj anketara manji je zbog određenog protokola. Također omogućuje znatnu kontrolu nad interakcijom između anketara i između anketara i kontrolora. Telefonska anketa dobar je kompromis između poštanske ankete i terenskih razgovora, ali odabir između nekoliko mogućih odgovora (više od šest) zahtijeva vizualizaciju, što je moguće samo kod drugih metoda prikupljanja podataka.

Glavni problemi s kojima se suočavamo kod ovakvog načina prikupljanja podataka i koji mogu u velikoj mjeri utjecati na rezultat istraživanja jesu neodgovor i netočna isporuka podataka. Neodaziv smanjuje učinkovitu veličinu uzorka, a to posljedično rezultira većom uzoračkom pogreškom. Međutim, u većini slučajeva skupine koje nisu odgovorile na upitnik mogu biti različite, i to onda može dovesti do pristranosti u rezultatima. Stopa odaziva stoga mora biti što veća jer što je veća stopa odaziva, to je manji rizik od pristranosti i velikih pogrešaka u metodi uzorkovanja. Ti se ciljevi mogu postići pažljivom pripremom procesa prikupljanja podataka, ispitivanjem mjernih instrumenata, kvalitetnim izvršavanjem i kontrolom terenskog rada te obradom podataka. Moraju se uzeti u obzir brojna ograničenja, od financijskih i ljudskih resursa, rokova izvršenja pa do izvora podataka. Posebnu brigu treba posvetiti organizaciji rada na terenu i komunikaciji s ispitanicima ili izvještajnim jedinicama. Pogreške kod zaprimljenih podataka mogu se otkriti samo uređivanjem postupaka koji se provode nakon faze prikupljanja.

Administrator baze podataka u svojoj evidenciji čuva podatke koji nisu prikupljeni za statističke svrhe. Upotrebom baza podataka postiže se učinkovitija upotreba te vrste izvora podataka. S druge strane ovakva primjena znači ovisnost o aktivnostima administratora baze podataka, i metodološki i tehnički. Mjesto i oblik pohrane podataka ovise o administratoru baze podataka. Pri preuzimanju baza podataka Državni zavod za statistiku primjenjuje interne upute za preuzimanje baza podataka.

Slijedom navedenog prikupljanje podataka treba biti dobro pripremljeno, plan rada potrebno je uskladiti i, konačno, podatke treba pretvoriti u elektronički format pogodan za daljnju obradu. Kako će ovaj proces izgledati ovisi o tome jesu li podaci prikupljeni samopopunjavanjem ili intervjuom. Kontrola kvalitete koja je u tijeku najvažnija je pri prikupljanju. Provjere upitnika često se provode prije unosa podataka u ulaznu bazu podataka. Preuzimanje administrativnih izvora podataka karakterizira primjena istih faza rada, od koji su neke puno kraće ili automatizirane, pa je ovaj dio obuhvaćen zasebnom točkom.

4.1. Izbor ciljne populacije / uzorka

Opis

Ako se želi pripremiti kvalitetan okvir za izbor uzorka, potrebno je primjenjivati više izvora podataka. Nakon što se ti izvori podataka utvrde i prikupe u odgovarajućem računalnom formatu, treba ih kombinirati s pomoću pripadajuće metodologije kako bi se izradile (dizajnirale) varijable koje određuju ključna obilježja jedinica koje su uključene u okvir za izbor uzorka. Krajnji je cilj pripremiti jedinstvenu tablicu podataka, koja će sadržavati popis jedinica koje odgovaraju teoretski definiranoj ciljnoj populaciji, a u njoj će svaka jedinica odrediti vrijednosti onih varijabli koje će biti potrebne u kasnijem procesu odabira jedinica promatranja.

Ključni korak u provedbi postupka pripreme okvira za izbor uzorka jest utvrđivanje postupka izbora jedinica koje će biti uključene u okvir za izbor uzorka. Polazna točka okvira za izbor uzorka jest baza podataka promatranih jedinica u odabranoj vremenskoj točki. Kvalitetan popis jedinica, koji je određen odabranom vremenskom točkom iz registra, nastoji se poboljšati upotrebom dodatnih administrativnih i statističkih izvora podataka. Primarni je cilj otkriti i zatim isključiti iz popisa jedinice koje su inače u registru, ali zapravo nisu dio ciljne populacije. Katkad se u okvir za izbor uzorka može uvrstiti i jedinica koja nije u registru, osobito ako je riječ o poslovnom istraživanju. Razlog postojanja takvih neodgovarajućih jedinica u registru jesu nedovoljan broj postupaka ažuriranja registra ili je razlog u administrativnoj prirodi registra. Kao što je već spomenuto, u nekim slučajevima pokušavaju se otkriti jedinice koje nedostaju, tj. dijelovi stanovništva koji nisu obuhvaćeni ovim registrom, upotrebom dodatnih izvora podataka. Razlog za takav nezadovoljavajući obuhvat obično proizlazi iz definicije populacije na koju se koristi registar jer ne udovoljava potpuno specifičnim potrebama statističkog istraživanja.

Ako okvir za izbor uzorka služi svojoj svrsi, mora se osigurati da za svaku od jedinica sadržava vrijednosti varijabli koje će biti potrebne u kasnijim postupcima. U postupku uzorkovanja to su uglavnom varijable stratifikacije, varijable koje određuju skupinu prve faze (u dvoetaptom planu uzorkovanja) ili druge varijable koje će biti potrebne u nacrtu uzorka ili koje definiraju područja na kojima želimo objaviti rezultate. Također je potrebno osigurati da svaka jedinica ima jedinstveni identifikator koji će se kasnije koristiti u procesu istraživanja.

Ako je vjerojatnosni uzorak odabran metodom trajnih Bernoullijevih brojeva⁴⁵, riječ je o postupku koji se povremeno ponavlja i kojim se smanjuje opterećenje na izvještajnim jedinicama. Tada je potrebno prebaciti sve slučajne brojeve iz prethodnog okvira za izbor uzorka u trenutačni okvir uzorka i nove jedinice unutar okvira. Odgovarajućim računalnim postupkom dodaje se slučajni broj (između 0 i 1). Primjena metode trajnih Bernoullijevih brojeva vrlo je korisna u koordiniranom uzorkovanju, u kojem se istodobno izabiru uzorci za nekoliko istraživanja. Koordiniranim uzorkovanjem istodobno se prati opterećenje izvještajnih jedinica u nekoliko istraživanja.

Okvir za izbor uzorka jest popis jedinica osnovnog skupa čija se obilježja opisuju statističkim rezultatima. To znači da je u sljedećem koraku potrebno između jedinica u okviru za izbor uzorka odabrati jedinice koje će se uključiti u istraživanje i o kojima će se u kasnijim fazama pokušati dobiti željeni podaci. Iz praktičnih i financijskih razloga rijetko se može priuštiti da se u istraživanje uključe sve jedinice iz okvira za izbor uzorka. Obično se popis jedinica za izbor uzorka treba prilagoditi veličini koja se može financijski podržati i istodobno može omogućiti dovoljno reprezentativan izračun rezultata za cijelu populaciju.

⁴ Bernoullijevi brojevi B_k predstavljaju niz racionalnih brojeva koje je otkrio Jakob Bernoulli, a vezani su za zbroj.

⁵ Jakob Bernoulli (Basel, 6. siječnja 1655. – Basel, 16. kolovoza 1705.) bio je švicarski matematičar. Od 1687. radio je kao profesor matematike u Baselu. Dao je važne priloge teoriji beskonačnih redova, riješio neke od osnovnih problema računa varijacija i znatno unaprijedio teoriju vjerojatnosti u posmrtno objavljenom djelu Umijeće pogađanja (lat. *Ars conjectandi*, 1713.), u kojem se kao glavni poučak pojavljuje zakon velikih brojeva.

Ako je riječ o vjerojatnosnoj metodi uzorkovanja i ako veličina uzorka nije određena pri planiranju istraživanja i ako u okviru uzorka postoje podaci za pomoćne varijable koje su u pozitivnoj korelaciji s istraženim varijablama ili ako su poznati podaci za istraženu varijablu (za jedinice uzorka) iz prethodnoga referentnog razdoblja, izračunava se odgovarajuća veličina uzorka. Kriteriji za izračunavanje odgovarajuće veličine uzorka (bez troškova) određuju se potrebnom točnošću procijenjenih parametara koji se objavljuju. Ti se zahtjevi obično daju na razini cijele populacije, a katkad i na razini regija.

Ako je pomoćna varijabla u okviru za izbor uzorka u pozitivnoj korelaciji s istraženom varijablom, tada se odgovarajuća veličina uzorka dobiva izravno iz formule za varijancu uzorka procjenitelja (glavnog) parametra. Naravno, u ovim se formulama koriste podaci o populaciji pomoćne varijable jer vrijednosti istraženih varijabli još nisu dostupne u ovoj fazi istraživanja. Ako su procjenitelji u složenom obliku, tada se izvodi Monte Carlo simulacija⁶. To je postupak u kojem se odabire velik broj vjerojatnosnih uzoraka za određene veličine uzoraka. Na temelju podataka iz ovih uzoraka izračunava se varijanca uzorka i odabire se najmanja veličina pri kojoj se postiže željena točnost. Ako postoji nekoliko ključnih parametara ili ključnih varijabli, tada se spomenuti postupci provode za sve parametre ili za sve varijable, a konačna veličina uzorka određuje se kao maksimum izračunanih veličina.

Ako se koriste podaci za predmetnu varijablu iz ankete za prethodno referentno razdoblje, tada se izračunava tzv. učinak uzorka (Deff)⁷ odgovarajuće veličine uzorka. Ovo je omjer između varijance uzorka primijenjenog dizajna uzorka i varijance uzorka jednostavnoga slučajnog uzorkovanja bez ponavljanja. U praksi se ovaj postupak koristi vrlo rijetko. Na temelju izračunane procjene varijance uzorka i procjene parametra iz podataka za prethodno referentno razdoblje i pod pretpostavkom jednostavnoga slučajnog uzorka, može se izračunati učinak uzorka Deff. Daljnjom pretpostavkom da je učinak uzorka jednak pri promjeni veličine uzorka (obično se želi povećati veličina uzorka) i željene varijance uzorka u stvarnom dizajnu uzorka izračunava se potrebna veličina uzorka. Izračunana veličina uzorka dijeli se s očekivanom stopom neodgovora kako bi se dobila konačna veličina uzorka.

Nakon što je okvir za izbor uzorka spreman i određena je veličina uzorka, počinje odabir jedinica uzorka koji je već utvrđen u fazi planiranja istraživanja. Za ovu fazu provedbe istraživanja potrebni su odgovarajući softver i određeni algoritmi. Posebno treba spomenuti stratificirani plan uzorkovanja, u kojem se prvo mora izračunati stratumska alokacija uzorka. To znači da se za svaki stratum mora odrediti broj jedinica koje će se odabrati na uzorku. Najčešće se koriste:

- jednaka alokacija: izabire se isti broj jedinica iz svakog stratuma
- proporcionalna alokacija: izabire se toliko jedinica iz svakog stratuma tako da je udio svakog stratuma u odnosu na veličinu okvira uzorka jednak udjelu odabranih jedinica iz ovog stratuma u odnosu na veličinu uzorka (to znači da se iz većeg stratuma odabire više jedinica u uzorak)
- optimalna alokacija: na temelju podataka za pomoćnu varijablu ili prema podacima iz prethodnoga referentnog razdoblja izračunava se odgovarajuća veličina uzorka u svakom stratumu i uzima se u obzir varijabilnost pomoćne odnosno istražene varijable prethodnog razdoblja (u stratumima s većom varijabilnošću utvrđujemo veću veličinu uzorka).

U svim spomenutim alokacijama vodi se računa o tome da broj odabranih jedinica nije premalen ni u jednom stratumu. Ako je nakon alokacije veličina uzorka u stratumu premala, određuje se minimalna fiksna veličina (obično oko deset jedinica) ili se u uzorak izabire cijeli stratum ako je njegov broj jedinica manji od ovoga fiksnog broja.

⁶ Metode Monte Carlo, odnosno simulacije, široka su klasa algoritama koji se oslanjaju na opetovano slučajno uzorkovanje za svrhu dobivanja numeričkih rješenja kod problema koje je teško riješiti.

⁷ Pokazatelj učinka dizajna uzorka (Deff)

Odabir uzorka rezultira težinom uzorka (u slučaju vjerojatnosnog uzorka) i popisom ključnih jedinica. Ponderi uzorka jesu brojevi koji se, zajedno s podacima dobivenima iz jedinica uzorka, koriste u izračunu procjena parametara populacije. Ključne jedinice jesu jedinice koje su važnije od ostalih jedinica za naše istraživanje, s obzirom na očekivani utjecaj na konačni rezultat. Stoga se s njima različito postupa i u fazi prikupljanja podataka i u fazi uređivanja podataka.

Upute za osiguranje kvalitete

Ako se različita istraživanja odnose na istu osnovnu ciljnu populaciju i isto referentno razdoblje, isti postupci trebaju se koristiti za uspostavljanje okvira za izbor uzorka jer to znatno pridonosi većoj dosljednosti statističkih rezultata.

Potrebno je uložiti sve napore i u procesu identificiranja izvora podataka i u procesu proizvodnje. Uklanjanjem neprimjerenih i udvostručenih jedinica može se znatno pridonijeti boljoj kvaliteti statističkih rezultata.

Ako se vrijednosti varijabli okvira za izbor uzorka određuju iz nekoliko različitih izvora podataka, postupci određivanja prioriteta pojedinih izvora podataka moraju se pažljivo razmotriti.

Kvalitetu postupaka za izradu okvira za izbor uzorka i kvalitetu korištenih izvora podataka potrebno je redovito i sustavno ocjenjivati na temelju povratnih informacija koje su prikupljene statističkim procesima. Ako se otkriju velika odstupanja od još uvijek prihvatljivih standarda, potrebno je provesti postupke poboljšanja kvalitete.

U slučaju vjerojatnosnog uzorka, za odabir jedinica treba se koristiti stratifikacijom (osim ako postoje opravdani razlozi da se ona ne koristi) jer to znatno pridonosi reprezentativnosti uzorka i istodobno većoj točnosti statističkih rezultata. Odabrane varijable stratifikacije trebaju biti u visokoj korelaciji s istraženim varijablama ili onima koje određuju područja publiciranja. Prema teoriji uzorkovanja, najučinkovitija je stratifikacija ona u kojoj su jedinice u stratumu što sličnije (s obzirom na ključnu varijablu), a jedinice iz različitih stratuma su što su moguće više različite.

U slučaju izrazito asimetrične raspodjele ciljnih varijabli u populaciji, neki stratumi trebali bi svakako biti uključeni u uzorak (*take-all stratum*). To se posebno preporučuje za poslovna istraživanja kod kojih se obvezno u uzorak uključuju velika poduzeća.

U vjerojatnosnom odabiru uzorka bitno je koristiti se računalno generiranim slučajnim mehanizmom. Za standardne planove uzorkovanja preporučuje se upotreba gotovih procedura. U povremenim istraživanjima mora se kontinuirano provjeravati točnost rezultata dobivenih na temelju slučajnog uzorka kako bi se zadovoljili traženi kriteriji i standardi. Ako nije, dizajn modela trebao bi biti izmijenjen u skladu s tim. Kao dio planiranih promjena, prvo treba istražiti mogućnosti korištenja učinkovitijeg dizajna modela. Samo ako se utvrdi da postupci uzorkovanja ne mogu poboljšati točnost, treba razmotriti mogućnosti za povećanje veličine uzorka.

Preporučuje se da nema previše ključnih izvještajnih jedinica (maksimalno jedna trećina ukupnog uzorka) ako se želi postići željeni učinak identificiranjem i upotrebom tih jedinica.

4.2. Priprema za prikupljanje podataka

Opis

Priprema za prikupljanje podataka obuhvaća postupke planiranja, pripremu za njihovo praćenje, razvoj i testiranje programa te obuku sudionika. Svaka metoda prikupljanja podataka ima svoje specifičnosti po pitanju izgleda i sadržaja upitnika, načina popunjavanja upitnika, primijenjenih postupaka i zahtjeva prema izvještajnoj jedinici.

Tijek i duljina priprema za prikupljanje podataka ovisi o nekoliko čimbenika, od kojih su najvažniji periodičnost statističkog istraživanja, način prikupljanja podataka i tehnologija prikupljanja podataka. U longitudinalnim studijama s mjesečnom periodičnošću priprema se obično izvodi jedanput tijekom duljeg razdoblja, npr. za tromjesečje, za pola godine ili za cijelu godinu, dok za unakrižne ankete nisu potrebne pripreme za dulje razdoblje prikupljanja. Longitudinalna istraživanja obično postaju rutina, tako da mogu biti učinkovitija, ali se mora voditi računa o tome da takva rutina ne prouzroči problem ponavljajućih pogrešaka.

Pri postupku **samopopunjavanja** potrebno je pripremiti sve materijale i pripremiti adresar:

Kod prikupljanja podataka metodom PAPI potrebno je pripremiti odgovarajući broj kopija dokumentacije (upitnika, obavijesti, metodologija i dr.) ovisno o tome koliko adresa sadržava adresar. Pisma obavijesti mogu pridonijeti većoj stopi odgovora te povećati vjerodostojnost ankete na izvještajnim jedinicama. Adresni dokumenti dostavljaju se poštom. Istodobno je potrebno pripremiti alate za bilježenje odgovora i programe za ručni unos podataka i/ili skeniranje.

Dokumenti koje će izvještajne agencije vratiti moraju se pravilno adresirati i pružiti im odgovarajuću identifikaciju za praćenje odgovora. U slučaju tiskanih materijala poželjno je da na svakoj stranici materijala postoji i barkod koji sadržava identifikaciju.

Pri prikupljanju podataka elektroničkim upitnikom izvještajne se jedinice uglavnom informiraju e-poštom (običnom poštom samo iznimno). Ta metoda prikupljanja također zahtijeva posebnu pripremu upitnika, integraciju s adresarima i njegovo testiranje. Pojednostavljen je unos preliminarnih podataka s popisa šifara ili registara.

Intervjui su najskuplji način prikupljanja podataka. Stoga ga je potrebno što detaljnije planirati. Uz pripremu dokumentacije kao i za samopopunjavanje mora se pripremiti sve što je potrebno za sudjelovanje anketara, od popratne dokumentacije, raspodjele izvještajnih jedinica na uzorku po anketarima i program praćenja rada anketara. Nadzor rada anketara trebao bi biti češći kod ove metode prikupljanja podataka nego kod samopopunjavanja. Za potrebe provedbe ankete na terenu pripremaju se radionice za anketere na kojima se provodi izobrazba o načinima pristupa izvještajnim jedinicama, vrsti upitnika i metodologiji za njegovo popunjavanje. Na radionici se također predstavljaju praktični primjeri popunjavanja upitnika i provjeravaju se vještine i stručnost anketara.

Za svako statističko istraživanje potrebno je izraditi i adresar. Adresar se koristi za opisivanje popisa izvještajnih i promatranih jedinica s adresama i ostalim potrebnim podacima. Takav se popis koristi za distribuciju upitnika, provjeru pristizanja upitnika, komunikaciju s izvještajnim jedinicama i za obradu podataka. Izvor za izradu adresara za pojedinačna statistička istraživanja je odgovarajući registar: Statistički poslovni registar Državnog zavoda za statistiku (istraživanje na poslovnim subjektima), baza podataka popisa stanovništva, kućanstava i stanova (ankete o osobama i kućanstvima) i Statistički registar poljoprivrednih gospodarstava (poljoprivredna istraživanja). Odabir jedinica promatranja i izvještajnih jedinica može se utvrditi metodom uzorkovanja, izravnim odabirom iz odgovarajućeg registra na temelju točno postavljenih pravila ili se može formirati na temelju adresara za isto istraživanje u prethodnom razdoblju dodavanjem ili brisanjem jedinica. U adresaru se također mora navesti popis ključnih jedinica, odnosno popis jedinica koje su najvažnije za istraživanje s obzirom na utjecaj na konačni rezultat i koje se tretiraju s većim prioritetom, a posebno u fazi prikupljanja i uređivanja podataka.

Kod periodičnih statističkih istraživanja adresar se mora redovito ažurirati. Treba uzeti u obzir trenutačne promjene u registru i povratne informacije iz statističkih istraživanja. U procesu ažuriranja adresara po potrebi se dodaju nove jedinice i prate se demografske promjene. Otkrivene promjene ne usvajaju se automatski, nego se promjena unosi od slučaja do slučaja. U vezi s adresarima, postoji i briga za rješavanje odbijene pošte s terena (nepoznata, premještena ili nepotpuna adresa).

Upute za osiguranje kvalitete

Način prikupljanja podataka treba prilagoditi ciljevima i metodologiji istraživanja, a ne obratno. Izbor načina provođenja ankete trebao bi biti prilagođen sadržaju. Primjerice, telefonske ankete ne bi trebale biti preduge i ne bi trebale sadržavati složena pitanja s velikim brojem mogućih odgovora. S druge strane, neracionalno je provoditi kratka i jednostavna istraživanja na terenu. Potrebno je dogovoriti raspored i radne upute. Prije ispisa završnog upitnika nositelj istraživanja ga je dužan koordinirati sa sudionicima po područjima.

Za potrebe naknadnih komunikacija podaci izvještajnih jedinica moraju biti jasno vidljivi na svim materijalima koji sadržavaju podatke preuzete od navedenih jedinica. Jednoobraznost i vizualna prepoznatljivost Državnog zavoda za statistiku bitni su u svim dokumentima. Rok za istraživanje trebao bi se odrediti ovisno o metodi uzorkovanja i trajanju istraživanja.

Metode i alate za prikupljanje podataka potrebno je temeljito provjeriti prije početka terenskog rada. Temeljito testiranje obaju programa brzog ulaska, kao i programa kontrole podataka i upitnika anketnog studija, potrebno je detaljno provjeriti. Testiranje provodi organizator istraživanja i ustrojstvene jedinice u kojima se aktivnosti provode. Tek nakon uspješnog završetka testiranja pripremljeni su programi prikladni za upotrebu. Treba provjeriti brzinu unosa podataka i vještinu anketara pri rukovanju telefonom. Prijenosna računala za ankete treba postaviti u skladu s usvojenim standardima. Također se u obzir mora uzeti zaštita podataka od moguće krađe.

U provedbenoj dokumentaciji potrebno je pravodobno definirati datume i osobe odgovorne za plan rada. Adresar mora sadržavati jedinstvenu identifikaciju jedinice za izvještavanje i promatranje te potpunu adresu. Prije dodavanja novih jedinica u adresar nositelj statističkog istraživanja mora pregledati prethodni adresar kako bi se izbjeglo udvostručivanje izvještajnih jedinica. Promatrane jedinice se ne brišu iz adresara. Ako pojedina promatrana jedinica više nije aktivna, onda joj se dodjeljuje odgovarajući status. Ako su podaci prikazani iz drugih izvora podataka koji mogu rasteretiti jedinicu ili poboljšati kvalitetu izvještavanja, potrebno je razmotriti mogu li se relevantni podaci upisati u upitnik prije slanja.

4.3. Prikupljanje primarnih podataka

Opis

Provodi se prikupljanje podataka s pomoću različitih instrumenata za prikupljanje podataka. Uključuje početni kontakt s davateljima podataka te bilo kakve aktivnosti vezane za daljnji kontakt ili podsjećanje davatelja podataka.

Državni zavod za statistiku prikuplja podatke u raznim formatima. Ako su zaprimljeni u elektroničkom obliku, trebali bi biti prikladni za agregiranje u ulaznoj bazi podataka za razliku od prikupljenih podataka iz tiskanih upitnika koji se trebaju najprije prebaciti u elektronički oblik za kasniju obradu. Prebacivanje podataka u elektronički oblik može se izvršiti na dva načina: ručnim unosom ili skeniranjem.

Odgovarajući način unosa podataka određuje se formatom polja za unos i općim formatom te sadržajem upitnika. U posebnim slučajevima provjera upitnika može se izvršiti prije unosa, dok u slučaju tiskanog upitnika izvorni podaci se trebaju čuvati u elektroničkom obliku. Opseg kontrole ovisi uglavnom o tome koji će se postupak unosa podataka koristiti: skeniranje ili ručni unos podataka. Ako se koristi skeniranje, pohranjuju se zapisi podataka u odgovarajuća polja i dodatno ih se šifrira.

Potrebni programi i postupci za ručni unos i skeniranje podataka pripremaju se u fazi pripreme prikupljanja podataka. Istodobno je potrebno testirati postupke i programe i zatim izvršiti sve ispravke.

