

**COMISIA CENTRALĂ PENTRU RECENSĂMÂNTUL
POPULAȚIEI ȘI AL LOCUINȚELOR - 2021
(CCRPL2021)**

***Metodologia de imputare aplicabilă pentru
recensământul populației și locuințelor
runda 2021***

CUPRINS

1. Introducere

2. Definirea și clasificarea non-răspunsurilor

3. Metode de tratare a non-răspunsurilor totale

4. Metode de imputare

5. Modalități de tratare a non-răspunsurilor parțiale

Bibliografie

1. Introducere

Operațiunea de colectare a datelor în cadrul unei cercetări statistice (în special în cazul celor exhaustive de tipul recensământului) presupune acțiunea și voința unui număr foarte mare de persoane atât din rândul populației cât și al personalului de recensământ. O consecință a acțiunilor acestora poate fi omiterea, voită sau involuntară, de a răspunde la unele întrebări ale recensământului sau chiar la toate.

Lipsa unui număr de înregistrări totale ale unor unități statistice ce aparțin populației țintă se numește în literatura de specialitate sub-acoperire, iar prezența în setul de date a unor înregistrări individuale aferente unor unități statistice care nu aparțin acesteia se numește supra-acoperire. Atât sub-acoperirea cât și supra-acoperirea determină probleme de calitate ale rezultatelor recensământului, prin urmare ele trebuie eliminate sau, cel puțin, diminuate în procesul de prelucrare a rezultatelor. În continuare, documentul tratează problematica sub-acoperirii.

Non-răspunsul reprezintă o realitate cu care se confruntă toate cercetările statistice, inclusiv recensămintele. Determinat de anumite cauze, dependente sau nu de unitățile eligibile, **non-răspunsul, sau lipsa de informație pentru o anumită unitate statistică sau pentru o anumită variabilă, generează riscul ca valorile indicatorilor obținuți pe baza datelor colectate de la unitățile (statistice) de observare să fie deplasate.** Se impune găsirea unor soluții de tratare a acestora.

2. Definirea și clasificarea non-răspunsurilor

Non-răspunsul reprezintă eșecul de a obține măsurări ale diferitelor variabile asupra unităților aflate în sfera de acoperire a recensământului. Tendința generală în toate țările este de creștere a ratelor de non-răspuns, fapt pentru care instituțiile/organizațiile care desfășoară cercetări

statistice trebuie să acorde o atenție din ce în ce mai mare acestui aspect, fiind imperios necesare măsuri de limitare a efectelor non-răspunsurilor.

O primă clasificare a non-răspunsurilor este cea realizată în funcție de cauzele de apariție a acestora. Motivele pentru care pot apărea non-răspunsuri sunt multiple, ele fiind determinate, în principal, de patru factori: mediul social, individul, personalul care realizează colectarea datelor și metoda de colectare a datelor.

O altă clasificare a non-răspunsurilor este aceea între:

- non-răspunsul total
- non-răspunsul parțial

Non-răspunsul total apare când pentru o unitate statistică (în cazul recensământului: persoană, gospodărie, locuință, clădire) nu se furnizează nici o informație în cadrul recensământului (nu se completează nici un răspuns la niciuna dintre întrebările recensământului). Dacă nu s-ar aplica metode de eliminare, acest tip de non-răspuns ar avea ca efect reducerea mărimii populației țintă, ceea ce ar conduce la o precizie mai mică a tuturor celorlați indicatori.

Non-răspunsul parțial intervine când se obțin de la unitatea observată răspunsuri doar la o parte din întrebările/variabilele recensământului, dar nu la toate.

3. Metode de tratare a non-răspunsurilor totale

Prevenirea și reducerea amplitudinii fenomenului de non-răspuns se poate realiza atât înainte cât și în timpul perioadei de colectare a datelor, prin diferite metode.

După etapa de colectare a datelor, tratarea non-răspunsurilor poate fi realizată în faza de prelucrare și analiză a datelor. Scopul urmărit este acela de a completa setul de înregistrări de unități statistice ce compun o populație țintă cu acele înregistrări care aparțin acesteia, dar care

nu s-au putut colecta, din diverse motive, prin metodele alese pentru recensământ (auto-recenzare sau interviu față-în-față cu recenzorul). În cazul recensământului, referitor la populație, prin aceasta se urmărește obținerea unui set de înregistrări unice de persoane care să asigure gradul cel mai ridicat de acoperire a populației rezidente a României. De asemenea, prin această etapă se realizează eliminarea sau măcar reducerea cât mai mult posibil a sub-acoperirii populației rezidente și creșterea calității statisticilor produse pe baza datelor colectate.

