

# PËRFSHIRJA E INOVACIONIT NË PROCESIN E MBLEDHJES SË TË DHËNAVE NË PRODHIMIN E STATISTIKAVE ZYRTARE

ALMA KONDI, INSTITUTI I STATISTIKAVE  
akondi@instat.gov.al

HELDA CURMA, INSTITUTI I STATISTIKAVE  
hmitre@instat.gov.al

## Abstrakti

Vizioni i Institutit të Statistikave (INSTAT) është të ofrojë të dhëna të besueshme dhe të krahasueshme, duke përshtatur metodologjitë dhe duke shtuar listën e treguesve statistikorë. Bazuar në këtë vizion, përdorimi i teknologjisë në procese të ndryshme statistikore është rritur gjatë viteve të fundit, duke filluar nga mbledhja, editimi, vlerësimi, përpunimi dhe shpërndarja e të dhënave. INSTAT ka avancuar në përdorimin e teknologjisë (metoda CAPI, mbledhja e të dhënave me internet, OCR, R, PxWeb).

Ky artikull shqyrton situatën në vitin 2016 lidhur me përdorimin e teknologjive të reja, përmirësimet e bëra duke i krahasuar me metodat tradicionale të grumbullimit, përpunimit dhe shpërndarjes së të dhënave,

si dhe analizon sfidat e ardhshme të INSTAT. Metoda e përdorur në këtë studim është një analizë e detajuar e kostos, shkallës së përgjigjes, mënyrës së kapjes si dhe cilësisë së të dhënave, nga metodat e ndryshme të grumbullimit të të dhënave.

Përdorimi i inovacionit mund të bëhet nga zyrat statistikore për të zvogëluar ngarkesën e të intervistuarve dhe koston, për të maksimizuar cilësinë e të dhënave të grumbulluara dhe për të përmirësuar afatet kohore.

## FJALË KYÇE:

INSTAT; Mbledhja e të dhënave;  
Teknologjia; Cilësia e të dhënave

## 1. HYRJE

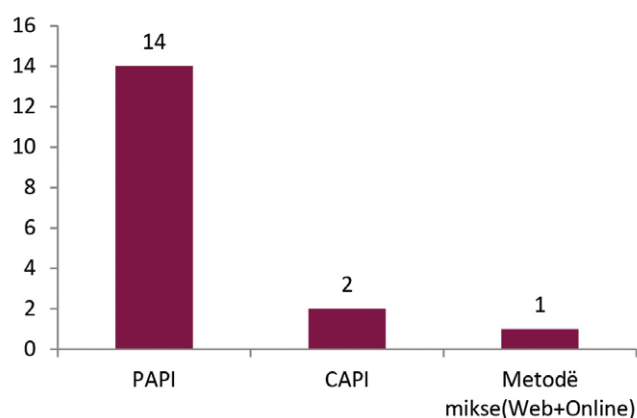
Një nga sfidat më të rëndësishme të zyrave të statistikave është rritja e cilësisë dhe përmirësimit të afatit kohor me efikasitet. Procesi i rritjes së të dhënave me cilësi të lartë për Institutet e Statistikave Kombëtare (ISK), duhet të përfshijë praktikën më të mira në grumbullimin, zvogëlimin e kostos dhe ngarkesën e të intervistuarve. Shpenzimet janë shumë të rëndësishme, veçanërisht për zyrat statistikore me buxhet të ulët. Ndryshimet e reja teknologjike mund të ndikojnë në çdo proces statistikor. Ndryshimet e shpejta në teknologji kanë krijuar mundësi të reja për përmirësimin e afateve kohore. Kostot e teknologjive të reja, të tilla si tableta ose pajisje të tjera celulare, janë në rënie, duke bërë të mundur përdorimin e tyre në procesin e grumbullimit të të dhënave përmes metodës së intervistimit personal me ndihmën e kompjuterit (CAPI).

Statistikat e Kanadasë analizojnë në Raportin e Metodologjisë (Statistics Canada, 2010) avantazhet e përdorimit të teknologjisë në prodhimin statistikor. Teknologjia mund të përdoret për të monitoruar dhe kontrolluar cilësinë e të dhënave nga mbledhja e të dhënave, kapja e të dhënave dhe në procesin e editimit dhe imputimit. Gjithashtu, kalimet automatike në pyetësor, mund të përdoren për të zgjidhur problemet e menjëhershme, duke zvogëluar ngarkesën e të intervistuarve. Është më e lehtë të zbatohet kontrolli i cilësisë së procesit të intervistimit dhe të krijohen raporte për menaxherët për statusin e intervistimit, për shembull shkalla e përgjigjes, numri i intervistave të përfunduara dhe kohëzgjatja për intervistë etj. Nga ana tjetër, përdorimet e teknologjisë kërkojnë investime për blerje ose rinovime. Gjithashtu, kërkohet që ekspertët e kompjuterëve të zhvillojnë programe për prodhimin statistikor.