Kada se podaci prikupe kombiniranim metodama izvještavanja (npr. tiskani, elektronički upitnici, upitnici CATI i CAPI) ili kada se unos podataka obavlja kombiniranim postupcima (npr. skener i ručni unos) ili se dio podataka preuzima iz administrativnog izvora podataka slijedi objedinjavanje podataka u jedinstvenu bazu podataka. Nadalje se podaci mogu provjeravati, uređivati i kasnije uspoređivati s distribucijom populacije u bazi podataka. Time se olakšava utvrđivanje pouzdanosti rezultata.

Upute za osiguranje kvalitete

Nositelj istraživanja mora pružiti jasne i nedvosmislene upute za provjeru podataka prije unosa. Upitnik mora biti razumljiv, transparentan i pravilno dizajniran te pregledan tijekom pripreme skupa podataka.

Program unosa podataka mora se pripremiti točno i pravodobno. Prije praktične primjene mora se temeljito testirati kako bi se izbjegle pogreške kod izvornih podataka.

Kako bi se smanjio broj pogrešaka kod digitaliziranja podataka, pri prikupljanju podataka prednost se daje elektroničkim upitnicima. Pri prikupljanju podataka u tiskanom obliku prikladnija je primjena skeniranja.

Ako koristimo skeniranje ili brzi unos podataka bez ponavljanja, upitnik treba biti dizajniran tako da se mogu obaviti interne kontrole (npr. posebno polje za ukupne vrijednosti). Podaci na upitnicima koji se čitaju na skeneru moraju biti napisani jasno i čitko. Prije bilježenja podataka iz tiskanih upitnika treba izbjegavati optičke provjere.

Inače se podaci moraju provjeriti i nakon optičkog čitanja jer skener može pogrešno pročitati podatke s obrasca.

Primjer organiziranja postupka skeniranja – Popis 2011.

1. Voditelj smjene

U okviru dobivenog rasporeda, a na traženje voditelja skeniranja, zadužen je za skeniranje određenoga grada / općine. Evidentira zaduženje u vlastitoj evidenciji i daje mu "Nalog za skeniranje".

2. Voditelj skeniranja

Brine o pravodobnoj opskrbi svakog skenera sa građom za skeniranje (svaki skener radi neovisno, tj. skenira drugi grad / općinu). Prema dobivenom "Nalogu za skeniranje" popunjava "Nalog za preuzimanje građe iz arhive", u koji obavezno uz ime voditelja navodi oznaku skenera za koji se građa "diže" i predaje ga voditelju arhive.

3. Voditelj arhive

Preuzima "Nalog za preuzimanje građe iz arhive" te provjerava u evidenciji je li tražena građa kompletirana u skladištu. Ispisuje "Popis stanja skladišta" traženoga grada / općine te ga predaje voditelju skeniranja. Zadužuje svoje zaposlenike da zahtijevanu građu rastućim redoslijedom obrasca PK smjeste na označene police skenera naznačenoga u "Nalogu za preuzimanje građe iz arhive". Građu iznesenu iz arhive evidentira u vlastitoj evidenciji, a zadržava "Nalog za preuzimanje građe iz arhive".

4. Voditelj skeniranja

Nakon što mu je tražena popisna građa složena na police, kontrolira njezinu kompletnost prema "Popisu stanja skladišta". Određuje početak skeniranja određenoga grada / općine.

5. Skupina za pripremu

Uzimaju redosljedom kutije s obrascima PK s police i odlažu ih na (posebno naručeni) stol za pripremu (na stolu se istodobno uvijek nalaze četiri kutije s obrascima PK). Otvaraju kutije te kontroliraju postojanje obrasca P-0, identifikaciju općine i popisnoga kruga. Spreman materijal za skeniranje nose i odlažu na stol ulazne strane skenera. Iz kutija s obrascima PK izdvajaju kontrolnike i slažu ih rastućim redosljedom u zaseban fascikl za svaku općinu na koji crnim flomasterom upisuju indentifikaciju (županija, grad/općina, MB) te raspon rednih brojeva u obrascima PK (npr. 1-0100). U svaki se fascikl ulaže do 100 kontrolnika (ovisno o veličini grada/općine), vodeći pritom računa da su svi višedjelni kontrolnici u istom fasciklu. Isto tako, izdvajaju se skice i opisi obrazaca PK u zasebne fascikle uz upis šifre u obrascima PK (po P-8) u desni gornji ugao. Za skice i opise koji nedostaju popunjava se "Evidencija nedostajućih skica", koja se ulaže na vrh fascikla.

6. Operateri na skeneru

Obavljaju skeniranje po proceduri, pazeći pritom na redosljed slaganja građe na izlazu. Po završetku svakoga pojedinog obrasca PK verificiraju broj pročitanih dokumenata.

7. Skupina za pripremu

Po skeniranju svakog obrasca PK evidentiraju završetak zaokruživanjem rednog broja u obrascima PK u "Popisu stanja skladišta". Uspoređuju je li broj skeniranih dokumenata identičan broju upisanu na obrascu P-0. U slučaju eventualne razlike u "Popisu stanja skladišta" upisuju broj obrazaca prema P-0, broj skeniranih obrazaca te razliku u odnosu na skener. Materijal vraćaju u pripadajuću kutiju s obrascima PK. Na kutiji označuju završetak skeniranja upisom oznake "S" crvenim flomasterom ispod naslova Popis 2011. Kutiju vraćaju natrag na isto mjesto na policu.

8. Voditelj skeniranja

O završetku skeniranja grada/općine izvještava voditelja smjene te popunjava "Nalog za spuštanje građe u arhivu", koji predaje voditelju arhive.

9. Voditelj arhive

Postupa prema dobivenu "Nalogu za spuštanje građe u arhivu" te zadužuje svoje zaposlenike za spuštanje zahtijevane građe s polica u arhivu. Također preuzima i arhivira fascikle s izdvojenim kontrolnicima. Zabilješku o vraćanju građe evidentira u vlastitoj evidenciji, a nalog zadržava.

U ovom procesu postojat će sitne preinake, ovisno o radu obradnog centra.

Prikupljanje podataka i komunikacija s izvještajnim jedinicama ovise o načinu popunjavanja, ciljnoj populaciji, promatranom fenomenu i dostupnim resursima. Provedeno vrijeme i neodgovor izvještajne jedinice u pravilu su obrnuto proporcionalni. Pri prikupljanju podataka posebnu pozornost treba obratiti na gubitak što manje podataka zbog ljudskih ili sistemskih pogrešaka anketara ili pogrešaka u obradi podataka.

Visoka stopa odgovora izvještajnih jedinica može se postići postavljanjem jasnih ciljeva, upotrebom odgovarajućih alata te odabirom prikladnog načina prikupljanja podataka. Posljedica provedbe istraživanje u nepovoljno vrijeme za izvještajne jedinice, npr. tijekom odmora ili većih sezonskih opterećenja (priprema godišnjih financijskih izvještaja) jest niža stopa odgovora. Na stopu odgovora može utjecati i vrsta istraživanja, npr. anketa o zdravlju ili računovodstvene ankete imaju veću stopu odgovora, dok ankete o stanovanju, kulturi i inovacijama obično imaju nižu stopu odgovora.

Neodgovor se može podijeliti u dvije glavne kategorije: neodgovor na varijablu i neodgovor jedinice. O neodgovoru pojedine varijable govorimo kada neka pitanja u upitniku, koji je inače popunjen, nemaju odgovor (vektor podataka je nepotpun). Ova se situacija može dogoditi kada ispitanik nema dovoljna znanja za popunjavanje odgovora na postavljeno pitanje ili je riječ o osjetljivu pitanju na koje ispitanik ne želi odgovoriti, a može biti riječ i o kontradiktornom odgovoru koji nije u skladu s pravilima kontrole podataka.

O neodazivu jedinice govori se kada nedostaju podaci za sve varijable. Najviša se stopa neodgovora javlja pri samopopunjavanju upitnika. Stoga se u anketnim istraživanjima na uzorku na kućanstvima i fizičkim osobama podaci obično prikupljaju tehnikom intervjua jer anketari znaju motivirati ljude i tako smanjiti stopu neodgovora.

Trajanje prikupljanja podataka također ovisi o načinu popunjavanja upitnika. Telefonska anketa obično je brža od ankete koja se provodi "licem u lice" ili poštom. U poštanskim anketama teško je procijeniti trajanje postupka prikupljanja jer to ovisi o broju podsjetnika koji se dostavljaju onima koji ne odgovaraju i, prije svega, o reakciji izvještajnih jedinica. Što se više vremena i izvora podataka potroši, to će biti veća i stopa odaziva. Postoji mišljenje da u jednom istraživanju možemo zadovoljiti samo dva od sljedeća tri zahtjeva: niži trošak, visoku stopu odaziva i malo utrošenog vremena. Praćenje razloga neodgovora provodi se češće i detaljnije u intervjuima nego pri samopopunjavanju.

Pri samopopunjavanju izvještajna jedinica sama tumači postavljena pitanja i upute te dostavlja odgovore u skladu s razumijevanjem dostavljenih uputa, tako da podaci često sadržavaju mnogo više pogrešaka nego kod drugih metoda prikupljanja podataka: takav način prikupljanja podataka kasnije zahtijeva više rada kod postupaka uređivanja podataka. Popunjeni upitnici ili komentari izvještajnih jedinica mogu se dobiti redovito poštom, elektroničkom poštom, elektroničkim izvješćivanjem ili telefonom. Evidencija odgovora i neodgovora prati se redovito. Izvještajnim jedinicama koje nisu dostavile odgovor na upitnik dostavljaju se najviše dva podsjetnika u pisanom obliku. Nekada je za ključne izvještajne jedinice potrebno obaviti i telefonski razgovor te zamoliti izvještajnu jedinicu za nužnost i svrsishodnost popunjavanja upitnika.

Kada se podaci prikupljaju metodom intervjua, anketar ispunjava upitnik u ime izvještajne jedinice. Poznata su mu pitanja i kakvi se odgovori očekuju na terenu, te u skladu s očekivanjima popunjava statistički upitnik. Budući da je anketaru upoznat s upitnikom, on može nesmetano popunjavati upitnik bez ikakva zastoja. Komunikacija između Državnog zavoda za statistiku i anketara odvija se načinima komuniciranja kao i kod samostalnog popunjavanja upitnika. Državni zavod za statistiku vodi računa o tome da je uvijek na raspolaganju odgovarajući broj anketara i da se moguća otkazivanja mogu kompenzirati što je prije moguće. Svaki se neodgovor kontinuirano prati slanjem anketara na teren u dogovorenim rokovima. Ako anketar ne ostvari plan dovoljnog broja popunjenih upitnika ili ako ispunjeni upitnici ne ispunjavaju odgovarajuće standarde kvalitete ili ako se otkriju metodološke pogreške, najprije ga se upozori, a zatim se nastavlja s provjeravanjem upitnika. Ako se ponove pogreške ili ozbiljna kršenja, anketar se jednostrano isključuje iz postupka prikupljanja podataka i zamjenjuje ga drugi anketar.

Kod telefonskog anketiranja kontrolor je zadužen za kontinuirano nadziranje rada anketera. Prati kako anketa napreduje, njezinu kvalitetu, upozorava anketera na pogreške i moguću nedovoljnu kvalitetu te se brine o pauzama tijekom rada. Podučavanje anketara, praćenje i distribucija dodatnih uputa jednostavnija je nego kod drugih metoda prikupljanja podataka. Uobičajena je praksa da svaki novi anketar ima nositelja ankete koji je u mogućnosti odgovoriti na metodološka pitanja iz ankete.

Organizacija i rad na terenu trebaju na odgovarajući način biti prilagođeni ciljnoj populaciji. Vozni red, informativna pisma, brošuru o istraživanju, upute, metode i alate za prikupljanje podataka, tehničke uređaje i druge alate potrebno je pažljivo planirati radi smanjenja tereta izvještavanja za izvještajne jedinice.

Pri prikupljanju i komuniciranju s izvještajnim jedinicama slijede se upute napisane u provedbenoj dokumentaciji. Pri radu se moraju poštivati rokovi zabilježeni u planu rada. U komunikaciji s izvještajnim jedinicama i anketarima potrebno je poštivati utvrđena pravila Državnog zavoda za statistiku. Probleme je potrebno rješavati brzo i strpljivo.

Za rad koji se odnosi na prikupljanje podataka te na komunikaciju s izvještajnim jedinicama kod anketara je jako bitno da poslužitelji (serveri) rade besprijekorno i nesmetano (sedam dana u tjednu, 24 sata dnevno). Potrebno je stalno težiti pronalaženju izvještajnih jedinica s kojima Državni zavod za statistiku nije uspostavio nikakav kontakt. Također je potrebno koristiti se odgovarajućim postupcima smanjenja odbijanja.

Broj odbijenica mora biti minimalan. Treba utvrditi segmente stanovništva s većom stopom neodgovora i povećati intenzitet kontakata s njima. Ako su promatrane varijable u populaciji izrazito asimetrično raspoređene, potrebno je osigurati odgovor najdominantnijih jedinica.

Jedinice trebaju prijaviti minimalan skup podataka kako bi se oni i dalje mogli smatrati potpunim odgovorom. Potrebno je osigurati da izvještajne jedinice ne snose nikakve troškove za potrebe statističkog izvještavanja, osobito kod anketiranja osoba ili kućanstava. Na primjer, uz upitnik je potrebno priložiti omotnicu s plaćenom poštarinom koju je potrebno popuniti i poštom vratiti u Državni zavod za statistiku.

Stopa neodgovora na upitnik također se može smanjiti korištenjem simboličnih darova. Međutim, pritom treba postupati racionalno u skladu s načelom dobroga financijskog upravljanja.

U Državnom zavodu za statistiku takav se pristup koristi samo u iznimnim slučajevima jer ako izvještajne jedinice dugoročno neracionalno koriste darove, to može imati negativan učinak. Kod poštanske metode potrebno je pomno nadzirati je li prava izvještajna jedinica zaprimila pravi statistički upitnik. Ako izvještajna jedinica nije zaprimila određeni upitnik, potrebno je ponovno dostaviti isti upitnik.

O situaciji u vezi sa zaprimanjem upitnika treba redovito izvještavati nositelja istraživanja. Mora postojati sustavna kontrola procesa odnosno sustavno praćenje napretka u prikupljanju podataka i kontroli podataka za koje nije preuzet nikakav odgovor na postavljeno pitanje. Cilj je postići razuman omjer između kvalitete podataka i cijene organiziranja i provođenja statističkog istraživanja. Važni pokazatelji kvalitete jesu: stopa odgovora, pogreške u obradi, potreban broj dostavljenih popunjenih obrazaca i stopa neodgovora ovisno o razlozima odbijanja.

Pri pripremi odgovarajućih postupaka mora se uzeti u obzir i razdoblje prikupljanja podataka, trajanje prikupljanja podataka i očekivano opterećenje za sve sudionike, kao što su duljina upitnika, kompleksnost sadržaja upitnika, periodičnost ankete i slično.

Metodološki gledano, potrebno je odrediti maksimalan broj kontakata s izvještajnom jedinicom. Izvještajnim jedinicama obično se dostavljaju najviše dva pisana podsjetnika. Za ključne jedinice drugi ili treći podsjetnik može biti i telefonski podsjetnik.

Za ključne izvještajne jedinice potrebno je prikupiti odgovore na postavljena pitanja. Ako ključne jedinice ne dostave nikakve odgovore, moli ih se da pismeno dostave razloge zašto nisu mogli dostaviti podatke. Kod ključnih jedinica treba voditi evidenciju o datumima odgovora ili ih prikazati u odgovarajućem statusu.

Na početku ankete nositelj statističkog istraživanja treba biti uključen u nadzor nad radom anketara. Pri prijenosu terenskih podataka u Državnom zavodu za statistiku primjenjuju se suvremene tehnologije. Uspostavom redovitog nadzora anketara postiže se bolja kvaliteta poslovnih procesa. S obzirom na to da je telefonski imenik osnovni alat za nesmetano odvijanje poslova u anketnom studiju, potrebno ga je redovito ažurirati.

Za rad anketara na terenu optimalno je da prijenosna računala koja se preuzimaju za obavljanje terenskog posla imaju ista funkcionalna i tehnička obilježja.

Kako bi se olakšalo i ujednačilo provođenje anketa i pribavili potrebni ljudski, financijski i informatički resursi, nužno je unaprijed isplanirati raspored istraživanja koja će se provoditi u sljedećoj godini.

4.4. Preuzimanje podataka iz administrativnih i drugih sekundarnih izvora

Opis

Preuzimanje administrativnih podataka obavlja se na temelju sklopljenih ugovora između institucija, dok se prijenos odvija putem različitih medija (optički mediji, mrežni protokoli, replikacije registra itd.). Pri preuzimanju podataka obavlja se kontrola strukture i formata podataka. Ova aktivnost uključuje planiranje primjene administrativnih izvora podataka, preuzimanje podataka od drugih institucija za različite svrhe i organiziranje suradnje s drugim institucijama kao i prijenos određene faze istraživanja u drugu instituciju.

U ovom dokumentu koristimo se izrazom administrativni izvor podataka za sve izvore podataka koji nisu prikupljeni statističkim istraživanjem. Riječ je o podacima koji se prikupljaju na temelju upravnog propisa i drugih sekundarnih izvora podataka, a mogu se koristiti za statističke svrhe kao pomoćni alat za provođenje klasičnih statističkih istraživanja, kao primarni izvor podataka ili za dopunu, uređivanje i kontrolu podataka koji se prikupljaju statističkim istraživanjima. Prednost primjene administrativnih izvora podataka jest u tome što se koriste već dostupni podaci, pa nije potrebno dodatno opterećivati izvještajne jedinice. Uz to, primjena administrativnih izvora podataka jeftinija je u usporedbi s klasičnom provedbom statističkog istraživanja. S obzirom na to da su administrativni izvori podataka uspostavljeni za nestatističke svrhe, potrebno je usredotočiti se na njihove nedostatke, kao što su primjena različitih definicija, nedostatak kontrole nad kvalitetom podataka i pravodobnost podataka. Primjena administrativnog izvora podataka za aktivnost prikupljanja podataka ne znači da ona potpuno zamjenjuje sve statističke postupke koji se odvijaju u klasičnome statističkom istraživanju pa se tako te vrste aktivnosti odnose na uređivanje podataka, povezivanje zapisa, usklađivanje sa statističkim identifikatorima itd.

Za potrebe provođenja nekih statističkih istraživanja Državni zavod za statistiku surađuje s drugim institucijama, bilo da je riječ o organizacijskim ili tehničkim razlozima. Ako se promatra cijeli poslovni proces, može se dogoditi da budu uključene institucije za provedbu određene faze istraživanja, npr. za elektroničko prikupljanje podataka.

Kada se koriste administrativni izvori podataka, potrebno je sklopiti sporazum o suradnji. Svrha sporazuma u vezi s prikupljanjem podataka jest utvrditi sadržaj administrativnog izvora podataka, način preuzimanja te uvjete i pravila korištenja; svrhu sklapanja sporazuma o suradnji te podjelu zadataka između potpisnika sporazuma.

Ovisno o vrsti podataka i posjedniku administrativnog izvora podataka od kojeg se preuzimaju podaci, sklapa se sporazum i dogovara protokol za razmjenu podataka između posjednika administrativnog izvora podataka i korisnika servisnog sučelja registra (primjer DGU – Registar prostornih jedinica) ili se sklapa sporazum i definira Katalog potraživanja statističkih podataka iz evidencija (primjer Porezna uprava). U slučaju preuzimanja podataka od Financijske agencije na godišnjoj razini potpisuje se Protokol o predaji i preuzimanju baze podataka iz financijskih izvještaja.

Prikupljanjem podataka statističkim istraživanjima nastoji se što je više moguće smanjiti stopa neodgovora. U anketama u kojima se podaci dobivaju na temelju sekundarnih izvora podataka (administrativni registri, evidencije itd.) problem neodgovora obično je manji. Međutim, pojavljuju se drugi problemi, posebno oni koji proizlaze iz metodoloških nedosljednosti sa statističkim konceptima.

Evidencije sekundarnih ili administrativnih izvora podataka treba čuvati na jednome mjestu, tako da se u slučaju dodatnih potreba najprije mogu provjeriti postojeći izvori podataka. Evidenciju je potrebno redovito i kontinuirano voditi i u njoj bilježiti datum posljednjeg prikupljanja podataka. Ako željeni izvor podataka još nije preuzet, ovisno o sadržaju administrativnog izvora podataka, potrebno je sklopiti sporazum, tehnički protokol i/ili definirati katalog podataka.

Prikupljanje podataka obavlja se u skladu s odredbama sporazuma i tehničkim protokolom. Prije prvog preuzimanja podataka potrebno je pripremiti i provjeriti upute, postupke i programe za preuzimanje podataka i provođenje formalnih kontrola, kao što su čitljivost podataka, primjerenost zapisa i njihova broja te provjera ispravnosti znakova. Uobičajeni način preuzimanja podataka jest da Državni zavod za statistiku dobije obavijest da su podaci pripremljeni. Nakon toga slijedi prihvaćanje i tehnička provjera podataka te predaja podataka na dogovoreno mjesto.

Prijenos podataka može se organizirati s pomoću raznih medija za pohranu podataka, kao što su CD-ovi, DVD-ovi, prijenosni diskovi, USB stickovi, FTP protokoli i slično. U slučaju nepravilnosti u prihvaćanju pogrešaka u strukturi podataka povratne informacije dostavljaju se izvoru podataka.

Administrativni izvori podataka mogu, kada se koriste kao osnovni izvor podataka, sadržavati pomoćne varijable s pogreškama. Kako bi se izbjegle neželjene situacije, iz predostrožnosti je potrebno provjeriti administrativni izvor podataka prije upotrebe podataka ili tijekom upotrebe otkriti pogreške i otkloniti ih u suradnji s vlasnikom administrativnih izvora podataka. Uređivanje administrativnih izvora podataka posebno je važno pri povezivanju više izvora podataka. Kako bi integracija bila što lakša i učinkovitija, svi resursi koji se koriste moraju se prvo urediti kako bi se utvrdilo imaju li ikakve pogreške koje bi mogle onemogućiti integraciju podataka.

Najčešće vrste kontrola koje se provode jesu provjera vrijednosti pridruženih popisima šifara, provjera raspona vrijednosti za numeričke varijable (npr. negativne vrijednosti), provjera procijenjenog broja zapisa i provjera prikladnosti referentnih razdoblja na koja se podaci odnose. U ovom se trenutku može provjeriti ispravnost veza između nekoliko varijabli u istom izvoru. Većina se provjera dosljednosti provodi kasnije u procesu statističkog uređivanja kada se također provjerava kompatibilnost varijabli iz različitih izvora, administrativnih i statističkih izvora podataka.

Upute za osiguranje kvalitete

Svi se izvori podataka po primitku moraju jasno i točno dokumentirati jer će to olakšati rad u kasnijoj fazi upotrebe administrativnih izvora podataka u statističkim istraživanjima.

Za svaku varijablu koja je povezana s popisom šifara potrebno je točno odrediti koja je verzija popisa šifara primjerena obradi i diseminaciji podataka. Za sve numeričke varijable potrebno je provjeriti obilježja podataka i zatim odrediti skup prihvatljivih vrijednosti. U većini slučajeva potrebno je odlučiti hoće li biti dopuštena nula ili čak negativne vrijednosti.

Sve kontrole smislenosti za pojedine varijable moraju se pažljivo dokumentirati. Ako je vrijednost ograničena za određene varijable, mora se navesti razlog takva "ograničavanja". Povratne informacije o otkrivenim pogreškama treba dostaviti administratoru baze ako je pogreška utvrđena isključivo tijekom kontrole izvora podataka.

U slučaju utvrđivanja pogrešaka trebali bi se definirati postupci za njihovo sustavno rješavanje na racionalan način. Tehničko rješenje trebalo bi biti što općenitije, jednostavno i detaljno dokumentirano.

Prije početka prikupljanja podataka moraju se pripremiti i potpisati sporazum i/ili katalog potraživanja. U katalogu potraživanja trebaju biti navedene sljedeće stavke: prijamni odjel u Državnom zavodu za statistiku, osoba zadužena za prijam u Državnom zavodu za statistiku, korisnički odjel u Državnom zavodu za statistiku, odgovorna osoba za korištenje datoteke u Državnom zavodu za statistiku, odgovorna osoba vlasnika administrativnog izvora podataka te odgovorna osoba – informatička podrška vlasnika administrativnog izvora podataka. Svi sudionici u procesu moraju biti upoznati sa sadržajem podataka prije prvog preuzimanja. Vlasnik administrativnog izvora podataka mora pravodobno i u dogovorenom roku obavijestiti korisnika podataka da su podaci spremni. Unos podataka o preuzimanjima u evidenciji podataka treba biti organiziran kontinuirano. U slučaju kašnjenja vlasnik administrativne baze podataka treba se pobrinuti da korisnici budu pravodobno obaviješteni o promjenama, mogućim kašnjenjima i drugim hitnim slučajevima.