Problema non-răspunsurilor totale se poate rezolva:

- prin metoda imputării totale
- prin folosirea unor coeficienți de extindere ca în cazul cercetărilor selective, pentru recensământ coeficienții de extindere având rolul de a compensa doar lipsa unor înregistrări care ar fi trebuit să fie prezente în baza de date, dar nu sunt.

Imputarea se poate defini ca fiind înlocuirea unei valori care lipsește cu o valoare artificială, dar de bună calitate, preluată direct din alte surse sau determinată / estimată pe baza unor metode statistice.

Metoda imputării totale presupune adăugarea unei înregistrări individuale unice în baza de date pentru o anumită unitate statistică pentru care s-a constatat că face parte din populația țintă, dar pentru care nu s-a reușit colectarea nici unei informații/ variabile prin metodele de colectare alese. Imputarea se realizează după etapa de colectare a datelor, ca rezultat al unui algoritm care determină, pe baza unor criterii pre-stabilite, apartenența la populația țintă (în cazul recensământului, la populația rezidentă). Înregistrarea individuală este adăugată în baza de date de înregistrări individuale rezultată în urma colectării datelor (prin auto-recenzare sau interviu față-în-față cu recenzorul), iar valorile fiecărei variabile de recensământ ce compun această înregistrare individuală adăugată se obțin prin colectare indirectă din surse administrative sau prin metoda imputării parțiale.

În acest fel se obțin informațiile lipsă pentru acele unități statistice pentru care metodele de colectare stabilite inițial au eșuat, baza de date de recensământ devine exhaustivă (cuprinde toate unitățile statistice ce aparțin populației țintă), indicatorii calculați pe datele cuprinse în

această bază de date nu mai sunt afectați de deplasare (reușind să determine cu o probabilitate neglijabilă de eroare statisticile de recensământ).

Persoanele pentru care se adaugă înregistrări individuale în baza de date de recensământ ca urmare a aplicării metodei imputării totale sunt acele persoane pentru care ar fi trebuit să se realizeze colectarea datelor la recensământ, dar aceasta nu s-a realizat, din diverse motive. Pentru a identifica ce unități statistice (persoane) lipsesc din setul de înregistrări ce compun populația țintă (populația rezidentă a României la momentul de referință al recensământului din runda 2021 - ora "0" a zilei de 1 decembrie 2021-) la recensământ, setul de înregistrări individuale stocate în baza de recensământ inițială, cea care conține înregistrările obținute din colectarea on-line a datelor (prin metoda CAWI) și din colectarea realizată de recenzori în teren (prin metoda CAPI), se compară cu alte seturi de înregistrări care conțin aceleași unități statistice (în cazul acesta, persoane) deținute de diverse ministere și instituții ale administrației publice centrale și locale.

Comparația și legătura între înregistrările aferente aceleiași unități statistice (persoane) aflate în surse diferite se realizează pe baza identificatorului statistic unic¹ (ISU).

Înregistrările lipsă rezultate în urma comparației se adaugă în baza de date a recensământului.

Seturile de înregistrări care conțin informații privind persoanele care sunt utilizate pentru completarea bazei de date de recensământ provin din surse administrative gestionate de următoarele ministere/instituții și se referă la următoarele caracteristici:

a) Ministerul Afacerilor Interne, Direcția pentru Evidența Persoanelor și Administrarea Bazelor de Date (D.E.P.A.B.D.):

- pentru datele privind cetățenii români cu domiciliul în România la 1 iulie și respectiv, decembrie 2021: CNP sau ISU, cod localitate administrativă, denumire localitate administrativă, cod județ, starea civilă, țara în care s-a născut, județul în care s-a născut,

¹ Ordonanța de Urgență a Guvernului nr. 19/2020 definește identificatorul statistic unic (ISU) ca fiind un cod numeric unic, generat în scopuri exclusiv statistice pentru fiecare CNP, în urma aplicării asupra fiecărui CNP a unui algoritm specific de anonimizare utilizat într-o bază de date sau în orice listă de unități statistice și folosit pentru a verifica că baza de date sau lista de unități statistice include o singură înregistrare pentru fiecare unitate statistică și pentru a se realiza legătura între înregistrări.