Puna e INSTAT bazohet në programin pesëvjeçar të Statistikave Zyrtare, i cili është dokumenti bazë që siguron prodhimin e të dhënave statistikore nga Sistemi Kombëtar i Statistikave, të nevojshme për vëzhgimin e situatës ekonomike, sociale dhe mjedisore në Republikën e Shqipërisë. Ky program punon në përputhje me parimet statistikore të parashikuara në ligjin statistikor, por gjithashtu ndjek parimet e përgjithshme të Kodit të Praktikës së Statistikave Evropiane.

Gjatë vitit 2016, INSTAT ka kryer 17 vrojtime. Në Figurën 1 janë paraqitur numri i vrojtimeve, sipas metodës së mbledhjes së të dhënave për vitin 2016.

**Figura 1: Numri i vrojtimeve sipas metodës së mbledhjes së të dhënave**



*Burimi: Programi INSTAT për vitin 2016, llogaritja e autorëve*

Tradicionalisht, INSTAT ka përdorur "Intervistën Personale me Laps e Letër" (PAPI), si një metodë për grumbullimin e të dhënave. Kjo është metoda më e përdorur edhe sot, siç tregohet në Figurën 1; rreth 82 për qind e vrojtimeve përdorin metodën PAPI për të mbledhur të dhëna, për shkak të disa faktorëve si, norma përgjithësisht e lartë e përgjigjes, sistemi i dobët i adresave, është i thjeshtë për t'u përdorur nga anketuesit dhe shpesh për shkak se është më i lirë.

Për herë të parë CAPI (Intervistën Personale me Kompjuter) është përdorur në Studimin Demografik dhe Shëndetësor në Shqipëri (ADHS) 2008-09. Një aspekt inovativ i ADHS 2008-09 ishte përdorimi i Personal Digital Assistants (PDAs) për mbledhjen e të dhënave, në vend të pyetësorëve me letër. Pyetësorët u programuan në PDA duke përdorur software Census and Survey Processing (CSPRO). Vrojtimi i plotë dhe menaxhimi i të dhënave, klasifikimet, kapërcimet (skips) dhe kontrolli i plotshmërisë ishin vendosur në sistemin e kapjes së të dhënave. Pyetësorët e letrës ishin në dispozicion të intervistuesve në rast të mosfunksionimit të pajisjeve. Në vitin 2016 metoda CAPI është përdorur nga 2 vrojtime.

Mbledhja e të dhënave me internet (WEB) është përdorur për herë të parë në vitin 2013, si një mënyrë për të ulur shpenzimet, duke shmangur përdorimin e intervistuesve dhe duke zvogëluar ngarkesën e të intervistuarve. Për këtë qëllim, është përdorur një software open source, Limesurvey Server. Forma të ndryshme janë përgatitur për mbledhjen e informacionit, për shembull për Anketën e Statistikave Afatshkurtra, studimin e Zhvillimit të Kërkimeve për Sektorin Publik etj. Kjo metodë është përdorur në 2016

në një vrojtim, por është përdorur gjithashtu me metodën CAPI, për të dhënë më shumë informacion.

## 2. KRAHASIMI I METODAVE TË GRUMBULLIMIT TË TË DHËNAVE

Ky seksion do të përshkruajë metodat e ndryshme të mbledhjes së të dhënave, avantazhet dhe disavantazhet e tyre në vrojtime e ndryshme të kryera nga INSTAT, në aspekte të tilla si kostoja, shkalla e përgjigjes, ngarkesa e të intervistuarve, treguesit e cilësisë etj. Këto aspekte analizojnë dhe krahasojnë vrojtmet e mëposhtme:

- Anketa e Forcave të Punës (AFP)
- Anketa e Statistikave Afatshkurtra (ATN)
- Anketa e Komunikimit dhe Teknologjisë së Informacionit (TIK)

AFP u krye për herë të parë në vitin 2007. Gjatë vitit 2007-2011, AFP është kryer nga INSTAT në bazë vjetore përmes intervistave të drejtpërdrejta personale, duke kontaktuar familjet në banesën e tyre. Në vitin 2011, për herë të parë, AFP filloi përdorimin e metodës CAPI. Metoda e optimizuar (në rastin tonë metoda CAPI), na lejon të kemi të dhëna më të sakta sepse kontrollet logjike janë bërë drejtpërdrejt në programin e laptopëve. Këto kontrolle zvogëlojnë gabimet që nuk vijnë nga procesi i zgjedhjes (për shembull kalimet nëpër pyetje kapërcehen automatikisht dhe jo në mënyrë manuale, siç ndodh në pyetësorët me letër). Kështu, nëse një person është më i vogël se 15 vjeç, pyetësori lidhur me statusin e punësimit kalohet automatikisht.

Anketa ATN është një studim tremujor i cili i drejtohet ndërmarrjeve. Nga viti 2003-2013 është përdorur metoda PAPI për mbledhjen e informacionit. Që nga viti 2013 është bërë përpjekje për të përdorur një metodë të përzier metodash: PAPI + pyetësorë online. Në vitin 2014, 91% e kompanive, me 10+ të punësuar, kanë pasur qasje në internet. Shkalla e lartë e depërtimit në internet, veçanërisht në mesin e kompanive të mëdha, është arsyeja pse qasja në internet është duke u përdorur për këtë studim. PAPI është aplikuar si metodë bazë, ndërsa format web janë përdorur vetëm për kompanitë e mëdha (rreth 300).