Kako bi se smanjile administrativne prepreke i pojednostavnilo poslovanje, nastoji se što je više moguće koristiti se administrativnim izvorima podataka. Pod pretpostavkom da su administrativni izvori podataka zadovoljavajuće kvalitete, ovaj način prikupljanja podataka ima prednost nad ostalima. Iako administrativni izvori podataka imaju prednosti, ipak treba voditi računa o nedostacima. Sve nedostatke potrebno je minimalizirati te administrativni izvor prilagoditi statističkim potrebama jer postoje razlike u obuhvatu, definicijama, pravodobnosti podataka, kvaliteti izvora podataka itd. Kako su pravne odredbe osnova za uspostavu administrativnog izvora podataka, potrebno je pomno proučiti sadržaj i metodološko značenje u smislu primjene za statističke svrhe. Uspostava kontinuirane suradnje s administrativnim posjednicima podataka bitna je i kod prvog preuzimanja podataka i kod provedbe svih sadržajnih promjena na administrativnim podacima. Primjena administrativnih izvora podataka katkada je zahtjevnija od provedbe uobičajenoga statističkog istraživanja jer uključuje drugačiji pristup kod obrade i uređivanja podataka. Prema članku 50. Zakona o službenoj statistici (NN, br. 25/20.), posjednici administrativnih izvora podataka i podataka prikupljenih metodom promatranja i praćenja dužni su prije uvođenja prikupljanja administrativnih izvora podataka i podataka prikupljenih metodom promatranja i praćenja o tome pisanim putem pravodobno obavijestiti Državni zavod za statistiku.

Kada prilike dopuštaju, Državni zavod za statistiku trebao bi sudjelovati ili u pripremi i u uvođenju novoga administrativnog izvora podataka. Pri preuzimanju administrativnog izvora podataka potrebno je detaljno poznavati sadržaj podataka koje će Državni zavod za statistiku preuzeti. Povjerljive podatke potrebno je zaštititi prebacivanjem u statističke identifikatore i kontroliranom dodjelom pristupa podacima. Ovisno o sadržaju podataka, pristupom ovoj vrsti podataka upravlja administrator podataka. Prije provođenja statističke obrade podataka podaci se moraju prilagoditi statističkim potrebama te se mora provjeriti njihova kvaliteta.

4.5. Unos prikupljenih podataka

Opis

Uključuje unos prikupljenih podataka i metapodataka za daljnju obradu u postupak. Može uključivati ručni unos podataka iz tiskanih upitnika ili s pomoću optičkog čitača, automatsko preuzimanje podataka putem mrežnog upitnika i aplikacija (CAWI, CAPI i CATI) *web scrapinga* ili pretvaranje podatkovnih datoteka primljenih od drugih institucija u drugi format.

Pripreme za automatsko šifriranje

Ako je na upitniku upisan tekst koji predstavlja vrijednost nekog obilježja, onda ga je najčešće potrebno prevesti u numeričku oznaku (šifru) pripadajuće klasifikacije. Taj se postupak zove šifriranje. Moguće ga je provesti u cjelini "ručno", tako da obučena skupina osoba čita upisane tekstove, pretražuje (ili poznaje) klasifikaciju i zatim upisuje šifre za koje pretpostavlja da su odgovarajuće.

Drugi način jest da se proces čitanja teksta, pretraživanja klasifikacije i određivanja odgovarajuće šifre provodi s pomoću računala. U tom slučaju potrebno je razviti algoritam koji obavlja te poslove umjesto ljudi i napisati odgovarajući program. Uobičajeno je da takve aplikacije omogućuju i potpuno automatsko šifriranje kada je vjerojatnost točnog odgovora vrlo velika, ali i da pružaju mogućnost konačnog izbora stručnjaku ako je vjerojatnost jednog izbora nedovoljno velika i/ili je podjednaka vjerojatnosti izbora neke druge šifre. Drugi slučaj naziva se računalno podržano šifriranje.

Svi programi za automatsko šifriranje imaju neku definiciju sličnosti (blizine) i pokušavaju odrediti koji je od tekstova klasifikacije nasličniji upisanom tekstu za koji se traži šifra. Kod malih klasifikacija (od jedne riječi) to je vrlo jednostavno, ali kod složenih velikih klasifikacija to je prilično složen postupak. Kvaliteta dobivene šifre, učinkovitost prepoznavanja i brzina rada mjere su kojima se mjeri ukupna uspješnost takvih algoritama i programa. Ulazni podaci u programe za šifriranje s jedne su strane tekstovi za koje je potrebno odrediti šifru, a s druge strane to su proširene klasifikacije obilježja koje se šifriraju. Klasifikacije su za tu potrebu proširene

novim redcima u kojima se uz već postojeće šifre upisuju očekivani odgovori koji se razlikuju od onih postojećih, službenih. Time se preuzimaju klasifikacije s više naziva uz istu šifru. Tako proširene klasifikacije nazivaju se tezaurusima.

Tezaurusi su iznimno važni za ukupnu uspješnost automatskog šifriranja i stoga ih je potrebno kvalitetno proširiti još prije početka rada. Naravno, pritom treba upisivati samo stvarno očekivane odgovore, a ne ići u krajnost pa upisivati neke iznimno rijetke i malo očekivane tekstove. Time se smanjuje i brzina rada i učinkovitost prepoznavanja. Tezauruse je poželjno puniti i dograđivati i tijekom procesa automatskog šifriranja, kao i na temelju rezultata takva posla. Ako primijetimo da se neki tekst javlja više puta, a nije uključen u tezaurus, potrebno ga je uključiti. Kvalitetna aplikacija automatskoga šifriranja mora podržavati i takvu funkciju koju obično nazivamo učenje. Osim toga, analizom dodijeljenih šifri može se primijetiti da se neke od prije dodijeljenih neslužbenih vrijednosti uopće ne koriste, pa ih je potrebno izbaciti iz tezaurusa.

Tezauruse je poželjno proširiti testovima u kojima su upisane neke očekivane kratice, sinonimi, stari nazivi koji se još koriste, opisi koji pripadaju nekoj šifri, a nisu navedeni u službenom tekstu i slično. Naravno, tezauruse je potrebno upisati na računalo u odgovarajućem formatu i tamo ih se kasnije održava, dopunjava i ispravlja.

Kako bi se velika količina prikupljenih podataka na upitnicima u tiskanom obliku mogla brzo očitati i prenijeti u digitalni medij, koriste se optički čitači. Osnovna obilježja optičkih čitača jesu brzina rada i razina pouzdanosti. Brzina rada ovisi izravno o brzini umetanja papira u stroj i o srednjoj brzini čitanja znakova. Na neke od tih parametara može se utjecati:

- broj pitanja i znakova koje je potrebno pročitati
- razina pouzdanosti koja se zahtijeva za svaki znak posebno
- čitljivost rukom pisanih znakova s obzirom na željeni standard
- broj i složenost kontrola
- vještina operatera koji posluhuje optički čitač (OCR).

Zahtjev za većom razinom pouzdanosti omogućuje i bolju kvalitetu čitanja, ali istodobno usporava čitanje. Osim toga, oni znakovi koji nisu prepoznati, tj. oni koji nisu zadovoljili razinu pouzdanosti dolaze kasnije operatoru na ispravak, a ti ispravci mogu ozbiljno usporiti cijeli proces. Vrlo je bitno održavati istu brzinu optičkog čitača i odgovarajućega ručnog ispravljanja.

Računalno potpomognuto internetsko anketiranje (CAWI) dio je metodologije koja se temelji na upitniku dostavljenom ispitaniku s poveznicom u panelu ili na mrežnome mjestu. U usporedbi s ostalim načinima prikupljanja podataka, najekonomičniji je način prikupljanja podataka jer nisu potrebni anketari, uređaji ili dodatni alati. Stoga su mrežne ankete, koje se još nazivaju i mrežni upitnici, jedna su od najčešćih metoda prikupljanja podataka. Kada se Državni zavod za statistiku odluči za metodu prikupljanja podataka metodom CAWI, sve aktivnosti bit će usmjerene na dizajn upitnika jer je stopa odgovora izravno povezana s kvalitetom upitnika.

Metoda CAWI može biti prikladna iz nekoliko razloga. Ako je upitnik dobro dizajniran, automatski će upravljati putem pitanja koristeći se logički postavljenim uvjetima, poput prikaza ili preskoka. Upitnik se može strukturirati tako da olakša razumijevanje i poveća stopa odgovora. Mogu se umetnuti upute kako bi se ispitaniku pomoglo da razumije postavljena pitanja i popuni mrežni upitnik. Ovim načinom prikupljanja podataka smanjuje se trošak prisutnosti anketara, nabave elektroničkih uređaja te vrijeme potrebno za analizu podataka jer su ti podaci dostupni u stvarnom vremenu u administratorskoj bazi podataka. Neki softveri omogućuju da se u stvarnom vremenu pripreme izvještaja o mrežnim anketama i statističkim podacima koji su prikupljeni na taj način. Glavni nedostatak mrežnih anketa jest nedostatak anketara koji su u nekim slučajevima itekako korisni jer mogu pomoći i voditi ispitanike kroz popunjavanje upitnika za određeno statističko istraživanje. Drugi nedostaci o kojima treba voditi računa što svi ispitanici nemaju internetsku vezu i može se dogoditi da će se mrežnom anketom izgubiti pozornost ispitanika, povećavajući mogućnosti dobivanja nekvalitetnih ili nepotpunih statističkih podataka.

Računalno potpomognuto anketiranje (CAPI) jest metoda prikupljanja podataka u kojoj anketar koristi tablet, mobitel ili računalo za bilježenje odgovora danih tijekom intervjua.

Prednosti metode CAPI jesu u tome što olakšavaju provjeru logičkih postavki, preskoke i provjeru smislenosti tijekom intervjua. Ovi postupci čine anketu učinkovitijom i pomažu u osiguravanju više razine kvalitete podataka. Također omogućuju uštedu vremena koja se odnosi na uređivanje podataka. CAPI je dobar alat za praćenje popisivača. Može automatski zabilježiti vrijeme početka, vrijeme završetka i GPS lokaciju intervjua, olakšavajući kontrolorima provjeru je li popisivač zaista obavio određeni intervju, uspoređujući svoje vrijeme i GPS podatke s vremenima ostalih intervjua na kojima je bila prisutna osoba zadužena za nadzor i kontrolu procesa. Ovdje treba voditi računa da je na nekim mjestima loš signal ili ga uopće nema stoga i da u ovakvim slučajevima računalo ne može očitati GPS lokaciju. S obzirom na čvrstu vezu s Wi-Fi mrežom ili podacima, podaci prikupljeni anketama CAPI odmah se prenose u elektronički format. To omogućuje kontrolorima koji se nalaze u područnim jedinicama da provjere određeni dio podataka prema uputama koje im je dao metodolog. Nedostatak ankete CAPI može se odnositi na teško pristupačna područja jer se CAPI uglavnom oslanja na dostupnost električne energije ili na podatkovnu povezanost, što možda nije izvedivo u nekim dijelovima Republike Hrvatske. Ako se podaci prikupljeni metodom CAPI oštete prije nego što se prenesu na poslužitelj, zauvijek će se izgubiti. Baterije i izvanmrežna pohrana podataka zasigurno mogu olakšati provedbu ankete CAPI u regijama bez dostupne električne energije ili podatkovne povezanosti. Anketar ima uputu da prije odlaska na takva mjesta mora provjeriti je li baterija na prijenosnom računalu napunjena i nakon povratka u područje u kojem ima internetsku vezu treba poslati podatke. Inače se svi podaci najprije spremaju lokalno i anketar mora sam pokrenuti slanje/preuzimanje podataka. Zasad nije postavljeno kao automatska funkcija. S nekim ispitanicima nije ugodno provesti intervju jer nemaju povjerenja, a to onda može rezultirati nepouzdanim podacima ili da određeni skup takvih ispitanika odbije sudjelovati u statističkom istraživanju. Ankete CAPI zahtijevaju tehnološki kompetentne popisivače ili dulja razdoblja obuke za popisivače. To može činiti problem anketama provedenima u hitnim situacijama ili u posljednji trenutak. Ako neki anketar odustane od anketiranja, njegove ankete prosljeđuju se drugom anketaru dok se ne pronađe zamjena.

Računalno potpomognuto telefonsko anketiranje (CATI) jest telefonsko prikupljanje podataka. Anketar telefonira na broj koji je prikazan na ekranu. Zatim, ako se ispitanik želi pridružiti anketi, anketar slijedi skriptu na ekranu odabirući odgovore. Alat CATI za istraživanje automatski će prijeći na sljedeće pitanje slijedeći logiku postavljenog upitnika. Na kraju upitnika anketar će prikazati novog ispitanika na poziv. Primjerice, sustav CATI nudi sljedeći broj koji anketar može nazvati.

Sustav upravlja kontaktima u skladu s pravilima koja su postavili administratori. Na primjer, kontakti sa zauzetoj linijom ponovno će se prikazati nakon 15 minuta. Aplikacija se može podesiti na intervale koji odgovaraju organizatoru ankete. Trenutačno je postavljena tako da u jednom danu zauzetu liniju može nazvati najviše šest puta. Inače, na početku dana osoba zadužena za nadzor i kontrolu procesa odabire broj anketa koje bi htio odraditi s anketarima koji su na raspolaganju, tzv. sustav *daybatch* odabire zadani broj anketa iz učitanoj uzorka. Ako se kontakt javi na poziv, ali nije dostupan za razgovor, anketar također može zakazati sastanak. U tom će slučaju softver CATI automatski ponovno prikazati taj kontakt na datum i vrijeme koji su predviđeni.

Prednosti metode CATI jesu dobivanje točnih odgovora te pouzdanih informacija o ispitanicima, postizanje anketiranja gotovo cijele populacije, precizno upravljanje uzorcima i stabilnost odgovora, lako izračunavanje vremena, izbjegavanje intervjua izvan kvote, ostvarivanje potpune kontrole napretka ankete, postojanje povratnih informacija u stvarnom vremenu, prisutnost anketara obeshrabruje odustajanje od intervjua, intervjui mogu biti i dulji jer je stopa odustajanja niža, omogućeno je potpuno zadovoljstvo ispitanika, dobro su prihvaćene osjetljive teme, telefonski anketar može pomoći u razumijevanju upitnika i nema neodgovara na postavljena pitanja. Ispitanik može na određena pitanja odgovoriti s "ne znam" ili "može odbiti odgovor".

Nedostaci metode CATI jesu visoki troškovi za telefonske anketere, troškovi PC stanica (trošak je izražen kod inicijalnog uspostavljanja centra CATI), troškovi brojača (softver sâm broji pozive i dodjeljuje im status), softvera i hardvera broj anketara izravno je proporcionalan dnevno obavljenim intervjuima. Nedostatak je i cijena telefonskih linija te otežano angažiranje novih anketara koje potom treba obučiti za rad u centru CATI i na svakom istraživanju koje se provodi.

Upute za osiguranje kvalitete

Kod optičkog čitanja razina pouzdanosti podrazumijeva da je vjerojatnost čitanja stopostotna odnosno da je jedan znak točno pročitano. Zahtjev za većom pouzdanosti omogućuje bolju kvalitetu čitanja, ali istodobno usporava čitanje. Osim toga, oni znakovi koji nisu prepoznati, tj. oni koji nisu zadovoljili razinu pouzdanosti, dolaze naknadno operatoru na ispravak, a ti ispravci mogu ozbiljno usporiti cijeli proces. Razina pouzdanosti postavlja se vrlo visoko za sva identifikacijska pitanja i za prvo slovo u svakoj riječi. Najniža razina zahtjevnija je za sve ostale numeričke odgovore. Uz postavljene uvjete izračunava se srednja brzina čitanja za prosječno ispunjeni upitnik svake vrste.

Ocjena optičkog čitanja i ocjena zajedničkog utjecaja optičkog čitanja i automatskog šifriranja provodi se na uzorku popisnih podataka. Za kontrolu unosa podataka korišten je uzorak s frakcijom 0.006. Primijenjeni su različiti omjeri uzorka, čime se omogućuje odgovarajući odnos za svaku vrstu jedinice uzorka unutar populacije iz koje se uzorak uzima.

Uzorak je planiran za ocjenjivanje:

- odstupanja popisnog materijala pročitana optičkim čitačem od ručno unesenoga
- odstupanja popisnog materijala pročitana optičkim čitačem od ispravno, odnosno točno unesenoga.

Prvi tip ocjenjivanja omogućuje ocjenu ili usporedbu optičkog čitanja u odnosu na današnju raspoloživu tehnologiju koja se koristi za unos podataka. Druga ocjena omogućuje mjerenje pouzdanosti i provjeru smislenosti. Svako razlikovanje među znakovima "unosa" i odgovarajućih znakova iz skupa "čitanje" označeno je kao "sumnjivo" na posebnom izvještaju. Do tog razlikovanja moglo je doći i zbog pogreške u ručnom unosu te je zbog toga posebna skupina izrađivala "autentični" skup podataka i prolazila kroz upitnike koji su bili označeni kao "sumnjivi". Bitno je napomenuti da oni nisu mogli vidjeti optički pročitane podatke uzorka. Ta je skupina provjeravala svaki sumnjivi znak i, ako je bilo potrebno, ispravljala ih da bi odgovarali onima na upitniku. Postupak usporedbe i ispravljanja ponavljao se sve dok svi znakovi označeni kao "sumnjivi" nisu odgovarali onima u upitniku. Konačni rezultat tog procesa bio je skup točno unesenih podataka koji se može nazvati "autentični" uzorak.

Nakon opsežnog posla u dobivanju "autentičnog" uzorka uočena su različita neslaganja, tj. odstupanja između autentičnog uzorka i onoga koji je bio nazvan "čitanje":

- razlikovanje na razini pozicije (jednostavno nabranje slučajeva u kojima je optički čitač pročitao jedno, a na tome mjestu u upitniku piše nešto drugo)
- razlikovanje na razini polja (pri čemu se pod poljem podrazumijeva skup znakova koji odgovaraju dogovoru na određeno pitanje).

Analiza automatskog šifriranja pokazala je da je ono bilo obavljeno na očekivanoj visokoj razini uspješnosti. Visoka kvaliteta algoritma za automatsko kodiranje rezultirala je visokim postotkom automatskog šifriranja varijabli.

Statistička istraživanja provedena metodom CAWI omogućuju praćenje kvalitete i pouzdanosti podataka. Dok je istraživanje u tijeku, preporučljivo je provjeriti cjelovite i djelomično popunjene upitnike. Time se može provjeriti mogu li složena (ili duga) pitanja prouzročiti veću stopu odustajanja ili ispitanika navesti da brzopletno i nesuvislo popuni upitnik. Kvaliteta upitnika može se procijeniti i tijekom postupka analize podataka. Bitna informacija jest i razina zadovoljstva ispitanika. Umetanjem ocjenjivačkog pitanja na kraju intervjua ili uspoređivanjem broja ispunjenih upitnika s brojem ispitanika koji nisu do kraja popunili upitnik omogućuje se

praćenje i otklanjanje nedostataka upitnika, što je presudno za dobivanje pouzdanih podataka za visokokvalitetnu analizu. Kako bi se uspostavilo učinkovito mrežno istraživanje, potrebno je uspostaviti sustav koji će omogućiti postizanje izvrsnih rezultata uz niske troškove.

Kod statističkih istraživanja provedenih metodom CAPI kvaliteta podataka prati se tijekom cijelog razdoblja prikupljanja podataka. Prikupljene informacije o kvaliteti obuhvaćaju i podatke povezane s ostalim popratnim procesima, kao što su vrijeme trajanja putovanja, vrijeme provedeno na licu mjesta i broj obavljenih / u tijeku / odbijenih intervjua.

Izvještaj o kvaliteti podataka CAPI pripremljeno je i pregledano. Izvještaj je uključivao, na primjer, za svakog anketara broj obavljenih intervjua, nepotpune podatke po stavkama iz statističkog istraživanja, iskrivljenost ili slučajeve u kojima su se odgovori ankete znatno razlikovali za jednog anketara u usporedbi s drugim anketarima unutar određene prostorne jedinice.

Statistička istraživanja koja se provode metodom CATI omogućuju brzo dosezanje ukupne ciljne populacije te individualni pristup ispitanicima i prilagodljivost. Prikupljeni podaci ovom se metodom cijelo vrijeme provjeravaju, čime su stvoreni preduvjeti da se odgovarajuća razina kvalitete održava i unapređuje u skladu s postavljenim statističkim standardima.

Kako bi se utvrdila razina kvalitete intervjuiranja, primjena metode CATI ima prednosti jer omogućuje praćenje anketara i, ako je potrebno, omogućuje unapređenja u poslovnom procesu. Metoda CATI nadzor pruža korisne informacije u vezi s anketarom, operativnim upravljanjem, obukom te razvojem i dizajniranjem instrumenata za prikupljanje podataka.

5. Obrada podataka

Podaci koji su osnova za izračun statističkih procjena u provedbi statističkih istraživanja mogu se preuzeti na različite načine ili iz različitih izvora podataka. Oni se prema načinu prikupljanja podataka mogu podijeliti na primarne i sekundarne izvore. Primarni izvori jesu oni podaci koji su primarno prikupljeni za svrhu statističkog istraživanja, a sekundarni su oni podaci koji su primarno prikupljeni za neke druge svrhe, npr. administrativne, a Državni zavod za statistiku ih je preuzeo za statističke svrhe. Ako se u anketi koriste podaci iz nekoliko različitih izvora podataka, najprije ih je potrebno pravilno povezati, a zatim statistički pravilno obraditi.

Statistička obrada podataka odnosi se na sve postupke kojima se podaci prikupljeni u statističkom istraživanju obrađuju i uređuju u obliku prikladnome za objavljivanje i diseminaciju, primjenjujući odgovarajuće statističke metodologije. Proces uključuje i obradu podataka na mikrorazini, kao što su uklanjanje pogrešaka i nepotpunih vrijednosti, matematički postupci na populaciji (ponderiranje, agregiranje) te obrada već prikupljenih podataka, kao što su sezonska prilagodba i zaštita podataka. U procesu statističke obrade podataka primjenjuju se više ili manje napredne ili zahtjevne statističke metode, čija je svrha prije svega omogućiti izračun točnih i nepristranih statističkih podataka. Svi dostupni izvori podataka trebaju biti sustavno povezani. Na podacima trebaju biti utvrđene i ispravljene pogreške. Nakon toga je potrebno definirati i primijeniti odgovarajući statistički model i definirati matematički postupak za procjenu populacije.

Zahvaljujući razvoju statističkih metoda i naprednim informacijskim tehnologijama, najviše se unaprijedila statistička obrada podataka od svih procesa provedbe statističkih istraživanja jer je postala brža i učinkovitija. Uspostavljeni su automatizirani potproces koji omogućuju brze i učinkovite postupke. Koliko god automatizirani postupci imaju svoje prednosti imaju i nedostatke u obliku tzv. "crne kutije", kada se postupci odvijaju bez ikakve mogućnosti utjecaja statističara. Kako bi se ublažili negativni učinci takve provedbe, obradu statističkih podataka trebalo bi pažljivo planirati i detaljno dokumentirati. Ponajprije bi trebalo omogućiti pripremu potrebnih metapodatkovnih procesa, poput pružanja uvida statističaru u automatizirane postupke.

5.1. Integriranje prikupljenih podataka

Opis

Obuhvaća integriranje podataka iz jednoga ili više izvora tijekom obrade podataka. U statističkim istraživanjima koja primjenjuju nekoliko različitih izvora podataka, isti bi se izvori podataka trebali objediniti i pripremiti u formatu koji omogućuje daljnju statističku obradu. Ako u izvorima podataka postoje jedinstveni identifikatori, agregiranje podataka provodi se izravno preko njih. Inače je potrebno definirati i pripremiti kombiniranje zapisa putem drugih parametara, tzv. neizravno povezivanje. Kako bi se pojednostavnili postupci, pretpostavka je da u procesu agregiranja uvijek postoji referentni skup jedinica promatranja koje su ujedno polazna točka za integracijske procese. Skup podataka koji proizlazi iz procesa integracije trebao bi sadržavati sve jedinice iz referentnog izvora podataka.

Izravna veza. Povezivanje zapisa iz različitih izvora podataka odvija se s pomoću jedinstvenog identifikatora (npr. OIB-a, MB-a ili MBS-a tvrtke) koji su sastavni dio obaju izvora podataka. Zapisi iz pridruženog izvora koje nije bilo moguće povezati s referentnim izvorom podataka s pomoću jedinstvenog identifikatora registrira se na zasebnom popisu. Proces izravnog povezivanja također je grafički prikazan.

