



Broj: 04-43-1 -1-57-2/26
Datum: 20. 01. 2026. godine

**STANDARDI KVALITETA
STATISTIČKOG POSLOVNOG PROCESA
- SMJERNICE -**

SADRŽAJ

strana

U V O D	5
KVALITET U SLUŽBENOJ STATISTICI	6
STANDARDI KVALITETA STATISTIČKOG POSLOVNOG PROCESA	11
1 IDENTIFIKACIJA POTREBA	12
1.1 Utvrđivanje potreba za podacima	12
1.2 Provjera i pregled izvora – raspoloživosti podataka	13
1.3 Priprema i prihvatanje poslovnog slučaja (business case)	13
2 PRIPREMA STATISTIČKOG ISTRAŽIVANJA	15
2.1 Dizajniranje proizvoda statističkog istraživanja	16
2.2 Priprema metodologije za prikupljanje podataka i provođenje istraživanja	16
2.3 Priprema izvora podataka za izradu okvira uzorka	21
2.4 Priprema metodologije statističke obrade podataka	23
3 USPOSTAVA OKVIRA ISTRAŽIVANJA	26
3.1 Uspostava kanala i instrumenata za prikupljanje podataka	26
3.2 Uspostava programske podrške	29
3.3 Uspostava komponenti diseminacije	29
3.4 Testiranje alata za prikupljanje i obradu podataka	30
3.5 Testiranje i konfiguracija statističkog poslovnog procesa	30
4 PRIKUPLJANJE PODATAKA	31
4.1 Izrada okvira uzorka i izbor uzorka	31
4.2 Organizacija prikupljanja podataka	35
4.3 Pokretanje prikupljanja podataka	43
4.4 Unos podataka	47
5 OBRADA PODATAKA	49
5.1 Povezivanje različitih izvora podataka	49
5.2 Šifriranje	50
5.3 Kontrola prihvatljivosti/validacija podataka – logičke kontrole	51
5.4 Editovanje i Imputacija	52
5.5 Proizvodnja izvedenih varijabli i jedinica	56
5.6 Ponderisanje	56
5.7 Izračunavanje agregata	57
5.8 Izrada finalnih datoteka	58

6. ANALIZA PODATAKA	59
6.1 Priprema nacрта rezultata	59
6.2 Analiza relevantnosti i potvrđivanje rezultata	60
6.3 Interpretacija rezultata	63
6.4 Statistička zaštita podataka	64
7. DISEMINACIJA PODATAKA	80
7.1 Ažuriranje izlaznih rezultata	80
7.2 Proizvodnja i prezentacija proizvoda za objavljivanje	80
7.3 Upravljanje objavom diseminacijskih proizvoda	85
7.4 Promoviranje diseminacijskih proizvoda	87
7.5 Podrška korisnicima	88
8. EVALUACIJA	89
8.1 Prikupljanje dokumentacije o istraživanju	89
8.2 Provođenje evaluacije - evaluacija rezultata	91
8.3 Akcioni plan za poboljšanje	93
IZVORI I LITERATURA	94

Uvod

Smjernice navedene u ovom dokumentu su važna orijentacija u procesu praćenja i implementiranja preporučenih standarda kvaliteta statističkog poslovnog procesa. Preporučeni standardi su jednim dijelom propisani međunarodnim regulativama, a drugim dijelom se temelje na primjeni najboljih metoda i statističke prakse (»Current Best Practises«).

U interesu proizvodnje pouzdane, relevantne i međunarodne uporedive statistike, nadležne statističke institucije u BiH imaju obavezu da učine sve što je potrebno, da u punom kapacitetu prihvate smjernice i preporuke koje su navedene u ovom dokumentu. Svi akteri koji su uključeni u proizvodnju statistike u našoj zemlji imaju veliku odgovornost, da u narednom periodu učine maksimalan napor da odgovore potrebama korisnika statističkih podataka kako u zemlji tako i inostranstvu.

Glavna namjena ovog dokumenta je da na jednom mjestu opiše pojedine faze u procesu provođenja statističkog istraživanja. Pored opisa pojedinih faza procesa, za svaku fazu i sam proces su date i okvirne upute (smjernice kvaliteta), koje proizvođači statistike trebaju u što većoj mjeri slijediti.

Proizvodnja statističkih rezultata je zahtjevan proces, u kojem kvalitet treba stalno pratiti i poboljšavati. Iz tih razloga, neophodno je svaki dokument koji opisuje taj proces i pruža smjernice za standarde kvaliteta, neprestano dopunjavati i poboljšavati.

DIREKTORICA

Vesna Ćužić

KVALITET U SLUŽBENOJ STATISTICI

Definicija kvaliteta

Kada govorimo o pojmu »kvalitet u statistici«, svjesni smo činjenice da se ovdje ne radi o jednodimenzionalnoj terminologiji. U ovom pojmu je sadržan veliki broj aspekata uključujući i s time vezane komponente kvaliteta.

ISO standard 8402 iz 1986. godine definira kvalitet na slijedeći način: „*Kvalitet je ukupnost svojstava i karakteristika jednog proizvoda ili usluge, koji se temelje na ispunjavanju utvrđenih ili implicitnih zahtjeva korisnika.*“

U skladu s ovom definicijom, a vezano za praksu statističke proizvodnje, postavljaju se dva pitanja:

1. Koje su bitne odlike i karakteristike statističkih proizvoda?
2. Koji zahtjevi se mogu utvrditi, a koji su implicitno dati?

Inicijalna razmatranja, kako kvalitet definirati u kontekstu službene statistike, urađena su već sredinom 90-tih od strane Eurostat-a. Radna grupa za kvalitet u Eurostatu je 1998. godine definirala pojam „kvaliteta“ i isti je ušao u zvaničnu upotrebu u službenu statistiku po prvi put u zemljama članicama EU-e. Radna grupa, (tzv. »Leadership Group Quality – LEG – Quality«) je utvrdila važeću verziju definicije 2003. godine u dokumentu *Assessment of quality in statistics, Methodological documents – definition of quality in statistics*.

Dakle, kvalitet u statistici se posmatra u odnosu na slijedećih šest kriterija/dimenzija kvaliteta:

- Relevantnost
- Tačnost
- Pravovremenost i aktuelnost objave
- Dostupnost i razumljivost
- Uporedivost
- Usklađenost

Relevantnost

Relevantnost je stepen do kojeg statistički proizvod zadovoljava sadašnje i potencijalne potrebe korisnika. To ovisi o tome, u kojoj mjeri statistika koja je proizvedena kao i korišteni koncepti (definicije, varijable, klasifikacije, itd) odražavaju potrebe korisnika.

Tačnost

Tačnost se definira kao stepen slaganja/bliskosti između procenjenih vrijednosti (koje dobijamo na kraju statističke obrade) i ispravnih, ali nepoznatih populacijskih vrijednosti.

Pravovremenost i aktuelnost objave

Pravovremenost statističkih rezultata je vremenski razmak između posljednjeg dana referentnog perioda na koji se odnose podaci i njegove raspoloživosti, tj. tačnog datuma objave.

Aktuelnost je vremenski razmak između stvarnog datuma objave podataka i planiranog datuma koji je naznačen u zvaničnom kalendaru publikovanja. Ako se gore spomenuti datumi poklapaju, smatramo da je objava bila aktuelna.

Dostupnost i razumljivost

Dostupnost statističkih rezultata/proizvoda se odnosi na konkretne fizičke okolnosti, u kojima su podaci dostupni korisniku: gdje se podaci fizički nalaze, kakve su mogućnosti naručivanja, kalendar objave, jasna politika plaćanja, dostupnost mikro - i makropodataka, različiti formati i mediji (npr. papir, datoteke, CD-ROM, internet).

Razumljivost se odnosi na statističko informaciono okruženje podataka, preko kojeg korisnik dolazi do informacija: da li su skupa s podacima dostupne i tekstualne informacije, metodološka objašnjenja, dokumentacija; da li su podaci opremljeni s grafičkim i drugim slikovnim gradivom; da li se daje informacija o kvalitetu podataka; da li su korisniku po potrebi dostupne dodatne informacije.

Uporedivost

Namjena komponente uporedivosti je izmjeriti razlike, koje se pojavljuju pri korištenju statističkih koncepata i definicija za izračunavanje uporednih statistika u različitim geografskim područjima ili različitim referentnim periodima. Uporedivost se izražava na temelju dvije podkomponente i to: *uporedivost s drugim zemljama* i *uporedivost u vremenu*.

Uporedivost podataka sa podacima drugih zemalja se postiže prihvatanjem temeljnih principa i definicija iz međunarodnih standarda i regulativa EU-e.

Uporedivost u vremenu se postiže korištenjem iste metodologije, definicija i koncepata u provođenju istraživanja tokom određenog niza godina.

Usklađenost

Usklađenost dva ili više statističkih proizvoda, odnosi se na stepen do kojeg su statistički procesi po kojem su statistički proizvodi proizvedeni koristili iste koncepte (klasifikacije, definicije i ciljnu populaciju) i harmonizirane metode.

Kao što je naglašeno, kvalitet u statistici obuhvata više faktora. Ova multidimenzionalnost oglada se takođe i kroz samu definiciju.

Posmatrajući dimenzije kvaliteta, postaje jasno da svaka dimenzija pojedinačno za sebe predstavlja važan faktor uticaja na ukupni kvalitet.

Idemo li pak korak dalje i posmatramo kriterije u interakciji, vidjet ćemo da se pri postizanju određenih ciljeva pruža konkurentna situacija.

Ova „Trade off“ situacija (ustupci - kompromisi između izlaznih komponenti/dimenzija kvaliteta) postaje najočiglednija ako posmatramo komponentu kvaliteta *tačnost* naspram komponente *aktuelnosti*. Pitanje je »da li je bolje dati tačan podatak sa zakašnjenjem ili dati na vrijeme ali pogrešan podatak«? Ali ovo je samo jedan upečatljiv primjer.

Može se čak tako daleko ići i kazati, *da pri proizvoljnom izboru bilo koje dvije ili više dimenzija kvaliteta dolazi do konkurencije ciljeva*. Ova situacija ne čini jednostavnim postizanje optimalnog rezultata kvaliteta za službenu statistiku.

U svakom slučaju, obzirom na sve gore navedene kriterije, cilj je optimizirati kvalitet. Naravno da važnu ulogu u ovom kontekstu ima aspekt *mjerenja kvaliteta*. *Posmatrajući ovdje multidimenzionalnost, vidljivo je da kvalitet u statistici ne može biti prikazan preko jednog pojedinačnog kvantitativnog broja*.

Postoji puno više sinteza svih komponenti ocjene, koje se moraju sažeti u ukupnu sliku. *Ukupna ocjena nekog statističkog proizvoda zavisi i od toga, kako su ponderirane pojedine dimenzije*.

Kvalitet proizvoda 1	Kvalitet proizvoda 2
Relevantnost	Relevantnost
Tačnost	Tačnost
Pravovremenost i aktuelnost objave	Pravovremenost i aktuelnost objave
Dostupnost i razumljivost	Dostupnost i razumljivost
Uporedivost	Uporedivost
Usklađenost	Usklađenost
Ostatak	Ostatak

Slika 1: Ponder za pojedine dimenzije kvaliteta može biti različit za dva statistička proizvoda (Širina pravougaonika odgovara dotičnom ponderu dimenzije kvaliteta)

Statističko istraživanje

U statističkom procesu centralno mjesto zauzima statističko istraživanje. Statističko istraživanje je veoma širok pojam i obuhvata mnogo različitih praktičnih izvedbi.

Prema »klasičnom« shvatanju, osnova statističkog istraživanja je neposredno prikupljanje podataka za slučajno izabran uzorak jedinica posmatranja; prema „širem“ shvatanju (koje sve više zamjenjuje »klasično«) statističko istraživanje uključuje različite izvedbe procesa, prije svega u fazi prikupljanja podataka.

Statistička istraživanja dijelimo u grupe, zavisno od načina provođenja, i to:

U pogledu **načina obuhvata**:

- *popis* – posmatramo cijelu populaciju
- *istraživanje na bazi uzorka* – posmatramo samo izabrani uzorak populacije
- *izvedene statistike* - statističke rezultate izvodimo iz već postojećih statističkih agregata (prije svega statistika nacionalnih računa i platnog bilansa)

U pogledu **načina prikupljanja podataka**:

- *Intervju/anketiranje* (telefonom ili na terenu)
- *samopopunjavanje* (poštom ili elektronski upitnik)
- *drugi načini prikupljanja* (posmatranje i praćenje ...)

U pogledu **periodike**:

- *mjesečna*
- *kvartalna*
- *polugodišnja*
- *godišnja*
- *drugi periodi* (sedmična, višegodišnja)

U pogledu **korištenog izvora**:

- *statistički (primarni) izvori*
- *administrativni i drugi sekundarni izvori*

U praksi se često susrećemo sa kombinacijama gore pomenutih kategorija. Kod brojnih istraživanja koristi se kombinacija primarnih i sekundarnih izvora.

Opće smjernice za osiguranje standarda kvaliteta statističkog poslovnog procesa

Opće smjernice, koje treba slijediti, kako kod uspostave novog tako i kod ažuriranja već postojećeg istraživanja.

1. Kod efikasne uspostave pojedinog procesa ključno je pravovremeno i pravilno planiranje. Prije svega, treba na odgovarajući način ocijeniti izvodljivost procesa,

obzirom na raspoložive ljudske i materijalne resurse, te pripremiti provodiv vremenski plan.

2. Sve faze procesa treba jasno i detaljno dokumentirati. Dokumentacija mora sadržavati kako sadržajni tako i tehnički dio. U dokumentaciji treba u što većoj mjeri biti uključena ocjena (vrednovanje, evaluacija) pojedinih faza procesa.
3. Treba osigurati da se u što većoj mjeri koriste opći standardi i primjeri dobre prakse. Osnova za osiguranje korištenja standardnih alata i postupaka jeste njihovo jasno i pregledno definiranje, te redovno obavještavanje svih zaposlenih (uključenih u provođenje statističkih procesa) o postojanju takvih standarda. To najlakše postizemo organiziranjem radionica, seminara i drugih oblika internog obrazovanja i obavještavanja.

STANDARDI KVALITETA STATISTIČKOG POSLOVNOG PROCESA

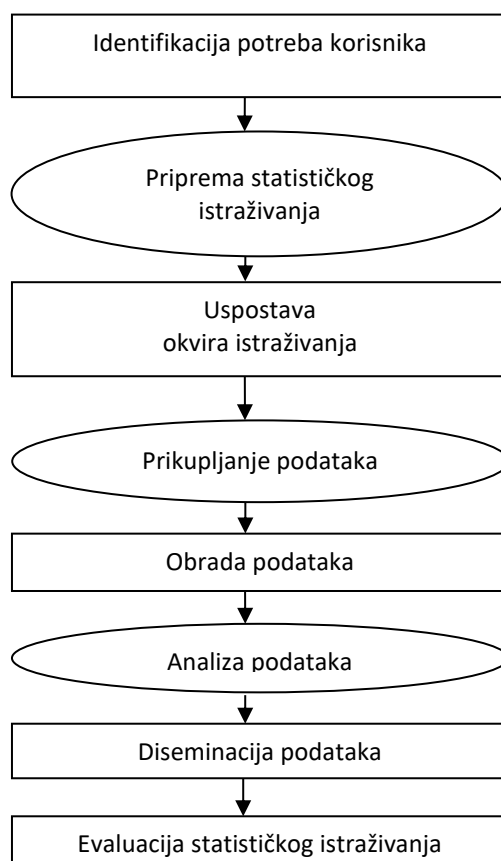
Statistički proizvodi u okviru proizvodnje službene statistike pokrivaju mnoštvo relevantnih socio-ekonomskih tema.

U zavisnosti od tema, metode izrade statističkih proizvoda, korišteni koncepti i definicije, kao i prikaz rezultata mogu biti različiti. Međutim, kao kod svake proizvodne forme postoje zajednički osnovni proizvodni koraci, koji leže u osnovi najvećeg broja službenih statistika.

U osnovnoj strukturi ovog dokumenta nalazi se opći poslovni procesni model, koji je neznatno prilagođena verzija međunarodno prihvaćenog Generičkog modela statističkog poslovnog procesa (*Generic Statistical Business Process Model - GSBPM*).

Model obuhvata 8 faza/procesa koji se dalje dijele na različit broj podfaza/podprocesu. Poglavlja i podpoglavlja u ovom dokumentu slijede strukturu općeg modela. U svakom poglavlju se daje kratak opis ukupnog procesa.

Sva podpoglavlja su sastavljena iz dva seta: prvi sadrži opći opis podprocesu a drugi smjernice za osiguranje kvaliteta procesu i podprocesu, koje će poslužiti kao kontrolni popis komponenti kvaliteta, na koje se treba obratiti pažnja prilikom provođenja pojedinih koraka.



Slika 3. Model općeg statističkog poslovnog procesa

1 IDENTIFIKACIJA POTREBA KORISNIKA

Kod identifikacije potreba i analize zahtjeva za statističkim podacima, pažnja uglavnom treba biti usmjerena prema vanjskim korisnicima statističkih podataka, kojima su podaci neophodni za donošenje odluka i analize, kao i zahtjevima EU regulativa. Analiza potreba korisnika i zahtjeva EU regulativa pruža temelj za planiranje budućih istraživanja.

Očekivanja i potrebe korisnika statističkih rezultata su u stalnom porastu i sve specifičnije iz razloga stalnih društvenih promjena, globalnih procesa i novih inicijativa na međunarodnom i nacionalnom nivou. Potrebe za podacima utvrđujemo u stalnom kontaktu sa korisnicima, pri tome uzimajući u obzir mogućnosti i okvire djelovanja statističkih institucija.

Trenutno je pažnja više usmjerena na harmonizaciju statistike BiH sa EU regulativama u cilju međunarodne uporedbe.

1.1 Utvrđivanje potreba za podacima

Utvrđivanje potreba za podacima se inicira, kada podaci još uvijek ne postoje, ili kada postojeći podaci nisu dovoljni, kako bi se zadovoljile sve korisničke potrebe za podacima.

Potreba za podacima dolazi od raznih korisnika; ministarstava i vladinih službi, CBBiH, od nacionalnih i međunarodnih institucija, te od stručne i šire javnosti.

U postupku utvrđivanja potreba za podacima saznajemo ono što korisnici očekuju od nadležnih statističkih institucija i koja njihova očekivanja za podacima se mogu osigurati. U tu svrhu, u ovaj proces treba uključiti konsultacije sa svim zainteresiranim korisnicima, kroz različite oblike saradnje.

SMJERNICE

- Kod utvrđivanja potreba za podacima potreban je cjelovit i sistematičan pristup, uz uključivanje svih zainteresiranih korisnika kao i zahtjeva EU regulativa.
- Potrebno je redovno organizirati različite oblike saradnje sa korisnicima, pri kojima se statističke institucije upoznaju sa potrebama korisnika.
- Prije svega treba doći do informacija, koje podatke korisnici trebaju, kada ih trebaju, za koje namjene i na kakav način.
- Prilikom donošenja odluke o novom prikupljanju podataka potrebno je procijeniti važnost pokrivanja određene potrebe za podacima. Treba »izvagati« troškove i koristi koje će rezultirati prikupljanjem podataka.

1.2 Provjera i pregled izvora – raspoloživosti podataka

Prije donošenja odluke o uvođenju novog statističkog istraživanja ili redizajniranja postojećeg istraživanja, neophodno je pregledati sva postojeća statistička istraživanja (koja se već provode), kao i sadržaj administrativnih izvora, te utvrditi - da li su ovi statistički podaci na raspolaganju već u postojećim izvorima.

Ako su na raspolaganju, slijedeći korak je ispitivanje usklađenosti postojećih izvora s potrebama novog istraživanja, odnosno ograničenja koja onemogućavaju korištenje ovih izvora kako bi se zadovoljili novi zahtjevi za podacima (razlike u metodologiji, periodici, svrsi prikupljanja podataka, itd.).

Pregled i proučavanje izvora provodi se prvenstveno s ciljem donošenja odluke o tome - mogu li se, i u kojoj mjeri administrativni izvori koristiti kao neposredni izvor podataka.

Ukoliko u pregledanim i dostupnim izvorima nema statističkih podataka koje korisnici zahtijevaju, potrebno je redizajnirati postojeća statistička istraživanja, ili uvesti novo statističko istraživanje.

SMJERNICE

- Potrebno je ispitati – da li je za dobijanje zahtjevanih podataka dovoljno dopuniti već postojeće istraživanje, ili možemo koristiti administrativni izvor, ili je potrebno planirati novo istraživanje.
- Ispitati kvalitet administrativnih izvora, u slučaju adekvatnog kvaliteta administrativni izvor ima apsolutnu prednost nad direktnim prikupljanjem.
- Korisnike podataka treba posebno upozoriti na metodološke posebnosti, pogotovo ako su statistički podaci u potpunosti preuzeti iz administrativnih izvora.
- Potrebno je osigurati poštivanje statističke povjerljivosti te fizičku i informacijsku zaštitu podataka iz administrativnih izvora (ograničenje pristupa podacima, praćenje pristupa).
- Sa vlasnicima administrativnih izvora treba sklopiti sporazum o prijenosu podataka statističkoj instituciji. U sporazumu se utvrđuje sadržaj prenesenih podataka, periodika i način prijenosa.
- Pri propisivanju i uspostavi administrativnih izvora, kao i predlaganju promjena u postojećim administrativnim izvorima, treba osigurati sudjelovanje nadležnih statističkih institucija u tim aktivnostima.

1.3 Priprema i prihvatanje poslovnog slučaja (business case)

Za uspješno provođenje statističkog istraživanja, potrebno je dobro isplanirati potrebne ljudske i materijalne resurse, te odrediti vremenski plan za obavljanje svih poslova.

Precizno planiranje sredstava i utvrđivanje liste aktivnosti i rokova je ključ za efikasno provođenje statističke aktivnosti. Svako statističko istraživanje treba unijeti u godišnji Plan rada (program statističkih istraživanja).

Istraživanje koje se tek uspostavlja uključujemo u Plan rada kao pilot istraživanje, a istraživanje koje se planira redovno provoditi, uključujemo u Plan rada kao redovno istraživanje.

Prije početka provođenja statističkog istraživanja potrebno je pripremiti kadrovski i finansijski plan, te isti pravovremeno uključiti u pripremu budžeta, kao i planirati potrebe za dodatnim materijalnim sredstvima. Treba »izvagati« troškove i koristi koje će rezultirati prikupljanjem podataka.

Prije početka provođenja istraživanja potrebno je izvesti probno/pilot istraživanje, pomoću kojeg možemo utvrditi neke metodološke aspekte istraživanja. Potrebno je jasno definirati ciljeve probnog istraživanja, jer treba prilagoditi dizajn uzorka za istraživanje.

Dizajn uzorka za probno istraživanje zavisi od namjene probnog istraživanja. Kod probnog istraživanja možemo testirati cijeli proces ili samo dio procesa, a najčešće testiramo *pitajna, potrebnu veličinu uzorka, dizajn uzorka i način prikupljanja podataka*.

Veličina uzorka je obično prilično manja nego kod redovnog istraživanja. Nakon provedenog probnog istraživanja rezultate detaljno analiziramo i na osnovu istih određujemo naredne postupke u provođenju istraživanja.

SMJERNICE

- Potrebno je jasno definirati ciljeve probnog istraživanja.
- U probnom istraživanju treba testirati različite metodološke aspekte: formulaciju pitanja, stopu odziva jedinice i varijable, itd.
- Rezultati iz probnog istraživanja se obavezno koriste kod pripreme glavnog istraživanja.
- Kod testiranja upitnika neophodan je neposredan kontakt sa izvještajnom jedinicom.
- U probnom istraživanju se po potrebi testiraju i postupci za statističku obradu i tabeliranje podataka.
- Utvrditi - da li je zahtjevani projekat u osnovi provodiv.
- Način prikupljanja podataka treba se unaprijed definirati i razjasniti. Pri tome elektronska forma prijenosa podataka ima prednost (kad god je to moguće). U slučajevima korištenja sekundarnih izvora, vlasnika podataka treba uključiti u proces planiranja.
- Za svaki projekat se najprije mora izračunati preliminarna procjena troškova tj. neophodna budžetska sredstva. Da bismo planirali realne vrijednosti za očekivane troškove, potrebno je uzeti u obzir iskustvene vrijednosti iz sličnih projekata koji su prije realizirani, odnosno uporedivih projekata.

2 PRIPREMA STATISTIČKOG ISTRAŽIVANJA

Temeljito planiranje je početna faza procesa proizvodnje, unaprijed osmišljenih budućih aktivnosti za pripremu i strukturiranje toka pojedinih podprocesa. Planiranje je neophodna pretpostavka za dobijanje kvalitativno visoko vrijednog krajnjeg proizvoda. Ovo, opće važeće pravilo vrijedi takođe i za proces proizvodnje statistike.

Planiranje se treba usmjeriti prema zadanom cilju, koji s jedne strane mora zadovoljiti legalističke norme, s druge strane zahtjeve za poštivanjem standarda kvaliteta. Cilj planiranja je operacionalizacija predmeta statističkog istraživanja.

Aktivnosti koje poduzimamo u postupku planiranja su:

- Specifikacija osnovnih teza u pogledu predmeta, cilja i rezultata istraživanja sa bitnim karakteristikama istog. Pomoću primjene određenih koncepata i metoda pristupamo izradi i definiranju projektne strukture faza istraživanja kao i raspoloživih resursa;
- Kada se donese odluka o statističkom istraživanju, poduzima se detaljno planiranje i daljnje diferenciranje pojedinih faza istraživanja. Moraju se uzeti u obzir i planirati svi bitni aspekti za provođenje istraživanja, kao što su npr. vremenski rokovi i trajanje, potrebni ljudski i tehnički resursi, pojašnjenje pojedinih metodoloških pitanja.
- Napraviti plan aktivnosti na provođenju istraživanja (priprema obrazaca i uputstava, definiranje okvira i izrada uzorka za pilot istraživanje, priprema varijabli i izlaznih tabela, provođenje pilot istraživanja, analiza rezultata pilota i priprema samog istraživanja)

U postupku planiranja posebna pažnja se poklanja samom procesu prikupljanja podataka. Dakle, treba odgovoriti na pitanje - odakle možemo dobiti potrebne podatke i kako ih koristimo u primjeru nekog istraživanja?

Nadležne statističke institucije najprije i bezuvjetno i dokle god je moguće, koriste sve raspoložive podatke iz drugih izvora, prije nego se ide na provođenje direktnog istraživanja.

Svaka institucija koja vodi javni registar, kao i vlasnici službenih evidencija i statističkih podataka su u obavezi da stave na raspolaganje podatke koji su neophodni za izvršavanje Plana i Programa statističkih institucija u BiH. Samo u slučajevima kada obezbjeđivanje podataka nije moguće na ovaj način, može se urediti provođenje statističkog istraživanja putem direktnog istraživanja.

Kod planiranja nekog istraživanja, potrebno je, što je moguće manje opteretiti izvještajne jedinice. Kad god je moguće, istraživanja treba provoditi na bazi dobrovoljne saradnje i

putem uzorka, i istraživanjima na bazi uzorka dati prednost nad istraživanjima s punim obuhvatom.

Raspoloživost budžetskih sredstava je osnovna pretpostavka da se uopće mogu poduzeti aktivnosti na provođenju statističkog istraživanja. Preliminarna procjena troškova mora obuhvatiti sve neophodne troškove koji su neophodni za provođenje određenog istraživanja.

2.1 Dizajniranje proizvoda statističkog istraživanja

Ovaj podproces sadrži detaljno oblikovanje statističkih rezultata, proizvoda i usluga koje treba proizvesti, uključujući povezane poslove razvoja i pripremu sistema i alata koji se koriste u fazi "Diseminacije".

Ovdje se takođe izrađuju postupci koji upravljaju pristupom svim povjerljivim rezultatima. Rezultati bi trebali biti dizajnirani kako bi se pratili postojeći standardi kad god je to moguće, tako da inputi u ovaj proces mogu uključivati metapodatke iz sličnih ili prethodnih prikupljanja podataka, međunarodnih standarda i informacija o praksama u drugim statističkim institucijama.

Ovaj je podproces usmjeren na oblikovanje rasporeda i sadržaja proizvoda (statističke tabele, pokazatelji, posebne publikacije itd...) planiranih za diseminaciju. Izrada proizvoda slijedi interno utvrđene procedure i politiku diseminacije.

2.2 Priprema metodologije za prikupljanje podataka i provedbu istraživanja

Ovaj podproces uključuje izradu svih potrebnih metodologija (metode, instrumenti za prikupljanje, varijable, definicije, upute, sporazumi i ugovori s davaocima podataka, sadržaj upitnika, plan diseminacije itd.). Priprema opisa metapodataka o prikupljenim i izvedenim varijablama i klasifikacijama ključan je preduvjet za sljedeće faze. Trebamo koristiti (koliko je moguće) usklađene koncepte, varijable i klasifikacije, ali definicije istih moramo prilagoditi specifičnim potrebama. Sve definicije koncepata, varijabli i klasifikacija treba dokumentirati, kao i sva moguća odstupanja od preporučenih standarda. Utvrđuju se najprikladnije metode i instrumenti za prikupljanje podataka. Aktivnosti ovise o načinu prikupljanja podataka (CAPI, PAPI, CATI, CAWI) uključujući testiranje instrumenata. Izrađuju se svi formalni ugovori o dostavljanju podataka (npr. tehnički protokoli).

U ovom podprocesu određujemo i ciljnu populaciju. Pri tom trebamo uzrati u obzir sljedeće: vrstu jedinice posmatranja i karakteristike koje definiraju populaciju; geografsku lokaciju jedinice posmatranja i vremensko (referentno) razdoblje na koje se odnose karakteristike populacije koja nas interesuje.

2.2.1 Koncepti, definicije i klasifikacije

Koncepti službene statistike čine temelj za numerički prikaz općih pitanja koja se odnose na socijalne, ekonomske ili druge društveno-političke fenomene. Statistička istraživanja

temelje se na odgovarajućim, u najvećoj mjeri međunarodno uporedivim harmoniziranim konceptima.

Definicije predstavljaju operacionalizirane opise unutar koncepata, koji su neophodni za konkretnu izradu statistika (npr. definicije varijabli). Dok koncept čini opći temelj za mnoge statističke projekte, definicija jasno određuje kako se utvrđuju relevantni temelji (npr. koje varijable se istražuju) za realiziranje nekog statističkog projekta.

U tom kontekstu, za ocjenu kvaliteta neke statistike od značaja je još jedan aspekt. Naime, različite statistike opisuju određenu temu sopstvenim definicijama, koje su bazirane na različitim konceptima. Na ovaj način, može doći do različitih rezultata i pri tome do problema usklađenosti. Da bismo korištene koncepte i definicije učinili transparentnim za korisnike statističkih proizvoda, moramo dodatno staviti na raspolaganje dovoljno *metapodataka*.