- adresa de domiciliu - denumire stradă, număr imobil, bloc, scară, etaj, apartament - data stabilirii la adresa curentă, reședința anterioară 1, data stabilirii la reședința anterioară 1, data valabilității actului de identitate; dacă informația există în RNEP, CNP sau ISU al tatălui, CNP sau ISU al mamei, CNP sau ISU al soțului sau soției, CNP sau ISU al copiilor;
- pentru datele privind cetățenii români care și-au stabilit sau restabilit domiciliul sau reședința obișnuită în România în cursul perioadei de referință: CNP sau ISU, țara de cetățenie, țara de naștere, țara de proveniență, data stabilirii domiciliului sau reședinței în România, județul de domiciliu, localitatea de domiciliu, județ de reședință, localitate de reședință, starea civilă;
 - b) Ministerul Afacerilor Interne, Inspectoratul General pentru Imigrări (I.G.I.):
 - pentru datele privind cetățenii străini și fără cetățenie care au un act care atestă dreptul de ședere în România la data de 1 ianuarie a anului de referință: CNP sau ISU, țara de cetățenie, țara de naștere, județul în care are reședința, localitatea în care are reședința, starea civilă, durata ultimului drept de ședere sau reședință;
 - pentru cetățenii străini și fără cetățenie care au solicitat azil sau au o durată efectivă de ședere de cel puțin 12 luni: CNP sau ISU, țara de cetățenie, țara de naștere, județul în care are reședința, localitatea în care are reședința, starea civilă;
 - pentru cetățenii străini și fără cetățenie care au obținut în cursul anului de referință un act care atestă dreptul de ședere în România: CNP sau ISU, sex, data nașterii, țara de cetățenie, țara de naștere, județul în care are reședința, localitatea în care are reședința, starea civilă, stare persoană;
 - pentru cetățenii străini și fără cetățenie cărora, în cursul anului de referință, le-a expirat dreptul de ședere în România, au decedat, au fost expulzați sau, din alte motive, nu mai beneficiază de un drept de ședere pe teritoriul României: CNP sau ISU, sex, data nașterii, țara de cetățenie, țara de naștere, data plecării din România, județul în care avea reședința, localitatea în care avea reședința, starea civilă, motivul excluderii, data și motivul încetării dreptului de ședere;
 - c) Ministerul Afacerilor Interne, Direcția Generală de Pașapoarte (D.G.P.):

- pentru datele privind cetățenii români care la data de 1 decembrie 2021 aveau stabilit domiciliul în străinătate: CNP sau ISU, data aprobării cererii de pașaport cu menționarea statului de domiciliu;

- pentru datele privind cetățenii români care, după anul 2011, și-au stabilit domiciliul în străinătate sau și-au reînnoit pașaportul cu menționarea țării de domiciliu: CNP sau ISU, data aprobării cererii de pașaport cu menționarea statului de domiciliu;

d) Ministerul Afacerilor Interne - Inspectoratul General al Poliției Române (I.G.P.R.): pentru datele privind persoanele condamnate sau arestate preventiv, aflate în centrele de reținere și arestare preventivă ale M.A.I., custodiate în perioada decembrie 2021- iulie 2022: CNP sau ISU, perioada privării de libertate;

e) Ministerul Finanțelor, Agenția Națională de Administrare Fiscală (A.N.A.F.) - pentru datele privind plătitorii de contribuții sociale, impozite pe venit:

- din declarația D112: CNP sau ISU, secțiune declarație, tip asigurat, cod CAEN declarație, an raportare, lună raportare, număr asigurați șomaj, inclusiv asimilați, număr asigurați sănătate, inclusiv asimilați, număr asigurați pensie, inclusiv asimilați, index încercare, cod obligație, tip declarație, zile lucrate condiții normale, zile lucrate condiții deosebite, zile lucrate condiții speciale, număr contracte, casa de sănătate, dacă persoana este pensionar, tip contract, ore normă zilnică, ore lucrate, ore suspendate, data început, data sfârșit, data extragere, index asigurat, funcție asigurat, perioada de raportare, luna an început, luna an sfârșit, ID adresă, cod țară, cod județ, cod localitate, nume localitate, sector, stradă, număr stradă, bloc, scară, etaj, apartament;

- din baza de date a sistemului informatic de administrare fiscală a impozitului pe venitul persoanelor fizice: CNP sau ISU, categorie de venit, anul fiscal de realizare a venitului, cod județ, data de început a contractului de închiriere sau asociere sau numărul autorizației, data de sfârșit a contractului de închiriere sau asociere sau a autorizației individuale, cod categorie venit, cod CAEN;

f) Ministerul Muncii și Protecției Sociale, Casa Națională de Pensii Publice (C.N.P.P.) - pentru datele conform prevederilor din Convenția de colaborare încheiată cu M.M.P.S. și pentru datele

privind beneficiarii de asigurări sociale și indemnizații prevăzute de legi speciale, după modalitatea de plată: CNP sau ISU, identificator CTP, cod fel pensie, tip cod fel pensie, data înscrierii la pensie, modalitatea de plată - M - mandat, C - cont intern, E - export prestație;

g) Ministerul Muncii și Protecției Sociale, Agenția Națională pentru Ocuparea Forței de Muncă (A.N.O.F.M.) - pentru datele privind șomerii care beneficiază de indemnizație de șomaj, șomerii neindemnizați, persoanele în căutarea unui loc de muncă în țară sau, după caz, în străinătate, înregistrate în evidență: CNP sau ISU, categorie persoană, adresă, localitate, sector;