Neizravna veza. Ako u referentnome i pridruženom izvoru podataka ne postoji zajednički jedinstveni identifikator, uspostavlja se povezivanje postupcima neizravnog povezivanja. U ovom slučaju pomaže se s drugim odabranim zajedničkim varijablama koje postoje u oba izvora. Jedinice referentnog izvora podataka u ovom su se slučaju "podijelile" na dva dijela.

U prvom su dijelu jedinice koje su jednoobrazno povezane s izvorom s pomoću neizravnih konektora. Za ove jedinice vrijednosti neizravnih konektora u oba izvora točno se podudaraju i u pridruženom izvoru te postoji samo jedan takav zapis.

U drugom dijelu postoje jedinice koje nije bilo moguće jednoobrazno povezati s izvorom jer svaki zapis iz referentnog izvora ima pridružene višestruke zapise iz pridruženog izvora. U ovom su slučaju pridruženi zapisi u kojima su vrijednosti neizravnih konektora u oba izvora najbliži.

Sličnost je potrebno definirati jednoobrazno. Uobičajeno je da se za tu svrhu koristi jedna od već postojećih funkcija u različitim programskim okružjima za numeričku procjenu podudaranja nizova.

Za daljnju statističku obradu u drugom skupu podataka potrebno je izabrati samo jedan zapis. Hoće li se taj odabir izvršiti ručno ili programski na temelju dogovorenog algoritma, mora se odlučiti za svaku anketu zasebno.

Kombiniranjem različitih izvora podataka preuzimaju se različite vrijednosti za istu varijablu. U takvim slučajevima potrebno je odrediti prioritet prikupljanja podataka i dodati novu varijablu s pokazateljem za tu varijablu, koja za svaku jedinicu prikazuje koji je izvor podataka određen kao prioritet pri utvrđivanju pouzdane vrijednosti varijable.

Upute za osiguranje kvalitete

Prije određivanja postupka nositelj ankete treba pažljivo ispitati dostupne administrativne izvore podataka i njihovu primjenjivost u određenome statističkom istraživanju. Kada se utvrđuje koje će se varijable preuzeti iz administrativnih izvora podataka, potrebno je ispitati ispunjavaju li varijable postavljene statističke kriterije u predmetnom izvoru podataka i hoće li podaci iz administrativnih izvora podataka biti pravodobno dostupni, tj. dovoljno rano za objavljivanje i diseminaciju statističkih rezultata.

Ako se za istu varijablu koristi nekoliko različitih izvora podataka, potrebno je utvrditi jesu li podaci iz tih različitih izvora međusobno smisleni. Za postavljanje modela podataka odgovara metodolog (uzorkaš), međutim, u cijeli poslovni proces trebaju biti uključeni opći metodolog i nositelj statističkog istraživanja jer se samo tako može osigurati primjena modela u provedbi statističkog procesa.

Ako se koriste podaci iz izvora podataka koji su osjetljivi s obzirom na prirodu osobnih podataka, moraju se organizirati postupci koji će u što većoj mjeri spriječiti zloupotrebu pristupa takvim podacima. Jedinstveni administrativni identifikatori moraju se stoga zamijeniti alternativnim statističkim identifikatorima prije procesa integracije podataka koji se kasnije koriste za provedbu statističkog postupka.

Ako se koriste postupci neizravnog povezivanja, oni trebaju biti postavljeni tako da udovoljavaju i zahtjevima potpunog obuhvata i zahtjevima relevantnosti i točnosti. Ponajprije treba izbjegavati manualne postupke jer bi to moglo ugroziti pravodobnost i poštovanje rokova za objavljivanje statističkih rezultata.

Na kraju provedenog postupka agregiranja podataka svaka jedinica promatranja u bazi treba imati odgovarajući status jedinice i svaka varijabla treba imati odgovarajući status varijable. Ako su vrijednosti određene varijable izvedene iz nekoliko različitih izvora, mora se voditi računa za svaku vrijednost iz kojeg je izvora podataka preuzeta.

Za svaku varijablu povezanu s popisom šifri potrebno je točno odrediti koja je inačica popisa šifri prikladna prema valjanosti podataka. Tada se ova verzija popisa šifara koristi za kontrolu vrijednosti varijable. Za sve numeričke varijable potrebno je pažljivo ispitati njihova obilježja i zatim odrediti skup prihvatljivih vrijednosti. U većini slučajeva potrebno je odlučiti hoće li biti dopuštene vrijednosti nula ili čak negativne vrijednosti.

Sve logičko-računske kontrole koje se koriste u ovom dijelu postupka potrebno je dokumentirati. Ako je postavljena vrijednost određenih varijabli ograničena, potrebno je opisati razloge postavljenog ograničenja. Postavljene logičko-računske kontrole trebaju biti detaljno dokumentirane.

5.2. Kontrola, uređivanje i ispravljanje podataka

Opis

Kontrola uključuje logičko-računsku kontrolu podataka prema zadanim pravilima kontrole, kontrolu agregiranih podataka ili skupine jedinica, netipičnih vrijednosti (outliera) i kritičnih vrijednosti. Uređivanje i ispravljanje uključuje automatsko uređivanje podataka ili aktiviranje upozorenja da treba ručno pregledati i ispraviti podatke. Kontrola, uređivanje i ispravljanje podataka može se provoditi više puta dok podaci ne dostignu zadovoljavajući stupanj kvalitete. Ovisno o razini primijenjenih podataka, postupci uređivanja podataka mogu se u grubo podijeliti na uređivanje podataka na mikrorazini i uređivanje podataka na makrorazini. Pri uređivanju podataka na mikrorazini provode se postupci na razini podataka pojedinih jedinica. Postoji i nekoliko postupaka za uređivanje podataka na mikrorazini, ovisno o tome kako su podaci dobiveni.

Pri terenskoj ili telefonskoj provedbi anketa primjenjuje se potpuni ili djelomični skup logičkih kontrola. Po završetku tako provedene ankete podaci se kombiniraju s dostupnim sekundarnim izvorima podataka u zajedničku datoteku. Nakon toga slijedi dodatna provjera podataka, a sve utvrđene pogreške ispravljaju se postupcima ručnoga ili automatskog ispravljanja.

U slučaju samopopunjavanja (tiskanoga ili elektroničkog upitnika) logičke se kontrole izvršavaju nakon završavanja unosa podataka. Provjere cjelokupnog skupa podataka provode se istodobno, automatskim ispravicima i generiranjem statistike pogrešaka. Otkrivene pogreške ispravljaju se ručnim ili automatskim postupcima korekcije.

Ako se u istraživanju koristi kombinacija nekoliko različitih izvora podataka (npr. statističkih i administrativnih), svaki se izvor podataka najprije treba urediti zasebno. Nakon toga mora se kontrolirati kombinirani skup podataka. Treba naglasiti kako je pogreške otkrivene u objedinjenoj datoteci podataka potrebno uklanjati na svakome pojedinačnom izvoru podataka (a ne u objedinjenoj datoteci) jer je to jedini način da se kod sljedećih prikupljanja i obrade podataka postigne održivost postupaka na dugi rok.

Budući da je uređivanje podataka vremenski dugotrajan postupak, potrebno ga je što je moguće više automatizirati. Računalni bi programi trebali biti dizajnirani tako da istodobno omogućuju utvrđivanje i ispravljanje pogrešaka.

Automatske se korekcije mogu ugrubo podijeliti u dvije skupine: determinističke i vjerojatnosne. Prve su korekcije određene jednostavnim determinističkim pravilima koja se mogu zapisati u obliku logičko-aritmetičkih izraza (npr. $AKO X > 10 \text{ I } Y < 1 \text{ ONDA } Y = 1$). Ostale se korekcije određuju na temelju vjerojatnosnih postupaka, obično na temelju minimalne promjene prijavljenih podataka (Fellegi-Holtov pristup⁸).

Upute za postizanje kvalitete

Logičke kontrole moraju se definirati i testirati u probnom programu. Treba izbjegavati praksu postavljanja velikog broja kontrola, s prestrogim kriterijima, u želji da se podaci prikupljeni istraživanjem što više "pročiste". Takva praksa dovodi do tzv. prekomjernog vrednovanja, a posljedica takve regulacije jest da dobivamo previše jedinica čije vrijednosti treba ponovno provjeriti.

Potrebno je istražiti mogu li se uvesti postupci automatskog uređivanja podataka jer bi se time moglo znatno smanjiti troškovi, skratiti vrijeme provedbe statističkog istraživanja i povećati učinkovitost. Prije nego što se takvi postupci uvedu u redoviti postupak, treba detaljno ispitati jesu li u sadržajnome i tehničkom smislu smisljeni i izvedivi.

⁸ Odnosi se na pretpostavke i uređivanje te ciljeve imputiranja koje su iznijeli Fellegi i Holth u svom radu objavljenome 1976. u časopisu "Journal of American Statistical Association".

Svi postupci za uređivanje podataka moraju biti detaljno i uredno dokumentirani. U povremenim anketama potrebno je omogućiti da se rezultati uređivanja podataka koriste za poboljšanje kvalitete. Uređivanje podataka trebalo bi biti organizirano tako da omogući održivost postupaka. U praktičnom smislu to znači da se moraju osigurati postupci uređivanja podataka koji se koriste istim ulaznim podacima i da iste metode uvijek daju isti krajnji rezultat.

Za sve se promjene na podacima treba voditi "povijest". Drugim riječima, izvorni podaci ne smiju se prekrivati ispravcima tijekom faze uređivanja. Međutim, treba pripremiti novu verziju podataka ako su podaci ispravljani. Svaki ispravljani podatak treba sadržavati pripadajuće metapodatke koji pružaju informacije o tome gdje se podatak nalazi u procesu i zašto su podaci ispravljani.

5.3. Imputiranje i utežavanje

Opis

Pojam imputiranja podataka odnosi se na sve postupke u kojima se nepotpune ili netočne vrijednosti utvrđene u procesu uređivanja podataka zamjenjuju statističkim procjenama. Pritom je potrebno razlikovati postupke u kojima se vrijednosti zamjenjuju statističkim procjenama i postupke u kojima se ispravljene vrijednosti dobivaju ponovnim uspostavljanjem kontakta s izvještajnim jedinicama. O umetanju podataka govori se isključivo u prvom slučaju.

Postupci imputiranja podataka trebali bi pomoći u poboljšanju statističkih procjena na ukupnoj razini. Kako bi se postigao taj cilj, metode umetanja podataka moraju se odabrati pažljivo i razborito. Te se metode moraju odabrati tako da imputirane vrijednosti budu što bliže valjanim, vjerodostojnim i pouzdanim vrijednostima, kako bi se osigurali dosljedni podaci (prema zadanom skupu logičkih kontrola) i što je više moguće održala osnovna distribucija dobivenih podataka.

Metode imputiranja dijele se na:

- **deduktivne metode** – metode imputacije u kojima se odluka o vrijednosti imputiranja donosi na temelju poznatih informacija o ispitaniku
- **determinističke metode** – metode imputacije kojima se jedinicama s istim određenim obilježjima uvijek imputira ista vrijednost
- **stohastičke metode** – metode imputacije kojima se jedinicama s istim određenim obilježjima u pravilu imputiraju različite vrijednosti.

U primjeni su različite metode umetanja podataka, a u nastavku teksta navodimo one koje se najčešće koriste:

Metoda logičkog imputiranja (deduktivna metoda) – njome se unosi vrijednost koja iz podataka logično slijedi, a riječ je o podatku koji se nije moglo prikupiti za predmetnu jedinicu. Ako je, npr. za pojedinog ispitanika poznat njegov datum rođenja, ali ne i njegova dob, može se za određenu vremensku točku izračunati dob koja nedostaje.

Metoda imputiranja prosječne vrijednosti (deterministička metoda) – njome se zamjenjuje vrijednost koja nedostaje prosječnom vrijednošću podataka onih jedinica čijim podacima raspolažemo. Prosječne vrijednosti obično se ne izračunavaju iz cijelog skupa podataka, nego iz podataka za područje u kojem se nalazi jedinica za koju se unose podaci (npr. podaci za ispitanike određene dobi).

Metoda unutarnjeg darivatelja (metoda hot deck; stohastička metoda) – u nju se ubacuje vrijednost koja se uzima od druge jedinice (darivatelja), a za koju imamo podatke. Darivatelj se može na temelju funkcije pomoćnih i drugih proučavanih varijabli precizno odrediti ili nasumično odabrati. Ovisno o provedbi metode, vrijednost darivatelja se ili preuzme ili prilagodi (npr. za omjer vrijednosti pomoćne varijable jedinice za koju se umeću podaci i darivatelja).

Metoda vanjskog darivatelja (stohastička metoda) – njome umećemo vrijednost koju uzimamo iz nekoga vanjskog izvora, npr. iz istog istraživanja za prethodno referentno razdoblje (za periodična istraživanja) iz drugog istraživanja ili iz administrativnih izvora podataka. Kao i kod metode unutarnjeg darivatelja vrijednost se može jednostavno preuzeti ili prilagoditi (npr. za koeficijent rasta pomoćne varijable).

Način dionica odnosno strukture podataka (deterministička metoda) koristi se kada je poznat zbroj i kada je potrebno umetnuti pojedine komponente tog zbroja. Potrebno je umetnuti vrijednost svake komponente množenjem zbroja s nekim razlomkom. Taj se udio može izračunati na više načina (ovisno o metodi provedbe), ali važno je da zbroj udjela svih sastavnica bude jednak 1.

Metoda regresije – u nju se unose vrijednosti koje se izračunavaju iz odgovarajućega regresijskog modela. Najjednostavniji regresijski model jest model linearne regresije. Ako se, npr. iza izračun vrijednosti varijable Y koriste dvije pomoćne varijable (X, Z), matematička jednadžba takva modela jest sljedeća: $Y = \alpha \cdot X + \beta \cdot Z + \varepsilon$.

Bez slučajne komponente (reziduala) ε metoda regresije jest deterministička metoda imputacije, a dodamo li modelu slučajnu komponentu, dobijemo stohastičku metodu.

Možemo reći da se **determinističkim metodama** umetnuta vrijednost izračunava nekim analitičkim postupkom u kojem se koristi odgovarajuća deterministička funkcija, a **stohastičkim metodama** umetnuta vrijednost izračunava se postupkom koji se koristi mehanizmom vjerojatnosti.

Kako bi pružatelji statističkih istraživanja imali bolji uvid u utjecaj postupaka umetanja podataka koji se koriste na statističke rezultate, potrebno je omogućiti izračun odgovarajućih pokazatelja kvalitete u procesu. Dva ključna pokazatelja kvalitete za postupak umetanja podataka jesu udio umetnutih podataka i utjecaj umetanja / unosa podataka na statistiku.

Pokazatelj udjela umetnutih podataka izračunava se za ključne varijable, kao omjer broja jedinica za koje su podaci umetnuti i odnose se na određenu varijablu i na broj svih jedinica koje bi trebale imati podatke za ovu varijablu.

Utjecaj umetanja podataka na pokazatelj statistike izračunava se za ključne statistike, kao (relativna) razlika između izračunane statistike prije postupka umetanja podataka i statistike izračunane prema ovom postupku.

U većini se anketa podaci o željenim obilježjima populacije dobivaju samo od dijela te populacije (čak i ako je riječ o popisu, tj. o cjelovitom promatranju, uobičajeno je da odgovore ne dostavljaju sve jedinice); to znači da je potrebno prilagoditi postupke za izračunavanje statistika i uzeti u obzir činjenicu da nema podataka za varijable od interesa za cijelu populaciju. Ako je riječ o vjerojatnosnom uzorku, koriste se formule za izračunavanje statističkih procjena. Jedan od najčešćih načina svođenja na cijelu populaciju jest utežavanje. Pri utežavanju se svaki prikupljeni podatak pomnoži s nekim pozitivnim faktorom većim od 1 ili jednakim 1. Ti se faktori nazivaju težine i označuju se s w_k , pri čemu indeks k označuje da ta težina pripada k -toj izvještajnoj jedinici, ako su izvještajne jedinice označene s 1,2 ... n .

Ako primjenjujemo vjerojatnosne obrasce, težine su zapravo određene nacrtom uzorkovanja. Kao što je već spomenuto, nacrtom uzorkovanja određena je za svaku jedinicu iz okvira uzorkovanja vjerojatnost izbora u uzorak različita od nule, koju označujemo s π_k . Najčešće korištene težine, koje se određuju za jedinice u uzorku, jesu recipročne vrijednosti vjerojatnosti izbora (ako zbroj populacije procjenjujemo procjeniteljem, tada ga u ovom slučaju nazivamo Horvitz-Thompsonovim procjeniteljem), tj.

$$w_k = \frac{1}{\pi_k}$$

Takvo izračunavanje težina vrlo je prikladno jer je jednostavno, težine su poznate prije postupka uzorkovanja (pod pretpostavkom da se prikupe odgovori od svih jedinica koje su uključene u uzorak), a teorija vjerojatnosnog uzorkovanja omogućuje da su formule korištene za izračun preciznosti procjena jednostavne, barem za najčešće korištene procjenitelje (prosjeck, zbroj, udio). Takav izbor težina istodobno omogućuje nepristranost procjene.

U praksi se pokušava poboljšati ovaj najjednostavniji postupak izračunavanja težina (i time povećati preciznost procjena) upotrebom tzv. proporcionalnog procjenitelja, kod kojeg težine w_k pomnožimo s nekim faktorom (naziva se proporcionalna korekcija ili kalibracijska korekcija), koji se izračunava na temelju podataka za pomoćnu varijablu koja je u dobroj korelaciji (linearni odnos čiji grafikon prolazi kroz ishodište koordinatnog sustava) s varijablom od interesa. Varijabla čije podatke znamo prije provođenja istraživanja (tj. imamo ih u okviru uzorkovanja) ili čija je vrijednost za ukupnu populaciju poznata, a pojedinačni podaci dobiveni su na uzorku, naziva se pomoćnom varijablom. Ako je pomoćna varijabla (označena kao x) u dovoljnoj korelaciji (najmanje 0,5) s varijablom od interesa, gornji faktor izračunava se dijeljenjem totala pomoćne varijable za cijelu populaciju (t_x) s uteženim zbrojem vrijednosti pomoćne varijable za jedinice u uzorku (x_k), pri čemu su nam težine recipročne vrijednosti vjerojatnosti izbora $\frac{1}{\pi_k}$

$$\text{popravak} = \frac{t_x}{\sum_{k \in S} \frac{1}{\pi_k} x_k}$$

Težina proporcionalnog procjenitelja jest

$$w_{k,\text{prop}} = \text{popravak} \cdot w_k$$

Težine proporcionalnog procjenitelja vrlo se često koriste u poslovnim anketama jer se u ovim anketama koriste pomoćne varijable iz administrativnih izvora podataka i drugih izvora (broj zaposlenih, promet tvrtke, plaće, investicije itd.).

U anketama u kojima se promatraju pojedinci i kućanstva često se koristi slična tehnika koja se naziva kalibracija. Cilj je kalibracije promijeniti težine uzorka tako da kada ih se koristi za utežavanje podataka za neku pomoćnu varijablu, dobit će se točna vrijednost pomoćne varijable za populaciju. Postupak izračunavanja kalibracijskih težina potpuno je isti kao i kod težina proporcionalnog procjenitelja. Često je namjera izvršiti kalibraciju na nekoliko pomoćnih varijabli (npr. spol i skupine prema dobi). Poseban primjer kalibracijskih metoda jest poststratifikacija. U ovom su postupku poznate vrijednosti populacije pomoćnih varijabli i za neke subpopulacije koje nismo planirali nacrtom uzorka (npr. jer nismo imali dovoljno podataka u okviru uzorkovanja). U poststratifikaciji se preračunavaju težine za svaku takvu subpopulaciju posebno, tako da su procjene za varijablu od interesa jednake njezinoj točnoj vrijednosti u svim subpopulacijama.

Ako se istraživanje ne provodi na vjerojatnosnom uzorku, tada se za određivanje težina koristi modeli. Najveći problem s tim je što se pri korištenju modela uglavnom teško može procijeniti točnost takvih procjena.

Na kraju je primjer nedostatka podataka zbog neodgovora jedinica. U praksi pri provođenju pojedinačnih statističkih istraživanja jedinice obuhvaćene statističkim istraživanjem nikada potpuno ne odgovaraju na upitnik. Dakle, postoji potreba za dodatnim utežavanjem jedinica koje su dostavile odgovor jer je u ovom slučaju potrebno prvo preračunati težine jedinica u uzorku koje su odgovorile na uzorak, a zatim uzorak na ukupnu populaciju okvira za izbor uzorka. Kako bi se pristupilo rješavanju ovog izazova, potrebno je pažljivo analizirati podatke koji su nepotpuni zbog neodgovora jedinice tako da se utvrdi postoje li razlike između odgovora i neodgovora. Najčešći postupci za ispravljanje težina uzoraka zbog neodgovora ispitanika jesu pretpostavka dvoetapnog plana uzorkovanja (model) i uporaba kalibracijskih metoda. U slučaju dvoetapnog plana uzorkovanja pretpostavlja se da je u prvoj etapi odabran stvarni uzorak jedinica, dok u drugoj etapi imamo jedinice koje su odgovorile i moguće neprikladne jedinice. Vjerojatnost izbora jedinice koja je odgovorila jest umnožak vjerojatnosti prve i druge etape, dok je težina recipročna vrijednost ovog produkta. Naravno, ove se težine koriste samo na podacima iz odgovora, pod pretpostavkom da su od interesa samo statistike o područjima relevantnih jedinica.

Upute za osiguranje kvalitete

Podatke koji nedostaju kao rezultat neodaziva varijable uvijek bi trebalo zamijeniti odgovarajućom statističkom procjenom. Održavanje nepotpunih vrijednosti u konačnom skupu podataka može, s jedne strane dovesti do pristranosti u rezultatima, a s druge strane može otežati provođenje složenijih statističkih analiza.

U slučaju da su podaci nepotpuni zbog neodgovora jedinice, metode umetanja trebaju se koristiti samo za mali broj varijabli ili ako postoje dobri razlozi da upotreba ovih metoda može znatno smanjiti pristranost. Inače, izračun težina neodgovora trebao bi se koristiti za podatke koji su nepotpuni zbog neodgovora jedinice.

Prije odabira metoda umetanja podataka treba izvesti precizne i dovoljno opsežne simulacijske studije kojima se može procijeniti odabir najprihvatljivije metode umetanja podataka. Provođenjem simulacijskih studija trebalo bi se izvršiti sljedeće:

- kreirati mehanizam stvaranja podataka koji nedostaju i koji će najbolje odgovarati pretpostavljenoj stvarnosti provedenog istraživanja
- pri testiranju metoda koristiti se svim pomoćnim podacima dostupnima tijekom statističke obrade podataka
- testirati metodu procjenjivanja koja najbolje čuva statistička svojstva rezultata na makrorazini, a ne metodu koja najbolje predviđa vrijednosti na mikrorazini.

Svaki umetnuti podatak mora biti označen u statističkom procesu s pomoću odgovarajućeg statusa varijable jer se time pokazuje koja se metoda umetanja koristila. U svakoj fazi postupka treba postojati mogućnost razdvajanja prikupljenih podataka od podataka koji su rezultat statističkih procjena.

Postojeća osnovna softverska rješenja potrebno je koristiti što je više moguće pri pripremi softvera za izvođenje postupaka umetanja podataka.

Potrebno je osigurati izračun pokazatelja kvalitete koji pružaju uvid u automatizirane postupke umetanja. Treba izračunati i pokazatelje na mikrorazini i makrorazini. Osobito treba paziti na udio umetnutih podataka i utjecaj umetanja podataka na konačne rezultate.

Relevantnost primijenjenih metoda potrebno je u analizi stalno provjeravati. Ako analiza pokaže da primijenjene metode nisu (više) primjerene, potrebno je postupak prilagoditi.

Ako je moguće, istraživanje se provodi na vjerojatnosnom uzorku jer se samo na temelju toga određuju težine koje daju vjerodostojne podatke o točnosti procjena. Kod utežavanja posebnu pozornost treba obratiti na netipične vrijednosti (outliere), tj. na jedinice koje imaju vrlo veliku (ili vrlo malu) vrijednost analizirane varijable u odnosu na ostale jedinice u uzorku. Ako je opravdana pretpostavka da je takva jedinica također izdvojena na razini cijele populacije (ili pripadajućeg stratuma), mijenja joj se težina tako da se težina smanjuje ili poprima vrijednost 1.

Nakon izračuna konačnih težinskih vrijednosti potrebno ih je zbrojiti za kontrolu jer njihov zbroj mora biti barem približno jednak broju jedinica u populaciji (ili u područjima populacije).