Klasifikacije predstavljaju osnovni alat za izradu službenih statistika. Nadležne statističke institucije primjenjuju određen broj nacionalnih i međunarodnih klasifikacija i nomenklatura.

SMJERNICE

- Definicije moraju biti razumljive i stavljene korisnicima na raspolaganje.
- Sve koncepte, varijable i klasifikacije, koje će biti korištene u istraživanju detaljno definiramo, a posebno opisujemo sva moguća neslaganja sa standardnim konceptima. Takođe, treba opisati moguće razlike između fenomena koje mjerimo i fenomena od interesa za korisnike (npr., razlika između ciljne i posmatrane populacije).
- Svi koncepti i definicije trebaju se temeljiti na međunarodnim standardima.
- Sve relevantne definicije, koncepti i klasifikacije trebaju biti objašnjene u »Izveštaju o kvalitetu« za pojedina statistička istraživanja. Ovo se posebno odnosi na namjenu statistike, statističke jedinice, prikupljene varijable, izračunate relativne brojeve i pokazatelje, korištene klasifikacije, itd... .
- Nove klasifikacije, odnosno revizije i modifikacije postojećih klasifikacija, moraju se odmah unijeti u bazu podataka klasifikacija.

2.2.2 Statistički poslovni registar

Statistički poslovni registar ima centralnu ulogu za sve poslovne statistike što je pretpostavka koherentnosti podataka, te se iz tih razloga registar ubraja u statističke proizvode. Statistički poslovni registar služi kao okvir za poslovnu (biznis) populaciju, za povezivanje administrativnih podataka i njihovo transformiranje za statističke potrebe, povezivanje mikro podataka na statističkoj jedinici iz različitih izvora i istraživanja. Vođenje statističkog poslovnog registra je jedno od najvažnijih i temeljnih poslova za službenu statistiku. Svrha je prikupiti, uspostaviti i održavati registerski relevantne podatke. Statistički poslovni registar se razlikuje od administrativnog registra po tome što se on uspostavlja i vodi prema EU Regulativi 177/2008 i isključivo za statističke namjene.

SMJERNICE

- Definicija jedinica registra, varijabli za jedinice registra, koncept vođenja, održavanje i ažuriranje registra mora biti dokumentirano i usaglašeno sa EU regulativom 177/2008.
- Pri ažuriranju registra moraju biti korišteni raspoloživi administrativni izvori podataka.
- Svaka jedinica u registru mora imati svoj identifikacijski broj, koji služi za internu upotrebu, i jednom dodijeljen identifikacijski broj ne smije više biti promijenjen.
- Podaci u registru su individualni podaci. Pridržavanje načela zaštite i povjerljivosti podataka u registru mora biti zagarantirano.
- U cilju praćenja pojedinačnih jedinica iz registra, neophodno je pratiti historijska stanja registra, koja se mogu rekonstruirati na određeni dan.
- Registar služi kao okvir za izbor jedinica posmatranja za statistička istraživanja. U cilju reduciranja opterećenosti potencijalnih izvještajnih jedinica, treba voditi evidenciju, koja jedinica iz registra je korištena u kojem statističkom istraživanju.
- Povratne informacije od strane svake organizacijske jedinice, koja koristi podatke registra služi za kontinuirano poboljšanje i održavanje istog.

2.2.3 Priprema preuzimanja administrativnih izvora

Korištenje administrativnih izvora podataka ima niz prednosti za službenu statistiku. Administrativni podaci već postoje i ne moraju se prikupljati na konvencionalan način. Pri tome, nema troškova i dugotrajnog rada, kao kod klasičnog istraživanja. Sve ovo vodi ka znatnom rasterećenju izvještajnih jedinica. Ovdje je također važno pomenuti da je davanje ovih podataka (ako su raspoloživi) besplatno i uglavnom se obavlja elektronskim putem, te su ovi podaci neposredno na raspolaganju za daljni proces obrade.

Mogućnosti primjene administrativnih podataka su mnogostruke i uključuju:

- direktno korištenje kao glavnog izvora podataka,
- korištenje kao dodatnog izvora podataka (dopuna specifičnih varijabli),
- korištenje kao okvira za izbor uzorka,
- mogućnost poređenja za rezultate istraživanja.

U cilju efektivnog korištenja ovih podataka, neophodno je također prepoznati i potencijalne slabosti pojedinih izvora podataka. Jedna od najvažnijih razlika naspram klasičnog istraživanja je u tome što statističke institucije nemaju kontrolu u proizvodnji tih podataka.

Mogući problemi su:

- različiti koncepti, definicije i klasifikacije,
- problemi obuhvata,

- slabo održavane varijable (varijable koje nisu relevantne za administrativne potrebe),
- podaci nisu dostupni pravovremeno.

SMJERNICE

- Potencijalnog vlasnika administrativnih podataka treba uključiti već u okviru procesa planiranja i procesa analize prikupljanja podataka.
- Potrebno je ispitati i prije samog korištenja administrativnih izvora podataka, da li koncepti i definicije korišteni u istima, korespondiraju sa zahtjevanom statističkom namjenom, kao i da li u dovoljnoj mjeri zadovoljavaju zahtjevene koncepte i definicije.
- Sve procesne korake koje poduzimamo (da bi administrativni podaci bili od koristi), treba unaprijed planirati i testirati. Potrebno je dokumentirati sve mjere, koje se odnose na transformaciju podataka iz administrativnih izvora za statističke namjene.
- Moraju se uspostaviti i kontinuirano održavati kontakti sa vlasnicima administrativnih izvora podataka. U tom cilju sklapati protokole koji osiguravaju redovnu i pravovremenu dostavu podataka. Ukoliko su podaci revidirani od strane vlasnika podataka, treba omogućiti prenos i tako korigiranih podataka.
- Putem stalnog kontakta sa vlasnicima administrativnih podataka treba osigurati redovne povratne informacije o kvalitetu dostavljenih podataka.
- Kod administrativnih podataka radi se uglavnom o evidencijama pojedinačnih setova podataka, koje se odnose na fizičke ili pravne osobe. Iz tih razloga, pri obradi ovih podataka mora se maksimalna pažnja posvetiti zaštiti istih.
- Korištenje administrativnih izvora mora biti predložen što transparentnije korisnicima. Sve implikacije, koje proizilaze iz korištenja administrativnih izvora, treba opisati u Izvještajima o kvalitetu. Ovo se posebno odnosi na pitanja o obuhvatu, pravovremenosti i aktuelnosti koje su posljedica konceptualnih razlika.

2.2.4 Priprema metodologije izbora jedinice posmatranja

Nakon što odredimo ciljnu populaciju, moramo obezbjediti listu jedinica ove populacije sa što više njihovih karakteristika. Listu ovih jedinica nazivamo okvir uzorka.

Varijable okvira uzorka, koje opisuju karakteristike jedinica, nazivamo pomoćnim varijablama. Kod pripreme metodologije izbora jedinica posmatranja uzimamo u obzir način prikupljanja podataka. Zahtjevane podatke možemo dobiti iz različitih administrativnih izvora, ili ih modeliramo iz postojećih statističkih izvora ili posrednim odnosno neposrednim kontaktom sa jedinicima posmatrane populacije.

Bez obzira na činjenicu kako prikupljamo podatke, statističko istraživanje možemo provoditi kao *istraživanje na bazi uzorka* ili *istraživanje s punim obuhvatom*.

Istraživanje s punim obuhvatom je uvijek izuzetno dugotrajno i stoga skupo. Međutim, ono je važno kako bi se mogle odrediti osnovne populacije bez praznina u dužim vremenskim intervalima. Ovo je jedini način da se kreiraju osnove selekcije i

ekstrapolacioni okviri za uzorke. Oni također mogu dati rezultate koji su detaljno strukturirani i regionalno i stručno.

Većina statističkih istraživanja se ne provodi kao potpuno istraživanje, već kao djelomično istraživanje na bazi slučajnog uzorka, iz razloga troškova i rasterećenja ispitanika.

Budući da nisu uzeti u obzir svi elementi osnovne populacije, rezultat uzorka može biti samo aproksimacija tražene, nepoznate vrijednosti u osnovnoj populaciji. Ograničavanje mjerenja na izbor nosilaca obilježja može uzrokovati različite probleme koji se moraju uzeti u obzir pri dizajniranju uzorka, tj. dizajnu izbora uzorka.

Problemi kod dizajna uzorka:

Selektivnost uzorkovanja – Problem selektivnosti uzorkovanja često se javlja kada se prilikom uzimanja uzorka preferiraju određeni nosioci obilježja ili je vjerojatnije da će biti uključeni u uzorak od drugih. Na primjer, poznato je da ljudi sa srednjim prihodima češće dobrovoljno sudjeluju u anketi, dok su vrlo bogati ili vrlo siromašni manje voljni.

Ako se anketar okrene nasumično odabranim ljudima koje je lako regrutovati za istraživanje, u uzorak će biti uključen nesrazmjerno veliki broj ljudi sa srednjim prihodima, dok će bogati i siromašni biti nedovoljno zastupljeni.

Dakle, ako želite da date izjave o prosječnom prihodu ukupne populacije, morate se pobrinuti da siromašne i bogate osobe podjednako budu uključene u uzorak. U suprotnom, zbog selektivnosti uzorka, rezultati uzorka mogu biti sistematski pristrasni i iz tog razloga netačni.

Problem selektivnosti uzorka uvijek se javlja kada je vjerovatnoća izbora povezana sa obilježjima ispitivanja.

Slučajna greška - Čak i ako ne postoji veza između vjerovatnoće izbora i obilježja ispitivanja, rezultati uzorka uvijek su podložni određenoj netačnosti. Na primjer, posebno veliki broj bogatih domaćinstava ili posebno veliki broj siromašnih domaćinstava može biti uključen u uzorak sasvim slučajno.

Ako uzmete veliki broj uzoraka iste veličine iz ukupne populacije i izračunate srednju vrijednost za određeno obilježje ispitivanja u svakom uzorku, tada će se na primjer prosječne vrijednosti određene u uzorcima raspršiti oko stvarne srednje vrijednosti ukupne populacije.

Prosječno očekivano odstupanje srednje vrijednosti uzorka od srednje vrijednosti ukupne populacije naziva se slučajna greška.

SMJERNICE

- Kod određivanja okvira uzorka treba procijeniti koji su izvori na raspolaganju (registri, baze podataka ...) i kako ih možemo međusobno povezati.

- Ako procijenimo da sastav (kompozicija) okvira ne zadovoljava kvalitet, potrebno je kod sastavljanja upitnika uključiti dodatna pitanja, pomoću kojih možemo lakše procijeniti kvalitet okvira uzorka. Ove rezultate treba uzimati u obzir kod izrade okvira uzorka za naredno referentno razdoblje.
- Kod uzorkovanja poslovnih subjekata, »najveća« preduzeća ulaze u potpun obuhvat (cenzus), u zavisnosti od istraživanja i posmatranih jedinica, a od ostalih jedinica iz okvira se bira slučajni uzorak. Kriteriji za određivanje »najvećih« preduzeća su broj zaposlenih, prihod ili kombinacija istih.
- Prednost uvijek treba dati što jednostavnijem dizajnu uzorka.
- Odluku o izboru kompleksnijeg dizajna uzorka treba pažljivo razmotriti. Korištenje višestepenog stratifikovanog dizajna uzorka može smanjiti troškove istraživanja, ali u isto vrijeme smanjuje preciznost procjene.
- Kod odlučivanja o veličini uzorka, treba uzeti u obzir predviđene stope odgovora kao i rezultate analize neodgovora iz prethodnog referentnog razdoblja (ako su podaci dostupni). Treba procijeniti, koliki odziv osigurava dovoljan kvalitet rezultata.

2.3 Priprema izvora podataka za izradu okvira uzorka

Veoma važno je razumijevanje razlike između ciljne populacije i okvira uzorka. *Ciljna populacija* statističkog istraživanja je skup jedinica, čije određene karakteristike posmatramo i na koju se odnose svi rezultati istraživanja. *Okvir uzorka* je stvarni popis jedinica ciljne populacije, koji imamo na raspolaganju u trenutku pripreme istraživanja i koji koristimo za izbor jedinica ciljne populacije.

Prije nego odredimo okvir uzorka, potrebno je tačno definirati ciljnu populaciju (osobe, domaćinstva, stanovi, farme, preduzeća ...) i to:

- u pogledu karakteristika koje nas zanimaju - odredimo ko ili šta je ciljna populacija i koje uvjete treba jedinica udovoljiti da bude dio ciljne populacije;
- treba odrediti geografsku lokaciju jedinice;
- treba odrediti referentno razdoblje (vrijeme) u kojem nas zanimaju karakteristike populacije.

Primjer:

Ciljna populacija su nam *osobe koje žive u privatnim domaćinstvima*. Na određeni datum, ove osobe su bile stanovnici BiH i bile su na ovaj datum stare 15 godina i više. Pri tome treba uvažavati definiciju »stanovnika BiH«; definirati uvjete koje osoba treba ispunjavati da je član privatnog domaćinstva. Ovim uvjetima definiramo ciljnu populaciju, i samim tim i uvjete koje mora ispunjavati neka jedinica da bi bila obuhvaćena istraživanjem.

U statističkom istraživanju razlikujemo (samo u nekim istraživanjima su sva tri svojstva sadržana u jednoj jedinici, što često to nije slučaj):

- *jedinicu uzorkovanja* (jedinica iz okvira uzorka, koja je izabrana u uzorak),

- *jedinicu posmatranja* (jedinica o kojoj prikupljamo podatke i stoga je dio ciljne populacije) i
- *izvještajnu jedinicu* (jedinica koja nam daje/ izvještava podatke koje prikupljamo).

Primjer:

Ciljna populacija u Anketi o radnoj snazi (LFS) su sve osobe starosti 15 i više godina. **Jedinica uzorkovanja** je stambena jedinica (sa određenom adresom), preciznije domaćinstvo koje *možda živi na izabranoj adresi, a koje je živjelo u vrijeme ažuriranja okvira*. **Jedinica posmatranja** je *domaćinstvo zatečeno na datoj adresi*, a **jedinice izvještavanja** su članovi zatečenog domaćinstva (stariji od 15 godina).

Kada znamo, šta je naša ciljna populacija i kako možemo pristupiti do jedinica te populacije, onda pristupamo određivanju okvira uzorka.

Ako korišteni izvor podataka pokriva referentno razdoblje istraživanja, tada tačnije obuhvatamo populaciju u referentnom razdoblju nego u slučaju kad takvog izvora nemamo na raspolaganju.

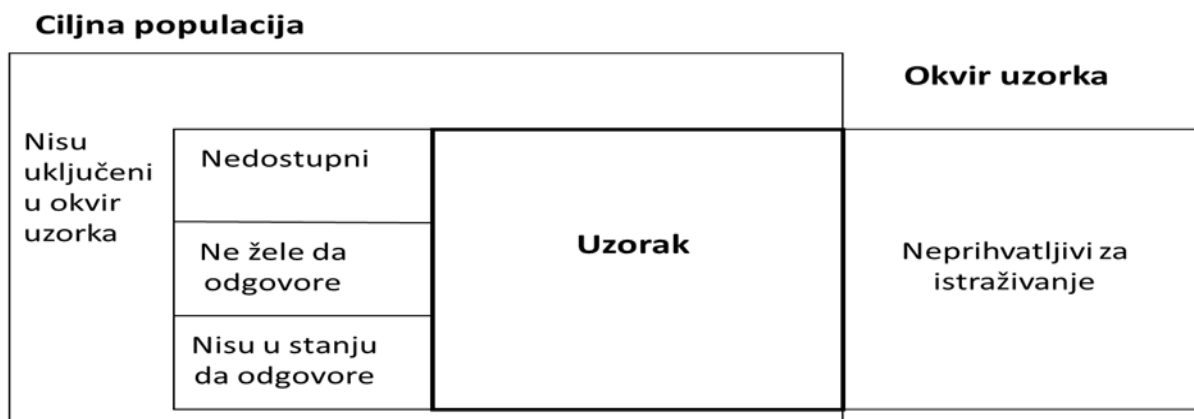
Obično izaberemo jedan glavni izvor podataka za okvir uzorka, a zatim poboljšamo kvalitet okvira sa dopunskim izvorima. Za glavni izvor izaberemo onaj koji nam omogućava da što bolje obuhvatimo ciljnu populaciju. Sa dopunskim izvorima dopunjujemo glavni izvor o jedinicima posmatranja.

Glavni izvor podataka kod poslovnih statističkih istraživanja je u pravilu *Statistički poslovni registar* (SPR), koji sadrži popis svih preduzeća.

Kod istraživanja, u kojima prikupljamo podatke o osobama i domaćinstvima, tj. za anketne upitnike koji se koriste za anketiranje na terenu, glavni izvor podataka je Popis stanovništva, domaćinstava i stanova.

Ponekad, jedan izvor podataka nije dovoljan, tako da moramo izraditi okvir uzorka iz raznih izvora (administrativni izvori, popis, druga statistička istraživanja). To je posebno slučaj, kada različiti izvori podataka pokrivaju različita razdoblja i različite dijelove populacije. Jako je važno, da u svim izvorima podataka imamo jedinstvenu identifikaciju jedinice. Za nas relevantni samo izvori podataka, u kojima možemo ovu identifikaciju utvrditi. U ovom slučaju se uvijek jedan izvor podataka određuje kao osnovni izvor.

Ciljna populacija i okvir uzorka se samo rijetko podudaraju u svim jedinicama. Postoje jedinice, koje su dio ciljne populacije, a nema ih u popisu jedinica okvira uzorka, a isto tako su u okviru uzorka jedinice, koje u stvarnosti nisu dio ciljne populacije. Oba slučaja uzrokuju greške obuhvata, koje u velikoj mjeri određuju kvalitet konačnih rezultata.



Slika 4: Šematski prikaz odnosa okvira uzorka, ciljne populacije i uzorka

SMJERNICE

- Referentna razdoblja, na koja se odnose podaci u izvorima, koje ćemo koristiti pri izradi okvira uzorka, trebaju biti što bliže referentnom razdoblju statističkog istraživanja. Ako su referentna razdoblja u različitim izvorima različita, razlike treba dokumentirati.
- Treba provjeriti, da li su varijable, koje se pojavljuju u više raličitih izvora, usklađene u pogledu metodoloških definicija. Ako se utvrdi da se ovim definicijama razlikuju odnosno ne podudaraju, te razlike treba evidentirati i dokumentirati.
- Okvir uzorka treba sadržavati ove podatke: jedinstveni identifikator jedinice; kontaktne podatke; klasifikacijske varijable. Korisno je, ako okvir uzorka sadrži datum stanja okvira i povezujuće varijable (kojima povezujemo okvir uzorka sa drugim izvorima podataka).
- Treba voditi računa o tome da svi izvori podataka prije kombiniranja (udruživanja) budu prečišćeni u pogledu formalnih kontrola (statusa aktivnosti pravnog subjekta, važećih šifarnika, odgovarajućeg raspona vrijednosti varijabli, problema duplih jedinica ...).
- Izvore podataka koje koristimo za izradu okvira (isto tako i zapise koje nećemo koristiti za izradu okvira), treba na odgovarajući način dokumentirati, uz kratak opis ovih izvora.
- Ako želimo da greške obuhvata budu što manje, potrebno je pri pripremi okvira uzorka koristiti sve izvore podataka koji su nam na raspolaganju, jer svaki od njih doprinosi tome da je okvir uzorka što konzistentniji sa ciljnom populacijom. Neslaganja između okvira uzorka i ciljne populacije treba dokumentirati.

2.4 Priprema metodologije statističke obrade podataka

Statistička obrada podataka obuhvata sve postupke koji se poduzimaju po okončanoj fazi prikupljanja ili preuzimanja podataka, s ciljem da konačni statistički rezultati što tačnije odražavaju karakteristike posmatrane populacije.

Različita statistička istraživanja u toku same implementacije obuhvataju različite postupke statističke obrade, a najčešći su:

2.4.1 Postupak sa neodgovorom

Kod provođenja istraživanja (osim kod nekih istraživanja s manjim brojem jedinica posmatranja) susrećemo se sa problemom neodgovora i problemom nedostajućih podataka (djelomično ili potpuno) kod dijela posmatranih jedinica.

Postupke za reduciranje udjela neodgovora treba uključiti u aktivnosti povezane sa prikupljanjem podataka, a u fazi planiranja statističke obrade je nužno predvidjeti, koji postupci će se koristiti po završetku prikupljanja podataka s ciljem da se u što većoj mjeri smanji uticaj neodgovora na konačne statističke rezultate (prije svega u odnosu na uzrokovanu pristrasnost).

Problem neodgovora se može pojaviti takođe i kod administrativnih izvora podataka.

2.4.2 Postupak uređivanja (editovanja) podataka

Uređivanje podataka označava postupke za traženje i korigiranje grešaka u podacima. Kod planiranja postupaka uređivanja podataka treba izabrati postupke, koji će u što većoj mjeri smanjiti zahtjevano vrijeme uređivanja i troškove.

Postupke »ručnog« editovanja treba poboljšati uvođenjem modernijih pristupa, kao što su *selektivno i automatsko uređivanje podataka*.

Postupak uređivanja identifikuje nedosljednosti u podacima. One su obično posljedica grešaka u originalnim podacima, ali mogu biti rezultat procesnih grešaka kod šifriranja i unosa podataka.

Stopa editovanja podataka je pokazatelj kvaliteta prikupljanja i obrade podataka, a ne kvaliteta konačnih podataka. Stopu editovanja treba izračunavati za ključne varijable.

2.4.3 Postupak agregiranja i tabeliranja

Ovim postupkom, iz odgovarajućih statistički uređenih podataka na mikro nivou, dobijemo statističke agregate (statistike) koji su konačni proizvod statističkog procesa.

Potrebno je već u fazi planiranja predvidjeti, koje statistike će se izračunavati, a zatim tačno i jasno napisati pravila za njihovo izračunavanje.

Takođe je potrebno predvidjeti i nivo detaljnosti tabeliranja rezultata.

Pri svemu tome treba poštivati (moguće) zahtjeve međunarodnih regulativa, zahtjeve domaćih korisnika, ali takođe treba uzeti u obzir ograničenja, koja određuju pravila za zaštitu podataka i pravila o neobjavljivanju rezultata sa nedovoljnim stepenom tačnosti.

Ponekad je potrebno u samom postupku statističke obrade planirati i slijedeće korake (ukoliko to zahtjeva sama priroda statističkog istraživanja ili potrebe korisnika):

- Ako se provodi istraživanje na bazi uzorka, potrebno je predvidjeti postupke za ocjenu greške uzorkovanja. Potrebno je prije svega odlučiti, koji će se pristup

koristiti (analitički pristup, pristup ponovljenog uzorkovanja ...) i koji programski alat;

- Kod finansijskih podataka korisniku se obezbjeđuju tzv. deflacionirani statistički rezultati. Deflacioniranje je postupak, kojim se iz rezultata, koji prikazuju promjene određene pojave (npr. indeks prosječne plaće), otkloni uticaj promjene cijena. Potrebno je već u fazi planiranja predvidjeti, koji pristup i koju metodologiju izračunavanja deflatora koristiti (deflacioniranje na mikro ili na makro nivou);
- Ako se provodi periodično istraživanje, rezultati istraživanja se mogu posmatrati kao vremenske serije, tako da treba predvidjeti da li je potrebno podatke desezonirati. Korištenje ovih postupaka ovisi od tome - da li postoji sezonski uticaj i uticaj broja radnih dana na rezultate.

SMJERNICE

- Kod planiranja statističkih alata koji će se koristiti za provođenje statističke obrade, potrebno je utvrditi - da li već postoje opći programski alati koje bi mogli koristiti, ili bi trebalo razviti rješenja po mjeri samog istraživanja.
- Ako je potrebno uvesti nova programska rješenja, treba izabrati najprikladniji programski alat. Pri tome treba uzeti u obzir i funkcionalne sposobnosti pojedinih programskih alata, kao i standardnu praksu u samim statističkim institucijama.
- Kod postupanja sa neodgovorom izvještajne jedinice treba odlučiti - da li koristiti postupak ponderisanja podataka ili postupak imputacije za sve varijable.

Pri tome, uzeti u obzir:

- broj varijabli koje treba imputirati (metode imputiranja koriste se kao rješenje za problem neodgovora jedinice samo onda, kada je u istraživanju malo varijabli);
- da li se radi o periodičnom istraživanju, kod kojeg se može za imputiranje podataka koristiti historijske podatke pojedinačnih jedinica;
- da li postoje vanjski izvori (npr. iz registra) na mikro nivou za imputaciju.
- Kod postupka uređivanja (editovanja) potrebno je u što većoj mjeri koristiti metode selektivnog i automatskog uređivanja podataka, jer ovo može znatno smanjiti troškove i opterećenje izvještajnih jedinica zbog ponovnog uspostavljanja kontakta s njima.
- Kod određivanja detaljnosti nivoa objavljivanja podataka, treba izbjegavati (ako je moguće) predetaljan nivo objavljivanja, jer to može imati za rezultat veliki broj praznih, skrivenih ili zaštićenih ćelija.

3 USPOSTAVA OKVIRA ISTRAŽIVANJA

Ova se faza uspostavlja i testira proizvodno rješenje do tačke gdje je spremna za upotrebu u "live" okruženju. Rezultat faze "Pripreme statističkog istraživanja" usmjerava izbor procesa, instrumenata, informacija i usluga koji se konfiguriraju u ovoj fazi kako bi se stvorilo potpuno radno okruženje za pokretanje procesa.

3.1 Uspostava kanala i instrumenata za prikupljanje podataka

Prikupljanje može koristiti jedan ili više načina za primanje podataka (npr. lični ili telefonski intervju, papir, elektronski ili web upitnici, itd...). Kanali za razmjenu mogu također biti postupci ekstrakcije podataka za korištenje prikupljenih podataka iz postojećih statističkih i administrativnih setova podataka. Ovaj podproces također uključuje pripremu i testiranje sadržaja i funkcioniranja tog kanala razmjene (npr. testiranje pitanja u upitniku). U ovom podprocesu pripremaju se tehničke specifikacije za izradu ili ažuriranje instrumenata za prikupljanje podataka.

Ovi postupci se provode na temelju konsultacija sa IT odsjekom i programerima. Instrumenti za prikupljanje su izrađeni ili ažurirani u saradnji između metodologa iz predmetnog odsjeka i IT odsjeka. Obavlja se testiranje izrađenih ili ažuriranih softvera za prikupljanje statističkih podataka. Obično ovo obavljaju IT i metodolozi iz predmetnog odsjeka. U nekim slučajevima osoblje za unos podataka ili odabrani ispitanici mogu biti uključeni u testiranje softvera za prikupljanje podataka.

3.1.1 Dizajniranje i testiranje upitnika

Upitnici/obrasci zauzimaju centralno mjesto u procesu statističkog istraživanja i prikupljanja podataka. Upitnici službene statistike sadrže u pravilu isključivo zatvorena pitanja, zatim pitanja sa preddefiniranim mogućim odgovorima, kao i pitanja o činjenicama, dok otvorena pitanja i pitanja o stavovima/mišljenjima o određenim stvarima dolaze samo u iznimnim slučajevima.

Takođe, kod pitanja o činjenicama dizajn pitanja je od centralnog značaja, te formuliranje nekog pitanja sa mogućim modalitetima odgovora neposredno utiče na rezultate

istraživanja. Kvalitet rezultata nekog istraživanja u velikoj mjeri ovisi od kvaliteta ovih koraka, koji se poduzimaju u pripremi samog upitnika.

Upitnik je najčešće podijeljen u više poglavlja (tematskih cjelina), pri čemu se izvještajna jedinica kroz upitnik vodi putem odgovarajućih „filtera“.

Kod dizajniranja samog upitnika od velikog značaja je - da li izvještajna jedinica sama ispunjava upitnik, ili anketar ispunjava odgovore koje daje izvještajna jedinica na pitanja iz upitnika.

Kod pitanja koja imaju više modaliteta odgovora (koji služe kao pomoć) izvještajnoj jedinici je moguće pri intervjuisanju dati da pročita moguće modalitete odgovora. Ova olakšica naravno nije moguća kod telefonskog istraživanja/intervjuisanja.

U posljednje vrijeme sve veći značaj ima davanje podataka putem elektronskog izvještavanja. S tim u vezi, u narednom periodu će se sve više raditi na razvoju elektronskih upitnika, kao veoma važnog metoda provođenja istraživanja i prikupljanja podatka. Elektronski upitnici imaju tu prednost, da je softverski podržano vođenje kroz upitnik putem filter pitanja, tako da nema mogućnosti za greške.

Svi smo svjesni činjenice, da je jednom proslijeđen upitnik na teren, moguće izmijeniti samo u iznimnim situacijama. Iz tih razloga, veliki značaj ima evaluacija upitnika na terenu, tj. njihovo testiranje.

Izbor metoda testiranja ovisi od raznih okolnosti, kao što su: predmet istraživanja, raspoloživi resursi ili vrsta i značaj određenog istraživanja.

Kod detaljnih upitnika, može biti od koristi uvođenje kontrolnih pitanja, pomoću kojih se može ispitati vjerodostojnost dobijenih informacija.

3.1.1.1 Priprema nacrtu upitnika

Kada se pripreme varijable i pitanja, pripremi se nacrt upitnika sa logičkim kontrolama i preskocima. Ako se priprema elektronski upitnik, o tehničkim mogućnostima i standardima se posavjetuje sa programerom. Zatim se izradi nacrt upitnika. Pitanja u upitniku se podijele na smislene setove pitanja.

3.1.1.2 Interni pregled upitnika

Veoma je važno da upitnik pregledamo, prije nego ga testiramo na ciljnoj populaciji. Ovaj interni pregled obuhvata: gramatičku ispravnost pitanja, poboljšanje opširno postavljenih i nerazumljivih pitanja. Interni pregled znači, da tekst upitnika pregledaju osobe, koje nisu neposredno uključene u statističko istraživanje. Osobe, koje pregledaju upitnik, mogu biti: stručnjaci za oblast koja je tema istraživanja, stručnjaci za dizajniranje upitnika, anketari ili predstavnici ciljne populacije.

3.1.1.3 Tehničko testiranje upitnika

Kada prikupljamo podatke elektronskim upitnikom (sa anketarom ili web upitnikom), isti testiramo sa tehničke strane. To uključuje testiranje preskoka i operacija logičkih kontrola.

3.1.1.4 Testiranje i revizija upitnika

Poželjno je, da se prije provođenja pilotnog istraživanja na terenu izvede testiranje upitnika. Ovaj postupak uključuje neformalna testiranja upitnika, kognitivne metode, fokusne grupe, izvještaje anketara, kodiranje ponašanja anketara ili izvještajnih jedinica, pilotna testiranja upitnika.