Anketa e TIK është kryer për herë të parë në vitin 2015, duke përdorur tableta (CAPI) për të mbledhur informacionin. Ky studim është përqendruar në disponueshmërinë e teknologjive të informacionit dhe komunikimit (TIK) dhe përdorimin e tyre nga ndërmarrjet.

### 2.1. KOSTO

Ekziston një kërkesë në rritje për të dhënat statistikore nga përdoruesit, jo vetëm nga politikëbërësit, por edhe nga përdorues të tjerë. Çdo agjenci statistikore duhet të marrë vendime ndërmjet prodhimit të të dhënave statistikore dhe kostove për ta realizuar atë. Mbledhja e të dhënave të tilla është e kushtueshme, veçanërisht në vendet me sisteme të dobët adresash dhe infrastrukture. Kostot ndikojnë në çdo proces statistikor. Në këtë artikull merren parasysh kostot e mbledhjes së të dhënave, pasi në përgjithësi ky proces është më i shtrenjtë. Zërat e buxhetit të marra në konsideratë për të analizuar koston për pyetësorin (në lekë) janë paraqitur në Tabelën 1.

Për të llogaritur kostot e vrojtimeve, me metoda të ndryshme të mbledhjes së të dhënave, janë marrë në konsideratë vrojtmet e mësipërme (AFP, ATN, TIK). Në këtë llogaritje, shpenzimet fillestare të pajisjeve (laptop / tableta) nuk janë marrë në konsideratë, duke supozuar që këto pajisje përdoren për disa vrojtime dhe shpenzimet duhet të shpërndahen. Gjithashtu, shpenzimet e aplikacioneve për mbledhjen e të dhënave nuk janë të përfshira.

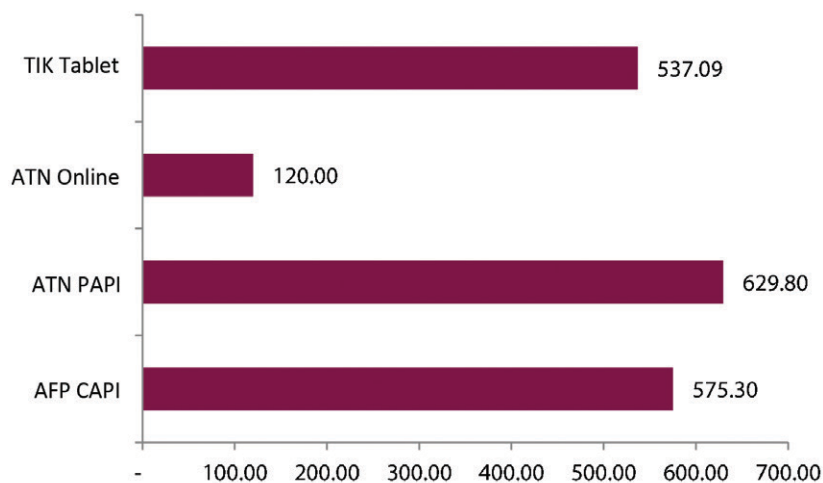
Figura 2 tregon kostot e vrojtimeve për pyetësorë. Siç pritet, qasja në internet, duke përdorur format e webit, është më e lirë. Kjo qasje është duke u përdorur deri tani vetëm për kompanitë e mëdha STS, duke e ditur me siguri se ata kanë infrastrukturën e nevojshme për t'iu përgjigjur pyetësorit dhe gjithashtu, INSTAT ka mundësinë t'i dërgojë kredencialet e tyre me postë ose email. Kostot për CAPI, duke përdorur laptop ose tableta, janë pothuajse të njëjta, duke përjashtuar kostot fillestare për pajisjet, të cilat mund të jenë shumë të ndryshme, pavarësisht nga teknologjitë e përdorura, edhe pse po zvogëlohet shumë shpejt. Metoda PAPI, e cila është akoma metoda më e përdorur në INSTAT, ka koston më të lartë nga metodat e tjera të mbledhjes së të dhënave.

Tabela 1: Zërat e buxhetit të vrojtimeve të marra në konsideratë

|                                     | AFP              | ATN PAPI         | ATN Online       | TIK Tablet       |
|-------------------------------------|------------------|------------------|------------------|------------------|
| <b>Shtypja e pyetësorëve</b>        | Jo e aplikueshme | Vlerë            | Jo e aplikueshme | Jo e aplikueshme |
| <b>Shtypja e materialeve shtesë</b> | Vlerë            | Vlerë            | Jo e aplikueshme | Vlerë            |
| <b>Intervistuesit</b>               | Vlerë            | Vlerë            | Jo e aplikueshme | Vlerë            |
| <b>Numri i Intervistuesve</b>       | Vlerë            | Vlerë            | Jo e aplikueshme | Vlerë            |
| <b>Kontrollorë</b>                  | Vlerë            | Vlerë            | Jo e aplikueshme | Jo e aplikueshme |
| <b>Mallra Konsumi</b>               | Vlerë            | Vlerë            | Jo e aplikueshme | Vlerë            |
| <b>Dieta për Intervistuesit</b>     | Jo e aplikueshme | Vlerë            | Jo e aplikueshme | Vlerë            |
| <b>Operatorët e të dhënave</b>      | Jo e aplikueshme | Value            | Jo e aplikueshme | Jo e aplikueshme |
| <b>Kartat e Internetit</b>          | Vlerë            | Jo e aplikueshme | Jo e aplikueshme | Vlerë            |
| <b>Shërbimi Postar</b>              | Jo e aplikueshme | Jo e aplikueshme | Vlerë            | Jo e aplikueshme |
| <b>Gjithsej</b>                     | Vlerë            | Vlerë            | Vlerë            | Vlerë            |