Kada se koristi proporcionalni procjenitelj ili bilo koja druga metoda kalibracije, preporučuje se da se ispravljene težine ne razlikuju mnogo od težine uzorka.

Ako je stopa neodgovora vrlo visoka, preporučuje se provesti dodatna istraživanja na uzorku jedinica bez odgovora i ponovno izračunati težine uzorka na temelju analize ovog uzorka.

Ako nisu dostupne pomoćne informacije za analizu neodgovora, tada se konačne težine u kojima se uzima u obzir odgovor izračunavaju koristeći se pretpostavkom dvoetapnog plana uzorkovanja. Ako su dostupne pomoćne informacije koje su u dobroj korelaciji s varijablom od interesa, onda se neodgovor može analizirati. Rezultati analize pomažu pri određivanju skupina u kojima se jedinice slično "ponašaju" te se na tim skupinama provodi pretpostavka dvoetapnog nacrtu uzorkovanja ili se odgovarajućom metodom uklanja utjecaj neodgovora.

U poslovnim anketama Državni zavod za statistiku raspolaže popisom statusa za sve jedinice koje se provode na uzorku koje se mogu grupirati prema statusu odgovora, statusu neodgovora i statusu za neprikladne jedinice. Na kraju istraživanja opći metodolog trebao bi dobiti popis u koji su uključene sve jedinice uzorka i u kojem svaka jedinica ima odgovarajući status.

Ako se promatrana jedinica doda u adresar nakon izbora jedinica, to se mora zabilježiti, kao i razlozi za to.

Ako se prijavi jedinica da pripada skupini koja nije dio stratumu u kojem je izabrana, a riječ je o slučaju kada je skupina i dalje dio promatrane populacije (npr. drugi stratum), tada se težine obično izračunavaju na razini stratumu u kojem je jedinica izabrana, dok se statistički objavljuje na razini skupine u kojoj je registrirano postojanje takve jedinice.

5.4. Proizvodnja izvedenih varijabli

Opis

Proizvodnja izvedenih varijabli i statističkih jedinica koje nisu izričito prikupljane tijekom prikupljanja podataka obavlja se primjenom aritmetičkih formula na jednu ili više postojećih varijabli. Nove statističke jedinice mogu se izvesti agregiranjem ili dijeljenjem podataka za jedinice prikupljanja ili prema različitim metodama procjene (npr. kreiranje kućanstava, kada su jedinice u prikupljanju podataka osobe ili poduzeća ili su jedinice pravne osobe).

Deflacioniranje se u statističkom procesu odnosi na postupak kojim se utjecaj promjena cijena u određenom razdoblju isključuje iz vrijednosnih podataka. Na teoretskoj se razini uvijek može zamisliti da se podaci o vrijednosti sastoje od dviju komponenata: količine (ili volumena) i cijene. Pri izračunavanju omjera podataka o vrijednosti u dvije vremenske točke izračunani omjer uključuje promjenu količine i cijene. Ako na temelju izmjerene promjene vrijednosti želimo dobiti najtočniju procjenu promjene količine, utjecaj promjene cijena mora se ukloniti iz omjera vrijednosti (indeksa). Indeks izračunan iz "izvornih" pojedinačnih podataka naziva se nominalni indeks, a indeks izračunan iz podataka iz kojih smo izuzeli utjecaj promjena cijena naziva se stvarni indeks.

Postoje dva osnovna pristupa izvođenju postupka deflacioniranja. Prvi je proces deflacioniranja na mikrorazini, a drugi je proces deflacioniranja na makrorazini. U deflacioniranju na mikrorazini najprije se dijeli svaki podatak na pojedinačnoj razini odgovarajućim deflatorom i tako se (u smislu promjena cijena) preračunavaju podaci u istu (fiktivnu) vremensku točku. Na primjer, ako se indeks potrošačkih cijena koristi na fiksnoj osnovi, prosjek godine, kao deflator 2005., tada svaki preračunani pojedinačni podatak o vrijednosti predstavlja vrijednost koja bi (s tada važećim cijenama) vrijedila u fiktivnome vremenskom trenutku u prosječnoj godini.

U slučaju deflacioniranja na makrorazini provode se postupci "uklanjanja" cijena na već izračunanim agregatima ili indeksima. Razlikuje se od izračuna na mikrorazini uglavnom po tome što je ovdje potreban deflator za isto razdoblje kao i indeks koji se deflacionira. Ako se deflacionira indeks u tekućem mjesecu u odnosu na prethodni mjesec, potreban je i deflator (tj. odgovarajući indeks cijena) za tekući mjesec u odnosu na prethodni mjesec, ali ako se deflacionira indeks u fiksnoj baznoj godini, potreban je i deflator u fiksnoj baznoj godini (npr. prosjek 2005.). U postupku deflacioniranja na mikrorazini potreban je samo jedan vremenski niz deflatora, a u deflacioniranju na makrorazini potrebno je onoliko vremenskih vrsta deflatora koliko postoji vremenskih serija koje želimo deflacionirati. To je ujedno i najveća prednost deflacioniranja mikrorazine. Prednost deflacioniranja na makrorazini jest u većoj postojanosti procesa jer je deflacija na makrorazini manje zahtjevnost za obradu zbog manjeg broja podataka koji se obrađuju.

Upute za osiguranje kvalitete

Pri odluci hoće li se koristiti postupci deflacioniranja na mikrorazini ili makrorazini moraju se uzeti u obzir sve prednosti i nedostaci obaju pristupa. Ponajprije je potrebno odabrati pristup koji će omogućiti jednostavnije i racionalnije smještanje postupka u cjelokupni statistički postupak koji sadržajno najbolje odgovara promatranoj pojavi.

Pri određivanju postupaka za pripremu deflatora prvi je korak provjeriti koji su indeksi cijena najprikladniji kao osnovni podaci za pripremu deflatora. U drugom koraku potrebno je provjeriti je li indeks cijena već prikladan deflator ili ga treba kombinirati s drugim strukturnim podacima. Cilj je izračunani deflator približiti pojavi stvarnoga kretanja cijena koji se mjeri na temelju vrijednosnih podataka.

Potrebno je osigurati najjednostavnije i, ako je moguće, automatizirano generiranje indeksa cijena. Također je potrebno osigurati ograničeni pristup.

Prikladnost korištenog deflatora mora se neprestano provjeravati. Informacije u tome mogu biti vrlo korisne za područje nacionalnih računa.

5.5. Izračunavanje agregata

Opis

Iz mikropodataka se izračunavaju agregirani podaci i ukupni rezultati (npr. agregiranje podataka prema ekonomskim, društvenim, geografskim i drugim klasifikacijama). Sumiraju se podaci sa zajedničkim obilježjima, određuju se mjere prosjeka i mjere disperzije te se primjenjuju težine dobivene u potprocesu imputiranja, utežavanja i procjenjivanja totala. Ako se primjenjuje metoda slučajnog uzorka, izračunavaju se standardne pogreške.

Pojmom izračun statističkih procjena ili agregiranje podataka označuje se onaj dio statističkog postupka u kojem se izračunavaju procjene iz konačnih mikropodataka, a nazivaju se i statistikama. Budući da se u većini statističkih istraživanja ne promatra cijela populacija, nego samo njezin dio (slučajni uzorak), u statistički izračun također je uključen izračun vrijednosti populacije ili izračunavanje procjena populacije. Proces ili funkcija kojom se iz podataka na uzorku računaju procjene populacije naziva se procjenitelj.

Općenito, postoji više različitih vrsta agregata i procjenitelja, ali u praksi Državnog zavoda za statistiku prevladava nekoliko najvažnijih agregata. Dat ćemo opis najčešće korištenih procjenitelja. Pretpostavlja se da je za svaku jedinicu uzorka već izračunana odgovarajuća težina populacije (w_i) i da je postupak utežavanja izveden tako da je ispoštovan nacrt uzorka. Nadalje se pretpostavlja da su $\{y_i, i = 1, \dots, n\}$ na uzorku izmjerene vrijednosti ciljne varijable Y . Procjenu za populaciju označujemo s \hat{Y} .

Procjena totala populacije. Procjenu ukupnog zbroja varijable Y na populaciji izračunamo prema formuli

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

Prosjeak populacije. Procjenu prosjeka varijable Y za populaciju izračunavamo prema formuli

$$\hat{\bar{Y}} = \frac{\sum w_i y_i}{\sum w_i}$$

Broj jedinica s određenim obilježjima. Neka bude D_i varijabla koja za svaku promatrana jedinicu poprima dvije vrijednosti: vrijednost 1, ako jedinica ima promatrana obilježja, i vrijednost 0 ako jedinica nema ovih obilježja

$$\hat{Y} = \sum w_i D_i$$

Udio jedinica s određenim obilježjima. Udio u populaciji jedinica s određenim obilježjima procjenjuje prema formuli

$$\hat{p} = \frac{\sum w_i D_i}{\sum w_i}$$

Omjer totala populacije. Omjer totala populacije za dvije varijable (X, Y) procjenjuje se formulom

$$\hat{R} = \frac{\sum w_i y_i}{\sum w_i x_i}$$

Ako se procjena vrijednosti populacije za odabrane statistike izračuna na temelju podataka dobivenih iz uzorka, takva procjena također sadržava uzoračku pogrešku.

U slučaju vjerojatnosnog uzorka, ova se pogreška mora procijeniti i prikazati na odgovarajući način kada se rezultati objave. Postupci za izračunavanje pogreške postupka uzorkovanja mogu biti prilično složeni. Stoga se određuju vrstom procjenitelja i nacrtom uzorka koji se koristi. Ugrubo govoreći, postupci za procjenu uzoračke pogreške mogu se podijeliti u tri skupine:

- primjena izravnih formula za procjenu standardnih pogrešaka linearnih procjenitelja i upotreba "približnih formula" Taylorova linearizacija za nelinearne procjenitelje (analitički pristup)
- procjena standardne pogreške na temelju odabira poduzorka i zatim primjena odgovarajućih formula za takvu primjenu (ponovno uzorkovanje)
- primjena regresijskog modela za procjenu standardnih pogrešaka procjene na podskupinama razmatrane populacije.

Postoji i nekoliko načina prezentiranja procijenjenih uzoračkih pogrešaka u publikacijama. U praksi Državnog zavoda za statistiku koriste se dvije metode: pogreška uzorka eksplicitno se (vrijednosno) objavljuje zajedno s procijenjenim statistikama ili se manje precizne procjene označuju posebnim oznakama.

Upute za osiguranje kvalitete

Pri izračunu procjena populacije na temelju uzorka trebaju se odabrati procjenitelji koji su barem približno nepristrani.

Svakako je potrebno izračunati procjene i izračunati uzoračke pogreške navedenih procjena u istome statističkom procesu. Nositelj ankete zajedno s procijenjenim statistikama dobiva i uzoračku pogrešku i u skladu s tim ocjenjuje koliko su pouzdane procijenjene statistike.

Upotrijebljeni nacrt uzorka mora se uvijek uzeti u obzir pri pripremi postupka za izračunavanje uzoračke pogreške. Pri pripremi uputa za prikazivanje uzoračke pogreške potrebno je uzeti u obzir usvojene standarde u Državnom zavodu za statistiku. Uzoračke pogreške potrebno je redovito analizirati, uglavnom zbog izračunanih procjena s nedovoljnom preciznošću kad je prevelika uzoračka pogreška. Ako analize pokažu da izračunane procjene nisu dovoljno precizne, treba uvesti mjere za smanjenje uzoračkih pogrešaka. Takve mjere mogu rezultirati ili povećanjem uzorka ako dostupni resursi to dopuštaju ili definiranjem učinkovitijeg procjenitelja koristeći se pomoćnim varijablama.

5.6. Izrada konačnih datoteka podataka

Opis

Objedinjuju se rezultati svih potprocesa u ovoj fazi, što rezultira konačnom datotekom podataka koja se koristi kao input za fazu analize. Katkad je to prijelazna, a ne konačna datoteka, osobito za statistička istraživanja koja su vrlo važna i postoji potreba za proizvodnjom preliminarnih i konačnih procjena.

Pojam uređivanja podataka na makrorazini u najužem smislu riječi znači utvrđivanje pogrešaka na već prikupljenim podacima. Takvi se postupci u procesu provođenja statističkog istraživanja postavljaju u fazu uređivanja podataka na mikrorazini. U pravilu se samo utvrđivanje pogrešaka događa na makrorazini, dok se ispravljanje pogrešaka uvijek treba izvoditi na mikrorazini. Razlikuje se više metoda za uređivanje podataka na makrorazini, a u nastavku je pregled metoda:

Metoda kontrole agregata. Osnovna ideja metode jest prvo provjeriti utežene i agregirane podatke (npr. na četveroznamenkastoj razini NKD-a), a zatim na mikrorazini kontrolirati sve jedinice iz skupina koje su upitne na makrorazini. Definicija kontrola na makrorazini uglavnom se temelji na raspodjeli omjera i razlika prema rezultatima iz prethodnog razdoblja.

Metoda postupnog izlučivanja. Osnovna ideja na kojoj se temelji ova metoda jest da se s pomoću odgovarajuće računalne aplikacije ograničava kontrola i uređivanje podataka samo na one jedinice čiji utjecaj na konačnu ocjenu u određenoj podskupini nije zanemariv. Postupak s pomoću odgovarajuće aplikacije prikazuje jedinice čiji utjecaj na agregate nije zanemariv. Nakon toga se u postupcima koji slijede kontroliraju vrijednosti tih jedinica.

Grafičke metode. Svrha primjene grafičkih metoda jest grafički prikazati raspodjelu uteženih vrijednosti podataka koje je potrebno kontrolirati. Grafički prikaz distribucije i ključnih parametara raspodjele (npr. kvantili) može olakšati određivanje granica ekstremnih vrijednosti. Grafičke metode uglavnom se koriste kao "komplementarni" alat kod provedbe drugih metoda uređivanja podataka.

U procesu tabeliranja strukturirani skupni podaci pripremaju se za različite vrste publikacija za elektroničke publikacije, bazu podataka Državnog zavoda za statistiku, tiskane publikacije, tablice koje udovoljavaju zahtjevima standardnih korisnika i za tablice namijenjene međunarodnom izvještavanju. Pripremljene tablice mogu se koristiti i u drugim potprocesima, osim u kontrolnim tablicama: u uređivanju na makrorazini, u statističkoj zaštiti podataka, u analizi prikladnosti i u validaciji rezultata. Važno je da su tablice za javno objavljivanje rezultat primjene unaprijed pripremljenih računalnih postupaka koji se mogu izvesti automatski ili na način koji kontrolira nositelj istraživanja. To povećava mogućnost za organiziranje i pohranu svih

izračunanih procjenitelja u jednu makrobazu, smanjuje rizik od vremenskih ograničenja koja ugrožavaju pravodobnost objavljivanja rezultata i u velikoj mjeri uklanja utjecaj ljudskog faktora na točnost rezultata.

Tablice moraju biti dizajnirane optimalno prema sadržaju, tehnologiji pripreme podataka, tehničkim mogućnostima alata i prihvaćenim standardima objavljivanja podataka. Postupak pripreme tablica može se provesti samostalno, ali većim dijelom ima smisla kombinirati ga s nekim drugim procesima, posebno sa statističkom zaštitom podataka i izračunavanjem statističkih procjena.

U potprocesu detaljno se definiraju rezultati koje smo okvirno odredili u procesu analize potrebnih zahtjeva.

U ovom je potprocesu općenito potrebno uzeti u obzir sljedeće: vrstu jedinica promatranja i obilježja koja određuju populaciju, geografski položaj jedinica koje su uključene u populaciju i referentno razdoblje za koje se primjenjuju važna obilježja populacije.

Za realizaciju navedenog potrebno je što je više moguće koristiti se usklađenim pojmovima, varijablama i klasifikacijama, a u nekim slučajevima navedene se definicije moraju prilagoditi specifičnim potrebama istraživanja. Sve definicije pojmova, varijabli i klasifikacija trebaju biti dokumentirane, kao i sva odstupanja od standarda.

Osim toga, potrebno je pripremiti detaljan popis svih varijabli koje se želi obuhvatiti statističkim istraživanjima. Za svaku varijablu potrebno je označiti sljedeće: izvor podataka, odnosno je li izvor podataka statistički upitnik ili administrativni izvor podataka; referentno razdoblje za podatke o upisu; informacije o tome kada je traženi izvor podataka dostupan.

Upute za osiguranje kvalitete

Skupne kontrole trebaju biti pažljivo definirane. Konkretno, treba izbjegavati situacije prekomjernoga i neučinkovitog upravljanja podacima na mikrorazini zbog preostalih kontrola na makrorazini.

Izvedivost kontrola treba ispitati prije upotrebe u redovitom postupku. Za testne podatke mogu se koristiti podaci iz prošlih statističkih istraživanja ili podaci iz srodnih statističkih istraživanja. Preporučuje se upotreba grafičkih metoda jer je vizualna prezentacija obično najučinkovitiji način percepcije "sumnjivih" vrijednosti. Uređivanje podataka na makrorazini treba se koristiti samo za utvrđivanje netočnih ili upitnih vrijednosti.

Sva ispravljanja podataka treba izvršiti na razini pojedinačnog zapisa. Postupci bi trebali biti što je više moguće automatizirani jer se time poboljšava učinkovitost cjelokupnoga statističkog procesa.

Za izradu tablica treba se koristiti standardnim uputama. Ako se kod izrade tablica koriste popisi šifara, uvijek se trebaju koristiti standardni popisi šifara koje se nalaze na klasifikacijskom poslužitelju (KLASUS). Tablice za javno objavljivanje trebaju se formirati na temelju unaprijed pripremljenih postupaka.

Za pripremu internih ad hoc i kontrolnih tablica nositelji statističkih istraživanja trebaju se koristiti standardnim interaktivnim analitičkim alatima. Budući da ovaj proces objedinjuje različite procese i obuhvaća različite ustrojstvene jedinice, potrebno je uspostaviti dobru koordinaciju te pravilno dokumentirati radne postupke na standardiziran način.

Kod postavljanja definicija koriste se standardni koncepti, varijable i klasifikacije. Svi pojmovi, varijable i klasifikacije koji se koriste u statističkom istraživanju propisuju se detaljno, dok se sva moguća odstupanja od standardnih pojmova opisuju odvojeno. Isto tako, opisuju se razlike između potencijalne pojave koja se prati i pojave važne za korisnike statističkih podataka, npr. razlika između ciljne⁹ i promatrane populacije¹⁰.

⁹ Ciljna populacija (Target population) – skup jedinica na koje se istraživanje odnosi i na koju se poopćavaju (generaliziraju) dobiveni rezultati

¹⁰ Promatrana populacija – popis jedinica osnovnog skupa iz kojeg se izabire uzorak, npr. registri poslovnih subjekata

Pri određivanju klasifikacija koriste se klasifikacije pohranjene u klasifikacijskom poslužitelju KLASUS, kao što su nacionalne te klasifikacije međunarodnih organizacija poput Eurostata, Ujedinjenih naroda, OECD-a, Svjetske banke i Međunarodne organizacije rada.

Klasifikacije bi trebale biti koncipirane tako da omoguće izvještajnim jedinicama jednostavno razvrstavanje pojave tako da istodobno dopuštaju objavljivanje i diseminaciju podataka na dovoljno detaljnoj razini bez ikakvih ograničenja s obzirom na odredbe povjerljivosti i točnosti podataka. Primjenom hijerarhijskih klasifikacija, odnosno objavljivanjem podataka na višim hijerarhijskim razinama može se zadovoljiti navedeni uvjet. Što se tiče procjenjivanja kvalitete rezultata, potrebno je imati dostupne barem osnovne pokazatelje kvalitete koji su rezultat provedbe statističkih istraživanja.

5.7. Proizvodnja i ažuriranje statističkih registara i baza podataka

Opis

Statistički registri jesu kontinuirano ili redovito ažurirani skupovi objekata određene populacije. Oni sadržavaju inoformacije potrebne za identifikaciju i pristup određenoj populaciji jedinica kao i druga obilježja koja su potrebna za provođenje statističkih istraživanja populacije. Statistički registri najčešće sadržavaju trenutačno i povijesno stanje populacije te uzroke, posljedice i izvore promjena podataka o populacijama statističkih jedinica. Integriraju podatke iz više statističkih i administrativnih izvora, povezujući ih s pomoću zajedničkih identifikatora, a pohranjeni su u strukturiranim bazama podataka.

Statistički registri imaju koordinacijsku ulogu u provedbi statističkih istraživanja, a njihova ažurnost i kvaliteta imaju veliki utjecaj na kvalitetu istraživanja i statističkih proizvoda.

Među mnogim funkcijama statističkih registara sljedeće su napose važne: 1) identifikacija i konstrukcija statističkih jedinica, 2) nedvosmisleno bilježenje veza među jedinicama, 3) evidencija populacija, 4) evidencija glavnih obilježja jedinica; bilježenje klasifikacije jedinica, 5) izgradnja okvira za uzorkovanje, tj. uspostavljanje zajedničkog okvira za organiziranje i koordinaciju statističkih istraživanja osiguravanjem usklađenog okvira uzorkovanja prema obilježjima stratifikacije, 6) pružanje osnove za agregiranje rezultata dobivenih iz istraživanja na temelju uzoraka i 7) praćenje stvaranja i zatvaranja te drugih demografskih događaja jedinica.

Statistički registri jesu okosnica organizacije statistike, a neki (kao što je SPR) temelje se na pravnim dokumentima.

Državni zavod za statistiku uspostavio je za potrebe službene statistike statističke registre koristeći se podacima iz administrativnih izvora podataka, popisa, statističkih istraživanja i podataka prikupljenih metodom promatranja i praćenja.

Uspostavljeni statistički registri jesu SPR, SRPG i PSR.

SPR je registar koji sadržava podatke o poslovnim subjektima (pravnim i fizičkim osobama), njihovim dijelovima i grupacijama, koji su institucionalno i formalno uključeni u proizvodne i financijske procese nacionalnoga gospodarstva. Poslovni subjekti kao administrativne jedinice pretvaraju se u statističke jedinice pogodne za statističko praćenje i analizu kako bi se, među ostalim, postigla usporedivost s drugim zemljama članicama EU-a. Najvažnija statistička jedinica jest poduzeće kao najmanja kombinacija pravnih jedinica koja čini organizacijsku jedinicu za proizvodnju roba i usluga. Poduzeće je krajnji rezultat metode profiliranja kojom se analizira pravna, operativna i računovodstvena struktura grupa poduzeća na nacionalnoj i svjetskoj razini kako bi se stvorile statističke jedinice radi najučinkovitijeg prikupljanja statističkih podataka.

Svrha SPR-a jest omogućiti učinkovito prikupljanje podataka o poslovnim subjektima, njihovu obradu i analizu, koju provode ustrojstvene jedinice Državnog zavoda za statistiku usmjerene na praćenje poslovnih subjekata te drugi nositelji službene statistike.

SPR je glavni izvor podataka za statistiku poslovne demografije jer promatra demografske događaje vezane za poduzeća.

Glavni izvori podataka za ažuriranje SPR-a jesu administrativni i statistički izvori, kao što su Europski registar grupa poduzeća te statistička istraživanja Državnog zavoda za statistiku.

SRPG je registar koji sadržava podatke o jedinicama koje se bave poljoprivrednom proizvodnjom: pravne osobe i njihove dijelove, obrtnike i obiteljska poljoprivredna gospodarstva. SRPG sastavljen je iz osnovnoga i statističkog dijela. Osnovni dio sadržava osnovne podatke o poljoprivrednom gospodarstvu (identifikacijski podaci, podaci o vlasniku / nazivu poslovnog subjekta, adresni podaci i drugo).

Statistički dio sadržava podatke iz područja poljoprivredne statistike (poljoprivredne površine po kategorijama, broj stoke i drugo). Svrha SRPG-a jest osigurati provođenje istraživanja primjenom pouzdanog okvira za izbor uzoraka za statistička istraživanja u području poljoprivrede, ispis i ažuriranje adresnih podataka za različita istraživanja u statistici poljoprivrede i bolji obuhvat poljoprivrednih jedinica u statistici poljoprivrede. Glavni izvori podataka za ažuriranje SRPG-a jesu administrativni izvori, SPR i redovita statistička istraživanja iz područja poljoprivrede.

PSR je registar koji sadržava podatke o prostornim jedinicama za statistiku 1., 2. i 3. razine Nacionalne klasifikacije prostornih jedinica za statistiku, županijama, gradovima/općinama, jedinicama mjesne samouprave, naseljima, statističkim krugovima, popisnim krugovima, ulicama/trgovima, kućnim brojevima, kućanstvima i drugo. Ustrojen je kao službena osnova za prikupljanje, evidentiranje, iskazivanje, razmjenjivanje i povezivanje različitih vrsta prostornih podataka iz područja službene statistike Republike Hrvatske. Podaci sadržani u PSR-u služe kao tehnička podloga za aktivnosti službene statistike.