Dizajniranje upitnika je ponavljajući (iterativni) postupak. Korisno je, da nakon svake veće promjene upitnik iznova testiramo.

SMJERNICE

- Postavke pitanja i instrukcije za popunjavanje upitnika trebaju biti jasne i razumljive za izvještajne jedinice. Kad god je to moguće, koristiti pitanja koja su se pokazala prikladnim u prethodnim istraživanjima.
- Na prvoj stranici pisanog upitnika treba da stoji naziv istraživanja kao i logo statističke institucije i pravni osnov za provođenje istraživanja. Uvodno pitanje, generalno treba da se odnosi na sve izvještajne jedinice.
- Izgled printanog upitnika treba (ovisno od diseminacijskog medija) biti takav, da različiti blokovi pitanja budu jasno razdvojeni jedni od drugih.
- Upitnici trebaju biti popraćeni sa uputstvom za popunjavanje, koje će olakšati popunjavanje. Putem uputstva treba posebno objasniti, kako upisati dozvoljene vrijednosti kod kvantitativnih pitanja.
- U cilju ispitivanja vjerodostojnosti dobijenih informacija, kod sveobuhvatnih upitnika je potrebno implementirati kontrolna pitanja.
- Potrebno je voditi računa da broj pitanja u upitniku bude što manji. Za svako pitanje koje se pojavi u upitniku mora postojati razlog.
- U slučaju da se od određenih izvještajnih jedinica prikupljaju različita obilježja, potrebno je da svaka grupa izvještajnih jedinica dobije različite upitnike („Tailor-made upitnici“).

3.1.2 Priprema materijala za komuniciranje sa izvještajnim jedinicama

Pored upitnika, prije samog provođenja statističkog istraživanja potrebno je pripremiti neke dodatne materijale, namijenjene za komuniciranje sa izvještajnim jedinicama. Dodatni materijali su: pismo obavijest, prospekt/brošura, opomena, pismo zahvale, kontrolno pismo, dopunsko pismo.

U *pismu obavijesti*, izvještajna jedinica se upoznaje sa statističkim istraživanjem, glavnim ciljem istraživanja, sa povjerljivošću podataka, itd... .

U *brošuri*, izvještajne jedinice se animiraju da sudjeluju u istraživanju, zato u brošuri prezentiramo za širu javnost što zanimljivije rezultate, koji mogu naglasiti smisao provođenja istraživanja i privući izvještajne jedinice da sudjeluju u istraživanju.

Kad je riječ o *opomeni*, obično se šalju dvije opomene, iznimno tri.

Pismom zahvale izvještajnoj jedinici se zahvaljuje za sudjelovanje, šalje se kada se neko istraživanje prestane provoditi (zbog ukidanja istraživanja), ili kada se prelazi na druge izvore podataka, te zbog toga sudjelovanje izvještajne jedinice više nije potrebno, ili kod istraživanja sa periodičnom rotacijom posmatranih jedinica, kada izvještajna jedinica nije više u obuhvatu.

Kontrolno pismo je pismo, namijenjeno kontroli rada anketara, i sadrži pitanja, koji se odnose na sadržaj upitnika i pitanja vezana za provođenje anketiranja.

Dopunsko pismo se koristi kada je ostvaren neuspješan prvi kontakt sa domaćinstvom, ili domaćinstvo zbog različitih razloga nije sudjelovalo u anketiranju (izabrane osobe nije bilo u stanu, nije imala vremena za anketiranje itd.).

Prije samog slanja, sve pripremljene materijale treba obavezno lektorisati.

SMJERNICE

- Kod statističkih istraživanja za osobe i domaćinstva, gdje u anketiranju sudjeluje anketar, potrebno je izabranu izvještajnu jedinicu obavijestiti pismom o dolasku anketara i o namjeni samog istraživanja.
- Izvještajnim jedinicama koje nisu odgovorile, šalje se opomena odnosno iznova se zamoli za sudjelovanje u istraživanju.
- Kod istraživanja za preduzeća obično se šalju dvije opomene.
- U slučaju, da se za istraživanje koristi kombiniran način anketiranja: npr. kombinacija telefonskog i terenskog anketiranja, praksa je da se nekontaktirane jedinice kod telefonskog anketiranja pokušaju anketirati na terenu, jer na taj način smanjujemo pristrasnost zbog neodgovora. U ovom slučaju ovim jedinicama se šalje dopunsko pismo.
- Kontrolna pisma su namijenjena naknadnoj kontroli rada anketara na terenu. U ova pisma se uključuju neka od ključnih pitanja iz istraživanja, i neka o radu anketara. Na pismu se ispisuje identifikacija anketara i anketiranog domaćinstva ili izabrane osobe.

3.2 Uspostava programske podrške

Potrebno je konfigurirati programsku podršku i sisteme koji se koriste u statističkim poslovnim procesima, od prikupljanja podataka do diseminacije. Opisujemo radnje koje treba poduzeti u fazi pripreme za uspostavu novih i unapređenje postojećih softverskih komponenti potrebnih za poslovni proces.

Komponente mogu uključivati kontrolne tabele i izvještaje, baze podataka, tabele rezultata, alate za transformaciju podataka, alate za upravljanje podacima i metapodacima. Ovo podrazumijeva pripremu tehničke specifikacije za uspostavu ili ažuriranje komponenti za obradu i analizu podataka. Uspostava programske podrške se izvodi na temelju konsultacija između IT odsjeka, metodologa predmetnih statistika i odgovornih osoba za uzorkovanje (po potrebi). Izrađuje se ili ažurira softver za obradu i analizu podataka u saradnji između metodologa, IT (ili vanjskih programera) predmetnih statističara i uzorkaša (po potrebi).

3.3 Uspostava komponenti diseminacije

Opisujemo aktivnosti za uspostavu novih komponenti ili ponovnu upotrebu postojećih komponenti i usluga potrebnih za diseminaciju statističkih proizvoda kao što je kreirano u podprocesu 2.1 (Dizajniranje proizvoda statističkog istraživanja).

Sve vrste komponenti diseminacija i usluga su uključene, od onih koje se koriste za proizvodnju tradicionalnih printanih publikacija do onih koji pružaju web usluge, (povezane) otvorene izlazne podatke, geoprostorne statistike i geografije, karte ili pristup mikro podacima.

Potrebno je pripremiti tehničke specifikacije za izgradnju ili ažuriranje komponenti za diseminaciju. Specifikacije uključuju popis proizvoda koji se diseminiraju, vrste i funkcionalnosti alata za diseminaciju, pravila i standarde za vizualizaciju, odnos sa izvještajima o kvalitetu i metapodacima i drugo. Ovaj podproces se provodi na temelju konsultacija između metodologa predmetnih odsjeka, osoblja odsjeka nadležnih za diseminaciju i IT-a. Potrebno je izraditi ili ažurirati softvere za diseminaciju.

3.4 Testiranje alata za prikupljanje i obradu podataka

Ovaj postupak podrazumijeva tehničko testiranje i odobravanje novih programa i postupaka. Uključuje i testiranje interakcije između komponenti te se brine da proizvodni sistem funkcionira kao usklađen skup komponenti. Obavljaju se inicijalni testovi izgrađenih ili ažuriranih komponenti za obradu i analizu. Ovo takođe uključuje prikupljanje podataka za probna istraživanja radi testiranja instrumenata za prikupljanje podataka. Nakon toga slijedi obrada i analiza prikupljenih podataka. Nakon probnog istraživanja možda se treba vratiti na prethodne korake i učiniti prilagođavanje.

3.5 Testiranje i konfiguracija statističkog poslovnog procesa

Konfiguriranje toka proizvodnih procesa odnosi se na prikupljanje podataka pa sve do arhiviranja konačnih statističkih rezultata. Te aktivnosti uključuju:

- proizvodnju dokumentacije o komponentama procesa, uključujući tehničku dokumentaciju i korisničke priručnike;
- upoznavanje i obučavanje korisnika s proizvodnim sistemom. Cilj je upoznati korisnike proizvodnih sistema sa njihovom strukturom, procedurama i uputama za rad s njegovim komponentama. Upoznavanje se obavlja kroz određene tehničke

i druge obuke kao i razvoj i upotrebu korisničkih priručnika i druge tehničke dokumentacije. Ove aktivnosti se često provode uz podršku IT odsjeka;

- premještanje komponenti procesa u proizvodno okruženje te osigurati da te komponente funkcioniraju prema očekivanjima u tom okruženju.

4 PRIKUPLJANJE PODATAKA

U ovoj fazi prikupljaju se svi potrebni podaci (podaci i metapodaci), koristeći različite načine prikupljanja podataka (uključujući ekstrakcije iz statističkih, administrativnih i drugih ne-statističkih registara i baza podataka) i učitava ih se u odgovarajuće okruženje za daljnju obradu. Iako može uključivati validaciju formata podataka, ne uključuje same transformacije podataka, jer se sve to događa u 5. fazi "Obrada podataka".

4.1 Izrada okvira uzorka za statističko istraživanje i izbor uzorka

4.1.1 Priprema okvira uzorka

Ako želimo pripremiti kvalitetan okvir uzorka, obično trebamo koristiti više izvora podataka. Nakon što se ti izvori identificiraju i pripreme u prikladnu elektronsku formu, potrebno ih je uz pomoć odgovarajuće metodologije povezati, te pripremiti varijable koje određuju ključne karakteristike jedinica uključenih u okvir uzorka.

Naš konačni cilj je pripremiti jedinstveni set podataka koja obuhvata listu jedinica koje su što bliže teoretski definiranoj ciljnoj populaciji, i u kojoj će svakoj jedinici biti dodijeljene vrijednosti onih varijabli koje trebamo u kasnijem postupku odabira jedinica posmatranja.

Ključni korak kod provođenja postupka pripreme okvira uzorka je utvrđivanje postupka za izbor jedinica koje će biti uključene u okvir uzorka. Listu jedinica koja je utvrđena u odabranoj vremenskoj tački potrebno je »poboljšati« korištenjem dopunskih administrativnih i statističkih izvora podataka.

Primarni cilj je otkriti i onda eliminirati iz liste jedinice koje su inače u registru, ali u stvarnosti nisu dio ciljne populacije. Razlog za postojanje nerelevantnih jedinica u registru može rezultirati zbog grešaka u postupcima ažuriranja registra ili zbog administrativne prirode registra. Kao što je već spomenuto, u nekim slučajevima, pokušavamo otkriti nedostajuće jedinice, tj. dijelove populacije koje nisu obuhvaćene u registru, korištenjem

dodatnih izvora podataka. Razlog za »nepokrivenost« obično proističe iz već definirane populacije na koju se odnosi korišteni registar, jer ona ne zadovoljavaju u potpunosti specifične potrebe istraživanja.

SMJERNICE

- Ako se različita istraživanja odnose na istu osnovnu ciljnu populaciju i isto referentno razdoblje, potrebno je kod određivanje okvira uzorka koristiti iste postupke, jer ovim značajno doprinosimo većoj skladnosti statističkih rezultata.
- Potrebno je (kako u postupku identificiranja izvora podataka tako i u postupku obrade) eliminirati neprimjerene i duple jedinice, jer na taj način bitno poboljšavamo kvalitet statističkih rezultata.
- Ako vrijednosti varijabli okvira uzorka određujemo iz više različitih izvora podataka, potrebno je pažljivo proučiti postupke za određivanje prednosti (prioriteta) pojedinih izvora podataka.
- Kvalitet postupaka izrade okvira uzorka i kvaliteta korištenih izvora podataka (na temelju povratnih informacija prikupljenih u statističkom postupku) treba redovno i sistematično ocjenjivati. Ako otkrijemo veća odstupanja od već prihvaćenih standarda, potrebno je provesti postupke za poboljšanje kvaliteta.

4.1.2 Izbor jedinice posmatranja - uzorka

Okvir uzorka predstavlja fizičku realizaciju jedinica ciljne populacije, čije karakteristike bi opisivale statistički rezultat.

U narednom koraku potrebno je izabrati jedinice (između jedinica u okviru uzorka), koje ćemo zaista uključiti u istraživanje, i o kojima u kasnijim fazama želimo dobiti podatke. Zbog praktičnih (stručnih) razloga rijetko u istraživanje uključujemo sve jedinice iz okvira uzorka.

Obično je neophodno (odgovarajućim postupkom) listu jedinica iz okvira uzorka reducirati na primjerenu veličinu koja nam još uvijek omogućava izračunavanje dovoljno tačnih rezultata za cjelokupnu populaciju.

Kada je okvir uzorka pripremljen i kada je određena veličina uzorka, počinjemo postupak izbora jedinica uzorka, koji je bio utvrđen još u fazi planiranja istraživanja. Za ovo trebamo odgovarajuće programsko okruženje i posebne algoritme.

U praksi se koristi više postupaka selekcije uzoraka. To su:

Jednostavni slučajni uzorak - gdje se unaprijed se utvrđuje kako će se odabrati nositelji obilježja za istraživanje u slučaju parcijalnog istraživanja. Jednostavan slučajni izbor (nasumični uzorak) je statistički postupak izbora u kojem se nosioci obilježja izboru iz osnovne populacije po slučajnom postupku (Primjer jednostavnog nasumičnog uzorkovanja je izvlačenje kuglice iz bubnja, kao u sedmičnom izvlačenju brojeva na lutriji). Svrha nasumičnog uzorkovanja je izbjeći problem selektivnosti uzorka što je više

moguće. Preduvjet je kompletan popis svih nosilaca obilježja osnovne populacije koja može poslužiti kao osnova za izbor.

Dvostepeno slučajno uzorkovanje – klaster - U dvostepenom nasumičnom izboru (uzorak klastera), ukupna populacija je razbijena u takozvane klastere, pri čemu svaki klaster predstavlja umanjenu sliku osnovne populacije (npr. pojedinačne škole u anketi učenika). Elementi unutar klastera trebaju da odražavaju čitav niz obilježja koja se istražuju. S toga bi elementi unutar klastera trebali biti što je moguće različitiji (heterogeni). S druge strane, različite klasteri treba da međusobno budu što sličniji jedni drugima. Za uzorke trebaju pojedinačni klasteri nasumično odabrani. Zatim se ponovo vrši jednostavan nasumični izbor u odabranim klasterima ili se obuhvataju svi elementi klastera (npr. anketiraju se svi učenici odabranih škola).

Kvotni uzorak - U slučaju kvotnog uzorka, ukupna populacija se prvo dijeli na kvotne ćelije, koje se, na primjer, mogu razdvojiti po godinama i spolu u slučaju upitnika za osobe. Za svaku kvotnu ćeliju, mora određen broj jedinica uzorka biti uključen u istraživanje. Predstavlja 'neslučajni' stratificirani uzorak, a najčešće se koristi kod različitih 'ad hoc' organiziranih istraživanja za potrebe tržišta, za prikupljanje mišljenja građana o nekom pitanju i sl. Izbor konkretnih nosilaca obilježja (npr. domaćinstva) za "punjenje" kvotne ćelije nije dalje preciziran i može ovisiti o troškovima istraživanja ili dostupnosti. Ova metoda se npr. koristi u uzorcima prihoda i potrošnje, kao i u tekućim ekonomskim proračunima.

Stratificirani uzorak - najčešće korišteni dizajn uzoraka, pri čemu se populacija dijeli u homogene međusobno isključive grupe zvane stratumi. Kod takvih uzoraka najprije moramo izračunati stratumsku raspodjelu uzorka. To znači, da moramo za svaki stratum utvrditi broj jedinica, koje ćemo izabrati u uzorak. Jednostavan slučajni uzorak se izvlači iz svake grupe posebno. Razgraničenje stratuma se zasniva na karakteristikama stratifikacije (npr. starost i pol). Primjer jednostavnog stratificiranog slučajnog uzorka bio bi nasumično izbor 100 stanovnika po iz svake regije u državi. Stratume treba formirati na način da jedinice unutar jednog stratuma budu što sličnije s obzirom na obilježja istraživanja. Jedinice iz različitih stratuma treba da budu što je moguće više različite.

Najčešće koristimo slijedeće tri raspodjele:

- *jednakomjernu raspodjelu*: u svakom stratumu izaberemo jednak broj jedinica;
- *proporcionalnu raspodjelu*: u svakom stratumu izaberemo onoliko jedinica, da je udio svakog stratuma u pogledu veličine okvira uzorka, jednak udjelu izabranih jedinica iz ovog stratuma u odnosu na veličinu uzorka (to znači da u većim stratumima izaberemo više jedinica u uzorak nego u manjim);
- *optimalnu raspodjelu*: izračunamo odgovarajuće veličine uzorka u svakom stratumu uzimajući u obzir pored veličine stratuma i varijabilnost dostupnih pomoćnih varijabli. U nekim slučajevima se, pored navedenog, koristi i varijabilnost podataka pomoćnih varijabli iz prethodnog istraživanja.

Kod svih navedenih raspodjela posebno trebamo pripaziti da imamo dovoljan broj izabranih jedinica. Ako je po raspodjeli veličina uzorka u nekom stratumu nedovoljna, određujemo neki minimalan fiksni broj (obično oko deset jedinica) ili u uzorak izaberemo cijeli stratum, ako je njegov broj jedinica manji od ovog fiksnog broja.

Izbor namjernog uzorka ili izbor na osnovu determinističkih pravila obično izvodimo s posebnim algoritmom. Determinističke metode se razlikuju od metoda slučajnog izbora po tome, da se kod ovih metoda za izbor uzorka koriste tačno određena pravila, u kojima nije uključen niti jedan mehanizam slučajnosti. U praksi koristimo prije svega dva deterministička metoda za izbor uzorka: *popis*, te *obuhvat s pragom*.

Kod *popisa*, postupak se završava samim utvrđivanjem okvira uzorka, jer se postupak odnosi na rijetke slučajeve kada u istraživanje uključujemo sve jedinice iz okvira uzorka.

Kod *obuhvata s pragom*, sužavamo listu jedinica okvira uzorka (kao kod slučajnog uzorka), i to tako da na osnovu unaprijed postavljenih kriterija eliminiramo neke jedinice iz okvira uzorka. Kriteriji za eliminiranje obično se utvrđuju na osnovu vrijednosti neke varijable (jedinice kod kojih je vrijednost varijable iznad određene granice uključujemo u izbor a druge ne).

Kod obuhvata s pragom postoje dva osnovna pristupa.

Prvi pristup je pristup **fiksnog praga**. Kod ovog pristupa, kriterij za izbor jedinice u uzorak se definira s tačno utvrđenom vrijednosti kriterijske varijable. Najjednostavniji primjer ovakvog izbora jedinica u uzorak je postupak pri kojem u istraživanje (na primjer kod poslovnih istraživanja) uključimo samo jedinice (npr. preduzeća) koje zapošljavaju veći broj osoba od određenog broja (npr. 20).

Drugi pristup je upotreba **varijabilnog praga**. Kod ovog pristupa, populaciju najprije dijelimo (slično kao kod slučajnog stratificiranog uzorka) na podgrupe, a zatim u svakoj podgrupi odredimo onoliko najvećih jedinica (obzirom na kriterijsku varijablu) koje obzirom na vrijednosti kriterijske varijable prelaze neki unaprijed utvrđeni udio (npr. 75 %).

U praksi se kod obuhvata s pragom često koristi kombinacija oba gore opisana pristupa.

Rezultat postupka izbora uzorka su uzorak i jedinice koje su izabrane u njemu, a pondere odredjujemo naknadno (zapravo inicijalne pondere u toku izbora uzorka a finalne pondere nakon prikupljanja podataka, jer se isti obično korigiraju na osnovi neodgovora).

Uzorački ponderi su vrijednosti, koje skupa sa dobijenim podacima od jedinica uzorka koristimo kod izračunavanja procjene populacijskih parametara.

Ključne jedinice su jedinice, koje su za naše istraživanje važnije nego ostale jedinice (obzirom na predviđeni uticaj na konačni rezultat), i iste zbog toga drugačije tretiramo u fazi prikupljanja podataka i fazi uređivanja podataka.

SMJERNICE

- Prema teoriji uzorkovanja najefikasnija je takva stratifikacija, kod kojih su jedinice unutar stratumu sličnije (homogenije) u odnosu na neku ključnu varijablu u odnosu na jedinice iz različitih stratumu.

- Kod izrazito asimetrične distribucije ciljnih varijabli u populaciji, potrebno je kod izrade dizajna uzorka neke stratume uključiti obavezno u uzorak (»take-all stratum«). Ovo se preporučuje prije svega kod poslovnih istraživanja, gdje obavezno uključujemo u uzorak velika preduzeća.
- Obuhvat s pragom umjesto slučajnog uzorka, koristi se samo kod izrazito heterogene populacije (u odnosu na posmatranu varijablu), i to onda kada populacija ispod praga posmatranja ima zanemariv uticaj na planirane statističke rezultate. Ako se uvede obuhvat s pragom, potrebno je prije toga izvesti studiju, čiji rezultati bi opravdali uvođenje ovakve metode izbora.
- Ako upotrebljavamo obuhvat sa pragom, važno je takođe pratiti populaciju ispod praga.
- U izračunavanju alokacije/raspodjele treba uzeti u obzir takođe predviđen neodgovor u pojedinim podgrupama. U podgrupama, u kojima je predviđena veća stopa neodgovora, dodijeli se više jedinica nego u onim u kojima je predviđena niža stopa neodgovora.

Kod obuhvata sa pragom, kriteriji za određivanje praga za podgrupe sa predviđenom visokom stopom neodgovora se prilagođavaju pomjeranjem samog praga.

- Kod periodičnih istraživanja treba stalno provjeravati, da li tačnost rezultata, dobijenih na osnovu slučajnog uzorka, zadovoljava zahtjevane kriterije i standarde. Ukoliko nije tako, potrebno je dizajn uzorka promijeniti na odgovarajući način. Samo ako se utvrdi, da s postupcima uzorkovanja ne bi bilo moguće poboljšati tačnost, potrebno je razmisliti o mogućnostima za povećanje veličine uzorka.
- Kod periodičnih istraživanja i kod obuhvata s pragom, treba stalno provjeravati, da li »ispušteni« dio populacije ne uzrokuje preveliku pristrasnost statističkih rezultata. Ako je procjenjena pristrasnost, obzirom na unaprijed određene kriterije i standarde previsoka, potrebno je na odgovarajući način promijeniti postupke za određivanje praga posmatranja.
- Preporučljivo je da nema previše ključnih izvještajnih jedinica ukoliko želimo korištenjem ovih jedinica postići željeni učinak.

4.2 Organizacija prikupljanja podataka

Podaci čine osnov za rad na izradi statističkih proizvoda. Obrada podataka („Data Processing“) polazi od tzv. sirovih podataka i završava sa izradom autentičnog skupa podataka. Sam način dobijanja podataka, kako bi se dobili sirovi podaci, mora biti utvrđen već u procesu planiranja.

Postoji više mogućnosti, kako se može urediti proces prikupljanja podataka:

- a) korištenjem već raspoloživih podataka,
- b) korištenjem administrativnih podataka i
- c) prikupljanjem podataka provođenjem primarnih statističkih istraživanja.

Strateški cilj statistike BiH u narednom periodu je - da što više koristi postojeće administrativne izvore podataka i da tradicionalni način prikupljanja podataka (putem

obrazaca) bude korišten samo onda - kada se na drugi način podaci ne mogu obezbjediti. Korištenje administrativnih podataka je nedovoljno, a razlozi za ovo leže u činjenici, da su službene zbirke podataka u BiH veoma oskudne ili nisu vođene na adekvatan način.

Prikupljanje podataka provođenjem primarnih statističkih istraživanja koristi se slijedećim metodama: Osobno intervjuiranje putem anketara i slanjem printanih upitnika putem pošte; kao i načini prikupljanja koji se tek primjenjuju ili su u pripremi u statistici: Intervjuiranje putem podrške kompjutera („CAPI“) i intervjuiranje putem telefon („CATI“).

U tabeli prikazujemo prednosti i nedostatke navedenih načina prikupljanja podataka.

Način prikupljanja podataka	Prednost (+)	Nedostatak (-)
Podaci raspoloživi	- o u procesu prikupljanja nema dodatnih troškova - o nema opterećenja izvještajnih jedinica - o dostupno znanje o podacima	o podaci prikupljeni i korišteni za druge statističke namjene
Prikupljanje iz administrativnih izvora	- o nema opterećenja izvještajnih jedinica - o uglavnom postizanje većeg obuhvata	- o podaci uglavnom nisu direktno statistički korisni - o ovisnost od vlasnika podataka
Osobni intervju (»face to face«)	- o osobna pažnja prema izvještajnim jedinicama - o mogućnost korekcije od strane anketara	- o eventualne greške od strane anketara - o vrijeme putovanja anketara
Istraživanje putem printanih upitnika	o interni dokaz otpreme instrumentarija za istraživanje	- o velika opterećenost izvještajnih jedinica - o potreban obiman instrumentarij za istraživanje - o intenzivni rad na terenu - o potreba stalnog aktueliziranja adresa

SMJERNICE

- Način prikupljanja podataka treba utvrditi već u fazi planiranja. Pri tome, treba pronaći prikladan balans između troškova/koristi dobijenih informacija i opterećenja izvještajnih jedinica.
- U slučaju osobnih intervjuova treba na adekvatan način izvršiti obuku anketara. Obavezno treba provesti pilot istraživanje sa odgovarajućim brojem anketara, gdje se može simulirati vođenje ankete od strane anketara. Najoptimalnije rješenje je intervjuiranje podržano kompjuterom.

- Vrijeme uspostave kontakta sa izvještajnim jedinicama treba pažljivo odabrati u odnosu na vrijeme izvještavanja.
- Za vrijeme prikupljanja podataka putem primarnog statističkog istraživanja potrebno je kontinuirano kontrolirati povrat i ulaz materijala. Kod male stope povrata materijala treba poduzeti odgovarajuće mjere.
- U svakom istraživanju je veoma važno informirati izvještajnu jedinicu (pomoću propratnog materijala i obavještenja) sa svrhom i namjenom istraživanja. Dalje, jedinice treba upoznati sa eventualnim sankcijama predviđenim zakonom u slučaju neodgovora. U svakom slučaju izvještajne jedinice trebaju imenovati kontakt osobu, kojoj se mogu obratiti u slučaju nastalih problema.
- U cilju omogućavanja budućeg poboljšanja projekata, potrebno je tačno dokumentirati cjelokupan proces prikupljanja podataka.
- Kod primarnog statističkog istraživanja i sekundarnog statističkog istraživanja prikupljaju se individualni podaci. Iz tih razloga, potrebno je posebnu pažnju posvetiti zaštiti i povjerljivosti podataka.
- Kada se steknu uvjeti, izvještajnim jedinicama treba omogućiti da dostavljaju svoje podatke putem elektronskih upitnika. Potrebno je objasniti prednosti korištenja elektronskog izvještavanja kao i obezbjediti adekvatnu tehničku podršku putem *Helpdeska*.
- U slučaju korištenja sekundarnih statističkih izvora, treba jasno definirati i dokumentirati tok podataka, od vlasnika podataka do statističkih institucija. Prenos ovih podataka od određenih registara i administrativnih izvora treba obavljati uglavnom elektronskim putem. Obaveza vlasnika sekundarnih izvora podataka prilikom davanja podataka obavezno podrazumijeva i slanje neophodnih metapodataka.

4.2.1 Organiziranje preuzimanja podataka iz administrativnih izvora

Kod prikupljanja podataka putem statističkog istraživanja, nastojimo u što većoj mjeri reducirati stopu neodgovora.

U istraživanjima, u kojima dobijamo podatke iz sekundarnih izvora (administrativni registri, evidencije ...), problem neodgovora je obično manji, mogu se javiti drugačije poteškoće, prije svega one koje rezultiraju iz metodoloških razlika sa statističkim konceptima.

Evidencija sekundarnih ili administrativnih izvora vodi se na jednom mjestu, zbog toga da možemo kod potencijalnih dodatnih potreba, najprije provjeriti već postojeće izvore. Evidencije treba redovno uređivati i u iste zapisivati datum zadnjeg preuzimanja podataka.

U vezi preuzimanja podataka iz administrativnih izvora potrebno je odgovoriti na nekoliko pitanja:

- Koje podatke preuzimamo i od koga;
- Na koji način ih preuzimamo;

- Kakva je struktura podataka;
- Kako su podaci zaštićeni;
- Kada podatke moramo preuzeti i sl.

Kod procjene trenutnog stanja procesa preuzimanja podataka iz administrativnih izvora, potrebno je:

- izvršiti uvid (sačiniti listu/popis) svih administrativnih izvora, pri čemu se pored vrste podataka utvrđuje – ko je vlasnik podataka, ko je davalac i ko je administrator podataka (osoba odgovorna za administrativne izvore u statističkoj instituciji);
- na koji način dolaze podaci;
- kakva je njihova tehnička struktura;
- postojeća pisana dokumentacija u kojoj su opisani postupci koje provodimo po preuzimanju podataka u statističkoj instituciji;
- način pohranjivanja (čuvanja) podataka i
- način daljnje upotrebe podataka u statističkom procesu.

Prilikom preuzimanja podataka iz administrativnih izvora, potrebno je uvesti jedinstvenu terminologiju.

Iz tih razloga predloženo je korištenje sljedećih termina:

- **Sporazum** – je dogovor između institucija (administrativnog izvora i statističke institucije), kojim se uređuju pitanja u vezi tehničko-organizacijske i pravne zaštite prenosa podataka. Dogovor priprema *administrator koordinator* (osoba odgovorna za administrativne izvore) u saradnju sa pravnom službom.
- **Tehnički protokol** - uređuje i definira tehnički i vremenski dio prenosa, preuzimanja i pripreme podataka iz administrativnih izvora. Priprema ga *tehnički administrator* (osoba za tehničku podršku za administrativne izvore).
- **Intranet stranica za administrativne izvore (portal)** će biti postavljena na intranet stranici. Stranica će sadržavati sve potrebne informacije o preuzimanju administrativnih izvora (zaključeni sporazumi i tehnički protokoli).
- **Baza podataka:** statistička institucija od administrativnih izvora dobija podatke koji su sadržajno povezani. Ove podatke nazivamo baza podataka. Primjer baze podataka su npr. baza PDV-a;
- **Set podataka:** Baza je sastavljena iz više setova podataka. Jedan set podataka rezultira u jednoj tabeli za preuzimanje u bazi podataka. Primjer seta podataka su podaci iz jedinstvenog registra obveznika indirektnih poreza (obrazac ZR1);
- **Vremenski period:** vremenskim periodom jedinstveno označavamo, na koji vremenski period se podaci odnose;
- **Administrator koordinator** je osoba odgovorna za administrativne izvore;
- **Tehnički administrator** je osoba odgovorna za tehničku podršku za administrativne izvore;

- **Osoba ovlaštena za fizičko preuzimanje podataka od administrativnog izvora** je osoba koju ovlasti direktor za obavljanje ovih poslova;
- **Administrativni izvor** je davalac podataka.