Burimi: Të nxjerë nga autorët bazuar në informacionin e INSTAT-it

Figura 2: Shpenzimet për pyetësor për vrotimet sipas metodave të mbledhjes së të dhënave (në Lekë)



Burimi: Llogaritjet e autorëve

## 2.2. SHKALLA E PËRGJIGJES

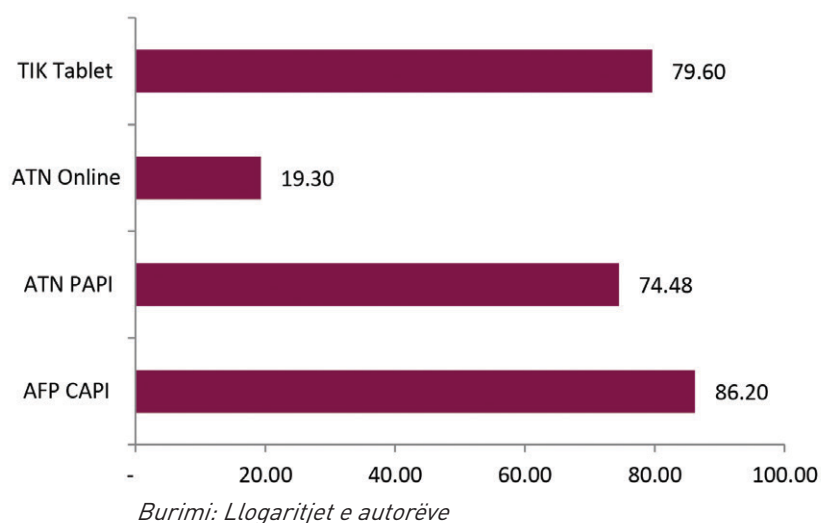
Mos-përgjigja është një tipar i përbashkët i vrojtimeve me zgjedhje. Një pjesë e elementëve të përzgjedhur nuk mund të kontaktohen, refuzojnë pjesëmarrjen ose nuk marrin pjesë në studim për arsye të tjera (Fannie Cobben, 2009).

Në Shqipëri, plotësimi vullnetarisht i një pyetësoi është relativisht një kulturë e re. Në përgjithësi njerëzit bashkëpunojnë me intervistuesit për të plotësuar pyetësoret. Ende është e vështirë për të gjetur të anketuarit në shtëpi gjatë orëve të ditës, pasi mënyra e jetesës ka ndryshuar. Popullsia është bërë më e lëvizshme dhe njerëzit nuk janë në vendin e tyre të zakonshëm të qëndrimit për periudha të gjata. Të gjitha këto ndryshime mund të ndikojnë në normën e përgjigjes. Për këtë qëllim, llogaritet shkalla e përgjigjes, të dhënë në Figurën 3. Shkalla e përgjigjes llogaritet duke pjesëtuar numrin e pyetësorëve të plotësuar me madhësinë e vërtetimit.

Bradburn (1978) sugjeroi që përkufizimi i ngarkesës së të intervistuarve duhet të përfshijë katër elemente; gjatësinë e intervistës, përpjekjet e të anketuarve, stresi i anketuarve, dhe frekuencën e intervistimit. Përpjekjet e të anketuarve mund t'i referohen sfidave njohëse të një detyre. Shkalla më e vogël e përgjigjes është për qasjen online, rreth 19.3%. Edhe në vërtetimet online, njësitë e anketuara mund të përgjigjen kur ata dëshirojnë, por ngarkesa ndaj të anketuarve është më e lartë, pasi ata duhet të kuptojnë, plotësojnë, kryejnë veprime dhe të mbajnë shënime.

Tradicionalisht, përdorimi i intervistuesve, për mbledhjen e të dhënave, ka rezultuar më efikase sa i përket shkallës së përgjigjes; rreth 80%. Të anketuarit u besojnë intervistuesve dhe bashkëpunojnë me ta më mirë, duke rezultuar në nivele më të larta përgjigjesh. Intervistuesit janë të pajisur me mjetet identifikuese dhe kanë me vete një letër zyrtare që shpjegon fushën e vërtetimit dhe aspektet e konfidencialitetit të të dhënave.