Navedene podatke PSR vodi u alfanumeričkome i grafičkom obliku. PSR je izvor podataka za statističke registre i statističke aktivnosti koje se koriste podacima o prostornim jedinicama. Podaci se redovito ažuriraju. Načini prikupljanja podataka mogu biti on-line pristup, elektronički medij i tiskani oblik. Glavni izvor administrativnih podataka jest Državna geodetska uprava, a glavni izvor statističkih podataka jesu istraživanja Državnog zavoda za statistiku.

KLASUS je alat za sve korisnike klasifikacija koji na jednome mjestu nudi mogućnost pregledavanja i pretraživanja klasifikacija prema nazivu i šifri, omogućuje prikazivanje pojedinačnih elemenata klasifikacija, njihove nazive, objašnjenja ili indekse, ako postoje, kao i tablica veza među različitim verzijama klasifikacija te pruža mogućnost preuzimanja klasifikacija u više formata sa svim razinama i elementima klasifikacije. Klasifikacije su grupirane unutar "obitelji" koje obuhvaćaju na određeni način povezane klasifikacije, a "prečice" omogućuju izravan pristup najčešće traženim klasifikacijama.

Ova aplikacija omogućuje tabeliranje i pregledavanje podataka koji su opisani u središnjoj bazi metapodataka, dok je u aplikaciji *Survey Processor* dopušteno tabeliranje tih podataka. Korisnici mogu sami izrađivati jednostavne tablice i spremati ih u više različitih formata, a agregirane podatke mogu spremati i kao nove izvore podataka za tabeliranje. U *Warehouse Browseru* mogu se koristiti i tablice pripremljene u *Survey Processoru* čiji se rezultati kreiraju u Excelu na temelju upita izravno iz podataka.

Statističko skladište podataka obuhvaća podatke spremne za tabeliranje koji se mogu podijeliti na mikropodatke, makropodatke i pripremljene podatke.

Makropodaci sadržavaju kocke. Kocka je sinonim za agregirane ili makropodatke koji su dobiveni obradom mikropodataka, a može biti generirana i iz drugih makropodataka. Najčešće se zbrajanjem kvantitativnih varijabli u mikropodacima po kategorijama koje pripadaju odabranim kvalitativnim varijablama dobivaju agregirane vrijednosti u kockama. Pripadajuće kategorije kvalitativnih varijabli nazivaju se dimenzije kocke.

Tablice (pripremljeni podaci) sadržavaju tablice agregiranih mikropodataka, ali se u njoj nalazi više agregiranih razina podataka, za razliku od kocke, u kojoj je samo jedna jedina razina. To znači da u tablici mogu biti prikazani i sumarni redci za neke dimenzije.

Upute za osiguranje kvalitete

Statistički registri trebaju biti relevantni za potrebe svojih korisnika podataka. Trebaju sadržavati relevantne jedinice i varijable te omogućiti odabir i generiranje relevantne populacije za provođenje postupka uzorkovanja za statistička istraživanja. Točnost je bitno obilježje kvalitete statističkih registara, a pripada najlakše mjerljivim načelima kvalitete.

Pravodobnost i poštovanje rokova objavljivanja te usklađenost i usporedivost također su bitna načela kvalitete za većinu korisnika statističkih registara. Jednostavnost pristupa podacima iz registara smatra se i komponentom kvalitete kao i tumačenje informacija koje su pohranjene u registru. Što se tiče usporedivosti, postoje dva aspekta usporedivosti, i to: usporedivost u prostoru i usporedivost u vremenu. Tako je, primjerice, usporedivost za statističke poslovne registre u EU propisana Uredbom o poslovnim statistikama. Stoga je u ovom smislu jedna od mjera kvalitete razina usklađenosti s važećim propisima. Ovakva vrsta usporedivosti vrlo je važna jer su svi statistički registri od vitalne važnosti za mnoga statistička područja. Usporedivi registri olakšavaju proces harmonizacije statistike izvedene iz takvih registara. Usporedivost jedinica i klasifikacija koje se koriste u statističkim poslovnim registrima također je jedan od postavljenih ciljeva europske statističke legislative. Usporedivost u vremenu možda je manje važna za mnoge korisnike, ali ipak postoji potreba za uvidom u podatke i jedinice kao i za usporedivosti ukupnih podataka u različitim razdobljima.

Na primjeru SPR-a usklađenost se može promatrati u smislu interne usklađenosti i usklađenosti s drugim registrima. Interna povezanost odnosi se na dosljednu evidenciju podataka u registru, kao što je dosljedno korištenje pravila ažuriranja i profiliranja. Povezanost s drugim registrima najbolje se održava u čuvanju i korištenju jedinstvenih identifikatora. Ovo omogućuje uspoređivanje podataka i jedinica među registrima, npr. unutar Republike Hrvatske: OIB, MB ili MBS u SPR-u omogućuje relativno lako povezivanje s administrativnim registrima (npr. Administrativni poslovni registar, Registar poreznih obveznika, Sudski registar, Obrtni registar i drugo). Za povezivanje SPR-a s EuroGroup Registrom koristi se jedinstveni međunarodni identifikator, tzv. LEID ID. Korištenje jedinstvenog identifikatora u svim poslovnim registrima (administrativnim i statističkim) jedan je od načina da se postigne bolja povezanost. Kada se i postigne ovakva povezanost, ona ima i ograničenja jer se definicije pohranjenih jedinica u registrima katkad razlikuju.

Za statistički registar može se reći da je potpun ako uključuje sve jedinice iz ciljane populacije i sve potrebne varijable. U stvarnosti je nemoguće postići potpuni obuhvat i zbog nemogućnosti pravodobne isporuke podataka. Stoga bi potpunost ipak trebala biti cilj, a njezino mjerenje prikazat će se u pokazateljima kvalitete.

6. Analiza

Primjenom raznih alata i tehnika provodi se statistička analiza podataka. Cilj provedenih analiza jest objasniti situacije i događaje, izdvojiti određena pravila koja su karakteristična za promatranu pojavu te ukratko protumačiti dobivene rezultate. Podaci se analiziraju na makrorazini, a kada takva analiza nije dovoljno precizna, onda se provodi analiza i na mikrorazini.

Analiza se provodi kako bi se potvrdila njihova smislenost ili otkrili mogući nedostaci. Otklanjanjem nedostataka postiže se bolja kvaliteta podataka. Ako analiza pokaže sustavne nedostatke, izračunani podaci koriste se za poboljšanje kvalitete odnosno za dopunu poslovnog procesa ili promjenu metodologije. Za ispravno tumačenje podataka provodi se sustavna analiza te se na konkretnim primjerima predstavljaju situacije iz prakse. Ovakvim se načinom rada mogu utvrditi određene statističke zakonitosti koje se sustavno javljaju u promatranom području.

Na makrorazini mogu se analizirati podaci koji se odnose na određeno vremensku točku ili dulje razdoblje. Ako se u analizi na makrorazini utvrde znatna odstupanja, te ih na toj razini nije moguće objasniti, provodi se analiza na mikrorazini jer se iz mikropodataka izvode rezultati na makrorazini.

Analiza podataka također obuhvaća analizu utjecaja postupaka koji se provode u statističkoj obradi podataka. Na primjer, analizom se provjerava utjecaj neodgovora i umetanja podataka na konačne rezultate te se provjerava koliki je bio utjecaj odabranog postupka umetanja podataka, zatim razlike između privremenih i konačnih podataka i slično. Sve su navedene informacije osnova za pripremu opisa kvalitete podataka, a osnova su i za dopunu i poboljšanje metodologije istraživanja. Za analizu i provjeru podataka koriste se, između ostalog, i drugi izvori podataka i informacije koji se izravno ili neizravno odnose na predmetno područje. Prije tumačenja statističke pojave analizom podataka potrebno je utvrditi što podaci pokazuju. U ovoj fazi poslovnog procesa statističar se mora zapitati što podaci prikazuju i odgovoriti na postavljena pitanja analizom i tumačenjem podataka.

6.1. Statistička analiza rezultata

Opis

Prikupljeni podaci transformiraju se u statističke rezultate. Statistička analiza rezultata uključuje proizvodnju dodatnih mjera, poput indeksa, trendova ili vremenski prilagođenih serija kao i navođenje kvalitativnih obilježja. Kako bi se povećala vrijednost i stvorili preduvjeti za analize statističkih podataka, u ovaj se potproces mogu uključiti i priprema karata, GIS rezultata te geostatističkih usluga.

Vremenske serije podataka odnose se na vremenske nizove podataka, u našem slučaju prikupljene statističke podatke za određena područja istraživanja (npr. industrijska proizvodnja, promet u trgovini na malo, indeksi troška rada i slično). Glavni dio analize vremenskih serija odnosi se na sezonsku prilagodbu i određivanje trenda, a manji dio odnosi se na prognoziranje. Kod sezonske prilagodbe uklanja se utjecaj sezonske i kalendarske komponente iz vremenskih serija, kada su te komponente izražene i važne. Tako dobivene vrijednosti nazivaju se sezonski prilagođene vrijednosti ili vrijednosti s isključenim sezonskim i kalendarskim učincima. Ovaj je postupak korisno provesti kad god je potrebno usporediti podatke iz različitih razdoblja istoga vremenskog niza ili podatke za isto razdoblje istoimenoga vremenskog niza iz različitih država jer su oni obično podložni utjecaju sezonskih oscilacija, broju radnih, odnosno trgovačkih dana i drugim utjecajima. Kada se uspoređuju podaci za isto razdoblje različitih godina u istome vremenskom nizu, provjerava se jesu li kalendarski učinci prisutni u vremenskom nizu i onda ih se isključuje jer je sezona za isto razdoblje u godini približno ista (npr. usporedba travnja 2010. s travnjem 2009.). Ovako prilagođeni podaci nazivaju se podaci s isključenim kalendarskim učincima ili podaci prilagođeni učinku broja radnih, odnosno trgovačkih dana.

Utjecaj broja radnih ili trgovačkih dana

Utjecaj praznika

Utjecaj Uskrsa

Utjecaj prijestupne godine

Za originalne (izvorne) vremenske serije (X_t) izrađuje se model sezonske prilagodbe iz kojeg se zatim određuje sezonska komponenta (S_t)

iregularna ili slučajna komponenta (I_t) i komponenta trend-ciklus (TC_t).

$$X_t = TC_t + S_t + I_t$$

Trend-ciklička komponenta sadržava dugoročnu tendenciju razvoja pojave, dok iregularna (slučajna) komponenta sadržava slučajne učinke, odnosno rezidualna odstupanja od trenda nakon uklanjanja ostalih komponenata. Modeli se u pravilu revidiraju jedanput godišnje, najčešće s prvim ili posljednjim podacima u kalendarskoj godini. Tada nositelj statističkog istraživanja dostavlja metodologu za vremenske serije sve datoteke s parametrima modela i podacima na kontrolu, tj. godišnji pregled, odnosno reviziju modela.

Postoje sljedeće vrste netipičnih vrijednosti:

- **aditivne netipične vrijednosti** (abnormalne vrijednosti u izoliranim točkama serije)
- **privremene promjene** (niz netipičnih vrijednosti s vremenski opadajućim učincima na razinu serije)
- **pomaci razine** (niz inovacijskih netipičnih vrijednosti s konstantnim dugoročnim učinkom na razinu serije, gdje se pod inovacijskim netipičnim vrijednostima podrazumijevaju netipične vrijednosti u inovacijskoj seriji)
- tzv. **ramp**, koji opisuje glatki, linearni ili kvadratični prijelaz između dviju vremenskih točaka, za razliku od nagle promjene povezane s pomakom razine
- **privremeni pomaci razine**, u kojima pomak razine ima kratkoročni, a ne dugoročni učinak.

Ako se netipična vrijednost pojavljuje na kraju vremenske serije, tada je to privremena netipična vrijednost i metodolog ga fiksira tek kad zaprimi barem sljedeća tri podatka u vremenskoj seriji jer tek tada može odrediti vrstu netipične vrijednosti. U tom slučaju svaki put kad se dodaju novi podaci, statističar treba dostaviti vremenske serije na kontrolu metodologu za vremenske serije dok se netipična vrijednost ne ustali.

Duljina vremenske serije koja je vremenski prilagođena mora biti najmanje tri godine za mjesečne vremenske serije i najmanje četiri godine za tromjesečne vremenske serije. Postojanje kalendarskih utjecaja (jedan ili dva regresora) utvrđuje se u vremenskim serijama koje imaju podatke za najmanje pet godina, a postojanje nekoliko regresora (šest ili sedam) samo u vremenskim serijama koje imaju podatke za najmanje sedam godina.

Upute za osiguranje kvalitete

Nositelj istraživanja trebao bi pažljivo pregledati i analizirati vremenske serije prije nego što ih preda metodologu na modeliranje jer se time može poboljšati kvaliteta modela vremenskih serija.

Nositelj statističkog istraživanja mora obavijestiti metodologa vremenskih serija o svim promjenama koje se odnose na vremenske serije, kao što su promjene u uzorkovanju, promjene u metodologiji, promjene na tržištu i slično.

Metodolog za vremenske serije trebao bi se posavjetovati s nositeljem statističkog istraživanja o važnosti rezultata sezonske prilagodbe (postojanje kalendarskih utjecaja, uzroci netipičnih vrijednosti). Usporedbe i prikazi sezonski prilagođenih serija nalaze se u publikacijama, npr. na godišnjoj razini (primjer: travanj 2010. u odnosu na travanj 2009.) kao i podaci o kalendarskoj prilagodbi, tj. o utjecaju broja radnih dana u pojedinim vremenskim serijama. Komponenta trend-ciklusa može se prikazati samo grafički i pritom treba obratiti pozornost na problem krajnjih točaka promatranja.

Kod dugih vremenskih serija, koje su dulje od 12 godina, katkad ih je potrebno skratiti, npr. na sedam godina kako bi se poboljšala kvaliteta modela. Naime, ako se osobitosti sezone mijenjaju u tako dugom razdoblju, rezultati sezonskog prilagođavanja mogu imati lošu kvalitetu. Podaci koji nisu uključeni u model sezonske prilagodbe, npr. stariji od sedam godina, dostupni su samo kao izvorni (originalni) podaci.

Bitno je da nositelj istraživanja ne zaboravi na godišnju provjeru modela i njihovu eventualnu reviziju i da jedanput godišnje metodologu za vremenske serije dostavi na kontrolu sve modele vremenskih serija i podatke.

6.2. Kontrola kvalitete rezultata

Opis

Provjerava se kvaliteta proizvedenih rezultata, u skladu s općim okvirom kvalitete i očekivanjima.

Analiza kvalitete rezultata postupak je u kojem se provjerava relevantnost, točnost, i usporedivost, dostupnost i jasnoća podataka u vremenu i prostoru te usklađenost i usporedivost s postojećim unutarnjim i vanjskim referentnim izvorima podataka.

Analiza kvalitete odnosno validacija rezultata provodi se nakon postupka uređivanja na makrorazini i može uključivati sljedeće postupke:

- provjeru usklađenosti rezultata ako već nije ugrađeno u postupak uređivanja na makrorazini. Na primjer, je li vrijednost proizvodnje veća od dodane vrijednosti.
- Provjerava se usklađenost rezultata s rezultatima iz prethodnih referentnih razdoblja, a ovo se posebno odnosi na ona istraživanja čija primarna svrha nije mjerenje promjena tijekom vremena.
- Potrebno je napraviti i internu provjeru rezultata na razini Državnog zavoda za statistiku te povremenu provjeru relevantnosti i usklađenosti rezultata s vanjskim stručnjacima.

Upute za osiguranje kvalitete

Prije konačne validacije rezultata svi podaci dostupni u vezi s provedbom procesa obrade podataka moraju se ponovno analizirati, osobito rezultati pripremljeni na makrorazini.

Pri odabiru metoda i kriterija za ocjenu usklađenosti i usporedivosti rezultata potrebno je uzeti u obzir obilježja osnovnih podataka i status pripremljenih rezultata, odnosno je li riječ o privremenim ili konačnim rezultatima.

Kod odabira metoda i kriterija za procjenu usklađenosti i usporedivosti rezultata potrebno je definirati postupke koji se provode prije svakog objavljivanja rezultata, a provode se povremeno zajedno s periodičnošću njihove provedbe.

Postupci provjere trebaju se prilagoditi razdoblju provedbe statističkog istraživanja, ciljnoj populaciji, načinu prikupljanja podataka i vrsti izvora podataka koji mogu biti primarni ili sekundarni.

Rezultati analize trebali bi se koristiti u programskim dokumentima za poboljšanje kvalitete.

6.3. Detaljna analiza i tumačenja podataka za objavljivanje

Opis

Ocjenjuje se koliko uspješno statistika odražava početna očekivanja, promatrajući statistiku s pomoću različitih alata i medija. Osigurava se da statistika dosegne potrebnu razinu kvalitete i spremnost za upotrebu.

Podaci moraju biti relevantni i primjenjivi, dok zaključci trebaju biti potkrijepljeni podacima koji su rezultat statističkog procesa. Pri tumačenju podataka treba uzeti u obzir načine prikupljanja podataka, odnosno je li riječ o punom obuhvatu, provedbi istraživanja na uzorku ili se koristila administrativna baza podataka kao i dodatne informacije koje su usko povezane s prikupljanjem podataka, kao što je, npr. stopa neodgovora. Sve informacije trebaju biti dostupne korisnicima s pripadajućim metapodacima, i to od metodoloških objašnjenja, šifrnika, naznaka izvora podataka itd. jer mogu pomoći kod tumačenja popratnoga komentara, tabličnih ili grafičkih prikaza. Posebnu pozornost treba posvetiti opisima nedostataka podataka, poput odstupanja od ciljne populacije i primijenjene populacije. Načelo povjerljivosti podataka također se mora uzeti u obzir pri tumačenju podataka. Kod većine publikacija ovakvi podaci navedeni su u završnom dijelu publikacije, dok se u početnom dijelu publikacije predstavljaju rezultati.

Aktualne teme, izvanredne situacije i tumačenje podataka i pojava koje su povezane s njima treba prilagoditi korisniku podataka, vodeći pritom računa o tome koja je vrsta informacija važna za korisnika.

Tumačenje podataka trebalo bi biti prilagođeno osnovnim obilježjima korisničke skupine, bilo da je riječ o široj javnosti ili o stručnjacima za određeno područje. Statističar bi trebao biti upoznat sa željama korisnika i znati predstaviti podatke tako da budu u skladu s korisničkim očekivanjima, a to znači da podaci trebaju biti razumljivi, zanimljivi i korisni. Širu javnost zanimaju najvažnije ili najzanimljivije općenite informacije iskazane statističkim podacima, za razliku od stručnjaka za određeno područje koji se uglavnom koriste detaljnim podacima za svrhu daljnje analize podataka.

Upute za osiguranje kvalitete

Tumačenje rezultata prilagođuje se ciljnoj populaciji i mediju u kojem se podaci naknadno objavljuju. Ono mora biti nepristrano, objektivno, točno, jasno i razumljivo. Tumačenje kratkoročnih statistika mora se razlikovati od tumačenja strukturnih statistika. Korištenje referentnih točaka ima snažan utjecaj na tumačenje podataka, zato je potrebno koristiti se referentnim točkama, osobito kod vremenskih usporedbi, kada su podaci stabilni i nepristrani. Kod tumačenja rezultata u obliku indeksa i drugih relativnih brojeva mora se odabrati reprezentativno razdoblje usporedbe koje će korisniku podataka omogućiti učinkovito tumačenje kretanja pojave. Kod predstavljanja rezultata u obliku indeksa i drugih relativnih brojeva mora se voditi računa o tome da se točno tumači promjena pojave kada je izražena u postocima.

6.4. Zaštita povjerljivih podataka

Opis

Statistička zaštita podataka jest skup metoda koje smanjuju rizik od razotkrivanja podataka koji su prikupljeni od izvještajnih jedinica (osoba, kućanstava, poduzeća i drugo), a obuhvaća:

- statističku zaštitu tablica (agregirane podatke)
- statističku zaštitu mikropodataka (pojedinačne podatke o izvještajnim jedinicama).

Službena statistika Državnog zavoda za statistiku organizirana je na temelju Uredbe (EU) br. 223/2009 o europskoj statistici i Zakona o službenoj statistici.

Razlikuju se dvije vrste metoda statističke zaštite podataka:

- perturbativne metode – modificiraju podatke do te mjere da je rizik prepoznavanja sveden na najmanju moguću mjeru (npr. postupak zaokruživanja)
- neperturbativne metode – njima se ne mijenjaju podaci, ali ih se prikriva (npr. metodom manjkavih vrijednosti).

Neke se metode mogu koristiti samo za zaštitu tablica (npr. pravilo dominacije), dok se druge metode mogu koristiti samo za zaštitu mikropodataka (npr. mikroagregacija).

Statističkom zaštitom podataka sprečava se razotkrivanje osjetljivih podataka svih izvještajnih jedinica koje se pojavljuju u ćelijama tablice. U statističkoj zaštiti podataka razlikuju se dvije skupine tablica: tablice učestalosti (broj izvještajnih jedinica koje pripadaju ovoj ćeliji pojavljuje se u svakoj ćeliji tablice) i tablice vrijednosti (u svakoj ćeliji nalazi se zbroj vrijednosti za određene varijable koje su prikupile izvještajne jedinice).

U statističkoj zaštiti tablica treba obratiti pozornost na hijerarhiju varijabli i na odnose između tablica i varijabli.

Posljednjih godina povećali smo i dostupnost raspoloživih baza podataka (npr. popisni podaci prije nisu bili dostupni kao ni mogućnost povezivanja baza različitih istraživanja, uključujući i vanjske baze podataka), a vidljivo je povećanje broja zahtjeva za korištenje povjerljivih statističkih podataka (mikropodataka odnosno podataka o individualnim statističkim jedinicama koji omogućuju neizravnu identifikaciju) za svrhu provođenja statističkih analiza za znanstvene svrhe. Naime, u odnosu na agregirane podatke, mikropodaci omogućuju detaljnije i preciznije analize potrebne za donošenje zaključaka predviđenih ciljevima istraživačkih projekata. Zaštita mikropodataka važna je zbog velikog broja varijabli koje se nalaze u datoteci mikropodataka jer su određene kombinacije varijabli rijetke, dok je rizik od razotkrivanja statističke jedinice visok. Ako su mikropodaci javno dostupni u obliku datoteke za javnu uporabu, razina zaštite bit će vrlo visoka, a gubitak podataka bit će puno veći. Ako se mikropodaci koriste za svrhu provođenja statističkih analiza za znanstvene svrhe na temelju potpisanog ugovora, razina zaštite bit će niža jer je vjerojatnost zloupotrebe manja.

Upute za osiguranje kvalitete

Ulazni podaci moraju biti pripremljeni u odgovarajućem obliku i formatu. Potrebno je detaljno ispitati odnose između tablica i/ili varijabli koje se štiti. Sve objavljene i prenesene tablice koje sadržavaju ili će sadržavati iste podatke (npr. tablice za nacionalnu publikaciju i tablice koje objavljuje Eurostat) moraju se uzeti u obzir.

Svaka ustrojstvena jedinica Državnog zavoda za statistiku utvrdit će pravila i metode zaštite povjerljivih statističkih podataka u skladu s odredbama Naputka o pravilima i metodama zaštite pri diseminaciji službenih statističkih podataka Državnog zavoda za statistiku (NN, br. 78/21.).

Preporučuje se da varijabla ili njezini razredi ne budu detaljnije definirani nego što je potrebno jer to može imati snažan utjecaj na gubitak informacija.

Broj varijabli u datoteci mikropodataka ne smije biti veći od broja koji je potreban za provođenje kvalitetnih statističkih analiza za znanstvene svrhe. Naputkom o pravilima i metodama zaštite pri diseminaciji službenih podataka uređeni su postupci i načini primjene fizičkih, tehničkih i organizacijskih mjera kojima se osigurava fizička, tehnička i logička zaštita povjerljivih statističkih podataka od neovlaštenog pristupa, izmjene, gubitka, iznošenja ili uništenja te ostala pitanja važna za sigurnost i zaštitu povjerljivih statističkih podataka prikupljenih za proizvodnju službene statistike iz djelokruga Državnog zavoda za statistiku. Mjere zaštite povjerljivih statističkih podataka obuhvaćaju pravila i postupke kojima se zaštićuju identifikacijski podaci o statističkim jedinicama u procesu prikupljanja, obrade i dostave, a osobito uključuju:

- zaštitu prostorija u kojima su pohranjeni statistički podaci koja obuhvaća kontrolu ulaska i izlaska zaposlenika
- zaštitu pristupa opremi informatičkih tehnologija (u nastavku teksta: IT oprema) te mrežnim poslužiteljima, uređajima i medijima koji se koriste za pohranu povjerljivih statističkih podataka u elektroničkom obliku
- zaštitu pristupa povjerljivim statističkim podacima putem sigurnosnih dozvola i ovlaštenja
- evidentiranje događaja koji se odnose na pristup, korištenje i iznošenje povjerljivih statističkih podataka
- vođenje evidencija o dozvolama i ovlaštenjima zaposlenika koji u obavljanju poslova imaju pristup povjerljivim statističkim podacima
- obvezu potpisivanja izjave o statističkoj povjerljivosti
- druga pitanja koja se odnose na sigurnost i zaštitu povjerljivih statističkih podataka.