4.2.2 Zaključivanje sporazuma sa administrativnim izvorom

Sa administrativnim izvorom statistička institucija zaključuje sporazum o preuzimanju podataka. Spotazume želimo standardizirati tako da u istim uvijek odredimo i sve parametre, koji su neophodni za uspješan i transparentan prijenos podataka.

Tehnički dio Dogovora je *Tehnički protokol*, i u njemu navodimo slijedeće:

- definiramo sadržaj podataka tj. vrlo precizno opišemo, koje podatke ćemo preuzimati,
- imenujemo i ovlastimo administratore: ADMINISTRATORA KOORDINATORA (osoba odgovorna za administrativne izvore) i TEHNIČKOG ADMINISTRATORA (osoba za tehničku podršku za administrativne izvore) na strani statističke institucije i na strani administrativnog izvora,
- utvrdimo periodiku preuzimanja podataka,
- utvrdimo tehničke specifikacije, koje sadrže:
 - tehničke parametre koji nam omogućavaju preuzimanje podataka,
 - tehničke parametre, koji nam omogućavaju da skinemo zaštitu sa podataka i iste dodatno obradimo prije punjenja u bazu podataka,
 - tehničke parametre, koji nam omogućavaju da podatke uspješno pročitamo i zapišemo u bazu podataka.

Određivanje tehničkih parametara kod preuzimanja podataka

Podatke možemo preuzimati na više načina, i to:

- preuzimanje podatka na prenosnom mediju (CD, DVD, prenosni disk, USB, ...),
- preuzimanje podataka preko FTP servera,
- prenosom datoteka preko e-maila,
- prenosom podataka pomoću web servisa (*engl. web service*),
- direktnim linkom na bazu podataka samog administrativnog izvora.

Tehničke specifikacije za preuzimanje podataka su specifikacije koje nam omogućavaju preuzimanje podataka u statističkoj instituciji. Pri tome je potrebno (pored dogovorenih podataka) obezbjediti i slijedeće podatke:

- identifikaciju davaoca podataka – administrativnog izvora
- jedinstvenu identifikaciju baze podataka,
- definirati vrijeme transfera podataka,
- odrediti verziju transferiranih podataka,
- odrediti nazive datoteka u slučaju prenosa datoteka i
- za svaku datoteku kontrolne podatke: broj zapisa u datoteci i broj definiranih polja

Preuzimanje datoteka na prenosnom mediju (CD, DVD, USB ...). Kod preuzimanja podataka u datotekama na ovim nosiocima, potrebno je definirati vrstu nosioca (CD, DVD, prenosni disk, USB, ...) i veličinu podataka koju prenosimo.

Kod dogovora sa administrativnim izvorom oko preuzimanje podataka na ovaj način, potrebno je definirati slijedeće:

- način obavještanja o novim podacima (administrativni izvor – Agencija),
- postići dogovor o dokumentu PRIMOPREDAJNI LIST kojim statistička institucija potvrđuje preuzimanje, a davalac potvrđuje dostavljanje podataka na ovom nosiocu,
- ovlastiti osobe, koje mogu dati podatke i osobe koje mogu podatke preuzeti,
- način identifikacije nosilaca podataka i
- nazive datoteka koje sadrže dogovorene podatke.

Ukoliko nije zaštićen način prenosa u tom slučaju datoteke moraju biti šifrirane. Za zaštitu se koriste programi za kriptovanje dogovoreni sa davaocem podataka.

Preuzimanje podatka preko FTP servera. Za uspješan prenos podatka preko FTP servera je potrebno da se administrativni izvor i statistička institucija dogovore o:

- IP adresi FTP servera ili FQDN servera,
- parametrima autentifikacije (korisničko ime i lozinka) na ovom serveru,
- načinu obavještanja o pripremljenim novim podacima na FTP serveru,
- definiraju potencijalnu direktorijsku strukturu na FTP serveru,
- nazivu datoteka koje sadrže dogovorene podatke.

Takođe je potrebno postići dogovor o tome, da li administrativni izvor odlaže podatke na svoj FTP server ili na FTP server statističke institucije. Ukoliko je transfer podataka nezaštićen, datoteke moraju biti šifrirane. Za zaštitu se koriste programi za kriptovanje dogovoreni sa davaocem podataka.

Preuzimanje podataka preko e-maila. Za uspješan prenos podatka preko maila potrebno je da se administrativni izvor i statistička institucija dogovore:

- sa koje adrese će pristizati mail sa podacima,
- odrediti adresu, na koj ćemo primati sve podatke iz administrativnih izvora,
- definirati sadržaj polja maila *Predmeta* (*engl. Subject*),
- nazivu datoteka sa dogovorenim podacima.

Ukoliko je veličina e-maila ograničena, odredimo da se mail koristi za prenos podataka, čija ukupna veličina ne prelazi 10MB. Za osiguranje zaštite pri prenosu podataka koriste se programi za kriptovanje dogovoreni sa davaocem podataka.

Preuzimanje podataka pomoću web servisa. Web servisi su internetski orjentirane aplikacije koje za razmjenu podataka koriste XML standard. Interfejs na strani davaoca podataka služi potrebama aplikacije primaoca podataka (administratoru ne dozvoljava

direktni pristup u bazu podataka na strani davaoca podataka), koji transferira samo unaprijed dogovorene podatke. Informatičari na obje strane se dogovore o tehničkim uvjetima preuzimanja podataka.

Za uspješan prenos podatka web servisom potrebno je da se administrativni izvor i Agencija dogovore o:

- namjeni web servisa,
- nazivu i adresi web servisa,
- XML šemi za poziv web servisa,
- XML šemi za odgovor web servisa i
- načinu autentifikacije.

Administrativni izvor na svojoj strani implementira web uslugu, statistička institucija na svojoj strani implementiraju klijenta, koji ima funkcionalnost za poziv dogovorene web usluge u dogovorenoj šemi i funkcionalnost preuzimanja podataka kad primi podatke koje mu dostavlja web servis u dogovorenom formatu.

Direktni link na bazu podataka izvora – Replikacije na nivou servera. Direktni link na server administrativnog izvora možemo obaviti na dva načina:

- potpunom replikacijom cjelokupne baze,
- linkom na bazu i preuzimanjem unaprijed dogovorenih podataka.

Prilikom prenosa podataka potrebno je definirati:

- koji podaci će se prenositi (cjelokupna baza ili samo određeni podaci),
- dogovor između adm. izvora i statističke institucije o sadržaju podataka. Pri tom se tačno utvrđuju tabele i polja s vrstom i detaljnim opisom sadržaja svakog polja,
- da li se će se prenositi stanje u određenom trenutku ili promjene,
- kada će se pronositi podaci (vremenski interval),
- izabrati tehnologiju za povezivanje baza podataka sa pripadajućim parametrima,
- administrativni izvor obzirom na način pristupa provodi svoju ili dogovorenu politiku zaštite,
- administrativni izvor uredi tabele, gdje se zapisuju svi važni podaci za prenos podataka,
- statistička institucija uredi intrefejs koji je dovoljan za link pristupa do podataka.

Određivanje tehničkih parametara za obradu podataka

Postavljanjem (određivanjem) tehničkih parametara za obradu podataka definiramo sve one postupke koji su potrebni da podatke pripremimo za punjenje u bazu podataka. U slučaju preuzimanja podataka u datotekama, ti parametri su slijedeći:

- dešifriranje preuzetih datoteka,
- kontrola datoteka:
 - da li smo preuzeli sve datoteke koje smo trebali preuzeti,
 - da li smo preuzeli pravilno broj zapisa u svakoj datoteci,

- provjera ostalih kontrolnih podataka.
- prevođenje (pretvaranje) identifikatora u statističke identifikatore.

Određivanje tehničkih parametara za punjenje podataka u bazu

U slučaju prenosa datoteka moramo u protokolu definirati format datoteke sa pripadajućim definicijama.

Proces određivanja tehničkih parametara za punjenje podataka u bazu omogućava nam da, nakon obrade možemo podatke iz datoteke pravilno pročitati, a potom ih puniti u bazu podataka.

Formati datoteka sa pripadajućim definicijama. Obzirom na različite formate datoteke, u fazi dogovaranja o razmjeni određujemo koji parametri nam omogućavaju uspješno preuzimanje podataka u bazu podataka.

Podaci u tekstualnoj datoteci. Podaci su u tekstualnoj datoteci zapisani u tabelarnom obliku, tako da svaki redak predstavlja jedan zapis.

Polja su definirana sa pozicijama u tom retku (od-do). Jedna datoteka može sadržavati jedan set podataka. U slučaju prenosa datoteke u tekstualnom formatu potrebno je odrediti slijedeće parametre:

- kodnu tabelu znakova u datoteci – standardna kodna tabela je Windows 1250,
- znakove za kraj zapisa (retka) – standard je <CR><LF>,
- znakove za kraj datoteke – standard je <CR><EOF>,
- struktura podataka:
 - naziv polja,
 - opis polja,
- tip polja:
 - tekstualni:
 - lijevo poravnanje teksta u polju.
 - numerički:
 - standardni zapis je bez znaka razdvajanja za hiljade;
 - znak razdvajanja za decimalno mjesto je zarez (,) (chr(44));
 - desno poravnanje broja
 - datumski:
 - standardni format je u obliku *ddmmyyyy* bez znakova razdvajanja.
 - vremenski:
 - standardni vremenski format je u obliku *hhmmss* bez znakova razdvajanja.

Prednost korištenja tekstualne datoteke je u njenoj jednostavnosti, davaocu podataka nije potrebno veliko znanje za izvoz podataka u takvoj strukturi, međutim mana je da se u tekstualnoj datoteci ne provjerava konzistentnost podataka.

CSV datoteka. CSV se od tekstualne datoteke razlikuje po tome da odredimo znak koji razdvaja polja jedna od drugih.

Za zapis podataka u CSV datoteci potrebne je odrediti slijedeće parametre:

- kodnu tabelu znakova u datoteci – standardna kodna tabela je Windows 1250,
- znakove za kraj zapisa (retka) – standard je <CR><LF>,
- znakove za kraj datoteke – standard je <CR><EOF>,
- znak razdvajanja između polja u zapisu – standard za znak razdvajanja je tačka i zarez (;) (chr(59)),
- struktura podataka:
 - naziv polja,
 - opis polja,
 - tip polja:
- tekstualni:
 - Sva tekstualna polja moraju početi i završavati sa navodnicima (") – (chr(34)). Ako tekstualno polje sadrži navodnike (chr(34)), važi pravilo da je pred svaki od njih potrebno dodati još jedne navodnike.
- numerički:
 - standardni zapis je bez znaka razdvajanja za hiljade;
 - znak razdvajanja za decimalno mjesto je zarez (,) (chr(44));
- datumski:
 - standardni format je u obliku *ddmmyyyy* bez znakova razdvajanja

Datoteka sa podacima u formatu Microsoft Excel. U ovom formatu možemo u jednoj datoteci preuzeti više setova podataka za razliku od tekstualnih datoteka, gdje smo jedan set preuzeli u jednoj datoteci. Microsoft Excel je program za obradu tabela. Tabele su razdvojene na više radnih listova.

U slučaju da preuzimamo podatke u ovom formatu, moramo odrediti slijedeće parametre:

- verziju Microsoft Excela u kojoj je bila kreirana datoteka,
- naziv lista (listova) na kojima su podaci,
- na svakom radnom listu može biti samo jedna vrsta podataka u obliku tabele, pri tome su u prvom retku napisani nazivi polja u tabeli,

Prednost zapisa podataka u ovom formatu je u tome što su podaci konzistentniji nego kod tekstualne datoteke, slabost je ograničenje (65536) broja zapisa na jednom listu u starijim verzijama Microsoft Excela.

Datoteka sa podacima možemo takođe preuzimati u Microsoft Access i u XML formatu. Tipovi podataka access baze moraju biti isti kao što je definirano u tabelama koje opisuju setove podataka za razmjenu. Podaci u XML formatu moraju pratiti strukturu definiranu u tabelama za razmjenu podataka.

SMJERNICE

- Prije početka preuzimanja podataka mora biti pripremljen i potpisan sporazum i tehnički protokol sa vlasnikom administrativnih podataka.
- Potrebno je imenovati administratora koordinatora i tehničkog administratora.
- Potrebno je poduzeti sve neophodne mjere koje će dovesti do reduciranja grešaka pri prijenosu podataka.
- Unošenje podataka o preuzimanju u evidenciju podataka treba biti redovno i ažurno.
- Potrebno je urediti evidenciju fizičkih nosilaca podataka. Svaki fizički nosilac podataka mora imati svoju šifru, i ona se bilježi u evidenciji, što omogućava da nosioce podataka lakše nađemo.
- Fizičke nosioce podataka treba pohraniti u vatrostalnim ormarima.
- Administrator koordinator mora se brinuti o obavještanju korisnika o promjenama, mogućim kašnjenjima i o drugih vanrednim događajima.

4.3 Pokretanje prikupljanja podataka

Prikupljanje podataka od izvještajnih jedinica (davalaca podataka) provodi se uz pomoć različitih instrumenata za prikupljanje. Samo pokretanje prikupljanja uključuje početni kontakt sa izvještajnim jedinicama (davaocima podataka) te sve naknadne aktivnosti praćenja pravovremenosti i vremena prijema izvještaja prema planiranom vremenskom rasporedu.

U ovom podprocesu se priprema preliminarna baza podataka. Odgovornom odsjeku i statističarima omogućava se pristup internim informatičkim alatima kako bi se pregledali prikupljeni preliminarni podaci. Takođe, obavlja se i provjera valjanosti prikupljenih podataka. Neke osnovne validacije strukture i integriteta primljenih informacija mogu se odvijati unutar ovog podprocesa, npr. provjera da su datoteke u traženom formatu i sadrže očekivana polja. U većini slučajeva ove automatizirane postupke valjanosti obavlja IT odsjek, a rezultati se dostavljaju predmetnim odsjecima za evaluaciju i daljnje postupke.

U slučaju identificiranih tehničkih problema i nakon što se poduzmu dodatne mjere, pristupi se izvještajnim jedinicama da ponovno podnesu svoje izvještaje. Ovaj podproces može uključivati i ručno unošenje podataka na mjestu kontakta ili upravljanje terenskim radom, ovisno o vrsti izvora i načina prikupljanja. Zapisuje se kada i kako su kontaktirane izvještajne jedinice i jesu li odgovorile na dostavljene upite. Upravlja se davaocima usluga, čime se osigurava da odnos između statističke institucije i davaoca podataka ostaje pozitivan.

4.3.1 Komunikacija sa izvještajnim jedinicama

Prikupljanje podataka i komuniciranje sa izvještajnim jedinicama zavisi od načina popunjavanja upitnika, od ciljne populacije, od posmatranih pojava i od raspoloživih izvora.

Kod prikupljanja podataka, posebnu pažnju treba obratiti na gubitak informacija zbog ljudske ili systemske greške anketara ili greške pri obradi podataka.

Visoku stopu odgovora izvještajnih jedinica možemo postići s jasnim ciljevima, sa odgovarajućim alatima i sa primjereno izabranim načinom prikupljanja podataka.

Istraživanja koja provodimo za izvještajne jedinice u nepovoljnom vremenu, npr. u vrijeme odmora ili većeg opterećenja jedinica (vrijeme izrade završnih računa) često imaju nižu stopu odgovora.

Na stopu odgovora utiče takođe oblast istraživanja: višu stopu odgovora imaju na primjer ankete o zdravlju, dok obično nižu stopu odgovora imaju ankete o žrtvama kriminala, itd...

Neodgovor možemo podijeliti u dvije glavne kategorije: neodgovor varijable i neodgovor izvještajne jedinice. O neodgovoru varijable govorimo onda, kada na neka pitanja u upitniku, (koji je inače ispunjen) nemamo odgovora.

Više je razloga za slučajeve neodgovora varijable, npr.: nije bilo moguće odgovoriti zbog nedovoljnog razumijevanja pitanja i znanja; odbijanja odgovora na osjetljivo pitanje; greška anketara, itd... .

O neodgovoru jedinice govorimo onda, kada podaci nedostaju kod svih varijabli.

Različiti načini popunjavanja imaju različite stope neodgovora. Najviša stopa neodgovora je kod samopopunjavanja, zato u istraživanjima, kod kojih su u okviru uzorka domaćinstva i osobe, podatke obično prikupimo intervjuom, gdje anketari mogu vrlo dobro motivirati ljude, te na taj način smanjiti stopu neodgovora.

Trajanje prikupljanja podataka je ovisno od načina popunjavanja upitnika. Obično je telefonsko anketiranje brže nego „face to face“ anketiranje ili anketiranje putem pošte. Kod istraživanja putem pošte, trajanje postupka prikupljanja je teško procijeniti, jer zavisi od broja opomena, koje šaljemo jedinicama koje nisu odgovorile, prije svega i od samog odziva jedinice. Ako koristimo više vremena i više izvora, stopa odgovora je obično viša.

Postoji mišljenje, da u jednom istraživanju možemo zadovoljiti samo dva od slijedeća tri zahtjeva: niži troškovi, visoka stopa odgovora i malo korištenog vremena. Praćenje razloga za neodgovor se provodi češće i detaljnije kod intervjua, nego kod samopopunjavanja.

Kod samopopunjavanja izvještajna jedinica sama pročitava pitanje i uputstva i odgovori tako kako to sama razumije, zbog toga ovi podaci često sadrže znatno više pogrešaka nego drugi načini prikupljanja podataka, što zahtjeva više rada u postupku uređivanja podataka.

Popunjeni upitnici ili komentari izvještajne jedinice mogu stići različitim kanalima: poštom, telefaksom, na elektronsku adresu, putem e-izvještavanja ili telefonom.

Evidenciju odgovora (neodgovora) redovno pratimo. Izvještajnim jedinicama, koje se nisu odazvale, pošaljemo najviše dvije pismene opomene. Po potrebi se (posebno kod ključnih jedinica) takođe provode i telefonske opomene.

Kod uspostave kontakta zaposlenka statističkih institucija sa izvještajnim jedinicama, one obično poručuju dodatne upitnike ili postavljaju tehnička ili metodološka pitanja. Zaposlenici (kontakt osobe) odgovoraju na ova pitanja u najkraćem mogućem vremenu.

Kod prikupljanja podataka putem intervjua, upitnik u ime izvještajne jedinice popunjava anketar. Anketar tačno zna, šta pita i kako će odgovori biti uneseni u upitnik, i osigurava, da dobije na sva pitanja prihvatljiv odgovor.

Uvijek vodimo računa, da je broj anketara adekvatan za određeno istraživanje. Mogući neodgovor se kontrolira tako, da anketari u dogovorenim rokovima pošalju podatke statističkim institucijama. Ukoliko anketar ne dostigne dovoljan broj popunjenih upitnika i kod popunjenih upitnika ne dostigne odgovarajući kvalitet, ako se otkriju metodološke greške, najprije ga upozorimo i upitnike provjeravamo. Ako se greške ponavljaju, anketaru jednostrano otkazemo ugovor i zamijenimo drugim anketarom.

Kod telefonskih anketa, anketare nadzire kontrolor koji je zadužen za praćenje njihovog rada. Kontrolor, kako anketiranje napreduje nadzire njegov kvalitet, anketare upozorava na greške i mogući nedovoljan kvalitet, te brine za odmor u toku rada. Obuka anketara, kontrola i distribucija dodatnih uputstava je ovdje jednostavnija nego kod drugih načina prikupljanja podataka. Poželjno je da je u svakom novom anketnom istraživanju prilikom anketiranja bude prisutan nosilac istraživanja i da odgovara na metodološka pitanja anketara.

Urgencija je postupak kojim pokušavamo izbjeći neodgovor odnosno minimizirati neodgovor jedinice izvještavanja. One su dakle bitan faktor za korelaciju načela pravovremenosti i tačnosti, te je iz ovih razloga jedna od ključnih mjera kvaliteta.

Formalne urgencije se vrše samo u slučajevima, kada do određenog zadanog termina nema odgovora od izvještajne jedinice. Osoblje statističke institucije se trudi (kada god je moguće) da izbjegne slanje urgencija. Kod istraživanja na dobrovoljnoj osnovi, treba izbjegavati slanje formalnih urgencija.

Općenito proizilazi, da uspjeh prikupljanja podataka puno ovisi od činjenice, da li postoji zakonska obaveza dostavljanja podataka ili se davanje podataka provodi na dobrovoljnoj osnovi. Stopa povratnih informacija kod prikupljanja podataka na dobrovoljnoj osnovi je osjetno niža, nego kod istraživanja za koje postoji zakonska obaveza davanja podataka.

Kod istraživanja na dobrovoljnoj osnovi (ankete) ponekad je potrebno telefonski animirati potencijalne izvještajne jedinice na sudjelovanje u istraživanju, u cilju postizanja odgovarajuće stope odgovora i osiguranja kvaliteta rezultata. Dakle, najprije se preko obnovljenih kontakata sa izvještajnim jedinicama pokušava iste uvjeriti u potrebu slanja traženih informacija.

Postupak slanja formalnih urgencija u pravilu se sastoji iz više faza, i to:

- prva opomena u formi pisanog podsjetnika, zatim druga opomena pomoću pisma, koje upućuje na zakonske konzekvence neodgovora po obavezi izvještavanja.

Termin za pojedine postupke urgencije se utvrđuje još u fazi planiranja, u ovisnosti od periodike istraživanja, obima procesa obrade i vremena diseminacije.

SMJERNICE

- Urgencije (kao instrument kojim se izbjegava ili minimizira neodgovor neodgovor jedinice izvještavanja) su veoma važan faktor za korelaciju načela pravovremenosti i tačnosti.
- Prije poduzimanja formalnog postupka slanja urgencije, treba pokušati stupiti u kontakt sa izvještajnim jedinicama i tražiti da dostave tražene informacije.
- Izbor vrste i učestalosti slanja urgencija, ovisi prije svega od stope odgovora respodenata koji se želi postići, ali i rokova naznačenih u projektima koji se odnose na obaveze diseminiranja podataka.
- Organizacija i rad na terenu moraju biti prilagođeni ciljnoj populaciji. Vremenski plan, obavještenja, brošura o istraživanju, uputstva, metode i alati za prikupljanje podataka moraju biti brižno planirani, i to s ciljem reduciranja opterećenja izvještajne jedinice.
- Kod komuniciranja sa izvještajnim jedinicama i s anketarima treba poštovati interna pravila komuniciranja. Probleme treba rješavati brzo i strpljivo.
- Za rad vezan za prikupljanje podataka, komuniciranje sa izvještajnim jedinicama odnosno anketarima, veoma je važno da telefonska centrala i serveri rade neometano i stalno.
- Potrebno je koristiti odgovarajuće postupke za smanjenje odbijanja davanja informacija. Broj odbijanja mora biti najmanje mogući. Treba identificirati segmente populacije sa većom stopom neodgovora i intenzivirati kontakte s njima.
- Treba utvrditi minimalan set podataka koji moraju dostaviti izvještajne jedinice, da iste možemo uzeti u obzir kao odgovor.
- Neodgovor možemo reducirati takođe korištenjem simboličnih poklona, mada pri tome moramo biti krajnje racionalni.
- Mora biti uspostavljena sistematska kontrola procesa: napredak u prikupljanju podataka, kontrola neodgovora, razuman odnos između kvaliteta i troškova. Značajni pokazatelji kvaliteta su: stopa odgovora, greške pri obradi, stopa neodgovora vezano za razloge odbijanja.
- Potrebno je utvrditi primjeren broj kontakata sa izvještajnom jedinicom. Obično izvještajnim jedinicama šaljemo najviše dvije pismene opomene.
- Od ključnih izvještajnih jedinica obavezno moramo dobiti odgovore. Ako nam to ne uspije, zatražimo od jedinice (pismom) informaciju, zašto nije poslala podatke. Za ključne jedinice moramo zbog toga imati zabilježen ili datum odgovora ili odgovarajući status.

- Na početku provođenja ankete, nosilac istraživanja obavezno sudjeluje u kontroli rada anketara.

4.4 Unos prikupljenih podataka

Statističke institucije prikupljanju podatke u raznim formatima i pripremaju za dalju obradu.

Podatke s printanih upitnika potrebno je za kasniju obradu najprije pretvoriti u elektronski format.

Pretvaranje u elektronski format može se obaviti na dva načina:

- ručnim unosom (rezultat ovog postupka su podaci u elektronskom obliku), ili
- pomoću optičkog čitača (rezultat ovog postupka su: slike upitnika i podaci u elektronskom formatu).

U posebnim slučajevima kontrola upitnika se može provesti prije unosa. Obim provjere u velikoj mjeri ovisi o tome koji će se postupak koristiti kod unosa podataka: skeniranje ili ručni unos podataka.

Ako namjeravamo koristiti optičko čitanje, podatke dopisujemo u odgovarajuća polja ili ih šifriramo u predviđena polja.

Podaci u upitniku ne smiju biti precrtani ili napisani pored polja za čitanje, jer to može dovesti do lošeg kvaliteta očitanih podataka; za ručni unos podataka, ova ograničenja ne važe.

Potrebni programi i postupci za ručni unos i optičko čitanje podataka se pripremaju još u fazi pripreme prikupljanja podataka. Potom je potrebno postupke i programe takođe testirati i izvesti neophodne korekcije.

Kada podatke prikupljamo korištenjem kombiniranih načina izvještavanja (npr. printani, elektronski upitnici, upitnici CATI, CAPI) ili pak obavljamo unos podataka korištenjem kombiniranih pristupa (npr. optički čitač i ručni unos) ili smo dio podataka preuzeli iz administrativne zbirke podataka, u daljnjim statističkim procesima treba prikupljene podatke udružiti u jedinstvenu bazu. Te podatke možemo kasnije u bazi dodatno provjeravati i urediti. Time je olakšano utvrditi pouzdanost rezultata.

SMJERNICE

- Nosilac istraživanja mora pripremiti jasna uputstva za provjeru podataka prije unosa.
- Program za unos podataka mora biti pripremljen npravovremeno. Prije upotrebe ga treba testirati, kako bi se izbjegle greške u izvornim podacima.
- Obzirom da je broj grešaka pri digitalizaciji podataka manji, kod prikupljanja podataka prednost treba dati elektronskim upitnicima. Kod prikupljanja podataka u printanom obliku primjerenije je optičko čitanje (zbog manje vjerovatnosti grešaka).

- Ako koristimo optičko čitanje ili brzi unos bez ponavljanja, upitnik treba oblikovati tako, da je moguće izraditi unutarnje kontrole (npr. posebno polje za iznose).
- Podaci na upitnicima, koji se čitaju na optičkom čitaču, moraju biti zapisani jasno i čitljivo.

5 OBRADA PODATAKA

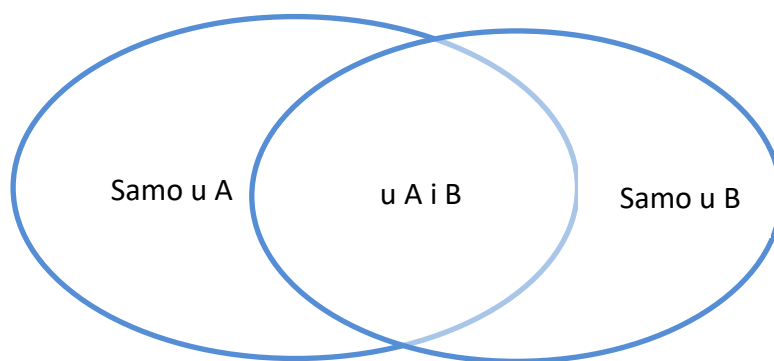
U ovoj fazi provodimo „čišćenje prikupljenih podataka“ i njihovu pripremu za analizu. Sastoji se od podprocesa koji provjeravaju, čiste i transformiraju ulazne podatke, tako da se mogu analizirati i diseminirati kao statistički rezultati. Podprocesi se mogu, ako je potrebno, ponavljati nekoliko puta. Za statističke proizvode koji se redovno proizvode, ova se faza pojavljuje u svakoj iteraciji. Podprocesi u ovoj fazi mogu se primijeniti na podatke iz statističkih i ne-statističkih izvora (uz mogući izuzetak podprocesa 5.6 „Ponderisanje“, što je obično specifično za podatke anketnih istraživanja). Faze "Obrada podataka" i "Analiza podataka" mogu biti iterativne i uporedne. Analiza može otkriti šire razumijevanje podataka, što bi moglo jasno pokazati da je potrebna dodatna obrada. Aktivnosti unutar faza "Obrada" i "Analiza" mogu započeti prije završetka faze "Prikupljanja podataka". To omogućava izradu privremenih rezultata u kojima je pravovremenost važna briga za korisnike i povećava raspoloživo vrijeme analizu.

5.1 Povezivanje različitih izvora podataka

Statistički proizvodni proces je doživio transformaciju u posljednjih nekoliko godina. U mnogim slučajevima, korištenje samo jednog izvora podataka nudi premalo informacija da bi zadovoljio analitičku svrhu nekog projekta. Kod statističkih istraživanja opravdano je uvoditi (iz razloga rasterećenja izvještajnih jedinica) obilježja iz raspoloživih administrativnih izvora. Efikasno i primarno korištenje postojećih administrativnih podataka za potrebe statističke proizvodnje podataka je prioritetni zadatak Statistike BiH. Za mnoge projekte **povezivanje različitih izvora podataka** je važan korak u procesu obrade podataka.

Opći uvjeti postupka povezivanja mogu biti različiti. Kod **neposrednog povezivanja**, dva seta podataka sa istim identifikacijskim obilježjima („ključne varijable“) međusobno se povezuju, pri čemu je važno da međusobno spojene jedinice sadržajno odgovaraju. Kako pokazuje *slika 5*, ovakva idealna situacija nije uvijek prisutna u praksi.

Stanje A



Stanje B

Slika 5: Kombiniranje dva izvora podataka

Nešto kompleksnije je **posredno povezivanje** tj. povezivanje podataka kada setovi podataka u referentnom i pridruženom izvoru nemaju isti identifikator (JMB, ID broj preduzeća), te je prije samog kombiniranja setova podataka potrebno uspostaviti vezu između različitih ključnih varijabli.

U ovom slučaju trebamo pomoć drugih odabranih varijabli, koje postoje u oba izvora. Jedinice referentnog izvora u ovom slučaju „razložimo“ na dva dijela. U prvom dijelu su jedinice, koje preko posrednih konektora, jednoobrazno povezujemo sa pridruženim izvorom.