**Figura 3: Shkalla e përgjigjes sipas metodave të ndryshme të grumbullimit të të dhënave dhe anketave (në%)**



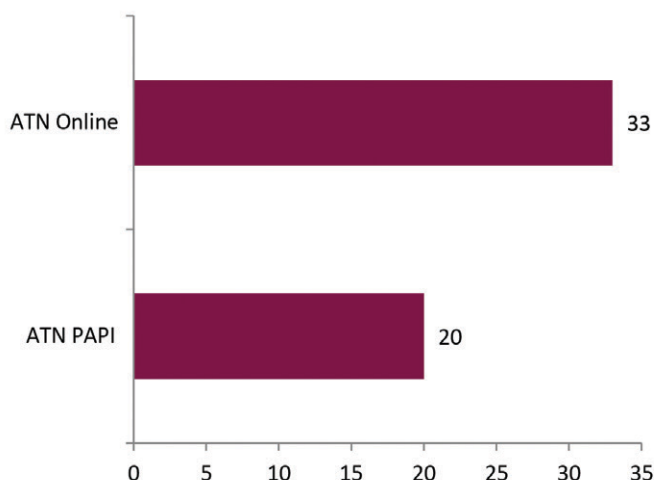
Ekziston një situatë specifike në Shqipëri, ku sistemi i adresave nuk funksion siç duhet. Kjo e bën shumë të vështirë dërgimin e pyetësorëve dhe kredencialeve për format online, me anë të postës. Për anketat e ekonomive familjare, INSTAT përdor sistemin e saj GIS, i dizajnuar për Censusin e Popullsisë dhe Banesave, 2011. Shkalla e përgjigjes për AFP është më e lartë, për shkak të këtij fakti. Shkalla e përgjigjes për AFP i referohet mesatares vjetore për vitin 2014, për ATN PAPI është mesatarja vjetore e vitit 2013, për ATN online është shkalla e përgjigjes për tremujorin e katërt 2015, ndërsa për TIK tablet është shkalla e përgjigjes për vitin 2015.

## 2.3. NGARKESA E TË INTERVISTUARVE

Matja e ngarkesës së të intervistuarve do të bëhet duke marrë parasysh kohëzgjatjen e intervistës, për të njëjtin studim, në metoda të ndryshme të mbledhjes së të dhënave. Vrojtimi ATN është testuar në dy metoda të ndryshme; PAPI dhe internet. I njëjti pyetësor përdoret për të dyja metodat. Në pyetësorin në internet janë shtuar shënime shtesë për të shpjeguar pyetjen; kjo është puna që bën intervistuesi në metodën ballë për ballë. Pyetësi online është drejtuar tek i njëjti individ, njësoj si për metodën PAPI (gjatë metodës PAPI kontakti me ndërmarrjen është vendosur më parë).

Natyrisht, metoda online ka një ngarkesë më të madhe për të intervistuarit (në rastit tonë kompanitë). Kjo lidhet kryesisht me faktin se i anketuari ka nevojë për më shumë kohë për të kuptuar pyetjet.

**Figura 4: Koha mesatare e shpenzuar për metodën PAPI dhe online (në minuta)**



Burimi: Llogaritjet e autorëve

## 2.4. KAPJA E TË DHËNAVE

Zhvillimet e mëdha dinamike në Teknologjinë e Informacionit (IT) dhe përparimet e shpejta në ndryshimet teknologjike, po ndikojnë në mënyrën e punës për zyrat statistikore. Sfida e përditshme e INSTAT-it është identifikimi i tendencave dhe gjetja e zgjidhjeve më të përshtatshme për të prodhuar statistika duke analizuar koston dhe efektivitetin. Pothuajse çdo vërtetim i kryer nga INSTAT përdor intervistën ballë për ballë. Për të lehtësuar procesin e kapjes së të dhënave, në tetor 2009 u instalua sistemi i skanimit për të përgatitur punën

në vijim për Censuset e ardhshëm.

Arritjet e INSTAT, në lidhje me kapjen e të dhënave janë:

- Në vitin 2011 skanimit dhe kapja e të dhënave për Censusin e Ndërmarrjeve Ekonomike
- Në vitin 2011-2012 skanimit dhe kapja e të dhënave për Censusin e Popullsisë dhe Banesave
- Në 2013 skanimit dhe kapja e të dhënave për Censusin e Bujqësisë
- Nga viti 2010 deri më sot skanimit dhe kapja e të dhënave për vërtetime të ndryshme

Qasja e teknologjisë së njohjes inteligjente të karaktereve (Intelligent Character Recognition-ICR) zvogëlon kohën për procesin e kapjes së të dhënave, veçanërisht për aktivitetet e mëdha statistikore. Institutet kombëtare të statistikave gjithmonë analizojnë kohëzgjatjen dhe cilësinë e të dhënave. Lidhur me këtë është bërë një analizë për të vlerësuar cilësinë e të dhënave, për Censusin e Ndërmarrjeve Ekonomike.

Kur analizohen të dhënat e skanuara, ekzistojnë disa mënyra për të ndarë sasinë e të dhënave sipas pyetësorit, faqes, nivelit të fushës ose karakterit. U zgjedh niveli më i detajuar - niveli i karakterit - ku karakteret e regjistruara, në këtë rast në skedarët e transferimit, krahasohen me karakteret nga pyetësi. Kutitë u përzgjedhën rastësisht dhe nga këto kuti u zgjedhën rastësisht disa pyetësorë. Të gjithë karakteret u krahasuan, pra ato karaktere që ishin të regjistruara dhe që kishin rezultuar të ishin zëvendësuar, interpretuar gabim ose të pasakta.