h) Ministerul Muncii și Protecției Sociale, Agenția Națională pentru Plăți și Inspecție Socială (A.N.P.I.S.) - pentru datele privind persoanele care beneficiază de asistență socială: CNP sau ISU al persoanei care încasează beneficiul social, CNP sau ISU al persoanei pentru care se plătește beneficiul social, luna plății beneficiului social, codul județului, codul beneficiului social, numele și prenumele persoanei care încasează beneficiul social; luna pentru care se calculează beneficiul social, cod județ, codul beneficiului social, tipul persoanei, adresa de domiciliu;

i) Ministerul Muncii și Protecției Sociale, Autoritatea Națională pentru Drepturile Persoanelor cu Dizabilități, Copii și Adopții (A.N.D.P.D.C.A.)

- din sistemul informatic CMTIS, dacă în decembrie 2021 acesta va fi funcțional: CNP sau ISU, data ultimei consultații - zi, lună și an, adresa - țara, județ, localitate, tipul adresei din interfața „Adrese“, numărul vizitelor părinților la copii, numărul întâlnirilor copiilor cu părinții în loc neutru, dacă copilul este înscris la o școală, tipul adresei din interfața „Situția școlară“;

- din sistemul informatic RNA: CNP sau ISU adoptator, țara de naștere, județ de naștere, localitate de naștere, data eliberare permis ședere adoptator, țara din interfața „Adresa domiciliu“, județ din interfața „Adresa domiciliu“, localitate din interfața „Adresa domiciliu“;

- date din sistemul informatic RENPH: CNP sau ISU, adresa - județ, localitate, comună, tip persoană, cetățenie, instituționalizare, tip eveniment, dată eveniment, cod centru;

j) Ministerul Muncii și Protecției Sociale, Inspecția Muncii (I.M.): pentru datele din Registrul General de Evidență a Salariaților, în forma transmisă de angajator: CUI angajator, denumire angajator, categorie angajator, forma juridică a angajatorului, tipul de proprietate angajator,

domeniul de activitate economică angajator, județ angajator, localitate sau sector angajator, CNP sau ISU salariat, cetățenia, cod ocupație conform COR, număr contract, dată contract, tip contract, durată contract, tip normă, stare contract, dată începere contract, data încetare contract, data sfârșit contract, data reactivare contract, temei reactivare contract, dată începere suspendare contract, dată sfârșit suspendare contract, dată începere detașare contract, dată sfârșit detașare contract, data ultimei modificări a contractului, radiat salariat, radiat contract;

k) Ministerul Justiției, Autoritatea Națională pentru Cetățenie (A.N.C.): pentru datele privind persoanele care dobândesc cetățenia română în țară: identificator stabilit de comun acord, țara de cetățenie, țara de naștere, data la care a dobândit cetățenia română și, după caz, adresa de domiciliu în România;

l) Ministerul Educației și Cercetării: pentru datele privind persoanele înmatriculate în instituțiile sau unitățile de învățământ din sistemul național de educație, precum și mobilitatea acestora: CNP-ul sau ISU, limba maternă, naționalitate, județul de domiciliu, localitatea de domiciliu, județul în care se află unitatea de învățământ urmată, numărul unității de învățământ urmate, localitate unitate, clasa în care se află elevul, formațiune asociată, limba de predare, categorie unitate;

m) Ministerul Sănătății, Institutul Național de Sănătate Publică (I.N.S.P.): pentru datele privind persoanele cuprinse în evidență privind bolile profesionale: CNP-ul sau ISU, cod CAEN, cod ocupație, dacă bolnavul a decedat, data completării fișei de declarare BP2;

n) Ministerul Sănătății, Casa Națională de Asigurări de Sănătate (C.N.A.S.): pentru datele din PIAS - platforma informatică din asigurările de sănătate: CNP-ul sau ISU, A - asigurat, CO - coasigurat, N - neasigurat, numărul evenimentelor medicale pentru fiecare lună din perioada 1 decembrie 2020 - 30 noiembrie 2021;

o) Ministerul Agriculturii și Dezvoltării Rurale, Agenția de Plăți și Intervenție pentru Agricultură (A.P.I.A.): pentru cetățenii români, străini și fără cetățenie, care sunt în evidența agenției: CNP-ul sau ISU, cetățenia, adresa de domiciliu, adresa reședinței obișnuite în România;

p) Ministerul Justiției, Administrația Națională a Penitenciarelor (A.N.P.): pentru datele privind persoanele condamnate, arestate preventiv sau internate, aflate în unitățile din sistemul administrației penitenciare;

q) Ministerul Lucrărilor Publice, Dezvoltării și Administrației, - Agenția Națională de Cadastru și Publicitate Imobiliară (A.N.C.P.I.): furnizarea limitelor administrativ-teritoriale stabilite conform Legii cadastrului și a publicității imobiliare nr. 7/1996, republicată, cu modificările și completările ulterioare, existente la 1 iulie 2020 și, respectiv, 1 ianuarie 2021 și datele disponibile din următoarele registre gestionate:

- Registrul Agricol Național – RAN

s) Firmele furnizoare de utilități, inclusiv cele de telefonie fixă sau mobilă: CNP sau ISU al titularului de contract, adresa de corespondență, adresa locuință, consumul lunar la adresa locuință.

Prin reuniunea înregistrărilor individuale unice la nivel de ISU din aceste surse administrative, se crează un registru statistic utilizat pentru a stabili, în urma aplicării unui algoritm pre-definit, persoanele care îndeplinesc criteriul de eligibilitate pentru populația rezidentă a României (acela ca persoana să fi avut reședința obișnuită pe teritoriul României cu cel puțin 12 luni înainte de momentul de referință al recensământului). Cu alte cuvinte, se determină eligibilitatea fiecărei înregistrări individuale prezentă în acestea, de a aparține sau nu populației rezidente a României la momentul de referință al recensământului populației și locuințelor runda 2021. Doar înregistrările eligibile pentru care nu s-au colectat date prin metodele pre-stabilite vor fi imputate total în baza de recensământ RPL2021.

Stabilirea apartenenței la populația rezidentă a României la momentul de referință al recensământului se bazează pe respectarea criteriului ca persoana respectivă să fi avut reședința obișnuită în România pentru cel puțin 12 luni înaintea lunii de referință.

Eligibilitatea de a aparține populației rezidente se determină prin metoda “**semnelor de viață**” care presupune luarea în considerare a cât mai multor informații existente în sursele

administrative disponibile pentru recensământ care, coroborate, pot proba că o anumită persoană a avut reședința obișnuită pe teritoriul României cu cel puțin 12 luni înainte de momentul de referință al recensământului.

Pentru minori, faptul că părinții cu care trăiesc în aceeași gospodărie aparțin populației rezidente poate fi considerat un criteriu suficient de puternic pentru includerea lor în populația rezidentă.

Algoritmul de determinare a apartenenței la populația rezidentă a României la momentul de referință a recensământului utilizează informația cuprinsă într-o serie de surse administrative. Sursele administrative de date sunt folosite “în cascadă”, una după alta. Coroborând informația din aceste surse la nivelul unei unități statistice (persoană) se determină, cu o probabilitate rezonabilă, dacă respectiva persoană face sau nu parte din populația rezidentă la recensământ. Dacă se trage concluzia este că aceasta face parte, se adaugă prin imputare totală o înregistrare individuală pentru acea persoană în baza de date inițială a recensământului.

Sursele de date care conțin informații utile recensământului, primite în baza art. 28 din OUG nr. 19/2020 aprobată și modificată prin Legea nr. 178/2020 privind organizarea și desfășurarea recensământului în România, sunt următoarele:

- fișierul de la DEPABD (**Evidența populației**)
- fișierul cu **declarațiile D112 de la ANAF**
- fișierul cu **declarațiile 70 de la ANAF**
- fișierul de la **Inspekția Muncii**
- fișierul de la **Casa Națională de Asigurări de Sănătate**
- fișierul de la **Casa Națională de Pensii Publice**
- fișierul cu date din Sistemul Informatic Integrat al Învățământului din România de la **Ministerul Educației și Cercetării (pre-universitar)**
- fișierul cu date din Registrul Matricol Unic din România, de la **Ministerul Educației și Cercetării (universitar)**
- fișierul de la **Agenția Națională de Plăți și Inspekție Socială**
- fișierul de la **Autoritatea pentru Drepturile Persoanelor cu Dizabilități, Copii și Adopții**

- fișierul de la **Agencia de Plăți în Agricultură**
- fișierul de la **Agentia Națională pentru Ocuparea Forței de Muncă**
- fișierul de la **Administrația Națională a Penitenciarelor**
- fișierul de la **Inspectoratul General al Poliției Române**
- fișierul de la **Inspectoratul General pentru Imigrări**
- fișierul de la **Institutul Național de Sănătate Publică**
- fișierul de la **Direcția Generală de Pașapoarte**
- fișierul de la **Agencia Națională pentru Cetățenie**

Criteriile pe baza cărora se determină apartenența la populația rezidentă, la nivelul fiecărei surse administrative disponibilă pentru recensământ, sunt descrise în Anexa 1.