Kod ovih jedinica vrijednosti posrednih konektora se u oba izvora potpuno slažu i u pridruženom izvoru postoji samo jedan takav zapis. U drugom dijelu su jedinice koje nismo uspjeli jednoobrazno povezati sa pridruženim izvorom što znači, da svaki zapis iz referentnog izvora „pridruženih“ ima više zapisa iz pridruženog izvora. Pridruženi zapisi su u ovom slučaju oni kod kojih su vrijednosti posrednih konektora u oba izvora najbliži.

U nekim slučajevima ne postoje ključne varijable putem kojih se može izvesti kombiniranje. U tom slučaju postoji mogućnost spajanja podataka preko tekstualnog poređenja pojedinih identifikacijskih obilježja (npr. adrese), pomoću sličnih metoda kao kod automatskog šifriranja. Ako ni ovo nije moguće, može se pokušati svakom zapisu podataka dodijeliti što sličniji duplikat.

SMJERNICE

- Prije nego se provede povezivanje podataka iz različitih izvora, treba obezbjediti i dokumentirati tačne informacije o korištenim konceptima i definicijama uključenih nosilaca podataka. Ovo se odnosi kako na količine i jedinice, tako i na sadržana obilježja - karakteristike.
- Pri povezivanju setova podataka na individualnom nivou mora se voditi računa o relevantnim zakonskim propisima vezanim za statističku povjerljivost.
- Ako je moguće, spajanja različitih izvora podataka treba provesti pomoću zajedničke ključne varijable.

- Pri povezivanju setova podataka sa različitim ključnim varijablama treba uspostaviti veze između identifikacionih obilježja. Memorisanje i održavanje ovih veza između važnijih nosilaca podataka treba kontinuirano provoditi u statističkom registru.
- U principu, poželjno je nosioce podataka u potpunosti povezati. Ne uspije li (sve jedinice spojiti preko identifikacionog obilježja), treba najprije pokušati ovo urediti pomoću poređenja teksta. Zatim, može se pokušati, pomoću „*Statistical Matching*“ postići što pouzdanija veza. Ostane li „nespojenih jedinica“, puna pokrivenost se osigurava primjenom odgovarajućih postupaka procjene.
- Svi aspekti koji su vezani sa spajanjem izvora podataka, trebaju biti opisani i dokumentirani. Ovo se odnosi kako na uključene izvore podataka tako i na korištene metode i kvantitativne informacije o procesu spajanja.

5.2 Šifriranje

Da bi se sa podacima moglo dalje raditi, trebaju biti na raspolaganju u elektronskoj formi.

Pod šifriranjem (signiranjem) se podrazumijeva dodjela alfanumeričkog ključa za tekstualni odgovor i najčešće je automatiziran postupak.

Za specijalne slučajeve takođe je neophodno djelomično manuelno šifriranje. Da bi se odgovori mogli automatski šifrirati, neophodna pretpostavka je da postoji dovoljno razvijen ključ za šifriranje. Svi zapisi koji se ne mogu automatski šifrirati trebaju biti po mogućnosti šifrirani od strane stručnog osoblja.

Signiranje korištenjem kompleksnih klasifikacija zahtjeva posebnu kvalifikovanost osoblja kako bi se postigli rezultati zadovoljavajućeg kvaliteta. Pri tome je potrebno poduzeti ciljanu obuku, da bi se prije svega osiguralo korektno identificiranje i razvrstavanje prema odgovarajućim pravilima klasificiranja, te postigla potrebna uporedivost.

SMJERNICE

Prikupljanje podataka

- Potrebno je najprije utvrditi raspone vrijednosti za svako obilježje. Isto tako mora se odlučiti, koji modaliteti, odnosno koje vrijednosti se koriste za situacije u kojima nema numeričkih obilježja (npr. nedostajuće vrijednosti, „bez odgovora“, itd.). Takve vrijednosti/modaliteti moraju biti jasno prepoznatljivi.
- U slučajevima manualnog prikupljanja podataka osoblje koje prikuplja, treba biti oprezno kod unosa na kritična polja. Ne smije biti dopušten unos izvan dozvoljenih

raspona vrijednosti. Putem dvostrukog popisivanja zapisa i naknadnog poređenja unosa, pogrešni unosi mogu biti svedeni na minimum.

Šifriranje

- Pri planiranju istraživanja potrebno je (što je više moguće) izbjegavati otvorena pitanja, na taj način šifriranje će biti potrebno samo tamo gdje je to bezuvjetno neophodno.
- Šifriranje treba, kad god je to moguće, obavljati automatski.
- Parametar za automatsko šifriranje treba biti izabran tako, da s jedne strane signira što veći broj zapisa, ali s druge strane da minimizira broj pogrešnih signiranja.
- Svi zapisi koji se ne mogu automatski šifrirati trebaju biti po mogućnosti šifrirani od strane stručnog osoblja. Osoblje treba biti u stanju (putem adekvatne obuke) preostale zapise korektno šifrirati.
- U slučajevima korištenja kompleksnih klasifikacija, potrebno je da se osoblje koje provodi šifriranje prije svega osigura korektnu identifikaciju i razvrstavanje po odgovarajućim klasifikacijskim pravilima.

5.3 Provjera prihvatljivosti/validacija podataka – logičke kontrole

Da bi statistički podaci bili adekvatno obrađeni, potrebno je osigurati da korišteni setovi podataka budu bez greške, logički konzistentni i sadržajno prihvatljivi. Svi podaci, dostavljeni od izvještajnih jedinica moraju proći provjeru prihvatljivosti.

Pri tome razlikujemo tzv. greške mjerenja i greške prepoznavanja:

- Kod *grešaka mjerenja* radi se o apsolutno pogrešnim podacima. Npr:
 - zbir podataka procentualne podjele ne daje 100%;
 - podaci nekog obilježja potpuno nedostaju /npr. nije dat podatak o prihodu/;
 - poslovni subjekt je razvrstan u trgovinu, ali se bavi 100% proizvodnjom, itd... .
- Kod *grešaka prepoznavanja* postoji sumnja u pogrešne podatke. Npr:
 - prihod jednog mjeseca je više nego dvostruk nego što je najviša vrijednost unutar zadnjih 12 mjeseci, ili je prihod manji od polovine nego što je najniža vrijednost unutar zadnjih 12 mjeseci;
 - odnos “prihod po zaposlenom” je izvan navedenog raspona, itd...

SMJERNICE

- Logičke kontrole moraju biti definirane dosljedno i kao takve obavezno testirane.
- U slučaju elektronskog prikupljanja podataka, važnija pravila logičkih kontrola trebaju biti već integrirana u aplikaciju prikupljanja i time onemogućiti pogrešan unos.

- Treba izbjegavati praksu (u želji da prikupljene podatke što bolje »pročistimo«) postavljanja previše kontrola, sa prestrogo postavljenim kriterijima.
- Pri formuliranju pravila prihvatljivosti na *mikro nivou* treba prepoznati:
 - vrijednosti izvan definiranih raspona vrijednosti,
 - nekonzistentne i neprihvatljive kombinacije vrijednosti i
 - nedostajuće vrijednosti.
- Za kvantitativne varijable postavlja se pitanje, da li su u okviru posmatranog predmeta prihvatljive ekstremne vrijednosti („grube greške“). Metode koje se koriste za otkrivanje grubih grešaka, trebaju biti u skladu s priznatim i poznatim standardima.

5.4 Editovanje i Imputacija

Postupak editovanja podataka možemo podijeliti na editovanje na mikro i makro nivou. Kod „mikro editovanja“ provodimo postupke editovanja na nivou podataka pojedinačnih jedinica.

Kod ovog postupka editovanja razlikujemo više postupaka u ovisnosti na koji način su podaci dobijeni.

- Kod intervjuisanja (terenska ili telefonska anketa) pojedinačne evidencije se testiraju na dosljednost sadržaja.

Po završenom anketiranju, podaci dobijeni na terenu ili telefonom se grupiraju skupa sa potencijalnim sekundarnim izvorima u zajedničku datoteku; u ovoj datoteci se provode još dodatne kontrole, eventualne otkrivene greške ispravljaju i to ručnim ili automatskim postupkom korekcije.

U ove svrhe se propisuju pravila vjerodostojnosti – pravila editovanja (*Editing Rules*), koja se primjenjuju na svaki pojedinačni set podataka.

Poseban aspekt kod mikro editovanja je prepoznavanje grubih grešaka – otkrivanje ekstremnih vrijednosti (*Outlier Detection*) kod kvantitativnih varijabli, koje mogu značajno uticati na rezultat. U slučaju prikupljanja podataka putem računara (npr. *CAP*) moguće je izvesti provjeru vjerodostojnosti i korekture već za vrijeme unosa podataka – uređivanje blizu izvora (*Editing near the source*).

- Kod samopopunjavanja (printani ili elektronski upitnik) logičke kontrole se provode po završenom unosu. Kontrole cjelokupnog seta podataka se provode istovremeno, i to sa automatskim korekcijama i generiranjem statistike pogrešaka. Otkrivene greške se u ovom slučaju ispravljanju ili s ručnim ili automatskim postupkom korekcije.
- Ako u istraživanju koristimo kombinaciju više različitih izvora podataka (npr. statističkih i administrativnih) najprije treba urediti svaki izvor posebno a zatim još izvesti kontrolu na zajedničkom setu podataka.

Kod poslovnih istraživanja se često koristi selektivno editovanje podataka. To je postupak kojim odredimo jedinicama različite prioritete za uređenje podataka. U fazi editovanja podataka odredimo jedinice koje su značajne za naše rezultate, i istima dajemo veću pozornost. Jedinice koje u fazi selekcije označimo s višim prioritetom, obično uređujemo

„ručno“ (sa mogućim ponovnim kontaktiranjem izvještajnih jedinica), a druge jedinice uređujemo automatskim postupkom.

Kod „makro editovanja“ se nakon stvarnog čišćenja podataka provjeravaju agregati i distribucija preko poređenja sa prethodnim periodom, poređenja sa drugim izvorima itd. Rezultati koji nisu vjerodostojni moraju biti detaljnije analizirani, te nakon razjašnjenja uzroka makropodatke treba korigirati.

Imputacija je postupak kojim se nedostajuće, nevažeće ili nekonzistentne vrijednosti zamjenjuju. Ovo se postiže putem izmjena odnosno umetanja modaliteta za pojedina obilježja u datoteci koja je identificirana da ima grešaka i/ili je nepotpuna. Sve ovo se poduzima s ciljem dobijanja potpunog, vjerodostojnog i koherentnog zapisa podataka.

U praksi, to znači da je imputacija popunjavanje nedostajućih vrijednosti (Missing Values) kod podataka iz istraživanja na nivou mikro podataka.

Imputacija se ne može identifikovati sa dodatnim procjenama neprikupljenog dijela populacije. Često se postupak imputacije smatra drugim dijelom čišćenja podataka i zbog toga je u uskoj sadržajnoj vezi sa provjerom prihvatljivosti (logičkom kontrolom).

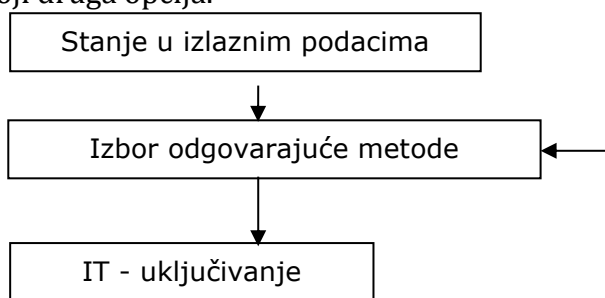
Uzroci za nedostajuće vrijednosti mogu biti različiti. Kao najčešći slučaj nedostajućih vrijednosti imamo slučaj odbijanja/uskraćivanja odgovora (*Non Response*). Ostali razlozi mogu biti npr. greška tehničke prirode u prijenosu podataka, greška anketara, itd... .

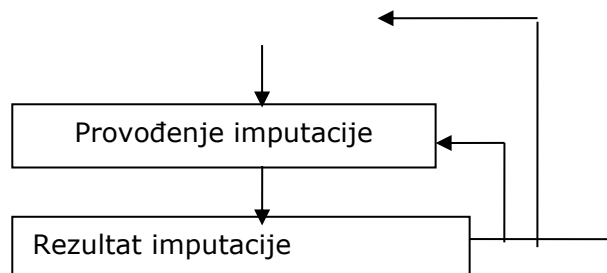
Imputacija služi prije svega tome da se izjednače nastala „izobličenja“ do kojih dolazi iz razloga odbijanja odgovora, tzv. pristranost izostanka odgovora (*Non Response Bias*). Pri tome se u osnovi preko imputacije samo korigira nedostatak pojedinog obilježja u datoteci – obilježje bez odgovora (*Item Non Response*).

Teoretski je takođe moguće, nadomjestiti preko imputacije cijele pojedinačne setove podataka koji nedostaju (*Unit - Non Response*). **Međutim, problem gubitka cijelih jedinica u službenoj statistici se najčešće nadoknađuje ponderisanjem.**

Karakteristika imputacije je da vrijednosti korištene za pojedine slučajeve ne odgovaraju određenim vjerovatnim vrijednostima. **Posebno treba naglasiti, da se imputacijom ne smatra korištenje neke vrijednosti od nekog pouzdanog administrativnog izvora.**

Važno je naglasiti, da su metode imputacije primjenjive samo onda kada je količina nedostajućih vrijednosti mala u odnosu na ukupan broj zapisa podataka. U nekim situacijama je neophodno dopuniti veliki dio ili čak sve vrijednosti nekog specifičnog obilježja nosioca podataka. Ovaj postupak (označen takođe kao *Masivna imputacija*) treba izabrati samo ako ne postoji druga opcija.





Slika 6: Postupak imputacije

Slika 6 pokazuje da se postupak imputacije sastoji iz više faza. Na samom početku se obavlja analiza nedostajućih (i neprihvatljivih) vrijednosti, u cilju boljeg razumijevanja problema sa podacima. Od ove analize i od distribucije podataka uglavnom ovisi izbor metode, koja se implementira u IT ili adaptira da bi se mogla primijeniti. Na kraju slijedi analiza rezultata. Ako je kvalitet imputacije loš, traže se moguće greške ili se izabere druga metoda imputacije.

Izbor metode imputacije ovisi od stanja podataka i najčešće se koriste slijedeće:

- **Metoda logičkih imputacija** - Kod ove metode imputiramo vrijednost koja logično slijedi iz podataka koje smo uspjeli dobiti za posmatranu jedinicu (npr. ako za anketiranu osobu znamo datum rođenja a ne starost, možemo nedostajuću starost izračunati);
- **Metod prosječne vrijednosti** - Kod ove metode nadomjestimo nedostajuću vrijednost s prosječnom vrijednosti onih jedinica čije podatke imamo. Prosječnu vrijednost obično ne izračunavamo iz cijelog seta podataka nego iz seta podataka domena u kojem se nalazi jedinica za koju imputiramo podatak;
- **Metoda internog donora** - Kod ove metode imputiramo vrijednost koju preuzimamo od neke druge jedinice (donora) za koju imamo podatak. Pri tome donor je najčešće slučajno izabran iz seta jedinica koje su po traženim karakteristikama najbliže nedostajućoj jedinici, ali može biti tačno određen;
- **Metoda eksternog donora** – Kod ove metode imputiramo vrijednost koju preuzimamo iz nekog vanjskog izvora, npr. iz istog istraživanja za prethodno referentno razdoblje, iz drugog istraživanja ili iz administrativnih izvora podataka. Vrijednost jednostavno prepíšemo ili je prilagodimo;
- **Metoda regresije** – Kod ove metode imputiramo vrijednost koju izračunamo iz odgovarajućeg regresijskog modela.

U cilju praćenja uticaja korištenih postupaka imputiranja na statističke rezultate, potrebno je omogućiti izračunavanje odgovarajućih pokazatelja kvaliteta.

- Stopa imputiranih podataka se izračunava barem za ključne varijable, kao odnos između broja jedinica za koje smo imputirali podatak za odgovarajuću varijablu i broja svih jedinica koje bi morale imati podatke za tu varijablu;
- Stopa imputiranih podataka za statistike se izračunava barem za ključne statistike, kao odnos između statistika izračunatih prije postupka imputiranja podataka, i statistike izračunate po ovom postupku.

SMJERNICE

- Za makroeditovanje trebaju biti što više uključene reference poređenja. Posebno poređenja u odnosu na prethodna razdoblja.
- Kad god je to moguće, provjera prihvatljivosti (logičkih kontrola) treba biti tehnički automatizirana.
- Sve postupke uređenja podataka treba detaljno dokumentirati. Kod svakog novog projekta koji se odnosi na imputaciju mora biti uključena oblast metodologije.
- Proces uređenja podataka treba postaviti tako kako bi se omogućila ponovljivost postupaka; što znači da postupci uređenja kod kojih budu korišteni isti ulazni podaci i iste metode, uvijek daju isti rezultat.
- Ako je neophodno, kod svakog nosioca podataka zamijeniti nedostajuće vrijednosti. Zadržavanje nedostajućih vrijednosti u setu podataka može dovesti do značajne pristrasnosti rezultata i da oteža provođenje zahtjevnijih statističkih analiza.
- Imputacije cijelog seta podataka treba provoditi samo u slučaju, kada je kompenzaciju ukupnih nedostataka (*Unit-Non Response*) nemoguće izvršiti ponderisanjem, a da pri tome naravno postoji adekvatan izvor na osnovu kojeg se može izvršiti imputacija.
- Masivna imputacija pojedinih obilježja za veliki dio setova podataka (>20%) može biti provedena samo u slučajevima kada se čine neizvodljivima ostale opcije.
- Rezultat toka imputacije treba analizirati, vrednovati i utvrditi - da li su imputacijom postignuti željeni efekti. Potrebno je posebno uporediti distribucije i izračunate setove prije i nakon imputacije.

5.5 Proizvodnja izvedenih varijabli i jedinica

Nove statističke jedinice mogu se izvesti agregiranjem ili dijeljenjem podataka za jedinice prikupljanja ili prema različitim metodama procjene (npr. izvođenje domaćinstava, kada su jedinice u prikupljanju podataka osobe, ili preduzeća, kada su jedinice pravne osobe). Izvedene nove varijable ili jedinice potrebne su za dobivanje agregata ili za izračunavanje pokazatelja, indeksa ili drugih statističkih rezultata.

Izvedene varijable i statističke jedinice se proizvode primjenom aritmetičkih formula na jednu ili više postojećih varijabli koje su već prisutne u setu podataka koje nisu izričito zahtjevane tokom prikupljanja podataka, ali su potrebne za isporuku zahtjevanih rezultata.

5.6 Ponderisanje

O uzorku govorimo kada se u nekom statističkom istraživanju koristimo samo dijelom populacije a ne ukupnom populacijom. Veličina uzorka je značajno manja od osnovne cjeline, pri čemu je za statističku instituciju od značaja da se istraživanja sa punim obuhvatom znatno reduciraju i na taj način štede rad i finansijska sredstva potrebna za provođenje samog statističkog istraživanja.

Postoje dva različita pristupa u uzorkovanju: Slučajni uzorak (ili uzorak uz primjenu vjerovatnoće – *probabilistic sample*) i uzorak bez primjene vjerovatnoće (*non-probabilistic sample*). Jednostavni slučajni uzorak se rijetko koristi u službenoj statistici. U Statistici BiH je u upotrebi slojeviti – stratificirani slučajni uzorak.

Slučajni uzorak je karakterističan po tome, što svaka jedinica uzorkovanja iz ciljne populacije (okvira uzorka) ima poznatu vjerojatnoću (različitu od nule) s kojom će biti izabrana u uzorak. Rezultati iz slučajnih uzoraka trebaju omogućiti izvođenje zaključaka o ukupnoj populaciji.

Često se u praksi primjenjuje (posebno u području ekonomskih statistika) tzv. *cut off - koncentracijski uzorak*, pri čemu se izabiru samo jedinice preko određenog praga vrijednosti, uzimajući u obzir u pravilu standardizirani kriterij reprezentativnosti za osiguranje kvalitetnog obuhvata.

Kod slučajnih uzoraka ekstrapolacija čini odlučujući korak u obradi za pripremu procjene. Pri tome se svakoj jedinici iz uzorka dodjeljuje ekstrapolacijski ponder, tj. podaci se ponderišu.

Postupak ponderisanja može uključivati:

1. Dizajn pondera (inicijalni ponder)
2. Ponderisanje uvjetovano neodgovorom (*Non Response Weighting*), odnosno prilagođavanje inicijalnog pondera stopom neodgovora (non-response faktor)

Dizajn pondera se unaprijed određuje planom uzorkovanja i svaka jedinica iz uzorka korespondira recipročnoj vrijednosti vjerovatnoće izbora.

Konačno, može se polazeći od prethodno određenih pondera (*Dizajn ponder $\times$ Non Response Faktori = bazni ponder*) provesti slijedeći korak, u cilju prilagođavanja rezultata ekstrapolacije vanjskim izvorima (*vezana ekstrapolacija*).

Kod svakog slučajnog uzorka pri procjeni rezultata pojavljuje se nepreciznost tj. greška uzorkovanja. Za svako istraživanje, koje se provodi na bazi uzorka izračunava se (barem za ključne varijable) greška uzorkovanja.

SMJERNICE

- Rezultati iz istraživanja koja se provode na bazi uzorka se uvijek smatraju procjenama, koje su samim tim opterećene greškama uzorkovanja.
- Procjene iz istraživanja na bazi uzorka baziraju se na ponderisanju. Pri tome moramo uzeti u obzir barem dizajn uzorka i stopu neodgovora (slobodna ekstrapolacija). Na kraju cijelog postupka (kad god je to moguće), rezultate treba prilagoditi eksternom izvoru (vezana ekstrapolacija).
- Nakon konačnog prilagođavanja treba kontrolirati rezultate. U svakom slučaju treba obratiti pažnju da se putem odgovarajućih metoda osigura, da varijabilitet pondera ne ode predaleko).
- Za svako istraživanje na bazi uzorka treba prikazati grešku uzorkovanja barem za ključne varijable. Metode, kojima se provode izračunavanja grešaka, moraju biti dokumentirane.
- Da bi se obezbjedilo razumijevanje, mora se pojasniti u kojem obliku će se prikazati greška uzorkovanja (standardna greška kao apsolutna mjera varijabilnosti, koeficijent

varijacije kao relativna mjera varijabilnosti, Interval povjerenja itd). Kod prikaza intervala povjerenja u svakom slučaju treba navesti odgovarajući nivo značajnosti (u pravilu 95%).

- Za sve aspekte izračunavanja grešaka nastalih uzorkovanjem, kao i za ponderisanje podataka odgovoran je odsjek za metodologiju uzorkovanja.
- Korisnicima treba biti saopćen nivo toleriranja greške uzorkovanja tj. da ista ne prelazi prihvatljivu vrijednost. Ovo se odnosi prije svega na regionalno raščlanjivanje. Ako se vrijednosti publiciraju i greška uzorkovanja pokazuje određeno prekoračenje prihvatljive vrijednosti, to mora biti označeno na odgovarajući način.
- Cijeli tok istraživanja na bazi uzorka mora biti dokumentiran. Prije svega od interesa su sljedeće veličine: izvorna bruto veličina uzorka, postignuta neto veličina uzorka i stopa odgovora za važnija obilježja (varijable).
- Svi aspekti vezani za uzorak, posebno plan uzorkovanja, metode ekstrapolacije kao i relevantne kvantitativne mjere trebaju biti navedene u Izvještaju o kvalitetu.

5.7 Izračunavanje agregata

Ova podfaza podrazumijeva kreiranje agregiranih podataka Ukupnoa populacije iz mikro podataka ili agregata nižih nivoa. Ovo uključuje sumiranje zapisa koji dijele određene zajedničke karakteristike, određuju se mjere prosjeka i disperzije te primjenjuju ponderi iz podprocesa 5.6 da bi se izveli odgovarajući Ukupnoi.

U slučaju istraživanja koja se provode na bazi uzorka, mogu se takođe izračunavati greške uzorkovanja i pridružiti relevantnim agregatima. Izračunavanje agregiranih podataka se može izvršiti putem zahtjeva primarnoj bazi podataka pomoću posebnog alata koji daje agregate ili izračunava nove agregate izvedene iz inicijalnih.

Ovdje treba uključiti sve podatke prikupljene za svrhe poređenja, tj. Ukupnoe, srednje vrijednosti, medijan, koeficijent varijacije, standardnu devijaciju itd.). Uključuje i svaku procjenu varijance za svrhe validacije, kao što su: intervalli pouzdanosti i greške u uzorkovanju koje se izračunavaju na nivou agregata samo za interne provjere (važno je napomenuti da se navodi metod, a ne aktuelna vrijednost za grešku). U praksi nije uobičajeno izračunati ove mjere u tako ranoj fazi, ali se može dogoditi, npr. u HBS izračunava se interval povjerenja za prihod i upoređuje s onim za SILC radi konzistentnosti.

5.8 Izrada finalnih datoteka

Ova podfaza povezuje rezultate ostalih podprocesa u ovoj fazi, što rezultira podatkovnom datotekom (obično makropodataka), koja se koristi kao ulaz u fazu "Analiza". Ponekad to može biti prelazna, a ne konačna datoteka, posebno za poslovne procese gdje postoje jaki vremenski pritisci i zahtjev za izradu preliminarne i konačne procjene. Znači u ovom

podprocesu obrađene podatkovne datoteke se pripremaju za analizu. Podatkovne datoteke bi trebale sadržavati mikro podatke, pondere i agregate.

6. ANALIZA PODATAKA

Statistična analiza podataka je postupak analiziranja podataka pomoću različitih alata i tehnika, i to s ciljem proučavanja stanja i utvrđivanja određenih zakonitosti koje su značajne za posmatranu pojavu s ciljem interpretacije rezultata. Uobičajeno, najprije analiziramo podatke na makronivou, a po potrebi i na mikronivou.

Podatke analiziramo iz razloga što analizom možemo potvrditi njihovu relevantnost, odnosno možemo otkriti moguće nedostatke i zatim iste eliminirati, te na taj način osigurati bolji kvalitet podataka. Ako analiza pokaže systemske nedostatke, rezultate koristimo za poboljšanje kvaliteta procesa ili za dopunu ili izmjenu cjelokupne metodologije.

6.1 Priprema nacrtu rezultata

U ovom podprocesu podaci se pretvaraju u statističke rezultate. Uključuje proizvodnju dodatnih mjera, kao što su indeksi, trendovi ili sezonski prilagođene serije, kao i evidentiranje obilježja kvaliteta. Pokazatelji kvaliteta kao što su stope neodgovara, greške uzorkovanja ili drugi potrebni pokazatelji također se izračunavaju paralelno s pripremanjem nacrtu rezultata.

Pokazatelji kvaliteta mogu se koristiti za dodatnu provjeru rezultata u podprocesu 6.2. ili kao dodatna informacija u podprocesu 8.2. Takođe, u ovom podprocesu se obavlja izračunavanje sezonski prilagođenih vremenskih serija.

6.1.1 Proizvodnja radnih i analitičkih tabela

Cilj proizvodnje statistika je publikovanje rezultata. Prije nego se definitivno odluči o vrsti i obimu objave, potrebno je proizvodnjom radnih i analitičkih tabela stvoriti osnovu koja omogućava procjenu dostupnih podataka pomoću detaljnih kombinacija obilježja.

Radne tabele prije svega služe internoj kontroli rezultata. Proizvodnja takvih tabela ne služi za objavljivanje, ali je bazirana na dobro definiranom i strukturiranom setu tabela.

Radne i analitičke tabele se generiraju iz autentične baze podataka. Ako analitičke tabele služe i za provjeru baza podataka (prije nego postanu autentične), onda one služe i čišćenju podataka.

Cilj proizvodnje radnih i analitičkih tabela je, odgovoriti na pitanja koja su relevantna za objavljivanje rezultata, a to prije svega uključuje: pitanje povjerljivosti, pitanje tačnosti rezultata i pitanje prikladnosti pohranjivanja u statističkim bazama podataka.

Radne i analitičke tabele se uglavnom izrađuju agregiranjem pojedinačnih podataka.

SMJERNICE

- Analitičke tabele trebaju obuhvatiti što više kombinacija obilježja, pri čemu prednost ima dvodimenzionalna tabela.
- Za izradu tabela treba koristiti raspoloživi standardni Software ili će odgovornost za proizvodnju tabela biti na IT-u. Proizvodnja radnih i analitičkih tabela uvijek mora slijediti iz autentičnih baza podataka.
- Interpretacija i kontrola radnih i analitičkih tabela mora biti savjesno provedena. Eventualni problemi trebaju biti prodiskutirani u širem krugu.
- Radne i analitičke tabele koriste se u interne svrhe. Važno je osigurati da je dostupnost tog materijala data samo ovlaštenim osobama. Potrebno je sve tabele (i elektronske) arhivirati.

6.2 Analiza relevantnosti i potvrđivanje rezultata

Analiza relevantnosti rezultata je proces, u kojem se provjerava „smislenost“ rezultata, njihova unutarinja skladnost, skladnost i uporedivost u vremenu i prostoru, te skladnost sa postojećim unutaršnjim i vanjskim referentnim izvorima podataka.

Analiza relevantnosti, odnosno potvrde rezultata se izvodi pri procesu uređivanja rezultata na makronivou i obuhvata slijedeće postupke:

- Provjera unutarnje skladnosti rezultata, npr. provjeravanje rezultata na osnovu poznatih odnosno očekivanih odnosa među rezultatima (da li je npr. vrijednost proizvodnje viša od dodane vrijednosti).
- Provjera skladnosti rezultata u odnosu na rezultate iz prethodnih referentnih razdoblja (prije svega ovo važi za ona istraživanja, kod kojih osnovna namjena nije mjerenje promjena u vremenu).
- Provjera skladnosti rezultata sa srodnim ili sa povezanim rezultatima iz drugih statističkih istraživanju, (rezultatima istraživanja koje provode statističke institucije u BiH, ili sa rezultatima istraživanja koje provode druge institucije).
- Interna provjera rezultata u statističkim institucijama.
- Povremeno provjera „relevantnosti i smislenosti« rezultata sa vanjskim stručnjacima.

SMJERNICE

- Prije konačnog potvrđivanja rezultata potrebno je još jednom analizirati sve procesne podatke (podatke, koji su na raspolaganju u vezi sa provođenjem procesa obrade podataka), prije svega rezultate uređivanja na makronivou.
- Postupke provjeravanja treba prilagoditi periodici istraživanja, ciljnoj populaciji, načinu prikupljanja podataka te vrsti izvora podataka (primarni ili sekundarni).
- Rezultate analize treba koristiti za poboljšanje kvaliteta kod narednog provođenja istraživanja.