Figura 5 tregon se si verifikuesi interpreton "6" si "0". Nëse verifikuesit e kanë pranuar "0" ose ta ndryshojnë atë në, për shembull "8", ky konsiderohet një zëvendësues i mbetur. Fushat u ndanë në tre lloje të ndryshme fushash; me shkrim dore, kutia e kontrollit dhe njohja e karakterit optik (OCR) (domethënë të para-shtypur).

**Figura 5: Shembull i një zëvendësimi**



Burimi: Censusi i Ndërmarrjeve Ekonomike, procesi i kapjes së të dhënave

Në Tabelën 2 janë paraqitur numrin e pyetësorëve të kontrolluar për të analizuar të dhënat e skanuara për Censusin e Ndërmarrjeve Ekonomike.

**Tabela 2: Pyetësorët e kontrolluar**

| Numri i kutisë | Numri i pyetësorëve | Numri i kontrolluar i pyetësorëve | Përqindja e pyetësorëve të kontrolluar |
|----------------|---------------------|-----------------------------------|----------------------------------------|
| <b>4101</b>    | 289                 | 14                                | 4.84                                   |
| <b>4601</b>    | 252                 | 7                                 | 2.78                                   |
| <b>5101</b>    | 245                 | 12                                | 4.90                                   |
| <b>5501</b>    | 296                 | 8                                 | 2.70                                   |
| <b>6201</b>    | 162                 | 4                                 | 2.47                                   |
| <b>3902</b>    | 259                 | 14                                | 5.41                                   |
| <b>4702</b>    | 68                  | 8                                 | 11.76                                  |
| <b>Total</b>   | <b>1571</b>         | <b>67</b>                         | <b>4.26</b>                            |

Burimi: Llogaritjet e autorëve

Në Tabelën 3 janë paraqitur numrin e pyetësorëve të zëvendësuar për të analizuar cilësinë e kapjes së të dhënave për Censusin e Ndërmarrjeve Ekonomike.

**Tabela 3: Rezultati i zëvendësimeve**

| Tipi i Fushës      | Numri i Karaktereve | Numri i zëvendësimeve | Përqindja e zëvendësimeve |
|--------------------|---------------------|-----------------------|---------------------------|
| <b>OCR numeric</b> | 1474                | 0                     | -                         |
| <b>Handwritten</b> | 18610               | 28                    | 0.15                      |
| <b>Checkbox</b>    | 4110                | 1                     | 0.02                      |
| <b>Total</b>       | 24194               | 29                    | 0.12                      |

Burimi: Llogaritjet e autorëve

Qëllimi kryesor i kontrollit të të dhënave të skanuara ishte matja e cilësisë duke numëruar zëvendësimet. U gjetën pak zëvendësimet, siç tregohet edhe në Tabelën 3. Fushat me shkrim me dorë përfaqësonin shumicën e zëvendësime të gjetura, të cilat nuk ishin shumë të habitshme. Një shembull tipik ishte shkronja "I" që ishte bërë numër "1". Fushat e OCR-së nuk kishin fare zëvendësime; vetëm një rast zëvendësimi u gjet për kutitë e kontrolluar. Tabela 3 pasqyron një cilësi shumë të mirë të të dhënave të kapura/të transferuara, vetëm 0,12 për qind e karaktereve ishin zëvendësuar.

## 2.5. CILËSIA E TË DHËNAVE

Përdoruesit e të dhënave duhet që të kenë gjithmonë të paktën disa informacione në çfarë shkalle vrojtimit janë modeluar apo përgjithësuar si pasojë e imputimit. Në fund të procesit të imputimit është e nevojshme të prodhohen indikatorët e mëposhtëm (Statistics Canada, 2010):

- Numri i rekordeve të cilat janë imputuar (për shembull numri i rekordeve që kanë qenë subjekt i imputimit);

- Numri i rasteve të cilat janë imputuar dhe nga cila metodë;

- Numri i rekordeve të cilat kanë shërbyer si donatorë;

- Numri i rekordeve të përdorur si donatorë dhe numri i fushave të imputuara;

- Një listë që tregon se cilët donatorë janë përdorur për secilin marrës (për të gjetur burimet e të dhënave të imputuara);

- Një listë e të gjitha të dhënave për të cilat imputimi nuk funksionon (për shembull, nëse nuk u gjet asnjë donator)

Eurostat propozon indikatorët standard të mëposhtëm të cilësisë që mund të përdoren nga prodhuesit e statistikave, për të përmbledhur cilësinë e statistikave të raportuara sipas Standardeve të Raportit të Cilësisë (ESTAT / 02 / Quality / 2005/9 / Quality Indicators). Ky grup treguesish mund të përdoret për të matur dhe ndjekur me kalimin e kohës cilësinë e të dhënave të prodhuara në Sistemin Evropian të Statistikave (ESS)<sup>1</sup>.