Mjere za zaštitu povjerljivih statističkih podataka dužni su primjenjivati svi zaposlenici Državnog zavoda za statistiku kao i druge fizičke ili pravne osobe kojima je na temelju ugovora ili ovlaštenja povjereno obavljanje pojedinih poslova ili zadataka u vezi s aktivnostima službene statistike.

Ako se izvještajne jedinice slažu s objavljivanjem podataka koji omogućuju njihovu identifikaciju, tim izvještajnim jedinicama dodjeljuje se poseban status, čime se smanjuje gubitak podataka u tablicama u kojima se objavljuju podaci koji se odnose na tu izvještajnu jedinicu.

7. Diseminacija

Objavlivanje statističkih podataka i informacija obuhvaća diseminaciju statističkih podataka javnosti odnosno svim korisnicima, upotrebom različitih medija i vrsta publikacija.

Rezultati statističkih istraživanja trebaju se objaviti pravodobno i uz poštovanje rokova objavljivanja. Podaci trebaju biti relevantni, točni, usklađeni, usporedivi, dostupni i jasni.

Najvažniji medij za objavljivanje podataka Državnog zavoda za statistiku jest mrežna stranica <https://dzs.gov.hr/>. Podaci trebaju biti objavljeni tako da svim korisnicima budu dostupni u isto vrijeme i na isti način. Podaci i metapodaci dostupni su korisnicima, uključujući opise metodologije i informacije o kvaliteti prikupljenih podataka. Bitan način objavljivanja podataka jest i informiranje korisnika o planiranim objavljivanjima podataka te o revizijama i ispravcima netočnih podataka.

Državni zavod za statistiku objavljuje podatke u sljedećim publikacijama, kartografskom portalu i bazama podataka:

STATISTIČKE INFORMACIJE

Ova publikacija džepnog formata na oko 105 stranica sadržava sažet izbor godišnjih podataka za Republiku Hrvatsku i pregled po županijama. Podaci su dani u trogodišnjim serijama i vrijedan su izbor osnovnih informacija o Republici Hrvatskoj.

HRVATSKA U BROJKAMA

Ova publikacija sadržava najvažnije podatke gospodarskih i društvenih kretanja za Republiku Hrvatsku prikazane u petogodišnjim serijama. Tablice i grafikoni smješteni su unutar 20 poglavlja. Namijenjena je najširem krugu korisnika.

PRIOPĆENJE

Riječ je o kratkim i brzim statističkim informacijama koje se objavljuju prema periodičnosti provođenja statističkih istraživanja (mjesečno, tromjesečno, polugodišnje, godišnje, dvogodišnje i trogodišnje). Priopćenje se objavljuje nekoliko dana po završetku obrade rezultata pojedinoga statističkog istraživanja.

STATISTIČKA IZVJEŠĆA

Prikazuju sveobuhvatne rezultate pojedinačnih istraživanja ili više istraživanja iz istog područja. Svako izvješće sadržava metodološka objašnjenja u kojima su dani izvori prikupljanja podataka, definicije i objašnjenja te usporedivost, tablični prikaz podataka u serijama za posljednju godinu te prikaz objavljenih publikacija. Podaci su prikazani za Republiku Hrvatsku ukupno, a u nekim izvješćima po županijama, gradovima i općinama.

METODOLOŠKE UPUTE

Metodološke upute sadržavaju podatke o izvorima i metodama prikupljanja podataka te o obuhvatu i definicijama statističkih istraživanja. Namijenjene su svima koji provode statistička istraživanja i korisnicima za bolje razumijevanje podataka.

STUDIJE I ANALIZE

Ovu seriju čine autorski radovi koji su rezultat primjene statističkih metoda i analiza iz pojedinih statističkih istraživanja. Autori su službenici Državnog zavoda za statistiku ili/i vanjski suradnici.

IZVANREDNA IZDANJA

Ova su izdanja posebno važna za daljnji razvoj statistike, a uglavnom su namijenjena stručnom krugu korisnika. Objavljuju se povremeno i obrađuju pitanja važna za razvoj metodologije i statističkih standarda usporedivih sa statističkim standardima UN-a, Eurostata i drugih međunarodnih organizacija. Neka od tih izdanja jesu prijevodi.

PROGRAM PUBLICIRANJA

Ovaj katalog na oko 120 stranica vodič je kroz naslove svih publikacija Državnog zavoda za statistiku planiranih za objavljivanje tijekom tekuće i do lipnja sljedeće godine. Uz broj i naslov većine publikacija dan je kratak sadržaj, rok i periodičnost objavljivanja, razina prikazivanja podataka, jezik i format te medij na kojem se publikacija objavljuje. Iza dijela Programa publiciranja koji najavljuje rokove objavljivanja nalazi se popis prethodno objavljenih izdanja po statističkim područjima. Cijene publikacija i usluga Državnog zavoda za statistiku navedene su na kraju Programa publiciranja i on je besplatan.

KALENDAR OBJAVLJIVANJA STATISTIČKIH PODATAKA

Ova publikacija u izravnoj je vezi s Programom publiciranja. Na 65-tak stranica Kalendara objavljivanja statističkih podataka točno su navedeni datum, dan i tjedan objavljivanja priopćenja, razdoblje na koje se podaci odnose te razina prikazivanja. Za ostale statističke publikacije navedeni su samo mjeseci u kojima će biti objavljene. Namijenjen je svim korisnicima publikacija Državnog zavoda za statistiku.

KARTOGRAFSKI PORTAL

Državni zavod za statistiku razvio je aplikaciju GeoSTAT, koja omogućuje kartografsko predstavljanje statističkih podataka, i to tako da se odabrana prostorna razina vrlo detaljno kombinira s odabranim statističkim podacima.

Točnije, korisnici mogu pratiti demografske podatke Popisa 2011., čak na mikrorazini čelija od 1 km². Na spomenutoj mikrorazini mogu se pratiti i podaci Registra poslovnih subjekata (svi pravno ustrojbeni oblici) te podaci iz područja turizma (smještajni kapaciteti, dolasci i noćenja turista). Time je olakšano praćenje dostupnih statističkih pokazatelja, osim u standardnom obliku putem Priopćenja i publikacija i interaktivnih karata na kojima svi zainteresirani mogu prostorno pratiti kretanje pojedinih statističkih pokazatelja. Također, na portalu GeoSTAT dostupni su podaci iz područja poljoprivrede, nacionalnih računa, zaposlenosti, obrazovanja, industrije, građevinarstva, trgovine, raspodjele rizika od siromaštva te poslovne demografije. Inače, svi podaci koji se nalaze na portalu GeoSTAT vidljivi su i u Katalogu metapodataka Državnog zavoda za statistiku, koji je razvijen u sklopu Nacionalne infrastrukture prostornih podataka. Katalog metapodataka razvijen je prema zahtjevima INSPIRE-a.

BAZE PODATAKA

Sektorska klasifikacija institucionalnih jedinica

Institucionalne jedinice u Republici Hrvatskoj klasificiraju se u sektore i podsektore u skladu s europskom regulativom za svrhu izrade službene statistike.

Europski sustav nacionalnih i regionalnih računa (ESA 2010) najnoviji je računovodstveni okvir za postizanje međunarodne usklađenosti nacionalnih računa koji, među ostalim, propisuje metodologiju za razvrstavanje institucionalnih jedinica u sektore i podsektore.

ESA 2010 objavljena je 26. lipnja 2013. u službenom glasilu Europske unije kao Aneks A Uredbe br. 549/2013.

Institucionalne jedinice razvrstane u skladu s metodologijom ESA 2010 mogu se pretražiti i preuzeti u csv formatu klikom na poveznicu <https://web.dzs.hr/sektorizacija.htm>.

Podaci o poslovnim subjektima (NKD 2007.)

Svaki poslovni subjekt upisan u Registar poslovnih subjekata može na mrežnoj stranici Državnog zavoda za statistiku s pomoću svoga matičnog broja vidjeti osnovne podatke koji se vode o njemu i o dijelu poslovnog subjekta.

STS baza podataka

Baza kratkoročnih poslovnih statistika ili STS baza podataka (Short-Term Statistics) jest relacijska baza konsolidiranih podataka o indeksima iz područja kratkoročne poslovne statistike. Aplikacija STS na jednome mjestu omogućuje korisnicima pretraživanje svih podataka u bazi. Svi podaci koji su spremljeni u STS bazu podataka mogu se u aplikaciji pregledavati u tabličnom obliku i u obliku grafikona.

PC AXIS baza podataka¹¹

- Poljoprivreda, lov, šumarstvo i ribarstvo
- Industrija
- Robna razmjena s inozemstvom
- Trgovina na malo
- Turizam
- Transport i komunikacije
- Okoliš
- Zaposlenost i plaće
- Cijene
- Osobna potrošnja i pokazatelji siromaštva
- Strukturne poslovne statistike
- Pravosuđe
- Stanovništvo
- Popis stanovništva 2011.
- Popis poljoprivrede 2003.
- Popis poljoprivrede 2020.
- Naselja i stanovništvo Republike Hrvatske 1857. – 2001.
- Subnacionalne statistike

Državni zavod za statistiku omogućuje sljedeće usluge svim korisnicima statističkih podataka:

- rješavanje korisničkih zahtjeva
- davanje osnovnih informacija o podacima koji se mogu pronaći na Eurostatovim mrežnim stranicama putem Eurostatova portala – European Statistical Data Support (ESDS) EUROSTAT na poveznici https://ec.europa.eu/eurostat/web/main/search/-/search/estatsearchportlet_WAR_estatsearchportlet_INSTANCE_bHVzuvn1SZ8J?p_auth=bQ6G7SQg&text=data+support
- korištenje knjižnične građe u čitaonici.

Važan zadatak Državnog zavoda za statistiku jest i pohranjivanje statističkih podataka za daljnju upotrebu. Državni zavod za statistiku vodi zbirne statističke podatke i u elektroničkom i tiskanom obliku. Pohranjivanje agregiranih podataka u elektroničkom obliku odvija se u sklopu provođenja politike izrade sigurnosnih kopija i arhiviranja elektroničkih podataka te u obliku mjesečnog arhiviranja mrežnih stranica Državnog zavoda za statistiku <https://dzs.gov.hr/>.

Sve publikacije koje objavljuje Državni zavod za statistiku čuvaju se u njemu u najmanje jednom primjerku. Tiskani primjerci publikacija dostavljaju se u dva primjeraka u arhiv NSK. Podaci o publikacijama unose se u katalog publikacija putem digitalnog sustava NSK.

Državni zavod za statistiku omogućuje istraživačkim subjektima i samostalnim istraživačima korištenje povjerljivih statističkih podataka za provedbu statističkih analiza isključivo za znanstvene svrhe.

Pohranjivanje statističkih mikropodataka u elektroničkom obliku odvija se kao dio provedbe politike izrade sigurnosnih kopija i arhiviranja elektroničkih podataka.

¹¹ PC AXIS je u početku razvijen za Popis stanovništva 1990. u Švedskoj. Daljnji razvoj vodi međunarodna referentna skupina zadužena za PC AXIS sa sudionicima iz drugih država licenciranih za PC-AXIS kojim se koriste nacionalni statistički uredi. Osobito ito je poznata bliska suradnja između statističkih ureda Švedske i Danske.

7.1. Dizajn i proizvodnja diseminacijskih proizvoda

Opis

Objavljivanje statističkih podataka i informacija slijedi standardizirane postupke u različitim tehnologijama. Standardizirani postupci temelje se na unaprijed propisanim strukturama, formatima i metapodacima te se koriste pri pripremanju tablica u fazi obrade statističkih podataka. Svi postupci podliježu istim, standardiziranim radnim postupcima u skladu s načelima dostupnosti i jasnoći podataka, vodeći računa o pravodobnosti. Kod pripremanja novog sadržaja i redovitog ažuriranja izlaznih podataka bitna je dobro pripremljena, strukturirana i organizirana dokumentacija, interna razmjena znanja, uhodani standardni komunikacijski kanali te arhiviranje dokumentacije i radnih postupaka.

Za svaki oblik objavljivanja statistika i informacija izlazni se podatak mora najprije ažurirati. Može biti riječ o podacima u obliku baza podataka ili o tablicama s konačnim objedinjenim podacima spremnima za objavljivanje.

Kada je riječ o kvaliteti revidiranih podataka, potrebno je slijediti upute Državnog zavoda za statistiku o reviziji statističkih podataka. Ako je pripremljena revizija statističkih podataka, potrebno ih je ažurirati i objaviti ovisno o statističkom području u skladu sa zadanim rokovima.

Način ažuriranja izlaznih podataka razlikuje se ovisno o vrsti medija objavljivanja ili vrsti potprocesa:

- dodavanje nove vremenske dimenzije datotekama baze podataka Državnog zavoda za statistiku
- dodavanje nove vremenske dimenzije u makroazu Državnog zavoda za statistiku za podatke vanjske trgovine
- podaci koje Državni zavod za statistiku priprema za prijenos u Eurostat
- pripremanje revidiranih podataka za interaktivne alate.

Upute za osiguranje kvalitete

Pri ažuriranju rezultata treba slijediti standardizirane procedure i provoditi ih na transparentan način. Bilo bi dobro automatizirati procese koji se ponavljaju u jednakim vremenskim intervalima.

Informacija o prijenosu podataka mora biti dostupna (putem Kalendara objavljivanja statističkih podataka ili izravno/osobno). Dokumentacija mora biti uređena, objavljena, dostupna i ažurna.

Kako je u ovaj poslovni proces uključeno nekoliko osoba iz različitih područja, unutarnja koordinacija mora biti dokumentirana.

Objavljene metodološke upute koje više nisu aktualne trebaju biti zamijenjene novom verzijom, dok stara verzija treba biti arhivirana.

Moraju se slijediti sva pravila koja se odnose na reviziju statističkih podataka.

7.2. Upravljanje objavljivanjem diseminacijskih proizvoda

Opis

Uključuje informiranje specifičnih skupina, poput predstavnika medija ili tijela državne uprave, kao i postupanje s eventualnim zabranama prije objavljivanja te dostavljanje podataka pretplatnicima.

Također uključuje upravljanje pristupom povjerljivim statističkim podacima za provedbu statističkih analiza isključivo za znanstvene svrhe koju obavljaju istraživački subjekti i samostalni istraživači.

Statistički sustav Republike Hrvatske proizvodi nepristrane statističke podatke o društvenim i ekonomskim procesima, osiguravajući činjeničnu osnovicu potrebnu za praćenje i analizu stanja hrvatskoga gospodarstva i usmjeravanje politika povezanih s razvojem hrvatskog društva i gospodarstva te politika Europske unije.

Objavljivanjem podataka kontinuiranom promidžbom Državni zavod za statistiku informira korisnike o važnosti statistike te potiče izvještajne jedinice na neometanu isporuku podataka.

Svako objavljivanje statističkih podataka mora se najaviti u skladu s planom objavljivanja (Kalendar objavljivanja statističkih podataka). Kalendar objavljivanja statističkih podataka priprema se na kraju godine, 18 mjeseci unaprijed, ovisno o završetku obrade pojedinoga statističkog istraživanja i o mogućnostima izrade pojedinog medija objavljivanja.

Jedanput godišnje objavljuje se Program publiciranja (na hrvatskome i engleskom jeziku) kako bi korisnici podataka bili što potpunije obaviješteni o medijima objavljivanja statističkih podataka Državnog zavoda za statistiku. Program je besplatan i nalazi se na mrežnim stranicama Državnog zavoda za statistiku. Na početku Programa publiciranja korisnici mogu naći informacije o vrstama publikacija koje se objavljuju.

Uz popis medija objavljivanja dan je kratak opis koji sadržava vrstu medija objavljivanja, broj i naslov, razinu prikazivanja podataka, rok i periodičnost objavljivanja te jezik na kojem se objavljuje. Na kraju Programa publiciranja korisnici mogu naći detaljne obavijesti o tome kako doći do statističkih informacija i podataka, koje su mogućnosti nabave, reklamacije, plaćanja, isporuke te obavijesti o cijenama publikacija i usluga.

Samo podaci koji su već objavljeni mogu se proslijediti međunarodnim organizacijama ili Eurostatu. Ako se povjerljivi podaci dostavljaju i Eurostatu, povjerljive ćelije moraju biti označene, a isti sadržaj na višoj razini agregiranja mora biti prethodno ili istodobno objavljen na mrežnome mjestu Državnog zavoda za statistiku.

Postupci objavljivanja podataka variraju ovisno o vrsti medija objavljivanja: objavljivanje vijesti na mrežnome mjestu; objavljivanje tiskanih publikacija; objavljivanje podataka u bazi podataka Državnog zavoda za statistiku; prijenos podataka Eurostatu.

Ispravljanje pogrešaka postupak je koji se odnosi na objavljivanje statistika. Svrha ispravljanja pogrešaka u objavljenim statistikama i informacijama jest pružanje točnih i visokokvalitetnih statistika i informacija korisnicima.

Detaljni postupci propisani su u uputama za ispravljanje pogrešaka u objavljenim statističkim podacima i informacijama.

Dobar primjer iz prakse naveden je u Metodološkim uputama br. 66, Mjesečno istraživanje o zaposlenima i plaći (RAD-1) na stranici 53. poglavlja 5.5. Upute za ispravak pogrešaka na poveznici https://podaci.dzs.hr/media/qf4jd3xw/metod_66.pdf.

Sadržaji za objavljivanje trebaju biti pripremljeni za potrebe raznovrsnih korisničkih profila. Predstavljanje podataka treba biti primjenjivo, razumljivo, jasno, zanimljivo i pripremljeno dvojezično na hrvatskome i engleskom jeziku.

Opća načela za predstavljanje podataka razlikuju se ovisno o vrsti i mediju objavljivanja.

Pri predstavljanju podataka treba imati na umu da svi korisnici nisu jednako statistički pismeni i da nemaju ista predznanja u vezi s upotrebom statistika i informacija. Stručna javnost uglavnom se koristi detaljnim podacima za svrhu daljnje analize podataka, stoga je uputno takve podatke objaviti u bazama podataka u formatima koji omogućuju daljnju obradu (elektronički obrazac). Širu javnost zanimaju najvažnije ili najzanimljivije općenite statistike i informacije objavljene na jasan i razumljiv način. Autor prezentacije podataka mora razmisliti tko je ciljna publika i što se želi saznati i u skladu s time treba predstaviti statistiku na razumljiv, zanimljiv i koristan način. Objavljeni podaci trebali bi sadržavati komentar i vizualizaciju podataka na razumljiv i jasan način. Komentar bi trebao biti primjenjiv, kratak, jednostavan, jasan, razumljiv i zanimljiv. Vizualizacija podataka trebala bi biti popraćena jednostavnim tablicama i grafičkim prikazima. Pri predstavljanju podataka potrebno je slijediti preporuke navedene u Uputama za uobličavanje i ujednačavanje publikacija¹².

¹² Upute za uobličavanje i ujednačavanje publikacija, DZS, Zagreb 2012. (poveznica [Upute za uobličavanje](#))

Tablice su sastavni dio statističkih medija objavljivanja jer su korisne za prikaz veće količine podataka. S pomoću tablica moguće je predstaviti bilo koju pojavu. Prikladne su za prikaz apsolutnih podataka i relativnih brojeva (koeficijenta, indeksa, prosjeka, struktura). Tablice sažetog prikaza prikadne su za objavljivanje u tiskanim publikacijama i vijestima na mrežnome mjestu, dok se detaljni podaci objavljuju u tablicama u bazi podataka Državnog zavoda za statistiku.

Grafikoni su grafički prikazi statističkih pojava u obliku stupaca, linija, pita itd. Obilježja pojava i podaci određuju koja je vrsta grafikona najprikadnija za prikaz podataka. Grafikoni mogu prikazati informacije učinkovitije i cjelovitije od komentara, dok tablica kao medij ima mogućnost prikazivanja veće količine podataka u obliku apsolutnih i/ili relativnih brojeva.

Grafički prikazani statistički podaci razumljiviji su i pregledniji u odnosu na njihovo predstavljanje tablicom, a veća preglednost grafičkog prikaza i snaga prvoga vizualnog dojma o obilježjima promatrane pojave prednosti su grafičkih prikaza.

Karte su najprikadniji alat za vizualizaciju pojava u prostornoj dimenziji. Omogućuju usporedbu prostornih jedinica različite veličine i prikazuju veliku količinu podataka na živopisan način. Statističke karte obično prikazuju relativne brojeve, posebno pokazatelje na 1 000 stanovnika i izračune na različitim razinama prostornih jedinica.

Pripreme za predstavljanje statističkih rezultata razlikuju se ovisno o mediju objavljivanja te u kojem će obliku biti predstavljeni rezultati, pa tako razlikujemo sljedeće:

- priprema informacija za objavljivanje na mrežnoj stranici
- priprema tiskane publikacije čiji je autor nositelj statističkog istraživanja
- priprema tiskane publikacije koju je pripremio urednik publikacije
- priprema novog sadržaja u bazi podataka Državnog zavoda za statistiku.

Pristup povjerljivim statističkim podacima za provedbu statističkih analiza za znanstvene svrhe može se omogućiti samo na temelju pisanog zahtjeva korisnika, i to na način i pod uvjetima koji su jasno definirani Pravilnikom o uvjetima i načinu korištenja povjerljivih statističkih podataka za znanstvene svrhe (NN, br. 137/13.). Koordinacija rješavanja zahtjeva za pristup povjerljivim statističkim podacima i pristup podacima trebaju biti definirani internim procedurama i pravilima koji će omogućiti kvalitetno upravljanje cijelim procesom.

Upute za osiguranje kvalitete

Prema datumu objavljivanja, koji je naznačen u programima publiciranja i kalendarima objavljivanja statističkih podataka, publikacije su dostupne točno u 11 sati, i to u elektroničkome i tiskanom obliku, slijedeći time Kodeks, koji nalaže pravodobnu diseminaciju, odnosno točan sat objavljivanja.

Kako bi se ispunio ovaj uvjet, potrebno je cijeli postupak statističkog istraživanja dizajnirati tako da se podaci pravodobno objavljuju. Bez obzira na to što su statistički podaci točni i detaljni, neće biti korisni ako se ne objavljuju pravodobno i uz poštovanje rokova objavljivanja.

Podaci moraju biti dostupni svim korisnicima na isti način. Korisnici isto tako trebaju imati na raspolaganju što više jednostavnih mogućnosti za iskazivanjem dodatnih zahtjeva i naručivanje podataka i usluga.

Načini objavljivanja podataka trebaju biti jasni korisnicima. Svi relevantni metapodaci, poput metodoloških objašnjenja, osnovnih pokazatelja kvalitete, upitnika itd. trebaju biti dostupni u skladu sa strukturom programa publiciranja na mrežnim stranicama u području kvalitete. Isto tako, potrebno je korisnicima na jednostavan način objasniti postupke ispravljanja pogrešaka i revidiranja podataka.

Potrebno je pripremiti i staviti na raspolaganje dokumentaciju o pravilima, uputama i preporukama za predstavljanje i objavljivanje podataka. Način predstavljanja statističkih podataka i pisanje jednoznačnih komentara te izrada tablica, grafikona i karata u medijima objavljivanja trebaju biti u skladu s dobrom praksom.

Autor, nositelj statističkog istraživanja, urednik publikacije i ustrojstvena jedinica zadužena za diseminaciju statističkih podataka i pripadajuće metodologije trebaju sudjelovati u pripremi podataka za objavljivanje.

Predlošci su dizajnirani prema sadržaju, tehnologiji pripreme podataka, tehničkim mogućnostima i alatima Državnog zavoda za statistiku te usvojenim standardima za objavljivanje.

Kod predstavljanja podataka treba se rukovoditi sljedećim načelima: relevantnost, točnost i pouzdanost, pravodobnost i poštovanje rokova objavljivanja, usklađenost i usporedivost te dostupnost i jasnoća podataka. Pri pripremi komentara o promjenama u pojavi obavezno treba navesti razdoblje promatranja i razdoblje usporedbe. U pripremu tekstova za objavljivanje treba angažirati prevoditelje i lektore.