6.2.1 Analiza vremenskih serija

Vremenska serija je vremenski uređen niz podataka, u našem slučaju statističkih podataka (npr. industrijska proizvodnja, indeks troškova rada ...). Glavni dio analize vremenskih serija je desezoniranje, manji dio je predviđanje - prognoziranje. Kod desezoniranja, odstranimo iz vremenske serije uticaje sezone i kalendara kada su oni specifični i značajni. Ovako dobijenim vrijednostima odredimo desezonirane vrijednosti ili vrijednosti sa eliminiranim uticajem sezone i kalendara.

Ovaj postupak treba izvoditi uvijek kada želimo uporediti podatke različitih vremenskih razdoblja iste vremenske serije ili podatke za isto vremensko razdoblje iste vremenske serije različitih država, jer oni tokom godine obično variraju ovisno o sezoni, broju radnih dana i drugim uticajima.

Kada poredimo podatke za isto razdoblje različitih godina iste vremenske serije, provjerimo samo - da li su u vremenskoj seriji prisutni uticaji kalendara i iste eliminiramo, jer je sezona za isto razdoblje u godini približno ista (npr. poređenje april 2023. u odnosu na april 2022.). Tako prilagođene podatke, nazivamo podaci sa eliminiranim uticajima kalendara ili podaci prilagođeni uticaju broja radnih dana.

Uticajima kalendara smatramo:

- uticaj broja radnih dana,
- uticaj praznika i
- uticaj prestupne godine

Međutim, ako poredimo podatke iste vremenske serije, ali različitog razdoblja potrebno je pored efekta uticaja kalendara ublažiti i uticaj sezone. Dakle, u tom slučaju, pri desezoniranju vremenske serije (X_t) utvrdimo model, iz kojeg zatim odredimo sezonsku komponentu (S_t), iregularnu komponentu (I_t) i komponentu trend-ciklus (TC_t). Drugim riječima, vremensku seriju možemo prikazati kao

- aditivan model navedenih komponenti $X_t = TC_t + S_t + I_t$, pri čemu su desezonirane vrijednosti (sezonski prilagođena vremenska serija) $X_t - S_t$ odnosno $TC_t + I_t$ ili
- multiplikativan model $X_t = TC_t * S_t * I_t$ kod kojeg su desezonirane vrijednosti X_t / S_t odnosno $TC_t * I_t$

Komponenta trend-ciklus TC_t opisuje dugoročno kretanje, sezonska komponenta S_t periodično kretanje (koje se ponavlja svake godine na isti ili sličan način), dok iregularna komponenta I_t sadrži slučajne uticaje, ostatke vremenske serije po uklanjanju ostalih komponenti.

Modeli se provjeravaju jednom godišnje (uobičajeno s prvim podatkom u kalendarskoj godini), nakon čega nosilac istraživanja pošalje metodologu za vremenske serije sve datoteke sa modelom parametara i podacima da iste pregleda (godišnji pregled).

Pri analizi vremenske serije su vrlo važne *netipične vrijednosti* (tzv. outlier), koji predstavljaju neobične promjene u vremenskoj seriji. Ako ne postoji greška u podacima, moramo pronaći razloge za postojanje netipičnih vrijednosti.

Obično imamo tri vrste netipičnih vrijednosti:

- *promjena nivoa*, odnosno prekid vremenske serije,
- *aditivna netipična vrijednost*, odnosno nagli skok ili pad, koji se sa slijedećim podatkom vraća na uobičajeni nivo vremenske serije,
- *prethodna promjena*, odnosno nagli skok ili pad vrijednosti vremenske serije, čije kretanje se zatim polako vraća nazad na uobičajeni nivo vremenske serije.

Ako se *netipična vrijednost* pojavi na kraju vremenske serije, onda je ista privremena *netipična vrijednost* i metodolog za vremensku seriju ga ispravlja jedino kada dobije bar sljedeća tri podatka u vremenskoj seriji, jer tek tada može odrediti vrstu *netipične vrijednosti*. U ovom slučaju nosilac istraživanja mora poslati vremensku seriju (svaki put kad je dodan novi podatak) metodologu vremenske serije na pregled, dok se ta *netipična vrijednost* ne popravi.

Dužina vremenske serije koju desezoniramo, mora biti kod mjesečnih vremenskih serija najmanje 3 godine, kod kvartalnih vremenskih serija barem 4 godine. Postojanje uticaja kalendara sa 1 ili 2 regresora; (1 regresor – radni/neradni dani (ponedeljak-petak/subota- nedjelja); 2 regresora – radni/neradni dani i prestupna godina) uzimamo u obzir kod vremenskih serija koje imaju podatke barem za 5 godina; sa više regresora (6 ili 7 regresora; 6 regresora – svaki dan u sedmici ima svoju tipičnost; 7 regresora – svaki

dan u sedmici i prestupna godina) jedino kod vremenskih serija, koje imaju podatke barem za 7 godina.

SMJERNICE

- Nositelj istraživanja mora vremenske serije pažljivo pregledati i analizirati, jer s ovim može poboljšati kvalitet modela vremenske serije.
- Nositelj istraživanja mora metodologu za vremenske serije dati sve promjene, povezane sa vremenskom serijom (promjene u uzorkovanju, u metodologiji ...).
- Metodolog za vremenske serije se mora posavjetovati sa nosiocem istraživanja u vezi značaja rezultata desezoniranja (uticaj kalendara, uzroci za netipične vrijednosti).
- U objavama podataka, kod međugodišnjih poređenja, treba uvijek porediti desezonirane vrednosti (npr. april 2023 u odnosu na april 2022) prilagođene uticaju broja radnih dana. Pri tome, komponentu trend-ciklus treba predstaviti samo grafički i upozoriti na problem ekstremnih vrijednosti.
- Vrlo duge vremenske serije (više od 12 godina) ponekad treba skratiti (npr. 7 godina), da bismo na taj način poboljšali kvalitet modela. Ako se zaista sezona u ovako dugom razdoblju promjeni, onda su rezultati desezoniranja slabiji. Podaci, koje ne uključujemo u model (npr. stariji od 7 godina), su na raspolaganju samo kao originalni (izvorni) podaci.
- Važno je da nositelj istraživanja ne zaboravi na godišnji pregled i da pošalje metodologu za vremenske serije jednom godišnje pregled svih modela vremenskih serija.

6.3 Interpretacija rezultata

U fazi interpretacije - podatke pretvorimo u informacije, odnosno statističke pojave pojasnimo korisniku i obrazložimo iste na jasan i razumljiv način.

Podaci moraju biti relevantni i korisni, svi zaključci moraju biti potkrijepljeni sa podacima, koje smo dobili u statističkom procesu.

Kod interpretacije podataka, moramo uzeti u obzir način prikupljanja podataka (obuhvat i izvor: istraživanje na bazi uzorka, administrativna baza) i druge informacije povezane sa dobivanjem podataka (npr. stopa neodgovora). Sve ovo treba dati korisniku zajedno sa odgovarajućim metapodacima. Posebno treba opisati sve nedostatke podataka, kao što su npr. odstupanja između ciljne populacije i populacije koju smo posmatrali. Pri interpretaciji podataka treba uzeti u obzir takođe načelo povjerljivosti podataka.

Pri interpretaciji podataka moramo procijeniti, koje pojave ćemo predstaviti korisniku. Pri tome uzimamo u obzir najrelevantnije pojave, aktuelnost tematike i izlazne podatke. Interpretacija podataka treba tome biti prilagođena u zavisnosti od korisničke grupe (da li je u pitanju stručna ili opća javnost).

Opću javnost zanimaju prije svega najvažniji ili najzanimljiviji opći (popularni) statistički podaci i informacije, prikazani na jasan i razumljiv način. *Stručna javnost* pretežno koristi detaljno raščlanjene podatke za potrebe provođenja daljnjih analiza podataka.

SMJERNICE

- Interpretacija rezultata mora biti prilagođena ciljnoj populaciji i medijima (korisnicima) u kojima će podaci biti kasnije objavljeni.
- Interpretacija rezultata mora biti nepristrasna, objektivna, tačna i razumljiva.
- Interpretacija kratkoročnih statistika mora biti drugačija nego interpretacija strukturnih statistika.
- Korištenje referentnih tačaka ima jak uticaj na interpretaciju podataka. Zbog toga moramo (posebno kod vremenskih poređenja) koristiti vremenske tačke koje podatke prikazuju najstabilnije i nepristrasno.
- Pri interpretaciji rezultata u formi indeksa i drugih relativnih brojeva, treba odabrati razumno razdoblje poređenja, koje će korisniku podataka omogućiti efikasnu interpretaciju kretanja bilo koje pojave.
- Kod predstavljanja rezultata u formi indeksa i drugih relativnih brojeva, treba biti oprezan za pravilnu interpretaciju promjena kada su pojave izražene u procentima (%) ili u procentnim poenima. Npr. ako je 2023. godini u poređenju sa 2022. godinom, broj domaćinstava koji su raspolagali PC-ijem povećan sa 21,3 na 23,3 - ovo povećanje ne iznosi 2 % nego 2 postotna poena ili u postotku 9,4 %.

6.4 Statistička zaštita podataka

Statistička zaštita povjerljivih podataka odnosi se na korake koje treba poduzeti s ciljem ublažavanja rizika da statistička jedinica može biti direktno ili indirektno identifikovana u setu podataka.

Aktivnosti na zaštiti povjerljivih podataka podrazumijevaju poduzimanje dva ključna koraka:

- de-identifikaciju podataka, to jest, uklanjanje bilo kakvog direktnog identifikatora (npr. ime/naziv, adresa, identifikacioni/matični broj) iz seta podataka i
- procjenu i upravljanje rizikom od indirektno identifikacije koja se pojavljuje u de-identifikovanom setu podataka.

Uklanjanje identificirajućih informacija (ime/naziv, adresa, identifikacioni/matični broj) zaštićuje statističku jedinicu od **direktno identifikacije**. Međutim, još uvijek je moguće indirektno identifikovati statističku jedinicu u de-identifikovanom setu podataka.

Ako je u setu podataka dostupno puno detalja, identitet određenih statističkih jedinica može biti izveden iz postojećih veoma rijetkih karakteristika ili kombinacijom posebnih karakteristika.

Primjer 1: Identitet osobe može se posredno izvesti, ako u setu podataka osoba u dobi 65 godina i više ima godišnji prihod veći od milion KM, a živi u gradu od 10 000 stanovnika.

Primjer 2: Identitet osobe s vrlo rijetkom bolesti može se otkriti i u vrlo agregiranim podacima.

Statistička zaštita povjerljivih podataka uključuje uklanjanje ili izmjenu informacija, kao i reduciranje (sažimanje) detalja, kako bi osigurali da niti jedna statistička jedinica ne bude identifikovana u podacima (bilo direktno ili indirektno).

Postoje razne metode koje za cilj imaju zaštititi identitet statističkih jedinica, a da u isto vrijeme omoguće objavljivanje dovoljno detaljnih i korisnih informacija s ciljem korištenja istih za statističke i istraživačke svrhe.

6.4.1 Upravljanje rizikom od identifikacije povjerljivih statističkih podataka

Najveći izazov izrade javno dostupnih statističkih podataka je osigurati, da se niti jedna statistička jedinica ne identifikuje u podacima.

Sam proces (koraci) upravljanja rizikom od identifikacije je isti za sve podatke koji se objavljuju (*mikropodatke* - gdje svaki zapis predstavlja opažanje za pojedinca ili pravnog subjekta ili *makropodatke* - gdje se agregirani podaci prikazuju u tabelama).

Prvi korak upravljanja rizikom od identifikacije je procjena potencijalnih rizika od identifikacije. **Drugi korak** je upravljanje rizikom od identifikacije pomoću odgovarajuće metode za zaštitu povjerljivih podataka.

Procjena potencijalnih rizika od identifikacije povjerljivih statističkih podataka

Vjerovatnoća da statistička jedinica bude identifikovana ovisi o nizu faktora:

- *broja detalja uključenih u podatke* (više detalja – veći rizik identifikacije),
- *osjetljivosti varijabli unutar podataka* (varijable kao što su informacije o finansijskim, zdravstvenim ili kriminalnim zapisima mogu povećati rizik od identifikacije) i
- *načina prezentiranja informacija* (makropodaci objavljeni u tabelama predstavljaju manji rizik identifikacije od objavljivanja mikropodataka).

6.4.2 Kako zaštititi povjerljive podatke – osnovni principi

Identifikacija rizika u diseminiranim podacima (kontrola otkrivanja) podrazumijeva poduzimanje koraka u cilju procjene i smanjenja rizika od otkrivanja identiteta određene statističke jedinice. Cilj je zaštititi identitet statističke jedinice, a u isto vrijeme maksimalno zadržati korisnost podataka.

U jednostavnim slučajevima, podaci mogu biti zaštićeni manuelno. Međutim, ponekad je neophodno korištenje softvera za zaštitu. U tom slučaju zahtijevaju se specijalizirane vještine i znanja o podacima, da bi se korektno zaštitili povjerljivi podaci.

Makropodaci

Makropodaci (uobičajeno tabele) su uobičajeni proizvodi statističke institucije i najčešći način prezentiranja agregiranih podataka. Tabele mogu sadržavati i individualne informacije.

Postoje dvije glavne vrste tabela:

- **Tabele frekvencija** - u svakoj ćeliji tabele upisan je broj statističkih jedinica koje pripadaju toj ćeliji (npr. broj osoba u različitim starosnim grupama, ili broj pravnih subjekata u nekoj djelatnosti).
- **Tabele vrijednosti** - u svakoj ćeliji upisan je zbir vrijednosti određene varijable za statističke jedinice koje pripadaju toj ćeliji (ćelije sadrže vrijednosti izračunate iz numeričkih odgovora - npr. ukupan prihod ili dobit).

Ponekad je moguće otkriti informacije o statističkoj jedinici iz ove dvije glavne vrste tabela. Npr., kada u *tabeli frekvencija* ćelija ima malu frekvenciju (tj. vrlo mali broj slučajeva), moguće je izvesti zaključak o određenoj statističkoj jedinici koristeći već poznate informacije i dodatne informacije prezentirane u tabeli.

U *tabeli vrijednosti*, riziku od identifikacije izložene su ćelije u tabeli u kojima se dominantne vrijednosti (više od 85% vrijednosti ćelije Ukupnoa) odnose na jednu ili dvije statističke jedinice.

Rizik od identifikacije postoji i ako korisnici imaju pristup višestrukim tabelama koje sadrže neke zajedničke elemente. Moguće je koristiti podatke iz jedne tabele za određivanje identiteta jedinice prema dostupnim podacima u drugoj tabeli, što ukazuje na to da je važno pratiti sve informacije iz seta podataka u svim tabelama koje se objavljuju.

Specijalizirani softverski alat koji se najčešće koristi u procjeni potencijalnih rizika od identifikacije u zapisima makropodataka je t-Argus.

Mikropodaci

Mikropodaci predstavljaju bazu individualnih podataka o statističkim jedinicama i veoma su vrijedan resurs za istraživače i kreatore politika. Međutim, veliki izazov za osobe koje upravljanju mikropodacima je održavanje prave ravnoteže između, s jedne strane ispunjavanja obaveze zaštite identiteta statističkih jedinica, a s druge strane, maksimiziranja informacija dostupnih za statističke i istraživačke namjene. To zahtijeva pažljivo balansiranje između rizika od identifikacije i koristi od dostupnih informacija.

Postoje dvije ključne vrste rizika od otkrivanja povezane sa mikropodacima:

- rizik identifikacije bez ikakve namjere da se identifikuje statistička jedinica – tz. spontana identifikacija (otkrivanje) i
- rizik povezan sa namjernim (zlonamjernim) pokušajem otkrivanja statističke jedinice.

Spontano otkrivanje - predstavlja identifikaciju učinjenu bez ikakve namjere, događa se kada su u podacima prisutne statističke jedinice sa rijetkim karakteristikama. Npr., baza podataka može uključivati osobe sa neuobičajenim poslom (npr. estradna zvijezda) ili veoma visoke prihode koji su veoma vidljivi u podacima i mogu voditi do njihove identifikacije.

Namjerno otkrivanje - predstavlja identifikaciju statističke jedinice u setu podataka koristeći podudarnost jedinstvenih zapisa u vanjskim datotekama kombinirajući obilježja koja su zajednička za oba seta podataka.

Rizik vezan uz pristup mikropodacima može se ublažiti na razne načine, uključujući:

- direktnu zaštitu podataka;
- odvrćanjem od bilo kakvog pokušaja identifikacije (npr. zakonskim mjerama koje ne dozvoljavaju pokušaj identifikacije, te propisivanjem kazne za kršenje propisa);
- ograničavanjem pristupa mikropodacima;
- edukovanje korisnika podataka o važnosti zaštite privatnosti, kao i njihovih obaveze u vezi zaštite (npr. davanjem obuka i detaljnih instrukcija); i
- pristup podacima je obezbjeđen preko primjerenog okruženja.

Već smo naglasili da nije moguće eliminirati sve rizike od identifikacije. Cilj je ublažiti rizike od identifikacije statističke jedinice, a da u isto vrijeme maksimiziramo korisnost podataka za statističke i istraživačke svrhe.

Različiti programski paketi su dostupni kao pomoć u procjeni potencijalnih rizika od identifikacije u zapisima mikropodataka, kao što su:

- μ - ARGUS - programski paket razvijen od strane statistike Holandije.
- SUDA - softver razvijen od strane Univerziteta u Manchesteru (UK).

6.4.3 Pravila identifikacije osjetljivih podataka i metode statističke zaštite povjerljivih podataka u makropodacima

Pravila identifikacije osjetljivih podataka u makropodacima/tabelama

Za identifikaciju ćelija u tabelama koje predstavljaju rizik od identifikacije, moraju biti određena pravila zaštite povjerljivih podataka i primijenjena na svaku ćeliju u tabeli.

Ako ćelija ne zadovoljava takva pravila, onda je potrebna daljnja analiza u cilju minimiziranja rizika od identifikacije.

Za procjenu rizika od identifikacije u ćelijama tabele koriste se sljedeća pravila za određivanje primarno osjetljivih ćelija:

- **pravilo praga (minimalnog broja slučajeva/statističkih jedinica)** - ćelija je osjetljiva, ako njenoj vrijednosti pridonosi manje od ***t*** slučajeva (određujemo vrijednosti praga za minimalni broj slučajeva u bilo kojoj ćeliji).
- **pravilo dominantnosti** (n, k) - ako ***n*** jedinica predstavlja više od ***k*** % vrijednosti ćelije, u tom slučaju ćelija je osjetljiva. Znači, ovo pravilo se primjenjuje za ćelije, gdje mali broj statističkih jedinica doprinosi velikom postotku Ukupnoa u ćeliji.

Pravilo praga

U *Tabeli 1*, u primjeru gdje imamo da sve osobe u dobi od 15 do 19 godina imaju nizak prihod (20) ova ćelija ne zahtijeva zaštitu, jer ovaj rezultat ne može biti neočekivan ili osjetljiv. Međutim, ako bismo imali slučaj da osobe u dobi 15 - 19 godina imaju visok umjesto niskog prihoda, onda ta ćelija mora biti zaštićena od rizika identifikacije, jer je rezultat neočekivan i smatra se osjetljivim. Pravilo minimalnog broja

slučajeva štiti rijetke i jedinstvene kombinacije karakteristika koje bi mogle dovesti do ponovne identifikacije. Kriterij je dakle ispunjen ako najmanje tri slučaja doprinose ćeliji tabele, što znači da vrijednost treba držati povjerljivom ako tome doprinose samo 1 ili 2 slučaja.

Primjer: U Tabeli 1, starosna grupa 50-59 godina ima frekvenciju **2** u ćeliji niskih prihoda i treba biti zaštićena. Primjenjujući pravilo praga/minimalnog broja slučajeva **t** za graničnu vrijednost 3, očigledno je da ćelija sa manje od 3 frekvencije predstavlja rizik od otkrivanja.

Tabela 1 - Ostvareni prihodi po starosnim grupama

Starosne grupe	Prihodi			
	Nizak	Srednji	Visok	Ukupno
15-19	20	0	0	20
20-29	14	11	8	33
30-39	8	12	7	27
40-49	6	18	24	48
50-59	2	5	14	21
60+	12	9	7	28
Ukupno	62	55	60	177

Napomena: Pravilo praga (minimalnog broja slučajeva) ne primjenjuje se ako postoji samo jedna karakteristika koja je po logici jedino moguća za grupu. Podatak da su npr., sve trudnice žene ne smatra se povjerljivim.

Pravilo dominantnosti

Pravilo *dominantnosti* naziva se također i **n, k** pravilo i koristi se za tabele vrijednosti. U skladu s pravilom dominantne ćelije, ćelija se smatra nezaštićenom ako kombinirani doprinos '**n**' najvećih članova ćelije predstavlja više od '**k**'% od ukupne vrijednosti ćelije. Vrijednosti '**n**' i '**k**' se određuju od strane vlasnika podataka.

Primjer: Posmatrajmo Tabelu 2 - *Ostvarena dobit po poslovnim subjektima* u proizvodnji namještaja. Unutar tabele postoji jedna ćelija koja pokazuje ukupnu dobit za sve poslovne subjekte u iznosu od 7 000 000 KM. Da bismo primijenili pravilo dominantne ćelije, dobit svih statističkih jedinica (u konkretnom slučaju poslovnih subjekata u proizvodnji namještaja) mora biti poznata.

Recimo da postoji osam poslovnih subjekata, obilježenih od A - H. Koristeći pravilo (2, 85), dvije najveće statističke jedinice (A i B) ne smiju sudjelovati sa više od 85% ćelije Ukupnoa (tj. ukupne dobiti u industriji namještaja).

Kombinirana dobit od dva najveća poslovna subjekta (A i B) iznosi 6 090 000 KM (3 760 000 KM + 2 330 000 KM) što je 87% ukupne dobiti u industriji namještaja (7 000 000 KM), što je više od 85% Ukupnoa (ukupne dobiti svih poslovnih subjekata u proizvodnji namještaja).

Iz ovih razloga, koristeći pravilo (2, 85), dobit ostvarena od strane ova dva poslovna subjekta je povjerljiva i treba biti zaštićena.

Tabela 2 – Ostvarena dobit (u KM) po poslovnim subjektima u proizvodnji namještaja

Poslovni subjekti	Dobit
A	3 760 000
B	2 330 000
C	300 000
D	250 000
E	150 000
F	100 000
G	75 000
H	35 000
Ukupna dobit u proizvodnji	7 000 000

Međutim, ako se ovi poslovni subjekti slažu (dali pismeni pristanak/odobrenje) da se njihovi individualni podaci objave, ili ako se dio podataka dobije iz javnog izvora podataka, tada ovi poslovni subjekti ne moraju biti zaštićeni i na taj način smanjujemo gubitak informacija u tabelama u kojima se pojavljuju ovi podaci.

Metode statističke zaštite povjerljivih podataka u makropodacima/tabelama

Metode koje se koriste za statističku zaštitu povjerljivih podataka koji predstavljaju rizik od identifikacije u nezaštićenoj ćeliji tabele, odnosno u makropodacima su:

1. Primarna i sekundarna supresija (potiskivanje) podataka u ćelijama
2. Preoblikovanje (redizajniranje) tabele
3. Modifikacija (perturbacija) podataka.

Primarna i sekundarna supresija ćelija

Supresija (potiskivanje) podataka u ćelijama obuhvata neobjavljivanje informacija za nesigurne ćelije, ili brisanje pojedinačnih zapisa iz seta mikropodataka.

Ako tabela sadrži Ukupno, moguće je izračunati vrijednost izostavljene ćelije oduzimanjem vrijednosti drugih ćelija iz Ukupnoa. Znači, najmanje jednu dodatnu ćeliju također moramo izostaviti, da bismo spriječili identifikaciju.

Supresija ćelije koja potpada pod jedno od pravila povjerljivosti naziva se **primarna supresija ćelija**. Supresija drugih ćelija (kako bi se spriječilo otkrivanje primarno potisnute ćelije) naziva se **sekundarna ili posljedična supresija**.

Primjer: U Tabeli 1 (na strani 10), ćelija koja u tabeli sadrži broj osoba koje zarađuju nizak dohodak u starosnoj dobi 50-59 godina je identifikovana kao nezaštićena ćelija koristeći pravilo 'praga 3'. Ova ćelija može biti zaštićena supresijom kako je prikazano u Tabeli 3 sa 'x'. Međutim, i pored toga još uvijek je moguće odrediti vrijednost ćelije koristeći preostale vrijednosti. Npr., osobe koje zarađuju nizak dohodak u dobi 50-59

mogu biti otkrivene oduzimanjem srednjeg i visokog dohotka od Ukupnoa ($21-14-5=2$). Da bi zaštitili ćeliju potrebno je izvršiti posljedičnu, odnosno sekundarnu supresiju ćelije.

Tabela 3 - Sekundarna supresija

Starosne grupe	Prihodi			
	Nizak	Srednji	Visok	Ukupno
5-19	20	0	0	20
20-29	14	11	8	33
30-39	8	12	7	27
40-49	6 Y	18 Y	24	48
50-59	X	5 Y	14	21
60+	12	9	7	28
Ukupno	62	55	60	177

U Tabeli 4, ćelije s najmanjim vrijednostima su potisnute. Ćelija označena sa **X** označava nezaštićenu ćeliju (primarna supresija). Ćelija označena sa **Y** se takođe treba izostaviti, kako bi se spriječilo da se nezaštićena ćelija ne otkrije indirektno iz Ukupnoa (*sekundarna supresija*).

Tabela 4 - Prikrivanje Ukupnoa

Starosne grupe	Prihodi			
	Nizak	Srednji	Visok	Ukupno
15-19	20	0	0	20
20-29	14	11	8	33
30-39	8	12	7	27
40-49	6	18	24	48
50-59	X	5	14	21 >19
60+	12	9	7	28
Ukupno	62	55	60	177

Koristeći ovu metodu, ćelija označena sa **X** će biti izostavljena, a sljedeći red i kolona Ukupnoa će biti prikriveni (na način prikazan u Tabeli 4) da bismo spriječili da se naknadno (indirektno) utvrdi vrijednost nezaštićene ćelije.

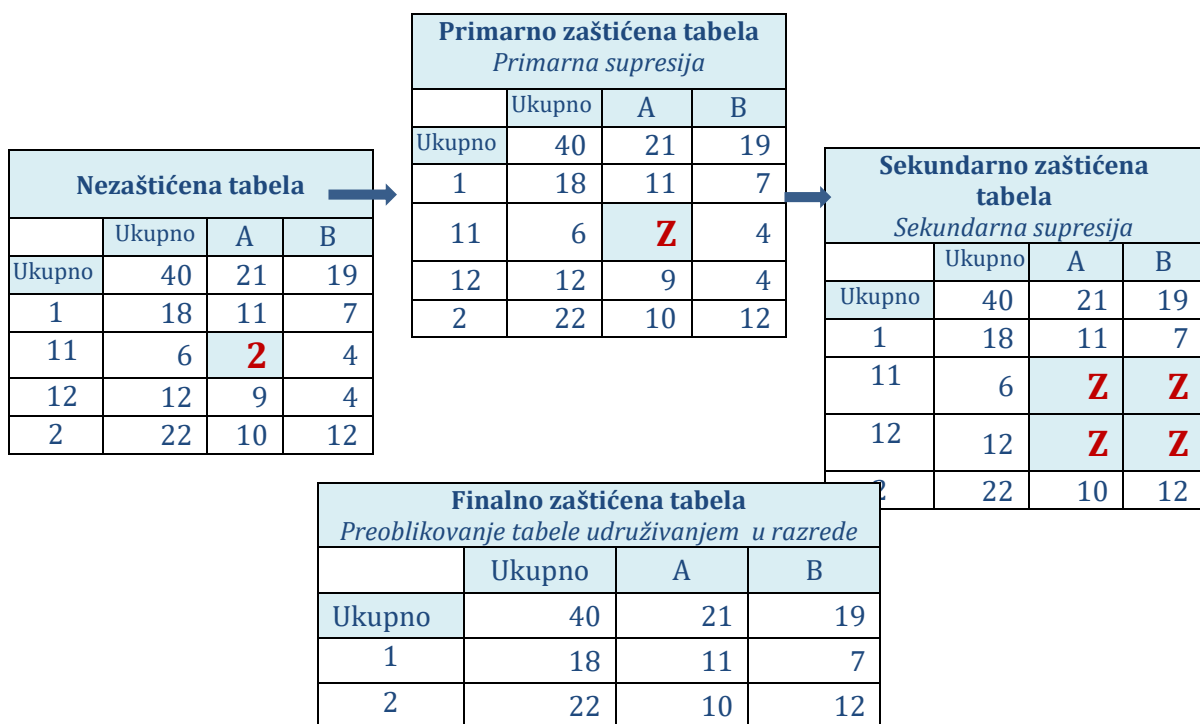
Primjer statističke zaštite tabele

Pretpostavimo da je pravilo za određivanje primarno osjetljivih ćelija prag 3. Znači, tabela sadrži samo jednu primarno osjetljivu ćeliju, tj. ćeliju [11, A] koja ima vrijednost 2.

Ova primarna osjetljiva ćelija može biti zaštićena na više načina.

Ako je štitimo kombiniranjem u razrede, smanjujemo broj primarno osjetljivih ćelija preoblikovanjem tabele: uklanjamo obje podkategorije kategorije 1.

Ako se odlučimo zaštititi metodom supresije - primarnom supresijom možemo prikriti ćeliju [11, A]; sekundarnom supresijom prikrivamo osjetljive ćelije [11, B], [12, A], [12, B] i time onemogućavamo sprječavamo korisnika da koristeći jednostavne matematičke funkcije otkrije pravu vrijednost primarne osjetljive ćelije.



Preoblikovanje (redizajniranje) tabele

Povjerljivost davaoca podataka štiti se izborom odgovarajuće agregacije. Ovaj metod uključuje agregiranje nekoliko kategorija odgovora u jedan, ili reduciranje klasifikacijskih detalja dostupnih u tabelama. Važno je dobro poznavanje materije, jer kombiniranje, udruživanje kategorija treba osigurati relevantnost novih kategorija grupisanja za korisnike podataka.

Klasifikacije koje su vrlo detaljne, kao što su zemlja rođenja, djelatnost ili zanimanje, mogu se sažimati ili udruživati na širem nivou. Na primjer, kategorije zanimanja za medicinske sestre i ljekare mogu se udružiti u kategoriju „medicinske profesije“.

Za kvantitativne i/ili kontinuirane podatke kao što su dohodak ili starosna dob, kategorije također mogu biti udružene u rangove.