Treguesit standard të cilësisë të propozuar nga Eurostat janë:

- Niveli i Kënaqësisë
- Shkalla e statistikave në dispozicion
- Koeficienti i variacionit
- Shkalla e përgjigjes nivel variabli (e pa-peshuar / e peshuar)
- Shkalla e përgjigjes nivel rekordi (e pa-peshuar / e peshuar)
- Shkalla dhe raporti i imputimit
- Shkalla e mbi-mbulimit dhe klasifikimit të gabuar
- Shkalla e nën-mbulimit gjeografik
- Madhësia mesatare e rishikimeve
- Përpikëria e botimeve
- Publikimi i rezultateve fillestare
- Publikimi i rezultateve finale
- Numri i publikimeve të shpërndara dhe / ose të shitura
- Konsultimet në bazat e të dhënave
- Shkalla e plotshmërisë së metadatave
- Gjatësia e serive kohore të të dhënave
- Numri i serive kohore të krahasueshme
- Shkalla e dallimeve në lidhje me konceptet dhe matjeve sipas normave evropiane
- Shkalla e asimetrive për statistikën e njëjta
- Shkalla e statistikave që plotësojnë kërkesat për përdorimin kryesor apo dytësor

Grupi i treguesve që u shqyrtua për të bërë vlerësimin e efekteve të procedurës së pastrimit në nivel të agreguar, mund të përmbledhet në tre tregues të ndryshëm (INSTAT Quality Dimensions, 2014):

1. Treguesit për sasinë e të dhënave që i nënshtrohen procedurës së imputimit si: numri i rekordeve, numri i variablave, numri i variablave objekt i procedurës së imputimit dhe numri i vlerave totale.

2. Treguesit për vlerësimin e efekteve të përgjithshme të procedurës së imputimit, si: i) shkalla e imputimit (I):  $(\text{numri i vlerave të imputuara} / \text{numri i vlerave totale}) * 100$ ; ii) shkalla e shtimit (Ia):  $(\text{numri i shtimeve} / \text{numri i vlerave totale}) * 100$ ; iii) shkalla e ndryshimeve (Im):  $(\text{numri i ndryshimeve} / \text{numri i vlerave totale}) * 100$ ; iv) shkalla e eliminimit (Ie):  $(\text{numri i eliminimeve} / \text{numri i vlerave totale}) * 100$

3. Treguesit sintetikë për shkallën e imputimeve të rekordeve, për shembull: numri i rekordeve me shkallë imputimi më të madhe sesa 2% dhe numri i rekordeve me shkallë imputimi më të madhe sesa 5%.

Nga treguesit e paraqitur në tabelën 4, shkalla e imputimit për LFS është rreth 0,01%, kështu që metoda e mbledhjes së të dhënave CAPI është

<sup>1</sup> [https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality\\_reports/reference\\_documents/STANDARD\\_QUALITY\\_INDICATORS.pdf](https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality_reports/reference_documents/STANDARD_QUALITY_INDICATORS.pdf)

metodë shumë e mirë për të monitoruar dhe kontrolluar cilësinë e të dhënave. PAPI, siç pritet, është metoda ku imputimi dhe editimi kanë shifrat më të larta, rreth 4,71%, të vlerave totale.

të mëdha që kanë rinovimet e licencave. Ndërkohë, para se procesi të realizohet plotësisht, ka ende nevojë

**Tabela 4: Treguesit e vlerësimit të cilësisë në nivel të agreguar**

|                                             | <b>AFP CAPI</b> | <b>ATN PAPI</b>  | <b>ATN Online</b> | <b>TIK Tablet</b> |
|---------------------------------------------|-----------------|------------------|-------------------|-------------------|
| <b>Numri i rekordeve</b>                    | 12,490          | 6,803            | 60                | 1,473             |
| <b>Numri i variablave</b>                   | 164             | 59               | 59                | 93                |
| <b>Numri i vlerave totale</b>               | 2,048,360       | 401,377          | 3,540             | 136,989           |
|                                             |                 |                  |                   |                   |
| <b>Numri i vlerave të imputuara</b>         | <b>20,484</b>   | <b>1,890,486</b> | <b>3,929</b>      | <b>56,165</b>     |
| <b>Numri i shtimeve</b>                     | -               | 545,873          | 2,443             | 5,480             |
| <b>Numri i eliminimeve</b>                  | -               | 8,028            | -                 | 16,439            |
| <b>Numri i ndryshimeve</b>                  | 20,484          | 1,336,585        | 1,487             | 34,247            |
|                                             |                 |                  |                   |                   |
| <b>Shkalla e imputimit (I)</b>              | <b>0.01</b>     | <b>4.71</b>      | <b>1.11</b>       | <b>0.41</b>       |
| <b>Shkalla e shtimeve (Ia)</b>              | -               | 1.36             | 0.69              | 0.04              |
| <b>Shkalla e eliminimit (Ie)</b>            | -               | 0.02             | -                 | 0.12              |
| <b>Shkalla e ndryshimeve (Im)</b>           | 0.01            | 3.33             | 0.42              | 0.25              |
| <b>Shkalla e mos-imputimit</b>              | <b>99.99</b>    | <b>95.29</b>     | <b>98.89</b>      | <b>99.59</b>      |
|                                             |                 |                  |                   |                   |
| <b>% e rekordeve me I më të madhe se 2%</b> | <b>5</b>        | <b>115</b>       | <b>26</b>         | <b>112</b>        |
| <b>% e rekordeve me I më të madhe se 5%</b> | <b>5</b>        | <b>13</b>        | <b>18</b>         | <b>9</b>          |

Burimi: Llogaritjet e autorëve

Është e qartë, nga të dhënat e mësipërme, se përfitimet e metodës CAPI nuk janë vetëm në kohën e përpunimit (pasi futja e të dhënave bëhet gjatë procesit të grumbullimit të të dhënave), por edhe në cilësinë e të dhënave.