4. Metode de imputare parțială

În practica statistică există diferite metode de imputare, fiecare având avantajele și dezavantajele ei și fiind pretabile unui anumit tip de variabilă.

În continuare trecem în revistă cele mai utilizate metode de imputare.

- **Imputare deterministică** – situația în care este posibilă doar o singură valoare corectă pentru o anumită variabilă, date fiind valorile specifice ale altor variabile din același set de date. Valoarea lipsă este astfel determinată din celelalte valori din același set de date.
- **Imputare statistică** – valoarea lipsă este determinată pe baza unui calcul sau altă metodă statistică.
- **Combinare/împerechere/ potrivire (matching) statistică** are scopul de a integra două sau mai multe surse de date referitoare la aceeași populație țintă în scopul obținerii unui set sintetic unic de date în care toate variabilele (ce provin din surse diferite) sunt disponibile împreună. Cuvântul sintetic se referă la faptul că setul de date este obținut din reuniunea informațiilor din diferite alte surse și nu din observarea statistică directă a unităților ce compun populația țintă. Împerecherea înregistrărilor din 2 surse diferite se realizează pe baza

unui identificator (același în toate sursele) sau pe baza unui set de variabile comune celor 2 surse, prezente în ambele surse. Setul sintetic unic de date obținut conține variabilele comune, dar și variabile care există doar într-una dintre surse.

- **Imputare bazată pe modelare** – valoarea lipsă a unei variabile este obținută prin utilizarea de medii, mediane sau ecuații de regresie etc.
- **Imputarea de la donor** – valorile lipsă ale unei variabile se obțin de la o înregistrare donor, un „geamăn” cu mai multe caracteristici similare cu cele ale unității statistice pentru care trebuie făcută imputarea.

Literatura de specialitate face distincția între imputarea Hot deck – donorul este în același set de date și imputarea Cold deck – donorul se află în alt set similar de date (sau un set precedent de date ale aceleiași cercetări statistice).

5. Modalități de tratare a non-răspunsurilor parțiale

Non-răspunsurile parțiale reprezintă o problemă practică la fel de severă ca și non-răspunsurile totale și necesită soluții de tratare. Acestea constau în lipsa răspunsului la o anumită întrebare/variabilă, în condițiile în care o unitate selectată acceptă interviul în cadrul unei cercetări statistice, răspunde la o parte din întrebări, însă, la un anumit moment, din diferite motive, refuză să răspundă la o întrebare sau la mai multe (însă nu la toate) din setul complet de întrebări.

De multe ori, cauzele non-răspunsurilor parțiale sunt diferite de cele ale non-răspunsurilor totale. Dacă acestea din urmă apar dintr-o decizie bazată pe informații foarte succinte și vagi cu privire la scopul cercetării statistice, non-răspunsurile parțiale apar după o informare exactă și o judecată completă a uneia sau alteia dintre întrebările cercetării. Astfel, se disting trei cauze majore care pot genera non-răspunsul parțial, și anume:

- interpretarea greșită a scopului unei întrebări,
- eșecul de a găsi răspunsul adecvat la o întrebare,
- lipsa de dorință sau de motivație de a oferi răspunsul/informația solicitată de întrebare.

Evident că aici intervine forma în care a fost formulată întrebarea, cât de ușor sau greu de înțeles este aceasta, la câte amănunte trebuie respondentul să se gândească pentru a răspunde la întrebare, cât de mare este efortul pe care îl solicităm respondentului să îl depună în găsirea răspunsului etc. În fine, nu este de neglijat nici forma în care solicităm răspunsul (de exemplu, folosirea unor intervale de timp în locul valorilor exacte pentru variabila anul construcției locuinței sau a întrebărilor deschise în locul răspunsurilor pretipărite pot avea drept efect reducerea numărului de cazuri de non-răspuns parțial).

Modalitatea de tratare a non-răspunsurilor parțiale o reprezintă **imputarea parțială**. Aceasta presupune substituția datelor lipsă cu valori artificiale de bună calitate, valoarea imputată fiind o predicție a valorii adevărate necunoscută. De multe ori se folosesc informații auxiliare pentru a determina valoarea ce trebuie imputată.

Câteva dintre cele mai folosite metode de imputare a valorilor lipsă sunt următoarele:

- *imputarea mediei* – pentru orice valoare numerică lipsă a unei variabile din setul de înregistrări individuale aparținând populației țintă se impută valoarea medie a variabilei determinată pe baza valorilor individuale ale unităților care au răspuns la respectiva întrebare/variabilă.
- *imputarea mediei claselor* – setul de înregistrări individuale pentru care s-a răspuns la o anumită întrebare/variabilă este partiționat în clase după anumite criterii astfel încât elementele din fiecare clasă să fie considerate similare. Datele lipsă dintr-o anumită clasă sunt imputate cu media calculată la nivelul respondenților care au răspuns din clasa respectivă.
- *imputarea hot-deck și cold-deck* – se poate îmbunătăți metoda imputării mediei prin introducerea unei variabilități mai autentice între valorile imputate. Astfel, dacă în imputarea hot-deck valorile lipsă sunt imputate de valori selectate de la respondenții din respectiva cercetare (de exemplu, de la cel mai apropiat donor sau de la un donor ales aleator din setul de respondenți pentru o anumită variabilă), în imputarea cold-deck valorile imputate provin din alte surse, de exemplu din cercetări precedente sau din date panel.