Tabela 5 - redizajniranje podataka u tabeli

KD	Veličina klase					
	Ukupn	0-4	5-9	10 -	15 -	20+
D	375	73	194	59	36	13
DA	90	29	55	1	5	-
DB	53	5	25	11	2	10
DC	30	3	4	12	11	-
DD	161	26	99	35	1	-
DE	41	10	11	-	17	3

Tabela 5.1 - redizajnirana tabela

KD	Veličina klase			
	Ukupno	0-	5-9	10+
D	375	73	194	108
DA	90	29	55	6
DB	53	5	25	23
DC	30	3	4	23
DD	161	26	99	36
DE	41	10	11	20

Modifikacija podataka (perturbacija)

Perturbacija je metoda modifikacije podataka koja podrazumijeva neznatnu promjenu podataka u cilju smanjenja rizika od otkrivanja, uz zadržavanje što je moguće više strukture.

Najčešći metod modifikacije podataka je *zaokruživanje podataka*, odnosno neznatno mijenjanje malih ćelija u tabeli, kako bi se osiguralo da analize koje se obavljaju na temelju ovakvih podataka ne utiču značajnije na rezultate iste (a izvorne vrijednosti sa sigurnošću ne mogu biti poznate).

Zaokruživanje podataka podrazumijeva zamjenu malih vrijednosti koji se prikazuju u tabeli s drugim malim slučajnim brojevima. To rezultira modifikacijom u nekim podacima, tako da zbir vrijednosti ćelije unutar ili između tabela nije jednak Ukupnoj tabeli. Nasumično zaokruživanje sa osnovicom X podrazumijeva nasumično mijenjanje svakog broja u tabeli.

Na primjer, nasumično zaokruživanje osnovom 3, znači da se sve vrijednosti u tabeli zaokružuju na najbliži broj koji je djeljiv sa 3. Svaka vrijednost, uključujući Ukupno, zaokružuje se nezavisno. Vrijednosti koje su već djeljive sa 3 ostavljaju se neizmjenjene.

Zaokruživanje sa osnovom 3

Tabela 6 - Prije zaokruživanja

Starosne grupe	Prihodi			
	Nizak	Srednji	Visok	Ukupno
5-19	20	0	0	20
20-29	14	11	8	33
30-39	8	12	7	27
40-49	6	18	24	48
50-59	2	5	14	21
60+	12	9	7	28
Ukupno	62	55	60	177

Tabela 6.1 - Nakon zaokruživanja

Starosne grupe	Prihodi			
	Nizak	Srednji	Visok	Ukupno
15-19	21	0	0	21
20-29	15	12	9	33
30-39	9	12	6	27
40-49	6	18	24	48
50-59	3	6	15	21
60+	12	9	6	27
Ukupno	63	54	60	177

Koraci u postupku zaštite podataka u tabeli

1. korak: definirati pravilo za mjerenje osjetljivosti tabele (tj. odrediti koje su ćelije osjetljive)
2. korak: ako je broj osjetljivih ćelija suviše veliki, redizajnirati tabelu
3. korak: izabrati metod zaštite
4. korak: zaštititi tabelu
5. korak: zadovoljstvo rezultatima zaštite (zaštita u odnosu na gubitak informacija)
6. korak: ako je 5 da, publikovati tabelu
7. korak: ako je 5 ne, ići ponovo na korak 2

REZULTAT: osjetljive ćelije ne mogu biti djelimično ili potpuno otkrivene

6.4.4 Pravila identifikacije osjetljivih podataka i metode statističke zaštite povjerljivih podataka u mikropodacima

Pravila identifikacije osjetljivih podataka u nezaštićenom zapisu u mikropodacima

Za procjenu rizika od identifikacije u nezaštićenom zapisu u mikropodacima koriste se sljedeća pravila za određivanje osjetljivih podataka:

- **Pravilo praga** (najmanja frekvencija određenih kombinacija pojedinih varijabli, koja može biti objavljena)
- Javne baze podataka se ne štite (jer je riječ o javno dostupnom izvoru podataka)

Metode statističke zaštite povjerljivih mikropodataka

Postoje razne metode (tehnike) koje se koriste za statističku zaštitu povjerljivih podataka koji predstavljaju rizik od identifikacije u nezaštićenom zapisu u mikropodacima.

Za zaštitu povjerljivosti mikropodataka se koriste metode perturbacije i redukcije podataka. Ovo su isti osnovni principi koji se koriste i za zaštitu agregiranih (makropodataka) podataka.

Popularne tehnike za zaštitu povjerljivosti mikropodataka uključuju:

- *ograničavanje broja varijabli* uključenih u set podataka;
- *udruživanje (spajanje) kategorija* za koje postoji vjerovatnoća da mogu omogućiti identifikaciju (npr. iskazivanje starosti u petogodišnjem rangu);
- *top/bottom šifriranje* ekstremnih vrijednosti kontinuiranih varijabli (prihod, starost);
- *potiskivanje (supresija) određene vrijednosti ili zapisa* koja se inače na drugi način ne može zaštititi od rizika od identifikacije; i
- *zamjena podataka* uključuje zamjenu vrijednosti u zapisu sa vrijednostima u drugom zapisu sa sličnim karakteristikama da bi sakrili jedinstvenost zapisa.

Ako se u postupku statističke zaštite mikropodataka odlučimo za metode korištene na osnovnim mikropodacima, prvo moramo izabrati statistike čije ćemo vrijednosti pokušati što manje izmijeniti.

Najprije trebamo definirati sljedeće setove varijabli:

- **identifikacijske varijable** - varijable na osnovu kojih se može otkriti jedinica:
 - direktni identifikatori - omogućavaju identifikaciju statističke jedinice direktno, to su varijable koje jedinstveno određuju statističku jedinicu (npr. matični/identifikacioni broj);
 - indirektni identifikatori - omogućavaju identifikaciju statističke jedinice indirektno, to su varijable koje mogu dovesti do otkrivanja statističke jedinice u kombinaciji s drugim varijablama (npr. Ime/naziv, godina rođenja, adresa, spol, dob, zanimanje, regija);
- **povjerljive izlazne varijable** - osjetljive varijable koje nose osjetljivu informaciju o statističkoj jedinici (npr. dohodak, religija, politička uvjerenja, zdravstveno stanje ...);
- **nepovjerljive izlazne varijable** - ostale varijable.

Moramo biti svjesni da i pored korištenja svih tehnika statističke zaštite, rizik od identifikacije (otkrivanja) povjerljivih podataka nikada nije nula.

U sljedećem koraku definiramo metode statističke zaštite i koristimo ih na podacima. Najprije, uklanjamo iz zapisa mikropodataka sve direktne identifikatore (matični/identifikacioni broj) i određene indirektne identifikatore (npr. ime/naziv, prezime, adresa). Takođe, po potrebi možemo zaštititi i određene povjerljive izlazne varijable (npr. dohodak, religija, zdravstveno stanje).

Preostali indirektni identifikatori (npr. dob, spol, godina rođenja, zanimanje, regija) su varijable koje zadržavamo, jer su važne za statističku analizu. Te se varijable nazivaju *ključnim varijablama*.

Zatim, izaberemo metode zaštite koje ćemo koristiti na ključnim varijablama. Metode zaštite izaberemo prema namjeni podataka (ovisi od toga da li je riječ o datoteci za javnost, datoteci za istraživače, datoteci za studente). Metode zaštite se koriste u mjeri u kojoj se smanjuje rizik otkrivanja statističkih jedinica u zaštićenoj datoteci mikropodataka na razumnu (prihvatljivu) mjeru.

Naravno, korisnicima podataka trebamo objasniti koje smo metode zaštite koristili na mikropodacima, jer određene podatke nužno mijenjamo, tako da se vrijednosti u zaštićenoj datoteci razlikuju od vrijednosti u izvornoj datoteci. Korisnicima ne smijemo objaviti vrijednosti parametara koje smo koristili prilikom primjene metoda zaštite (npr. set indirektnih identifikatora, minimalni prag kod metode nedostajućih vrijednosti, parametar pravila dominantnosti), jer bi otkrivanjem parametara smanjili zaštitu statističkih jedinica.

Primjer statističke zaštite mikropodataka

Uzmimo datoteku osoba (koju želimo statistički zaštititi) sa određenim demografskim karakteristikama i podatkom o ukupnim mjesečnim prihodima.

GODINA ROĐENJA	SPOL	MJESTO PREBIVALIŠTA	ZANIMANJE	UKUPNI MJESEČNI PRIHODI
1985	M	ZENICA	EKONOMISTA	1800
1985	M	ZENICA	EKONOMISTA	1900
1985	M	ZENICA	EKONOMISTA	2100
1985	M	ZENICA	EKONOMISTA	2500
1985	M	ZENICA	EKONOMISTA	1900
1985	M	ZENICA	EKONOMISTA	1850
1985	M	ZENICA	PRAVNIK	1800
1985	M	ZENICA	PRAVNIK	1890
1985	M	ZENICA	PRAVNIK	1550
1985	M	ZENICA	PRAVNIK	1800
1985	M	DOBOJ	EKONOMISTA	1950
1985	M	DOBOJ	EKONOMISTA	1650

1985	M	DOBOJ	PRAVNIK	1850
1985	M	DOBOJ	PRAVNIK	1600
1985	M	DOBOJ	PRAVNIK	1550
1985	M	TUZLA	EKONOMISTA	1900
1985	M	TUZLA	PRAVNIK	2100
1985	M	TUZLA	PRAVNIK	1850
1985	M	TUZLA	PRAVNIK	1750

Prije svega, primjetno je da su već uklonjeni direktni (npr. matični broj) i neki od indirektnih identifikatora (npr. ime, prezime, adresa), čije bi kombinacije u većini slučajeva omogućile neposrednu identifikaciju.

U sljedećem koraku za preostale indirektno identifikatore određujemo prag 3, odnosno za godinu rođenja, spol, mjesto prebivališta, zanimanje. Definiranjem ovog praga osiguravamo da se u zaštićenoj datoteci moraju pojaviti najmanje 3 identične kombinacije. Izračunamo kombinacije preostalih indirektnih identifikatora (godina rođenja x spol x mjesto prebivališta x zanimanje).

KOMBINACIJE PREOSTALIH INDIREKTNIH IDENTIFIKATORA	BROJ KOMBINACIJA
1985 - M - ZENICA - EKONOMISTA	6
1985 - M - ZENICA - PRAVNIK	4
1985 - M - DOBOJ - EKONOMISTA	2
1985 - M - DOBOJ - PRAVNI	3
1985 - M - TUZLA - EKONOMISTA	1
1985 - M - TUZLA - PRAVNIK	3

Nalazimo da imamo dvije kombinacije (1985 - M - DOBOJ - EKONOMISTA i 1985 - M - TUZLA - EKONOMISTA), koje se pojavljuju rjeđe od granične vrijednosti (praga 3).

Za zaštitu koristimo metod supresije (potiskivanja). Kod ove metode zaštite, prikrivamo vrijednost jednog od indirektnih identifikatora. U navedenom primjeru, statistički zaštićenu datoteku ćemo dobiti tako da zamijenimo mjesto prebivališta s nepoznatom vrijednosti (**99**).

GODINA ROĐENJA	SPOL	MJESTO PREBIVALIŠTA	ZANIMANJE	UKUPNI MJESEČNI PRIHODI
1985	M	ZENICA	EKONOMISTA	1800
1985	M	ZENICA	EKONOMISTA	1900
1985	M	ZENICA	EKONOMISTA	2100
1985	M	ZENICA	EKONOMISTA	2500
1985	M	ZENICA	EKONOMISTA	1900
1985	M	ZENICA	EKONOMISTA	1850
1985	M	ZENICA	PRAVNIK	1800
1985	M	ZENICA	PRAVNIK	1890
1985	M	ZENICA	PRAVNIK	1550

1985	M	ZENICA	PRAVNIK	1800
1985	M	99	EKONOMISTA	1950
1985	M	99	EKONOMISTA	1650
1985	M	DOBOJ	PRAVNIK	1850
1985	M	DOBOJ	PRAVNIK	1600
1985	M	DOBOJ	PRAVNIK	1550
1985	M	99	EKONOMISTA	1900
1985	M	TUZLA	PRAVNIK	2100
1985	M	TUZLA	PRAVNIK	1850
1985	M	TUZLA	PRAVNIK	1750

Kada iznova izračunamo sve kombinacije u statistički zaštićenoj datoteci, možemo vidjeti da se sada sve kombinacije pojavljuju najmanje tri puta.

KLJUČ	BROJ KOMBINACIJA
1985 - M - ZENICA - EKONOMISTA	6
1985 - M - ZENICA - PRAVNIK	4
1985 - M - 99 - EKONOMISTA	3
1985 - M - DOBOJ - PRAVNIK	3
1985 - M - TUZLA - PRAVNIK	3

Koraci u postupku statističke zaštite mikropodataka

1. korak - definirati pravilo za mjerenje osjetljivosti mikropodataka
2. korak - izabrati metodu zaštite
3. korak - zaštititi fajl mikropodataka
4. korak - jesmo li zadovoljni rezultatima (zaštita u odnosu na gubitak informacija)
5. korak - ako je 4 da, dati korisniku fajl sa mikropodacima
6. korak - ako je 4 ne, redizajnirati varijable, vratiti se na korak 3

REZULTAT: individualni podaci u fajlu sa mikropodacima ne mogu biti prepoznati

Aktivna zaštita povjerljivosti

Sve statistike (izuzev statistike međunarodne trgovine robama i statistike prodaje proizvoda prerađivačke industrije) treba da primijenjuju tehnike aktivne zaštite povjerljivosti. Najvažnije pravilo o povjerljivosti je primjena kriterija praga broja slučajeva, tj. minimalni zahtjev od tri opažanja (prag 3), za relevantne podatke u ćeliji tabele koja se objavljuje.

Pored primjene kriterija praga broja slučajeva u tehnikama aktivne zaštite povjerljivosti, primjenjuje se i tzv. *pravilo dominantnosti* za ekonomske varijable. To znači, da dvije najveće statističke jedinice zajedno ne smiju sudjelovati sa više od 85% ćelije Ukupnoa (npr. ukupne dobiti u nekoj djelatnosti).

Pri donošenju odluke o objavljivanju podataka, prvo primijenjujemo *primarno potiskivanje (supresiju) podataka odnosno primarnu zaštitu povjerljivosti*, gdje se pojedina grupisanja posmatraju odvojeno i gdje su kriteriji broj i/ili dominantnost. Ovo dopunjujemo *sekundarnim potiskivanjem odnosno sekundarnom zaštitom povjerljivosti*, koju izvodimo manuelno, a čiji je cilj spriječiti korisnika statistike da rekonstruira povjerljive podatke za grupu putem objavljenih brojki za druge grupe.

Ovo je posebno prisutno u publikacijama detaljnih statistika sa raznim kombinacijama u hijerarhijskim strukturama kao što su industrija, veličine grupa i geografske oblasti.

Pasivna zaštita povjerljivosti

Statistika međunarodne trgovine robom/uslugama ima posebne zahtjeve za povjerljivošću koji se razlikuju od onih u drugim poslovnim statistikama, jer se zaštita povjerljivosti provodi samo na zahtjev pravne osobe. Za ovu statistiku primjenjujemo tzv. *pasivnu zaštitu povjerljivosti* gdje pravne osobe same moraju zahtijevati (tražiti) zaštitu povjerljivosti. To je zbog činjenice da je u ovom slučaju statistička jedinica roba/usluga, a ne pravna osoba. Osim toga, nivo detaljnosti klasifikacije proizvoda potrebnih za EU je takva, da bi aktivna zaštita povjerljivosti na nivou pravne osobe značila da se ova statistika ne može objaviti na bilo koji smislen način. Argument za ovakav postupak je da set podataka ove statistike sadrži robu/uslugu kao statističku jedinicu, a ne pravnu osobu. Iz ovog razloga, teško je otkriti informacije o pojedinim pravnim osobama u setovima podataka za ove statistike.

U načelu, pasivna zaštita povjerljivosti implicira da se tabelarni rezultati ne tretiraju povjerljivim. Međutim, moguće je da pravne osobe mogu držati u tajnosti rezultate za određeni proizvod u cijelosti ili djelomično, ukoliko prema određenim kriterijima, oni imaju dominantan položaj na tržištu za određeni proizvod. Drugim riječima, pravna osoba mora zatražiti od Agencije da zaštiti povjerljivost ovih podataka. Ako pravna osoba ne poštuje kriterije dominantnog položaja na tržištu, zahtjev za tajnošću (zaštitom) će biti odbijen.

U slučaju da je statistika međunarodne trgovine robom/uslugama razvrstana prema karakteristikama pravnih osoba, onda je povjerljiva na isti način kao i druge poslovne statistike, osim što se kalkulacija dominantnosti temelji na prometu.

Pravila pristupa korisnika povjerljivim podacima

Pristup vanjskih korisnika povjerljivim podacima

Vanjski korisnici mogu dobiti pristup anonimiziranim individualnim podacima (statistički zaštićenim mikropodacima) ako ispunjavaju određene uvjete.

Dozvoljen je pristup mikropodacima za istraživačke namjene i to: registriranim znanstveno-istraživačkim subjektima i registriranim istraživačima. Pristup mikropodacima za istraživačke namjene se odobrava isključivo nakon:

- popunjavanja *Obrasca za pristup mikropodacima za istraživačke namjene*, odnosno *Obrasca za pristup mikropodacima za analitičke namjene u državnim tijelima*,
- potpisivanja *Izjave o zaštiti podataka* i
- zaključivanja *Ugovora o pristupu i korištenju podataka*.

Mikropodaci za istraživačke namjene su trenutno dostupni samo na *prijenosnom mediju*, a kada se osiguraju tehničke pretpostavke pristup mikropodacima će biti dostupan u *zaštićenoj prostoriji*, ili putem *pristupa s daljine*.

Prije samog odobravanja pristupa mikropodacima za istraživačke namjene, neophodno je da istraživački subjekt, odnosno registrirani istraživač podnese na uvid dokaz o registraciji.

Znači, mikropodaci za istraživačke namjene dostupni su samo u anonimiziranoj formi. Potpuna anonimizacija nije potrebna, ali određena ograničenja u pristupu podacima (npr. u formi ograničenog broja varijabli) su neophodna ako je rizik prepoznavanja statističkih jedinica previše očigledan.

Pristup internih korisnika povjerljivim podacima

Da bi zaposlenici Agencije dobili pristup podacima, njihovi nadređeni moraju pristup odobriti i to mora biti opravdano zahtjevom za podacima u strogom odnosu prema setu statističkih podataka s kojima dotični zaposlenik radi.

SMJERNICE

- Zaštita tajnosti je ključni element u očuvanju povjerenja statističkih jedinica.
- Prije nego se publiciraju rezultati istraživanja, obavezno provjeriti da li je došlo do narušavanja povjerljivosti. Ovo vrijedi kako za publikacije i posredovanje podataka u agregiranom obliku (tabele distribucije i indeksi), tako i za posredovanje mikropodataka.
- Imena i adrese treba što je prije moguće izbrisati iz setova pojedinačnih podataka. Sigurno postupanje i čuvanje pojedinačnih setova podataka, pridržavanjem svih propisa vezanih za zaštitu podatka, obaveza je svih uposlenika u statističkim institucijama u BiH.
- Statističke jedinice treba upoznati da će statističke institucije poduzeti sve neophodne mjere na očuvanju povjerljivosti i zaštite podataka koji su im dostavljeni, da će iste koristiti isključivo u statističke namjene i da neće biti posredovane trećim osobama.
- Jedan od ciljeva statističke povjerljivosti je spriječavanje objavljivanja povjerljivih ćelija u tabelama. U statističkim institucijama u BiH predviđeno je da se ne objavljuju podaci u tabelama kada iste u ćelijama sadrže manje od tri statističke jedinice. Grupe u ćelijama moraju imati najmanje tri jedinice, a udio jedne jedinice u grupi ne smije preći 85%.
- Korisnike statističkih podataka treba detaljno informirati o metodama koji se odnose na povjerljivost.
- Deidentifikovan podatak ne znači nužno da je zaštićen identitet statističke jedinice. Kako bi u potpunosti zaštitili identitet statističke jedinice, potrebno je uzeti u obzir neposredne i posredne identifikacije.
- Veze između tabela i/ili varijabli koje zaštićujemo treba detaljno proučiti.

- Preporučljivo je, da se varijabla odnosno njeni razredi ne definiraju detaljnije nego je to potrebno, jer sa ovim možemo uticati na gubitak informacija.
- Broj varijabli u datoteci mikropodataka ne bi trebao da bude veći nego je potrebno za planirane analize.
- Ako statističke jedinice pristanu na objavu podataka, iz kojih je potencijalno moguće otkriti njihove podatke, istoj dodijelimo poseban status i tako reduciramo gubitak informacije u tabelama, u kojima te statističke jedinice sudjeluju.

7. DISEMINACIJA PODATAKA

7.1 Ažuriranje izlaznih rezultata

Objavljivanje statističkih podataka i informacija se odvija prema standardiziranom postupku u različitim tehnologijama.

Standardizirani postupci se temelje na unaprijed propisanim strukturama, formatima i metapodacima, i isti se uzimaju u obzir kod pripreme tabela (u fazi statističke obrade podataka). Za sve sadržaje vrijede isti, standardizirani postupci rada u skladu sa načelom transparentnosti procesa i uzimajući u obzir pravovremenost. Kod pripreme novih sadržaja (kao što je redovno ažuriranje izlaznih podataka), važna je uređena dokumentacija, interna razmjena znanja, standardni komunikacijski kanali, te arhiviranje gradiva i postupaka rada.

Za svaki oblik objave statističkih podataka i informacija treba unaprijed ažurirati izlazne podatke; to mogu biti podaci u formi baza podataka ili u formi tabela sa finalnim agregiranim podacima, pripremljenim za objavu.

Za statistička istraživanja, kod kojih se podaci revidiraju u cilju osiguranja boljeg kvaliteta, u skladu sa *Uputstvo za reviziju podataka* potrebno je u fazi provođenja revizije ažurirati sve izlazne podatke te revidirane podatke objaviti.

SMJERNICE

- Pri ažuriranju izlaznih rezultata treba uzeti u obzir standardizirane postupke i iste provoditi transparentno (pregledno, jasno). Procese, koji se ponavljaju periodično (u jednakim vremenskim intervalima) potrebno je automatizirati.
- Dokumentacija mora biti uređena, objavljena, dostupna i redovno ažurirana.
- Ukoliko je u ovaj proces uključeno više osoba iz različitih oblasti, potrebno je dokumentirati internu koordinaciju.
- Sadržaji koji su uklonjeni zbog promjene metodologije moraju biti primjereno arhivirani.
- Uzeti u obzir pravila o reviziji statističkih podataka predviđena u Uputstvu.

7.2 Proizvodnja i prezentacija proizvoda za objavljivanje

Elemente sadržaja objave treba pripremiti uzimajući u obzir ciljne korisnike i samu namjenu. Podatke je potrebno prezentirati i objaviti na primjeren način. Prezentacija rezultata treba biti efikasna, razumljiva, jednostavna i zanimljiva.

Kod prezentacije rezultata treba uzeti u obzir načelo višejezičnosti, gdje u pravilu objavljujemo sve publikacije višejezično (bh jezici i engleska verzija).

Opća načela prezentacije rezultata se razlikuju obzirom na vrstu i medije objave. Agencija diseminira statističke informacije putem: Web stranice na adresi: www.bhas.gov.ba; Printanih i elektronskih publikacija; Saopćenja; Pres konferencija; Medija i društvenih mreža i Direktnim odnosom sa korisnicima na temelju standardnih zahtjeva i upita.

Kod prezentacije podataka treba imati na umu statističku pismenost korisnika i različito razumijevanje statističkih podataka i informacija. *Stručna javnost* pretežno koristi detaljno raščlanjene podatke za potrebe provođenja dodatnih analiza podataka, zato takve podatke ima smisla objaviti u bazama podataka tj. u formatima koji omogućavaju daljnje obrade (elektronska forma). *Opću javnost* zanimaju prije svega najvažniji ili najzanimljiviji opći statistički podaci, prikazani na razumljiv (popularan) način.

Autor prezentacije podataka mora razmišljati o tome, šta želi korisnik vidjeti i na kakav način to predstaviti a da je to razumljivo, zanimljivo i korisno. Podaci, koji se objavljuju u elektronskoj prvoj objavi – *Saopćenju* i popularnim publikacijama trebaju trebaju biti prezentirani kao statistička priča. Ova priča treba da sadrži komentar i vizualizaciju podataka. Komentar mora biti učinkovit - kratak, jednostavan, razumljiv i zanimljiv.

Vizualizacije podataka moraju biti izrađene i sažete u preglednoj formi sa jednostavnim tabelama i grafikonima. Prezentiranje treba rezultate statističkog poslovnog procesa učiniti dostupnim javnosti i/ili korisnicima. Pri tome razlikujemo slijedeće načine prezentacije:

- Tabelarni
- Tekstualni
- Grafički i kartografski

Priprema tabela

Tabela je klasični prikaz statističkih rezultata. Publikacija u službenoj statistici (na papiru ili u elektronskom obliku) sastoji se dobrim dijelom iz tabela. Tabele se ne proizvode samo za printane publikacije. Prikaz rezultata u formi tabela na internetu je prijeka potreba službene statistike. Tabela je alat za prenošenje obimnih i složenih informacija u razumljivom i sažetom obliku. Tabela se sastoji od prve kolone, zaglavlja tabele i tijela tabele. Distribucija obilježja u prvoj koloni i zaglavlju tabele ovisi o tome šta će se međusobno upoređivati. Tabela ne smije biti previše komplicirana. Zaglavlje tabele ne smije imati više od četiri nivoa raščlanjivanja. Najviše tri nivoa raščlanjivanja imaju smisla u prvoj koloni. Previše nivoa detalja narušava je razumljivost. Informacije koje su raspoređene na više od tri nivoa raščlanjivanja više nisu korisne za bilježenje. Prva kolona često je prikladnija za smještanje dužih tekstova. Provjerite postaje li tabela lakša za razumijevanje ako zamijenite prednju kolonu i zaglavlje tabele.

Priprema teksta

Kod publikovanja statističkih rezultata tekst ima ulogu podrške. Načelno, u tekstu treba biti objašnjeno sve što je neophodno za razumijevanje prezentiranih tabelarnih rezultata. Ovo također uključuje kratak opis korištenih koncepta i metoda. Iscrpna metodološka objašnjenja i metapodaci će biti navedeni u *Izveštajima o kvalitetu za pojedina statistička istraživanja*. Konferencije za štampu najčešće služe za prvo objavljivanje aktuelnih statističkih podataka.

One nude (u formi *Saopćenja*) informacije za medije u vrlo sažetom obliku i sastavljeni su koncizno (par stranica, metodološke upute i 1-2 tabele). Struktura teksta je u skladu s uobičajenim pravilima za ovu formu objavljivanja.

Grafički prikaz

U cilju prikaza statističkih podataka, koriste se razne forme grafičkog prezentacije.

Za strukturu i ukupne količine prikladne su vrste grafike "Stupčasti dijagram (vertikalni)", "Stupčasti dijagram (vodoravni)" i "Tortni dijagram".

Za rangiranje prikladna vrsta grafika je "stupčasti dijagram (vodoravni)".

Za vremenski tok ili trend prikladne su vrste grafike "Stupčasti grafikon (vertikalni) i "Linijski grafikon".

Za regionalna poređenja idealne su „Karte“.

Za raspršivanje idealna vrsta grafika je "Scatter dijagram, box plot".

SMJERNICE

Priprema tabela

- Sve tabele trebaju sadržavati najmanje slijedeće komponente: naziv tabele, zaglavlje, predkolone, polje za brojčani unos i futer.
- Naziv tabele treba da sadrži predmet (podatke) prezentacije, zatim strukturu podataka iz kolona kao i vremenski period (ili datum) na koji se podaci odnose. Vremensko, predmetno i prostorno razgraničenje prikazanog sadržaja u tabeli treba proizilaziti iz samog naziva tabele.
- Iz zaglavlja treba jasno da proizilazi, na koji dio sadržaja se odnose brojke svake kolone. Dalje, treba navesti mjernu jedinicu u kojoj se prezentiraju podaci.
- Prezentirani brojevi u ćelijama trebaju biti čitljivi. Kod izbora jedinice prikaza treba paziti da ne prikazemo previše brojki. U datom slučaju broj zaokružujemo.

Osim toga važe slijedeće konvencije za prezentaciju u tabelama:

- ako je polje (u ćeliji) prazno, pišemo poprečnu crtu (-),
- ako je vrijednost manja od polovine predviđene najmanje jedinice, pišemo nulu (0 ili 0.0),
- ako relevantne informacije još nisu dostupne, ali se očekuju, pišemo tri tačke (...),
- ako se brojka neće uopće saopćiti ili njeno je objavljivanje protivno odredbama statističke povjerljivosti, ova ćelija se označava sa slovom „T“ i time se prava vrijednost se prikriva.
- U fusnoti treba uvijek navesti izvor podataka npr. „Agencija za statistiku BiH“. Također u fusnoti trebaju naći mjesto i eventualni komentari za pojedine brojčane vrijednosti unutar tabele.

Priprema teksta

- Tekstualni dodatak tabelarne publikacije treba kratko i jasno obuhvatiti sve neophodne aspekte za razumijevanje prezentiranih brojeva. Ni u kojem slučaju, ne trebamo prezentirane rezultate u tabelama ponovno prezentirati i navoditi u tekstualnom dijelu publikacije. Međutim, dobro je tekstualno naglasiti samo najvažnije glavne rezultate, odnosno najupečatljivije promjene. S tim u vezi, pri iznošenju komentara u saopćenjima i ostalim publikacijama važno je posebno obratiti pažnju na sljedeće: izbjegavati korištenje termina - porast, rast, pozitivna i negativna stopa promjene, itd..., a koristiti termine, povećanje – smanjenje; veća – manja; viša - niža itd. (npr. pogrešno je navesti: „broj prevezenih putnika vazdušnim saobraćajem u 2023. godini u poređenju sa 2022. godinom porastao je za 7,7%“. Ne može **porasti** broj putnika – može se samo **povećati**. Ne može: „**Stopa porasta** proizvodnje u odnosu na april 2022. godine iznosi 3,3%“; treba: „**Povećanje** proizvodnje u odnosu na april 2022. godine iznosi 3,3%“. Ne može: „negativna stopa promjene proizvodnje tj. **stopa pada** proizvodnje u odnosu na april 2022. godine iznosi – **5,3%**“; treba: „**smanjenje** proizvodnje u odnosu na april 2022. godine iznosi **5,3%**“, jer sam termin smanjenje sugerira da je riječ o negativnoj stopi promjene. Samo u slučajevima kada se govori o najvećem rastu ili najvećem padu može se koristiti termin rast umjesto povećanje i pad umjesto smanjenje.
- Saopćenje za štampu (konferencija za štampu), kao najsazetija forma prezentiranja rezultata za javnost, podliježe posebnim pravilima:
 - glavni naslov i prvi stavak odlučuje o tome - da li je medijsko saopćenje privuklo pažnju medija,
 - glavni naslov treba u sebi da sadrži najvažnije rezultate,
 - prvi stavak je sažetak onog što će se saopćiti u daljnjem toku konferencije za štampu,
 - kod svih medijski opravdanih skraćivanja teksta mora se osigurati tačnost sadržaja.
- Za sve statističke publikacije treba primijeniti sljedeća načela:
 - u pravilu treba primijeniti "načela struke"(preciznost, jasnost i stručnost),
 - izbor riječi treba prilagoditi „tipičnom čitatelju/slušatelju“(bez stručnih žargona, po mogućnosti malo stranih riječi, nezaobilazne stručne izraze jasno definirati),
 - za često korištene skraćenice potrebno je izraditi odgovarajući popis/listu.
- Pred svaku objavu publikacije, potrebno je svaki tekst proći intenzivnu uredničku provjeru i provjeru sadržaja. Odgovornost za objavu publikacije snosi direktor ili po njegovom dopuštenju, vođa sektora publicistike.