Potrebno je jasno odrediti uvjete i načine pristupa pod kojima se može omogućiti korištenje povjerljivih statističkih podataka za provedbu statističkih analiza za znanstvene svrhe te definirati interne procedure i pravila za rješavanje takvih zahtjeva.

7.3. Promidžba diseminacijskih proizvoda

Opis

Područje diseminacije i komunikacije službene statistike iznimno je važno za približavanje statističkih podataka korisnicima na njima jednostavniji i razumljiviji način, tj. popularizacijom službene statistike.

Promicanje statističkih proizvoda pripremljenih tijekom proizvodnog procesa bitno je učiniti dostupnim putem različitih komunikacijskih kanala i na različite načine širokom krugu korisnika.

Uključuje, između ostalog, korištenje društvenih mreža kao važnoga komunikacijskog kanala za diseminaciju statističkih podataka, održavanje korisničkih vijeća s određenim ciljnim skupinama korisnika kako bi iz prve ruke dobili korisne informacije o potrebama korisnika statističkih podataka; održavanje edukacijskih radionica za učenike na temu može li statistika biti zanimljiva i koja je njezina važnost u suvremeno doba.

Statistička pismenost može se promicati organizacijom statističkih natjecanja kako bi potaknuli interes srednjoškolaca i nastavnika za statistiku te srednjoškolcima pokazali kakve im sve mogućnosti može otvoriti primjena stečenoga statističkog znanja, izradom interaktivnih statističkih portala koji, osim teorijskog dijela i objašnjenja, sadržavaju i primjere riješenih zadataka kao i zadatke za samostalno vježbanje usvojenoga statističkog znanja, a uključuju i interaktivne prikaze te izradu brošura u kojima se na jednostavan način objašnjavaju često korišteni, ali nedovoljno razumljivi statistički pojmovi.

Osim toga, važne komponente jesu funkcionalna i modernizirana mrežna stranica kao glavni izvor statističkih podataka i kanal za diseminaciju statističkih proizvoda te komunikaciju s predstavnicima medija. Unaprijed pripremljeni zanimljivi članci u povodu obilježavanja različitih međunarodnih, svjetskih i europskih dana te njihovo slanje predstavnicima medija može rezultirati većim brojem medijskih objava u kojima se spominje Državni zavod za statistiku i njegov djelokrug rada.

Upute za osiguranje kvalitete

Uz postizanje većeg stupnja zadovoljstva korisnika upotrebom modernih informacijskih tehnologija i smanjenjem potrebe za tiskanim publikacijama te statističkim obrascima i uputama postiže se i daljnja racionalizacija troškova poslovanja.

Potrebno je aktivno se koristiti društvenim mrežama kao važnim komunikacijskim kanalom za diseminaciju statističkih podataka (Facebook, Twitter, Instagram, YouTube), omogućavajući proaktivnu komunikaciju s postojećim i budućim korisnicima te jačajući imidž i ugled institucije. Trebalo bi proaktivno raditi na razvojnim aktivnostima kojima je cilj popularizacija službene statistike i približavanje statističkih podataka korisnicima na njima jednostavniji i razumljiviji način.

Potrebno je kontinuirano modernizirati komunikacijske kanale, ponajprije mrežnu stranicu Državnog zavoda za statistiku povećanim obuhvatom objavljenih statističkih podataka koji su dostupni u formatima prilagođenima potrebama korisnika.

Trebalo bi intenzivno nastaviti posvećivati pozornost radu s određenim ciljnim skupinama korisnika, održavajući korisnička vijeća te edukacijske radionice.

Trebalo bi kontinuirano provoditi promocijske aktivnosti radi povećanja stope odaziva ispitanika na istraživanja koje provodi Državni zavod za statistiku korištenjem svih komunikacijskih kanala (priopćenja za javnost, društvene mreže i slično), ali i modernizirati letke, najavna pisma i dopise koji se šalju izvještajnim jedinicama.

7.4. Upravljanje odnosima s korisnicima

Opis

Vodeći računa o zaštiti statističke povjerljivosti, korisnicima treba pravodobno biti omogućen pristup statističkim podacima koji su rezultat provedbe statističkih istraživanja. Podrška korisnicima pruža se elektroničkom poštom, telefonom, osobnim kontaktom te putem društvenih mreža.

Podrška korisnicima obuhvaća ustupanje statističkih podataka i informacija, savjete o pristupu podacima te pomoć u pronalaženju statističkih podataka i informacija na mrežnim stranicama Državnog zavoda za statistiku.

Javno dostupni statistički podaci i informacije na raspolaganju su svim korisnicima u skladu s uvjetima korištenja. Priprema podataka na zahtjev korisnika naplaćuje se prema važećem cjeniku Državnog zavoda za statistiku.

Evidencija pisanih zahtjeva za podacima, uključujući i zahtjeve za pristup povjerljivim statističkim podacima za znanstvene svrhe, čuva se u ustrojstvenoj jedinici nadležnoj za odnose s korisnicima. Korišti se za izradu analize korisničkih zahtjeva, na temelju koje se određuje struktura korisnika, načini komuniciranja s njima te koje su najtraženije statistike po statističkim područjima.

Korisnicima je također omogućen pristup povjerljivim statističkim podacima za provedbu statističkih analiza isključivo za znanstvene svrhe, ali pod posebnim uvjetima i na temelju sklopljenog ugovora. Pristup povjerljivim statističkim podacima za znanstvene svrhe određen je Pravilnikom o uvjetima i načinu korištenja povjerljivih statističkih podataka za znanstvene svrhe¹³.

Međunarodno izvještavanje i ispunjavanje obveza službene statistike obuhvaća, između ostalog, i prijenos podataka drugih tijela koja su Zakonom o službenoj statistici i Programom statističkih aktivnosti Republike Hrvatske ovlaštena za obavljanje poslova službene statistike, tzv. nositelji službene statistike.

Načini podrške korisnicima mogu se razlikovati ovisno o vrsti i obliku zahtjeva, i to kako slijedi:

- pismeni zahtjev za podacima
- telefonski upit u vezi s podacima
- pristup povjerljivim statističkim podacima za znanstvene svrhe
- međunarodno izvještavanje
- pristup korisnika knjižničnoj građi.

¹³Poveznica <http://www.propisi.hr/print.php?id=12692>

Upute za osiguranje kvalitete

Podaci koji se dostavljaju korisnicima trebaju biti točni, relevantni i nedvosmisleni. Korisnici trebaju pravodobno dobiti informacije, a u slučajevima kašnjenja isporuke trebali bi biti unaprijed obaviješteni. Sve statističke informacije besplatne su, dok usluge obrade statističkih podataka na zahtjev korisnika naplaćuju odnosno nastale troškove snosi korisnik koji je takvu uslugu tražio, a u skladu s Cjenikom publikacija i usluga Državnog zavoda za statistiku.

Državni zavod za statistiku ima pravo na naknadu stvarnih materijalnih troškova koji nastanu pružanjem informacije korisniku prava na pristup informacijama i ponovnu uporabu informacija, kao i na naknadu troškova dostave tražene informacije, koja se naplaćuje u skladu s Kriterijima za određivanje visine naknade i način naplate naknade.

Informacije o dostupnosti statistika objavljene su na mrežnoj stranici Državnog zavoda za statistiku te u publikacijama. Ustrojstvena jedinica nadležna za odnose s korisnicima kontinuirano pruža podršku korisnicima tijekom radnog vremena.

Korisnicima treba omogućiti besplatan pristup podacima na mrežnome mjestu, tiskanim publikacijama i dokumentaciji u knjižnici tijekom radnog vremena.

Osnovne informacije o pristupu povjerljivim statističkim podacima trebaju se objaviti na mrežnim stranicama Državnog zavoda za statistiku.

Briga o kvaliteti kod međunarodnog izvještavanja temelji se na standardiziranim radnim uputama, transparentnosti, urednoj dokumentaciji, radnim dokumentima objavljenima na intranetu i na internoj arhivi Državnog zavoda za statistiku.

8. Vrednovanje

Prioritet je kvaliteta statističkih istraživanja i podataka. Kako bi se osigurao zajednički okvir za kvalitetu statistike, pripremljen je Kodeks, koji je prvi put u 2005. usvojio SPC. Kodeks je revidirao ESSC u 2011. i 2017., a posljednju reviziju potvrdio je 16. studenoga 2017. Revidirani Kodeks ima 16 načela koja se tiču institucionalnog okružja, statističkih procesa i statističkih rezultata.

Cilj Kodeksa jest osigurati da su statistike proizvedene unutar ESS-a relevantne, pravodobne i točne te da su u skladu s načelima profesionalne neovisnosti, nepristranosti i objektivnosti. Skup pokazatelja najbolje prakse i standarda za svako načelo pruža upute i referenciju za ocjenu primjene Kodeksa. Usvajanjem Kodeksa statistički uredi država članica Europske unije i Eurostat obvezali su se provoditi aktivnosti koje će osigurati visoku kvalitetu statistike.

Prikupljanje podataka o kvaliteti statističkih podataka odvija se sustavno tijekom provedbe svih statističkih procesa. Pripremljena dokumentacija o provedbi statističkih istraživanja po svim fazama pomaže u otkrivanju mogućih sustavnih pogrešaka u pojedinim potprocesima. S pomoću takvih podataka moguće je procijeniti kvalitetu statističkih podataka i kritički procijeniti njihov izračun. Osim toga, oni su bitni i za korisnike jer stječu dodatni uvid u postupak prikupljanja, obrade i diseminaciji podataka. Objavljivanje podataka o njihovoj kvaliteti transparentan je način informiranja korisnika o različitim aspektima statistike. Na temelju prikupljenih podataka prema potrebi se priprema plan poboljšanja i uvode se promjene u provedbeni postupak u skladu s postavljenim planom.

8.1. Prikupljanje informacija za vrednovanje

Opis

Proces provođenja statističkog istraživanja obuhvaća potprocese koji mogu utjecati na kvalitetu statističkih rezultata. Kako bi se što bolje mogla procijeniti kvaliteta statističkih rezultata, trebalo bi prikupiti što više podataka o svakom potprocesu.

Podaci koji se prikupljaju za ocjenu kvalitete mogu se podijeliti u dvije skupine. Prva skupina obuhvaća informacije izvedene izravno iz statističkih procesa i potprocesa, dok drugu skupinu čine podaci preuzeti iz rezultata statističkih istraživanja. Ovdje treba posebno voditi računa o zadovoljstvu korisnika, informacijama o novim ili promijenjenim zahtjevima ili o neprimjenjivosti objavljenih rezultata za potrebe korisnika.

Što se tiče informacija koje se prikupljaju izravno iz statističkog procesa, posebnu pozornost treba obratiti na kvantitativne informacije, tj. na one informacije koje nazivamo pokazateljima kvalitete. Dva tipična primjera takvih pokazatelja jesu pogreška na uzorku i stopa neodgovora. Ti su pokazatelji bitni jer omogućuju empirijsko-analitički pristup kod praćenja i osiguravanja kvalitete statističkih proizvoda i procesa. Izračun tih pokazatelja potrebno je uključiti u statistički proces, dok je u kasnijoj fazi potrebno osigurati da se vrijednosti pokazatelja redovito analiziraju i objavljuju u izvještajima o kvaliteti.

Upute za osiguranje kvalitete

Statistički se procesi trebaju pažljivo analizirati kao i svi mogući izvori informacija koji bi se mogli iskoristiti za procjenu kvalitete. Popis standardnih pokazatelja kvalitete treba uzeti u obzir pri određivanju pokazatelja kvalitete koji će se izračunati i naknadno uključiti u izvještaje o kvaliteti. Ako je potrebno, u popis treba dodati pokazatelje koji su karakteristični za određeno statističko područje (npr. zrcalna statistika za područje vanjske trgovine ili turizma).

Pri pripremi uputa za izračun standardnih pokazatelja kvalitete strogo se treba pridržavati Priručnika za izračun pokazatelja kvalitete pod poveznicom https://podaci.dzs.hr/media/1cmjru1a/metod_73.pdf. Trebaju se koristiti svi dostupni izvori podataka za preuzimanje informacija i podataka od korisnika. Kako bi informacije bile korisne za daljnju analizu, potrebno ih je prilagoditi i uključiti u izvještaje o kvaliteti. Isto tako, potrebno je osigurati da sve informacije budu što je prije moguće dostupne jer je to jedini način da njihova upotreba bude korisna i učinkovita.

8.2. Vrednovanje rezultata

Opis

Materijal za vrednovanje može proizaći iz bilo koje faze ili potprocesa. Proces pripreme dokumentacije za statističko istraživanje uključuje detaljan opis statističke aktivnosti od opisa pojmova, definicija, primijenjenih metoda, proizvodnog procesa, informacijskog sustava pa do radnih uputa. Koliko je kvalitetno pripremljena dokumentacija za provedbu statističkog istraživanja, u toj mjeri će biti postavljena i kvaliteta statističkog istraživanja. Dokumentacija statističkog istraživanja važan je alat u komunikaciji između različitih sudionika u procesu organiziranja i provedbe statističkih istraživanja kao i između proizvođača i korisnika statističkih podataka. Dokumentacija o istraživanju dio je metapodataka.

Općenito, dokumentaciju možemo podijeliti na dokumentaciju za korisnike rezultata statističkih istraživanja i dokumentaciju za provedbu statističkih istraživanja. Prva vrsta dokumentacije opisuje i dokumentira statističke rezultate i obično se objavljuje u javnosti. Druga vrsta dokumentacije opisuje statističke postupke i procese koji se koriste tijekom statističkih procesa i potprocesa. Ta dokumentacija nastaje tijekom organiziranja i provedbe pojedinih procesa i potprocesa za statistička istraživanja i uglavnom je namijenjena za internu upotrebu.

Svrha korisničke dokumentacije jest pomoći korisnicima da bolje razumiju što prikazuju podaci i statističke metode. Kvalitetnom pripremom korisničke dokumentacije korisnicima je omogućeno lakše razumijevanje, pronalaženje i obrada podataka.

Primjeri sadržaja dokumentacije za korisnike jesu upitnici za ankete, prezentacijski letci, metodološka objašnjenja i izvještaji o kvaliteti statističkih istraživanja (poveznica <https://dzs.gov.hr/istaknute-teme-162/kvaliteta/izvjestavanje-o-kvaliteti/izvjestaji-o-kvaliteti-prema-statistickim-podrucjima/841>).

Svrha dokumentacije za nositelje statističkih istraživanja jest slijediti načela relevantnosti, točnosti, pravodobnosti i poštivanje rokova objavljivanja, usklađenosti i usporedivosti, dostupnosti i jasnoće podataka. U dokumentaciji za nositelje statističkih istraživanja trebaju se detaljno opisati pojedini postupci koji se koriste u provedbi istraživanja (npr. određivanje ciljne populacije, odabir uzorka, dizajn upitnika, uređivanje podataka, objavljivanje podataka itd.). Opisi bi trebali sadržavati izvještaj o tome što je učinjeno u svakom koraku i objašnjenje zašto je odabran određeni način provedbe. Sve su to korisne informacije za razvoj i unapređenje procesa i potprocesa uhodanih statističkih istraživanja i/ili za planiranje novih istraživanja. Katkad su pojedini tekstovi iz takve dokumentacije zanimljivi i vanjskim korisnicima podataka, stoga ga je, ovisno o potrebi, katkada potrebno i objaviti.

Državni zavod za statistiku čuva statističke mikropodatke u elektroničkom obliku za daljnju upotrebu u Državnom zavodu za statistiku i za upotrebu za istraživačke i analitičke svrhe. Arhiviranje statističkih mikropodataka u elektroničkom obliku odvija se kao dio provedbe politike izrade sigurnosnih kopija i pohrane elektroničkih podataka.

Sigurnosna kopija podijeljena je na sigurnosnu kopiju baze podataka i sigurnosnu kopiju različitih poslužitelja. Sve baze podataka Državnog zavoda za statistiku koje sadržavaju podatke iz statističkih istraživanja i metapodatke trajne su vrijednosti te podliježu jedinstvenom sustavu sigurnosnih kopija. Sigurnosne kopije izrađuju se samo za podatke koji se svakodnevno unapređuju. Podaci se čuvaju u sigurnom prostoru za pohranu medija.

Statistički mikropodaci pohranjuju se zajedno s datotekama i metapodacima u elektroničkom obliku na za to predviđenim poslužiteljima.

Državni zavod za statistiku arhivira agregirane statističke podatke za daljnju upotrebu i u elektroničkom obliku i u tiskanom obliku. Arhiviranje agregiranih podataka u elektroničkom obliku odvija se u sklopu provođenja politike izrade sigurnosnih kopija i pohrane elektroničkih podataka te u obliku mjesečnog arhiviranja mrežnog portala Državnog zavoda za statistiku <https://dzs.gov.hr/>.

Agregirani statistički podaci u tiskanom obliku mogu se objavljivati u publikacijama koje objavljuje Državni zavod za statistiku ili u publikacijama koje Državni zavod za statistiku preuzima od drugih institucija. Sve se publikacije nalaze u arhivi knjižnice Državnog zavoda za statistiku. Publikacije koje objavljuje Državni zavod za statistiku arhiviraju se u najmanje jednom primjerku.

Upute za osiguranje kvalitete

Dokumentacija o istraživanju mora biti točna, sveobuhvatna i razumljiva ciljnoj publici kojoj je namijenjena. Pri pripremi dokumentacije trebaju se koristiti standardni obrasci i predlošci. Sadržaj pojedinih procesa treba koliko je moguće slijediti standardnu strukturu. Kako bi se izbjegle nejasnoće i nesporazumi, dokumentacija za korisnike trebala bi biti pripremljena jednostavnim jezikom.

Dokumentacija bi za korisnike trebala sadržavati jasnoću i ponovljivost primijenjenih postupaka i procesa, osobito opise svih odstupanja od standardnih postupaka. Sadržaj i razina detalja dokumentacije trebali bi biti prilagođeni ciljnoj publici.

U slučaju utvrđenih pogrešaka dokumentacija svih procesa bi trebala sadržavati opise postupaka koje treba slijediti i imena osoba ili ustrojstvenih jedinica s kojima je potrebno kontaktirati u slučaju nejasnoća.

Izradu sigurnosnih kopija i pohranu treba izvršiti u skladu s uputama za izradu i pohranu sigurnosnih kopija.

Vlasnici poslovnih procesa odgovorni su u svojim ustrojstvenim jedinicama za sustavan nadzor nad upravljanjem informacijskim rizicima te za usklađivanje potrebnih aktivnosti s propisanim postupcima. Kontrole navedenih aktivnosti obavljaju se jedanput godišnje ili ovisno o potrebi.

8.3. Akcijski plan za poboljšanje

Opis

Akcijski plan za poboljšanje treba uključivati vrednovanje proizvodnog procesa s prijedlozima za njegovo poboljšanje temeljenog na izvještaju o kvaliteti. Podatke prikupljene za procjenu kvalitete potrebno je analizirati te pripremiti izvještaj koji sadržava najviše identificiranih kritičnih točaka u procesu ili u širem kontekstu provedbe statističkog istraživanja. Na temelju provedene analize procesa potrebno je dostaviti prijedloge za uvođenje poboljšanja kod provedbe statističkih istraživanja u sljedećemu programskom razdoblju.

Kod statističkih istraživanja čija se organizacija i provedba odvija s vremena na vrijeme, posebnu pozornost treba obratiti na vrijednosne pokazatelje kvalitete koje je potrebno usporediti s prethodnim razdobljima. Ako se utvrde veće razlike u odnosu na vrijednosti iz prethodnih razdoblja, potrebno je analizirati razloge zbog kojih su nastale takve razlike i pripremiti akcijski plan koji će omogućiti unapređenje kvalitete statističkog područja čija se kvaliteta mjeri.

Na temelju prikupljenih i pripremljenih informacija izrađuje se izvještaj o kvaliteti koji obuhvaća opise svih dimenzija kvalitete, zajedno s vrijednosnim pokazateljima kvalitete.

Posebnu pozornost potrebno je posvetiti analizi informacija koje se preuzimaju od korisnika podataka. Sažeti izvještaj trebao bi sadržavati opise svih uočenih neispunjenih zahtjeva korisnika i predvidjeti koji bi se od neispunjenih zahtjeva mogao ispuniti kod provedbe statističkog istraživanja u sljedećemu programskom razdoblju.

Prikupljene informacije o kvaliteti i njihova analiza koristi se za pripremu plana poboljšanja. Ako se utvrde glavni nedostaci u prikupljenim informacijama o tijeku organiziranja i provedbe istraživanja, treba pripremiti akcijski plan za poboljšanje i na temelju tog plana uvesti u proceduru potrebna poboljšanja.

Upute za osiguranje kvalitete

Prikupljeni se podaci trebaju analizirati na stručnoj i nepristranoj razini. U analizi treba utvrditi nedostatke. Rezultati analize kvalitete trebaju biti predloženi jednostavnim jezikom koji će biti razumljiv, čak i nekome tko nije izravno uključen u predmetno istraživanje.

U analizi pokazatelja kvalitete posebnu pozornost treba obratiti na usporedbu s pokazateljima kvalitete iz prethodnih razdoblja ili s drugim statističkim istraživanjima. Izvještaj o kvaliteti izrađuje se i objavljuje nakon objavljivanja statistike. Sve prikupljene odgovore korisnika bi trebalo pregledati i analizirati te dati kritički osvrt. Važno je u fazi analize posebno istaknuti one zahtjeve koji se mogu uključiti u planove poboljšanja sljedećega programskog razdoblja. Sveobuhvatan pregled poboljšanja koja se mogu uvesti u proces istraživanja trebali bi biti sastavni dio programskog plana.

LITERATURA

1. Cox, B. G. et al. (1995). *Business Survey Methods*. New York: Wiley.
2. Sarndal, C.E., Svensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
3. Cochran, W. G. (1963). *Sampling Techniques*. London: John Wiley and Sons, Inc.
4. Design your questions right. How to develop, test, evaluate and improve questionnaires. (2004). Stockholm: Statistics Sweden. Pristupljeno 19. lipnja 2012. s mrežne stranice na adresi: https://share.scb.se/ov9993/data/publikationer/statistik/publikationer/ov9999_2004a01_br_x97op0402.pdf
5. Dillman, D. A. (2003). *Mail and Internet Surveys: The Tailored Design Method*. London: John Wiley and Sons, Inc.
6. Fellegi, I.P., Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*. 71(353), 17-35.
7. Granquist, L. (1991). Macro Editing – A Review of Some Methods for Rationalizing the Editing of Survey Data. *Statistical Journal*, 8, 137-145.
8. Hidiroglou, M.A., J.M. Berthelot, (1986). Statistical Editing and Imputation for Periodic Business Surveys. *Survey Methodology*, 12, 73-83.
9. Cochran, W. G. (1963). *Sampling Techniques*. London: John Wiley & Sons, Inc.
10. Levy, P. S. Lemeshow, S. (2004). *Sampling of Populations, Methods and Applications*. New York: Wiley.
11. 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.
12. Quality Guidelines for Official Statistics. (2007). Helsinki: Statistics Finland. Pristupljeno 19. lipnja 2012. s mrežne stranice na adresi: http://www.stat.fi/org/periaatteet/qg_2ed_en.pdf
13. Sarndal, C. E. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
14. 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.
15. Snijkers, G. (1999). *Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality*. Deen Haag: Statistics Netherlands.
16. Snijkers, G.J.M.E. (2002). *Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality*. Utrecht: Sveučilište Utrecht.
17. Thompson, S. K. (1992). *Sampling*. London: John Wiley & Sons, Inc.
18. *Survey Methods and Practices*. (2003). Ottawa: Statistics Canada.
19. *Quality Guidelines*. (2003). Ottawa: Statistics Canada.

20. Quality Guidelines for Official Statistics. (2007). Helsinki: Statistics Finland.
21. Tourangeau, R., Rips, L. J., Rasinski, K. (2000). The Psychology of Survey Response. Cambridge: Cambridge University Press.
22. Wallgren A., Wallgren B. (2007). Register-based Statistics; Administrative Data for Statistical Purposes. London: John Wiley & sons.
23. Cochran, W.G. (1977): Sampling Techniques. London: John Wiley & Sons, Inc.
24. ESS guidelines on seasonal adjustment, 2015 edition, Eurostat
25. Handbook on Seasonal Adjustment, 2018 edition, Eurostat
26. Seasonal Adjustment | CROS (europa.eu)
<https://www.idsurvey.com/en/methods-web-survey-cawi/>
[https://dimewiki.worldbank.org/wiki/Computer-Assisted_Personal_Interviews_\(CAPI\)](https://dimewiki.worldbank.org/wiki/Computer-Assisted_Personal_Interviews_(CAPI))
<https://www.idsurvey.com/en/>
https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=252469