Donorul este respondentul care are un set de caracteristici asemănătoare cu cel pentru care se dorește imputarea unei valori lipsă. Cu cât sunt definite mai multe caracteristici, cu atât procesul de căutare este mai complicat și șansa de găsire a unui donor este mai mică, însă crește probabilitatea ca valoarea imputată să fie mai apropiată de cea reală. Căutarea donorului se desfășoară începând din "imediata vecinătate" în cadrul setului de respondenți. În cazul recensământului, căutarea donorului se face în cadrul sectorului de recensământ din care face parte unitatea statistică care are date lipsă, iar în caz că nu este găsit, căutarea continuă în cadrul localității, județului, mediului de rezidență sau chiar la nivelul întregului set de respondenți.

- *imputarea multiplă* – în această procedură, un anumit număr de seturi din toate valorile lipsă se selectează din setul de date. Pentru fiecare dată lipsă se calculează mai multe valori (de exemplu m valori). Aceste m valori sunt ordonate astfel încât primul set de date pentru care s-au imputat date lipsă este folosit pentru formarea primului set complet de date și așa mai departe. Din cele m estimări se calculează o singură valoare estimată, împreună cu o valoare estimată a varianței. În acest mod varianța estimată conține o componentă care reflectă variabilitatea dintre seturile de date.
- *imputarea cu ajutorul regresiei* – folosește relațiile estimate dintre variabile.

La Recensământul populației și locuințelor runda 2021 se vor aplica următoarele metode de imputare parțială:

- **Imputare deterministă** – pentru variabile cum ar fi existența și localizarea bucătăriei în cazul locuințelor situate în blocuri de apartamente; sex, data nașterii, starea civilă legală pentru persoanele cu vârsta sub 16 ani, numărul copiilor născuți vii pentru persoanele de sex feminin cu vârsta sub 12 ani, folosirea internetului pentru persoanele cu vârsta cuprinsă între 15 - 29 ani, nivelul de educație urmat la momentul de referință al recensământului pentru persoanele de 50 ani și peste etc.
- **Imputare bazată pe modelare** - pentru variabilele numerice de genul numărul și suprafața camerelor de locuit etc.
- **Imputarea de la donor** – pentru variabilele calitative cum ar fi combustibilul folosit pentru gătit, instalație de alimentare cu apă caldă, instalație de canalizare, modul de încălzire a

locuinței în cazul caselor individuale; data și motivul stabilirii în localitatea de recensare, statutul activității curente, ocupația, activitatea principală a unității, statutul profesional etc.

- **Combinare statistică** – pentru următoarele variabile se preiau valorile existente în setul de date al recensământului din 2011, pe baza prezumției că nu s-au modificat până la momentul de referință al recensământului din runda 2021: caracteristicile locuințelor situate în blocuri de apartamente (număr de camere și suprafața acestora, existența băii, bucătăriei, instalației de alimentare cu apă caldă, instalației de canalizare, anul de construcție și materialul de construcție al pereților exteriori ai clădirii), anul de construcție și materialul de construcție al pereților exteriori ai clădirii pentru casele individuale; nivelul de educație absolvit pentru persoanele de 50 ani și peste, numărul de copii născuți-vii de către persoanele de sex feminin cu vârsta peste 50 ani etc..

Pentru persoanele pentru care se găsesc informații în sursele administrative, valorile lipsă ale variabilelor de recensământ se preiau din sursele administrative (județul, localitatea și localitatea componentă de domiciliu, cetățenia, țara, județul, localitatea și localitatea componentă de naștere, ocupația persoanelor ocupate care nu sunt funcționari publici, timpul de lucru, activitatea principală a locului de muncă, județul, localitatea și localitatea componentă în care persoana a avut reședința anterioară, nivelul de educație urmat la momentul de referință al recensământului, etc.).

Pentru unele variabile se vor folosi mai multe metode de imputare combinate, în funcție de situația identificată pentru fiecare persoană în parte. Mai exact, pentru acele persoane care sunt identificate ca prezente în sursele administrative se aplică *combinarea statistică* (preluarea valorii variabilei dintr-o anumită sursă administrativă), iar dacă acest lucru nu este posibil, se aplică apoi *imputarea deterministă* sau *imputarea de la donor*.