### 3. PROCESET E TJERA STATISTIKORE, PËRVEÇ MBLEDHJES SË TË DHËNAVE

INSTAT synon të zëvendësojë software statistikorë SAS dhe SPSS me softuerit statistikor R, i cili është një software open source, i cili është i aftë të përpunojë statistika, si edhe software e tjerë. Ky zëvendësim do të bëhet për shkak të shpenzimeve

për përdorimin e SAS dhe SPSS. Pra, plani i kalimit nga SAS / SPSS në R do të realizohet gradualisht. Në shtator të vitit 2012, INSTAT ka zhvilluar një faqe interneti të re dhe funksionale për t'u përdorur sa më lehtë nga përdoruesit. Përdorimi i Modelit Common Nordic Meta Model (CNMM) do t'i lejojë përdoruesit të përdorin të dhënat statistikore në mënyrë dinamike. Baza e të dhënave statistikore e INSTAT-it është ndërtuar me ndërfaqen PX-Web. Nga janari 2016 ky software është open source, ndërsa PX-Web është pa pagesë që nga janari 2015. PX-Web është zhvilluar dhe është pronë e Statistikave të Suedisë, dhe është përdorur për të krijuar tabelat dinamike.

Çdo përpjekje e përmendur në seksionet e mësipërme është realizuar për zhvillimin e një sistemi statistikor të qëndrueshëm në Shqipëri. Ky sistem duhet të lehtësojë vendimmarrjen bazuar në të dhënat statistikore relevante dhe të besueshme, që të plotësojnë nevojat e përdoruesve.

## 4. KONKLUZIONE

Metoda PAPI ka një shkallë përgjigjeje të mirë; intervistuesi ka një efekt pozitiv në këtë aspekt. Intervistuesit e INSTAT-it janë të pajisur me materiale shtesë (harta, letra prezantuese, etj.) për të përmirësuar cilësinë e të dhënave. Bazuar në vlerësimin e cilësisë, pyetësorët kompleks (kalimet në pyetësor, listat) janë të vështira për t'u plotësuar.

Metoda CAPI ndikohet pozitivisht nga përdorimi i intervistuesve, veçanërisht në shkallën e përgjigjeve. Intervistuesit e INSTAT-it janë të pajisur me materiale shtesë (harta, letra prezantuese, etj.). Procesi i hedhjes së të dhënave bëhet gjatë fazës së mbledhjes së të dhënave dhe mund të përdoren edhe pyetësorët më kompleks. Lidhur me cilësinë e të dhënave të grumbulluara, kjo metodë redukton gabimet e intervistuesit, pasi gjatë grumbullimit të të dhënave zbatohen edhe rregullat e plotshmërisë

dhe kalimet në pyetësor. Pajisjet e përdorura për këtë metodë, edhe pse çmimet janë në rënie, janë ende të kushtueshme. Përdorimi i laptopëve ose pajisjeve të tjera në terren nuk është shumë e lehtë dhe nevojitet një trajnim i mirë. Intervistuesit dhe të anketuarit kanë reaguar pozitivisht ndaj përdorimit të laptopëve dhe tabletave.

Të dhënat e prodhuara për këtë artikull mbështesin supozimin se kostoja e laptopëve ose tabletave të përdorura për procesin e grumbullimit të të dhënave është e ndarë nga vrojtime të ndryshme që përdorin këtë pajisje. Metoda në internet ka koston më të ulët dhe kjo metodë ka edhe shkallën më të ulët të përgjigjes. Ngarkesa e të intervistuarit është më e lartë në metodën online, për shkak se ajo kërkon më shumë kohë për të lexuar dhe kuptuar pyetjet.

CAPI dhe metoda online lejojnë transmetimin e menjëhershëm të të dhënave në zyrat qendrore, duke mundësuar përpunimin e menjëhershëm të të dhënave. Përfshirja e inovacionit në procesin e mbledhjes së të dhënave për prodhimin e statistikave zyrtare duhet të bëhet duke analizuar koston, cilësinë e të dhënave dhe kohën e përpunimit.

## BIBLIOGRAFI

Bradburn, N. M. "Respondent Burden", Paper presented at the 138th Annual Meetings of the American Statistical Association, San Diego, CA., 1978

ESTAT/02/Quality/2005/9/Quality Indicators,

Fannie Cobben, Statistics Netherlands, Nonresponse in Sample Surveys, 2009

[https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality\\_reports/reference\\_documents/STANDARD\\_QUALITY\\_INDICATORS.pdf](https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality_reports/reference_documents/STANDARD_QUALITY_INDICATORS.pdf)

INSTAT, Quality Dimensions of the 2011 Population and Housing Census of Albania, May 2014

Statistics Canada (2010) Survey Methods and Practices. Statistics Canada, Catalogue no. 12-587-X, Ottawa. <http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.htm>