Grafička priprema

- Kod grafičke prezentacije treba izabrati tip grafa (barchart, linijski dijagram, kolač, itd.) koji najbolje odgovara određenom rješenju.
- Svi grafovi ili kartogrami moraju biti adekvatno označeni. Ovo se odnosi na ose, serije podataka, važne podatke i legendu. U samom nazivu grafa treba navesti - šta sadrži graf, odnosno kartogram.

7.2.1 Statistička baza podataka

Publikovanje statističkih rezultata u printanoj formi ili pomoću fiksiranih tabela na internetu ima nedostatak, jer se time ograničava broj tabela i procjena koje stoje na raspolaganju. Iz ovih razloga, korisnicima se pruža mogućnost da iz postojećih podataka sami kreiraju analitičke tabele. U ove svrhe statističke institucije pohranjuju podatke u statističke baze podataka.

Uspostavljanjem statističke baze podataka pruža se mogućnost korisnicima da sami kreiraju određene izlazne tabele i rezultate.

Statističke podatke koji su pohranjeni u bazu podataka, u načelu karakteriziraju dvije stvari. S jedne strane pohranjene su određene statističke jedinice (npr. poslovni subjekti), za koje su zatim agregirana određena obilježja (npr. promet, zaposleni). S druge strane je izvedivi proces agregacije raščlanjen prema određenim obilježjima (kriterijama raščlanjivanja). Ova obilježja omogućavaju, da, posmatrane veličine razgraničimo vremenski, prostorno i sadržajno.

Mogućnost proizvodnje višedimenzionalnih tabela je povezana sa određenim problemima.

Tako je moguće da kod proizvodnje ćelija sa manjim brojem jedinica nastupi problem vezan za povjerljivost. Slijedeći problem leži u činjenici da su podaci koji potiču iz anketnih istraživanja jednim dijelom nesigurni. Naime, većom mogućnosti raščlanjivanja baze podataka, mogu nastati ćelije, čije vrijednosti imaju veću vjerovatnost greške i time njihova vrijednost može biti problematična.

SMJERNICE

- Načelno, treba težiti ka tome da se svi publicirani statistički materijali pohrane u baze podataka.
- Prilikom pohrane materijala u baze podataka, treba odlučiti koja obilježja i u kojoj detaljnosti treba staviti na raspolaganje. U svakom slučaju, potrebno je da postoje obilježja koja dozvoljavaju prostorno i vremensko razgraničenje materije.
- Prije pohranjivanja materijala u baze podataka treba obratiti pažnju da ne dođe do problema vezano za povjerljivost. Znači, treba definirati šta raditi sa ćelijama, čije vrijednosti iz razloga povrede pravila povjerljivosti ne mogu biti prikazane u originalnom obliku. Sve smjernice koje se odnose na statističku povjerljivost vrijede i za publiciranje baza podataka.
- Novi materijal koji se pohranjuje u baze podataka, mora biti provjeren prije nego se da na korištenje. Posebno se mora provjeriti, da li su vrijednosti identične prije i nakon tehničkog postupka konverzije koji prethodi pohranjivanju.
- Pohranjeni materijal mora se redovno održavati. Posebno je potrebno (što je više moguće) aktuelizirati segmente koji sadrže materijal koji se periodično obnavlja. Vremenski razmak između publikovanja rezultata i pohranjivanja u baze podataka treba biti što kraći.

- Kada se materijal po prvi put pohranjuje u baze podataka, najvažnije korisnike treba obavijestiti da su podaci na raspolaganju u bazama podataka.
- Korištenje baza podataka mora biti stalno nadzirano. Treba snimati broj pristupa bazama podataka. Potrebno je obratiti pažnju da tehničke okolnosti dozvole adekvatno korištenje baza podataka. Ovo se odnosi na dostupnost, zadovoljenje korisnika i brzinu obrade.

7.3 Upravljanje objavom diseminacijskih proizvoda

Misija Statistike BiH je pružanje pouzdanih, kvalitetnih, razumljivih, pravovremenih i međunarodno uporedivih statističkih podataka koje zadovoljavaju potrebe donosilaca odluka, istraživača i ostalih domaćih i stranih korisnika i odražavaju stanje i promjene u ekonomskom, demografskom i socijalnom području, području okoliša i prirodnih resursa.

Prikupljanje, obrada, analiza i diseminacija statističkih podataka vrši se na osnovu statističkih standarda i savremene tehnologije, uz zaštitu statističke povjerljivosti, optimalno korišćenje resursa i razumno opterećenje davalaca podataka.

Svi objavljeni podaci i informacije su korisniku dostupni na internet stranicama Agencije i entitetskih statističkih zavoda.

Svaka objava statističkih podataka mora biti najavljena u skladu sa Kalendarom objavljivanja. Kalendar objavljivanja se priprema krajem godine za jednu godinu unaprijed, svaka objava mora biti potvrđena najkasnije u petak prije sedmice objave.

Diseminacija statističkih proizvoda i usluga Agencije provodi se u skladu sa sljedećim načelima:

Dostupnosti svih korisnika rezultatima statističkih istraživanja uključenih u Plan rada; Jednakosti svih korisnika kod upoznavanja sa rezultatima statističkih istraživanja uključenih u Plan rada; Objektivnosti i nepristrasnosti u prezentiranju rezultata statističkih istraživanja uključenih u Plan rada; Strogog pridržavanja rokova za publikovanje rezultata statističkih istraživanja, prema ranije najavljenom Kalendaru publikovanja; Zaštite statističke povjerljivosti individualnih podataka u skladu sa Zakonom o statistici BiH i Pravilnikom o zaštiti statističkih podataka u Agenciji; Razumljivosti i jasnoće u objavi rezultata statističkih istraživanja uključenih u Plan rada i Orijentacije prema potrebama korisnika, uključujući publikovanje glavnih informacija i na engleskom jeziku.

Status podataka (privremeni ili konačni) mora biti utvrđen još kod planiranja objave.

Međunarodnim organizacijama ili u Eurostat mogu biti proslijeđeni samo podaci, koji su već objavljeni. Ako u Eurostat šaljemo povjerljive podatke, povjerljive ćelije moraju biti označene.

Postupci objavljivanja podataka su za različite vrste objava različiti. U skladu s tim razlikujemo:

- objavljivanje vijesti na internet stranici,
- objavljivanje printanih publikacija,
- objavljivanje podataka u bazi,
- proslijeđivanje podataka Eurostatu,
- ažuriranje podataka u interaktivnim alatima,
- ispravka grešaka u objavama

Aktivnost, povezana sa objavljivanjem statističkih podataka, je takođe korigiranje grešaka. Svrha korigiranja grešaka u objavljenim statističkim podacima je obezbjeđivanje tačnih i kvalitetnih statističkih podataka i informacija za korisnike. Sistem korigiranja grešaka mora biti uređen tako, da je pregleden i razumljiv za korisnike. Tačni postupci su propisani u *Uputstvu za korekciju grešaka u publiciranim izdanjima*.

SMJERNICE

- Podaci moraju biti objavljeni tačno u skladu sa najavama u Kalendaru objavljivanja, koji je dostupan korisniku na internet stranicama.
- Cjelokupni postupak statističkog istraživanja mora biti tako isplaniran da se podaci pravovremeno objave.
- Statistički podaci i informacije, iako mogu biti iznimno tačni i detaljni, nisu od velike koristi ukoliko nisu pravovremeno objavljeni.
- Podaci moraju biti dostupni svim korisnicima na isti način. Isto tako, korisniku mora biti na raspolaganju što više mogućnosti za dodatnu potražnju i naručivanje podataka.
- Objavljanje podataka mora biti što transparentnije. Skupa sa podacima moraju biti na raspolaganju i odgovarajući metapodaci, kao što su metodološka objašnjenja, osnovni pokazatelji kvaliteta, upitnici itd. Takođe mora biti omogućena preglednost i jasnost postupaka korigiranja grešaka i postupaka revidiranja podataka.

7.3.1 Dostavljanje podataka Eurostatu

Značajan dio statistika Agencija proizvodi na temelju preporučenih europskih standarda i dobrih praksi. Ovo se prije svega odnosi na statistike, gdje je dat jači naglasak na harmonizaciju koncepata na evropskom nivou.

Eurostat je mjesto prema kojem se dostavljaju podaci prema ugovorenim procedurama i načinu dostave. Obim, termini i sadržaj isporučenih podataka utvrđuje se odgovarajućim EU-propisima. Načelno, prenos podataka mogu uključivati slijedeće:

- a) Prenos programa tabela;
- b) Prenos pojedinačnih podataka i
- c) Prenos metapodataka u formi *Izvještaja o kvalitetu* (detaljnog opisa metodologije).

Prenos tabela Eurostatu je uglavnom dio ukupnog programa publikovanja. Ovdje je prije svega riječ o dostavi izabranih rezultata. Izvještaji o kvalitetu služe Eurostatu kao dokaz da se pridržavamo dogovorenih preporuka i standarda kvaliteta.

SMJERNICE

- Agencija će poduzeti sve da ispune zahtjeve Eurostata. Ovo se odnosi kako na sadržaj i obim programa dostave podataka, tako i na predviđene rokove i način dostave.
- Dostava podataka, odnosno metapodataka Eurostatu treba biti dokumentovana. Posebno trebaju biti opisani svi relevantni postupci u Izvještaju o kvalitetu.
- Pojedinačni podaci se dostavljaju Eurostatu samo u anonimiziranom obliku (statistički zaštićen mikropodatak). Dostava slijedi u unaprijed dogovorenom formatu.
- Zaposlenici Agencije će zajednički raditi sa Eurostatom, da sadržaj i format dostavljenih metapodataka tokom vremena ostane stabilan, u cilju visokog stepena vremenske uporedivosti i kvaliteta informacija.

7.3.2 Dostavljanje podataka na međunarodne upite

Međusobna komunikacija između nacionalnih statističkih ureda i ostalih nacionalnih i međunarodnih organizacija putem koje se razmjenjuje statistika od najvećeg je značaja.

Ova saradnja se odvija ne samo u radnim grupama, sastancima i konferencijama, nego se odvija i putem zahtjeva koji se direktno šalju Agenciji. Agencija će se truditi da na ove međunarodne upite odgovori na adekvatan način.

SMJERNICE

- Na upite međunarodnih organizacija treba biti odgovoreno na vrijeme i (po mogućnosti) što potpunije.
- Kada se primi zahtjev od međunarodne organizacije, o tome odmah treba informirati Odjeljenje za međunarodnu saradnju.
- Svi kontakti sa mjesta gdje se obrađuju zahtjevi, trebaju biti proslijeđeni na znanje i Odjeljenju za međunarodnu saradnju. Arhiviranje slijedi centralizirano i potrebno je da se o tome brine služba kojoj to organizacijski pripada (Sektor diseminacije).
- Zahtjev treba proslijediti stručno kompetentnim zaposlenicima.

7.4 Promoviranje diseminacijskih proizvoda

Dok se marketing općenito može smatrati sveobuhvatnim procesom, ovaj podproces se odnosi na aktivnu promociju statističkih proizvoda proizvedenih u određenom statističkom poslovnom procesu, kako bi im se pomoglo da dosegnu najširi krug korisnika. To uključuje upotrebu alata za upravljanje odnosima s korisnicima, kao i korištenje alata uključujući web stranice, i blogove kako bi se olakšao proces komuniciranja statističkih podataka korisnicima. Najbolji način za dobijanje povratnih informacija od korisnika je ispitivanje njihova zadovoljstva. Na temelju povratnih informacija je moguće izvršiti analizu i evaluaciju korisničkih potreba.

7.5 Podrška korisnicima

Korisniku treba omogućiti pristup do podataka koji su rezultat statističkih istraživanja, pri čemu mora biti osigurana zaštita statističke povjerljivosti. Podrška korisniku se osigurava prije svega putem elektronske pošte i telefona, kao i osobnim kontaktom.

Podrška korisniku obuhvata posredovanje statističkih podataka i informacija, savjetovanje o pristupu do podataka i savjetovanje o korištenju alata za pripremu podataka, te pomoć pri traženju i pripremi statističkih podataka i informacija.

Statistički podaci i informacije koji su objavljeni i podaci čija priprema ne zahtjeva dodatne obrade, besplatni su. Priprema podataka po želji korisnika se obračunava po važećem cjenovniku. Davanje podataka, koje spada u okvir izvještavanja međunarodnim organizacijama, je besplatno.

Obavezno se vodi evidencija pismenih zahtjeva koja je neophodna za pripremu analize korisnika, načina komuniciranja i koja su najviše tražena statistička područja.

Korisniku je omogućen takođe pristup do statistički zaštićenih mikropodataka, ali pod posebnim uvjetima i posebnim ugovorom. Statistički zaštićene mikropodatke imaju pravo dobiti registrirane istraživačke institucije i registrirani istraživači.

Načini podrške korisnicima mogu biti različiti ovisno od toga, o kojoj vrsti zahtjeva se radi. Razlikujemo slijedeće forme zahtjeva:

- pisani upit za podacima,
- upiti putem telefona,
- osobna posjeta korisnika,
- pristup do statistički zaštićenih mikropodataka,
- pristup korisnika do bibliotečkih materijala.

SMJERNICE

- Podaci koje posredujemo korisniku moraju biti tačni, relevantni i primjereno predstavljeni.
- Korisnici moraju pravovremeno dobiti podatke. O mogućim kašnjenjima treba korisnika odmah obavijestiti.
- Svi podaci su besplatni. Za zahtjevnije upite zaračunavaju se samo eventualno nastali dodatni troškovi.
- Informacije o tome, kako pristupiti do statističkih podataka, objavljuju se na nekoliko mjesta (internet stranici, publikacijama).
- Telefonski broj za informacije putem telefona mora biti dostupan u radnom vremenu statističke institucije.
- Osnovna objašnjenja u vezi pristupa do mikropodataka za naučnoistraživačke namjene, moraju biti objavljena na internet stranicama statističke institucije.

8. EVALUACIJA

Statistička istraživanja se u pravilu provode periodično, tako da se cjelokupni statistički proces ponavlja. Važno je da taj proces uključuje povratnu vezu, koja omogućuje uvođenje promjena i poboljšanja. Za ove namjene potrebno je svako statističko istraživanje (nakon što se završi) cjelovito ocijeniti a zatim u ovoj ocjeni kritički procijeniti uspješnost cjelokupnog provedenog istraživanja i iznaći mogućnosti za poboljšanje.

Prikupljanje informacija o kvalitetu statističkih podataka odvija se tokom cjelokupnog statističkog procesa. Sistematsko dokumentiranje pojedinih dijelova istraživanja predstavlja važan dio informacija o toku istraživanja i pomaže u otkrivanju potencijalnih sistematskih pogrešaka u ovom procesu. Pomoću ovih informacija možemo ocijeniti kvalitet statističkih podataka i kritički vrednovati dobijene rezultate koji su važni za korisnike, jer tako dobijaju dodatni uvid u proces prikupljanja podataka. Objavljivanje informacija o kvalitetu podataka predstavlja transparentan način informiranja korisnika o različitim aspektima statističkih podataka.

8.1 Prikupljanje dokumentacije o istraživanju

Postupak pripreme dokumentacije o istraživanju obuhvata detaljan opis statističke aktivnosti, uključujući opis koncepata, definicija, korištenih metoda, korištenog informacijskog sistema te radnih uputstava. Kvalitet dokumentacije o istraživanju je važan pokazatelj kvaliteta istraživanja. To je također važan alat za komunikaciju između različitih istraživanja, kao i između proizvođača i korisnika statistike. Dokumentacija o istraživanju je dio metapodataka.

Dokumentaciju generalno možemo podijeliti na dokumentaciju za korisnike rezultata istraživanja i na dokumentaciju za proizvođače statističkih istraživanja.

Dokumentacija za korisnike rezultata istraživanja opisuje i dokumentira statističke rezultate i javno se objavljuje. Namjena dokumentacije za korisnike je da se upoznaju koje podatke i korištene statističke metode mjerimo, da razumiju same podatke, da iste lakše zatraže i da ih po potrebi mogu sami dalje obrađivati. Primjeri dokumentacije za korisnike su: upitnici za istraživanja, metodološka uputstva, korisnički usmjereni izvještaji o kvalitetu statističkih istraživanja.

Dokumentacija za proizvođače statističkih istraživanja opisuje statističke postupke i procese, koje smo koristili u cjelokupnom statističkom procesu. Ova dokumentacija nastaje prilikom provođenja pojedinih koraka u procesu statističkog istraživanja i uglavnom je namijenjena za internu upotrebu. Namjena dokumentacije za proizvođače statističkih istraživanja je osiguranje efikasne i pouzdane proizvodnje podataka. U dokumentaciji za proizvođače statističkih istraživanja moraju biti prije svega detaljno opisani pojedini koraci koje smo koristili u provođenju istraživanja (npr. određivanje ciljne populacije, izbor uzorka, dizajniranje upitnika, uređivanje podataka, objava podataka, itd).

SMJERNICE

- Dokumentacija o istraživanju mora biti tačna, detaljna i razumljiva ciljnoj grupi kojoj je namijenjena.
- Kod pripreme dokumentacije treba koristiti standardne obrasce i predloške (ako postoje). Takođe, sadržaj pojedinih procesa treba u najvećoj mogućoj mjeri slijediti standardnu strukturu.
- Dokumentacija za korisnike i proizvođače treba biti dostupna svim zainteresiranim stranama.
- Dokumentacija treba biti napisana tako da u što većoj mjeri osigura transparentnost statističkih postupaka i proizvoda.
- U dokumentaciji za korisnike treba biti dokumentirano sve što bi moglo pri korištenju eventualno izazvati nejasnoće i nerazumijevanje.
- Sadržaj i nivo detaljnosti dokumentacije se prilagođava ciljnim korisnicima.
- Dokumentacija o procesima treba sadržavati takođe opis predviđenih postupaka - ako dođe do grešaka, zatim imena osoba ili odsjeka s kojima treba uspostaviti kontakt u slučaju problema ili nejasnoća.

8.1.1 Pohranjivanje agregiranih statističkih podataka i statističkih mikropodataka

Statističke institucije pohranjuju agregirane statističke podatke za daljnju upotrebu, kako one u elektronskom tako i one u printanom obliku.

Arhiviranje agregiranih podataka u elektronskom obliku odvija se u sklopu provođenja politike sigurnosnog kopiranja i pohranjivanja elektronskih podataka. Agregirani statistički podaci u printanom obliku mogu biti objavljeni u publikacijama koje izdaju Agencija i entitetski statistički zavodi, ili u publikacijama koje preuzimaju od drugih institucija. Sve publikacije, koje publikuju Agencija i entitetski statistički zavodi se pohranjuju u barem u jednom primjerku.

Printani izvodi publikacija se šalju u arhive. Podaci o publikacijama se unose u katalog publikacija. Strane serijske statističke publikacije, arhiviraju se svaka u jednom primjerku. Odgovorni zaposlenik nadležne statističke institucije priprema godišnji spisak bibliotečkog materijala za otpis. Materijal pregleda komisija i određuje listu materijala za otpis.

Nadležne statističke institucije pohranjuju statističke mikropodatke za daljnju upotrebu u elektronskom obliku za upotrebu za istraživačko-analitičke namjene. Pohranjivanje statističkih mikropodataka u elektronskoj formi odvija se u sklopu provođenja politike zaštitnog kopiranja i pohranjivanja elektronskih podataka.

Sve baze podataka, koje sadrže podatke iz različitih statističkih istraživanja i metapodatke, koji su za nadležne statističke institucije od trajne vrijednosti, podliježu jedinstvenom sistemu zaštitnog kopiranja na osnovu dobrih stručnih praksi.

Dnevno se zaštitno kopiraju samo podaci koji su tog dana bili izmjenjeni. Sedmično se zaštitno kopira cijeli server. Oni se čuvaju na sigurnom mjestu za pohranu medija. Najmanje jednom mjesečno se izvede test postupka ažuriranja baza podataka. Statističke mikropodatke u elektronskoj formi pohranimo na zajedničkom serveru datoteka, na koji se pohranjuju datoteke sa podacima i metapodacima.

SMJERNICE

- Potrebno je redovno pratiti zahtjeve struke na području sigurnosnog kopiranja i arhiviranja.
- Potrebno je redovno slati izvode u arhiv i upisivati publikacije u KATALOG PUBLIKACIJA.
- Zaštitno kopiranje i pohranjivanje treba provoditi u skladu sa uputstvima za izradu i pohranjivanje zaštitnih kopija.

8.2 Provođenje evaluacije - evaluacija rezultata

Da bi primjenu korištenih metoda pri proizvodnji podataka učinili razumljivim za korisnike statističkih proizvoda, potrebno je staviti na raspolaganje dovoljno metapodataka.

U statističkim bazama podataka metapodacima se smatraju oni podaci koji nisu direktno prikazani u sadržaju statistika, ali su za razumijevanje procesa proizvodnje statističkog proizvoda od iznimne važnosti (npr. naziv djelatnosti ili zanimanja, popis općina, itd).

Statističkim metapodatkom smatramo također i opise polja podataka u upitnicima, definicije varijabli, objašnjenja i uputstva za popunjavanje obrazaca. Stvarni statistički podaci nazivaju se (za razliku od metapodataka) mikropodacima i makropodacima. Bez dovoljno metapodataka moguće je da korisnik nema sve potrebne informacije, tako da može doći do pogrešnih interpretacija i pogrešnih odluka.

Iz ovih razloga od posebnog značaja je izrada Izvještaja o kvalitetu za pojedina statistička istraživanja. Oni ne služe samo za interne potrebe, nego sadrže metapodatke, koji su neophodni za potpuno korištenje statističkog proizvoda.

Dokumentiranje samog statističkog procesa i njegovih rezultata, stavljanje na raspolaganje metainformacija i izvještavanje o kvalitetu putem odgovarajućih pokazatelja su direktno povezani. U tom smislu, u Agenciji i entitetskim statističkim zavodima postoji standardizirani sistem izvještavanja o kvalitetu statističkog proizvoda.

Izvještaj o kvalitetu za statistička istraživanja je strukturiran po utvrđenoj fiksnoj šemi i **Struktura Izvještaja o kvalitetu za statistička istraživanja** i ima sljedeći izgled.

- 1 STATISTIČKI PROCES I STATISTIČKI PROIZVOD
- 2 RELEVANTNOST
 - 2.1 Korisnici podataka statističkog istraživanja
 - 2.1.1 Ključni korisnici podataka iz statističkog istraživanja
 - 2.1.2 Procjena korisničkih potreba
 - 2.1.3 Mjerenje percepcije i zadovoljstva korisnika
 - 2.2 Potpunost podataka
 - 2.2.1 Indikator kvaliteta i učinka - Stopa kompletnosti podataka (R1)
- 3 TAČNOST I PRECIZNOST
 - 3.1 Greška uzorkovanja
 - 3.1.1 Indikator kvaliteta i učinka – Greška uzorkovanja (A1)
 - 3.1.2 Aktivnosti za smanjenje grešaka uzorkovanja
 - 3.2 Neuzoračke greške
 - 3.2.1 Neuzoračke greške - Greške obuhvata
 - 3.2.1.1 Indikator kvaliteta i učinka - Stopa prekomjernog obuhvata (A2)
 - 3.2.1.2 Indikator kvaliteta i učinka – Udio zajedničkih jedinica (A3)
 - 3.2.1.3 Greška nedovoljnog obuhvata
 - 3.2.1.4 Mjere za smanjenje grešaka obuhvata
 - 3.2.2 Neuzoračke greške - Greške mjerenja
 - 3.2.2.1 Razlozi za nastanak grešaka mjerenja
 - 3.2.2.2 Mjere za smanjenje broja grešaka mjerenja
 - 3.2.3 Neuzoračke greške - Greške neodgovora
 - 3.2.3.1 Indikator kvaliteta i učinka-Stopa neodgovora jedinice(A4)
 - 3.2.3.2 Indikator kvaliteta i učinka - Stopa neodgovora varijable (A5)
 - 3.2.3.3 Postupci u slučaju neodgovora
 - 3.2.3.4 Postupci za smanjenje stope neodgovora
 - 3.3 Revizije
 - 3.3.1 Indikator kvaliteta i učinka - Prosječna veličina revizije podataka (A6)
 - 3.4 Imputacija
 - 3.4.1 Indikator kvaliteta i učinka - Stopa imputiranih podataka (A7)
- 4 PRAVOVREMENOST I TAČNOST OBJAVLJIVANJA
 - 4.1 Pravovremenost objavljivanja
 - 4.1.1 Indikator kvaliteta i učinka - Pravovremenost prvih rezultata (TP1)
 - 4.1.2 Indikator kvaliteta i učinka - Pravovremenost konačnih rezultata (TP2)
 - 4.2 Tačnost objavljivanja
 - 4.2.1 Indikator kvaliteta i učinka – Tačnost objavljivanja (TP3)
 - 4.3 Razlozi za veća kašnjenja i mjere za poboljšanje pravovremenosti i tačnosti
- 5 USKLAĐENOST I UPOREDIVOST
 - 5.1 Usklađenost
 - 5.1.1 Indikator kvaliteta i učinka - Skladnost sa rezultatima iz referentnog izvora (CH1)
 - 5.1.2 Razlozi za veća odstupanja

- 5.2 Uporedivost
 - 5.2.1 Indikator kvaliteta i učinka – Nepodudarnost uporedivih statistika (CC1)
 - 5.2.2. Indikator kvaliteta i učinka - Dužina uporedivih vremenskih serija (CC2)
 - 5.2.3 Prekidi u vremenskim serijama
- 5.3 Geografska uporedivost
 - 5.3.1 Uporedivost s ostalim članicama evropskog statističkog sistema
- 6 DOSTUPNOST I RAZUMLJIVOST, DISEMINACIJSKI FORMAT
 - 6.1 Saopćenja u kojima se objavljuju podaci
 - 6.2 Publikacije u kojima se objavljuju podaci
 - 6.3 On – line baza podataka
 - 6.4 Pristup mikropodacima
 - 6.5 Dostupnost metodološke dokumentacije
 - 6.6 Mjere za poboljšanje razumljivosti diseminiranih rezultata
 - 6.7 Indikator kvaliteta i učinka - Konsultovanje setova podataka (AC1)
 - 6.8 Indikator kvaliteta i učinka – Konsultovanje meta podataka (AC2)
 - 6.9 Indikator kvaliteta i učinka – Stopa kompletnosti meta podataka (AC3)
- 7 TROŠKOVI ISTRAŽIVANJA I OPTEREĆENOST DAVALACA PODATAKA
 - 7.1 Troškovi provođenja statističkog istraživanja
 - 7.2 Opterećenost davalaca podataka
 - 7.3 Mjere za smanjivanje troškova i opterećenosti
- 8 POVJERLJIVOST
 - 8.1 Povjerljivost – politika
 - 8.2 Povjerljivost – postupanje sa podacima
- 9 STATISTIČKA OBRADA
 - 9.1 Izvor podataka
 - 9.2 Učestalost prikupljanja podataka
 - 9.3 Prikupljanje podataka
 - 9.4 Validacija podataka
 - 9.5 Kompilacija podataka
 - 9.6 Prilagođavanja
 - 9.6.1 Sezonsko prilagođavanje

8.3 Akcioni plan za poboljšanje

Ova podfaza okuplja sve potrebne snage odlučivanja kako bi se formirao i dogovorio akcijski plan na temelju izvještaja o evaluaciji. Uključuje korake ili akcije za praćenje provođenja preporuka u budućnosti kroz predloženi mehanizam na institucionalnom nivou.

DIREKTORICA

Vesna Čužić

POPIS LITERATURE

- Eurostat 2005, *Handbook on data quality assessment methods and tools*
http://www.cmi.edu/research/rmi_government/EC/DatQAM_handbook_final.pdf
- Eurostat 2009, *ESS Handbook for Quality Reports*
http://epp.eurostat.ec.europa.eu/portal/page/portal/ver/quality/documents/EHQR_FINAL.pdf
- Kako podacima dati smisao (2009), prvi dio - Priručnik za pisanje o brojevima – UNECE, Ekonomska Komisija Ujedinjenih naroda za Europu
https://unece.org/DAM/stats/documents/writing/MDM3_Croatian_version.pdf
- Kako podacima dati smisao (2009), drugi dio - Priručnik za prikazivanje statistike - UNECE, Ekonomska Komisija Ujedinjenih naroda za Europu
https://unece.org/DAM/stats/documents/writing/MDM_Part2_Croatian.pdf
- ABS 2010, Information Paper:
Quality Management of Statistical Processes Using Quality Gates, Dec 2010, cat. no. 1540.0, ABS, Canberra.
- Statistik Austria *Qualitätsrichtlinien Version 1.1 - Stand 14.12. 2010*
http://www.stat.at/web_de/ueber_uns/aufgaben_und_grundsaeetze/qualitaet/index.html
- Statistični ured Republike Slovenije – Smernice za zagotavljanje kakovosti (2012)
<http://www.stat.si/doc/pub/Smernice.pdf>
- UN (2014) *Fundamentale Prinzipien der amtlichen Statistik der Vereinten Nationen*
https://unstats.un.org/unsd/dnss/hb/E-fundamental%20principles_A4-WEB.pdf
- Eurostat (2017) - *European Statistics Code of Practice*
<https://op.europa.eu/en/publication-detail/-/publication/661dd8ef-7439-11e8-9483-01aa75ed71a1/language-hr/format-PDF>
- Statistics Canada: Statistics Canada's Quality Assurance Framework (2018)
<https://www150.statcan.gc.ca/n1/pub/12-586-x/12-586-x2017001-eng.htm>
- *A guide to good survey design Fifth edition* by Statistics New Zealand (2019):
<https://www.stats.govt.nz/assets/Uploads/Methods/A-guide-to-good-survey-design-fifth-edition/a-guide-to-good-survey-design-fifth-edition.pdf>
- European Statistical System (ESS) - Manuals and guidelines (2020)
European Statistical System (ESS) handbook for quality and metadata reports — 2020 edition