Datorită caracterului special al unor variabile sau a inexistenței informațiilor în sursele administrative utilizate pentru colectarea indirectă a datelor, **vor rămâne neimputate** următoarele variabile:

1. În secțiunea **CHESTIONAR PENTRU RECENZAREA PERSOANEI**

- numărul de copii născuți-vii de o femeie de vârstă fertilă (15-49 ani)
- etnia

- limba maternă
- religia
- a doua cetățenie.

2. În secțiunea CHESTIONAR PENTRU RECENZAREA LOCUINȚEI

- forma de proprietate a locuinței
- numărul de camere folosite pentru scopuri profesionale

Bibliografie

***- *Statistical matching: a model based approach for data integration*, Eurostat Methodologies and working papers, editia 2013, ISBN 1977-0375

***- *Glossary of Terms on Statistical Data Editing imputation occurs through a human review process* , Conference of European Statisticians Methodological Material, United Nations 2000
https://ec.europa.eu/eurostat/ramon/statmanuals/files/UN_editing_glossary_2000.pdf

Marcello D’Orazio - *Statistical Matching and Imputation of Survey Data with StatMatch*, ESSnet project on DataIntegration

S.F.Buck – *A method of estimation of missing values in multivariate data suitable for use with an electronic computer*, Journal of Royal Statistical Society B 22

Criteriile pe baza cărora se determină apartenența la populația rezidentă pe baza datelor disponibile din sursele administrative

Nr. crt.	Sursa administrativă din care se preiau informațiile utile RPL2021	Criteriul îndeplinit în perioada de referință*
1.	Direcția pentru Evidența Persoanelor și Administrarea Bazelor de Date	Existența domiciliului/ reședinței pe teritoriul României
2.	ANAF - declarațiile D112	Dovada obligațiilor de plată a contribuțiilor sociale, impozitului pe venit și evidența nominală a persoanelor asigurate pentru activități economice desfășurate
3.	ANAF - declarațiile D70	Dovada declarației fiscale a persoanelor fizice care desfășoară activități economice în mod independent sau exercită profesii libere
4.	Inspecția Muncii	Existența unui contract de muncă activ (nesuspendat sau încetat)
5.	Casa Națională de Pensii Publice	Primirea pensiei de asigurări sociale de stat prin mandat poștal sau într-un cont bancar pe teritoriul țării timp de cel puțin 12 luni
6.	Casa Națională de Sănătate	Dovada existenței unor evenimente medicale (fără precizarea acestora)
7.	Ministerul Educației și Cercetării (pre-universitar) – Sistemul Informatic Integrat al Învățământului din România	Dovada înscrierii unei persoane într-o formă de învățământ pre-universitar în România
8.	Ministerul Educației și Cercetării (universitar) - Registrul Matricol Unic din România	Dovada înscrierii unei persoane într-o formă de învățământ universitar în România
9.	Agenția Națională de Plăți și Inspecție Socială	Dovada plății/ încasării unor beneficii de asistență socială
10.	Autoritatea pentru Drepturile Persoanelor cu Dizabilități, Copii și Adopții	Locuirea în cadrul unor unități ale sistemului de protecție a copilului (servicii de tip rezidențial și servicii de tip familial) de pe teritoriul României
11.	Agenția de Plăți în Agricultură	Dovada plății unor subvenții către persoane fizice pentru desfășurarea unor activități în agricultură

Nr. crt.	Sursa administrativă din care se preiau informațiile utile RPL2021	Criteriul îndeplinit în perioada de referință*
12.	Agenția Națională pentru Ocuparea Forței de Muncă	Dovada că persoana s-a aflat în evidența ANOFM pentru furnizarea de servicii specifice persoanelor în căutarea unui loc de muncă
13.	Administrația Națională a Penitenciarelor	Dovada că persoana a locuit într-un penitenciar de pe teritoriul României
14.	Inspectoratul General al Poliției Române	Dovada că persoana s-a aflat în arestul unei unități aparținând IGPR pe teritoriul României
15.	Institutul Național de Sănătate Publică	Dovada că persoana se află în evidență privind bolile profesionale
16.	Inspectoratul General pentru Imigrări	Dovada existenței unui act care atestă dreptul de ședere în România pentru cetățenii străini și fără cetățenie
17.	Direcția Generală de Pașapoarte	Dovada existenței domiciliului în străinătate pentru cetățeni români
18.	Agenția Națională pentru Cetățenie	Dovada dobândirii cetățeniei române

*Perioada de referință reprezintă ultimele 12 luni precedente momentului de referință a recensământului, respectiv perioada 1 decembrie 2020 - 30 noiembrie 2021.